

IntechOpen

Digital Video

Edited by Floriano De Rango



DIGITAL VIDEO

EDITED BY
FLORIANO DE RANGO

Digital Video

<http://dx.doi.org/10.5772/173>

Edited by Floriano De Rango

© The Editor(s) and the Author(s) 2010

The moral rights of the and the author(s) have been asserted.

All rights to the book as a whole are reserved by INTECH. The book as a whole (compilation) cannot be reproduced, distributed or used for commercial or non-commercial purposes without INTECH's written permission.

Enquiries concerning the use of the book should be directed to INTECH rights and permissions department (permissions@intechopen.com).

Violations are liable to prosecution under the governing Copyright Law.



Individual chapters of this publication are distributed under the terms of the Creative Commons Attribution 3.0 Unported License which permits commercial use, distribution and reproduction of the individual chapters, provided the original author(s) and source publication are appropriately acknowledged. If so indicated, certain images may not be included under the Creative Commons license. In such cases users will need to obtain permission from the license holder to reproduce the material. More details and guidelines concerning content reuse and adaptation can be found at <http://www.intechopen.com/copyright-policy.html>.

Notice

Statements and opinions expressed in the chapters are those of the individual contributors and not necessarily those of the editors or publisher. No responsibility is accepted for the accuracy of information contained in the published chapters. The publisher assumes no responsibility for any damage or injury to persons or property arising out of the use of any materials, instructions, methods or ideas contained in the book.

First published in Croatia, 2010 by INTECH d.o.o.

eBook (PDF) Published by IN TECH d.o.o.

Place and year of publication of eBook (PDF): Rijeka, 2019.

IntechOpen is the global imprint of IN TECH d.o.o.

Printed in Croatia

Legal deposit, Croatia: National and University Library in Zagreb

Additional hard and PDF copies can be obtained from orders@intechopen.com

Digital Video

Edited by Floriano De Rango

p. cm.

ISBN 978-953-7619-70-1

eBook (PDF) ISBN 978-953-51-5489-1

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,200+

Open access books available

116,000+

International authors and editors

125M+

Downloads

151

Countries delivered to

Our authors are among the
Top 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Meet the editor

Floriano De Rango is Associate Prof. in Telecommunication and Networking at DIMES Dept., University of Calabria, ITALY. In October 2000 he got his Master Degree Thesis cum laude in Electrical and Computer Science Engineering at University of Calabria. During his thesis he spent nine months at C.S.E.L.T (Centro Studi e Laboratorio Telecomunicazioni) in Turin where he worked on Access Network of Telecom Operator gaining experience in the management of Network Access Server (NAS) and on the interworking and integration of multi-vendor and multiplatform infrastructures. During this period he worked also on security issues and on Authentication, Authorization and Accounting (AAA) architecture in order to offer differentiation and security at access network in Dial-up phase. His Master Degree thesis was titled "Services Differentiation at IP level in Access Network through network management policies. Reliability, functionality and interoperability analysis on multi-vendor platforms". In 2001 he started his Ph.D in Systems and Information Engineering at DEIS Dept., University of Calabria. In March 2004 he went to Los Angeles at UCLA University where he spent 8 months at Network Research Lab. In February 2005 he got his Ph.D with a thesis titled "Hybrid Location Based Routing Protocols for Scalable Wireless Ad Hoc Networks with Group Motion". From November 2004 to September 2007 he worked at DEIS Dept. as Research Fellow. He has been also involved in many professional activities and in particular he has been TPC member of many International Conferences such as IEEE VTC, ICC, WCNC, Globecom, Med Hoc Net, SPECTS, WirelessCOM, WinSys and reviewer in many journals such as IEEE Communication Letters, JSAC, IEEE Trans.on Vehicular Technology, Computer Communication, Eurasip JWCN, WINET etc. In September 2007 he become Assistant Professor at DEIS Dept., University of Calabria, where he teaches Telecommunication Systems, Designing of Advance Cellular Networks and Radiomobile Networks II. He is now Associate Prof. in Telecommunication and Networking at DIMES Dept., University of Calabria. He was recipient of Young Researcher Award in 2007 for a Project on Vehicular Ad Hoc Networks (VANETs) called ATENA (Autoconfiguring intelligent vehicular Network for urban Area). He co-authored till now more than 200 papers in International Conferences and Journals.

Preface

Digital video broadcasting (DVB) interest aroused in recent years and the birth of DVB projects some years ago led to the development of many new technologies and architecture able to deliver video and multimedia traffic over heterogeneous platforms. In the first phase, the DVB projects focused on the development of technical specifications relevant for the more traditional broadcasting of audio and video services by satellite, in cable networks and via terrestrial transmitters. Later and more recently attention has focused on the IP paradigm over DVB networks, solutions for interaction and return channels, the software environment called Multimedia Home Platform (MHP) and the distribution of video and audio over handheld devices and smartphones. This book tries to address different aspects and issues related to video and multimedia distribution over the heterogeneous environment considering broadband satellite networks and general wireless systems where wireless communications and conditions can pose serious problems to the efficient and reliable delivery of contents. In this context, specific chapters of the book have been related to different research topics covering the architectural aspects of the most famous DVB standard (DVB-T, DVB-S/S2, DVB-H etc.), the protocol aspects and the transmission techniques making use of MIMO, hierarchical modulation and lossy compression. In addition to these topics, also research issues related to the application layer and to the content semantic, organization and research on the web have also been addressed in order to give a complete view of the problems. The network technology faced in this book are mainly broadband wireless and satellite networks and the book can be read by intermediate students, researchers, engineers or people with some knowledge or specialization in network topics.

All chapters have been written individually by different authors and this allows for each chapter to be read independently from the other chapters of the book. We decided to adopt this approach to permit researchers, students or engineers to select the arguments of major interest and that fall in the expertise area of any reader that is interested in the video distribution topic with the possibility to understand the approaches proposed by the authors without being forced to read the entire book. Obviously, in order to go deeply into the video distribution and multimedia transmission topic over broadband networks such as Satellite, overall reading of the book is suggested.

In the following, brief snapshots of the single chapters are presented in order to give the reader the opportunity to select the chapters that better fall in the area of interest.

Chapter 1: Multimedia Traffic over Wireless and Satellite Networks

The rapid growth of multimedia applications and the development of advanced digital mobile terminals able to connect to multiple network segments lead to the proposal and

design of novel network architectures and protocols where video distribution, IPTV, audio transmission and multimedia traffic in general can be supported. Multiple technologies such as satellite DVB, 3G networks, wireless systems and so on can converge in an overall framework and architectures where fixed and mobile users can transparently communicate maintaining a good quality of service (QoS).

Chapter 2: Adaptive Video Transmission over Wireless MIMO System

Multimedia transmission and, in particular, video transmission that needs variable bit rate support can be obtained through advanced transmission techniques and technologies such as MIMO systems and source channel coding that permits an increase of channel capacity and diversity. In this context, adaptive video transmission and scalable video coding become an important issue to be addressed in order to allow video transmission over wireless networks.

Chapter 3: Transmission Optimization of Digital Compressed Video in Wireless Systems

The transmission of multimedia streams over wireless networks determines new issues to be addressed considering the variable nature of wireless channels and the possibility of user movements. New compression techniques able to reduce the network requirements or data rate to be supported consents the maintaining of acceptable QoS requirements also over wireless channels. Thus, compressed multimedia transmission together with dynamic optimized bandwidth assignment strategies and rate based and channel condition based scheduling algorithms can become essential to support and respect QoS constraints of multimedia traffic.

Chapter 4: Resilient Digital Video Transmission over Wireless Channels using Pixel-Level Artefact Detection Mechanisms

In order to support a resilient digital video transmission over noisy wireless channels advanced video coding techniques and machine learning algorithms to offer robust signal detection are suitable. Also Support Vector Machines (SVM) techniques can increase the capability of the decoder in detecting visual distorted regions increasing the QoS in the video transmission and its robustness.

Chapter 5: Digital Video Broadcasting via Satellite – Challenge on IPTV Distribution

Recent satellite technologies and in particular Digital Video Broadcasting via Satellite (DVB-S) penetrated the market in recent years and it has allowed the delivery of high quality digital video. With this technology it can be interesting to see how the new IPTV distribution paradigm can be mapped over Satellite architecture in order to obtain a fully integrated IP network where video distribution can be delivered to the end users.

Chapter 6: The Deployment of Intelligent Transport Services by using DVB-based Mobile Video Technologies

Mobile video technologies are becoming so popular and useful that a possible use for new Intelligent Transport Services (ITS) is creating interest. The capability to warn drivers in time to avoid collision or the detection of the exact accident location can offer useful information for road traffic management and to reduce further incident risks. Novel standards such as Dedicated Short Range Communication (DSRC), GPRS, WiMAX and UMTS networks and broadcast technologies such as RDS, DAB and DVB specifications (DVB-T, DVB-S, DVB-H etc) can be integrated in order to offer a ubiquitous ITS service. In this context, DVBH and DVB-SH are examined to focus on the support of novel services on handheld devices and on the possibility of integrating these broadcast technologies with heterogeneous systems to realize a general ITS framework.

Chapter 7: A New Waveform based on Linear Prediction Multicarrier Modulation for Future Digital Video Broadcasting Systems

The OFDM technique represents an efficient scheme of transmission for propagation over multiple channels and, for this reason, it has been adopted in different standards relative to Digital Video Broadcasting (DVB). The success of this technique is in the capacity of subdividing the single channel into parallel free sub-channels and then to rebuild the signal using the reverse Fast Fourier Transform (FFT) at the transmitter and FFT at the receiver. In order to reduce the overhead associated with the standard OFDM technique a linear pre-coding based on OFDM modulation is considered in this chapter.

Chapter 8: Performance Analysis of DVB-T/H OFDM, Hierarchical Modulation in Impulse Noise Environment

This chapter regards DVB-T/H, introducing basic DVB characteristics and modulation techniques. The hierarchical modulation and the COFDM proposed in different digital standard proposals are considered to learn about the capability to modulate in a single bit-stream separate and multiple bit-streams. Moreover, the overall system performance in the presence of impulse noise, and not only Gaussian noise, is also evaluated.

Chapter 9: IP Datacast (IPDC) over DVB-H and the Repair Costs of the IPDC File Repair Mechanism

This chapter briefly introduces the DVB technology and the transmission systems adopted by mobile terminals (DVB-Handhead). Then, the IP datacast architecture (IPDC) and the protocols used in the DVB-H are also applied focusing mainly on data transmission and error signaling mechanisms in the reception phase.

Chapter 10: DVB-T2: New Signal Processing Algorithm for a Challenging Digital Video Broadcasting Standard

The novel standard DVB-T2 is introduced focusing also on the difference between DVB-T and DVB-T2 and the new possibility to increase the transmission data rate.. De-mapping, iterative decoding, and antenna diversity etc. are the main characteristics of DVB-T2 able to increase its performance in comparison with DVB-T.

Chapter 11: Passive Radar using COFDM (DB or DVB-T) Broadcaster as Opportunistic Illuminator

In this chapter the advantages and benefits of COFDM in broadcast transmission are introduced. At the beginning the main concepts of COFDM are provided and later two stop filters used for the COFDM signal transmission are explained. Many of the results of this chapter have been experimentally obtained.

Chapter 12: Reliable and Repeatable Power Measurement in DVB-T System

In this chapter, after briefly introducing the DVB technique, measurement approaches and methodologies to apply on this broadcast technology are explained such as bandwidth and power measurements. The specific instrumentations and techniques such as specified by the European Telecommunication Standards Institute (ETSI) are described.

Chapter 13: MidField: An Adaptive Middleware System for Multipoint Digital Video Communication

In order to support digital video transmission through multiple intermediate nodes, it is necessary to design and implement a novel middleware level called MidField. In particular, in this chapter the routing of multiple streams among intermediate nodes for the delivery of Digital Video (DV) and High Definition Video (HDV) are introduced and described. Architectures, protocols and the working of MidField, communication modalities and streaming modules are also described.

Chapter 14: Video Content Description using Fuzzy Spatio-Temporal Relation

Sometimes, it can be interesting to find video content on the basis of some description of the video track. With this regard, to obtain the task of an efficient research of video contents novel techniques based on Fuzzy logic and the spatio-temporal relation is explained and proposed.

Chapter 15: Trick play on Audiovisual Information for Tape Disk and Solid - State based Digital Recording System

This chapter is dedicated to the storage techniques to memorize video content in an efficient manner in order to speed up the video content research. A solution called Digital Consumer Storage Standard is described and in addition storage and compression techniques for different physical supports are also provided.

Chapter 16: Video Quality Metrics

Compression and transmission techniques are two basic aspects in video transmission. However, in order to know how much a video content can be compressed offering good quality perception to the end user, it is very important to define good video quality metrics. It is possible to define subjective and objective video metrics that are able to correlate the network characteristics with human perception. In this chapter a brief description of the Human Visual System and the most important video quality metrics will be provided.

Chapter 17: Video Analysis and Indexing

In recent years, the increasing number of applications and traffic based on video content are leading to a particular interest in the design of (semi) automatic ways to describe, organize, and manage video data with greater understanding of its semantic contents. With this regard, the design of efficient databases, storing techniques, indexing, semantic extraction by video content and time-efficient query can become essential in video management and they will be object of this chapter.

Chapter 18: Video Editing based on Situation awareness from Voice Information and Face Emotion

Since video camera systems are becoming ever more popular in the home environment and many other places, it is important to record different types of events during the course of daily life. In this context, in order to automatize the recording process and to capture significant events, it is possible to design recording strategies that make use of voice detection and face recognition to capture and zoom situations of particular interest. In this chapter, the authors, after describing a video editing system based on audio and face emotion, describe the specific techniques adopted in their system concerning voice and voice direction detection and facial emotion recognition algorithms.

Chapter 19: Combined Source and Channel Strategies for Optimized Video Communications

In the context of Universal Media Access (UMA), one of the main challenges is to flexibly deliver video content with the best perceived image quality for end-users having different available resources, access technologies and terminal capabilities. This chapter will introduce different coding and compression techniques managed on the basis of the channel condition in order to improve efficiency in video and multimedia distribution. The first part of the chapter will describe the MPEG-2 and H.264/AVC standard with subjective and objective quality metrics. In the second part of the chapter, instead, channel coding and error control techniques in video distribution, scalable video transmission and hierarchical modulation are discussed.

Chapter 20: Amplitude Phase Shift Keying Constellation Design and Its Applications to Satellite Digital Video Broadcasting

This chapter focuses on the Amplitude Phase Shift Keying (APSK) and M-APSK applied to Digital Video Broadcasting and in particular to DVB-S2 and DVB-SH. This modulation technique and its extension to multidimensional modulation (M-APSK) present higher spectral efficiency and a higher data rate providing good support for multimedia and video distribution.

Chapter 21: Non-Photo Realistic Rendering for Digital Video Intaglio

In recent years, rendering algorithms have been introduced to mimic various classic art forms, ranging from pen-and-ink illustrations and line art drawings to expressive paintings. In this context, non photo realistic (NPR) rendering techniques can become an important instrument to consider and this chapter will give an idea of NPR techniques proposed in literature, focusing later on an innovative approach proposed by the authors.

Chapter 22: Building Principles and Application of Multifunctional Video Information System

In this chapter structural principles of perspective television system, video informational system (VIS) for multipurpose application or multifunctional video informational system (MFVIS) are described.

Editor

Floriano De Rango

Associate Prof.

*at DEIS Dept., University of Calabria,
Italy*

Contents

| | |
|---|-----|
| Preface | IX |
| 1. Multimedia Traffic over Wireless and Satellite Networks <i>Floriano De Rango, Mauro Tropea and Peppino Fazio</i> | 001 |
| 2. Adaptive Video Transmission over Wireless MIMO System <i>Jia-Chyi Wu, Chi-Min Li, and Kuo-Hsean Chen</i> | 031 |
| 3. Transmission Optimization of Digital Compressed Video in Wireless Systems <i>Pietro Camarda and Domenico Striccoli</i> | 045 |
| 4. Resilient Digital Video Transmission over Wireless Channels using Pixel-Level Artefact Detection Mechanisms <i>Reuben A. Farrugia and Carl James Debono</i> | 071 |
| 5. Digital Videos Broadcasting via Satellite – Challenge on IPTV Distribution <i>I Made Murwantara, Pujianto Yugopuspito, Arnold Aribowo and Samuel Lukas</i> | 097 |
| 6. The Deployment of Intelligent Transport Services by using DVB-Based Mobile Video Technologies <i>Vandenbergh, Leroux, De Turck, Moerman and Demeester</i> | 111 |
| 7. A New Waveform based on Linear Precoded Multicarrier Modulation for Future Digital Video Broadcasting Systems <i>Oudomsack Pierre Pasquero, Matthieu Crussi re, Youssef Nasser, Eddy Cholet and Jean-Fran ois H lard</i> | 125 |
| 8. Performance Analysis of DVB-T/H OFDM Hierarchical Modulation in Impulse Noise Environment <i>Tamgnoue Val ry, V ronique Moeyaert, S bastien Bette and Patrice M gret</i> | 145 |

| | |
|--|-----|
| 9. IP Datacast and the Cost Effectiveness of its File Repair Mechanism <i>Bernhard Hechenleitner</i> | 163 |
| 10. DVB-T2: New Signal Processing Algorithms for a Challenging Digital Video Broadcasting Standard <i>Mikel Mendicute, Iker Sobrón, Lorena Martínez and Pello Ochandiano</i> | 185 |
| 11. Passive Radar using COFDM (DAB or DVB-T) Broadcasters as Opportunistic Illuminators <i>Poullin Dominique</i> | 207 |
| 12. Reliable and Repeatable Power Measurements in DVB-T Systems <i>Leopoldo Angrisani, Domenico Capriglione, Luigi Ferrigno and Gianfranco Miele</i> | 229 |
| 13. MidField: An Adaptive Middleware System for Multipoint Digital Video Communication <i>Koji Hashimoto and Yoshitaka Shibata</i> | 263 |
| 14. Video Content Description Using Fuzzy Spatio-Temporal Relations <i>Archana M. Rajurkar, R.C. Joshi, Santanu Chaudhary and Ramchandra Manthalkar</i> | 285 |
| 15. Trick play on Audiovisual Information for Tape, Disk and Solid-State based Digital Recording Systems <i>O. Eerenberg and P.H.N. de With</i> | 303 |
| 16. Video Quality Metrics <i>Mylène C. Q. Farias</i> | 329 |
| 17. Video Analysis and Indexing <i>Hui Ding, Wei Pan and Yong Guan</i> | 359 |
| 18. Video Editing Based on Situation Awareness from Voice Information and Face Emotion <i>Tetsuya Takiguchi, Jun Adachi and Yasuo Arika</i> | 381 |
| 19. Combined Source and Channel Strategies for Optimized Video Communications <i>François-Xavier Coudoux, Patrick Corlay, Marie Zwingelstein-Colin, Mohamed Gharbi, Charlène Mouton-Goudemand, and Marc-Georges Gazelet</i> | 397 |

20. Amplitude Phase Shift Keying Constellation Design and its Applications to Satellite Digital Video Broadcasting 425
Konstantinos P. Liolis, Riccardo De Gaudenzi, Nader Alagha, Alfonso Martinez, and Albert Guillén i Fàbregas
21. Non-Photo Realistic Rendering for Digital Video Intaglio 453
Kil-Sang Yoo, Jae-Joon Cho and Ok-Hue Cho
22. Building Principles and Application of Multifunctional Video Information System 463
Mahmudov Ergash B. and Fedosov Andrew A.
23. Digital Video Image Quality 487
Tomáš Kratochvíl and Martin Slanina

Multimedia Traffic over Wireless and Satellite Networks

Florianio De Rango, Mauro Tropea and Peppino Fazio
*DEIS Dept., University of Calabria,
Italy*

1. Introduction

The rapid growth of multimedia application together with advanced development in digital technology and with the increased use of mobile terminal has pushed the research toward new network technologies and standards for wireless environment. Moreover, ever-increasing computing power, memory, and high-end graphic functionalities have accelerated the development of new and exciting wireless services. Personal video recorders, video on demand, multiplication of program offerings, interactivity, mobile telephony, and media streaming have enabled viewers to personalize the content they want to watch and express their preferences to broadcasters. Viewers can now watch television at home or in a vehicle during transit using various kinds of handheld terminals, including mobile phones, laptops computers, and in-car devices. The concept of providing television-like services on a handheld device has generated much enthusiasm. Mobile telecom operators are already providing video-streaming services using their third-generation cellular networks. Simultaneous delivery of large amounts of consumer multimedia content to vast numbers of wireless devices is technically feasible over today's existing networks, such as third-generation (3G) networks. As conventional analog television services end, broadcasters will exploit the capacity and flexibility offered by digital systems. Broadcasters will provide quality improvements, such as high-definition television (HDTV), which offer many more interactive features and permit robust reception to receivers on the move in vehicles and portable handhelds. Mobile TV systems deliver a rich variety of content choice to consumers while efficiently utilizing spectrum as well as effectively managing capital and operating expenses for the service provider. Mobile TV standards support efficient and economical distribution of the same multimedia content to millions of wireless subscribers simultaneously. Mobile TV standards reduce the cost of delivering multimedia content and enhance the user experience, allowing consumers to surf channels of content on a mobile receiver. Mobile TV standards address key challenges involved in the wireless delivery of multimedia content to mass consumers and offer better performance for mobility and spectral efficiency with minimal power consumption. An important aspect of multimedia delivery contest is the possibility of make integration between different networks in order to be able of reaching users every-time every-where. Then, it is possible to use a multi-layer hierarchic platform that use satellite [1] segment together with wireless network based on 802.11 standard [2,3,4] and with cellular network in order to have an ubiquitous coverage. In

order to grapple with the continuously increasing demand on multimedia traffic over high speed wireless broadband networks it is also necessary to make network able to deal with the QoS constraints required by users. In order to provide quality of service to user applications, networks need optimal and optimized scheduling and connection admission control algorithm. These mechanisms help to manage multimedia traffic guaranteeing QoS to calls already admitted in the system and providing QoS to the new connection. In order to evaluate the quality of video traffic with the mobility it is important to examining the quality assessment techniques: Subjective and Objective quality assessment.

2. Video quality in multimedia broadcasting

In the last few years multimedia applications are grew very fast in all networks typologies and in particular in wireless networks that are acquiring a big market slice in the telecommunication field. This trend of applications has pushed the researchers to perform a lot of studies in the video applications and in particular in the compression field in order to be able of transporting this information in the network with a low impact in the system resources. In literature a lot of studies exist on video compression and new standard of compression are proposed in order to be able to transmit video traffic on different network technologies that often have resource problem in terms of bandwidth capacity and, then, a very performance compression algorithm can give a great support to network service provider in respecting the quality constrains otherwise nor achievable.

The ubiquitous nature of all multimedia services and their use in overall telecommunications networks requires the integration of a lot of technologies that aim to improve the quality of the applications received by the users. First of all, it is clear that the traditional concept of best effort paradigm in the delivery of multimedia contents is not possible to adopt because it does not match with the users requirements. This type of approach try to do its best but it is unable of guaranteeing any form of users requirements. In order to address this type of problem, recently different Quality of Services architectures have been proposed capable of guaranteeing to the multimedia streams the users constrains. The most famous architectures are Integrated Services and Differentiated Services that manage different class of services in order to allow a traffic differentiation able to discriminate also a cost differentiation for the customers. It is easy to understand that this new type of applications is based on a constantly reliable reception of information that it is possible to have only through an appropriate network management.

The new end users are always more quality aware and then they are more exigent and accordingly networks have to guarantee always more capacity in order to satisfy their users. As a consequence, there is a continuous and extensive research effort, by both industry and academia, to find solutions for improving the quality of multimedia content delivered to the users; as well, international standards bodies, such as the International Telecommunication Union (ITU), are renewing their effort on the standardization of multimedia technologies. There are very different directions in which research has attempted to find solutions in order to improve the quality of the rich media content delivered over various network types [5,6,7,8,9].

Moreover, it is very important to determine efficient quality assessment. It is really important to know how the network behaves in terms of parameters of service for end users in order to take the correct decisions during the development, evaluation, construction and operation of network services.

Over the years have developed many models, more efficient and less expensive, able to consider more types of factors involved in the telecommunication networks. An overview of models of measurement is reported in the Figure 1 below. They are divided substantially in two categories: subjective assessment and objective assessment.

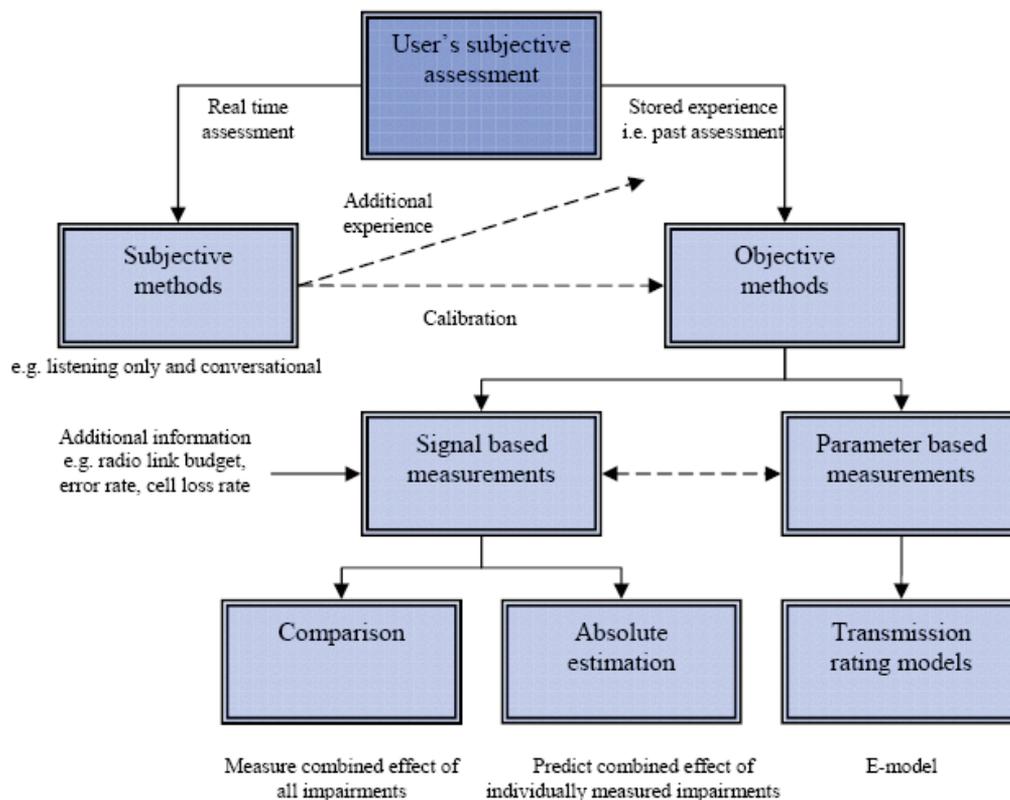


Fig. 1. Overview of models of measurement

2.1 Subjective quality assessment

A method to perform quality evaluation for multimedia applications is called subjective assessment and it represents an accurate technique for obtaining quality ratings. It is basically based on experience done by a number of users. Typically, the experiments are made in a room where there are a certain number of persons (about 15-30) that they have to watch a set of video streams and, then, they have to give a rate based on own quality perception. On the basis of all rates given by all subjects included in the experiment it is formulated an average rate called Mean Opinion Score (MOS). It is clear that, being a subjective evaluation determined by the subjectivity and variability of the involved persons, this test is affected by the personal opinion that cannot be eliminated. In order to avoid this type of problem the experiments are made through precise instructions that give to the subject a set of precautions that they have to follow. Moreover, also the environment used for the test is a controlled environment. In this way it is possible to perform a set of tests and provide a quality score that is a number that results from a statistical distribution. There are

a wide variety of subjective testing methods. In literature different methods exist for giving a measure of the perceptual performance of subjects. The ITU has formalized a series of conditions that are used for quality assessment in various recommendations [10,11,12,13,14]. They suggest standard viewing conditions, criteria for the selection of observers and test material, assessment procedures, and data analysis methods. Recommended testing procedures exist as Absolute Category Rating (ACR), Degradation Category Rating (DCR), Comparison Category Rating (CCR).

In the test the presentation order of the video clips can be randomized between viewers in order to obtain a more realistic statistical score. After each video presentation, viewers are asked to judge its overall quality using the rating scale shown in Figure 2.



Fig. 2. Subjective assessment rating scale

Normally voting period was not time-limited. After choosing their quality rating, assessors had to confirm their choice using an 'OK' button. Furthermore, this eliminated the possibility of missing ratings in the test. After each video sequence and after to have gave a vote, a neutral gray background is, often, displayed on the video terminal during some second before the next sequence is presented. The test procedure and monitor selection adhered to the latest findings and recommendations for best practice from the Video Quality Experts Group (VQEG), a technical body supporting ITU standardization activities.

The use of this type of test has some disadvantages, first of all, the result of the test depend from uncontrollable attributes like experience, the mood, the attitude and culture, then, they are very expensive and impractical if you want to do frequently because of the number of subjects and tests are necessary to give reliable results. Anyway, subjective assessment are invaluable tools for evaluating multimedia quality. Their main shortcoming is the requirement for a large number of viewers, which limits the amount of video material that can be rated in a reasonable amount of time. Nonetheless, subjective experiments remain the benchmark for any objective quality metric.

2.2 Objective quality assessment

The subjective methods are not feasible during the design of a network. These are limited, impractical and very expensive. To overcome these problems have been developed, new methods that allow the calculation of values that represent the different combinations of

factors of damage that could affect the network. The primary purpose of these methods is to produce an estimate of quality, providing the results as comparable as possible to MOS values. The ITU is proposing a method of objective testing, automatic and repeatable, which takes into account the perceived quality. Different types of objective metrics exist [15]. For the analysis of decoded video, we can distinguish data metrics, which measure the fidelity of the signal without considering its content, and picture metrics, which treat the video data as the visual information that it contains. For compressed video delivery over packet networks, there are also packet- or bitstream-based metrics, which look at the packet header information and the encoded bitstream directly without fully decoding the video. Furthermore, metrics can be classified into full-reference, no-reference and reduced-reference metrics based on the amount of reference information they require.

The most known video quality metric is called Peak Signal to Noise Ratio (PSNR) that is calculated simply as mathematical difference between every pixel of the encoded video and the original video. The other popular metric is the classical Mean Square Error (MSE) is one of many ways to quantify the amount by which an estimator differs from the true value of the quantity being estimated.

3. Multimedia over wireless networks

Nowadays multimedia communication over wireless and wired packet based networks is growing up. In the past, many applications were used for video downloads, while now they take up the share of all traffic on the Internet. Most mobile devices can actively download and upload photos and videos, sometimes in real time. In addition, Voice over IP (VoIP) is heavily changing the voice telecommunications world and the enhanced television is also being delivered into the houses over IP networks by Digital Subscriber Line (DSL) technologies. Another issue in the multimedia revolution takes place inside the home environment: the electronics manufacturers, the computer industry and its partners are distributing audio and video over local-WiFi networks to monitors and speakers around the house. Now that the analog-to-digital revolution is going to be complete, the "all media over IP" revolution is taking place, with radio, television, telephony and stored media all being delivered over IP wired and wireless networks. Figure 3 shows an example of different multimedia applications in a home environment.

The growing and the emergence of communication infrastructures, like the Internet and wireless networks, enabled the proliferation of the above mentioned multimedia applications (music download to a portable device, watching TV through the Internet on a laptop, viewing movie trailers posted on the web through a wireless link). New applications are surely revolutionary, like sending VoIP to an apparently conventional telephone, sending television over IP to an apparently conventional set top box or sending music over WiFi to an apparently conventional stereo amplifier. The exposed applications include a big variety of new multimedia related services but, unfortunately, the Internet and the wireless networks do not provide full support for multimedia applications. The Internet and wireless networks have stochastic and variable conditions influenced by many factors, however variations in network conditions can have heavy consequences for real-time multimedia applications and can lead to unacceptable user experience, because multimedia transmissions are usually delay sensitive, bandwidth intense and loss tolerant. The theory that has traditionally been taught in information theory, communication and signal processing may not be directly applied to highly time-varying channel conditions and, as a

consequence, in recent years the area of multimedia communication and networking has emerged not only as a very active and challenging research topic, but also as an area that requires the definition of new fundamental concepts and algorithms that differ from those taught in conventional signal processing and communication theory.

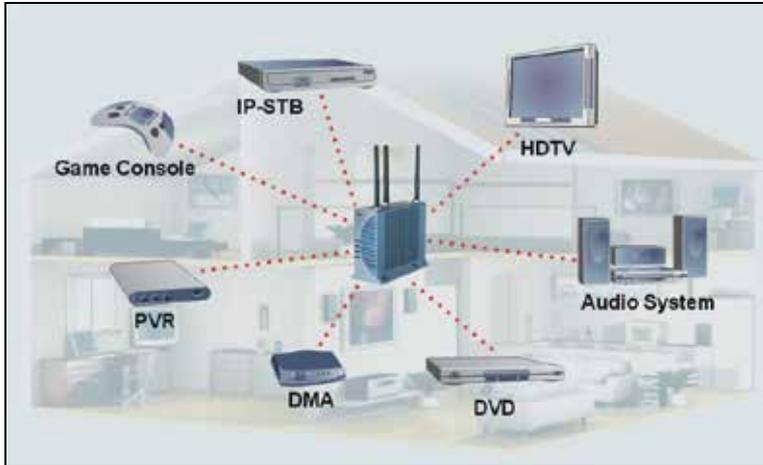


Fig. 3. Different Wireless IP-based multimedia applications (STP – Set Top Box for on-demand video, HDTV – High Definition TV, DMA – Digital Media Adapter for multimedia home extensions, PVR – Personal Video Recorder, etc.).

It is clear that Best-Effort (BE) IP networks are unreliable and unpredictable, especially in wireless networks, where there can be many factors that affect the Quality of Service (QoS), that measures the performance of a transmission system via parameters that reflect its transmission quality, such as delay, loss and jitter. In addition, congested network conditions result in lost video packets, which, as a consequence, produces poor quality video. Further, there are strict delay constraints imposed by streamed multimedia traffic. If a video packet does not arrive before its “playout time”, the packet is effectively lost. Packet losses have a particularly devastating effect on the smooth continuous playout of a video sequence, due to inter-frame dependencies. A slightly degraded quality but uncorrupted video stream is less irritating to the user than a corrupted stream. Controlled video quality adaptation is needed to reduce the negative effects of congestion on the stream whilst providing the highest possible level of service and quality. The applications based on streaming operations are able to split the media into separate packets, which are transmitted independently in the network, so that the receiver is able to decode and play back the parts of the bit stream that are already received. The transmitter continues to send multimedia data packets while the receiver decodes and simultaneously plays back other already received parts of the bit stream.

The philosophy of playing back received packets allows the reduction of the delay between the transmission time instant and the moment at which the user views the multimedia content. Having a low delay is of primary importance in such kind of systems, where interactive applications are dominant (for example, a video conference or a video on-demand architecture).

The transmission of multimedia content can be categorized into three main classes: unicast, multicast and broadcast, depending on the number of senders and receivers. Unicast

transmission connects one sender to one receiver (point-to-point connection, p2p), as downloading, on-demand streaming media and p2p telephony. The main advantage of unicast is that a feedback channel can be established between the receiver and the transmitter, so the receiver can return information to the sender about the channel conditions which can be used accordingly by the transmitter to change transmission parameters. In a multicast communication the sender is connected to multiple receivers that decide to join the multicast group; multicast is more efficient than multiple unicast flows in terms of network resource utilization, since the information must not be replicated in the middle nodes (it is obvious that in a multicast communication the sender cannot open a session toward a specific receiver). In a broadcast transmission, the sender is connected to all receivers that it can reach through the network (an example is a satellite transmission); the communication channel may be different for every receiver.

One possible approach to the problem of network congestion and resulting packet loss and delay is to use feedback mechanisms to adapt the output bit rate of the encoders, which, in turn adapts the video quality, based on implicit or explicit information received about the state of the network. Several bit rate control mechanisms based on feedback have been presented in the last few years. As the Real-Time Control Protocol (RTCP) provides network-level QoS monitoring and congestion control information such as packet loss, round trip delay, and jitter. Many applications use RTCP to provide control mechanisms for transmission of video over IP networks. However, the network-level QoS parameters provided by RTCP are not video content-based and it is difficult to gauge the quality of the received video stream from this feedback.

Now, in the following paragraphs, multimedia transmission techniques will be deeply introduced, as well as different policies dedicated to evaluate the quality (and the distortion) of the perceived content.

3.1 Background and coding

The transmission of video over wireless media is becoming very popular for the different variety of applications and networks. The advantages of these transmissions over wireless media are evident, but the transmission rate will always be limited due to the limitations introduced by physical layer. So, having the possibility of compressing the video before transmission is crucial in such kind of environments, especially for real-time traffic, where some constraints are required. For these reasons, video decoder must tolerate delay and packet losses: the standards in video coding (like MPEG-4 and H.264/AVC) [16,17,18,19] are today very popular, because of their capability to adapt to these environments.

These standards, like the previous well known ones, use a hybrid coding approach. The Motion Compensated Prediction (MCP) is combined with transform coding of the residual components in order to obtain a codec that can be very useful nowadays. As described in previous paragraph, video communications can be categorized into unicast, multicast and broadcast services, with different peculiarities, depending on the desired service, like the on-line generated or pre-encoded content, video telephony and downloading.

Figure 4 illustrates a generic scheme of a video transmission system, including the main components that go from the transmitter device to the receiver device. Data compression is feasible also because video content presents significant redundancy (in time and in space) that reduces the amount of transmitted packets significantly. According to Figure 4, the video encoder generates data units containing the compressed video stream, which is stored in the encoder buffer before the transmission. In general, the transmission system damages

(quantitatively or qualitatively) individual data units, also introducing some delay; then, the encoder/decoder buffer are used in order to balance the bit-rate fluctuations produced by the encoder and by the wireless channel. In general, a coded video stream can be considered as composed by a sequence of data units, called Access Units (AU) in MPEG-4 or Network Abstraction Layer Units (NALU) in H.264. The AUs/NALUs can be labeled as data unit specific information (as in MPEG, where they are labeled on the basis of relative importance for video reconstruction). On the other hand due to spatial and temporal prediction the independent compression of data units cannot be guaranteed without significantly losing compression efficiency.

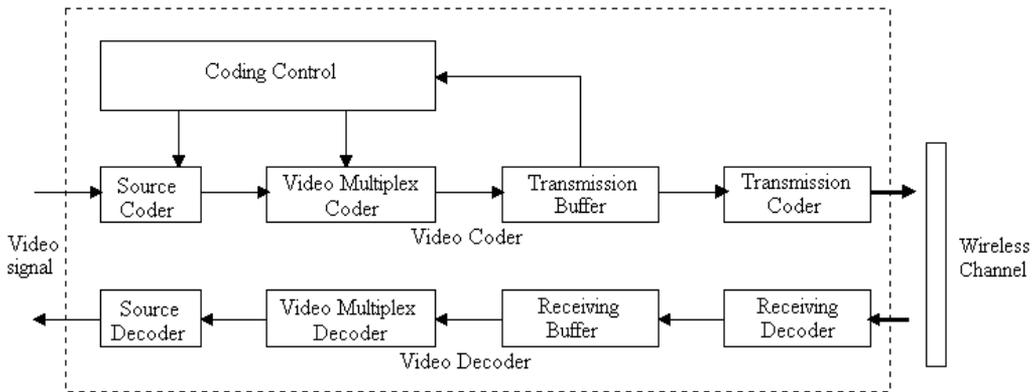


Fig. 4. Blocks diagram of a generic audio/video transmission system.

Errors introduction and its effects are considerably different in wired or in wireless networks, because of the different phenomena that impact on the medium. For wireless networks, fading and interference cause burst errors in form of multiple lost packets. Moreover congestion can result in lost packets in a wired IP network. Nowadays, even for wireless networks, systems are able to detect the presence of errors in a packet on physical layer and the losses are reported to higher layers. These techniques usually use Cyclic Redundancy Check (CRC) mechanisms. By consequence the video decoder will not receive the entire bit-stream. Intermediate protocol layers such as User Datagram Protocol (UDP) might decide to completely drop erroneous packets so deleting all the encapsulated data units. In fact, video data packets are considered lost if their delay overcomes a fixed and tolerable threshold, defined by the user video application.

Figure 5 illustrates a typical simplified version of an end to end video system when the MCP compressed video is transmitted over non error-free channels. In this environment, t represents the time, S_t is a single video frame composing a network packet P_t , C_t (with values 1 or 0) indicates if P_t is correctly received or discarded, while $S_t(C)$ represents the decoded frame as a function of the channel error pattern. Let us assume that a packet is transmitted over a channel that forwards correct packets to the decoder, in case of successful transmission, the packet is forwarded to the normal decoder operation such as Entropy Decoding and Motion compensated Prediction.

The prediction information and transform coefficients are reloaded from the coded bit-stream to reconstruct the current frame S_{t-1} . After that, the frame is forwarded to the display buffer and also to the reference frame buffer to be used in the MCP process to reconstruct

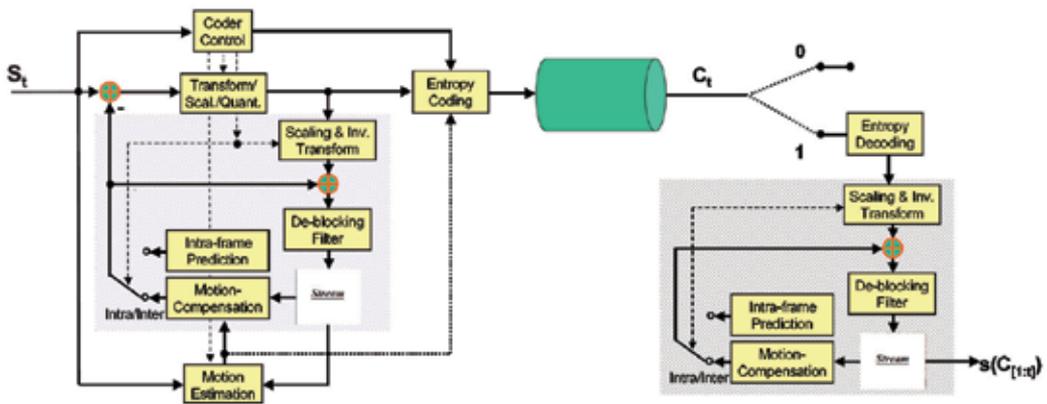


Fig. 5. End to end video transmission system.

the following inter-coded frames (i.e. the frame s_t frame). When the packet P_i is lost, i.e. at the reference time t , $C_t = 0$, the Error Concealment (EC) is necessary to be enabled, so the decoder just avoids the decoding operation and the display buffer is not updated, so the displayed frame is still S_{t-1} . In this case the viewer will understand that there has been a loss of motion, since the continuous display update is not maintained. Also the reference frame buffer is not updated as a result of this data loss.

In case of successful reception of packet P_{t+1} , the inter-coded frame S_{t+1} , reconstructed at the decoder, will in general not be identical to the reconstructed frame S_{t+1} at the encoder side, because as the encoder and the decoder refer to a different reference signal in the MCP process, there will be a reconstruction mismatch in reference signal when decoding S_{t+2} . For this reason it is obvious that the loss of a single packet P_i affects the quality of all the inter-coded frames: S_{t+1} , S_{t+2} , S_{t+3} , etc. This phenomenon is present in any predictive coding scheme and is called error propagation. If predictive coding is applied in the spatial and temporal domains of a sequence of frames, it is referred to as spatio-temporal error propagation.

The MCP technique makes the reconstructed frame S_i not only depending on the actual channel behavior C_i , but also on the previous channel evolution $C[1..t]$.

After these considerations, remarking that the error propagation has direct consequences on the perceived video content, it is preferable that a video coding system provides some of the following features:

- a reliable communication mean, in order to avoid transmission errors;
- an algorithm or device dedicated to the reduction of the visual effects of errors in the received frames;
- an algorithm or device for the minimization of the error propagation effect.

Generally, in MCP coding, the video content that belongs to a single frame is not encoded as a single entity: it is composed by a macro-block and individual data units are syntactically accessible and independent. The features that are employed in these systems to correct the introduced errors are the Forward Error Correction (FEC) and Backward Error Correction (BEC) or any combinations of those. It is very important, at the receiver side, that the erroneous and missing video content is observed, localized and, if possible, eliminated or minimized.

The modern video coding systems also have the capability to inform the video encoder about the loss of the video content, so the encoder can adapt itself to enhance the

transmission quality. The macro-blocks assignments, error control methods and feedbacks transmission, for example, must be used if we desire a robust, efficient and 'error-free' application.

As known in literature, if Shannon's separation principle [20] is observed, the main goal in video transmission (that is low error or error-free propagation) can be reached: the compression (or coding) features and the transmission (or link layer) ones can be optimized separately in order to avoid losses in video transmissions.

However, in several applications and environments (low delay situations), error-free transport may be impossible and the features like channel loss correction, detection and localization of unavoidable errors, their minimization, reduction of distorted frames impact become essentials. When we are dealing with 'no-wired' systems and QoS guarantees must be given, error control such as FEC and retransmission protocols are the primary features that should be provided.

The rules that are provided by the video coding standards, such as H.263 [21], MPEG-4 [22] and H.264, beyond the syntax and the semantics of the transmitted data, specify the decoder behavior in case of reception of an error-free bitstream, so the deployment of video coding standards still provides a significant amount of freedom for decoders that have to process erroneous bitstreams. In this way, a processing algorithm may be better than another one, in terms of computational complexity or quality of the received content. In the last years, video compression tools have evolved significantly over time: previous standards, like H.261, MPEG, MPEG-2 [23, 24], are very limited in error resilience capabilities, while the latter ones (starting from the H.263 [25]), heavily influenced the applications, with a lot of improvements and tools for error resilience. In the meantime, the new emerging standard MPEG-4 Advanced Simple Profile (ASP) introduced a radically different approach, based on some resilience tools, like Reversible Variable Length Coding (RVLC) and Re-synchronization Markers (RM) were introduced [26]. Figure 6 shows the evolution in time of the video compression standards.

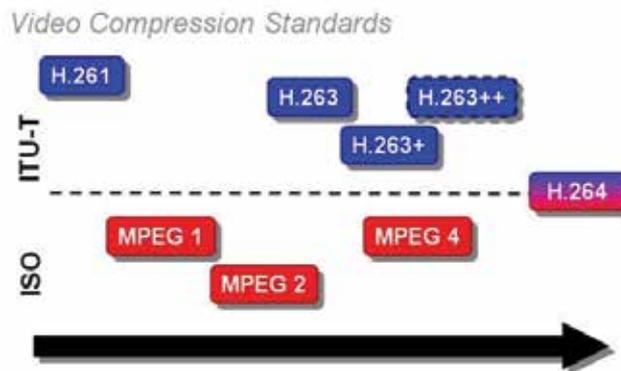


Fig. 6. The evolution of video compression standards.

As exposed in previous paragraphs, it is evident that the main aim of a communication system is to transfer the data generated by an information source efficiently and reliably over a non-ideal channel, respecting some QoS constraints. Among the different employed components (encoder, modulator, demodulator, source decoder) a channel encoder is used to insert additional and redundant data to the information sequence, so that channel errors can be detected or corrected. Shannon's channel coding theorem affirms that if there is

channel capacity (that is to say it is larger than the data rate), a coding scheme can be useful to reduce error probabilities. In general, the Bose-Chadhuri-Hocquenghem (BCH) codes are used and they represent an Error Control scheme based on block coding. They are a generalization of the Hamming codes and they add redundancy bits to the payload, in order to form code words; they also introduce the capacity of error correction (limited in number) [27]. The non-binary BCH-family codes that are massively used are the Reed Solomon (RS) codes. RS codes groups the bits into symbols, achieving good burst error suppression capability.

In order to achieve low-error (or error-free) communication, channel coding schemes must be implemented for the worst case channel characteristics. Obviously, when the channel is currently good, the channel protection, dimensioned for the worst case conditions, results in inefficient utilization of the wireless link. In addition, complex hardware or software structures may be required to realize the powerful long codes, required to defeat the worst case error patterns.

These drawbacks can be overcome by using the *Adaptive Coding* (AC) [28,29], by adding the redundancy on the basis of channel conditions and characteristics. Different adaptive algorithms have been proposed for video streaming applications, as the one proposed in [28]: in that work the fading level is estimated and, then, the algorithm evaluates the proper coding ratio in order to actively protect packets from losses. In this way, channel utilization can be immediately and efficiently improved.

3.2 Error perception

As discussed earlier, packet losses affect the quality of multimedia communications over wireless packet networks and the amount of quality degradation strongly vary on the basis of the meaning of the lost data. For the designing of an efficient loss protection mechanism, a reliable estimation method for multimedia data is needed. Providing an accurate estimation algorithm is required and in [30] the importance of a video coding element is outlined: for example, it is very useful to introduce the macro-block or packet concept as a value directly related to the distortion that would be introduced at the decoder by the loss of that specific element.

In Figure 7 the main phases of the "Analysis By Synthesis" (ABS) algorithm are illustrated; it is useful to compute the possible distortion of each element. It consists of the following steps, applied for each packet:

- decoding, with concealment, of the bitstream simulating the loss of the packet being analyzed (synthesis stage), by adding some errors;
- quality evaluation, in order to compute the distortion caused by the loss of the packet. The source and the reconstructed pictures are compared using Mean Square Error (MSE);
- storage of the obtained value as an indication of the perceptual importance of the analyzed video packet.

Little modifications of the standards are necessary in order to implement the modified encoding process. The algorithm reconstructs the encoded frames by the simulation of the decoder operations. It can be surely used for video coding, but the obtained values depend on the adopted encoding (if the video will be compressed with a different encoder or if a different packetization policy is used, values will be very different. The main disadvantage of the exposed technique consists of the interdependencies usually present between data

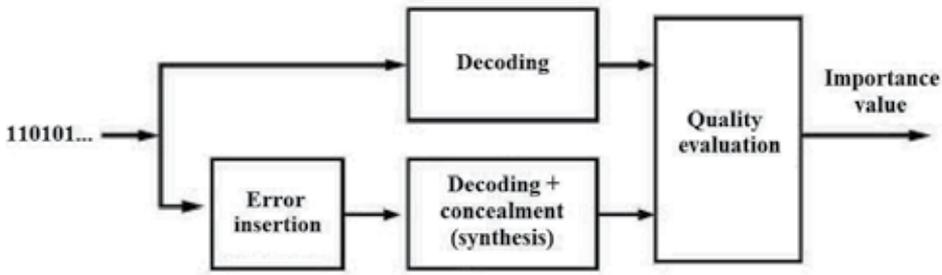


Fig. 7. Analysis by synthesis blocks diagram.

units, so the simulation of the loss of an isolated data unit is not completely realistic, particularly for high packet loss rates: all possible combinations of the events should be considered, weighted by its probability, and its distortion computed by the ABS technique, obtaining the expected importance value. The application of the ABS algorithm is easier when considering elements of the video stream which do not influence next frames (that is to say, no error propagation is present). If propagation is present, the distortion introduced in next frames should be evaluated until it is negligible (for example until a *I* frame is reached in a MPEG stream). In this case, the application of the ABS scheme is harder, due to the need of pre-computation of the perceptual distortion.

Another scheme for distortion evaluation is represented by the Distortion Matrix (DM) [31]: it allows to compute the distortion introduced when some frames are dropped in a sequence of Group Of Pictures (GOP) (as in an MPEG-2 stream). It is calculated under the assumption that once a specific P frame or I frame is lost, all the depending frames in the current GOP are replaced with the latest successfully decoded frame. This assumption makes the (DM) model unsuitable, because today all video decoders tend to mitigate the error propagation, reducing the distortion after a single loss. The GOP structure of video k is described by the GOP length L^k and the number of B-frames B^k between two *I* or *P* frames. For example, with $L^k = 9$ and

$B^k = 2$, the GOP structure will be *I B B P B B P B B*. The obtained distortion matrix as in [31] is shown in Figure 8.

$$\begin{array}{l}
 R : \\
 I : \\
 P_1 : \\
 P_2 : \\
 B_1 : \\
 B_3 : \\
 B_5 :
 \end{array}
 \left[\begin{array}{cccccccccc}
 D_I^R & D_{B_1}^R & D_{B_2}^R & D_{P_1}^R & D_{B_3}^R & D_{B_4}^R & D_{P_2}^R & D_{B_5}^R & D_{B_6}^R \\
 / & D_{B_1}^I & D_{B_2}^I & D_{P_1}^I & D_{B_3}^I & D_{B_4}^I & D_{P_2}^I & D_{B_5}^I & D_{B_6}^I \\
 / & / & / & / & D_{B_3}^P & D_{B_4}^P & D_{P_2}^P & D_{B_5}^P & D_{B_6}^P \\
 / & / & / & / & / & / & / & D_{B_5}^{P_2} & D_{B_6}^{P_2} \\
 / & / & D_{B_2}^{B_1} & / & / & / & / & / & / \\
 / & / & / & / & / & D_{B_4}^{B_3} & / & / & / \\
 / & / & / & / & / & / & / & / & D_{B_6}^{B_5}
 \end{array} \right]$$

Fig. 8. DM for $L_k = 9$ and $B_k = 2$.

The entries in the distortion matrix are the MSE values observed when replacing frame F_{loss} as part of the concealment strategy. The column left to the distortion matrix shows the

replacement frame for every row of the matrix. R is a frame from the previous GOP that is used as a replacement for all frames in the current GOP if the I -frame of the current GOP is lost. From this matrix, the resulting distortion for any possible loss pattern can be determined. The total distortion for the GOP is computed as the sum of the individual frame loss distortions. This matrix can be determined during the encoding of the video. The number of columns of the distortion matrix corresponds to the GOP length L^k . For more details see [31].

The ABS and DM end to end distortion estimation techniques can be categorized as “frame-based” methods, because they concern the distortion introduced at the frame level; other distortion evaluation techniques are well-known in the literature and they can be classified as “pixel-based” methods. In particular, the block-based approach generates and recursively updates a block-level distortion map for each frame [32,33,34]. Nevertheless, since inter-frame displacements influence sub-block motion vectors, a motion compensated block may inherit errors propagated from prior frames. In contrast, pixel-based policies estimate the distortion on a “pixel-basis”, so they have the advantage of providing high accuracy. Obviously, on the other hand, the complexity goes increasing. An innovative approach has been proposed in [35], where the distortion on pixel-basis is calculated by exhaustive simulation of the decoding procedure and averaging over many packet loss patterns. Another scheme is illustrated in [36], where only the two most likely loss events are considered.

However, in [37] it has been demonstrated that, by using the ROPE scheme, low complexity can be preserved, without losing the optimality of the distortion estimation algorithm. ROPE recursively calculates the first and second moments of the decoder reconstruction of each pixel, while accurately taking into account all relevant factors, including error propagation and concealment.

4. MPEG Standards

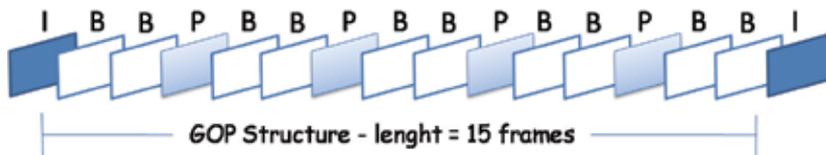


Fig. 9. GOP Structure

MPEG is the acronym of Moving Picture Experts Group, a working group which has the role of developing video and audio encoding standards [38].

MPEG video traffic is characterized by constant transmission rate of two groups of picture (GOP) (see Figure 9) per second and 15 frames per GOP. Since the number of bytes in a frame is dependent upon the content of the video, the actual bit rate is variable over time. However, the MPEG video supports also the constant bit rate (CBR) mode. There are three types of frames:

- I-Frames (intraframes) – encoded independently of all other frames;
- P-Frames (predictive frames) – encoded based on immediately previous I or P frames;
- B-Frames (bidirectionally predictive) – encoded based on previous and subsequent frames.

Many works in literature faced the problem of analyzing and describing the structure of MPEG-2 traffic streams, trying to find a way to emulate them through statistic and stochastic streams generators, preserving the proper and intrinsic nature of the original stream. In [39] the input process model is viewed as a compromise between the Long Range Dependent (LRD) and short range dependence (SRD) models. Simulation results were found to be better than those of a self-similar process when the switch buffer is relatively small. The MPEG video model presented in [40] is a Markov chain model based on the 'Group of Pictures' (GOP) level process rather than the frame level process. This has the advantage of eliminating the cyclical variation in the MPEG video pattern, but at the expense of decreasing the resolution of the time scale. Typically a GOP has duration of a half second, which is considered long for high speed networks. Of particular interests in video traffic modeling are the frame-size distribution and the traffic correlation. The frame size distribution has been studied in many existing works. Krunz [41] proposed a model for MPEG video, in which, a scene related component is introduced in the modeling of I frames, but ignoring scene effects in P and B frames. The scene length is i.i.d. with common geometric distribution. I frames are characterized by a modulated process in which the local variations are modulated by an Auto-Regressive (AR) process that varies around a scene related random process with log-normal distribution over different scenes; i.e., two random processes were needed to characterize I frames. The sizes of P and B frames were modulated by two i.i.d. random processes with log-normal marginal. This model uses several random process and need to detect scene changes, thus complicating the modeling process. In [42,43] the adaptive source is modeled by means of a discrete-time queuing system representing a virtual buffer, loaded by the video source when its quantizer scale parameter is changed according to the feedback law implemented in the encoder system. The whole paper is based on the Switched Batch Bernoulli Process (SBBP) that has been demonstrated to be suitable to model an MPEG video source; in fact, being a Markov modulated model, an SBBP is able to capture not only the first-order statistics but also the second-order ones which characterize the evolution of the movie scene.

In this paper we introduce a new concept of GOP-rate modeling, based on the discretisation method originally proposed in [44] for the wireless channel study. After the GOP-rate trend has been analyzed for the whole duration of the stream, it has been discretised in a certain number of states. Then the associated Markov Chain parameters have been evaluated.

MPEG algorithms compress data to form small bits that can be easily transmitted and then decompressed. It achieves its high compression rate by storing only the changes from one frame to another, instead of each entire frame. The video information is then encoded using a technique called Discrete Cosine Transform (DCT). MPEG uses a type of lossy compression, since some data is removed. But the diminishment of data is generally imperceptible to the human eye. The major MPEG standards include the following:

- MPEG-1: The most common implementations of the MPEG-1 standard provide a video resolution of 352-by-240 at 30 frames per second (fps). This produces video quality slightly below the quality of conventional VCR videos.
- MPEG-2: Offers resolutions of 720x480 and 1280x720 at 60 fps, with full CD-quality audio. This is sufficient for all the major TV standards, including NTSC, and even HDTV. MPEG-2 is used by DVD-ROMs. MPEG-2 can compress a 2 hour video into a few gigabytes. While decompressing an MPEG-2 data stream requires only modest computing power, encoding video in MPEG-2 format requires significantly more processing power.

- MPEG-3: Was designed for HDTV but was abandoned in place of using MPEG-2 for HDTV.
- MPEG-4: A graphics and video compression algorithm standard that is based on MPEG-1 and MPEG-2 and Apple QuickTime technology. Wavelet-based MPEG-4 files are smaller than JPEG or QuickTime files, so they are designed to transmit video and images over a narrower bandwidth and can mix video with text, graphics and 2-D and 3-D animation layers. MPEG-4 was standardized in October 1998 in the ISO/IEC document 14496.
- MPEG-7: Formally called the Multimedia Content Description Interface, MPEG-7 provides a tool set for completely describing multimedia content. MPEG-7 is designed to be generic and not targeted to a specific application.
- MPEG-21: Includes a Rights Expression Language (REL) and a Rights Data Dictionary. Unlike other MPEG standards that describe compression coding methods, MPEG-21 describes a standard that defines the description of content and also processes for accessing, searching, storing and protecting the copyrights of content.

5. Multimedia over satellite

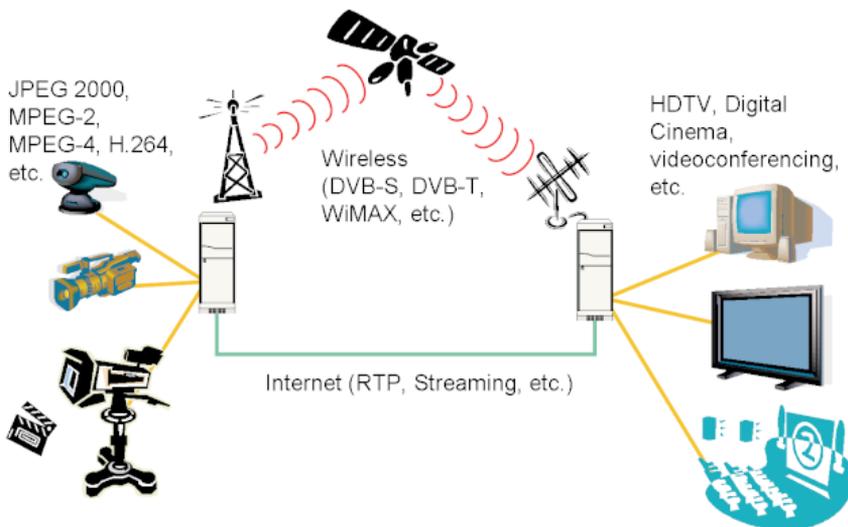


Fig. 10. Multimedia Broadcast Satellite Scenario

The new tendency of the multimedia applications on telecommunication networks is a full interactivity between users and network (see Figure 10). Thanks to this interactivity, users are able to manipulate what they receive on their terminal requiring to the network what they want. The interactivity has changed the way to exploit the network that pass from an asymmetry to a symmetry network. Then it is possible to use a return channel in order to make the users choices. This evolution in telecommunication also have consequences for satellite communications that, of sure, is the most oriented broadcast medium. Originally, the return channel of the satellite networks was designed as a terrestrial channel through different technologies such as PSTN, ISDN, GSM. Nowadays the most promising technology used for the satellite interactivity is a satellite one, in fact, the new standards for

satellite communications propose a return channel via satellite (RCS) like the DVB-RCS standard. The reasons for this choice are a lot, firstly customers prefer have a single technology for simplicity of management. It is very simple to have a single box in which all technical equipment are concentrates without have the necessity of interact with different objects. Another important reason regards the increased traffic in terrestrial networks that can produce problem in services providing such as service blocks and consequently reduction in quality of service. Moreover, the use of return satellite channel guarantees a greater bit rate available for the applications reaching a 2 Mbps against the few hundred of kbps of the terrestrial solutions. An example of this greater available bit rate can be seen on a file transfer application that for a 100 Mbyte file it will need about 7 minutes against the about 3 ½ hours of the terrestrial lines that operates a 64 kbps. There is also an advantage both for the users and the operators, that is to have both channels on the same medium. This produces a better control in QoS applications and in network management by operator, in fact, the terrestrial infrastructure is often not controlled by the same operator as for satellite, and this is certainly true when national borders are crossed.

Until 60 years ago, each individual communicated with about 100 other persons, of which 80–90 percent lived in closed vicinity. Twenty years ago this picture changed: each individual communicates with about 500 other individuals, of which 80–90 percent do not live in close vicinity. This was made possible by the emergence of advanced communications systems and by the integration of different technologies of communication that have allowed to reach users in very isolated areas. The emergence of new technologies has allowed the integration of different devices that originally worked separately as computers, TVs, telephones. Thanks to the Internet and to communication technologies always more sophisticates all consumers devices are able to work in a merged manner. The rapid technological advances will allow in a next future to have a exchange of information in every places and in every time.



Fig. 11. DVB Satellite Terminal

Satellite plays a key role in the telecommunications networks. It is able to resolve the problem of last mile providing connections to those areas where no investment return will be possible because a large investment is required to bridge services between the local exchange and the customers. Satellite, thanks to its broadband nature, is able to provide

connection in the rural areas and isolated areas with the same investment of other areas. Users has only to install a satellite terminal (see Figure 11) and subscribe to the service and they are able to receive satellite information. Moreover, thanks to the new RCS standard the customers can also use the network interactivity exploiting the return satellite channel that is faster than terrestrial one which is based on telephony network that are limited in providing high bit rates to subscribers. It will be necessary to perform a fiber cabling also between local exchange and subscribers in order to make faster the connections. Big investments are required by telephone companies to perform a fiber cabling in order to guarantee high bandwidth to each subscriber.

It is clear that the increase of multimedia services poses the problem of provide high bandwidth to the users by operators and in this context the use of satellite platform can represents an optimal solution in terms of costs. Moreover the satellite segment guarantees also an overall integration with all communication technologies.

5.1 Digital Video Broadcasting (DVB) system

In the mid-eighties, the ability to transmit digital images was still remote and it was thought that it was neither technically nor economically feasible, at least in the short term. The main reason was the high bit rate required, especially for the transmission of digital images in motion (from 108 to 270 Mbps). The most important issue was to improve the quality of the TV, so huge amounts of capital were invested in research and development of IDTV (improved Definition Television) and HDTV (High Definition Television). The situation in the early nineties has completely changed: the creation of efficient compression algorithms has resulted in immediate result, the birth of the standard JPEG image for fixed and then the MPEG standard for moving images; thanks to MPEG compression the amount of data required for transmission of digital images up to decrease a bit rate of between 1.5 and 30 Mbps has been reduced drastically, depending on the resolution chosen and the type of content of the transmitted images. Soon we saw that digital television would have satisfied the requirements of suppliers of services, but only if they had adopted a common standard, which in 1993 gave rise to the DVB Project [45]. DVB, short for Digital Video Broadcasting, is used with reference to digital television services in accordance with the standards developed by a consortium of 300 organizations (both public and private) of more than 50 countries, operating in different sectors: from production to broadcast TV of television sets by the rules of the frequency spectrum to the study of protocols for access to a network. The members of this organization work with the DVB project to develop a set of standards, technical recommendations and guidelines available to the various manufacturers. Once given the specifications, these standards are published by ETSI (European Telecommunication Standards Institute), and then made accessible to all. So thanks to these open standards, manufacturers can create interoperable DVB systems and can also easily adapt to different transmission channels (satellite, terrestrial, cable, etc.). Although it was born for the European landscape, the DVB platform is beginning to be accepted as a world standard solution, a number of radio and television programs based on DVB standards are currently operating in North and South America, Africa, Asia and Australia.

DVB means is based on the standard ISO 13818 encoded MPEG-2 and multiplexing specifications. It defines how to transmit the signals using MPEG-2 satellite, cable and terrestrial repeaters, and as transmitting the information system, the program guide, etc.. DVB allows the broadcasting of "data containers" that can include digital data of any type, a

key element of DVB is the source encoding of MPEG-2 data to be transmitted in such containers. According to the channel through which must be sent the data stream it is possible to make a differentiation of DVB in:

- DVB-S (Satellite), for transmission via satellite;
- DVB-C (Cable), for transmission over coaxial cable;
- DVB-T (Terrestrial), for terrestrial wireless transmissions.

Some of the most important advantages of the DVB standard are:

- Integration of data with audio and video
- Multi-programming, i.e. the ability to transmit multiple programs on the same channel (multiplexing more than 8192 streams);
- variable transmission capacity to meet quality and quantity of programs;
- use of transmission capacity for the introduction of additional data services such as teletext and / or other multimedia services;
- high data rates (> 48Mbps);
- optimal exploitation of the bandwidth of channels on satellite radio that is terrestrial;
- High reliability of the service accessed through an efficient system of modulation and coding for error correction.
- Security of transmission

Existing DVB IP systems can be classified into two categories:

- Unidirectional DVB IP Systems called one way
- Bidirectional system called DVB RCS (Return Channel System).

While the uni-directional systems are aimed primarily at a consumer market, the bi-directional systems are directed to a business market. Both systems are based on the common standard, however, DVB. The DVB IP systems one way may use the terrestrial return channel via PSTN or ISDN lines (typically 64 Kbps). DVB RCS systems are on a satellite channel dedicated to increased capacity (up to 2 Mbps). DVB IP platforms are based on existing standard DVB-S (Digital Video Broadcasting) for a number of reasons. First, the DVB technology has become an industry standard, it is well tested and there are a large number of manufacturers producing DVB devices both transmitters and only receivers. This makes hardware prices very low compared to other via satellite systems. The standard DVB-S is indispensable for communication over long distances and enjoys all the advantages of the satellite medium:

- Large coverage
- Rapid activation link
- Broadcast and multicast can be ready
- Bypassing the congested networks
- Data rates asymmetric
- High scalability

The software fairly robust and well tested is capable of supporting the transmission system DVB services and streaming data delivery based on IP protocols. Finally, the IP over DVB is a system optimized for the distribution of high volumes of data via satellite. This system provides significant benefits for the relatively easy integration with DVB equipment used for receiving satellite TV and the widespread use of such equipment in the consumer.

The specification of the DVB IP data distribution is defined by the ETSI standard EN 301 192 for Data Broadcasting and illustrates a variety of ways for the dissemination of data. This standard is also based on standard ETSI / DVB DVB-S (ETS 300 421 - modulation and

channel coding for satellite) and DVB-SI (ETS 300 468 - Information Service). This specification identifies four modes of spread: the data piping, data transmission through streams (data stream), the encapsulation of several protocols (Multi Protocol Encapsulation MPE), and data carousels. MPE is the format standardized by ETSI, which ensures interoperability between different hardware that provide this function. The protocol supports MPE data broadcasting services that require the encapsulation of communication protocols, such as IP. This affects applications Unicast (each datagram is labeled to be directed to a single user) and Multicast (the datagram is directed to a group of users). In the MPE packets are "tagged" by the MAC address (Media Access Control), but the DVB standard does not deal with as it should be allocated and maintained that address. The technology that lies at the basis of DVB IP is the transport of data through the encapsulation of the IP datagram within the DVB. The following figure (Figure 12) shows the protocols stack provided for the DVB IP standard.

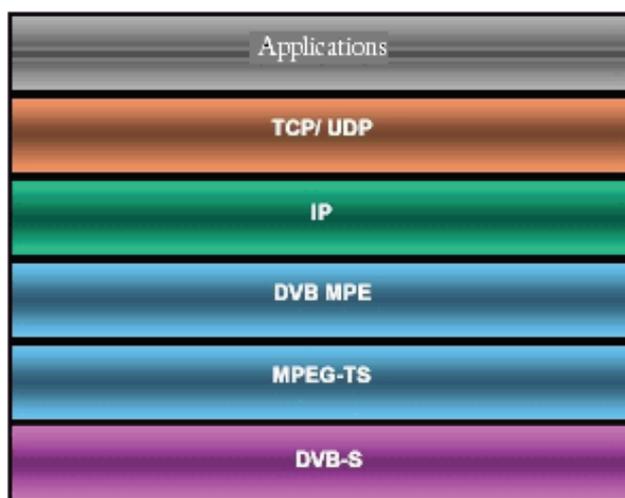


Fig. 12. Protocols stack of DVB IP standard

The MPEG-2 transport stream (MPEG-2-TS) arises from the need to create a layer of data for a fault tolerant environment (see Figure 13); and precisely, because this uses small packets. Although originally targeted for the spread of video and audio encoded in MPEG2 for digital television, the MPEG2-TS is particularly suitable for the transport of IP datagram. The transport stream consists of packets in time division multiplexed belonging to different flows of information. As already mentioned, each packet is 188 bytes, four of which belong to the header and the rest are left to the payload of the data. The header contains 4 bytes of synchronization and identification (PID) of the package, The payload (184 bytes) contains the data to be transmitted, such as data, audio and video format in small packages elementary (Packet Stream Element - PES). Each PES is of variable length and contains a header and a body of data called PES Data. The header of the PES includes a start code prefix, an identifier of flow, an optional length field, a PES-header and an optional number of bytes to fill. The remaining bytes are for data.

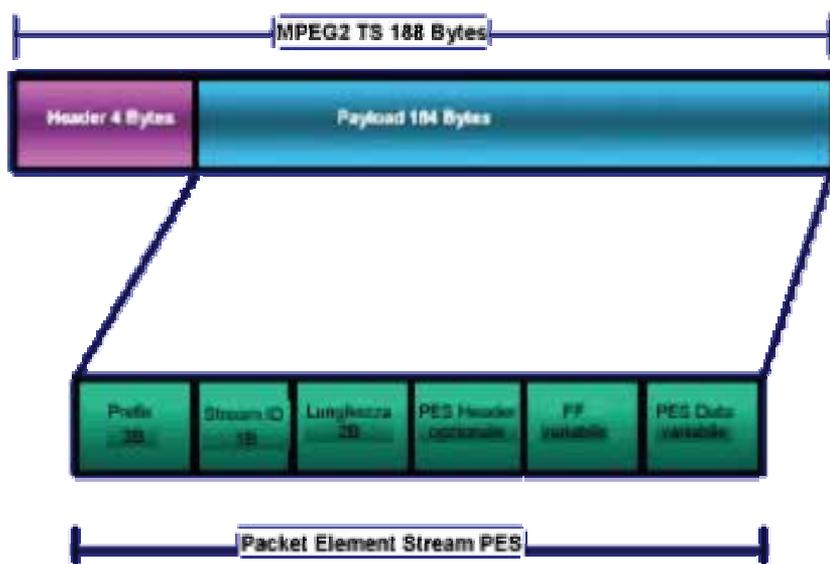


Fig. 13. MPEG 2 – Transport Stream

The large worldwide success of the DVB is due to the fact that it consists of a set of open standards, upon which they have agreed the entire DVB-Community made up of industrialists, traders, users, research institutes, etc. The DVB-RCS standard [46,47] was presented and approved ETSI (European Telecommunications Standards Institute) in 1999. Standard DVB-RCS is born as an evolution of TDMA networks where the carrier TDM (with bit rates up to 38Mbps), transmitted from the Master Station, transports the data packets in time division multiplexed addressed to all terminals of the network, instead, TDMA carriers (typically with bit rates up to 256Kbps) are shared between peripheral stations that have to talk with the Master Station. In order to make interoperable different technologies it has felt the need, at the European level, to define a standard which regulates the proliferation of networks with a star standard and proprietary hard interfaced with other technologies. The other reason that has necessitated the establishment of a standard DVB-IP was the proliferation of interactive applications with higher volumes of information that could not be implemented with the DVB-IP standard in which the return channel, made by a terrestrial link with a modem, did not allow an adequate bit rate (up to 64Kbps). The channel DVB-IP is referred to as direct channel or Forward Channel and the return channel RCS, Return Channel. The return channel has variable bit rate from 128Kbps up to 204Kbps and operates with TDMA in multi-access (MF-TDMA) in a dynamic way. The MF-TDMA access allows to give a slot of TDMA burst to terminals that request it in the pool of available frequencies. The channel DVB-IP or Forward Channel has a variable bit rate from 2Mbps up to 45Mbps and works with access TDM. The standard DVB-RCS can be implemented using a combination of Ku or Ka frequency for transmission and reception. The Ka-band offers to the Server Provider economic transponders to get a higher spectral efficiency. Such a system is designed to meet the needs of a wide section of users that can be distinguished into three categories:

- Prosumer
- Corporate
- Consumer

The main target is the Prosumer, a term used to describe a "professional user", which requires broadband, high-quality services and having the economic means to invest in relatively expensive equipment. The corporate customer is a further objective of the market, represented by a group of users who are behind a single terminal connected through a LAN for example, to a single RCST (Return Channel Satellite Terminal). The consumer will probably be the last profile that feels the need to use a similar system, but it is true the fact that the rapid technological development, together with the growing need for high capacity bandwidth and services, it will certainly make a category of users realistic in the near future. Let's look now at the applications on DVB-RCS:

- A first category of applications is made by "popular" services of the Internet, such as, for example, electronic mail, web-browsing, file transfer, and newsgroups.
- A second category of applications, so DVB-RCS exceed benefits in terms of access to the terrestrial network, is based on the multicast capability of the satellite such as multicast file transfer or streaming multicast. Many research results have led to a standard that supports the IP Multicast on the Internet. One of the major advantages of DVB-RCS standard is that it makes possible the Multicasting at low cost using existing Internet standards. The multicast data are channeled through an Internet Multicast Streaming Feeder Link, streaming from a source to a Multicast Streaming Server, then this data is transmitted on the satellite and sent to that particular group to which they were intended. The DVB-RCS system supports bandwidths relatively large for streaming compared to existing terrestrial solutions (from 64Kbps to 1Mbps).
- A third category of applications may be "Voice over IP." Broadband connections of DVB-RCS allows a good control of the flow constant rate. The biggest drawback is that the configuration of a star-RCS will require a double jump to a satellite connection between two users. This problem disappears if one user is connected to a hub through terrestrial links (PSTN or ISDN).

5.1.1 DVB Advantages and disadvantages

As already mentioned, the advantages of broadband satellite are undoubtedly significant: in addition to the ubiquitous availability of the service and the high speed of navigation available to users, in fact, are also to remember the cost lower than terrestrial connections, the possibility of subscribing services also with foreign companies, which means you have greater choice and range of commercial offers, and the opportunity to receive satellite TV channels on your personal computer at home or office. Another technical advantage is the equipment concentrated in a single box, called Set Top Box. Another reason for choosing a return channel via satellite is the significant increase in traffic of terrestrial networks that often reduces the quality of service (QoS). Finally, the forward channel and return channel is available on the same medium. This enables better control of QoS and network management by a single operator, which is not the case with terrestrial infrastructures that are not always handled by the same operator. QoS parameters include point-to-point delay, delay variation and packet loss. These parameters are measured on a path point-to-point, where the delay of propagation of the satellite was taken into account properly. In contrast, however, is also worth pointing out those that may be the defects of this type of connection as the inability to upload direct from most of the services, the excessive investment for purchasing the

necessary equipment for connection (satellite dish, LNB device, etc.), the problem of utilizing the phone line every time you make a web browsing (and the cost of calls made), the possibility of jamming signal that may occur due to the repositioning of satellites and weather situations, short delays in signal transmission due to the distance between the satellite and earth, and, not least, the difficulties of implementation and management of the connection that you can present to users are not particularly expert in the use of pc . The choice of the satellite connection is recommended, as well as to residents in areas not yet reached the service of terrestrial broadband, especially for users and companies that use the network to browse or download files of large size, but just recommended to all those realities, both professional and private, that are able to choose lines of traditional fast (even fiber optic) and require to send files on the network particularly heavy in terms of kb.

5.2 Platform integration: DVB-SH

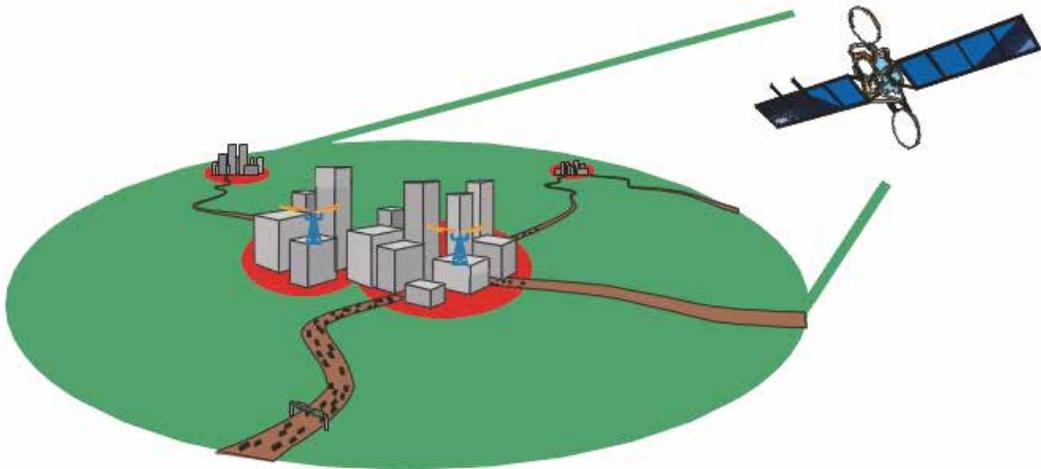


Fig. 14. DVB-SH Scenario

A new standard that is emerging in the DVB panorama is the DVB-SH whose first specification has been published in 2006 [48] (see Figure 14). It is a standard that includes the features of the well-known DVB-S and DVB-H. It is a hybrid system born to provide services through satellite and terrestrial platform to the users handheld, such as mobile phones, PDAs, vehicle-mounted, nomadic (laptops, palmtops...) and to fixed terminals. The hierarchical structure is composed of satellite platform and UMTS terrestrial repeater. This multi-layer structure allows to use the satellite communication when the conditions of LoS exist between satellite and device and, in the NLoS case, it can switch on the terrestrial repeaters in order to guarantee continuously connection to the subscribers. This standard has been designed to use frequencies around 3 GHz that fall in the S band (2.2 GHz) suitable for Mobile Satellite Service (MSS) and that are adjacent to the 3G terrestrial frequencies (see Figure 15).

DVB-SH inherits from the standard DVB-S many features such as turbo coding for forward error correction and a highly flexible interleaver in an advanced system designed to cope with the hybrid satellite/terrestrial network topology. The advantage of satellite channel is a

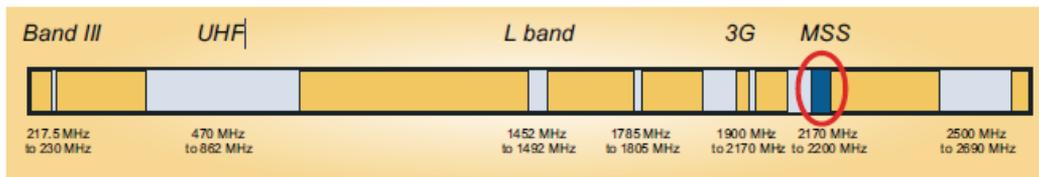


Fig. 15. 3G and S-Band spectrum

wide area coverage whilst the terrestrial component are able to provide coverage where the satellite signal cannot be received, as in urban canyon areas. The advantages of the terrestrial segment can be exploited for the deployment of repeaters for good indoor coverage otherwise not possible from satellite connection.

This new standard has to face with one of the most important application that in the last years is strongly emerging, the mobile TV; moreover it has to provide a lot of new multimedia services such as digital video streaming, voice over IP, Radio content delivery, interactive services, video on demand. The advantage of this system is the possibility of reach users that are in everywhere on the country and it guarantees access to users services that are moving in the region both walking or travelling in a car, train, ship. For this purpose the mobile terminal have to respect some required constraints in order to have the right compatibility with the mobility in term of size, weight and power consumption.

5.3 Satellite Digital Multimedia Broadcasting (SDMB) system

The Satellite Digital Media Broadcasting (S-DMB) system [49] consists in an overlay network based on satellite communications dedicated to terrestrial UMTS segments as it is possible to see in Figure 16. It is used when multimedia content, like TV programs, or non real-time multimedia services should be delivered to mobile nodes. This can be done by the use of geostationary satellites and low power terrestrial stations, which act like gap-fillers in order

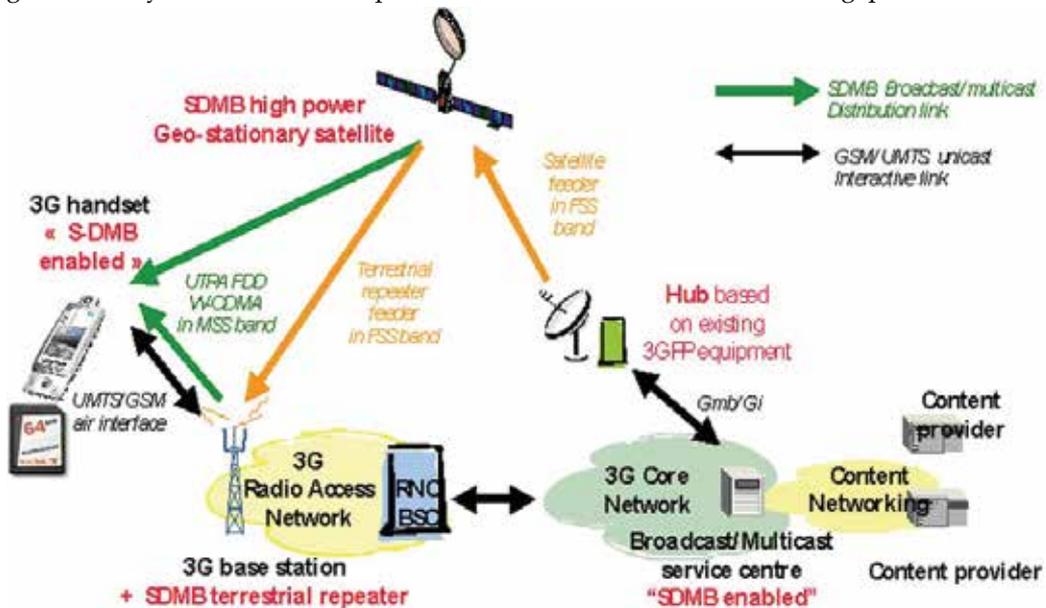


Fig. 16. SDMB Scenario

to cover urban and indoor environments. They can be located in the same places of mobile base stations. Satellites and terrestrial repeaters communicate with synchronized signals and the end-user nodes can use such kind of signals to improve reception quality. The S-DMB system is based on 18 channels (at 128kbps) in 15MHz, fully compliant with the UMTS Multimedia Broadcast/Multicast Service (MBMS). In this way, the integration among S-DMB/UMTS can become very useful. The streaming of TV programs on the own handset is suitable for end-users: it can be done wherever they are (waiting for a bus, for a TAXI, relaxing in a park, etc.); the only needed thing is a 3G compatible phone, although new technologies are taking place, like DVB-H or T-DMB.

The DMB is the natural extension of the Digital Audio Broadcasting (DAB) and there are two different versions of it: terrestrial or satellite DMB. Figure 17 represents the countries where the broadcasting technologies are employed (Europe, Canada, Australia, South Africa, New Zealand, India, China, South Korea and Turkey).

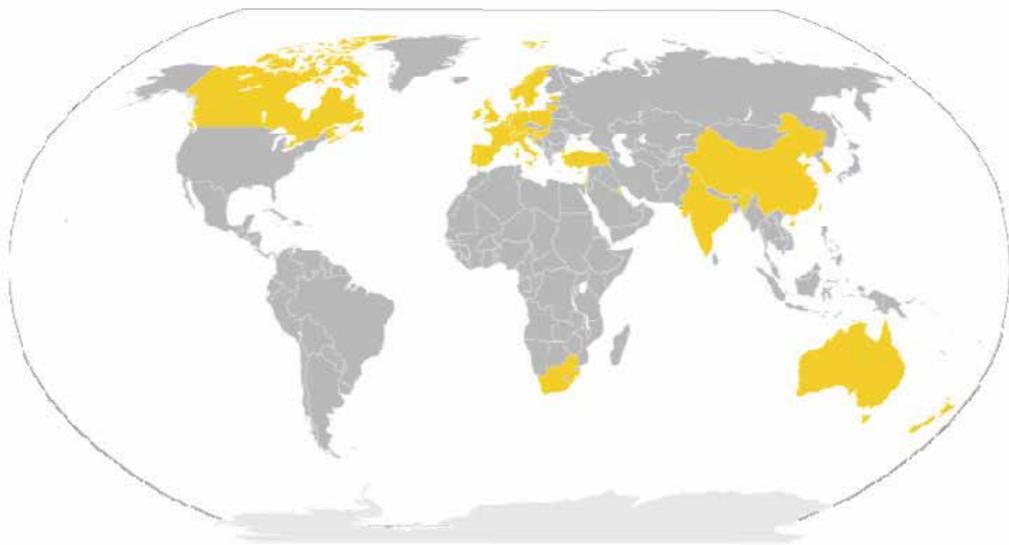


Fig. 17. DAB/DMB use in the world

5.4 Radio Resource Management in satellite environment

The Radio Resource Management (RRM) functionalities implemented at the Satellite broadcasting access layer comprise two main separated but cooperated parts: packet scheduling and connection admission control (CAC).

The physical channels are multiplexed in the satellite gateway through a radio resource allocation procedure. That is responsible of the radio bearer configuration at the time of the admission for each session, which includes the estimation of the required number of logical, transport and physical channels and their mapping from logical channels to the transport, physical channels.

The main function of a scheduling algorithm is to perform a time-multiplexing together with a QoS differentiated service flows and adjust the transmit power setting for each physical channel according.

The admittance decision of each incoming requested session is handled by the admission control function.

5.4.1 Packet scheduling schemes

The scheduling algorithm is responsible for managing the situations of contention for shared resources and limited, such as a buffer, a CPU, etc., In order to guarantee certain QoS requirements and fairness. In the telecommunications sector is often used algorithms to solve scheduling problems of access to the medium and, for example, to determine which users can transmit data over a communication channel shared. In this application, the main constraints to which the scheduling policies must be followed, among others, compliance with constraints on the delay, end-to-end, the jitter, the level of drop packages, or implementing a policy of "fair share" of resources among various users. The design of a scheduling algorithm is through the definition of objectives and constraints that the algorithm must satisfy. It is often necessary to adopt a scheduling algorithm in application scenarios characterized by response times are very low, so it is necessary that the algorithm can make decisions in a much reduced and with the least possible computational load. For example, in a network to 1Gbps with 1Kbyte packets, we will send a packet on average every 8 μ sec, the scheduling algorithm must be extremely fast in taking its decisions. Moreover, it is necessary to restrict the use of sophisticated data structures to implement the algorithm, as this would limit its use in real systems, with computing resources and memory are often limited. Many scheduling policies have a requirement to provide a breakdown "right" of resources, namely the maintenance require a certain level of "fairness" among users. An allocation of resource is "fair" if it meets the criterion of max-min allocation. " In an informal test this is to first meet the demand of small users, dividing among all the possible resource remaining. A policy of fair share, of course, cannot be applied to connections with guaranteed service, where users pay to receive a certain level of service [50].

Let us suppose to have a resource with capacity C and n service users with applications for resource x_1, \dots, x_n . Formally, the steps to follow to obtain a "fair share" of the resource according to the criterion of max-min allocation is as follows:

1. Let order the resource requests for increasing capacity, so that $x_1 \leq x_2 \leq \dots \leq x_n$.
2. It chooses the smallest request not yet satisfied, it assigns, initially, the amount C/n of the resource.
3. If $C/n > x_1$ it distributes the surplus in an equitable manner among all the other $n-1$ requests.
4. The algorithm takes from point 1 until all requests were taken into account.

Other algorithm features may include the presence of priority levels, the respect or not of the law of work conservation, rules for the connections aggregation on the basis of priority level, etc. ..

A scheduler is called "work-conserving" when the server is free only when the queue is empty.

5.4.2 Call admission control algorithms

Efficient radio resource management and CAC strategies are key components in wireless system supporting multiple types of applications with different QoS requirements. CAC tries to provide QoS to multiple types of applications with different requirements considering both call level and packet level performance measures. A CAC scheme aims at maintaining the delivered QoS to different calls (or users) at the target level by limiting the number of ongoing calls in the system. Call admission control (CAC) schemes have been investigated extensively in each type of network. Different approaches of CAC exist in literature, centralized, distributed, Traffic-Descriptor-Based, Measurement-Based and so on [51,52].

In satellite networks, different types of admission control have been studied. In [53] the authors have presented a novel strategy for handling ATM connections of different natures, traffic profile, and QoS requirements in enhanced satellite systems. CAC represents a module of Network Operation Center (NOC) disposed on a terrestrial station. Its task is to regulate the access to satellite segment. It permits a flexible handling of the bandwidth and avoids the a priori partitioning of the resources among different types of service. The CAC algorithm has been designed also to fulfill the objectives of minimizing the signaling exchange between the on-board and on-earth segments of the system. In order to reduce delays due to the processing of the call requests on board, the relevant parameters of the processed calls are stored and elaborated within the ground segment. The method is based on the concept of reserving buffer resources to each virtual circuit as long as data are sent. The decision on the call acceptance is taken following the evaluation of the excess demand probability, i.e., the probability that the accepted calls during their activation periods request more buffer resources than those available.

In [54,55] the authors propose an adaptive admission control strategy, which is aimed at facing link congestion and compromised channel conditions inherent in multimedia satellite networks. They present the performance comparisons of a traditional (fixed) admission control strategy versus the new adaptive admission control strategy for a Direct Broadcast Satellite (DBS) network with Return Channel System (DBS-RCS). Fixed admission control uses the same algorithm independent of the past traffic characteristics. The Bandwidth Expansion Factor (BEF) for VBR traffic is determined such that the probability of the aggregate instantaneous rate exceeding the fraction of the capacity assigned to the admitted VBR services will not be greater than a pre-specified probability value (ϵ). The dynamic approach recognizes that the admission control can only approximately estimate the statistical multiplexing and attempts to use the characteristics of past traffic streams to better estimate the gain that can be achieved. Unlike the fixed admission control, the adaptive admission control adjusts the BEF such that the actual value of is close to the desired value ϵ that is restricted by the acceptable QoS limits.

Concerning the Video Broadcasting delivery and scalability properties to be offered for large scale and heterogeneous networks, the authors in [56] adopted a Video on Demand (VoD) scheme where VBR videos are mapped over CBR channels and a traffic smoothing scheme with a buffering delay control are proposed. The same authors in [57], proposed novel broadcasting and proxy caching techniques in order to offer more scalability to the video delivery and to increase the overall performance of the system. In [58], the authors proposed a scheme to reduce the waiting time of the video application client side. The video traffic considered by authors was MPEG2.

In [59] the author performs a comparison between Quality oriented adaptation scheme (QOAS) against other adaptive schemes such a TCP Friendly Rate Control Protocol (TFRC), Loss-Delay-based Adaptation Algorithm (LDA+) and a non adaptive (NoAd) solution when streaming multiple multimedia clips with various characteristics over broadband networks.

The purpose of this study in [60] is to propose a quality metric of video encoded with variable frame rate and quantization parameters suitable for mobile video broadcasting applications. In [61] the authors present the results of a study that examine the user's perception of multimedia quality when impacted by varying network-level parameters (delay and jitter).

In contribution [62] they considered the GOP loss ratio as QoS parameter to be respected and VBR traffic has been considered in a DVB-RCS architecture.

6. Conclusions

In this chapter an analysis of the multimedia traffic over Wireless and Satellite networks has been shown. The importance of managing multimedia applications nowadays in the overall networks is an incontrovertible fact of our life. Moreover, the rapid increased use of mobile terminal together with video and audio services have pushed the research and the researchers toward new standards and technologies capable of dealing with these new users requirements. An important aspect of multimedia delivery contest is the possibility of make integration between different networks in order to be able of reaching users every-time every-where. Personal video recorders, video on demand, multiplication of program offerings, interactivity, mobile telephony, and media streaming have enabled viewers to personalize the content they want to watch and express their preferences to broadcasters. Viewers can now watch television at home or in a vehicle during transit using various kinds of handheld terminals, including mobile phones, laptops computers, and in-car devices. The concept of providing television-like services on a handheld device has generated much enthusiasm. Mobile telecom operators are already providing video-streaming services using their third-generation cellular networks. Simultaneous delivery of large amounts of consumer multimedia content to vast numbers of wireless devices is technically feasible over today's existing networks, such as third-generation (3G) networks. The concept of mobility has push toward wireless solutions such as terrestrial wireless networks and satellite one. Moreover, in the recent years new standards have been proposed for a integration of this two types of platforms giving the birth of hybrid solution like DVB-SH and SDMB standard. These standards provide integration between satellite and 3G networks in order to guarantee services also to those areas where the terrestrial infrastructures are impossible to install both for the particular territorial morphology and for economical issues.

7. References

- [1] Erling Kristiansen, Roberto Donadio, "Multimedia over Satellite - the European Space Agency Perspective", 2004
- [2] Wang Jian-ming, Zhan Shi Xian, and Zhao Xu Dong, "Improving the Multimedia traffic Performance over Wireless LAN", IEEE 2005
- [3] Vincent Chavoutier, Daniela Maniezzo, Claudio E. Palazzi, Mario Gerla, "Multimedia over Wireless Mesh Networks: Results from a Real Testbed Evaluation", Med Hoc Net 2007.
- [4] M-Etoh, T.Yoshimura, "Advances in Wireless Video Delivery," Prof.of the IEEE, vol.93, no.1, Jan.2005, pp.111-122.
- [5] Quan Huynh-Thu, Mohammed Ghanbari, "Termportal Aspect of Perceived Quality in Mobile Video Broadcasting" IEEE Transaction on Broadcasting, vol. 54, no. 3, September 2008.
- [6] Gabriel-Miro Muntean, Gheorghita Ghinea, Pascal Frossard, Minoru Etoh, Filippo Speranza, Hong Ren Wu, "Advanced Solution for Quality-Oriented Multimedia Broadcasting", IEEE Transaction on Broadcasting, vol. 54, no. 3, September 2008.
- [7] Sfehan Winkler, Praaven Mohandas, "The Evolution of Video Quality Measurement: From PSNR to Hybrid Metrics", IEEE Transaction on Broadcasting, vol. 54, no. 3, September 2008.

- [8] Noriki Uchida, Kazuo Takahata and Yoshitaka Shibata, "Optimal Video Stream Transmission Control over Wireless Network", IEEE International Conference on Multimedia and Expo (ICME) 2004
- [9] Ron Shmueli, Ofer Hadar, Revital Huber, Masha Maltz, and Merav Huber, "Effects of an Encoding Scheme on Perceived Video Quality Transmitted Over Lossy Internet Protocol Networks", IEEE Transaction on Broadcasting, vol. 54, no. 3, September 2008.
- [10] ITU-R Recommendation BT.500-11, "Methodology for the subjective assessment of the quality of television pictures," International Telecommunication Union, Geneva, Switzerland, 2002.
- [11] ITU-T Recommendation P.910, "Subjective Video Quality Assessment Methods for Multimedia Applications," International Telecommunication Union, Geneva, Switzerland, 1999.
- [12] ITU-T Recommendation P.911, Subjective Audiovisual Quality Assessment Methods for Multimedia Applications International Telecommunication Union, Geneva, Switzerland, 1998.
- [13] P. Corriveau, "Video quality testing," in Digital Video Image Quality and Perceptual Coding, H. R. Wu and K. R. Rao, Eds. Boca Raton, FL: CRC Press, 2006, ch. 4
- [14] ITU-T P911: "Subjective audiovisual quality assessment methods for multimedia applications".
- [15] S. Winkler, "Video quality and beyond", in Proc. European Signal Processing Conference, Poznan, Poland, Sept. 3-7 2007, invited paper.
- [16] Advanced Video Coding for Generic Audiovisual Services," ITU-T and ISO/IEC JTC 1., 2003.
- [17] G. Sullivan and T. Wiegand, "Video compression from concepts to the H.264/AVC standard," in *Proceeding of the IEEE*, vol. 93, June 2005.
- [18] A. M. Tourapis, K. Sühring, G. Sullivan, *Revision of the H.264/MPEG-4 AVC Reference Software Manual*, Dolby Laboratories Inc., Fraunhofer-Institute HHI, Microsoft Corporation, JVT-W041, April 2007.
- [19] A. Luthra, G. Sullivan, T. Wiegand, *Special Issue on the H.264/AVC video coding standard*. IEEE Transaction on Circuits and Systems for Video Technology, July 2003, vol. 13, no. 7.
- [20] C. E. Shannon, "A Mathematical Theory of Communications," *Bell Systems Technology*, p. 379-423, 1948.
- [21] "Video coding for narrow telecommunication channels at 64 kbit/s," ITU-T Recommendation H.263, 1996.
- [22] S. Bauer, J. Kneip, T. Mlasko, B. Schmale, J. Vollmer, A. Hutter, and M. Berekovic, "The mpeg-4 multimedia coding standard: Algorithms, architectures and applications," *J. VLSI Signal Process. Syst.*, vol. 23, no. 1, 1999.
- [23] "Video Codec for audiovisual services at p x 64 kbits/s," ITU-T Recommendation H.261, 1990.
- [24] D. Banks and L. A. Rowe, "Analysis tools for mpeg-1 video streams," EECS Department, University of California, Berkeley, Tech. Rep. UCB/CSD-97-936, Jun 1997. [Online]. Available: <http://www.eecs.berkeley.edu/Pubs/TechRpts/1997/5497.html>
- [25] Bernd Girod, "Comparison of the H.263 and H.261 Video Compression Standards."
- [26] T. Chujoh and T. Watanabe, "Reversible variable length codes and their error detecting capacity," in *Proceeding of the IEEE*. Portland, (OR): Picture Coding Symposium, 1999, pp. 341-344.

- [27] H. Liu, H.Ma, M. E. Zarki, and S. Gupta, "Error control schemes for networks: An overview," *Mobile Networks and Applications*, no. 2, p. 167-182, June 1997.
- [28] W. Kumwilaisak, J. Kim, and C. Kuo, "Reliable wireless video transmission via fading channel estimation and adaptation," in *Proceedings of IEEE WCNC, Chicago, IL*, September 2000, p. 185-190.
- [29] S. Lin and D. C. Jr., *Error Control Coding*, 2nd ed., ed., Ed. Prentice Hall, 2004.
- [30] E. Masala and J. C. De Martin, "Analysis-by-synthesis distortion computation for rate-distortion," in *IEEE International Conference on Multimedia and Expo*, Baltimore, MD, Ed., vol. 3, July 2003, p. 345-348.
- [31] W. Tu, W. Kellerer, and E. Steinbach, "Rate-Distortion Optimized Video Frame Dropping on Active Network Nodes," in *IEEE Proceedings of the International Packet video Workshop*, Irvine, CA, Ed., December 2004.
- [32] G. Cote, S. Shirani, and F. Kossentini, "Optimal mode selection and synchronization for robust video communications over error prone networks," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 952-965, June 2000.
- [33] S. Ekmekci and T. Sikora, "Recursive decoder distortion estimation based on source modeling for video," in *Proceeding of International Conference on Image Processing(ICIP)*, 2004, p. 187-190.
- [34] Y. Zhang, W. Gao, H. Sun, Q. Huang, and Y. Lu, "Error resilience video coding in H.264 encoder with potential distortion tracking," in *Proceeding of International Conference on Image Processing(ICIP)*, vol. 1, 2004, p. 173-176.
- [35] T. Stockhammer, T.Wiegand, and S.Wenger, "Optimized transmission of H.26L/JVT coded video over packet-lossy networks," in *Proceeding of International Conference on Image Processing(ICIP)*, Rochester, NY, vol. 2, 2002, p. 173-176.
- [36] T.Wiegand, N. Farber, K. Stuhlmuller, and B. Girod, "Error-resilient video transmission using long-term memory motion-compensated prediction," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 1050-1062, June 2000.
- [37] R. Zhang, S.L. Regunthan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 966-976, June 2000.
- [38] B.Haskell, A.Puri, and A.Netravali, *Digital MPEG: An Introduction to MPEG-2*, London, U.K.: Chapman and Hall, 1997.
- [39] M. M. Krunz and A. M. Makowski, "Modeling video traffic using M/G/1 input processes: a compromise between Markovian and LRD models", *IEEE J. Select. Areas Commun.*, vol. 16, no. 5, pp. 733-749, Jun. 1998.
- [40] O. Rose, "Simple and efficient models for variable bit rate MPEG video traffic", *Performance Evaluation*, vol. 30, pp. 69-85, 1997.
- [41] M. Krunz and S. K. Tripathi, "On the characterization of VBR MPEG streams", in *Proc. SIGMETRICS'97*, Cambridge, MA, Jun. 1997, pp. 192-202.
- [42] A.Lombardo,G. Morabito,G. Schembra, "Modeling intramedia and intermedia relationships in multimedia network analysis through multiple timescale statistics," *IEEE Transactions on Multimedia*, Vol.6, no.1, Feb. 2004, pp.142-157.
- [43] A.Lombardo, G.Schembra, "Performance evaluation of an adaptive-rate MPEG encoder matching IntServ traffic constraints," *IEEE/ACM Trans.on Networking*, Vol.11, no.1, Feb. 2003, pp.47-65
- [44] X. S. Wang, Moayeri, "Finite-state Markov channel - a useful model for radio communication channels", *IEEE Transactions on Vehicular Technology*, Volume 44, Issue 1, Feb. 1995 Page(s):163 - 171.
- [45] www.dvb.org

- [46] ETSI TR 101-790 V1.1.1 Digital Video Broadcasting (DVB); Interaction Channel for Satellite Distribution Systems; Guidelines for the use of EN 301-790.
- [47] ETSI, "Digital Video Broadcasting (DVB); Interaction channel for Satellite Distribution System," DVB-RCS 333, REV6.0, 22Feb. 2002.
- [48] Philip Kelley, Christian Rigal, "DVB-SH - mobile digital TV in S-Band", EBU TECHNICAL REVIEW - July 2007.
- [49] Christophe Selier, Nicolas Chuberre, "Satellite Digital Multimedia Broadcasting (SDMB) system presentation", 14-th IST Mobile and Wireless Communication Summit, Dresden, 19-23 June 2005.
- [50] Linghang Fan, Hongfei Du, Upendra Mudugamuwa, Barry G. Evans, "A Cross-Layer Delay Differentiation Packet Scheduling Scheme for Multimedia Content Delivery in 3G Satellite Multimedia System", IEEE Transaction on Broadcasting, vol. 54, no. 4, December 2008.
- [51] L.Huang, C.-C.Jay Kuo, "Joint Connection-Level and Packet Level Quality-of-Service Support for VBR Traffic in Wireless Multimedia Networks," IEEE Journ.on Selected Area in Comm.,vol. 23, no.6, June 2005, pp.1167-1177.
- [52] D.Niyato, E. Hossain, "Call Admission Control for QoS Provisioning in 4G Wireless Networks: Issues and Approaches", IEEE Network, Sept./Oct. 2005
- [53] A.Iera, A.Molinaro, S.Marano, "Call Admission Control and resource management issues for real-time VBR traffic in ATM-satellite," IEEE Journ.on Selected Area in Comm., vol.18, pp.2393-2403, Nov.2000.
- [54] F.Alagoz, et al., "Fixed versus Adaptive Admission Control in Direct Broadcast Satellite Networks with Return Channel Systems," IEEE Joun.on Selected Areas in Comm., Vol.22, No.2, Feb.2004, pp.238-249.
- [55] F.Alagoz, "Approximation on the aggregate MPEG traffic and their impact on admission control," Turkish J.Elec.Eng.Comput.Sci., vol.10, pp.73-84, 2002.
- [56] W.-F.Poon, K.-T.Lo, J.Feng, "Interactive Broadcasting System for VBR Encoded Videos," in IEEE Transaction on Broadcasting, Vol.53, No.2, June.2007, pp.459-467.
- [57] K.-M.Ho, W.-F.Poon, K.-T.Lo, "Performance Study of Large-Scale Video Streaming Services in Highly Heterogeneous Environment," in IEEE Transaction on Broadcasting, Vol.53, No.4, Dec.2007, pp.763-773.
- [58] T.Yoshihisa, M.Tsukamoto, S.Nishio, "A Broadcasting Scheme Considering Units to Play Continuous Media Data," in IEEE Transaction on Broadcasting, Vol.53, No.3, Sept.2007, pp.628-636.
- [59] G.-M.Muntean, "Efficient Delivery of Multimedia Streams Over Broadband Networks Using QOAS," in IEEE Transaction on Broadcasting, Vol.52, No.2, June.2006, pp.230-235.
- [60] R.Feghali, et al., "Video Quality Metric for Bit Rate Control via Joint Adjustment of Quantization and Frame Rate," in IEEE Transaction on Broadcasting, Vol.53, No.1, March.2007, pp.441-446.
- [61] S.R.Gulliver, G.Ghinea, "The Perceptual and Attentive Impact of Delay and Jitter in Multimedia Delivery," in IEEE Transaction on Broadcasting, Vol.53, No.2, June.2007, pp.449-458.
- [62] De Rango F., Tropea M., Fazio P., Marano S., "Call Admission Control for Aggregate MPEG-2 Traffic over Multimedia Geo-Satellite Networks," to be published on IEEE Transaction on Broadcasting.

Adaptive Video Transmission over Wireless MIMO System

Jia-Chyi Wu, Chi-Min Li, and Kuo-Hsean Chen

National Taiwan Ocean University

Department of Communications, Navigation and Control Engineering

Taiwan

1. Introduction

There has been an interesting issue in multimedia communications over wireless system in recent years. In order to achieve high data rate wireless multimedia communications, spatial multiplexing technique (Foschini & Gans, 1998; Wolniansky et al. 1998) has recently developed as one of the most noteworthy techniques as multiple-input and multiple-output (MIMO) systems. If the channel state information is perfectly available at the transmitter (Driessen & Foschini, 1999; Burr, 2003), we can maximize the channel capacity to design a realizable video transmission system. Under channel capacity limitation, the chapter presents how to employ joint source-channel coding algorithm with adequate modulation techniques to get the possibly best performance in the system design. Adaptive video coding to the varying channel conditions in real-time is well matched to MIMO systems for an optimized video transmission. An important matter in designing adaptive video transmission system is how often the feedback of the channel state information should be carried out. In fact, the feedback interval is mainly decided by the channel characteristics. For wireless fading channels, the feedback information needed to be able to capture the time varying channel characteristics for a true adaptive transmission. Song & Chen (2007, 2008) proposed adaptive algorithm design to utilized partial channel state information from receiver for layered scalable video coding (SVC) transmission over MIMO system. There are some interesting topics related in adaptive video transmission over wireless multimedia communication systems can be found in (Chen & He, 2006).

In our proposed system, we investigate the system performance of a joint MPEG-2 coding scheme with convolutional channel coding and space time block coding (STBC) techniques, associated with suitable modulation method (BPSK or QPSK), for video data transmission over a wireless MIMO system with Rayleigh fading noises. Rates assigned to MPEG-2 source code and convolutional channel code as well as space-time block code schemes are based on the feedback information from Performance Control Unit (PCU) under system channel capacity limitation, which ensures the proposed system achieved the best performance compared to a conventional designed system. In a conventional way, source coding and channel coding are designed to accomplish the best system performance respectively. With simply combining the best source coding scheme with the best channel coding scheme together, the system does not promise a better overall performance.

Consequently, the present algorithm employs joint source-channel coding scheme and MIMO concept to get the best performance in the system design over fading channel.

We are interested in the joint source-channel coding with modulation scheme design under the channel capacity constraint consideration in a MIMO system. Figure 1 shows joint source-channel codes under the combination of various source coding rates and various channel coding rates. Source coding is concerned with the efficient representation of a signal. While bit errors in the uncompressed signal can cause minimal distortion, in its compressed format a single bit error can lead to significantly large errors. For data, channel coding is necessary to overcome the errors resulted from transmission channel. We have noticed that combining source coding with adequate channel coding, we should be able to achieve a better system performance. Assuming that the overall system transmission rate $r = k/n$, where k is the source coding rate and n is the channel coding rate. In Fig. 1, we have found that a better performance (with a lower distortion) can be promised when we increase the source coding rate k , while we increase the channel coding rate n , a higher bit error rate (a lower system performance) happened under the same E_b/N_0 , signal-to-noise ratio (SNR) criterion. Therefore, we would like to design a transmission system with higher source coding rate k but lower channel coding rate n to achieve a higher overall transmission rate r . Since the overall transmission rate r is under channel capacity limitation, we have to justify the concept with proper method to design transmission system.

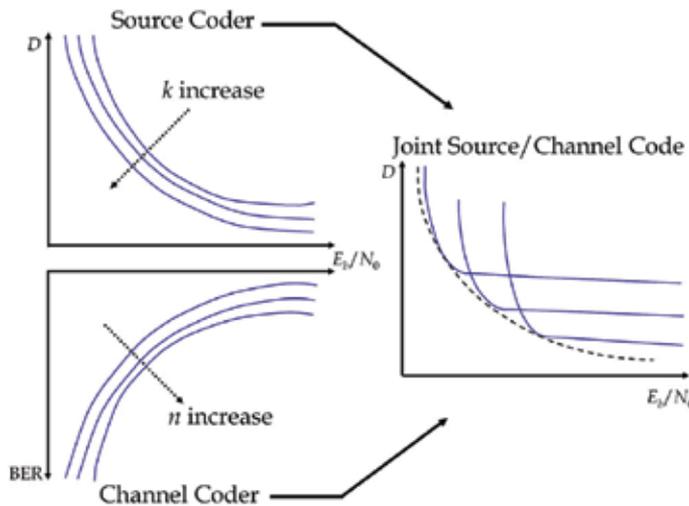


Fig. 1. Joint source and channel coding with different rates

The overall transmission rate r can be obtained by source coding rate k cooperated with channel coding rate n . We will not satisfy the system performance while we have a high source compression ratio (lower k) with strong channel protection (lower n). In turns, if we apply a low source compression ratio (higher source coding rate k with low distortion) but a high channel coding rate n (weak channel protection capability) to the system, which may result in higher bit error rate (BER) performance, we are not satisfactory with the reconstructed signal from the received high BER data. It is quite clear that we have to find a better match between source coding rate and channel coding rate to assure an acceptable system performance.

The most significant criterion in designing a transmission system is the channel capacity limitation. The available channel capacity restricts the overall transmission rate r , which is the rate between source coding rate k and channel coding rate n . We have to consider source coding rate, channel coding rate, and the corresponding modulation type all together simultaneously to cope with the channel capacity limitation. Assuming channel capacity limitation is one bit/transmission, we are asked to keep overall transmission rate $r \leq 1$ bit/channel-use, which can be achieved only with $k \leq n$. Therefore, we will keep our system rate design with $r \approx 1$ and $r \leq 1$, that is, $k \approx n$ and $k \leq n$. Furthermore, space-time block coding (STBC) algorithm was introduced (Alamouti, 1998; Tarokh et al., 1999) as an effective transmit diversity technique to resist fading effect. For a fixed number of transmit antennas, its decoding complexity increases exponentially with the transmission rate. The proposed algorithm employs joint source-channel coding scheme with STBC technique to get the best performance in MIMO systems design.

1.1 Outline

The rest of the chapter is organized as follows. Section 2 describes the system configuration adopted in this proposed algorithm. Experimental results for performance of the overall adaptive video transmission system compared with a conventional scheme over Rayleigh fading channel are shown in Section 3. Finally, a summary and conclusions are presented in Section 4.

2. System configuration

We are interested in the joint source-channel coding with modulation scheme design under the channel capacity constraint consideration in a MIMO system. We have applied the integrated transmission system design method (Daut & Ma, 1998) for digital transmission of video signals over noisy channels. To transmit a given video bit stream efficiently, we propose a joint source-channel coding system as shown in Figure 2.

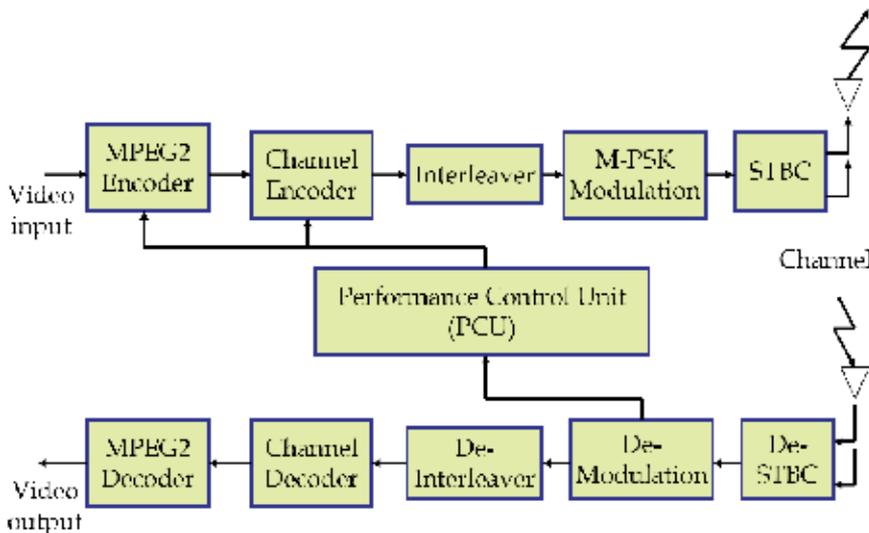
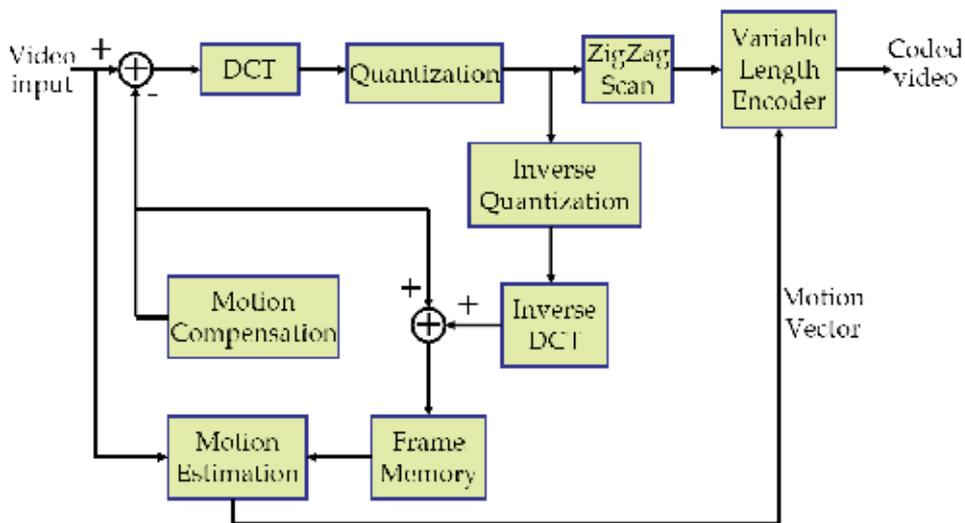


Fig. 2. Proposed adaptive video transmission system block diagram

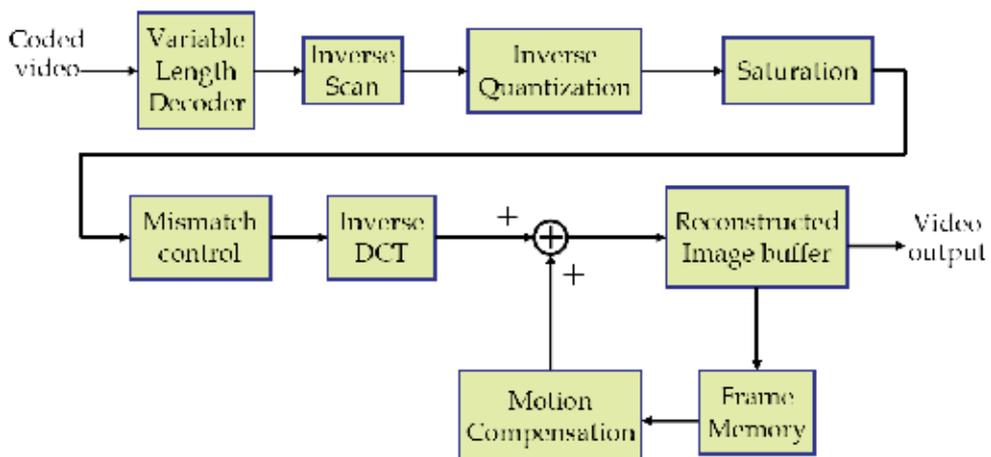
In this proposed system, the video sequence is first source coded by a MPEG2 scheme (Sikora, 1997). In order to reduce the system complexity of decoding, after the source coding stage, we use convolutional code and STBC in channel coding. The interleaver is adopted which is effected resisting burst error in wireless channel. There are two modulation techniques employed to be selected, BPSK or QPSK. The channel capacity is limited to one bit/transmission.

2.1 MPEG2 video source coding

The proposed adaptive video transmission system has been experimentally tested using an MPEG2 source coding algorithm provided from MPEG.ORG website.



(a) MPEG2 Encoder



(b) MPEG2 Decoder

Fig. 3. MPEG2 video coding block diagram

Figure 3 shows the video coding and decoding block diagram of MPEG2 scheme, in which we can change the coding bit rate of MPEG2 to obtain the required source compression ratio. In the proposed system, there are three MPEG2 video coding rates adopted and the resulted compressed video frames compared with the original test video is shown in Fig. 4. The source coding scheme is MPEG2 format and there are 160×120 pixels in every frame. It can be seen that the video quality is better with MPEG2 coding rate 0.6659 bit/pixel (bpp) among the three tested coding rates.

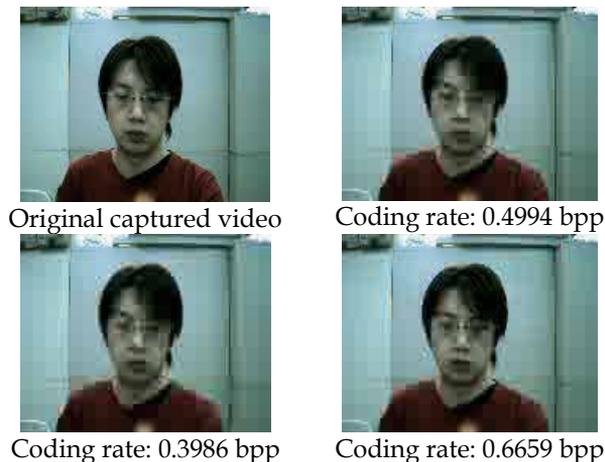


Fig. 4. Video quality comparison of original and MPEG2 compressed test video frames

2.2 Channel coding – convolutional coding and space-time block coding

In order to reduce the channel error effect and to improve the system performance while transmitting video signals over wireless channel, we have employed the convolutional encoder and maximum-likelihood Viterbi hard decision decoder for channel error correction. Figure 5 shows a typical $1/2$ recursive systematic convolutional (RSC) code scheme with generator function $G(D) = [1, (1+D^2)/(1+D+D^2)]$ (Proakis, 2001). After the convolutional encoding, the processed data is fed into a random interleaver to reduce burst error effect in wireless channel. The convolutional coding rates provided in this proposed system are set as $2/5$, $1/2$, and $2/3$, respectively. The channel coding rate selected is corresponding to the MPEG2 source coding rate to satisfy the channel capacity limitation to one bit/transmission. For the system simulation, we have adopted two modulation types: BPSK and QPSK. The corresponding coding rates and modulation types are listed in Table 1. In order to receive a decent quality video sequence over wireless MIMO system with good coding gain real time, we have selected convolutional code and space time block code (STBC) for the channel coding. It has been known that a transmission system with antenna diversity can achieve reliable communication over wireless channel. Antenna diversity is achieved by employing spatially separated antennas at the transmitter and/or receiver. The advantage with multiple antennas scheme is that it results in a drastic increase in the channel capacity (Foschini & Gans, 1998). Alamouti (1998) introduced an efficient scheme which involves using two transmit antennas and one receive antenna (2×1 STBC code) for a wireless communication system. Tarokh et al. (1999) generalized the Alamouti scheme with STBC code to an arbitrary number of transmit antennas. STBC codes do not in general

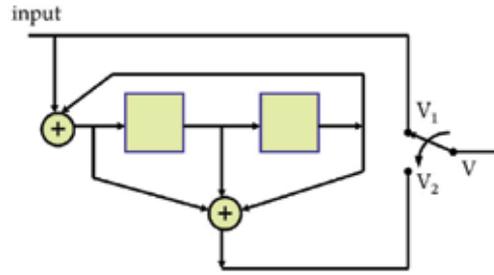


Fig. 5. Recursive systematic convolutional (RSC) encoder, coding rate = $\frac{1}{2}$

| Type | Transmission rate ($r = k/n$) | Channel coding rate, n (Convolutional) | Source coding rate, k (MPEG2) | Modulation type |
|------|---------------------------------|--|---------------------------------|-----------------|
| A | 0.9965 bit | 2/5 | 0.3986 bpp | BPSK, QPSK |
| B | 0.9988 bit | 1/2 | 0.4994 bpp | BPSK, QPSK |
| C | 0.9989 bit | 2/3 | 0.6659 bpp | BPSK, QPSK |

Table 1. Corresponding source-channel coding rate to achieve transmission rate, $r \approx 1$ bit.

provide any coding gain, therefore, it may need to be combined with an outer channel coding scheme to provide such coding gains. We then select convolutional code and space time block code for the channel coding. To simplify the analysis, we consider the simple STBC G_2 encoder (Alamouti, 1998). The input symbol vector of the STBC encoder is denoted as $S = [S(0), S(1), \dots, S(2N-1)]^T$, where N is the number of the subcarriers. Let $S_1 = [S(0), S(1), \dots, S(N-1)]^T$ and $S_2 = [S(N), S(N+1), \dots, S(2N-1)]^T$, after the STBC encoder, the generated STBC G_2 coded data symbols are

$$G_2^{STBC} = \begin{matrix} \text{antenna} \\ \left[\begin{array}{cc} S_1 & S_2 \\ -S_2^* & S_1^* \end{array} \right] \begin{matrix} \downarrow \\ \text{time} \end{matrix} \end{matrix} \quad (1)$$

It can be found that, at time T , the first antenna sends out symbol S_1 and the second antenna sends out symbol S_2 ; at time $T+1$, the first antenna sends out symbol $-S_2^*$ and the second antenna sends out symbol S_1^* . It is easy to show that the inner product of matrix G_2 is zero, which means the data within matrix G_2 are orthogonal to each other. The space diversity types applied in this proposed system are 2×2 , 2×1 , and 1×2 , respectively. Figure 6 suggests a 2×2 STBC G_2 coded diversity transmission block diagram. The received signals can be represented as follows,

$$r_1(t) = S_1 h_1 + S_2 h_2 + n_1 \quad (2)$$

$$r_2(t+\tau) = -S_2^* h_1 + S_1^* h_2 + n_2 \quad (3)$$

$$r_3(t) = S_1 h_3 + S_2 h_4 + n_3 \quad (4)$$

$$r_4(t+\tau) = -S_2^* h_3 + S_1^* h_4 + n_4 \quad (5)$$

where S_1 and S_2 are the transmitted signals (* represents complex conjugate), $h_1 \sim h_4$ are the channel fading coefficients between transmitter antennas and receiver antennas as shown in

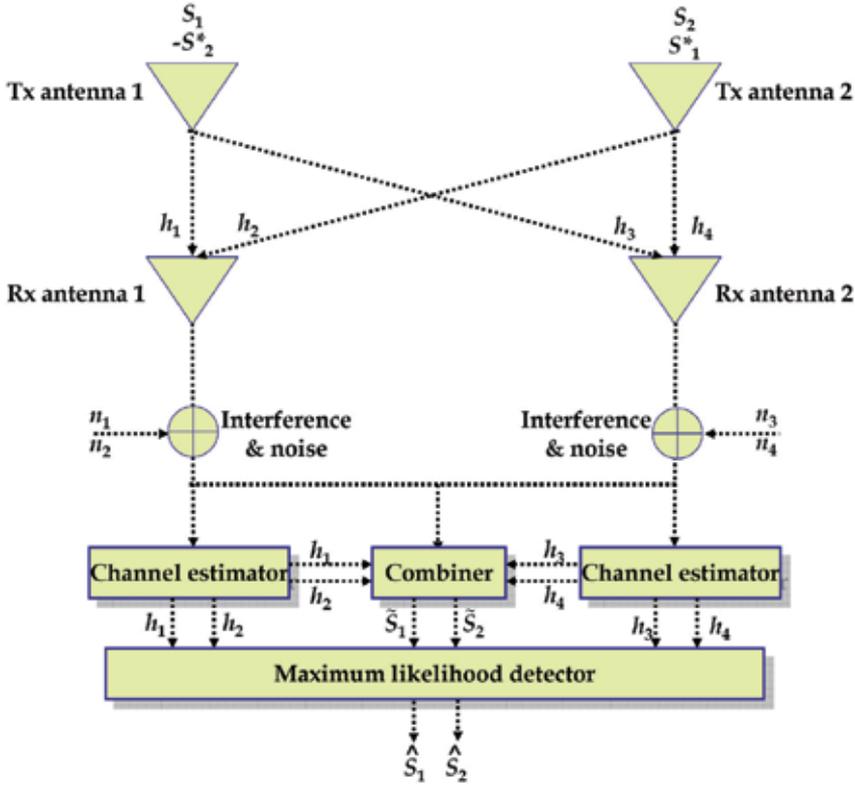


Fig. 6. A 2x2 STBC \$G_2\$ coded diversity transmission block diagram

| Antennas | Rx antenna 1 | Rx antenna 2 |
|--------------|--------------|--------------|
| Tx antenna 1 | h_1 | h_3 |
| Tx antenna 2 | h_2 | h_4 |

Table 2. Channel impulse response between transmitter and receiver

Table 2, and $n_1 \sim n_4$ are corresponding AWGN channel noise. The coefficients h_1, h_2, h_3 , and h_4 can be represented as the following,

$$h_i = \alpha_i e^{j\theta_i}, i = 1, 2, 3, 4 \quad (6)$$

Assuming that the channel impulse responses can be fully estimated, we then are able to reconstruct \tilde{S}_1 and \tilde{S}_2 with the received signals, r_1, r_2, r_3 , and r_4 by the following equations:

$$\tilde{S}_1 = h_1^* r_1 + h_2^* r_2 + h_3^* r_3 + h_4^* r_4 \quad (7)$$

$$\tilde{S}_2 = h_2^* r_1 - h_1^* r_2 + h_4^* r_3 - h_3^* r_4 \quad (8)$$

Substituting r_1, r_2, r_3 , and r_4 from equations (2) ~ (5) into equations (7) and (8), we have,

$$\tilde{S}_1 = (\alpha_1^2 + \alpha_2^2 + \alpha_3^2 + \alpha_4^2) S_1 + h_1^* n_1 + h_2^* n_2 + h_3^* n_3 + h_4^* n_4 \quad (9)$$

$$\tilde{S}_2 = (\alpha_1^2 + \alpha_2^2 + \alpha_3^2 + \alpha_4^2)S_2 - h_1n_2^* + h_2^*n_1 - h_3n_4^* + h_4^*n_3 \quad (10)$$

Finally, we can decode S_i by applying Maximum Likelihood Detector (MLD) rule if and only if:

$$\begin{aligned} (\alpha_1^2 + \alpha_2^2 + \alpha_3^2 + \alpha_4^2)|S_i|^2 - \tilde{S}_1S_i^* - \tilde{S}_1^*S_i \\ \leq (\alpha_1^2 + \alpha_2^2 + \alpha_3^2 + \alpha_4^2)|S_k|^2 - \tilde{S}_1S_k^* - \tilde{S}_1^*S_k, \quad \forall i \neq k \end{aligned} \quad (11)$$

To realize the channel coding rate effect under wireless Rayleigh fading channel with AWGN noise conditions, we have performed the experiment for 2x2 system antenna structure with three convolutional coding rates: 2/5, 1/2 and 2/3, respectively. The resulted system performance is shown in Figure 7. The system performance is improved with lower channel coding rate in the experiment. It can be found from Fig. 7, rate 2/3 convolutional coded system is with the worst bit error rate (BER) performance, where rate 2/5 convolutional coded system shown the best BER performance at the same SNR conditions. On the other side, system with lower channel coding rate (2/5 in this case) resulted in slower overall transmission rate. Therefore, if we may alternate the source coding rate corresponding to the channel coding rate, we are able to remain a consistent transmission rate which achieves channel capacity with considerable system BER performance.

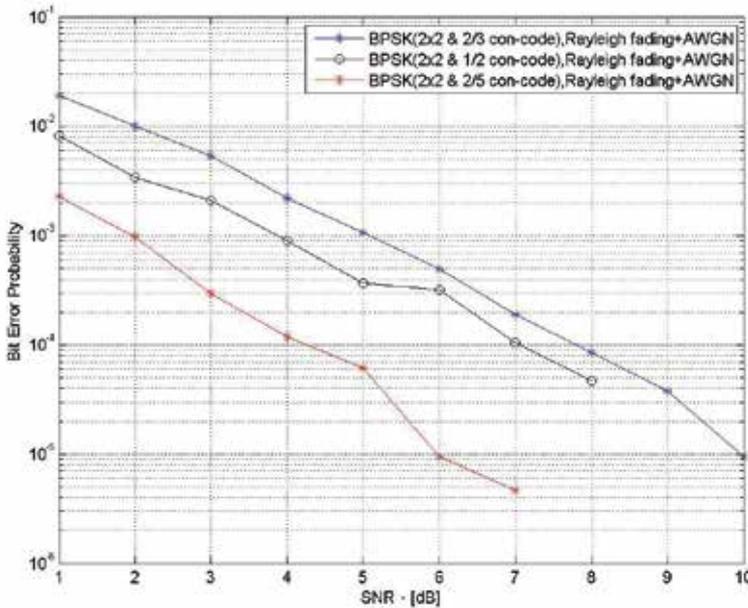


Fig. 7. Bit error rate (BER) performance of three convolutional channel coding rates under Rayleigh fading channel with AWGN noise in a 2x2 MIMO system.

2.3 Performance Control Unit (PCU)

Rates assigned to MPEG2 source coding and convolutional channel coding schemes as well as STBC space diversity selection are based on the feedback information from Performance Control Unit (PCU) under system channel capacity limitation, which ensures the given

system achieved the best performance compared to conventional systems. PCU is the key components in the adaptive system design, where we have assigned three PCU states to report the changeable overall transmission status as shown in Table 3.

| PCU state | BER after CTS state feedback | Convolutional code rate | MPEG2 code rate | No. of receiver antenna |
|-------------------|------------------------------|-------------------------|-----------------|-------------------------|
| State A: $H = 0$ | BER = 0 % | 2/3 | 0.6659 bpp | 2 |
| State B: $H = 1$ | BER \leq 20% | 1/2 | 0.4994 bpp | 2 |
| State C: $H = -1$ | BER > 20% | 2/5 | 0.3986 bpp | 1 |

Table 3. System state assignment of PCU

We adopt first-order Markov chain to describe the system states transfer (Daut & Ma, 1998). The present state is associated with the one-step adjacent state as shown in Figure 8. We have set-up three states to collaborate with the variable Rayleigh fading channel conditions. The three states is arranged to form a circular situation where the state transition is made according to Table 3, the system state assignment of PCU. On Fig. 8, H is the output status index of the PCU. We have assigned that, $H = 0$ is the "state A" index, system with good channel condition and a fast channel coding rate ($n = 2/3$) is assigned; $H = 1$ is the "state B" index, the channel is in an "OK" condition and the transmission data need more protection (channel coding rate $n = 1/2$); $H = -1$ is the "state C" index, channel condition has degraded and the channel code with good protection has to be utilized ($n = 2/5$).

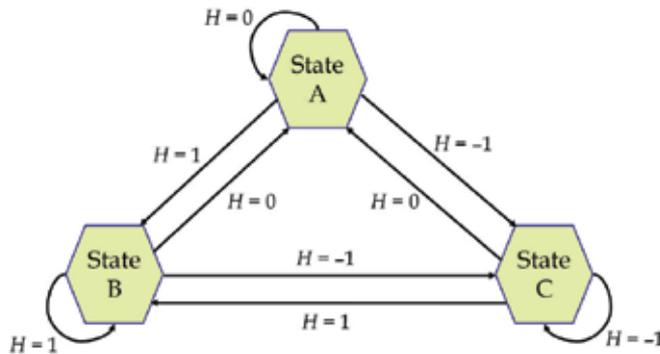


Fig. 8. System state transition diagram

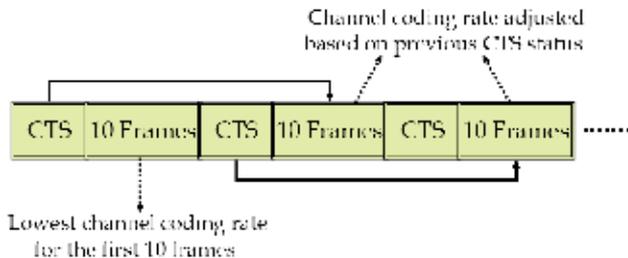


Fig. 9. Transmission rate adjustable for MPEG2 video frames

In the simulation experiment, we first send Command Testing Sequence (CTS), consisting of 10 bits stream of "1", which is attached in front of the transmitted data stream as shown in

Fig. 9. After receiver been channel decoded the received data sequence, the BER information of the CTS is fed-back to the transmitter site as a status index H to adjust the next transmission status as shown in Table 3.

3. System performance analysis

Under the channel capacity constraint, which we assumed in the simulation experiment is 1 bit/transmission; we have proposed an adaptive MPEG2 video transmission over wireless MIMO system. Based upon the feedback information index from Performance Control Unit, we adjust the compression rate of MPEG-2 video coder jointly with associated convolutional channel code rate to reach the 1-bit channel capacity limitation. We have also utilized the space diversity with 2×2 , 2×1 or 1×2 antenna configurations to obtain an accessible system performance. We have adopted three types of rate assignment as shown in Table 1. According to the feedback error rate information from test sequence as shown in Table 3, we can choose adequate joint code rate assignment (system state, Fig. 8) with suitable space diversity STBC to achieve the best system performance.

3.1 Experiment procedure

System transmission simulation is based on the system block diagram shown in Fig. 2. The experiment procedures are provided as follows, (1) capture video image streams from video camera and applied MPEG2 video coding scheme to produce a better data compression ratio and then stored as an MPEG file (*.mpg); (2) the MPEG2 coded video file is fed into a convolutional encoder, a random interleaver (size = 1024), M-PSK modulation and STBC encoder, consecutively; (3) the resulted data streams are transmitted through wireless Rayleigh fading channel with AWGN noises. For the proposed adaptive video transmission system, we have assigned three combinations of the joint source-channel code transmission rate adjusted to be nearly but not greater than 1 bit/transmission ($r \approx 1$ and $r \leq 1$) associated with proper M-PSK modulation scheme (as listed in Table 1).

The system simulation transmitted a total of 30 video frames, each transmission (10 video frames) will be added up a 10 bits Command Testing Sequence before transmitted over wireless channel consisting of Rayleigh fading and AWGN noises. With appropriate selection of receiver antenna numbers (as shown in Table 3), we have gained space diversity to improve system performance. At the receiver end, de-modulation, de-interleaving, and de-coding procedures are provided to reconstructed MPEG video. After received the feedback error rate information of the CTS sequence, a status index H of the PCU is fed-back to the transmitter site to adjust the next 10 video frames transmission status as shown in Table 3.

3.2 Simulation results

The bit error rate (BER) performance versus SNR for different space diversity schemes over Rayleigh fading and AWGN channel is shown in Figure 10. It is assumed that the amplitudes of fading noise from each transmitter antenna to each receiver antenna are mutually uncorrelated Rayleigh distributed and that the average signal power at each receiver antenna from each transmitter antenna is the same. Furthermore, we assumed that the receiver has perfect knowledge of the channel conditions. The simulation results of two transmitter antennas and two receiver antennas (2×2) STBC coded system shows the best

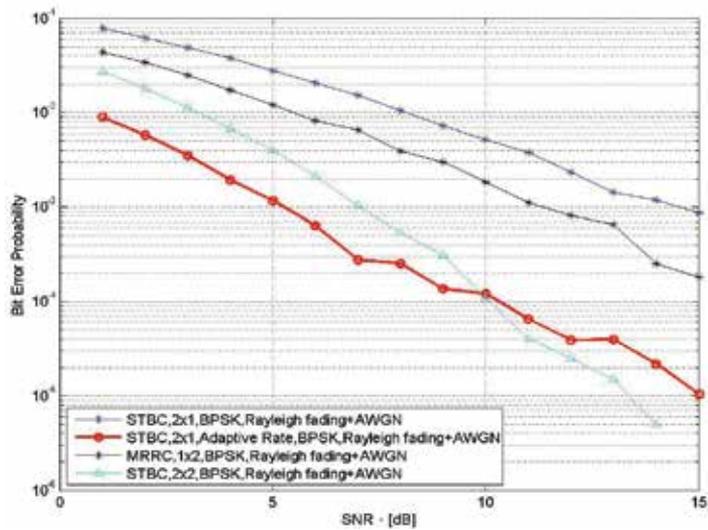


Fig. 10. The BER performance comparison of STBC systems (2/3 convolutional coded) and the proposed adaptive system (with 2×1 antennas structure) over Rayleigh fading channel.

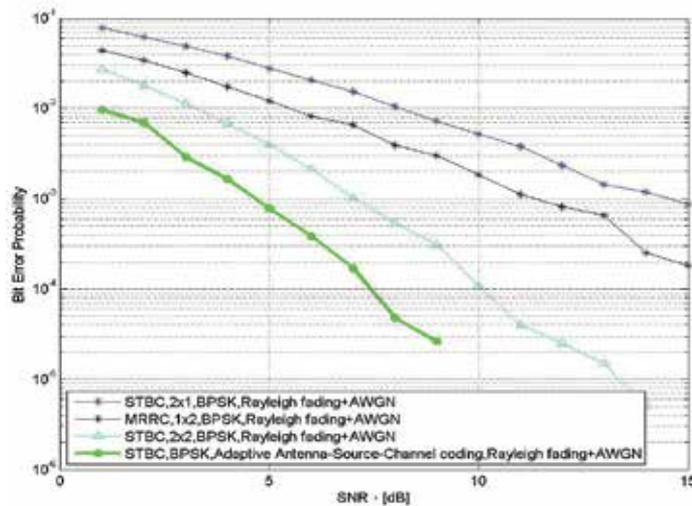


Fig. 11. The BER performance comparison of STBC systems and the proposed adaptive system over Rayleigh fading channel (the number of receiver antenna is adaptive).

BER performance at higher SNR values (> 10 dB), while the worst performance goes to the 2x1 STBC coded system. The proposed adaptive coding system with 2x1 (in this case, number of antenna is fixed) space diversity can improve the system performance especially in lower SNR (< 10 dB) situation, and with close performance as the 2x2 STBC coded system in SNR > 10 dB environment. The BER system performance can be improved more with adaptive receiver antenna numbers (as given in Table 2) of the proposed scheme as shown in Figure 11. It is about 2.5 dB SNR gain at the same BER condition for the proposed system. We have extended the total transmitted video frames to 100 for the experiment, each transmission (10 video frames) is added up a 10 bits CTS before transmitted over wireless

Rayleigh fading channel. From Figure 12, we have noticed that the proposed adaptive coding system with 2x2 (fixed antenna numbers) space diversity is outperformed the conventional systems: 2x2 STBC coded scheme, 2x1 STBC coded scheme, and 1x2 maximal ratio receive combining (MRRC) scheme. Figure 13 shows the reconstructed video frames of the proposed PCU controlled adaptive system (in Fig. 10, SNR = 9 dB, BER $\approx 7 \times 10^{-4}$).

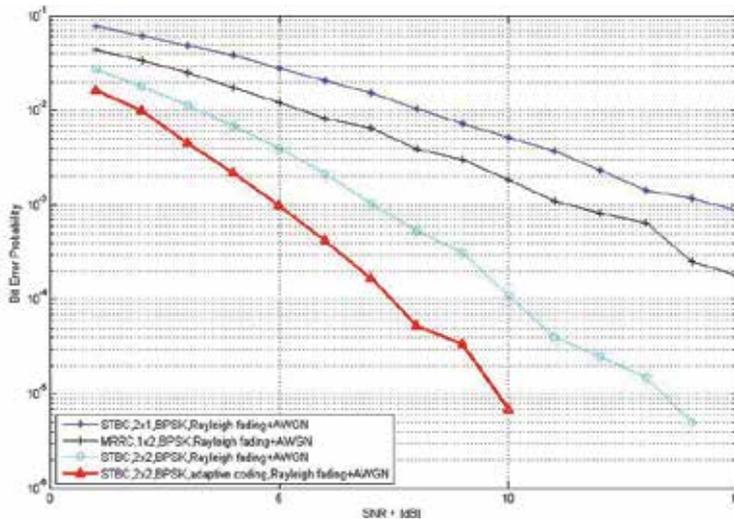


Fig. 12. The BER performance comparison of STBC systems (2/3 convolutional coded) and the proposed adaptive system (with 2x2 antennas structure) over Rayleigh fading channel. (100 video frames transmitted)

4. Conclusions

Video transmission over MIMO systems offers numerous new research opportunities. Lots of new researches are originated from the fundamental change in data transmission from single link to multiple simultaneous links in the MIMO systems. In this study, we applied joint source-channel coding with modulation scheme to design an adaptive video transmission over wireless MIMO system. The bit rates of MPEG2 video can be adaptive to associate with the convolutional channel codes and space-time block code (STBC) under the channel capacity constraint consideration. In order to be consistent with the channel capacity constraint (which is set to be 1 bit/transmission), there are three rate combination types of the joint source-channel coding algorithm as shown in Table 1. From the simulation results, we found that the space diversity and the channel code rate both are important factors influenced reconstructed video quality. The simulation results shows that two transmitter antennas and two receiver antennas (2x2) STBC coded system demonstrates the best BER performance, while the worst performance goes to the 2x1 STBC coded system. The system performance is also improved with lower channel coding rate in the experiment. It can be found from the simulation, rate 2/3 convolutional coded system is with the worst bit error rate performance, where rate 2/5 convolutional coded system shown the best BER performance at the same SNR conditions.

In this study, the proposed adaptive system can choose an adequate transmission rate and the number of receiver antennas based on the channel condition. With the feedback BER



Fig. 13. The reconstructed video frames of the proposed PCU controlled adaptive system.

information provided by the performance control unit (PCU), the proposed system is able to choose an appropriate source-channel rate to transmit video. Therefore, the transmitted video quality may keep at an almost uniform level over a Rayleigh fading channel condition. The study has ensured the proposed system achieved a better BER performance compared to conventional systems.

5. References

- Alamouti, S. (1998). A Simple Transmit Diversity Technique for Wireless Communications, *IEEE Journal Selected Areas on Communications*, Vol. 16, 1451-1458
- Burr, A. (2003). Capacity Bounds and Estimates for the Finite Scatterers MIMO Wireless Channel, *IEEE Journal on Selected Areas in Communications*, Vol. 21, 812-818
- Chen, C. & He, Z. (2006). Signal Processing Challenges in Next Generation Multimedia Communications, *China Communications*, Vol. 4 (5), 20-29
- Daut, D. & Ma, W. (1998). Integrated Design of Digital Communication Systems for Fixed Channel Conditions, *Proceedings of International Telecommunication Symposium*, Vol. I, 10-15
- Daut, D. & Modestino, J. (1983). Two-Dimensional DPCM Image Transmission over Fading Channel, *IEEE Transactions on Communications*, Vol. 31 (3), 315-328
- Driessen, P. & Foschini, G. (1999). On the Capacity Formula for Multiple Input-Multiple Output Wireless Channels: a Geometric Interpretation, *IEEE Transaction on Communications*, Vol. 47 (2), 173-176
- Foschini, G. & Gans, M. (1998). On the Limits of Wireless Communications in a Fading Environment when using Multiple Antenna, *Wireless Personal Communications*, Vol. 6 (3), 311-335
- Proakis, J. (2001). *Digital Communications*, Fourth Edition. ISBN 0-07-232111-3, McGraw-Hill, New York, USA.
- Sikora, T. (1997). MPEG Digital Video-Coding Standards, *IEEE Signal Processing Magazine*, Vol. 14, 82-100
- Song, D. & Chen, C. (2008). Maximum-Throughput Delivery of SVC-based Video over MIMO Systems with Time-Varying Channel Capacity, *Journal of Visual Communication and Image Representation*, Vol. 19, 520-528
- Song, D. & Chen, C. (2007). Scalable H.264/AVC Video Transmission over MIMO Wireless Systems with Adaptive Channel Selection Based on Partial Channel Information, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 17 (9), 1218-1226
- Tarokh, V.; Jafarkhani, H. & Calderbank, A. (1999). Space-Time Block Codes from Orthogonal Designs. *IEEE Transactions on Information Theory*, Vol. 45 (5), 1456-1467
- Wolniansky, P. ; Foschini, G.; Golden, G. & Valenzuela, R. (1998). V-BLAST: An Architecture for Realizing Very High Data Rates over the Rich-Scattering Wireless Channel, *Proceedings of IEEE International Symposium on Signals, Systems and Electronics*, 295-300
- Yin, L., Chen, W., Lu, J. & Chen, C. (2004). Joint Source Channel Decoding for MPEG-2 Video Transmission, *Proceedings of SPIE Conference on Multimedia Systems and Applications*, Vol. 5600, 128-136
- Zeng, X. & Ghayeb, A. (2006). Performance Bounds for Combined Channel Coding and Space-Time Block Coding with Receive Antenna Selection, *IEEE Transactions on Vehicular Technology*, Vol. 55 (4), 1441-1446

Transmission Optimization of Digital Compressed Video in Wireless Systems

Pietro Camarda and Domenico Striccoli

*Politecnico di Bari
Italy*

1. Introduction

Digital multimedia transmission is one of the most significant technological challenges of our time. In the last few years, new innovative services have been developed allowing users to benefit of high quality multimedia contents through a wide variety of transmission standards (e.g., Digital Video Broadcasting, 3rd Generation cellular networks, etc.). The modern video compression techniques, like the MPEG or H.264/AVC standards (Mitchell et al., 1996), allow the final users to experience high quality video decoding with a consistent bandwidth saving. At the same time, the significant progress in wireless communication networks with high Quality of Service (QoS) guarantees has brought to a development of innovative multimedia services, whose flexibility can satisfy the mobility requirements peculiar to the ever growing number of wireless terminal users. This process has been also encouraged by an integration between fixed and mobile networks where final users can combine telecommunications, information technology and entertainment services offered from various operators. The most highlighting and well known example is the Universal Mobile Telecommunication System (UMTS) network, where streaming servers providing video are typically located in wired packet-data or telephone networks. Video data cross the core network and are then broadcasted or multicasted to mobile users through the wireless Radio Access Network. The development of value added services on cellular networks like UMTS are regulated by the 3rd Generation Partnership Project (3GPP), that defines the standard for 3rd generation cellular networks (Holma & Toskala, 2004), dedicating also a special work group for streaming video (3GPP TSG-SA; 3GPP TS 26.234; 3GPP TR 26.937). Another example of wired-wireless integrated network infrastructure is the Digital Video Broadcasting for Handheld terminals (DVB-H) system, where multimedia contents are transmitted through the IP network, exploiting the same DVB Terrestrial (DVB-T) infrastructure (ETSI TR 101 190; ETSI EN 300 744; ETSI TR 102 377), and are received on several kind of terminals like smartphones, Personal Digital Assistants (PDAs), notebooks, etc.

Video transmission over wireless terminals has introduced several new issues, peculiar to these environments, that need to be investigated, like Doppler and multipath noises with consequent signal degradation, handover, etc. These issues become also more critical in the case of video streaming, where frame losses should be minimized to keep a high quality video decoding. Power saving has to be maximized for an efficient multimedia reception

because of terminals batteries with limited capacity. Another important aspect is that compressed video is generally characterized by a bit rate highly variable in time. This makes the optimal allocation of bandwidth resources very difficult: a static bandwidth assignment with data highly fluctuating in time could easily bring to an underutilized channel, or consistent frame losses.

In this chapter we will face the problem of the dynamic optimization of Variable Bit Rate (VBR) video transmission in wireless environments. Schedule at transmission side is varied during stream running according to the varying system conditions. The goal is to maximize the network channel utilization and reduce lost bits, to perform a continuous video playback on wireless terminals without any quality degradation.

Two different, but structurally similar, environments will be analyzed: the DVB-H system and the UMTS network. Transmission plans at server side will be proposed that take into account not only the VBR data, but also the channel bandwidth and the buffering capacity of mobile terminals. In the UMTS scenario data transmission will be varied by taking into account the user interactivity. The proposed algorithms are compared with the transmission techniques already known by literature and guidelines; their effectiveness will be validated in several simulation scenarios.

2. Video transmission frameworks

2.1 Digital Video Broadcasting for Handheld terminals

In the recent years, several aspects have contributed to the exponential growth of Digital Video Broadcasting and innovative services delivered through wireless networks: an improved digital video quality with consistent bandwidth saving thanks to the MPEG video compression standard; a better allocation of frequency resources, the possibility of user interactivity, and the reception of high-quality services during terminal motion. In this context, DVB for Handheld terminals (DVB-H) is able to carry multimedia services through digital terrestrial broadcasting networks. Since the Terrestrial DVB (DVB-T) infrastructure is able to serve both fixed and mobile terminals, DVB-T and DVB-H systems share almost the same common physical layer of the ISO-OSI protocol stack; nevertheless, DVB-H introduces additional features in the Physical Layer, Data Link Layer and Network Layer of the stack (ISO/IEC 7498-1), as represented in Fig. 1. This system is designed for VHF and UHF frequency bands, in the same spectrum assigned to analogue TV.

At Physical Layer, single channels can occupy 6MHz, 7MHz or 8MHz bandwidths. To multiplex a wide number of signals the Orthogonal Frequency Division Multiplexing (OFDM) is adopted, because of its high efficiency in rejecting the multipath noise typical of wireless links. DVB-T allows two different transmission modes: the 2K and 8K. They refer to the number of OFDM subcarriers multiplexed in a "symbol" for transmission. The 2K mode is suitable for small Single Frequency Networks (SFNs). In the SFNs all transmitters carry the signal in the same frequency range, and in small SFNs there is a reduced distance among transmitters. The 8K mode is adopted for small, medium and large SFN networks. In addition, DVB-H allows also the 4K mode. It can be used both for single transmitter operation and for small and medium SFNs, providing also a higher robustness towards noise and allowing very high speed reception.

Each OFDM subcarrier is modulated with the Quadrature Amplitude Modulation (QAM) or Quaternary Phase Shift Keying (QPSK). The transmission system is hierarchic since two

different MPEG streams are multiplexed in the modulated signal: a High Priority (HP) stream and a Low Priority (LP) stream.

Video data are grouped into fixed-length packets of 188 bytes, including 1 byte of synchronization header. The 187 information bytes are then coded through a proper pseudo-random bit redistribution that minimizes the energy dispersion (the Adaptation Energy Dispersion). In the Outer Coder a Error Correction Code is applied; the adopted code is the Reed-Solomon that adds 16 parity bytes generating packets of 204 bytes. The so obtained data are then interleaved in the Outer Interleaver; a bit coding in the Inner Coder allows a further protection against transmission errors in the aerial channel. A further interleaving performed by the Inner Interleaver allows the correct associations among bits that will then be mapped by the Mapper on the OFDM subcarriers.

The Transmission Parameter Signalling (TPS) information is added to the OFDM symbol. It consists of an adding number of OFDM subcarriers conveying only signalling information about modulation techniques, code rates, transmission modes, etc.; this operation is called Frame Adaptation. (ETSI EN 300 744). The OFDM symbols are then organized in frames and a guard time interval is set between consecutive symbols. The so packetized data are then sent to a D/A converter and then to the terrestrial channel.

At the Data Link Layer, the data broadcast service shall insert data directly in the payload of MPEG-2 Transport Stream packets (ISO/IEC 13818-1). DVB standard specifies four methods for data broadcasting: Data Piping, Data Streaming, Multi-Protocol Encapsulation (MPE) and Data Carousel, together with additional Program Specific Information (PSI) and Service Information (SI) that are proper data descriptors (ETSI EN 301 192). Specifically, MPE is well suited to the delivery of streaming services. MPE supports the delivery of other protocols, giving more flexibility, and is also suitable to implement additional features specific to DVB-H environments, that is, Multi Protocol Encapsulation - Forward Error Correction (MPE-FEC) and Time Slicing.

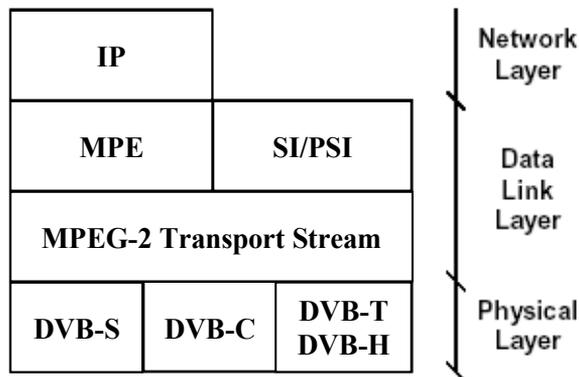


Fig. 1. Protocol stack (layers 1-3) for DVB-H transmission

MPE-FEC is introduced to improve the system robustness towards Doppler noise and impulse interferences. To this aim, an additional level of error correction is introduced at the MPE layer. By calculating parity information and sending it in MPE-FEC sections, error-free datagrams can be received also in very bad reception condition (ETSI EN 302 304).

Time Slicing is instead introduced to improve terminal power saving and manage handover. At network layer, IP protocol is adopted. The DVB-H payload consists of IP datagrams incapsulated into MPE sections and multiplexed together with classical MPEG-2 services (ETSI EN 301 192). An example of a DVB-H system for IP service transmission is illustrated in Fig. 2, where DVB-H services are multiplexed with traditional MPEG-2 services.

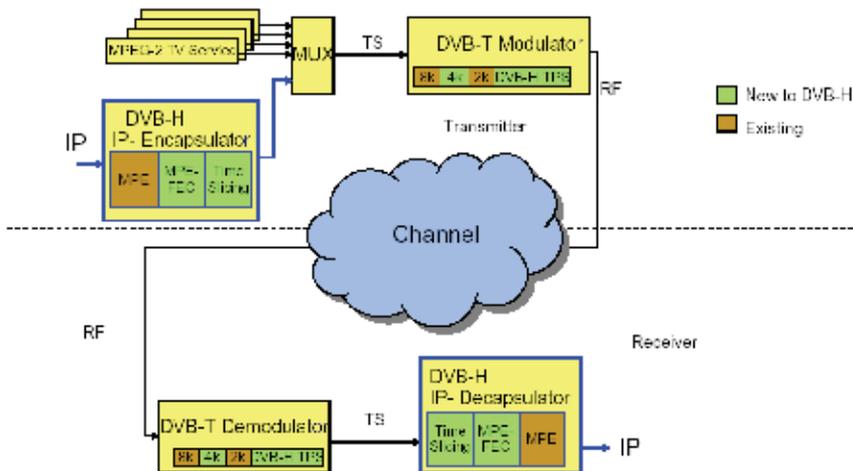


Fig. 2. Conceptual DVB-H system structure

2.1.1 Time slicing

Time Slicing is a transmission technique thought to improve mobile terminal performance in terms of energy saving and handover. Service data are transmitted in “packets” or “bursts” periodically repeating in time with the classical Time Division Multiplexing (TDM) technique. Service Bursts are interspaced by “off-times”, i.e., time intervals between two consecutive bursts of the same service, where no service data are decoded but data of other multiplexed services are transmitted. The energy saving is achieved because receiver remains active only for a fraction of time, during the reception of the chosen service, and switches off during that service off-times. Furthermore, off-times can be exploited to manage handover, that is, cell changes, through proper signal monitoring, without service interruptions. Stream decoding should always be continuous and without losses, so the transmitted bit rate in a burst must be consistently higher than the average bit rate required for the classical DVB-T transmission. Fig. 3 illustrates the comparison between DVB-T and DVB-H transmission of four services.

The main parameters for the single time-sliced service are illustrated in Fig. 4. The Burst Size (BS) represents the total Network Layer bits transmitted in a burst, included the MPE-FEC and Cyclic Redundancy Check code (CRC-32) headers. The Burst Bitrate (BB) is the constant bit rate of the time-sliced stream during the burst transmission. The Constant Bitrate (CB) is the average bit rate required by the stream not time sliced, as would be transmitted in classical DVB-T systems. The Burst Duration (BD) is the time interval from the burst beginning to its end. The Off-time (Ot) is instead the time interval between two consecutive bursts.

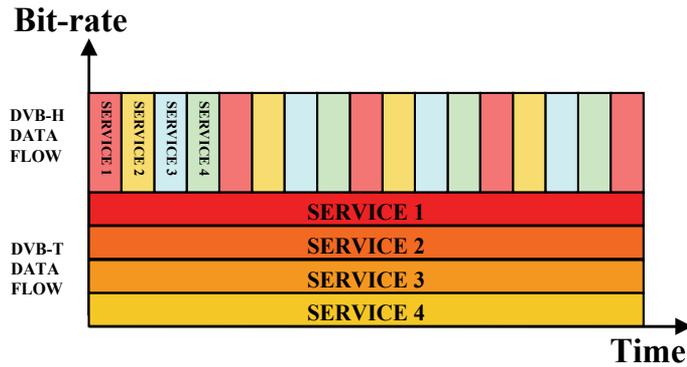


Fig. 3. Comparison between DVB-T and DVB-H transmission

In the sequel, we will define the Burst Cycle (BC) as the time interval between the beginning of two burst of the same service. For each service it holds that $BC = BD + Ot$.

The time slicing implementation allows a consistent power saving at receiving side since the receiver remains inactive during off-times, and the power consumption is reduced. Furthermore, off-times can be exploited to perform handover without any service interruption.

Continuous data decoding is theoretically always guaranteed by receiver data buffering. The burst size is in fact buffered in the client memory during burst durations, and consumed during the subsequent Off-times. Obviously, sufficient buffering is needed at receiving side for continuous and lossless decoding; this condition can be reached by properly regulating the main burst parameters (burst duration and burst bitrate).

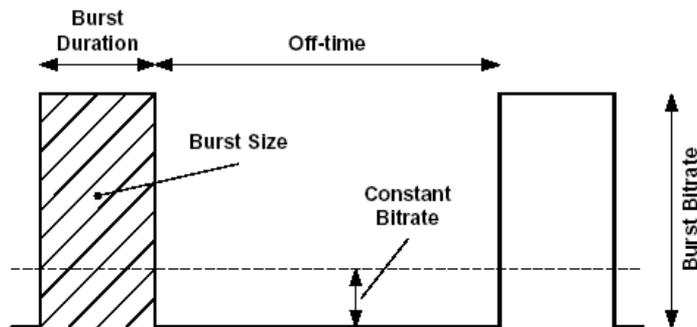


Fig. 4. Time Slicing parameters for the single service

2.2 UMTS networks

Another scenario of great interest in video transmission context is the stream delivery over 3rd generation wireless networks. The 3G networks bring some important enhancements to the previous 2G cellular networks, improving at the same time the already existing services. Specifically, it should:

- Support broadband services, both point-to-multipoint and point-to-point;
- Allow new low-cost terminals of small weight and size, simple to use for the users;
- Support different QoS levels for the wide variety of available services;

- Support an efficient network resource assignment;
- Provide a more flexible rating, depending from the session duration, the transferred data and the desired QoS;
- Implement a wider set of bit rates, depending from the specific service and mobility issues.

The UMTS network is thought as a group of logical networks, each one accomplishing specific functions, as illustrated in Fig. 5 (Uskela 2003).

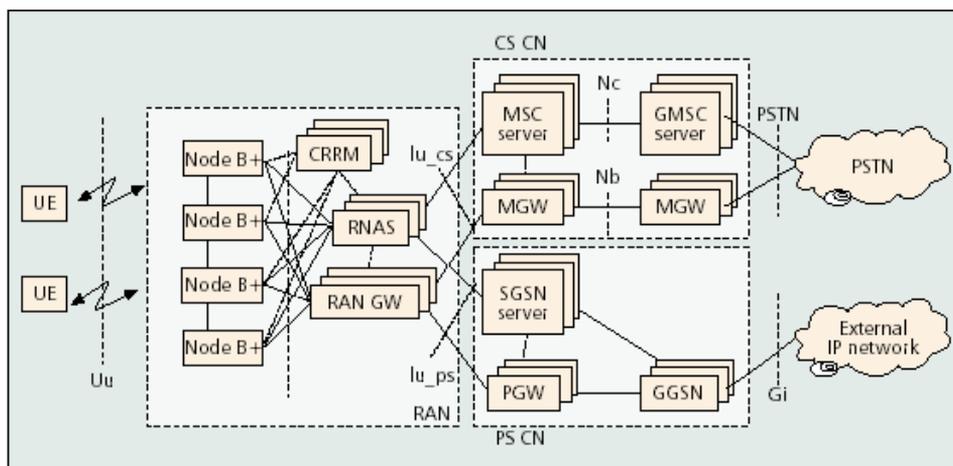


Fig. 5. The 3G network architecture

The Core Network (CN) is the part of the UMTS network that provides services to final users; it can be connected to different types of networks supporting different communication protocols. The CN is composed by the Circuit Switched CN (CS CN) and a Packet Switched CN (PS CN). Regarding the CS CN, the Mobile services Switching Centre (MSC) switches the CS transactions and manages the user mobility. It is interconnected to external networks (like for example the PSTN networks in Fig. 5) through the Gateway MSC (GMSC). The MSC is subdivided into MSC server and Media GateWay (MGW), where all user data are managed. A MSC server can manage more MGWs, allowing a higher network scalability when new implemented services increase data flow. In fact, whenever this happens, it is sufficient to increment the MGWs number.

The PS CN is composed by two elements: the Serving GPRS Support Node (SGSN), that manages the different sessions (including various QoS requirements) and mobility, and the Gateway GPRS Support Node (GGSN), that performs, among the other things, the QoS mapping among different networks, the packet filtering and the connection with other packet switching external networks (such as Internet). The streaming server providing multimedia services can reside either in the external packet data network, or just before the GGSN.

The PS CN is connected to the Radio Access Network (RAN) through the RAN GateWays (RAN GW) and the RAN Access Server (RNAS). The RAN GW is a routing point for the user traffic between CN and UTRAN. All radio resource management functionality is pooled in the Common Radio Resource Manager (CRRM). It is thus possible to provide the radio resource management network-wide to all radio technologies, which allows an efficient utilization optimization. The RNAS main task is to provide an interface to the CN.

RAN GWs and RNAS form the Radio Network Controller (RNC). Each SGSN can manage hundreds of RNCs.

All processing related to the radio interface of the RNC is relocated into the Node B+, whose main functions are, among others: macro diversity control, handover related functions, power control and frame scheduling. Data are provided to the User Equipment (UE), which is the radio terminal utilized by the user to exploit the UMTS network services.

2.2.1 Feedback information through RTCP

To guarantee a good quality multimedia streaming while minimizing losses at client side, there is a need for scheduling at transmission side. As previously explained, when the network or the client conditions vary during a session, bad effect can appear at client side due to the buffer overflows and underflows events. A very good way for scheduling optimization at transmission side is that the streaming server has knowledge of the "status" of the network and the client terminal. This information is performed through the Real Time Control Protocol (RTCP) (3GPP TS 26.234), that carries control information between the media player on the user equipment and the streaming server. RTCP packets carry the main statistics on the media stream, that can be exploited by the streaming server to regulate the transmission rate whenever network congestion and/or losses at receiving side occur. RTCP main role is to provide a feedback information on the quality of the stream delivery.

Streaming content is instead carried through the Real-time Transfer Protocol (RTP) (Schulzrinne et al., 2003), and the streaming sessions are administered by the Real Time Streaming Protocol (RTSP), an application protocol that allows external actions (pause, record, rewind, etc.) on stream data.

RTCP can be a solution for the wireless multimedia streaming over IP networks, since it could help improving the quality of the delivered data, that depends on the variable conditions of the wireless links and the user equipments limited amount of memory. To this aim, the 3GPP Packet-switched Streaming Service (3GPP PSS) specifications give a guideline for transmission rate optimization through the RTCP feedback (3GPP TS 26.234). RTCP packets are usually sent with a periodicity of 5 seconds.

The streaming standard introduces the following RTCP parameters:

- *HTSN (Highest Transmitted Sequence Number)*, that is the sequence number of the last packet sent by the server;
- *HRSN (Highest Received Sequence Number)*, that is the sequence number of the last packet arrived at the client buffer;
- *NSN (Next Sequence Number)*, that is the sequence number of the next packet to be decoded by the client.

The 3GPP TS 26.234 standard considers also the RTCP packet evolution, called the Next Application Data Unit (NADU) packet, whose structure is illustrated in Fig. 6.

The NADU packet fields are:

- *SSRC*, that is the source identifier of the stream this packet belongs to;
- *PD (Playout Delay)*, representing the time interval between two consecutive reports (in milliseconds);
- *NSN (Next Sequence Number)*, the sequence number of the next packet to be decoded;
- *NUN (Next Unit Number)*, the next video frame to be decoded;
- *FBS (Free Buffer Space)*, representing the free residual buffer space at receiving side (in multiple of 64 bytes);
- *Reserved*; these bits are not used.



Fig. 6. A NADU packet

Through a NADU packet a high number of information can be derived, such as the number of packets that reached the client, the number of packets stored in the client buffer and decoded by the client. Fig. 7 clearly illustrates these parameters.

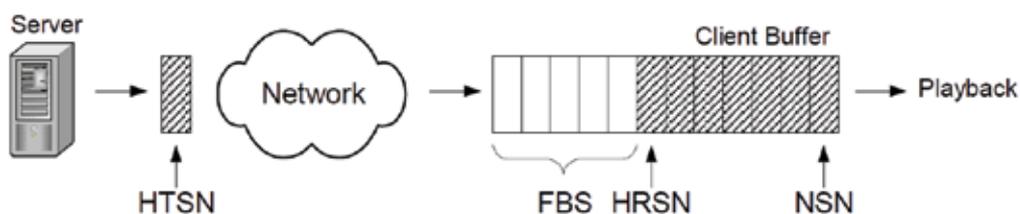


Fig. 7. Feedback information parameters

Packets with a sequence number less than NSN have already been decoded. If the streaming server knows the size of each transmitted packet, the HRSN and NSN parameters provide the occupancy level of the client buffer. It can be noted that this parameter can also be derived by the FBS information if the client sent the buffer size information to server during the connection setup.

3. Principles of video compression

The need of compressing digital video has been felt since 1988, when the standardization effort in this direction was initiated by the Moving Picture Experts Group (MPEG), a group formed under the auspices of the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC). A sequence of pictures with accompanying sound track can occupy a vast amount of storage space when represented in digital form, and also needs a huge amount of channel bandwidth to be transmitted without any quality degradation. This amount of storage and bandwidth can be greatly reduced through specific video compression techniques. In this section we only give a brief overview of the MPEG (further distinguished in the MPEG-1, MPEG-2 and MPEG-4) and H.264/AVC compression techniques, actually the most widely adopted standards.

MPEG video compression is applied to temporally consecutive images (the video frames) of a multimedia stream, exploiting the concept of the similarities between temporally adjacent frames. In fact, except for the special case of a scene change, these pictures tend to be quite similar from one to the next. Intuitively, a compression system should take advantage of this similarity. To this aim, each video sequence is divided into one or more Groups of Pictures (GoP), each one composed of one or more pictures of three different types: I-pictures (Intra-coded pictures), P-pictures (Predictive-coded pictures) and B-pictures (Bidirectionally predictive-coded pictures). Specifically, I-pictures are coded independently, entirely without reference to other pictures. P-pictures obtain predictions from temporally preceding

I-pictures or P-pictures in the sequence, whereas B-pictures obtain predictions from the nearest preceding and/or upcoming I- pictures or P-pictures in the sequence. An example of GoP is represented in Fig. 8. On average, P-pictures have a double size if compared with B-pictures, and size 1/3 if compared with I-pictures (Mitchell et al., 1996). Let us note that MPEG sometimes uses information from future pictures in the sequence; so the order in which compressed pictures are found in the bitstream, called “coding order”, is not the same as the “display order”, the order in which pictures are presented to a viewer.

The compression efficiency is quantified by the compression ratio, defined as the ratio between the original uncompressed image and the compressed one. The compression ratio depends on the compression level and the frame complexity: a more detailed image will have a lower compression ratio and a higher number of bits, than a less detailed image, both coded with the same compression level.

Each frame is subdivided in groups of samples of luminance (to describe grayscale intensity) and chrominance (to describe colors), called *macroblocks* that are then grouped together in *slices*. Each macroblock is transformed into a weighted sum of spatial frequencies through the Discrete Cosine Transform (DCT) to enhance the compression degree. Further compression is then reached through Motion Compensation (MC) and Motion Estimation techniques. Motion Compensation is introduced because if there is motion in the sequence, a better prediction is often obtained by coding differences relative to areas that are shifted with respect to the area being coded. Motion Estimation is simply the process of determining the motion vectors in the encoder, to describe direction and the amount of motion of the macroblocks (Mitchell et al., 1996).

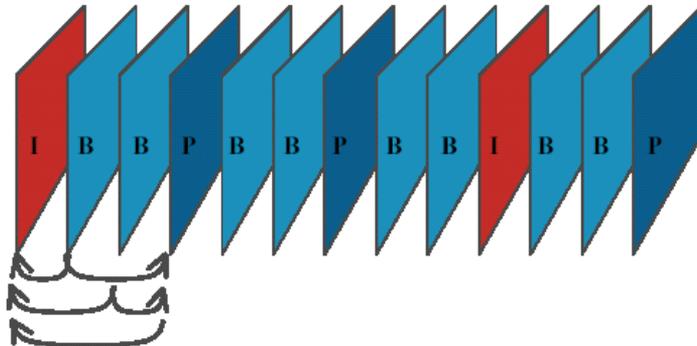


Fig. 8. A Group of Pictures in display order.

Globally, all MPEG standards are based on these concepts. More specifically, MPEG-1 is designed for mean bit rates of 1.5 Mbps, with a video quality comparable to VHS and a frame rate of 25 frames per second (fps) for PAL and 30 fps for NTSC.

MPEG-2 is an extension of MPEG-1. It has been thought for interlaced video coding and utilized for transmission of Standard Definition (SD) and High Definition (HD) TV signals over satellite, cable and terrestrial broadcasting, and the storage of high quality SD video signals onto DVDs (Puri et al., 2004). If compared with MPEG-1, MPEG-2 has a lower compression ratio and a higher mean bit rate.

MPEG-4 was intended for new multimedia applications and services such as interactive TV, internet video etc. (ISO/IEC 14496-2, 2000). It is a living standard, with new parts added continuously as and when technology exists to address evolving applications. The

significant advances in core video standard were achieved on the capability of coding video objects, while at the same time, improving coding efficiency at the expense of a modest increase in complexity.

The H.264/AVC standard (ISO/IEC JTC 1, 2003) is the new state of the art of video coding. It promises significantly higher compression than earlier standards. It is also called "MPEG-4 Advanced Video Coding (AVC)". Since the standard is the result of collaborative effort of the VCEG and MPEG standards committees, it is informally referred to as Joint Video Team (JVT) standard as well. The standard achieves clearly higher compression efficiency, up to a factor of two over the MPEG-2 video standard (Horowitz et al., 2003). The increase in compression efficiency comes at the cost of substantial increase in complexity. Moreover, H.264/AVC includes a high number of application-dependent profiles, ranging from low bit rate to very high bit rate applications. The resulting complexity depends on the profile implemented. Like MPEG-2, a H.264/AVC picture is partitioned into fixed-size macroblocks and split into slices. Each slice can be coded by using I, P, B, SP and SI frames. The first three are very similar to those in previous standards, even if the concept of B slices is generalized in H.264 when compared with prior video coding standards (Flierl & Girod, 2003). SP slices aim at efficient switching between different versions of the same video sequence whereas SI slices aim at random access and error recovery.

All the samples of a macroblock are either spatially or temporally predicted. There are also many new features and possibilities in H.264 for prediction, with several types of intra coding supported (Wiegand, 2002). H.264 standard is also more flexible in the selection of Motion Compensation block sizes and shapes than any previous standard (ISO/IEC JTC 1, 2003). The accuracy of Motion Compensation is in units of one quarter of the distance between luminance samples. Prediction values are obtained by applying specific filters horizontally and vertically in the image. The motion vector components are then differentially coded using either median or directional prediction from neighboring blocks.

The H.264 supports also multi-picture motion-compensated prediction (Flierl et al., 2001), in which more than one previously coded picture can be used as reference for Motion Compensation prediction. This new feature requires both encoder and decoder to store the reference pictures used for inter prediction in a multi-picture buffer. Multiple reference pictures not only contribute to the improvement of the compression efficiency, but also to error recovery.

4. Scheduling video transmission

From Section 3 we know that MPEG or H.264 compressed video is characterized by a high bit rate variability due to I, P or B frames, and the scene complexity in terms of details and the amount of motion. Frames are transmitted with a constant frame rate (25 fps for PAL and 30 fps for NTSC). For this reason, these kind of videos are called Variable Bit Rate (VBR) videos. The VBR video transmission introduces a consistent complexity degree for resource assignment, to keep a high QoS degree. A lossless transmission should require a bandwidth assigned to the single video flow equal to its peak rate. Nevertheless, this assignment surely would bring to a bandwidth waste for almost all the video duration, because of the high video bit rate variability (Zhang et al., 1997).

To reduce the high bit rate fluctuation typical of VBR streams, video smoothing techniques have been introduced for video transmission from a server to a client through a network, through the system model illustrated in Fig. 9. Smoothing algorithms exploit client buffering

capabilities to determine a “smooth” rate transmission schedule, while ensuring that the client buffer neither overflows nor underflows, and achieving significant reduction in rate variability.

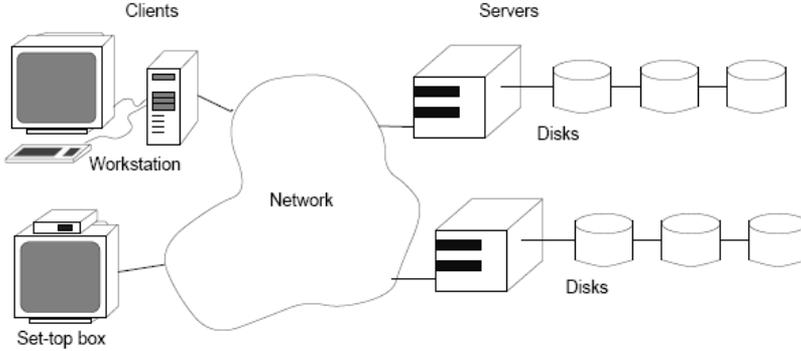


Fig. 9. The server-client model for video delivery.

The basic principle of video smoothing is the “work-ahead”, that is, the transmission of video frames ahead of their playback time. As described in (Salehi et al., 1998) a VBR video stream is composed by N video frames, each of them of size f_i bytes ($1 \leq i \leq N$). At server side, data are scheduled according to the specific algorithm. At client side, smoothed video data enter the buffer and the original unsmoothed video frame sequence leaves it for decoding and playout. Let us now consider the client buffer model in the k^{th} discrete frame time. A frame time is defined as the time to decode a frame (1/25 s for PAL and 1/30 s for NTSC), and is assumed as the basic time unit. Two curves are built:

$$D(k) = \sum_{i=1}^k f_i \quad (1)$$

$$B(k) = b + \sum_{i=1}^k f_i = D(k) + b \quad (2)$$

representing, respectively, the cumulative data consumed by the client in k , and the maximum amount of data that can be received by the client in k without overflowing the buffer. It derives that $D(k)$ is the underflow curve, whereas $B(k)$ is the overflow curve. They are both non decreasing curves. A feasible transmission plan $S(k)$ must verify the condition:

$$D(k) \leq S(k) = \sum_{i=1}^k s_i \leq B(k) \quad (3)$$

to guarantee a lossless transmission. $s(i)$ represents the smoothed stream bit rate in the i^{th} frame time. The smoothed stream transmission plan will result in a number of CBR segments, and the cumulative smoothed transmission plan is given by a monotonically increasing and piecewise linear path $S(k)$ that lies between $D(k)$ and $B(k)$ (Salehi et al., 1998).

Different types of smoothing algorithms have been implemented (Feng & Rexford, 1999). They are applied to stored video traffic, where all source video data are all available at server side and can be optimally scheduled “off-line”. Nevertheless, there is a growing number of VBR live interactive video applications requiring “on-line” smoothing algorithms, to reduce bit rate variability on-the-fly during stream transmission. Online smoothing is effective to reduce peak rate and rate variability in temporal windows of limited size, at the same time performing a on-the-fly computation of smoothing transmission plans. A further optimization of smoothed transmission plan can be obtained by smoothing videos over consecutive windows partially overlapped in time (Rexford et al., 2000). The impact of available bandwidth on scheduling has also been developed in several works (Bewi et al., 2002; Le Boudec & Thiran, 2004; Lai et al., 2005; Lee, 2006). Bandwidth dependent smoothing considers the additional available bandwidth information for improving the efficiency of the stream schedule.

The basic smoothing principles previously mentioned are the starting point for the implementation of the novel transmission techniques proposed in this chapter. Two scenarios will be analyzed in detail: the VBR video transmission in DVB-H systems and the impact of user interactivity on VBR video delivering in UMTS terminals. For DVB-H systems, a online schedule for all the TDM services is generated at server side to prevent buffer overflows and underflows. It is calculated on consecutive temporal observation windows partially overlapped in time and takes into account available bandwidth resources, burst and receiving buffer sizes. The same smoothing principles can be also used to implement an on-line scheduling that takes into account the user interactivity, that can consistently change the status of receiving terminal, in terms of decoded frames and buffer occupancy level. The on-line schedule, performed over partially overlapping temporal windows, reschedules data according to the feedback RTCP information on the terminal status.

4.1 Optimal scheduling of multiservice VBR video transmission in DVB-H systems

As stated in Section 2.1.1, service data are stored in bursts at transmission side and buffered in the client memory during burst times for continuous playing also during the subsequent off-times. We conventionally say that a burst “covers” the off-time if the burst size guarantees the continuous playback on the DVB-H terminal. Obviously, sufficient buffering is needed at receiving side to cover off-times. In fact the burst size must always be less than the memory available in the receiver and the burst bitrate and burst duration can be set accordingly. Let us note that it is quite easy to statically set the burst parameters for CBR streams; it is instead more difficult for VBR videos, e.g. coded with MPEG or H-264/AVC. In these cases, since the video bit rate is highly variable in time, data stored in the client buffer can heavily vary during video reproduction. Statically setting the burst size for the whole video transmission could easily bring to losses at receiving side, because of the insufficient amount of buffered data and/or because of the relatively small receiving buffer size.

This critical aspect in transmission scheduling is further complicated by the available bandwidth information. Service data in fact could be more effectively scheduled if available bandwidth reserved for the single service is a known parameter. A reduced available bandwidth reduces the maximum amount of data filling a burst because it limits the burst bitrate, under the same burst duration. So when the bandwidth assigned to the service is

relatively small, burst bitrate is limited and data stored into bursts could not cover the off-time. The proposed Variable Burst Time (VBT) algorithm, an on-line scheduling algorithm suitable for transmission of time-sliced high quality VBR services, tries to overcome these problems. Transmission optimization is performed by dynamically and simultaneously varying the burst durations of the whole set of services transmitted in a temporal window of fixed size, sliding in time. The optimization method takes into account available bandwidth resources, burst parameters and receiving buffer sizes. The goal is the loss minimization of the whole service set. To test its effectiveness, VBT is compared with the classical DVB-H transmission as recommended by the standard (ETSI TR 102 377), that statically sets all service burst durations.

The VBT schedule exploits the same basic smoothing concepts illustrated in Section 4. Its transmission plan aims to prevent buffer overflows and underflows, the only two conditions supposed to bring to bit losses at receiving side. In the specific DVB-H scenario, a service buffer underflow occurs if the burst size is not enough to cover off-time. A buffer overflow occurs instead if the free buffer level is too small to store the burst size.

4.1.1 The VBT implementation

To fully understand VBT, let us first consider the single service schedule, assuming a discrete time evolution with the basic time unit of a frame time (1/25 s for PAL). VBT schedules as many data as possible in each service burst in advance respect to its playback time avoiding buffer overflows and underflows. To perform this step, supposed the receiving buffer size of b bits and f_i the i^{th} frame size (in bits), the two overflow and underflow curves are built according to (1) and (2):

$$F_{\text{under}}(k) = \sum_{i=0}^k f_i ; F_{\text{over}}(k) = b + \sum_{i=0}^k f_i \quad (4)$$

The resulting schedule $S(k) = \sum_{i=0}^k s_i$ is represented in Fig. 10 in a generic burst cycle, where it is supposed that the service burst duration starts in T_{bi} and ends in T_{bs} , and that the service off-time starts in T_{bs} and ends in T_{cycle} . The burst duration is $T_{on} = T_{bs} - T_{bi}$ and the Off-time $T_{off} = T_{\text{cycle}} - T_{bs}$. Consequently the scheduled bitrate s_i will be $s_i = \text{Burst Bitrate}$ in $[T_{bi}, T_{bs}]$, during data transmission, and $s_i = 0$ in $[T_{bs}, T_{\text{cycle}}]$. The cumulative schedule $S(k)$ will thus increase only in $[T_{bi}, T_{bs}]$, with slope equal to burst bitrate, and will remain constant in $[T_{bs}, T_{\text{cycle}}]$. Furthermore, let us assume $q_b = S(T_{bi}) - F_{\text{under}}(T_{bi})$ as the buffer fill level in T_{bi} . This is the amount of data stored in the client buffer during the previous burst duration and not yet consumed by client at the end of the previous burst cycle. Similarly, $q_e = S(T_{\text{cycle}}) - F_{\text{under}}(T_{\text{cycle}})$ is the buffer fill level at the burst cycle end.

The schedule will be feasible without losses if and only if $F_{\text{under}}(k) \leq S(k) \leq F_{\text{over}}(k) \quad \forall k$. If there are some k so that $S(k) > F_{\text{over}}(k)$ a buffer overflow occurs in that k ; if instead $S(k) < F_{\text{under}}(k)$ there will be a buffer underflow. Both these critical conditions are supposed to generate losses at receiving side. Nevertheless, a buffer overflow can easily be avoided by properly regulating the burst bitrate at transmission side, where the receiving buffer size is

supposed to be known, so that $S(k)$ cannot never cross $F_{over}(k)$ in $[T_{bi}, T_{bs}]$. A buffer underflow occurs instead because the burst size is relatively small and cannot cover the off-time, both because available bandwidth limits the burst bitrate or because the off-time is relatively long if compared with the burst duration even without any bandwidth limitation. The complication is that a service off-time strongly depends on the other service burst durations; so *all* the burst durations must be simultaneously adjusted in a burst cycle to minimize losses.

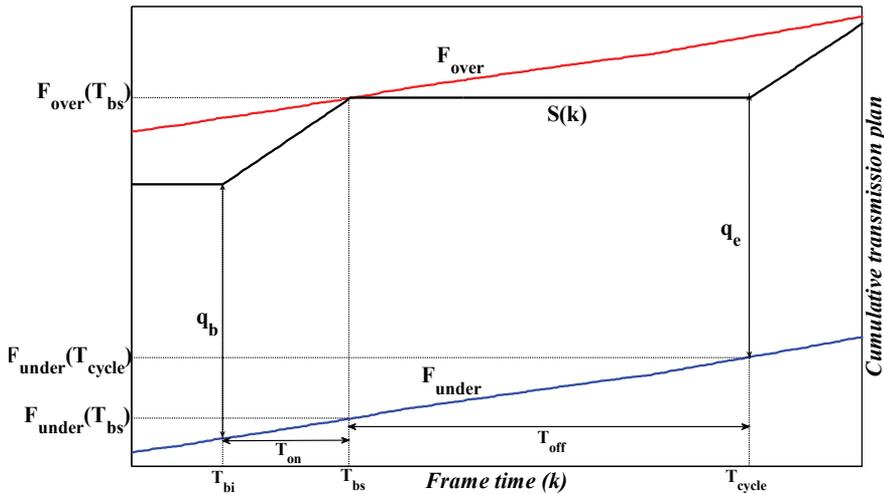


Fig. 10. Single service cumulative transmission plan in a burst cycle

Since VBT is an on-line algorithm, loss minimization is performed on a Temporal Observation Window (TOW) whose length is chosen as a integer number of burst cycles. As explained in (Rexford et al., 2000) a more efficient scheduling can be obtained if the TOWs are partially overlapped in time. The higher the overlapping degree, the better the on-line schedule, even if this comes at a cost of a computational overhead. For a more efficient schedule optimization, VBT also considers partially overlapping TOWs sliding in time. More precisely, we suppose that there are W_s burst cycles in a TOW, and that there are N_s services in a burst cycle. Loss optimization is performed in the considered TOW, that then slides by N_B burst cycles, repeating the optimization procedure in the new TOW. The first N_B burst cycles in the previous step are transmitted and N_B new burst cycles are introduced in the optimization process of the following step. In Fig. 11 this sliding procedure is illustrated, together with the main TOW parameters.

The parameters W_s and N_B influence VBT performance. The TOW length in fact introduces also an initial delay in video playback, since the server must know all service frames to be stored in bursts before calculating the optimal burst durations. On the other side, a larger TOW allows VBT to span a higher number of burst duration configurations, increasing the probability to find a lower minimum for losses.

Also the slide length N_B influences VBT performance. A smaller N_B allows to more efficiently optimize the service burst durations as new service data are scheduled in bursts. In fact, the service data to be scheduled in the last N_B new burst cycles included in the

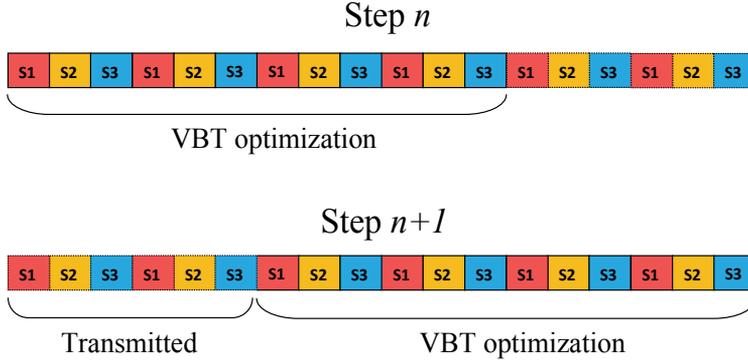


Fig. 11. The TOW sliding procedure with $W_s = 4$, $N_s = 3$ and $N_B = 2$

$(n+1)^{th}$ step of Fig. 11 refine also the calculation of the service burst durations of the previous $W_s - N_B$ burst cycles already calculated in the n^{th} step, even if the computational overhead increases. The VBT computational complexity also increases with the number of services N_s .

Supposed a generic TOW, we define as *configuration* the n-uple of burst durations (where $n = N_s \cdot W_s$):

$$\bar{T}_{on} = (T_{on}^{(1,1)}, \dots, T_{on}^{(N_s,1)}, \dots, T_{on}^{(1,W_s)}, \dots, T_{on}^{(N_s,W_s)}) \quad (5)$$

where each $T_{on}^{(i,j)}$ is the i^{th} service burst duration in the j^{th} burst cycle. $T_{on}^{(i,j)}$ is a positive integer multiple of the frame time unit. To find the optimal configuration $\bar{T}_{on,opt} = (T_{on,opt}^{(1,1)}, \dots, T_{on,opt}^{(N_s,W_s)})$ that minimizes all service losses a *Total Loss Function* (TLF) is introduced, that considers all the service losses for buffer underflow in the TOW. The first step is to calculate the i^{th} service losses by considering the buffer state $q_e^{(i,j)}$ at the end of the j^{th} burst cycle (see Fig. 10). Let us suppose $T_{on}^{(i,j)}$ the i^{th} service burst duration in the j^{th} burst cycle. We can say that cumulative data filling the buffer (namely, the burst size) are:

$$D_{in}^{(i,j)}(T_{on}^{(i,j)}, W_{av}^{(i,j)}) = BB^{(i,j)} \cdot T_{on}^{(i,j)} \quad (6)$$

Where $BB^{(i,j)}$ is the i^{th} service burst bitrate in $T_{on}^{(i,j)}$. $BB^{(i,j)}$ must be properly calculated to avoid buffer overflow and to be less than the i^{th} service available bandwidth $W_{av}^{(i,j)}$, a constant value assigned to the specific service. That is:

$$BB^{(i,j)} = \min \left\{ W_{av}^{(i,j)}, \frac{F_{over}(T_{bs}^{(i,j)}) - (F_{under}(T_{bi}^{(i,j)}) + q_b^{(i,j)})}{T_{on}^{(i,j)}} \right\} \quad (7)$$

where $q_b^{(i,j)}$ is the buffer fill level in $T_{bi}^{(i,j)}$.

Cumulative data leaving the buffer at the end of the burst cycle are instead:

$$D_{out}^{(i,j)}(T_{on}^{(i,j)}, T_{off}^{(i,j)}) = F_{under}(T_{cycle}^{(i,j)}) - F_{under}(T_{bi}^{(i,j)}) \quad (8)$$

as clearly visible in Fig. 10.

The buffer fill level in $T_{cycle}^{(i,j)}$ is thus:

$$q_e^{(i,j)}(T_{on}^{(i,j)}, T_{off}^{(i,j)}, W_{av}^{(i,j)}) = q_b^{(i,j)}(T_{bi}^{(i,j)}) + D_{in}^{(i,j)}(T_{on}^{(i,j)}, W_{av}^{(i,j)}) - F_{under}(T_{cycle}^{(i,j)}) \quad (9)$$

Losses will occur if and only if the result of (9) is negative, that is, S crosses F_{under} in $T_{cycle}^{(i,j)}$, that is:

$$L^{(i,j)}(T_{on}^{(i,j)}, T_{off}^{(i,j)}, W_{av}^{(i,j)}) = \max\{-q_e^{(i,j)}(T_{on}^{(i,j)}, T_{off}^{(i,j)}, W_{av}^{(i,j)}), 0\} \quad (10)$$

$$1 \leq i \leq N_s, 1 \leq j \leq W_s - 1$$

Let us note that the N_s burst durations of the last burst cycle in a TOW are only used to evaluate the N_s streams off-times for losses evaluation in the $(W_s - 1)^{th}$ burst cycle.

$L^{(i,j)}(T_{on}^{(i,j)}, T_{off}^{(i,j)}, W_{av}^{(i,j)})$ can be derived only after the available bandwidth $W_{av}^{(i,j)}$ and the \bar{T}_{on} vector as defined in (5) have been set, since the generic $T_{off}^{(i,j)}$ of the i^{th} service off-time in the j^{th} burst cycle is given by:

$$\left\{ \begin{array}{l} T_{off}^{(i,j)} = \sum_{k=2}^{N_s} T_{on}^{(k,j)} \text{ if } i = 1 \\ T_{off}^{(i,j)} = \sum_{k=i+1}^{N_s} T_{on}^{(k,j)} + \sum_{k=1}^{i-1} T_{on}^{(k,j+1)} \text{ if } i > 1 \end{array} \right. , 1 \leq j \leq W_s - 1 \quad (11)$$

and depends on the burst durations of the other services.

To guarantee fairness among services, the TLF normalizes each service losses to the amount of service data transmitted in the TOW; otherwise, services with lower mean bit rates would be penalized. So the i^{th} service losses are evaluated as:

$$L^{(i)}(\bar{T}_{on}, \bar{W}_{av}) = \left(\sum_{j=1}^{W_s-1} L^{(i,j)} \right) / \left(\sum_{j=1}^{W_s-1} D_{in}^{(i,j)} \right), 1 \leq i \leq N_s \quad (12)$$

where $D_{in}^{(i,j)}$ is given by (6), \bar{T}_{on} by (5) and $\bar{W}_{av} = (W_{av}^{(1,1)}, \dots, W_{av}^{(N_s,1)}, \dots, W_{av}^{(1,W_s)}, \dots, W_{av}^{(N_s,W_s)})$ is the available bandwidth vector of the N_s services in a TOW.

Finally, to obtain the TLF the normalized losses calculated in (12) are averaged over the services:

$$M(\bar{T}_{on}, \bar{W}_{av}) = \sum_{i=1}^{N_s} L_{fac}^{(i)} / N_s \quad (13)$$

providing the total amount of losses to be minimized.

The VBT optimal solution must verify the condition:

$$TLF(\bar{T}_{on,opt}, \bar{W}_{av}) = \min\{TLF(\bar{T}_{on}, \bar{W}_{av}), \bar{T}_{on} \in \mathbb{N}_+^n\} \quad (14)$$

where \mathbb{N}_+^n simply indicates the subset of \mathbb{N}^n of all strictly positive natural numbers. The solution to (14) can be found iteratively by numerical methods that find the minimum of a nonlinear bounded multivariable function.

4.1.2 Numerical results

To test its effectiveness, VBT has been compared with the stream transmission as recommended by the DVB-H standard (ETSI TR 102 377) that considers constant burst durations and burst cycles. We call this implementation "Constant Burst Time (CBT) algorithm" for simplicity of notation. Comparison has been performed by multiplexing $N_s = 4$ video streams and reproducing different simulation scenarios. The four chosen video streams, all of length 5.000 video frames, have different quality coding degrees; their main statistics have been listed in Table 1.

| Video streams | Jurassic Park | Video Clip | Star Wars IV | The Silence of the Lambs |
|-----------------------------|---------------|------------|--------------|--------------------------|
| Compression ratio (YUV:MP4) | 9.92 | 38.17 | 27.62 | 43.43 |
| Mean frame size (bytes) | 3.8e+03 | 1e+03 | 1.4e+03 | 8.8e+02 |
| Min frame size (bytes) | 72 | 31 | 26 | 28 |
| Max frame size (bytes) | 16745 | 9025 | 9370 | 11915 |
| Mean bit rate (bit/s) | 7.7e+05 | 2e+05 | 2.8e+05 | 1.8e+05 |
| Peak bit rate (bit/s) | 2.4e+06 | 8.5e+05 | 1.2e+06 | 1.8e+06 |
| Peak/Mean of bit rate | 3.15 | 4.29 | 4.29 | 10.07 |

Table 1. Main video streams statistics

The first proposed experiment shows the influence of the TOW length in VBT losses calculation for three different N_B . The TOW length has been varied from $W_s = 4$ to $W_s = 8$ burst cycles, keeping a constant available bandwidth of 3 Mbps and a receiving buffer size of 320 kB for all services. Fig. 12 depicts the total losses experimented for the services presented in Table 1. As expected, losses decrease with TOW increase since loss optimization for each service is performed over a larger number of burst durations. Furthermore, losses decrease with slide length decrease, because a smaller slide length allows a better refinement in burst durations calculation among subsequent steps. Let us also note that for $W_s = 4$ and $N_B = 3$ losses are evaluated over non overlapped and uncorrelated TOWs; they are thus proportionally much higher than the other experimented cases, since the burst durations refinement through TOW overlapping is not possible.

Total CBT experimented losses are much higher than VBT ones, so they have not been reported in the figure to improve its readability. They have been calculated as follows. The TOW length has been set to the whole streams length to eliminate its influence in CBT loss calculation. The same available bandwidth of 3 Mbps has been considered for all services. The burst duration, the same for all services, has been increased from 3 to 100 frame times (with step 1 frame time) and the minimum of losses calculated in each step has been observed, resulting in 103.2 Mbits. This minimum has been found for a service burst

duration of 17 frame times and a burst cycle of 68 frame times. Let us note that CBT losses found in the best scheduling conditions are approximately an order of magnitude higher than the maximum amount of VBT losses (observed for $W_s = 4$ and $N_B = 3$).

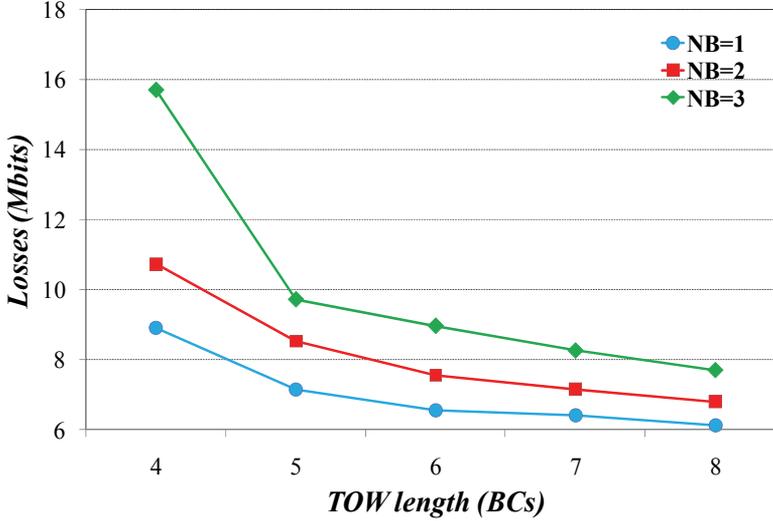


Fig. 12. VBT losses vs TOW length for three different N_B values.

The second proposed experiment illustrated in Fig. 13 shows the impact of the available bandwidth over losses for CBT and VBT. The same four pieces of video streams have been scheduled for transmission in a TOW of $W_s = 8$ burst cycles, using a constant available bandwidth ranging from 1 to 5 Mbps. The receiving buffer size adopted is 320 kB. Numerical results have been displayed for two different slide lengths ($N_B = 1$ and $N_B = 7$). CBT losses have been evaluated as the minimum among all service burst durations ranging from 3 to 100 frame times, as previously explained.

As expected, losses increase with available bandwidth decrease and VBT losses are smaller than CBT ones in all experimented scenarios. Differences between CBT and VBT, and between the two VBT slide lengths, are almost imperceptible for 1 Mbps of available bandwidth, because the dynamic variation of burst durations and the overlapping degree have an almost null effect in reducing losses, that are instead almost exclusively due to the very stringent bandwidth limitation. For increasing available bandwidth, VBT burst duration adjustment is much more effective if compared with the static CBT burst duration assignment.

The last experiment illustrated in Fig. 14 shows the impact of the receiving buffer size over losses for CBT and VBT schedules. Regarding VBT, the TOW length has been set to 8 burst cycles with a slide length of $N_B = 1$ burst cycle. Regarding CBT, the whole services have been scheduled a time with the same procedure illustrated in Fig. 12 for loss minimization. The client buffer sizes range from 128 to 1024 kB, with step of 128 kB. VBT losses have been evaluated for two different available bandwidth values (3 and 4 Mbps), while the available bandwidth for CBT has been set to 5 Mbps to exclude the influence of the bandwidth limitation in loss calculation.

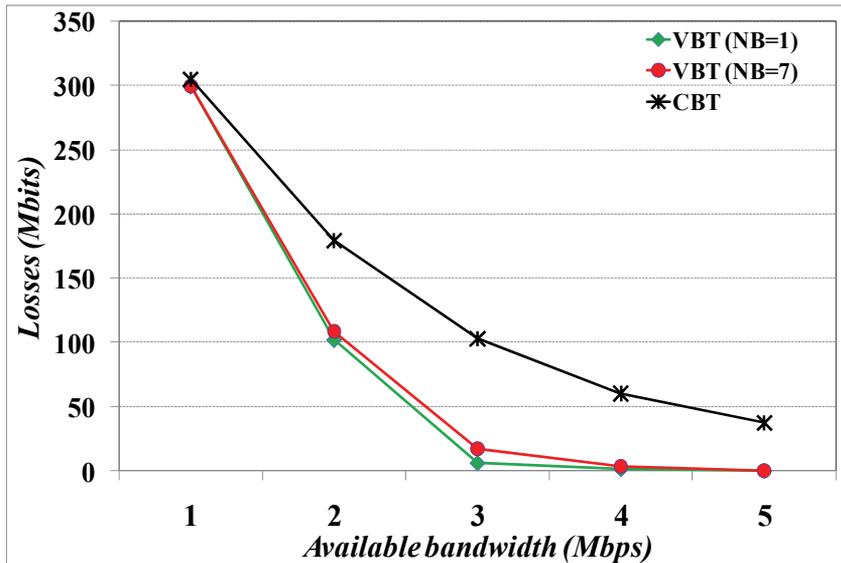


Fig. 13. Losses vs available bandwidth for VBT and CBT schedules.

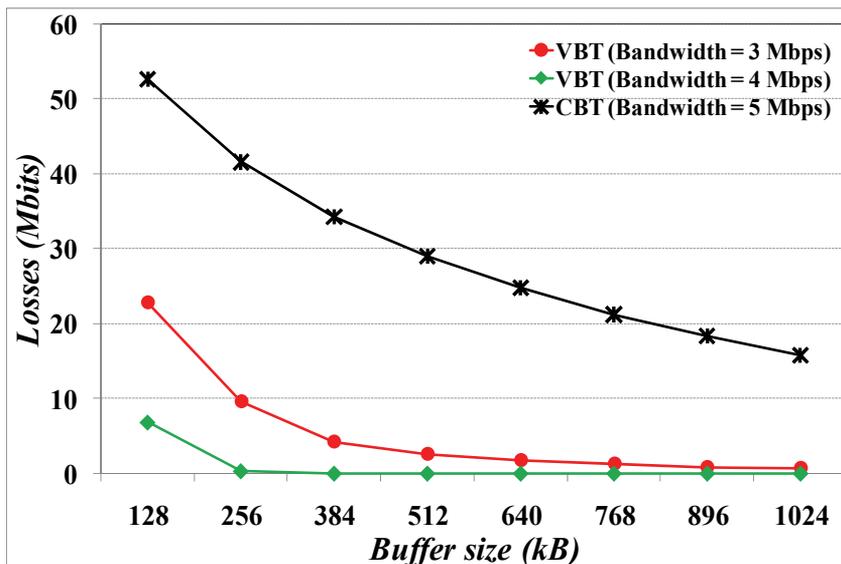


Fig. 14. Losses vs receiving buffer size for VBT and CBT schedules.

As expected, losses decrease with buffer increase for both VBT and CBT. This happens because a larger buffer size allows to store more data at receiving side, reducing the loss probability for buffer underflow. Again VBT losses are always smaller than CBT ones; furthermore, CBT losses are still present even if the available bandwidth is relatively high: with a buffer size of 1024 kB, CBT total losses are still of about 16 Mbits. VBT experimented losses, with a 4 Mbps bandwidth, are instead null right from a buffer size of 384 kB up, testifying that video scheduling through dynamic burst duration adjustment, together with proper buffering at receiving side, can drastically abate losses.

4.2 Scheduling VBR video streams in 3G networks

The main factors that influence the high quality delivery of audio and video contents over UMTS terminals are the highly fluctuating conditions of wireless links and the limited amount of buffering on mobile terminals. UMTS systems should guarantee lossless data delivery despite of the highly variable bandwidth conditions of wireless channel and the high fluctuating data bit rates. Furthermore, a certain degree of user interactivity, that strongly influences data buffering and decoding, should be taken into account to avoid continuous rebufferings and bad media quality, respectively due to receiving buffer underflows and overflows. These problems can be faced through dynamic scheduling at transmission side. In this work we focus on the effects of the user interactivity that modifies the status of the mobile terminal, ignoring the influence of other aspects like the fluctuating channel bandwidth.

In this section we present an on-line scheduling algorithm, the Dynamic-Buffer Dependent Smoothing Algorithm (D-BDSA) suitable for interactive multimedia applications in 3G wireless networks. Scheduling is performed over partially overlapped temporal windows, sliding in time according to the feedback information carried to the streaming server by RTCP packets. Rescheduling is performed because of external user actions (pause, or fast forward) that change the client buffer fill level. The goal is the frame losses reduction.

Let us suppose a streaming server providing VBR video to a 3G terminal. The server reduces the video bit rate variability through the work-ahead MVBA smoothing as developed in (Rexford et al., 2000) calculated over partially overlapped TOWs of N frames, sliding by α frames each step. α is the so called *slide length*. In each step, the first α frames scheduled in the previous step are sent to the client, while α new frames are scheduled together with the remaining $N - \alpha$ frames already scheduled in the previous step. In the on-line smoothing algorithms with a statically assigned α analyzed in (Rexford et al., 2000), it has been experimented that $\alpha = N / 2$ is a good compromise between an optimized schedule and a reduced computational overhead. D-BDSA exploits the feedback information provided by RTCP packets to dynamically adjust the slide length according to the user actions. In this analysis, we suppose that there is enough available bandwidth in the UMTS network, so that delays in data transmission and control information across the network can be assumed almost null. For this reason the FBS information completely describes the client status (see Fig. 7). In fact in the generic k^{th} frame time it holds:

$$\begin{aligned} HTSN(k) &= HRSN(k) \\ NSN(k) &= HTSN(k) + FBS(k) - b + 1 \end{aligned} \quad (15)$$

where b is the client buffer size.

Let us now suppose that a NADU packet containing the FBS information arrives to the server in the k^{th} frame time. Let us call this information $FBS_c(k)$ for simplicity, to indicate that it is sent by the client. The server compares the value of $FBS_c(k)$ with the free expected buffer level derived by the schedule, that we call $FBS_s(k)$, that is simply given by:

$$FBS_s(k) = B(k) - S(k) > 0 \quad (16)$$

With $B(k)$ and $S(k)$ given by (2) and (3) respectively. The server knows this information because it calculates the transmission plan for each k . Let us point out that the lower bound

$D(k)$ defined in (1) is the cumulative amount of bits consumed by the client during video playback, so the server calculates the transmission plan always assuming continuous playback at client side. So, if the user performs a video playback, it surely will be $FBS_s(k) = FBS_c(k)$. If instead the user performs other actions like for example fast forward, or pause, it will be $FBS_s(k) \neq FBS_c(k)$ depending on the specific user action. For example, if the user pauses the stream, $FBS_c(k) < FBS_s(k)$ because data only enter the client buffer, while the server supposes that frames also leave the buffer for playback. And vice versa for the fast forward action. In this case:

- Video data must be rescheduled by the server according with the new updated $FBS_c(k)$ information coming from the client, to prevent frame losses;
- The frequency of the feedback information must be increased as a function of the $|\Delta B| = |FBS_s(k) - FBS_c(k)|$ displacement, to more quickly adjust the schedule according to the real client buffer status.

The two critical conditions experimented at client side are a buffer underflow and a buffer overflow. Nevertheless, whenever the client buffer underflows because of a fast forward action, when the server performs a rescheduling in k it knows exactly the number of the last decoded packet $HTSN(k)$, since the buffer is empty, so it can send scheduled frames starting by $HTSN(k)$. The user will thus experiment only rebufferings with consequent delays in frame decoding; frame losses for buffer underflow will *never* occur. In the case of a buffer overflow instead losses surely will occur because the buffer is full and the server continuously sends data. This implies that $HTSN(k) > HRSN(k)$, and the server will not be able to send lost frames again. Both these critical conditions can be prevented as more as α is small, because the server can more quickly react to a buffer variation through rescheduling. This suggests the implementation of the relationship $\alpha(\Delta B)$. Remembering that:

$$|\Delta B| = |FBS_s(k) - FBS_c(k)| \Rightarrow 0 \leq |\Delta B| \leq b \quad (17)$$

by (17) it is clear that the higher $|\Delta B|$ the more likely a user action that changes the free client buffer level compared with the one calculated by the server. Rescheduling should thus be as more frequent as $|\Delta B|$ increases. The proposed solution is hyperbolic relationship between α and $|\Delta B|$:

$$|\Delta B| = \frac{a_1}{\alpha} + a_2 \quad (18)$$

This kind of relationship has been chosen because small $|\Delta B|$ increments bring to high α decrements, which is typical of a hyperbolic relationship. a_1 and a_2 can easily be derived by imposing the two bound conditions:

$$\begin{cases} |\Delta B| = 0 \Rightarrow \alpha = N / 2 \\ |\Delta B| = b \Rightarrow \alpha = 1 \end{cases} \quad (19)$$

The maximum slide length has been chosen $\alpha = N / 2$ since, as previously mentioned, it is the best trade-off between the optimality of the schedule and the algorithm computational overhead. Imposing the (19) and solving the system:

$$\begin{cases} a_1 + a_2 = b \\ \frac{2a_1}{N} + a_2 = 0 \end{cases} \Rightarrow \begin{cases} a_1 = \frac{Nb}{N-2} \\ a_2 = -\frac{2b}{N-2} \end{cases} \quad (20)$$

it derives:

$$\alpha = \frac{Nb}{(N-2)|\Delta B| + 2b} \quad (21)$$

The (21) is exploited by the server to dynamically update the frequency of rescheduling. D-BDSA is summarized as follows:

1. When a RTCP packet comes to the server in k , the server compares $FBS_c(k)$ with $FBS_s(k)$ and calculates $\Delta B = FBS_s(k) - FBS_c(k)$;
2. The server calculates the next scheduled transmission time $k_1(\Delta B) = k + \alpha$ of the RTCP packet through (21), and sends this information back to the client through specific RTCP packets (Schulzrinne et al., 2003), supposed to be immediately available to client;
3. The server updates the $NSN(k)$ information exploiting the second of (15) calculated in k ;
4. The server reschedules video data in a TOW of N frames, considering $NSN(k)$ calculated in the step 2 as the first frame to be decoded. Then sends the first α scheduled frames, until a new $FBS_c(k_1)$ information arrives from the client.

4.2.2 D-BDSA performance

In this Section, we test D-BDSA effectiveness by comparing it with the same BDSA scheduling algorithm, but with a constant slide length α . We call this version of BDSA Static-BDSA (S-BDSA), setting $\alpha = N/2$. All time units are expressed in frame times. Comparison between S-BDSA and D-BDSA has been made by simulating the transmission of 70.000 video frames of the "Jurassic Park" video, MPEG-4 coded with high quality, with a sequence of the simulated user external actions summarized in Table 2.

The first proposed experiment is illustrated in Fig. 15. It shows the influence of the TOW length N on losses for D-BDSA and S-BDSA. N is varied from 500 frame times (20 seconds) to 1.000 frame times (40 seconds) with step 50 frame times (2 seconds). The client buffer size has been chosen of 1 Mbyte. D-BDSA losses are always smaller than S-BDSA ones, thanks to the D-BDSA dynamic change of the slide length. For both schedules, losses decrease with N decrease because a smaller N means a smaller α , both for S-BDSA ($\alpha = N/2$) and even more for D-BDSA (where $1 \leq \alpha \leq N/2$) and a resulting increased feedback frequency that reduces loss probability. On the other side, let us note that a smaller N increases the algorithm computational complexity since a higher number of TOWs is needed to schedule the whole video stream.

The second proposed experiment illustrated in Fig. 16 shows the S-BDSA and D-BDSA performance for different client buffer sizes. Losses have been calculated by choosing the same piece of video stream used in the previous simulation and the same sequence of user actions listed in Table 2, with $N=600$ frame times. The buffer size has been increased from 64 kbytes until 2 Mbytes, with increasing powers of 2.

| Action Number | Action type | Duration (frame times) | Starting time (frame time) | Ending time (frame time) |
|---------------|-----------------|------------------------|----------------------------|--------------------------|
| 1 | Playback | 5.000 | 1 | 5.000 |
| 2 | Pause | 2.000 | 5.001 | 10.000 |
| 3 | Fast Forward 4x | 500 | 10.001 | 10.500 |
| 4 | Playback | 2.000 | 10.501 | 12.500 |
| 5 | Fast Forward 2x | 2.000 | 12.501 | 14.500 |
| 6 | Playback | 2.000 | 14.501 | 16.500 |
| 7 | Fast Forward 4x | 700 | 16.501 | 17.200 |
| 8 | Playback | 2.000 | 17.201 | 19.200 |
| 9 | Pause | 2.000 | 19.201 | 21.200 |
| 10 | Playback | 2.000 | 21.201 | 23.200 |
| 11 | Fast Forward 2x | 1.000 | 23.201 | 24.200 |
| 12 | Playback | 10.000 | 24.201 | 34.200 |
| 13 | Pause | 5.000 | 34.201 | 39.200 |
| 14 | Playback | Until stream end | 39.201 | End of Stream |

Table 2. Sequence of the user actions on the client terminal

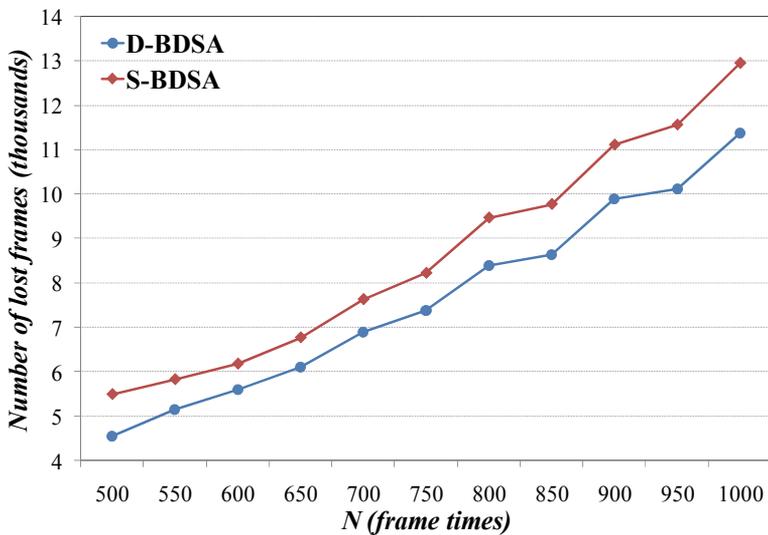


Fig. 15. Lost frames vs TOW length for S-BDSA and D-BDSA

As shown by Fig. 16, losses decrease with buffer increase since larger client buffers reduce the buffer overflow probability. D-BDSA losses are always smaller than S-BDSA ones. For smaller buffer sizes (64, 128 and 256 kB) losses are high in both cases, even if they decrease

more quickly for D-BDSA. This happens because the video flow is compressed with high quality, with a relatively high number of bits per frame that cannot be stored by the client buffer, on average increasing the buffer overflow probability. The same result has been experimented also for other values of N and/or other stream types. For a 4 Mbyte buffer and higher, losses are null in both cases.

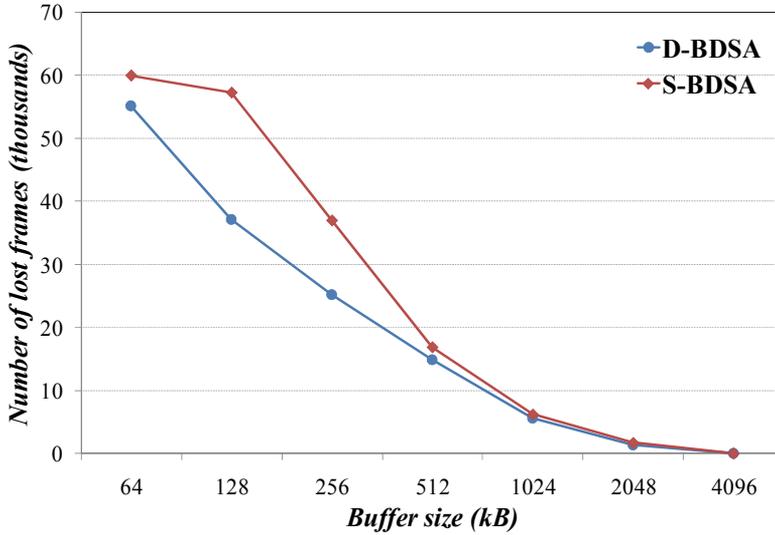


Fig. 16. D-BDSA and S-BDSA losses vs client buffer size.

5. Conclusions and future work

Studies illustrated in this chapter show that dynamic scheduling of VBR streams in wireless systems is very effective in reducing losses. Two different transmission scenarios have been analyzed where dynamic scheduling can be fruitfully applied: the DVB-H system, where a number of time multiplexed services share the same channel resources, and the UMTS network where schedule calculated by the streaming server is influenced by user actions reported back to server by RTCP packets.

Regarding DVB-H, the simultaneous dynamic variation of all service burst durations is of great help in reducing losses when VBR videos are transmitted. It is performed by taking into account service data, receiving buffer size and available bandwidth. Further work in this direction can be done in the improvement of the optimization method that finds the minimum of a nonlinear function of several variables. The method implemented in the proposed study finds local minimum, that in any case provides very good results. Nevertheless the optimization method could be further refined by finding a global TLF minimum, possibly with a relatively lower computational cost.

Regarding the UMTS network, the dynamic schedule derives from user actions that modify the client buffer status. Simulation have been performed over different types of video streams, TOW lengths and buffer sizes, testifying the effectiveness of the proposed method. Performance results are strongly influenced by the sequence of the user actions and especially by the slide length values calculated by the server. Further improvements in this direction can be done by testing other methods for dynamic α calculation and adopting

different scheduling algorithms than MVBA, that can more quickly react to the varying terminal conditions to further reduce frame losses. The real status of the UMTS core network, together with its buffers, could also be modeled to analyze the network behavior towards losses.

6. References

- 3GPP Technical Specification Group Services and System Aspects (TSGS-SA). TR 26.937. *Transparent end-to-end Packet Switched Streaming Service (PSS). RTP usage model (Release 6)*. Version 6.0.0, March 2004.
- 3GPP Technical Specification Group Services and System Aspects (TSGS-SA). TS 26.234. *Transparent end-to-end Packet Switched Streaming Service (PSS). Protocols and codecs (Release 6)*. Version 6.3.0, March 2005.
- Bewi, C., Pereira, R., Merabti, M. (2002). *Network Constrained Smoothing: Enhanced Multiplexing of MPEG-4 Video*. Proc. 7th International Symposium on Computers and Communications (ISCC'02), pp. 114-119, July 2002.
- ETSI EN 300 744. *Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for digital terrestrial television*.
- ETSI EN 301 192. *Digital Video Broadcasting (DVB); DVB specification for data broadcasting*.
- ETSI EN 302 304. *Digital Video Broadcasting (DVB); Transmission System for Handheld Terminals (DVB-H)*.
- ETSI TR 101 190. *Digital Video Broadcasting (DVB); Implementation guidelines for DVB terrestrial services; Transmission aspects*.
- ETSI TR 102 377. *Digital Video Broadcasting (DVB); DVB-H Implementation Guidelines*.
- Feng, W.-C., Rexford, J. (1999). *Performance Evaluation of Smoothing Algorithms for Transmitting Prerecorded Variable-Bit-Rate Video*. IEEE Transactions on Multimedia, Vol. 1, No.3, , pp. 302-313, September 1999.
- Flierl, M., Wiegand, T., Girod, B. (2001). *Multihypothesis Pictures for H.26L*. IEEE ICIP 2001, Greece, 2001.
- Holma, H. & Toskala, A. (2004). *WCDMA for UMTS : Radio Access for Third Generation Mobile Communications*. John Wiley & Sons, 3rd edition, ISBN 978-0471720515, New York.
- Horowitz, M., Joch, A., Kossentini, F., Hallapuro, A. (2003). *H.264/AVC Baseline Profile Decoder Complexity Analysis*. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 13, no. 7, pp. 704-716, 2003.
- ISO/IEC 13818-1. *Information Technology - Generic Coding of Moving Pictures and Associated Audio Information - Part 1: Systems*.
- ISO/IEC 14496-2. (2000). International Standard: 1999/Amd1:2000, 2000.
- ISO/IEC 7498-1. *Information Technology - Open System Interconnection - Basic Reference Model: The Basic Model*.
- ISO/IEC JTC 1.(2003). *Advanced video coding. ISO/IEC FDIS 14496-10*. International Standard, 2003.
- Lai, H., Lee, J.Y., Chen, L. (2005). *A Monotonic Decreasing Rate Scheduler for Variable-Bit-Rate Video Streaming*. IEEE Transactions on Circuits and Systems for Video Technology, vol.15, n.2, pp.221-231, February 2005.
- Le Boudec, J.-Y., Thiran, P. (2004). *Network Calculus: A Theory of Deterministic Queueing Systems for the Internet*. Book Springer Verlag, May 2004.

- Lee, M. (2006). *Video Traffic Prediction Based on Source Information and Preventive Channel Rate Decision for RCBR*. IEEE Transactions on Broadcasting, vol.52, n.2, pp.173-183, June 2006.
- Mitchell, J.L., Pennebaker, W.B., Fogg, C.E. & LeGall, D.J. (1996). *MPEG video compression standard*. Chapman & Hall, ISBN 978-0412087714, London.
- Puri, A., Chen, X., Luthra, A. (2004). *Video Coding Using the H.264/MPEG-4 AVC Compression Standard*. Elsevier Science, Signal Processing: Image Communication, Sept. 2004.
- Rexford, J., Sen, S., Dey, J., Kurose, J., Towsley, D. (2000). *Online Smoothing of Variable-Bit-Rate Streaming Video*. IEEE Transactions on Multimedia, vol.2, n.1, pp.37-48, March 2000.
- Salehi, J.D., Zhang, Z.-L., Kurose, J., D. Towsley. (1998). *Supporting Stored Video: Reducing Rate Variability and End-to-End Resource Requirements Through Optimal Smoothing*. IEEE/ACM Transactions On Networking, Vol.6, N.4, pp. 397-410, August 1998.
- Schulzrinne, H., Casner, S., Frederick, R., Jacobson, V. (2003). *RTP: A transport protocol for real time application*. RFC 3550, July 2003.
- Uskela, S. (2003). *Key Concepts for Evolution Toward Beyond 3G Networks*. IEEE Wireless Communications, pp.43-48, February 2003.
- Wiegand, T. (2002). *Joint Final Committee Draft*. Doc. JVT-E146d37ncm, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6), November 2002.
- Zhang, Z.L., Kurose, J., Salehi, J.D., Towsley, D. (1997). *Smoothing, statistical multiplexing, and call admission control for stored video*. IEEE Journal on Selected Areas in Communications, Vol.15, no.6, pp. 1148-1166, August 1997.

Resilient Digital Video Transmission over Wireless Channels using Pixel-Level Artefact Detection Mechanisms

Reuben A. Farrugia and Carl James Debono
*University of Malta
Malta*

1. Introduction

Recent advances in communications and video coding technology have brought multimedia communications into everyday life, where a variety of services and applications are being integrated within different devices such that multimedia content is provided everywhere and on any device. H.264/AVC provides a major advance on preceding video coding standards obtaining as much as twice the coding efficiency over these standards (Richardson I.E.G., 2003, Wiegand T. & Sullivan G.J., 2007). Furthermore, this new codec inserts video related information within network abstraction layer units (NALUs), which facilitates the transmission of H.264/AVC coded sequences over a variety of network environments (Stockhammer, T. & Hannuksela M.M., 2005) making it applicable for a broad range of applications such as TV broadcasting, mobile TV, video-on-demand, digital media storage, high definition TV, multimedia streaming and conversational applications.

Real-time wireless conversational and broadcast applications are particularly challenging as, in general, reliable delivery cannot be guaranteed (Stockhammer, T. & Hannuksela M.M., 2005). The H.264/AVC standard specifies several error resilient strategies to minimise the effect of transmission errors on the perceptual quality of the reconstructed video sequences. However, these methods assume a packet-loss scenario where the receiver discards and conceals all the video information contained within a corrupted NALU packet. This implies that the error resilient methods adopted by the standard operate at a lower bound since not all the information contained within a corrupted NALU packet is un-utilizable (Stockhammer, T. et al., 2003).

Decoding partially damaged bitstreams, where only corrupted MBs are concealed, may be advantageous over the standard approach. However, visually distorted regions which are not accurately detected by the syntax analysis of the decoder generally cause severe reduction in quality experienced by the end-user. This chapter investigates the application of pixel-level artefact detection mechanisms which can be employed to detect the visually impaired regions to be concealed. It further shows that heuristic thresholds are not applicable for these scenarios. On the other hand, applying machine learning methods such as Support Vector Machines (SVMs) can significantly increase the decoder's capability of detecting visual distorted regions. Simulation results will show that the SVMs manage to detect 94.6% of the visually impaired MBs resulting in Peak Signal-to-Noise (PSNR) gains of up to 10.59 dB on a

frame-by-frame basis. This method can be adopted in conjunction with other standard error resilient tools without affecting the transmission bit-rate required. Furthermore, the additional complexity is manageable which makes it applicable in real-time applications.

2. The effect of transmission errors

The H.264/AVC can achieve high compression efficiency with minimal loss of visual quality (Gonzalez, R.C., & Woods R.E., 2008). However, the resulting bitstream is susceptible to transmission errors, a phenomenon common in wireless environments, where even a single corrupted bit may cause disastrous quality degradation for an extensive period of time. This is mainly because video and image compression standards employ variable length codes (VLCs) to maximise the compression efficiency. Transmission errors may cause a decoder to lose synchronization and fail to decode subsequent VLC symbols correctly, as shown in Fig. 1. Thus, a single corrupted bit may result in a burst of corrupted pixels within the decoded frame. Furthermore, the spatial and temporal prediction algorithms adopted in block-based video compression standards such as H.264/AVC employ neighbouring macroblocks (MBs) and regions from reference frames respectively for prediction. If these regions are distorted, the reconstructed frame will be distorted as well, thus causing spatio-temporal propagation of errors (Richardson I.E.G., 2003).

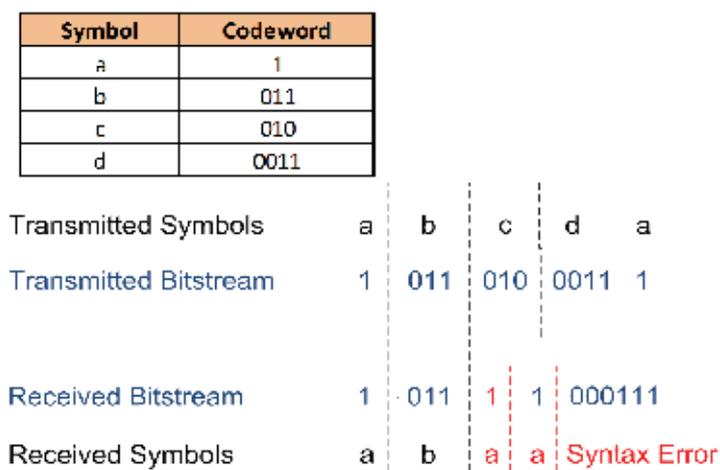


Fig. 1. Effect of transmission errors on variable length encoded sequences

In the standardization of H.264/AVC, corrupted packets are considered as being discarded by the receiver. Therefore, even a single corrupted bit within a packet will cause all the MBs contained within that packet to be dropped and concealed, thus, in general, the decoder will conceal a number of uncorrupted MBs. Furthermore, the spatio-temporal propagation of the superfluously concealed regions will propagate in both space and time, which by and large results in a significant reduction in perceptual quality. Fig. 2 illustrates the effect of a single corrupted bit in frame 41 of the *Foreman* sequence on the perceptual quality of the reconstructed frame and the effect of spatio-temporal propagation in the following frames.

Decoding of partially corrupted payloads may be beneficial for some applications. This is particularly true when considering damaged video bitstreams, since most of the MBs contained within a corrupted packet are either not corrupted or else provide imperceptible



Fig. 2. Error propagation in standard H.264/AVC caused by a single bit error

artefacts. Driven by this observation a set of protocols which allow the delivery of damaged packets are available. An overview of these methods is provided in (Welzl, M., 2005). However, as shown in Fig. 3, transition errors which are not detected by the syntax analysis of the H.264/AVC decoder may cause significant visual distortions which propagate in the spatio-temporal domain, thus reducing the end-user experience. The syntax analysis only manages to detect 57% of the corrupted MBs (Superiori, L. et al., 2006). Thus, to make decoding of partially damaged data applicable for video applications, the design of decoder-based algorithms which better cope with transmission errors without affecting the transmission bit rate are required.

3. Standard error resilient tools

The H.264/AVC video coding standard specifies several error resilient mechanisms aimed at minimizing the effect of transmission errors on the perceptual quality of the received video content. The following error resilient mechanisms are included in the standard:

- Slice Structuring
- Intra Refresh
- Flexible Macroblock Ordering (FMO)
- Redundant Slices (RS)
- Data Partitioning (DP)

3.1 Slice structuring

Delivery of video content over wireless networks is generally provided through frames having small maximum transmission units (MTU). This is mainly because the probability of



Fig. 3. Error propagation of artefacts caused by transmission errors when decoding partially damaged slices

bit errors affecting a small packet is lower than that for larger packets (Stockhammer, T. et al., 2003). Wireless video transmission is not an exception, where each frame is segmented into a number of independent coding units called slices. Slices are coded using limited spatial prediction and thus can be considered to provide spatially distinct synchronization points (Kumar, G. et al., 2006).

Furthermore, small packets reduce the amount of lost information, and hence the error concealment method can be applied to smaller regions, providing a better quality of experience. In (Kumar, G. et al., 2006) it was reported that slice structuring provides Peak Signal-to-Noise Ratio (PSNR) gains of around 7.33 dB when compared to single picture per packet with no error resilience.

3.2 Intra refresh

The application of slice structures limits the spatial propagation of errors within slice boundaries. However, due to the hybrid design of the H.264/AVC coding engine which adopts information from previously decoded frames, temporal propagation of artefacts caused by transmission errors is still severe. Instantaneous decoding refresh (IDR) pictures can be used to eliminate the temporal propagation of distorted regions where the entire picture is intra encoded. However, for real-time conversational video applications, it is not advisable to insert I-frames due to bit rate constraints and the resulting long delays involved (Kumar, G. et al., 2006).

Random intra MB coding is more acceptable for real-time applications. In addition to reducing the temporal propagation of distorted regions, it allows the encoder to maintain a

constant bit rate with the help of an H.264/AVC integrated rate-distortion control mechanism. Additionally, this technique is an encoder-based tool, which increases the data rate requirement but provides no additional overheads at the decoder.

3.3 Flexible Macroblock Ordering (FMO)

A more advanced error resilient tool specified in the H.264/AVC standard is the flexible macroblock ordering (FMO) which allows the specification of MB allocation within slice groups. The objective behind FMO is to scatter possible errors to the whole frame as equally as possible to avoid error clustering in a limited region. FMO is particularly powerful in conjunction with appropriate error concealment when the samples of a missing or corrupted slice are surrounded by many correctly decoded ones (Stockhammer, T. et al., 2003).

FMO can provide a significant gain in quality even at high error rates. However, FMO disallows the intra-frame prediction to exploit spatial redundancy in neighbouring MBs, since they are not enclosed within the same slice group. This will reduce the coding efficiency of the codec, thus limiting the applicability of FMO to low bit rate applications. Experimental results have shown that when transmitting over an error free channel, FMO encoded video sequences incur a data rate increase of 10% compared to when FMO is switched off (Wenger, S. & Horowitz, M., 2002).

3.4 Redundant Slices (RS)

A redundant slice is a new error resilient feature included within the H.264/AVC standard which allows the encoder to send redundant representations of various regions of pictures. Redundant slices may use different coding parameters, such as quantization levels, reference pictures, mode decisions, and motion vectors, than those used to encode the primary picture. If the primary slice is received correctly the redundant slice is simply discarded. However, if it is corrupted, the redundant slice is used instead in order to limit the visual distortion caused by transmission errors (Zhu, C. et al, 2006).

The enhanced error resilience provided by this strategy is achieved at the cost of additional overheads in terms of the excess bit rate requirements. There exists a trade-off between the degradation in the picture quality of the recovered video due to redundant slices and the available bandwidth. The more information introduced to describe the secondary slices the better is the performance of the decoder. Even though there is no restriction on the amount of information to be included in the redundant slices, in most applications bandwidth draws the limit and thus affects the performance of the redundant slices mechanisms.

3.5 Data Partitioning (DP)

The H.264/AVC codec generally encodes each frame to provide one single bitstream which forms a slice. However, since some coded information is more important than others, H.264/AVC enables the syntax of each slice to be separated into three different partitions for transmission (Wenger, S., 2003):

- Header information, including MB type, quantization parameters and motion vectors. This information is the most important, because without it, symbols of the other partitions cannot be used. This partition is called type A.
- The Intra partition (Type B) carries intra coded pictures and intra coefficients. This partition requires the availability of the type A partition of a given slice in order to be useful.

- The Inter partition contains only inter coded pictures and inter coefficients. This partition contains the least important information and in order to be useful requires the availability of the type A partition, but not the type B partition.

All partitions have to be available to execute standard conformant reconstruction of the video content. However, if the inter or intra partitions are missing, the available header information can still be used to improve the performance of the error concealment. Data partitioning is not included in the H.264's baseline profile and thus cannot be adopted for typical videoconferencing and mobile applications (Liu, L. et al., 2005).

4. Related work

Several extensions to the standard and novel error resilient strategies have been proposed in literature. Fragile watermarking was adopted in (Nemethova, O. et al., 2006), (Chen, M. et al., 2005) and (Park, W. & Jeon, B., 2002) to embed information that aids the detection and concealment of distorted regions. A low resolution version of each video frame was embedded in itself in (Adsumilli, C.B. et al., 2005) using spread-spectrum watermarking and is used to aid concealment of distorted regions. However, embedding information contributes to a reduction in the quality of the transmitted video content even when transmission is performed over an error free channel.

In order to better protect the transmitted bitstreams, error control strategies can be employed. Instead of including structured redundancy at the encoder, which reduces the compression efficiency, the authors in (Buttigieg, V. & Deguara, R., 2005) have replaced the VLC tables of the MPEG-4 video compression standard with variable length error correcting (VLEC) codes. A soft-decision sequential decoding algorithm was then implemented to decode the MPEG-4 bitstreams. However, the adoption of VLEC codes reduces the compression efficiency of the codec since the codewords produced have a longer average length.

The redundancy in compressed image and video was analyzed in (Nguyen, H. & Duhamel, P., 2003), where it was concluded that significant gain in performance can be achieved when additional video data properties are taken into consideration while decoding. The same authors have later adopted a modified Viterbi decoder algorithm to recover feasible H.263 video sequences in (Nguyen, H. & Duhamel, P., 2003), (Nguyen, H. & Duhamel, P. et al., 2004). Sequential decoding methods were adopted for H.264/AVC encoded sequences where Context Adaptive Variable Length Codes (CAVLC) (Weidmann, C. et al, 2004) and (Bergeron, C. & Lamy-Bergot, C., 2004) and Context Adaptive Binary Arithmetic Codes (Sabeva, G. et. al, 2006) coding modes were considered. However, sequential decoding algorithms introduce variable decoding delays which can be problematic in real-time applications (Lin, S. & Costello, D.J., 1983). A survey on joint-source channel coding techniques is provided in (Guillemot, C. & Siohan, P., 2005).

Various pixel-level artefact detection mechanisms based on heuristic thresholds were proposed in (Superiori, L. et. al, 2007), (Farrugia, R.A. & Debono, C.J., 2007) and (Ye, S., et. al, 2003). However, these methods only manage to detect between 40~60% of the corrupted MBs and have limited applications in practice since the optimal thresholds vary from sequence to sequence. An iterative solution which was presented in (Khan, et. al, 2004) attained substantial gain in quality at the expense of significantly increasing the complexity of the decoder, making it unsuitable for real-time mobile applications.

Scalable video coding (SVC) is another solution that can provide resilient delivery of video content over wireless channels (Schwarz, H. et. al., 2006). A scalable representation of video

consists of a base layer providing basic quality and multiple enhancement layers serving as a refinement of the base layer. The enhancement layers however are useless without the base layer. The benefit of SVC for wireless multiuser was demonstrated in (Liebl, G. et. al, 2006). However, this is achieved at the cost of higher encoding complexity and higher bit rate demand (Roodaki, H. et. al, 2008), (Ghanbari, M., 2003)).

Multiple Description Coding (MDC) (Goyal, V.K., 2001) is another alternative approach whose objective is to encode a source into two bitstreams such that high-quality reconstruction is derived when both bitstreams are received uncorrupted, while a lower but still acceptable quality reconstruction is achieved if only one stream is received uncorrupted. Several methods which adopt MDC to enhance the robustness of the video codec can be found in literature (Wang, Y. & Lin, S., 2002), (Tilio, T. et. al, 2008) and (Wang, Y. et. al, 2005). Again, the increase in error resilience is achieved by increasing the data rate required to deliver the same quality criterion as conventional single description coding methods in the absence of transmission errors.

Unequal Error Protection (UEP) schemes were extensively investigated in order to give higher protection to more important information. UEP was successfully combined with Data Partitioning (Barmada, B. et. al, 2005) and SVC (Zhang, C. & Xu, Y., 1999), (Wang, G. et. al 2001). Unbalanced MDC methods which allocate less bit rates to channels operating in bad conditions were proposed in (Ekmekci Flierl, S. et. al, 2005). The authors in (Rane, S. et. al, 2006) have proposed the transmission of a Wyner-Ziv encoded version of the image using the redundant slice option of H.264/AVC. Finally, Encoder-Decoder interactive error control approaches were presented in (Wang, J.T. & Chang, P.C., 1999), (Girod, B. & Färber, N., 1999) and (Budagavi, M. & Gibson, J.D., 2001). These methods however introduce additional delays making them unsuitable in wireless real-time applications and are not applicable for typical broadcasting/multicasting applications.

5. Pixel-level artefact detection mechanism

The standard H.264/AVC video coding standard was built with the assumption that corrupted slices are discarded and therefore does not allow partial decoding of corrupted payloads. This forces these mechanisms to operate at a lower bound since they assume a worst case scenario, where all the MBs contained within a corrupted slice are discarded and concealed. This assumption is most of the time untrue since in the majority of cases the MBs contained within a corrupted slice are either not corrupted or else provide imperceptible visual distortions. Therefore, decoding of partially damaged payloads may be beneficial when considering damaged compressed video content. For this reason, a set of syntax and semantic violations, presented in (Superiori, L. et al., 2006) were integrated within the H.264/AVC decoder to enable the decoding of damaged video content. However, these set of rules do not manage to accurately detect and localize a number of corrupted MBs resulting in severe visual impairments which propagate in the spatio-temporal domain, significantly reducing the quality of the reconstructed video sequences.

The pixel-level artefact detection mechanism is included as a post-process of the H.264/AVC sequences, as shown in Fig. 4, to detect the residual visually distorted regions to be concealed. The decoder is informed through the NALU header of the presence of transmission errors within the slice being decoded. To minimise computational complexity, the pixel-level artefact detection mechanism is only invoked to detect artefacts within corrupted slices. Therefore, no extra computation is required to decode uncorrupted slices.

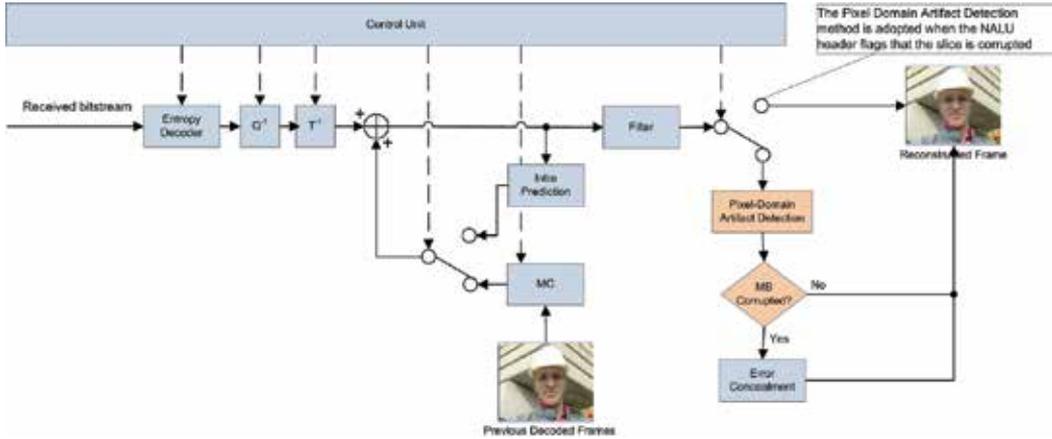


Fig. 4. Modified H.264/AVC Decoding Strategy using Pixel-Level Artefact Detection

The pixel-level artefact detection mechanism exploits the spatio-temporal smoothness of video scenes to detect visually impaired MBs. A number of dissimilarity metrics which exploit the redundancy present at pixel level are considered in subsection 5.1. These dissimilarity metrics generally provide large values to corrupted MBs and small values to uncorrupted MBs. Heuristic thresholds can be applied to discriminate between corrupted and uncorrupted MBs. However, the selection of these thresholds is sequence dependent and is thus not flexible enough to cater for a wide range of video sequences. This section introduces the application where a set of dissimilarity metrics were combined to form a feature vector, which is used to describe the reliability of each MB contained within a corrupted slice. The discrimination between corrupted and uncorrupted MBs is then provided through supervised machine learning algorithms which are discussed in subsection 5.2. These classifiers are trained and evaluated using subjective results provided through a case study discussed in subsection 5.3.

5.1 Dissimilarity metrics

5.1.1 Average Inter-sample Difference across Boundaries (AIDB)

In an image, there exists sufficient similarity among adjacent pixels, and hence across MB boundaries, even in the presence of edges. The *AIDB* dissimilarity metric is used to detect artefacts which affect the entire MB. Considering Fig. 5, let M denote a potentially corrupted MB under test with its four neighbouring MBs N , S , E and W . Further, let $p^{in} = \{p_1^{in}, p_2^{in}, \dots, p_K^{in}\}$ represent boundary pixels inside the MB M and $p^{out} = \{p_1^{out}, p_2^{out}, \dots, p_K^{out}\}$ represent boundary pixels in one of the neighbouring MBs $X \in \{N, S, E, W\}$. Then, the *AIDB*($M:X$) distance measure is given by:

$$AIDB(M:X) = \begin{cases} \frac{1}{K} \|p^{in} - p^{out}\|_2 & \text{if } X \text{ available} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where K is the size of the MB and $\|\bullet\|_2$ is the Euclidean distance. The *AIDB* dissimilarity metric is then computed by evaluating the average of *AIDB*($M:X$) over the available neighbouring MBs.

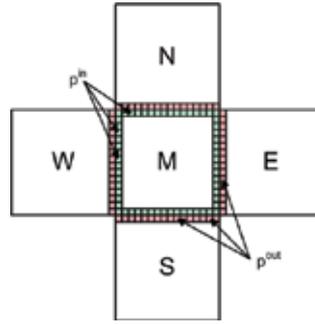


Fig. 5. Graphical representation of the $AIDB/IAIDB_{block}$ Dissimilarity Metrics

5.1.2 Internal AIDB per block ($IAIDB_{block}$)

The $IAIDB_{block}$ dissimilarity metric is based on the principle that, in non-corrupted MBs, the pixel transition from one 4×4 block to the adjacent block is generally smooth and thus the pixels at 4×4 boundaries are very similar. Opposed to the $AIDB$ dissimilarity metric, the $IAIDB_{block}$ was designed to detect artefacts within an MB. For this purpose, each MB is divided into a grid of 16 4×4 blocks and the $IAIDB_{block}(M:X)$ metric is computed using (1), but this time the parameter M represents the 4×4 block within an MB under test and $K = 4$. The $IAIDB_{block}$ dissimilarity metric of each 4×4 block is then derived by averaging the $IAIDB_{block}(M:X)$ over the available neighbouring 4×4 blocks. At the end of the computation we have a set of 16 $IAIDB_{block}$ dissimilarity metrics.

5.1.3 Internal AIDB ($IAIDB$)

The $IAIDB$ dissimilarity metric is based on the same spatial smoothness property described above. However, during experimentation, it was observed that some of the artefacts provide internal vertical/horizontal boundary discontinuities which could not be classified by the spatial features described above. The $IAIDB$ dissimilarity metric is designed to provide a measure of the dissimilarity across the internal horizontal/vertical boundaries, as illustrated in Fig. 6. The metric is computed using:

$$IAIDB_{h/v} = \frac{1}{K} \|p_{h/v}^{in} - p_{h/v}^{out}\|_2 \quad (2)$$

where $K = 16$, $p_{h/v}$ represent the horizontal/vertical boundary pixels and $\|\bullet\|_2$ is the Euclidean distance.

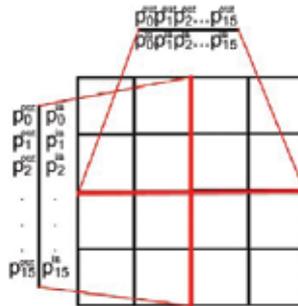


Fig. 6. Graphical representation of the $IAIDB$ Dissimilarity Metric

5.1.4 Average Internal Difference between Subsequent Blocks (AIDSB)

Generally, the pixel transition of an MB and the corresponding MB in the previous frame varies smoothly. Again, since the H.264/AVC design is based on 4x4 transform blocks, each MB is dissected into 16 4x4 blocks.

Let M_t represent the potentially corrupted MB under test and M_{t-1} be the corresponding MB in the previous frame. The AIDSB dissimilarity metric for each 4x4 block b_t and b_{t-1} , shown in Fig. 7, is computed using:

$$AIDSB = \frac{1}{K^2} \|p_t - p_{t-1}\|_2 \tag{3}$$

where K represents the size of the block (in this case $K = 4$), $\|\bullet\|_2$ is the Euclidean distance, and p_t and p_{t-1} represent the pixel of the 4x4 block under test b_t and the corresponding block in the previous MB b_{t-1} respectively. Once the computation is terminated, a set of 16 AIDSB dissimilarity metrics, one for each 4x4, is available.

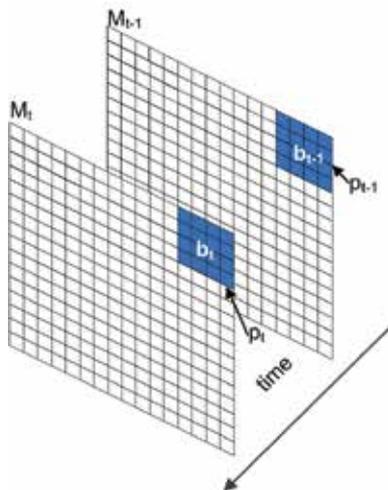


Fig. 7. Graphical representation of the AIDSB Dissimilarity Metric

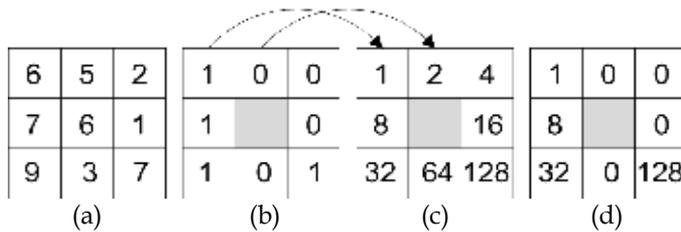


Fig. 8. Computation of the Local Binary Pattern

5.1.5 Texture Consistency (TC)

The Local Binary Pattern (LBP) operator (Ojala, T., et. al, 1996) is a powerful grey-scale invariant texture measure. To understand how it works let us consider Fig. 8. The original 3x3 neighbourhood (Fig. 8(a)) is thresholded by the value of the centre pixel. The values of the pixels in the thresholded neighbourhood (Fig. 8(b)) are multiplied by the binomial

weights given to the corresponding pixels (Fig. 8(c)). The result of this example is illustrated in Fig. 8(d). Finally, the values of the eight pixels are summed to obtain the LBP metric. The LBP histograms of the current MB, h_t , and the corresponding MB in the previous frame h_{t-1} are first computed. The TC dissimilarity metric is then computed by evaluating the histogram insertion method given by:

$$TC = 1 - \sum_{i=0}^{B-1} \min(h_t, h_{t-1}) \quad (4)$$

where B is the number of bins which is set to 256.

5.1.6 Feature vector

The dissimilarity metrics described above exploit both colour and texture consistencies of the MB under test. This set can be reduced to eight dissimilarity metrics without compromising the performance of the classifiers described in the following section. These are:

1. $AIDB$
2. Mean of $IAIDB_{block}$
3. Standard Deviation of $IAIDB_{block}$
4. Vertical $IAIDB$
5. Horizontal $IAIDB$
6. Mean of $AIDSB$
7. Standard Deviation of $AIDSB$
8. TC .

These dissimilarity metrics are then combined together to form the feature vector, which solely describes the reliability of the MB under test. After extensive simulation and testing it was found that these eight dissimilarity metrics provide the best compromise between complexity and performance and therefore adopting higher dimensional feature vector resulted futile.

5.2 Classification methods

Pattern recognition techniques are used to classify some objects into one of the pre-defined set of categories or classes c . For a specific pattern classification problem, a classifier is developed so that objects are classified correctly with reasonably good accuracy. Inputs to the classifier are called features, which are composed of vectors that describe the objects to be classified. The features are designed according to the problem to be solved.

The aim of the pixel-level artefact detection mechanism is to detect the visually distorted MBs to be concealed. The feature extraction module extracts the feature vector which solely defines the reliability of the MB under test. The pattern recognition method then tries to categorize the MB under test in one of the categories based on the statistical information made available by the feature vector.

One of the simplest classification methods is to derive the probability density functions (PDFs) of the dissimilarity metrics representing the uncorrupted and corrupted MBs. The aim of the dissimilarity metrics is to have PDFs similar to the one illustrated in Fig. 9, at which point heuristic thresholds can be employed to detect visually distorted MBs. However, as it will be shown in the simulation results, these dissimilarity metrics provide

limited discriminative power when applied on their own. Furthermore, the distribution of uncorrupted MBs varies with varying video sequences and thus the optimal threshold is sequence dependent.

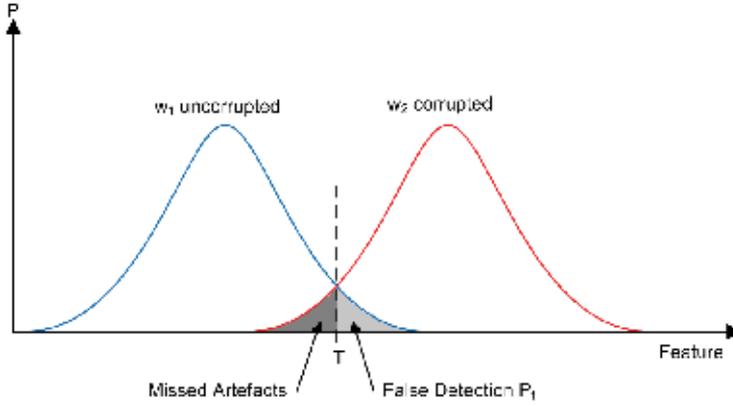


Fig. 9. Probability Density Function of a typical Dissimilarity Metric

Supervised learning classification methods can be employed to solve this problem. These algorithms adopt a set of l training samples (\mathbf{x}, y) , where $\mathbf{x} = [x_1, x_1, \dots, x_n]$ represent the extracted feature vectors, y their corresponding labels (corrupted MB = -1, uncorrupted MB = +1), and n represents the number of dimensions of the feature vector. In supervised learning algorithms, the classifier attempts to learn the input/output functionality from examples to derive an optimal hyperplane to discriminate between the two classes.

5.2.1 Backpropagation Neural Network (BPNN)

A feed-forward neural network is a simple structure where multiple hidden layers are employed. Since it has the potential of approximating a general class of nonlinear functions with a desirable degree of accuracy, it has been employed in many pattern recognition applications (Gupta, M. M. et. al, 2003), (Duda R. O. et. al, 2000). The architecture of the feed-forward neural network is illustrated in Fig. 10.

Every neuron in the hidden layer receives an input vector \mathbf{x} . The output of all the neurons in the hidden layer, represented by a p -dimensional vector $\mathbf{z} = [z_1, z_1, \dots, z_p]$ is fed forward to the neurons in the output layer. The output neurons generate an output vector $\mathbf{f} = [f_1, f_2, \dots, f_m]$. Further, consider the weights corresponding to the i^{th} neuron in the hidden layer to be $\mathbf{w}_i^{(1)} = [w_{i1}^{(1)}, w_{i2}^{(1)}, \dots, w_{in}^{(1)}]$, $i = 1, 2, \dots, p$ and the weights corresponding to the j^{th} neuron in the output layer to be $\mathbf{w}_j^{(2)} = [w_{j1}^{(2)}, w_{j2}^{(2)}, \dots, w_{jp}^{(2)}]^T$, $j = 1, 2, \dots, m$. The input/output relation of the neurons in the network can be expressed as:

$$\text{hidden layer} \begin{cases} s_i^{(1)} = \sum_{k=0}^n w_{ik}^{(1)} x_k \\ z_i = \sigma(s_i^{(1)}) \\ i = 1, 2, \dots, p \end{cases} \quad \text{output neuron} \begin{cases} s_j^{(2)} = \sum_{q=0}^p w_{jq}^{(2)} z_q \\ f_j = \sigma(s_j^{(2)}) \\ j = 1, 2, \dots, m \end{cases} \quad (5)$$

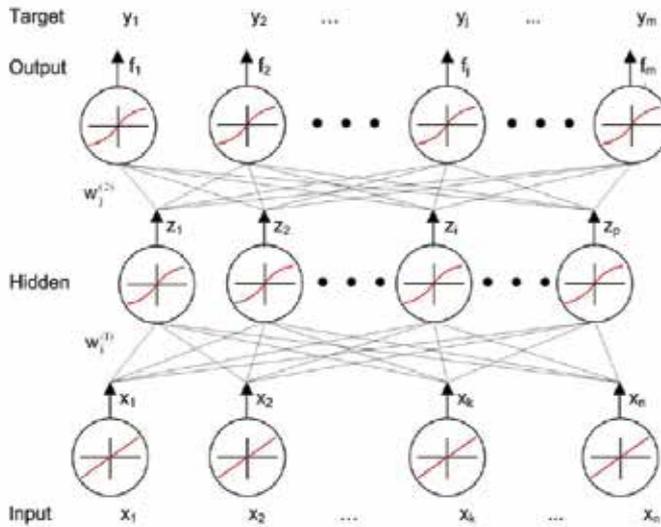


Fig. 10. Fully Connected Feed-Forward Neural Network

where $\sigma(\bullet)$ is a nonlinear activation function which is normally modelled by the sigmoid function as given by:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (6)$$

One of the most popular methods for training feed-forward neural networks is the backpropagation neural network (Rumelhart, D.E., et. al, 1986) (BPNN). The basic approach in learning starts with an untrained feed-forward neural network which is presented with an input training pattern \mathbf{x} and the corresponding output targets \mathbf{y} . The BPNN algorithm then derives the weights which provide the best separating hyperplane which minimises the error function given by:

$$E = \frac{1}{2} \sum_{j=1}^m [y_j - f_j]^2 \quad (7)$$

where E is the cost function and is the measure of the learning performance of the network. The BPNN algorithm is an iterative method which adapts the weight vectors in the direction of decreasing error E (gradient decent) with respect to the weight vectors. The predicted weight difference is given by:

$$\Delta w_i^{(1)} = -\eta \frac{\partial E}{\partial w_i^{(1)}}, i = 1, 2, \dots, p \quad (8)$$

$$\Delta w_j^{(2)} = -\eta \frac{\partial E}{\partial w_j^{(2)}}, j = 1, 2, \dots, m \quad (9)$$

where $0 < \eta < 1$ is the learning rate constant. The weight adapting formulae for the hidden and output layer are as follows:

$$w_i^{(1)}(t+1) = w_i^{(1)}(t) + \eta \sigma'(s_i^{(1)}(t)) x(t) \sum_{l=1}^m \delta_l^{(2)}(t) w_l^{(2)}(t) \quad (10)$$

$$w_j^{(2)}(t+1) = w_j^{(2)}(t) + \eta [y_j(t) - f_j(t)] \sigma'(s_j^{(2)}(t)) z(t) \quad (11)$$

where

$$\delta_j^{(2)}(t) = (y_j(t) - f_j(t)) \sigma'(s_j^{(2)}(t)) \quad (12)$$

The design of the BPNN to solve the artefact detection problem is not trivial. Whereas the number of inputs and outputs is given by the feature space and number of categories respectively, the total number of hidden neurons in the network is not that easily defined. After extensive simulation and testing the best recognition rate was registered when applying a single hidden-layer of 50 neurons and applying a learning rate $\eta = 0.1$.

5.2.2 Probabilistic Neural Network (PNN)

The Probabilistic Neural Network (Specht, D.F., 1988) (PNN) is another feed-forward neural network commonly used for pattern recognition. The PNN classifier forms a Parzen estimate based on l samples, where each sample is represented by a normalized n -dimensional feature vector. The PNN architecture, as illustrated in Fig. 11, consists of n input units, where each unit is connected to each of the l pattern units. Each pattern unit is, in turn, connected to one and only one of the category units. The connections from the input to pattern units represent modifiable weights, which will be derived during the training phase. On the other hand, the connections between the pattern units and the output units have weights which are set to unity.

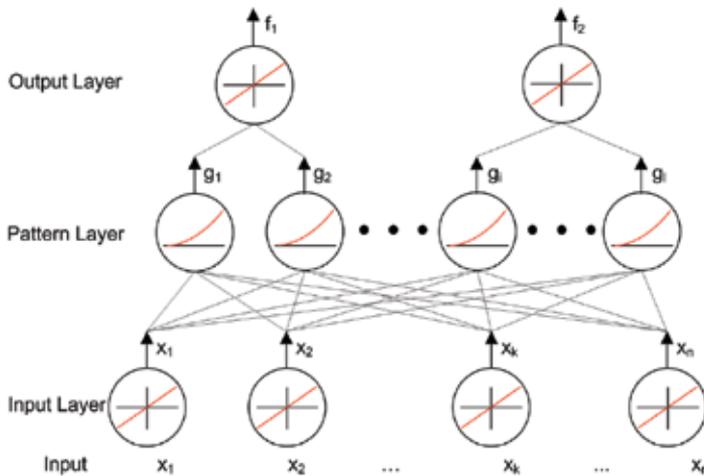


Fig. 11. Architecture of a Probabilistic Neural Network

The training of a PNN is quite straightforward, where at first each pattern x of the l training feature vectors is normalized to have unit length. The first normalized training pattern is then placed on the input units. The modifiable weights linking the input units and the first pattern unit are set such that $w_1 = x_1$. A single connection from the first pattern unit to the output unit corresponding to the known class of that training pattern is created. The process

is repeated for all the remaining training patterns, setting the weights to the successive pattern units such that $w_k = x_k$ for $k = 1, 2, \dots, l$. After this training, we have a network that is fully connected between the input and pattern units, and sparsely connected from pattern to category units.

The trained PNN can now be used for classification by computing the inner product between the inputted feature vector \mathbf{x} and the weight vector w of the k^{th} pattern unit as follows:

$$z_k = \langle w_k, x \rangle \quad (13)$$

where $\langle \bullet \rangle$ denotes the inner product operator. Each pattern unit emits a nonlinear activation function given by:

$$g_k = \exp\left(\frac{(z_k - 1)}{\sigma^2}\right) \quad (14)$$

where σ is the smoothing parameter, which after extensive simulation and testing was set to 0.12. Each output unit then accumulates the values of these activation functions of a given class. The feature vector is simply categorised with the largest output neuron value.

5.2.3 Support Vector Machine (SVM)

A Support Vector Machine (Cortes, C. & Vapnik, V., 1995) (SVM) is a very powerful method that in the few years since its inception has outperformed most other machine learning algorithms in a wide variety of applications. The aim of SVM classification is to derive a separating hyperplane which optimises the generalisation bounds. The generalisation theory gives clear guidance on how to control capacity and hence prevent overfitting by controlling the hyperplane margin measures, while optimisation theory provides the mathematical techniques necessary to find the hyperplane which optimises these measures.

The SVM classifier is a linear learning machine and therefore it may not manage to classify nonlinearly separable data at an acceptable accuracy. Thus, to enhance linear separability, the SVM employs an implicit nonlinear mapping of the data onto a higher dimensional feature space via a positive semi-definite kernel $K(x, y)$, where the SVM tries to derive an optimised separating hyperplane. Several algorithms (Platt, J.C., 1998), (Keerthi, S.S., et. al, 2001) can be found in literature which can be used to train the SVM. These methods try to solve the following quadratic optimisation problem:

$$\begin{aligned} \max_{\alpha} W(\alpha) &= \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j K(\bar{x}_i, \bar{x}_j) \alpha_i \alpha_j, \\ 0 &\leq \alpha_i \leq C, \forall i, \\ \sum_{i=1}^l y_i \alpha_i &= 0 \end{aligned} \quad (15)$$

where α are the Lagrange multipliers and C is the finite penalization constant. The Lagrange multipliers have nonzero values to support vectors (SV) which solely determine the optimal hyperplane. The decision function is given by:

$$f(x) = \text{sign} \left(\sum_{i \in \text{SV}} \alpha_i y_i K(\tilde{x}_i, \tilde{x}) + b \right) \quad (16)$$

where b is the bias term. Following this, the unknown data are classified using the decision function as follows:

$$x \in \begin{cases} \text{Class 1} & \text{if } f(x) > 0 \\ \text{Class 2} & \text{if } f(x) < 0 \\ \text{Unclassifiable} & \text{if } f(x) = 0 \end{cases} \quad (17)$$

For the application at hand, a modified version of the Sequential Minimal Optimization (SMO) algorithm was adopted (Keerthi, S.S., et. al, 2001) to train the SVM classifier. This algorithm was reported to be faster and provides better convergence when compared to the other methods (Platt, J.C., 1999). This classifier utilises the Radial Basis Function (RBF) Kernel which is given by:

$$K(x, y) = \exp \left(\frac{-\|x - y\|^2}{2\sigma^2} \right) \quad (18)$$

where σ is the smoothing parameter. Following extensive testing, the penalisation constant C was set to 60 while the smoothing parameter was set to 1.5.

5.3 Training the classification methods

The syntax analysis check rules used to detect syntax and semantic violation in the H.264/AVC bitstreams only manage to detect 57% of the transmission errors. The errors which do not cause syntax or semantic violations produce different levels of visually impaired regions. As shown in Fig. 12, some of the distorted MBs are very annoying while others are imperceptible. Most of the artefacts caused by transmission errors generally provide imperceptible visual distortions and thus concealing all the MBs contained within a corrupted slice results in concealing a number of MBs which provide minimal or no visual distortion.



Fig. 12. Typical residual artefacts undetected by the syntax analysis rules

In natural video sequences there exists sufficient correlation between spatio-temporal neighbouring MBs. As shown in Fig. 12, the statistics of annoying artefacts significantly

differ from those of uncorrupted MBs. This observation suggests that the design of an artefact detection mechanism which exploits the spatio-temporal redundancies available at pixel level after decoding can be used to detect distorted regions. The aim of the designed method is to maximise the detection of highly distorted MBs which significantly degrade the quality of the decoded frame while being more lenient with imperceptible ones.

To design a robust artefact detection mechanism a set of five video sequences (*Foreman*, *Carphone*, *Mobile*, *Coastguard* and *News*) were used to derive the training and testing data to be used for experimentation. A third set, called the cross-validation set, was made up from another four video sequences (*Miss-America*, *Salesman*, *Akiyo* and *Silent*). These sequences were encoded at QCIF resolution at 30 frames per second and were transmitted over a Binary Symmetric Channel (BSC) at different error rates. From the resulting distorted frames, a population of 3000 MBs is extracted at random. This population is divided into three distinct groups with each group consisting of 500 corrupted MBs and 500 uncorrupted MBs. The first group is used for training while the remaining two groups are used for recognition and cross-validation respectively.

To better analyze the performance of the artefact detection method, the distorted MBs were scaled according to the five-level distortion scale given in Table 1. The experiment consisted of 21 reliable viewers who did not have any experience in image quality evaluation. A subset of 74 corrupted MBs were chosen at random from the training set and were supplied to the viewers for assessment. The viewers have classified these MBs using a methodology similar to the single stimulus test (ITU-T Rec. P.910, 1999). Two experts in the area of multimedia communications have scaled the same images and were used as a reference.

| Distortion Level (DL) | Definition |
|-----------------------|--|
| 4 | Very annoying artefacts |
| 3 | Annoying artefacts |
| 2 | Slightly annoying artefacts |
| 1 | Perceptible but non annoying artefacts |
| 0 | Imperceptible or non-corrupted MB |

Table 1. Definition of the Distortion Levels

Applying the single stimulus test methodology to all the 1500 distorted MBs was prohibitive due to excessive time required. It is much less time consuming to derive this subjective evaluation through the judgement of a group of experts. However, before doing so, it had to be ensured that the results provided by the group of experts represented the opinion of normal users. For this reason, the One-Sample *t*-test was used to compare the opinions provided by the viewers to the reference results based on the subset of 74 corrupted images. This test confirmed that the difference between the means for 83.74% of the images considered during the test is not significant at a 95% confidence level. Furthermore, all the remaining samples have a negative mean difference indicating the judgement provided by the group of experts caters for the most demanding users.

6. Simulation results

The performance of the classification methods employed by the pixel-level artefact detection module is dependent on the dissimilarity metrics. The aim of the dissimilarity metrics employed is to provide small metrics for uncorrupted MBs and large metrics for corrupted

MBs in order to increase the separability in the input space, thus facilitating classification. Since these dissimilarity metrics (except the Texture Consistency) measure colour differences, it becomes clear that the performance of the dissimilarity metrics and thus of the classification method is dependent on the colour space model where these dissimilarity metrics are computed.

To identify the colour space model where to compute the dissimilarity metrics, the *AIDSB* and *AIDB* dissimilarity metrics are considered. These distance measures are computed in the *YUV*, *HSI*, *CIELAB*, and *CIELUV* colour systems, where the heuristic thresholds T are derived using the training set. The heuristic thresholds are selected in such a way that we achieve an acceptable error detection rate P_D at a false detection rate P_F of around 5%. The dissimilarity based classification method is then evaluated based on the data contained in the training set and the results obtained are summarised in Table 2 and Table 3.

| Colour System | Threshold T | P_D (%) | P_F (%) |
|---------------|-------------|-----------|-----------|
| YUV | 7.50 | 65.20 | 5.20 |
| HSI | 0.25 | 71.80 | 5.60 |
| CIELAB | -0.10 | 68.00 | 6.60 |
| CIELUV | 0.25 | 72.80 | 4.60 |

Table 2. Performance of the *AIDSB* dissimilarity metric using different colour systems

| Colour System | Threshold T | P_D (%) | P_F (%) |
|---------------|-------------|-----------|-----------|
| YUV | 8.00 | 31.40 | 4.60 |
| HSI | 0.30 | 40.20 | 4.60 |
| CIELAB | -0.20 | 34.00 | 5.40 |
| CIELUV | 0.20 | 41.80 | 4.20 |

Table 3. Performance of the *AIDB* dissimilarity metric using different colour systems

These results evidence that the dissimilarity-based classification methods perform better when computed in colour systems which better describe the human perception. In particular, computing these dissimilarity measures in the perceptually uniform CIELUV colour space model, generally adopted in television and video display applications, seems to be more beneficial. For this reason, in the remaining part of this section the dissimilarity metrics (except Texture Consistency) are computed in the CIELUV colour space model.

The three supervised learning algorithms described in the previous section can now be trained on the training set using the eight dissimilarity metrics described above to represent each component of the feature vector. The performance of these algorithms is compared to the *AIDB* and *AIDSB* based classifiers and the results obtained are summarized in Table 4. From these results, it can be concluded that the three supervised learning algorithms manage to detect all severely distorted MBs whereas dissimilarity-based classifiers do not. Furthermore, an overall recognition gain of around 20% and 50% is achieved relative to the *AIDSB* and the *AIDB* based classifiers, respectively. Finally, it can be observed that the SVM achieves the best result, where it manages to detect 94.6% of the visual artefacts at P_F smaller than 5%. The gain achieved by the SVM classification method over the other neural approaches occurs mainly because this technique manages to detect more DL2 and DL1 artefacts.

| Classifier | P_D (%) | P_{DL4} (%) | P_{DL3} (%) | P_{DL2} (%) | P_{DL1} (%) | P_F (%) |
|------------|-----------|---------------|---------------|---------------|---------------|-----------|
| AIDB | 41.80 | 86.57 | 45.32 | 18.18 | 2.74 | 4.20 |
| AIDSB | 72.80 | 97.76 | 87.77 | 62.34 | 21.55 | 4.60 |
| BPNN | 92.60 | 100.00 | 100.00 | 93.51 | 63.01 | 7.40 |
| PNN | 92.20 | 100.00 | 100.00 | 92.21 | 63.01 | 4.80 |
| SVM | 94.60 | 100.00 | 100.00 | 94.81 | 73.97 | 4.60 |

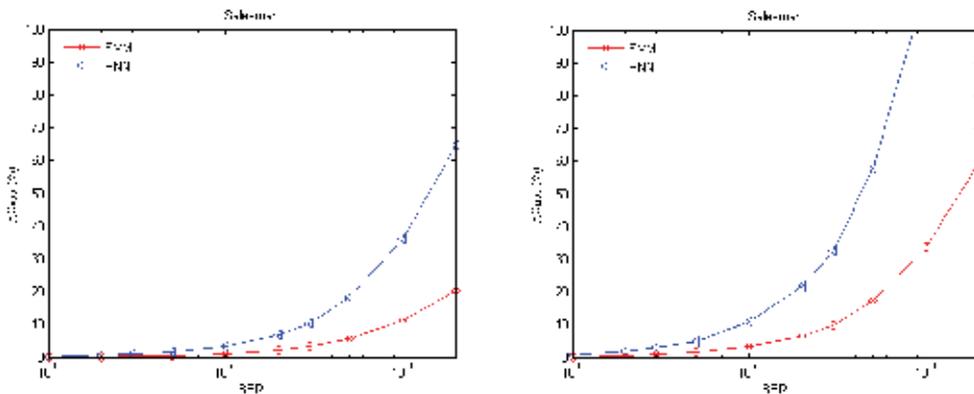
Table 4. Performance of the different classification methods

Both SVM and PNN solutions achieve good artefact detection capabilities with the SVM classifier performing best. However, in order to avoid overfitting, these classifiers are tested on a cross-validation set which contains feature vectors extracted from different video sequences than the ones considered in the training phase. The results are summarised in Table 5, where it can be noticed that the SVM still performs well on these video sequences while the performance of the PNN degrades significantly. Another point in favour of the SVM classifier is that, as shown in Fig. 13, the classification is less computational intensive. This is attributed to the fact that only 156 support vectors are used to derive the separating hyperplane. Given these results, the SVM was integrated within the pixel-level artefact detection method.

| Classifier | P_D (%) | P_{DL4} (%) | P_{DL3} (%) | P_{DL2} (%) | P_{DL1} (%) | P_F (%) |
|------------|-----------|---------------|---------------|---------------|---------------|-----------|
| PNN | 78.89 | 100.00 | 96.23 | 67.69 | 65.45 | 4.80 |
| SVM | 90.95 | 100.00 | 100.00 | 90.77 | 78.18 | 5.20 |

Table 5. Performance of the different classification methods on cross-validation set

The pixel-level artefact detection module is then integrated within the JM software model, where the syntax and semantic violation test procedures are enabled to allow the decoding of partially damaged H.264/AVC bitstreams. The raw video sequences are encoded at QCIF resolution at 15 frames per second at a data rate of 64 kbps with the format IPPP.... The encoder employs slice structuring, where each slice is forced to a size strictly smaller than 100 bytes. The resulting slices are encapsulated within RTP/UDP/IP packets and transmitted over an AWGN channel at different noise levels.

Fig. 13. Complexity analysis of the SVM and PNN methods (left) ΔC_{Add} and (right) ΔC_{Mul} for the *Salesman* sequence

The pixel-level artefact detection method is tested on the *Salesman* video sequence, which is not used during the training phase. The performance of the standard decoder and the modified decoder which employs an SVM classifier at its core is shown in Fig. 14. These results confirm that concealing only those MBs which provide visually distorted regions is beneficial over the standard decoding method, where the PSNR gains over the whole sequence are higher than 0.5 dB at moderate to high error rates. Additionally, PSNR gains of up to 3.90 dB are observed. This superior performance is attributed to the fact that the pixel-level artefact detection method localises the MBs which need to be concealed and thus the over-concealment problem is minimised. Furthermore, reducing the area to be concealed results in improved performance of the error concealment method employed.

The gain achieved by the pixel-level artefact detection method is consistent even when using video sequences which contain fast moving objects. However, in such sequences, it was observed that in the presence of abrupt shots a number of false detections occur which force undistorted MBs to be concealed. Since this method is only employed on distorted slices, the pixel-level artefact detection method will perform at worst like the standard decoder (i.e. detecting all the MBs contained within a slice to be corrupted) hence such cases do not deteriorate the performance compared to the standard.

The performance of the pixel-level artefact detection method is further tested on a frame-by-frame basis and the results are provided in Fig.15 - Fig. 16. These results confirm the superiority of the pixel-level artefact detection method which employs the SVM classifier at its core. The PSNR on a frame-by-frame basis of this method is consistently superior to the standard decoder, where PSNR gains of up to 10.59 dB are achieved. The gain in subjective quality is even more impressive, where it can be seen that concealing undistorted regions provides artefacts which reduce the quality of the video sequence. On the other hand, the pixel-level artefact detection method manages to reconstruct the video sequence at an acceptable level of quality which is quite similar to the original undistorted video sequence.

The artefact detection method can also be employed in conjunction with other standard error-resilient tools such as intra-refresh and FMO. As shown in Fig. 17 and Fig. 18 the pixel-level artefact detection method boosts the performance of the standard error resilient tools by more than 0.3 dB in PSNR at moderate to high error rates, where PSNR gains of up to 2.08 dB are observed. The flexibility of the proposed solution is attributed to the designed features which are used by the classifier to detect distorted regions and the generalisation achieved by the SVM classifier.

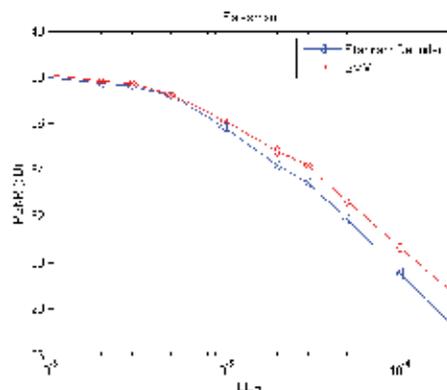


Fig. 14. Performance of the Pixel-Level Artefact-Detection method for the *Salesman* sequence

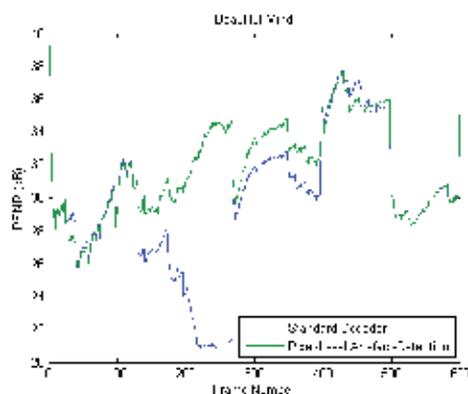


Fig. 15. Performance of this method at a BER of 1.00E-005 for the *Beautiful Mind* sequence

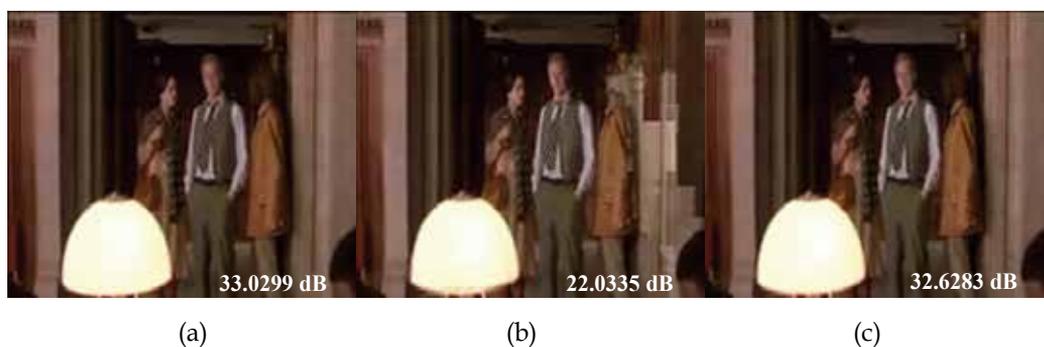


Fig. 16. Frame 215 from the sequence *Beautiful Mind* at 64 kbps (a) reference sequence without errors, (b) standard decoder, and (c) Pixel-Level Artefact Detection method

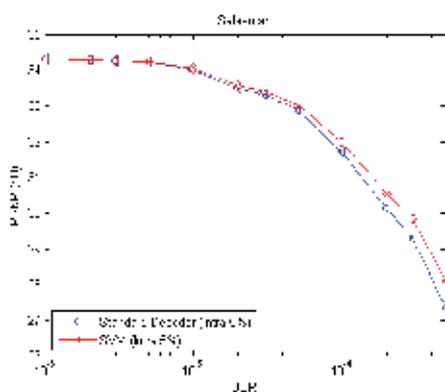


Fig. 17. Performance of the Pixel-Level Artefact-Detection method using Intra Refresh 5% for the *Salesman* sequence

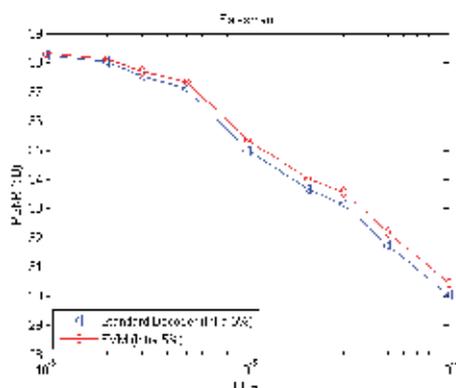


Fig. 18. Performance of the Pixel-Level Artefact-Detection method using Dispersed FMO for the *Salesman* sequence

7. Comments and conclusion

Decoding of partially damaged H.264/AVC generally results in distorted regions which severely affect the quality of the reconstructed video sequence. In this chapter, it was shown that robust pixel-level artefact detection methods can be used to detect those distorted MBs which provide major visual distortion, which are later concealed. In this way, only visually impaired regions are concealed by the decoder resulting in an improved quality of experience. Several classification methods have been considered to solve this problem. After several testing and cross-validation, it is concluded that the SVM classifier achieves the best performance, where it manages to detect 94.6% of the visually distorted regions at a false detection rate of around 5%. More importantly, it provides an unequal important artefact detection strategy, where all annoying artefacts (DL4, DL3) are detected but at the same time the solution is more lenient with slightly annoying artefacts (DL2, DL1).

The method discussed in this chapter makes the H.264/AVC decoder more resilient to transmission errors with overall PSNR gains higher than 0.5 dB being observed at high error rates. This increased robustness does not incur additional bit-rate. In fact the algorithm is computed entirely at the decoder, where it exploits the inherent redundancies available at pixel-level between spatio-temporal neighbouring MBs to detect distorted MBs. Furthermore, the complexity introduced by this method is manageable even at high error-rates making it more applicable to real-time mobile applications. This method can be further applied in conjunction with other error-resilient methods adopted by the standard to boost their performance. Moreover, it is possible to extend this concept to decode other block-based video coding systems such as H.263.

The performance of the classifier adopted by the pixel-level artefact detection method is dependent on the generalisation achieved by the classifier. During testing it was confirmed that the SVM classifier did generalise better than the other classification methods especially when considering video sequences which were not used during the training phase. In fact, the performance of the SVM classifier has managed to outperform the PNN classifier in the cross-validation tests for the H.264/AVC encoded sequences.

During the experiments it was noticed that the method suffered in presence of abrupt screen changes. In fact, due to the temporal dissimilarity metrics, the SVM classifies most of the

MBs affected by the scene change to be artefacts and thus are concealed. However, since the SVM classifier is adopted only on damaged slices, at worst the pixel-level artefact detection method will perform like the standard H.264/AVC implementation where all MBs contained within a corrupted slice are concealed.

8. References

- Adsumilli, C.B., Farias, M.C.Q., Mitra, S.K. & Carli, M., 2005. A Robust Error Concealment Technique using Data Hiding for Image and Video Transmission over Lossy Channels, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 15, no. 11, Nov. 2005, pp. 1394-1406, ISBN 1051-8215
- Barmada, B., Ghandi, M.M., Jones, E.V. & Ghanbari, M. (2005). Prioritized Transmission of Data Partitioned H.264 Video with Hierarchical QAM, *IEEE Signal Processing Letters*, Vol. 12, no. 8, Aug. 2005, pp. 577-580, ISSN 1070-9908
- Bergeron, C. & Lamy-Bergot, C., (2004), Soft-Input Decoding of Variable-Length Codes applied to the H.264 Standard, *IEEE Workshop on Multimedia Signal Processing*, Siena, Italy, 2004
- Budagavi, M. & Gibson, J.D., (2001). Multiframe Video Coding for Improved Performance over Wireless Channels, *IEEE Transactions on Image Processing*, Vol. 10, no. 2, Feb. 2001, pp. 252-265, ISBN 1057-7149
- Buttigieg, V. & Deguara, R., (2005). Using Variable Length Error-Correcting Codes in MPEG-4 Video, *Proceedings of the International Symposium on Information Theory*, Adelaide, Australia, Sep. 2005
- Chen, M., He, Y. & Lagendijk, R.L., (2005). A Fragile Watermark Error Detection Scheme for Wireless Video Communications, *IEEE Transactions on Multimedia*, Vol. 7, no. 2, Apr. 2005, pp. 201-211, ISBN 1520-9210
- Cortes, C. & Vapnik, V., (1995). Support Vector Networks, *Machine Learning*, Vol. 20, pp. 279-297, 1995
- Duda, R.O., Hart, P.E. & Stork, D.G., (2000). *Pattern Classification (Second Edition)*, Wiley-Interscience, ISBN 978-0-471-05669-0, New York
- Ekmekci Flierl, S., Sikora, T. & Frossard, P., (2005). Coding with Temporal Layers or Multiple Descriptions for Lossy Video Transmission, *Proceedings of the International Workshop Very Low Bit-rate Video*, Sardinia, Italy, Sep. 2005
- Farrugia, R.A. & Debono, C.J., (2007). Enhancing the Error Detection Capabilities of the Standard Video Decoder using Pixel Domain Dissimilarity Metrics, *Proceedings of the IEEE International EUROCON Conference*, Warsaw, Poland, Sep. 2007
- Ghanbari, M. (2003). *Codec Standards: Image Compression to Advanced Video Coding*, IET, ISBN 978-0852967102, London
- Girod, B. & Färber, N. (1999). Feedback-Based Error Control for Mobile Video Transmission, *Proceedings of IEEE*, Vol. 97, Oct. 1999, pp. 1707-1723, ISSN 0018-9219
- Gonzalez, R.C., & Woods R.E., (2008). *Digital Image Processing*, Pearson Prentice Hall, ISBN 978-0-13-505267-9, New Jersey
- Goyal, V.K. (2001). Multiple Description Coding: Compression Meets the Network, *IEEE Signal Processing Magazine*, Vol. 18, pp. 74-93, Sep. 2001
- Guillemot, C. & Siohan, P., (2005). Joint Source-Channel Decoding of Variable-Length Codes with Soft Information: A Survey, *Eurasip Journal on Applied Signal Processing*, Vol. 2005, no. 6, (2005), pp. 906-927

- Gupta, M.M, Jin, L. & Homma, N., (2003). *Static and Dynamic Neural Networks: From Fundamentals to Advanced Theory*, IEEE Press, ISBN 978-0-471-21948-4, New Jersey
- ITU-T Rec. P.910, (1999), *Subjective video quality assessment methods for multimedia applications*
- Keerthi, S.S., Shevade, S.K., Bhattacharyya, C. & Murthy, K.R.K., (2001). Improvements to Platt's SMO Algorithm for SVM Classifier Design, *Neural Computing*, Vol. 13, no. 3, pp. 637-649, Mar. 2001
- Khan, E., Lehmann, S., Gunji, H. & Ghanbari, M., (2004). Iterative Error Detection and Correction of H.263 Coded Video for Wireless Networks, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 14, no. 12, Dec. 2004, pp. 1294-1307, ISSN 1051-8215
- Kumar, G., Xu, L., Mandal, M.K. & Panchanathan, S., 2006. Error Resiliency Schemes in H.264/AVC Standard, *Elsevier Journal of Visual Communications and Image Representation*, Vol. 17, no. 2, Apr. 2006, pp. 425-450
- Liebl, G., Schierl, T., Wiegand, T. & Stockhammer, T., (2006). Advanced Wireless Multiuser Streaming using the Scalable Video Coding Extensions of H.264/MPEG4-AVC, *Proceedings of the IEEE International Conference on Multimedia and Expo*, Toronto, Canada, Jul. 2006
- Lin, S. & Costello, D.J., (1983), *Error Control Coding: Fundamentals and Applications*, Prentice-Hall, ISBN 9787-0132837965, New Jersey
- Liu, L., Zhang, S., Ye, X. & Zhang, Y., (2005), Error Resilience Schemes of H.264/AVC for 3G Conversational Video Services, *IEEE Proceedings of International Conference on Computer and Information Technology*, Shanghai, China, 2005
- Nemethova, O., Forte, G.C & Rupp, M., (2006). Robust Error Detection for H.264/AVC using Relation based Fragile Watermarking, *Proceedings of International Conference on Systems, Signals and Image Processing*, Budapest, Hungary, Sep. 2006
- Nguyen, H. & Duhamel, P., (2003). Estimation of Redundancy in Compressed Image and Video data for Joint Source-Channel Decoding, *IEEE Global Telecommunications Conf.*, San Francisco, USA, Dec. 2003
- Nguyen, H. & Duhamel, P., (2005). Robust Source Decoding of Variable-Length Encoded Data taking into account Source Constraints, *IEEE Transactions on Communications*, Vol. 53, no. 7, Jul. 2005, pp. 1077-1104, ISSN 0090-6778
- Nguyen, H., Duhamel, P., Broute, J. & Rouffet, F., (2004). Optimal VLC Sequence Decoding Exploiting Additional Video Stream Properties, in *IEEE Proceedings International Conference on Acoustic, Speech and Signal Processing*, Montreal, Canada, May 2004
- Ojala, T., Pietikäinen, M. & Harwood, D., (1996). A Comparative Study on Texture Measures with Classification based on Feature Distributions, *Pattern Recognition*, Vol. 29, no. 1, Jan. 1996, pp. 51-59
- Park, W. & Jeon, B., (2002). Error Detection and Recovery by Hiding Information into Video Bitstreams using Fragile Watermarking, *Proceedings of SPIE Visual Communications and Image Processing*, Vol. 4671, Jan. 2002, pp. 1-10
- Platt, J.C., (1998). Fast Training of Support Vector Machines using Sequential Minimal Optimization, *Advances in Kernel Methods: Support Vector Learning*, MIT Press, ISBN 0-262-19416-3, Cambridge
- Platt, J.C., (1999). Using Sparse and Analytic QP to Speed Training of Support Vector Machines, *Advances in Neural Information Processing Systems*, MIT Press, ISBN 0-262-11245-0, Cambridge

- Rane, S., Baccichet, P. & Girod, B., (2006). Modeling and Optimization of a Systematic Lossy Error Protection System based on H.264/AVC Redundant Slices, *Proceedings on IEEE Picture Coding Symposium*, Beijing, China, Apr. 2006
- Richardson, I.E.G (2003). *H.264 and MPEG-4 Video Compression: Video Coding for Next Generation Multimedia*, Wiley, ISBN 0-470-84837-5, New York
- Roodaki, H., Rabiee, H.R. & Ghanbari, M., (2008). Performance Enhancement of H.264 Coded by Layered Coding, *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing*, Las Vegas, USA, Apr. 2008
- Rumelhart, D.E., Hiltin, G.E. & Williams, R.J., (1986). Learning Internal Representations by Error Propagation, *Parallel Processing: Explorations in the Microstructure of Cognition*, MIT Press, Vol. 1, ISBN 0-262-68053-X, Cambridge
- Sabeva, G., Ben Jamaa, S., Jieffer, M. and Duhamel, P., (2006). Robust Decoding of H.264 Encoded Video Transmitted over Wireless Channels, *IEEE Workshop on Multimedia Signal Processing*, Victoria, Canada, Oct. 2006
- Schwarz, H., Marpe, D. & Hannuksela, M.M., (2006). Overview of the Scalable H.264/MPEG4-AVC Extension, *Proceedings of the IEEE Conference on Image Processing*, Atlanta, USA, Oct. 2006
- Specht, D.F., (1988). Probabilistic Neural Networks for Classification, Mapping or Associative Memory, *Proceedings of the IEEE Conference on Neural Networks*, San Diego, California, Jul. 1988
- Stockhammer, T. & Hannuksela, M.M. (2005). H.264/AVC Video for Wireless Transmission, *IEEE Wireless Communications*, Vol. 12, no. 4, Aug. 2005, pp. 6-13, ISBN 1536-1284
- Stockhammer, T., Hannuksela, M.M. & Wiegand T. (2003). H.264/AVC in Wireless Environments, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 13, no. 7, Jul. 2003, pp. 657-673, ISBN 1051-8215
- Superiori, L., Nemethova, O. & Rupp, M., (2006). Performance of a H.264/AVC Error Detection Algorithm based on Syntax Analysis, *Proceedings of International Conference on Advances in Mobile Computing and Multimedia*, ISBN, Yogyakarta, Indonesia, Dec. 2006
- Superiori, L., O. Nemethova, O. & Rupp, M., (2007). Detection of Visual Impairments in the Pixel Domain of the Corrupted H.264/AVC Packets, *IEEE Proceedings of the International Picture Coding Symposium*, Lisbon, Portugal, Nov. 2007
- Tilio, T., Grangetto, M. & Olmo, G., (2008). Redundant Slice Optimal Allocation for H.264 Multiple Description Coding, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 18, no. 1, Jan. 2008, pp. 59-70, ISSN 1051-8215
- Wang, G., Zhang, Q., Zhu, W. & Zhang, Y-Q. (2001), Channel-Adaptive Unequal Error Protection for Scalable Video Transmission over Wireless Channel, *Proceedings of SPIE Visual Communications Image Processing*, San Jose, USA, Jan. 2001
- Wang, J.T. & Chang, P.C., (1999). Error Propagation Prevention Technique for Real-Time Video Transmission over ATM Networks, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 9, no. 3, Apr. 1999, pp. 513-523, ISBN 1051-8215
- Wang, Y. & Lin, S., (2002). Error-Resilient Video Coding using Multiple Description Motion Compensation, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 12, no. 6, Jun. 2002, pp. 438-452, ISBN 0-7803-7025-2
- Wang, Y, Reibman, A.R. & Lin, S., (2005). Multiple Description Coding for Video Delivery, *Proceedings of IEEE*, Vol. 93, no. 1, Jan. 2005, pp. 57-70, ISSN 0018-9219

- Weidmann, C., Kadlec, P., Nemethova, O. and Al-Moghrabi, A., (2004). Combined Sequential Decoding and Error Concealment of H.264 Video, *IEEE Workshop on Multimedia Signal Processing*, Siena, Italy, Oct. 2004
- Welzl, M., (2005). Passing Corrupted Data Across Network Layers: An Overview of Recent Development and Issues, *EURASIP Journal on Applied Signal Processing*, Vol. 2005, Jan. 2005, no. 2, 2005, pp. 242-247, ISBN 1110-8657
- Wenger, S. & Horowitz, M. (2002). Flexible MB ordering – a new error resilience tool for IP-based video, *Proceedings of Tyrrhenian International Workshop on Digital Communications*, Capri, Italy, Sep. 2002
- Wenger, S., (2003). H.264/AVC over IP, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 13, no. 7, Jul. 2003, pp. 645-656, ISBN 1051-8215
- Wiegand, T. & Sullivan, G.J.,(2007). The H.264/AVC Video Coding Standard, *IEEE Signal Processing Mag.*, Vol. 24, no. 2, Mar. 2007, pp. 148-153, ISBN 1053-5888
- Ye, S., Lin, X. & Sun, Q., (2003). Content Based Error Detection and Concealment for Image Transmission over Wireless Channels, *Proceedings of the IEEE International Symposium on Circuits and Systems*, Bangkok, Thailand, May 2003
- Zhang, C. & Xu, Y., (1999). Unequal Packet Loss Protection for Layered Video Transmission, *IEEE Transactions on Broadcasting*, vol. 45, no. 2, Jun. 1999, pp. 243-252, ISBN 0018-9316799
- Zhu, C., Wang, T-K., & Hannuksela, M.M., (2006). Error Resilient Video Coding using Redundant Pictures, *Proceedings of IEEE International Conference on Image Processing*, Atlanta, GA, USA, Oct. 2006

Digital Videos Broadcasting via Satellite – Challenge on IPTV Distribution

I Made Murwantara, Pujiyanto Yugopuspito,
Arnold Aribowo and Samuel Lukas
*Faculty of Computer Science,
Universitas Pelita Harapan
Indonesia*

1. Introduction

Digital Video Broadcasting via Satellite (DVB-S) is achieving its new order. DVB-S penetration which are reaches a lot of viewer around the world was the evidence. Its capability to deliver high quality digital video primary, with reasonable price was the key success factor. Main factor that made DVB-S technology more mature was the innovation. In recent years, DVB-S is capable to deliver internet traffic. One-way downlink is transmission from satellite to customer location and others media transmission for uplink from customer to the internet. This kind of activities is being done for many years as an inexpensive internet services and as a redundant network link for low speed internet provider.

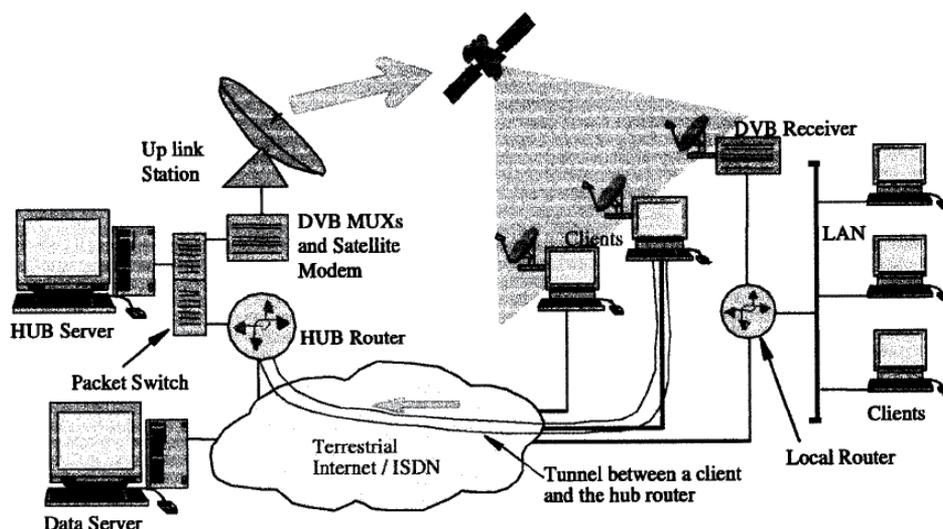


Fig. 1. DVB-S Satellite Network (Nihal 1999)

As the satellite technology increased, two-way satellite communication is served the community everywhere around the globe. The uplink transmission established by L-band

and the opposite direction served by Ku-Band or C-Band. It have to be established the communication on two different frequency with specific wavelength to prevent signal interference and to meet the International Telecommunication Union (ITU) standard. The transmission reception is considered by the climate and geographical of customer location. For high humidity area, such as Asia, it is wise to use C-Band frequency as its downlink transmission. On the other hand, European countries will be effective to use Ku-Band frequency, otherwise the services can not be delivered on maximal standard.

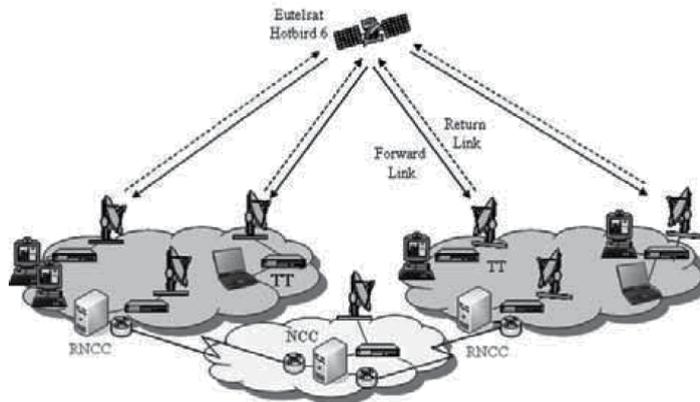


Fig. 2. DVB-RCS Network Architecture (Helmut, rklin et al. 2007)

The Internet Protocol Television (IPTV) is a mechanism of transmitting its information to their customer using internet protocol as media transmission. Because IPTV can be broadcast on the internet protocol, then every media that is capable to communicate using internet protocol is also the media of IPTV distribution. Those are including wired and wireless media.

The main interesting part on internet is the development of IP-based broadcasting of "digital" television. This issue bring a lot of investor and technologist to become the pioneer on such technology. And their invention makes a big leap of telecommunication technology, especially how people communicate on low cost (Alberto, Francesco et al. 2006). Delay Timing is also constrain that influence the performance of satellite transmission, exclude the weather condition for specific region.

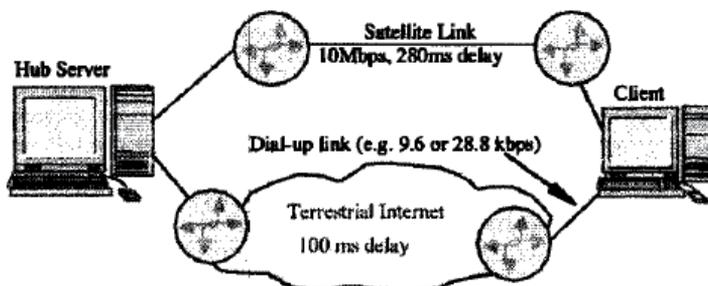


Fig. 3. Satellite Transmission Time Delay(Nihal 1999)

This chapter will be organized as follows. It starts with an overview of IPTV and DVB-S standard and its latest technology in Section 2. In Section 3, the DVB-S technology in relation

to IPTV distribution is explained. Then on Section 4, the challenge of IPTV distribution over DVB-S will be given and Section 5 concludes the chapter.

2. Review on DVB-S and IPTV

Digital Video Broadcasting via Satellite (DVB-S) has been served more than a decade. This technology has been emerged as a leading standard to provide video streaming services through the satellite media. Since its introduction, DVB has been used on terrestrial (DVB-T), cable (DVB-C) and satellite (DVB-S) (Helmut, rklin et al. 2007). Those outcomes are the result of demands that communication technology is an unlimited area.

DVB Return Channel via Satellite (DVB-RCS) is established as the answer of demand of Internet technology via satellite. DVB-RCS provide a new era of IP relay via satellite. This breakthrough is bringing good news to the internet providers company. It could enlarge the covering area of most providers in the entire world.

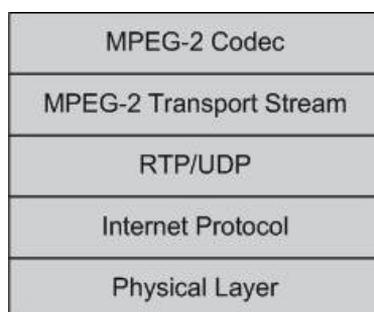


Fig. 4. DVB-IP Protocol Simplified

The development of internet protocol of version 4 (IPv4) for satellite communication had been established since the introduction of DVB-S (Helmut, rklin et al. 2007). Quality of Services (QoS) and packet management are being improved everyday. But the most important achievement is the streaming reliability that fully demanded by customers. In addition, the transformation from IPv4 to IPv6 was on going at satellite network. This challenge has come to DVB protocol to support internet services.

DVB-S systems typically make use of multicast. Multicast communication is predicated on the need to send the same content to multiple destinations simultaneously. Groups or individual who receive the transmission would dynamically change their channel or keep tuned on one favourite channel. Meanwhile the multicast transmission on DVB-S is transmitting hundreds of channel on single frequency. User management on multicast transmission is using the MAC address of the receiver.

The distribution of new technology above DVB-S protocol is a great challenge. How this protocol satisfied their consumer by carrying IPTV on its transmission will be answer on the following pages.

Table 1. shows the DVB specs for E_b/N_0 (energi per bit over noise) required for various modulation and coding rates. An E_b/N_0 of 6-9 dB is desirable. In this satellite design jitter is absent. A set of international standards for digital TV has been developed to follow up on the demand of the digital video quality standard by the DVB Project. It is an industrial consortium with about 300 members and published by a Joint Technical Committee (JTC) of ETSI, CENELEC and European Broadcasting Union (EBU). These are collectively known as

| Coding/Modulation | | Required Eb/No (dB) |
|-------------------|-------|---------------------|
| DVB/QPSK | R 1/2 | 4.5 |
| | R 2/3 | 5.0 |
| | R 3/4 | 5.5 |
| | R 7/8 | 6.4 |
| DVB/8PSK | R 2/3 | 6.9 |
| | R 5/6 | 8.9 |
| Turbo QPSK | R 1/2 | 2.7 |
| | R 5/8 | 3.3 |
| | R 3/4 | 3.9 |
| | R 5/6 | 4.5 |
| | R 7/8 | 4.7 |
| Turbo 8PSK | R 2/3 | 5.4 |
| | R 3/4 | 6.4 |
| | R 5/6 | 7.7 |

Table 1. Link Performance Requirements for Satellite Links Operating with Various Modulation Technologies (Minoli 2008)

Digital Video Broadcast. IPTV using some standards of DVB technology, particularly on satellite links.

Standards of DVB exist for the last decade is defining for the physical layer and data-link layer of a distribution system, as follows:

- Satellite video distribution (DVB-S)
- Cable video distribution (DVB-C)
- Terrestrial television video distribution (DVB-T)
- Terrestrial television for handheld mobile devices (DVB-H)

Distribution systems vary mainly in the modulation schemes used:

- DVB-S uses QPSK.
- DVB-S2 uses QPSK, 8PSK, 16APSK.
- DVB-C (VHF/UHF) uses Quadrature Amplitude Modulation (QAM).
- DVB-T (VHF/UHF) uses 16-QAM or 64-QAM combined with Coded Orthogonal Frequency Division Multiplexing (COFDM).

DVB devices interact with the physical layer via Synchronous Serial Interface or Asynchronous Serial Interface (ASI). All information is transmitted in MPEG-2 transport with some additional constraints.

The DVB is using MPEG2-TS as its transport stream. The MPEG2-TS is mapping the Packet Elementary System (PES) which was produced by MPEG-2 processor. MPEG2-TS are encapsulated in the User Datagram Protocol (UDP) afterwards in IP.

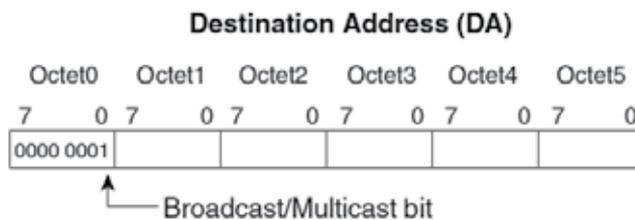


Fig. 5. MAC Address (Minoli 2008)

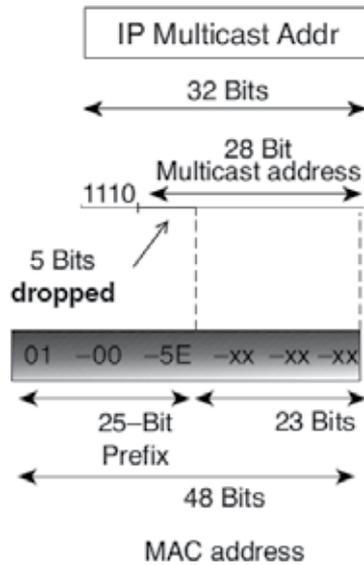


Fig. 6. Scheme of IP Multicast to MAC Address (Minoli 2008)

Each IPTV stream carried a specific Packet Identification Data (PID). PID is used as an identification parameter for specific receiving group. It's have to be exactly the same as setup on the receiver. If this parameter has been setup not properly, then the transmission could not be seen on the screen. Technically it is called the out of tuned condition. DVB-based applications will configure the driver in the receiver to passing up the packet of specified PID. Afterwards, the receiver will extract streams from the TS partially by looking the same PID existed on the packet.

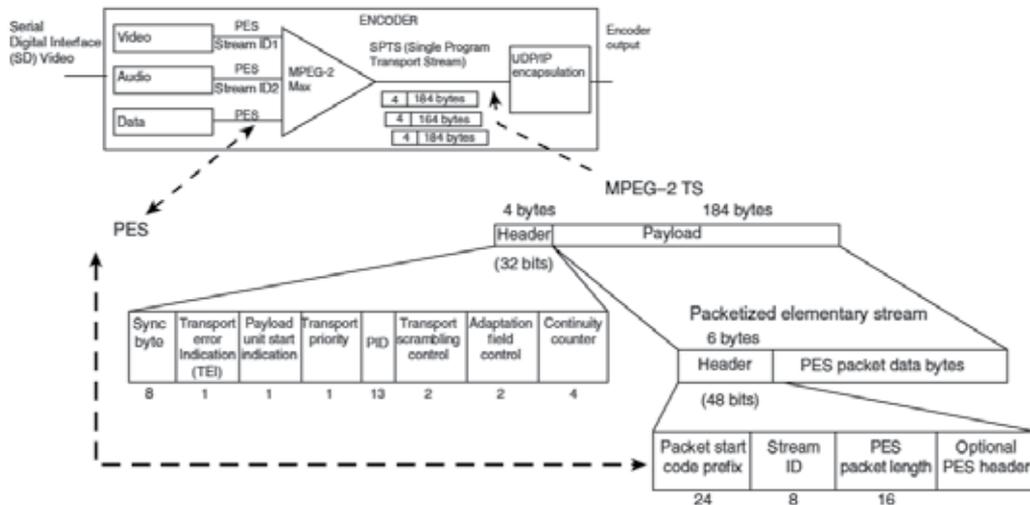


Fig. 7. TS and PES Multiplexing (Minoli 2008)

IPTV stream consist of packets of fixed size, each of which carries a stream-identifying number called PID. These packets are aggregated into an IP packet, the IP packet is

transmitted using multicast methods. Each PID contains specific video, audio or data information.

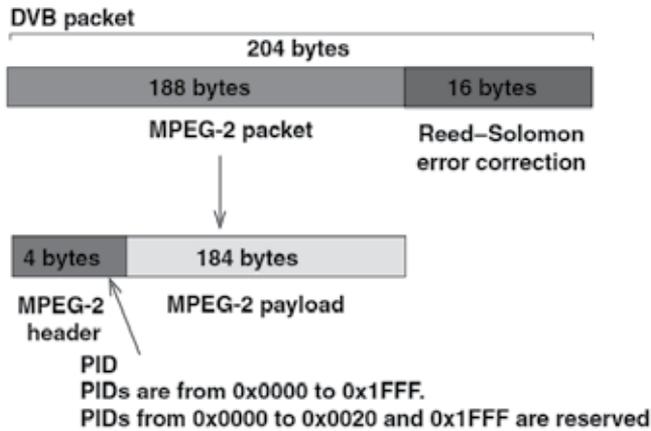


Fig. 8. DVB Packet (Minoli 2008)

For satellite transmission and to remain adequate with existing MPEG-2 technology, TS are encapsulated in Multiprotocol Encapsulation (MPE) and then segmented and placed into TSs via a device called IP Encapsulator (IPE). MPE is used to transmit datagram that exceed the length of the DVB packet size.

IPE handle statistical multiplexing and facilitate coexistence. IPE receives IP packets from and Ethernet connection and encapsulates packets using MPE and then maps these streams into an MPEG-2 TS. Once the device has encapsulated the data, the IPE forwards the data packets to satellite link.

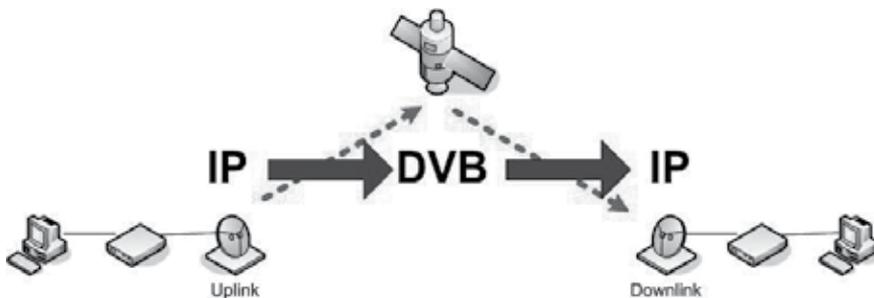


Fig. 6. IP to DVB encapsulation

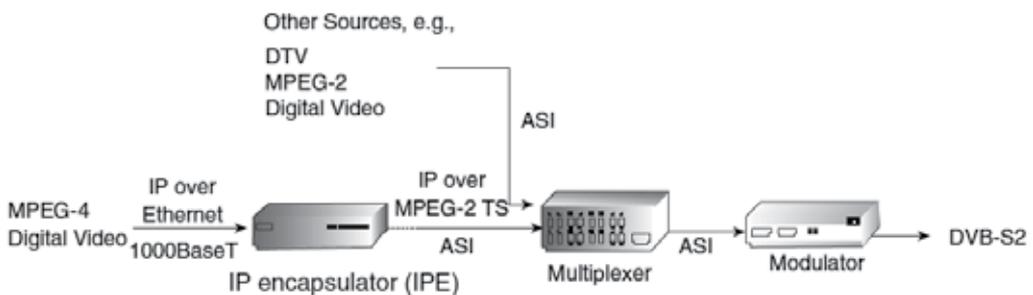


Fig. 7. Encapsulator Function

Data for transmission over the MPEG-2 transport multiplex is passed to an encapsulator that commonly receives Ethernet frames, IP datagram, or other network layer packets. It formats each PDU into a series of TS packets, which sent over a TS logical channel.

3. IPTV distribution on DVB-S

The two basic types of multicast distribution trees provided for IPTV distribution are source trees and shared trees (O'Driscoll 2008). Messages are replicated only where the tree branches. Both trees are loop-free topologies. A source tree is based on the principle of identifying the shortest path through network from source to destination. Due to the fact that source trees identify shortest path, they called shortest path trees (SPT). A new source tree is generally configured when new source servers are added to the IPTV network as can be seen on figure 8. The configuration of a shared three is different to a source tree in the sense that the shared multicast distribution locates the root at a chosen point on the network called a rendezvous point (RP). The RP acts as an intermediate device between IPTV sources and IPTV user. This contrasts with source trees that locate their routes at the source of the IPTV content. The shared tree is shown on figure 9.

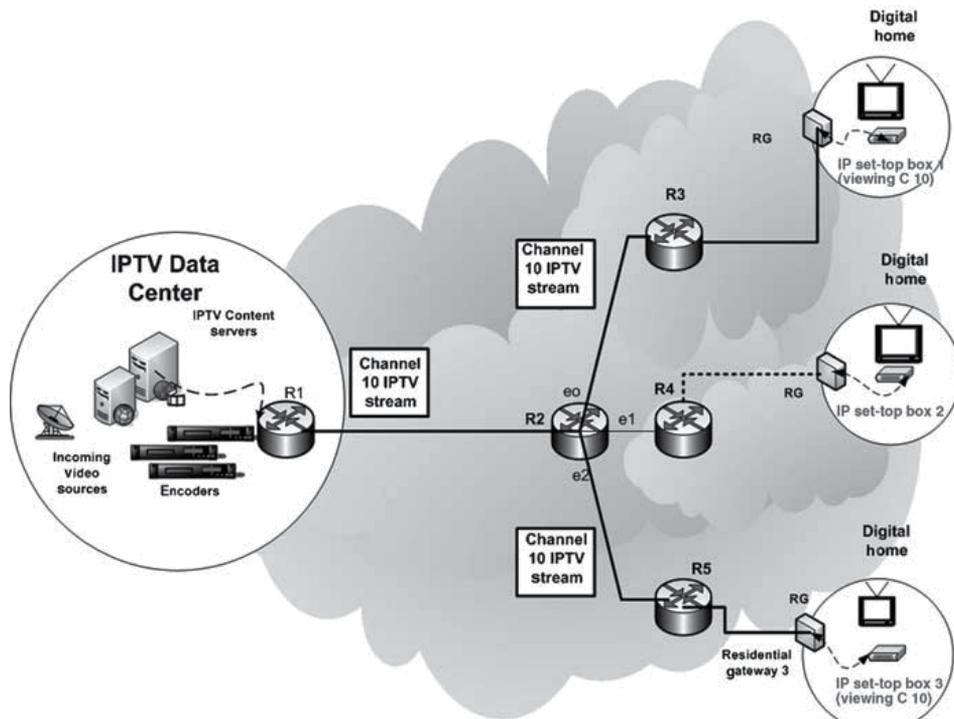


Fig. 8. Source Tree (O'Driscoll 2008)

Distribution of a high resolution stream is not easy to manage and high bandwidth consume. On satellite stream distribution, IPTV could not distribute on effective and efficient manner. Mainly about delay that occur on satellite transmission. To deal with this problem, we propose the cache stream management that provide IPTV streams management on one-way DVB-S environment.

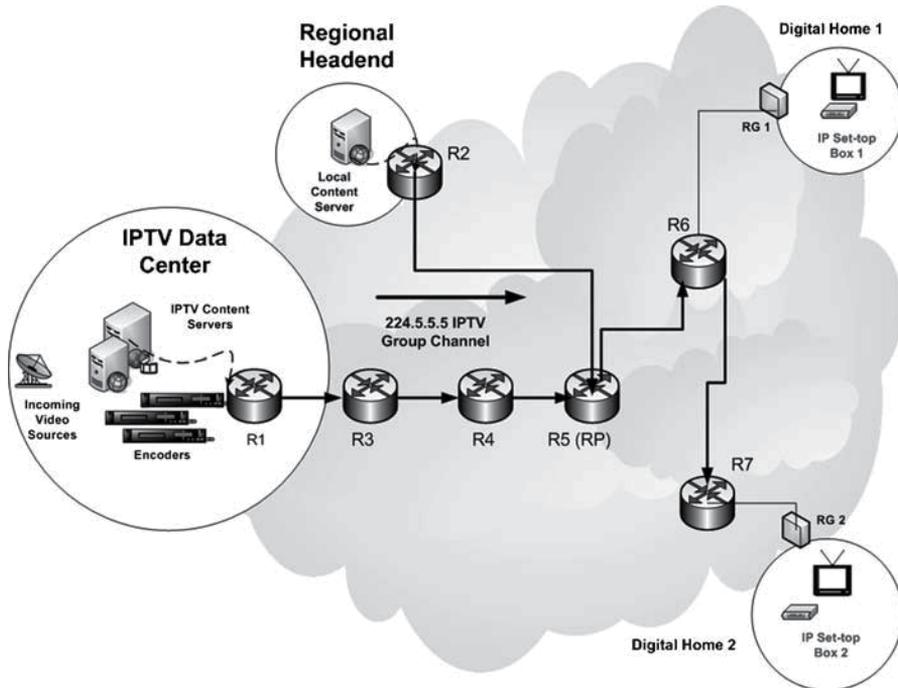


Fig. 9. Shared Tree(O’Driscoll 2008)

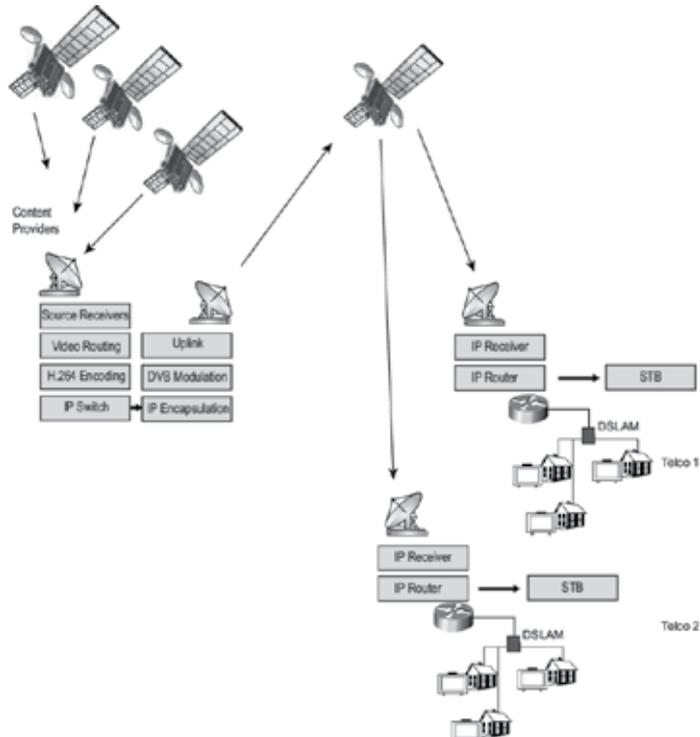


Fig. 10. Single Source IPTV (Minoli 2008)

Cache Stream Management (CSM) is an IPTV cache stream system. CSM do not utilize the network performance overall. Indeed, it is a simple solution for IPTV distribution over DVB-S. The CSM consists of grid computing facility which is doing request analyzes, and then makes a schedule for each request specifically to the cache manager.

In this chapter, IPTV distribution is only cover single source downlink via satellite and uplink via local network. The local network capacity is considered fulfil the request signalling to the IPTV provider, besides cost for stream uplink via satellite is very expensive.

| Application | Design Concern | | | | | | | Current Technology | Future Technology |
|------------------------|----------------|----------------------|------------------|--------------|---------------|----------|------------------|-------------------------|------------------------|
| | Low Bitrate | Multiformal Decoders | Right Management | High Quality | Random Access | Low Cost | Error Robustness | | |
| Internet Streaming | ■ | ■ | ■ | | | | | WMV9, MPEG-4, Quicktime | |
| Digital terrestrial TV | | | | ■ | | | ■ | MPEG-2 | MPEG-2, H.264, AVS 1.0 |
| Satellite TV | ■ | | | ■ | | | ■ | MPEG-2 | H.264 HP |
| IPTV | ■ | | ■ | ■ | | | ■ | MPEG-2, MPEG-4 | MPEG-4 |

Table 1. Compression Characteristics

Gotta(Alberto, Francesco et al. 2006), shows the output of their experiment on figure 9 using a single channel transmission delay via satellite. Their experiment was using a multi frame which is made of 8 frames, each containing 6 time slots. Every time slot is composed by 8 cells (MPEG-2 packets).

The response to an offered traffic impulse is a throughput transient with amplitude R_{min} lasting for T_a , followed by a throughput transient with amplitude $R + R_{min}$ ending at T_t , after which the throughput is equal to the offered traffic. All the following results are valid for $R > R_{min}$.

The duration of the throughput transient T_t depends on the impulse amplitude. In fact, the moment the source starts sending packets at rate R , the service rate allocated to the TT is R_{min} , meaning that the queue at the TT fills up at a rate $R - R_{min}$. After an allocation delay of T_a , the scheduler assigns the bandwidth $R + R_{min}$ and the queue empties at a rate R_{min} (that is, the difference between the output and the input rates). The queue length reaches its maximum $T_a(R - R_{min})$ after a time equal to the allocation delay T_a . Since the total duration T_t of the transient is the sum of the allocation delay and the time required to empty the queue.

$$T_t = T_a + T_a \frac{R - R_{min}}{R_{min}} = T_a \frac{R}{R_{min}} \quad (1)$$

Bottom graphs in Fig. 8 show results for the one-way delay. The peak amplitudes during the transient weakly depend on transmission rates. Indeed, the maximum queuing delay T_d is given by the ratio between the maximum queue size and the offered traffic rate where τ is the satellite system latency.

$$T_d = \tau + \left(1 - \frac{R_{min}}{R}\right) \quad (2)$$

Gotta's et. al. experiment results emphasize the delay that could not be avoided on satellite transmission. Their result shows the need of CSM to be implemented on IPTV distribution via satellite.

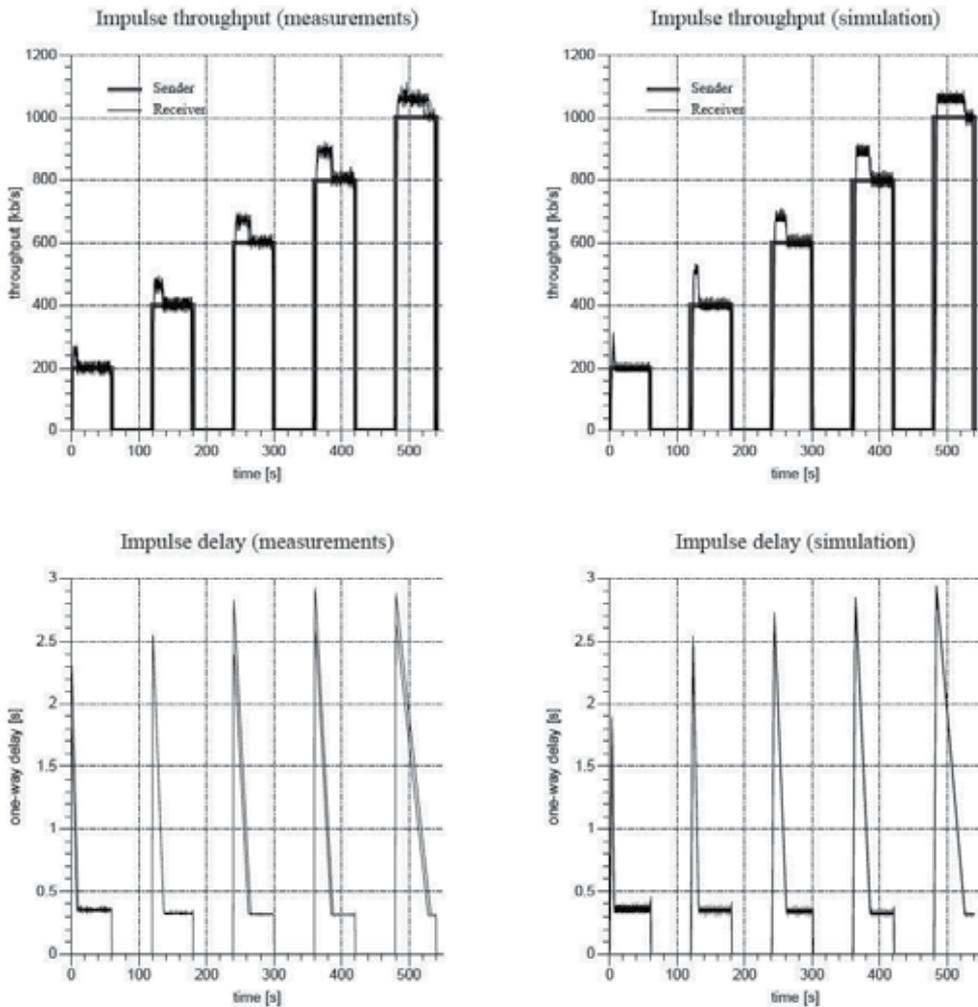


Fig. 11. Throughput and one-way delay (Alberto, Francesco et al. 2006)

4. Challenge on IPTV distribution

Problems on satellite transmission are always about compression and delay. As compression playing a big role on device manufacturing and transmission standard, its existence must comply on DVB organizational standard. While delay, it is can not be anticipated on current satellite technology.

And we proposed CSM as a framework that is projected to overcome such problem. On Cache Stream Management, its system is separated into three critical sections. First section is Request Manager (RM), this RM must ensure that every incoming request are served. Each

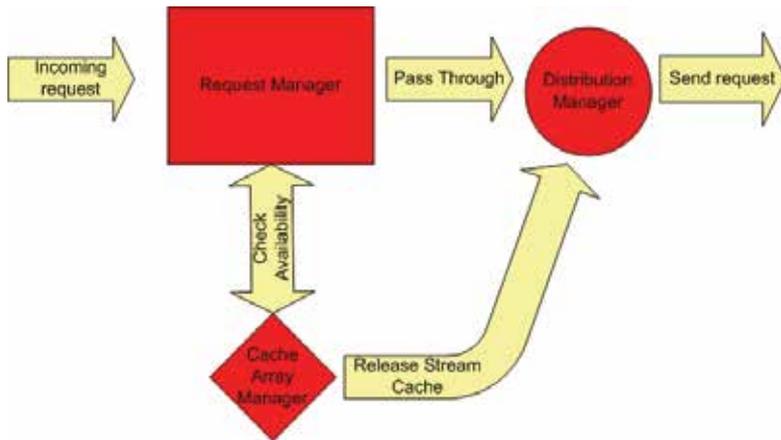


Fig. 12. Cache Stream Management

request in particular will be check to the availability on Cache Array Manager (CAM), if the request is match then the content will be delivered based on the cache source. If it is not, then distribution manager will retrieved the data from source. For real time events (i.e. sports), CAM will act as a validation content to user that want to replay the events. In particular, CSM is not useful if the main backbone is a terrestrial or cable network because the aim of CSM is to support remote area with low upstream network capacity. This happen based on the capability of those both technologies to deliver a multicast session without significant delay. But to the satellite technology, cache system is playing a vital role on stream delivery.

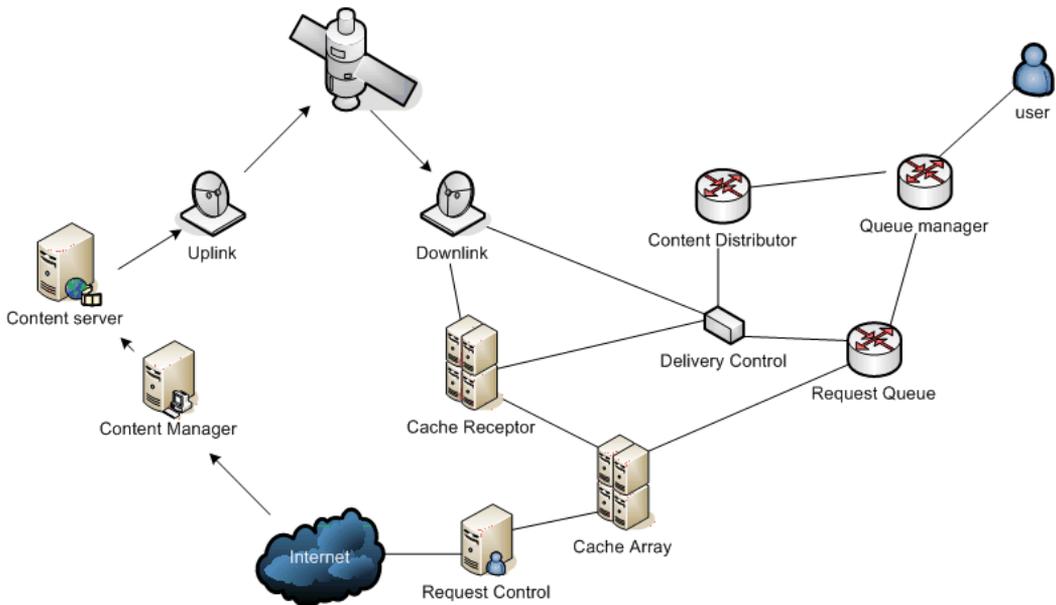


Fig. 13. Cache Stream Management Implementation.

In CSM implementation, incoming content is managed by Cache Receptor. The Cache Receptor is acting like a grid computing storage system. It will store every incoming content

and waiting for an instruction from the Delivery Control. If there is no request then the content will be record as there are no content send to the Delivery Control.

Cache Array is a control request manager. It will checking for a request whether it is exist on the cache request or not, if it is exit then the Cache Receptor will send that request via delivery control. On IPTV this request is a channel. The Cache Array keeps updating its content information database by Cache Receptor. Its operation supported by Request Queue that is triggering the Delivery Control to pass the request.

Particularly, every channel request would not delay interfere other channel. The Content Distributor which is functioning as a distributor mechanism manager will arrange the allocation bandwidth. It will manage the change of the channel that is requesting by end user. If a specific channel is highly demand by user then the Content Distributor will allocate that particular channel with higher bandwidth and priority. In fact, the Queue Manager is the front end sensor for each decision made by the system in relation to channel statistics.

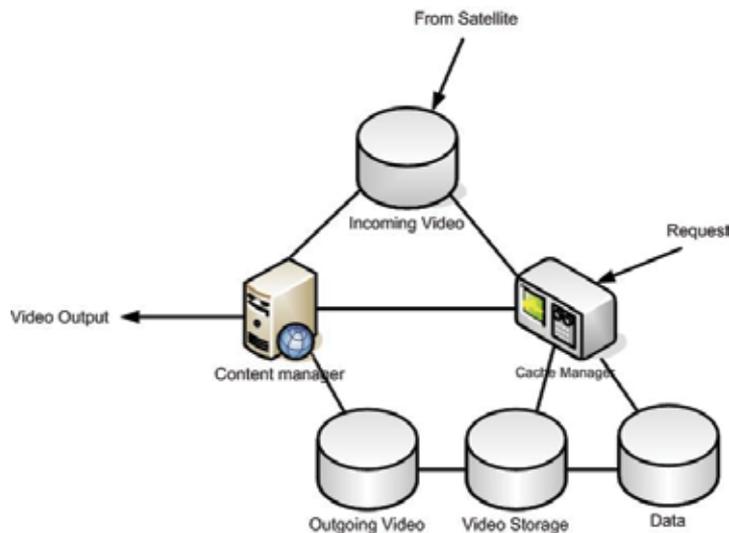


Fig. 14. Cache Receptor and Cache Array

In detail of Cache Array and Cache Receptor can be seen on figure 14. The Cache Array in this configuration is working as the Cache Manager, it manage the incoming request by updating the Content Manager to release every available channel by checking its availability to Data (Database Storage).

Incoming content stream from satellite is collected by a content storage. For short time, the content will be send to the user via the Content Manager if it is tagged live streaming by cache manager and will keep stream until the Cache Manager inform the Content Manager to stop. If a stop signal is given to the Content Manager by the Cache Manager, then that particular channel will be saved on video storage. On this state the Content Manager will retrieve all its content from storage array that represent by Outgoing Video, Video Storage and Data.

IPTV is delivering television stream via IP and each channel have to be served on real-time system in accordance to its critical mission. Channels were served by Content Manager which is not different to grid front manager function. Each channel is delivered on real-time

basis. To handle this condition the Content Manager can not working on a single queue mechanism. It must be managed into array system which will be effectively transfers channels on it services.

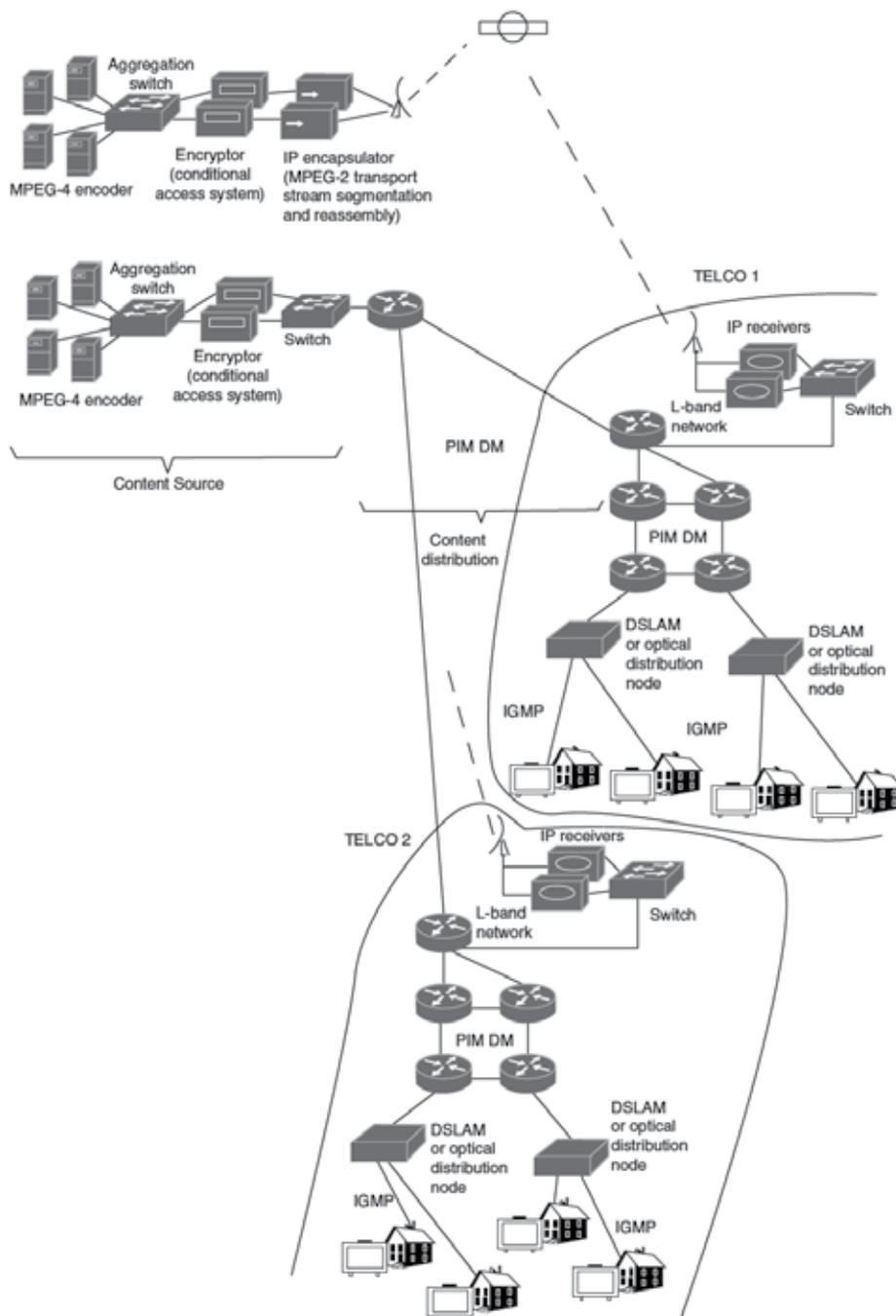


Fig. 15. IPTV Multicast Implementation(Minoli 2008)

The incoming content, content delivery and caching are working on a dense system. Implementation of CSM can be seen on figure 15. The CSM exact position is between the switch and first router on satellite content reception. Each user would be request via local network whether the contents are ready on the cache system. If it is not available then the request will be passed through the router. The router will contact the IPTV head operation channel system to send their specific request channel via satellite.

5. Conclusion

In this chapter, we already propose our design of Cache Stream Manager as a solution to IPTV distribution technology via one-way satellite communication. The main protocol distribution is DVB-S. The Cache management is the main agenda which was resonance by this chapter. This mechanism supported by grid technology would produce a great pleasure for IPTV operator. In particular, IPTV operators which would like to make wider of its services on low cost. The cost is always the bigger barrier for IPTV operator to spread their services to some remotes area.

One-way DVB-S is discussed as a protocol and mechanism to send IPTV services. Channel request and user interaction were manage through local internet. The user interaction and request is not necessary to use high speed internet, because the schedule is only xml data which is updated every time user receiving the video services. It was sending under the video signal. And user interaction will be sends to request control that manage every request if the demand is not exist on the cache array.

6. References

- Alberto, G., P. Francesco, et al. (2006). Simulating dynamic bandwidth allocation on satellite links. Proceeding from the 2006 workshop on ns-2: the IP network simulator. Pisa, Italy, ACM.
- Fatih, A., W. David, et al. (2001). "Adaptive rate control and QoS provisioning in direct broadcast satellite networks." *Wirel. Netw.* 7(3): 269-281.
- Frank, K., B. Torben, et al. (2005). A Synthesizable IP Core for DVB-S2 LDPC Code Decoding. Proceedings of the conference on Design, Automation and Test in Europe - Volume 3, IEEE Computer Society.
- Helmut, B., rklin, et al. (2007). "DVB: from broadcasting to ip delivery." *SIGCOMM Comput. Commun. Rev.* 37(1): 65-67.
- Hyoung Jin, Y., P. Jang Woong, et al. (2008). Data-aided algorithm based frequency synchronizer for DVB-S2. Proceedings of the 2nd international conference on Ubiquitous information management and communication. Suwon, Korea, ACM.
- Minoli, D. (2008). IP multicast with applications to IPTV and mobile DVB. Canada, John Wiley & Sons, Inc.
- Nihal, K. G. S. (1999). "Return link optimization for internet service provision using DVB-S networks." *SIGCOMM Comput. Commun. Rev.* 29(3): 4-13.
- O'Driscoll, G. (2008). NEXT GENERATION IPTV SERVICES AND TECHNOLOGIES. New Jersey, WILEY-INTERSCIENCE, JOHN WILEY & SONS, INC., PUBLICATION.
- Xu, N., S. Li, et al. (2005). The design and implementation of a DVB receiving chip with PCI interface. Proceedings of the 2005 conference on Asia South Pacific design automation. Shanghai, China, ACM.

The Deployment of Intelligent Transport Services by using DVB-Based Mobile Video Technologies

Vandenbergh, Leroux, De Turck, Moerman and Demeester
*Ghent University
Belgium*

1. Introduction

ITS systems combine (wired and wireless) communication systems, innovative applications, integrated electronics and numerous other technologies in a single platform. This platform enables a large number of applications with an important social relevance, both on the level of the environment, mobility and traffic safety. ITS systems make it possible to warn drivers in time to avoid collisions (e.g. when approaching the tail of a traffic jam or when a ghost driver is detected) and to inform them about hazardous road conditions. Navigation systems can take detailed real-time traffic info into account when calculating their routes. In case of an accident, the emergency services can be automatically informed about the nature and the exact location of the accident, saving very valuable time in the first golden hour. In case of traffic distortions, traffic can be immediately diverted. These are just a few of the many applications that are made possible because of ITS systems, but it is very obvious that these systems can make a significant positive contribution to traffic safety. In literature it is estimated that the decrease of accidents with injuries or fatalities will be between 20% and 50% (Bayle et al., 2007).

Attracted by the high potential of ITS systems, the academic world, the standardization bodies and the industry are all very actively involved in research and development of ITS solutions. The pillars of these systems are the communication facilities connecting the vehicles, the roadside infrastructure and the centralized safety and comfort services. Several wireless technologies can be considered when designing ITS architectures, and they can be divided into three categories: Dedicated Short Range Communication (DSRC) of which IEEE 802.11p WAVE, ISO CALM-M5 and the ISO CALM-IR standard are typical examples, wireless Wide Area Networks (WAN) such as GPRS, WiMAX and UMTS and finally digital broadcast technologies like RDS, DAB and the DVB specifications (DVB-T, DVB-S, DVB-H, etc.).

Since so many suitable technologies exist or are in development today, it is very hard to decide on which technologies future ITS architectures should be based. This problem is the starting point of several major ITS research projects, where much attention is given to solutions based on DSRC and wireless WAN networks. In the CVIS project, the implementation focuses on CALM-M5, CALM-IR, GPRS and UMTS technology (Eriksen et al., 2006). The Car2Car Communication Consortium aims to create and establish an open European industry standard for car2car communication systems based on the WAVE

standard (Baldessari et al., 2007). The COOPERS project evaluates the GPRS, CALM IR and DAB communication media (Frötscher, 2008). Although broadcast technologies are not neglected by the research community, it is harder to find examples focused on this category. As already mentioned, the COOPERS project has some attention for DAB, and in Korea a trial implementation of a TPEG based traffic information service system was deployed on their T-DMB network (Cho et al., 2006).

In this book chapter, we focus on the usage of DVB-H and DVB-SH for ITS systems. This approach is driven by the lower cost for the end user compared to wireless WAN solutions, by the lack of scalability issues and by the high provided bandwidth. Section 2 introduces the mobile broadcast technologies that are used in our architecture, and explains what the advantages are of using them in ITS systems. Section 3 describes how heterogeneous communication in mobile environments can be realized by means of the ISO TC204/WG16 CALM standard. This standard enables the seamless combination of DVB-H/SH with other wireless communication technologies such as IEEE 802.11p WAVE or an UMTS internet connection. In section 4 the functional description of our architecture is elaborated, the service architecture is described and a more in depth explanation of the implementation details is given. In section 5, the conclusions are drawn and section 6 finishes with the acknowledgment of the enablers of our research.

2. Mobile broadcast technologies

In this section we will elaborate on the broadcast specifications on which our architecture is based upon. We will first shortly introduce each specification and then explain why these technologies are used in our architecture and what the advantage is of using them instead of other communication standards.

2.1 Broadcasting to Handhelds (DVB-H)

DVB-H (Digital Video Broadcasting - Handheld) is a technical specification (ETSI, 2004) for bringing broadcast services to handheld receivers. It adapts the successful DVB-T (Terrestrial) system for digital terrestrial television to the specific requirements of handheld, battery-powered receivers. The conceptual structure of a DVB-H receiver is depicted in Fig. 1. It includes a DVB-H demodulator and a DVB-H terminal. The DVB-H demodulator includes a DVB-T demodulator, a time-slicing module and a MPE-FEC module.

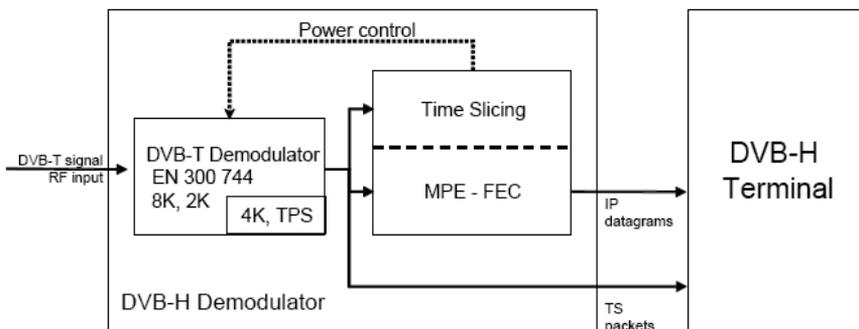


Fig. 1. Conceptual structure of a DVB-H receiver.

The DVB-T demodulator recovers the MPEG-2 Transport Stream packets from the received DVB-T RF signal. It offers three transmission modes 8K, 4K and 2K with the corresponding

Transmitter Parameter Signaling (TPS). Note that the 4K mode, the in-depth interleavers and the DVB-H signaling have been defined while elaborating the DVB-H standard. It aims to offer an additional trade-off between single frequency network (SFN) cell size and mobile reception performance, providing an additional degree of flexibility for network planning.

The time-slicing module, provided by DVB-H aims to save receiver power consumption while enabling to perform smooth and seamless frequency handover. Power savings of up to 90% are accomplished as DVB-H services are transmitted in bursts (using a high instantaneous bit rate), allowing the receiver to be switched off in inactive periods. The same inactive receiver can be used to monitor neighboring cells for seamless handovers. Time-slicing is mandatory for DVB-H.

The objective of Multi-Protocol Encapsulation - Forward Error Correction (MPE-FEC) is to improve the mobile channel tolerance to impulse interference and Doppler effect. This is accomplished through the introduction of an additional level of error correction at the MPE layer. By adding parity information calculated from the datagrams and sending this parity data in separate MPE-FEC sections, error-free datagrams can be output (after MPE-FEC decoding) even under bad reception conditions.

DVB-H is designed to be used as a bearer in conjunction with the set of DVB-IPDC (see Section 2.3) systems layer specifications. DVB-H has broad support across the industry. Currently, more than fifty DVB-H technical and commercial trials have taken place all over the world and further commercial launches are expected. In March 2008 the European Commission endorsed DVB-H as the recommended standard for mobile TV in Europe, instructing EU member states to encourage its implementation.

2.2 Satellite Services to Handhelds (DVB-SH)

DVB-H is primarily targeted for use in the UHF bands (but may also be used in the VHF- and L-band), currently occupied in most countries by analogue and digital terrestrial television services. DVB-SH (ETSI, 2007a; ETSI, 2008) seeks to exploit opportunities in the higher frequency S-band, where there is less congestion than in UHF. The key feature of DVB-SH is the fact that it is a hybrid satellite/terrestrial system that will allow the use of a satellite to achieve coverage of large regions or even a whole country. This is shown in Fig. 2. TR(a) are broadcast infrastructure transmitters which complement reception in areas where satellite reception is difficult, especially in urban areas; they may be collocated with mobile cell site or standalone. Local content insertion at that level is possible, relying on adequate radio frequency planning and/or waveform optimizations.

TR(b) are personal gap-fillers of limited coverage providing local on-frequency retransmission and/or frequency conversion; typical application is indoor enhancement under satellite coverage; no local content insertion is possible.

TR(c) are mobile broadcast infrastructure transmitters creating a "moving complementary infrastructure". Depending on waveform configuration and radio frequency planning, local content insertion may be possible.

DVB-SH's major enhancements when compared to its sister specification DVB-H are:

- The availability of more alternative coding rates.
- The inclusion of support for 1.7 MHz bandwidth.
- FEC using Turbo coding.
- Improved time interleaving.

As mentioned above, DVB-H systems have already been widely deployed, mostly on a trial basis so far. DVB-SH will be a complement to DVB-H and could potentially be used as such

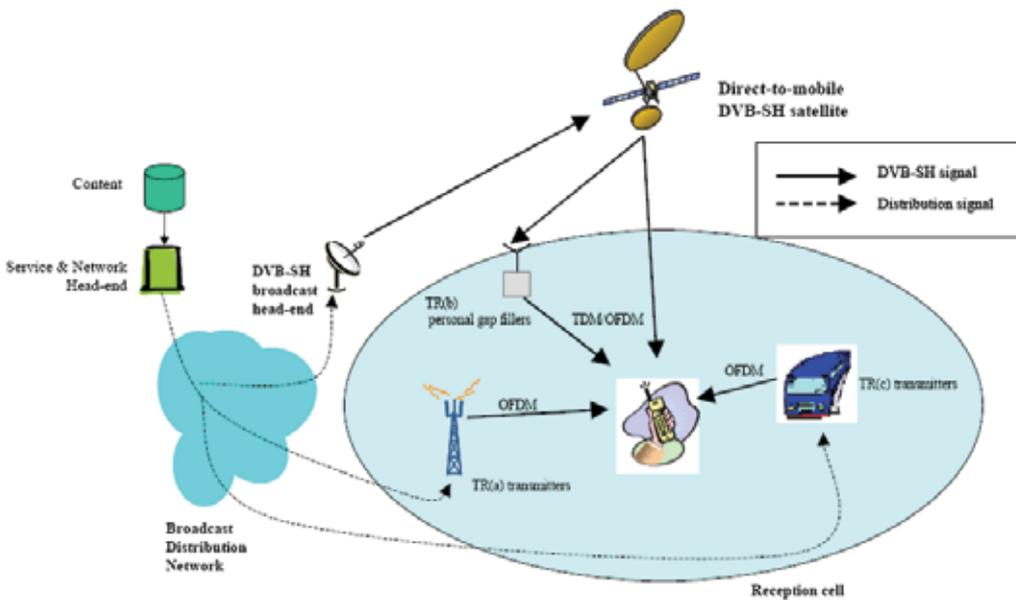


Fig. 2. Overall DVB-SH system architecture.

in a number of ways. Nationwide coverage could be achieved with the satellite footprint. Terminals that are in development will be dual mode, receiving DVB-SH in S-Band and DVB-H at UHF, and the over-lapping use of the DVB-IPDC specifications ensures that the two systems will be complementary.

2.3 Internet Protocol Datacast (DVB-IPDC)

Many commercial mobile TV networks are likely to be hybrid networks combining a uni-directional broadcast network, typically involving a wide transmission area and high data throughput, with a bi-directional mobile telecommunications network, involving much smaller transmission areas (cells). The set of DVB specifications for IP Datacasting (DVB-IPDC) (ETSI, 2007b) are the glue that bind these two networks together so that they can cooperate effectively in offering a seamless service to the consumer.

DVB-IPDC is originally designed for use with the DVB-H physical layer, but can ultimately be used as a higher layer for all DVB mobile TV systems. Currently, work is ongoing to make the necessary additions and adaptations to the DVB-IPDC specifications to allow interfacing with the DVB-SH standard. This work already resulted in the recent document "DVB Document A112-2r1: IP Datacast over DVB-SH". DVB-IPDC consists of a number of individual specifications that, taken together, form the overall system. The way the different elements fit together is defined in a reference architecture of the IPDC system whilst a further specification sets out the various use cases that are allowed for within the system.

The protocolstack of an IP Datacast over DVB-H system is shown in Fig. 3. The integration of an IP layer in the broadcaststack is one of the key concepts of an IPDC system. These IP datagrams are encapsulated inside the MPEG Transport Stream (TS) using MPE and MPE-FEC to improve mobile performance. For the delivery of streaming media, IP Datacast specifies the use of the Real-time Transport Protocol (RTP) and file delivery is performed by using File Delivery over unidirectional Transport (FLUTE) (IETF, 2004). FLUTE is a protocol

for unidirectional delivery of files over the Internet. In Section 3.3 we elaborate on the DVB-IPDC specifications as we point out how ITS services are incorporated into our architecture.

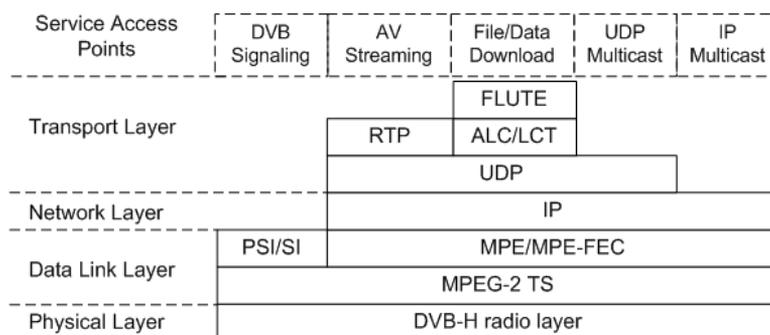


Fig. 3. The DVB-IPDC protocol stack.

2.4 Advantages of using DVB-H/SH as bearer technology

The goal of this section is to point out why DVB-H/SH technology is a very well suited candidate for the implementation of ITS systems. First, the advantages of using a digital broadcast technology are described. Second, we compare DVB-H and DVB-SH with other (mobile) broadcast technologies.

As already mentioned in the introduction, several wireless technologies can be considered when designing ITS architectures. There are roughly three categories of wireless systems that may be used: DSRC, wireless WAN and digital broadcast technologies. DSRC systems typically have a limited range of a few up to a few hundred meters. They were originally designed for direct link communications such as toll collect, but newer technologies support multi-hop communications. Examples are the IEEE 802.11p WAVE standard, or the ISO CALM-M5 standard. Wireless WAN technologies have a much larger range, and typically provide internet connectivity to mobile devices. Examples are GPRS, UMTS and WiMAX. Digital broadcast technologies can also cover large areas, but they do not offer two-way communications, only broadcast services. Examples technologies are RDS, DAB and the on DVB based technologies such as DVB-T or DVB-S.

When selecting a technology for the implementation of ITS systems, it is important to know that using broadcast technologies instead of wireless WAN solutions has some important advantages:

Scalability – Using a broadcast medium offers independence of the number of users that are connected to the system and thus the number of users that is able to receive ITS services. Antennas of non-broadcast systems could become overloaded when e.g. there is a traffic jam and all the car terminals would retrieve the same traffic info from the same antenna that covers the traffic jam's region.

Low cost - high user adoption – Recent large-scale motorist surveys have revealed that although users find ITS systems very useful, they are not very willing to pay for these services (RACC Automobile Club, 2007). This means that the cost of wireless WAN solutions could be a major stumbling block in the adoption of ITS systems by motorists. As broadcast media may provide free-to-air services, the cost to end users is kept much lower. When ITS systems use e.g. UMTS as the bearer technology then even if the service itself is free, the user (or the terminal manufacturer) still has to pay for the UMTS data connection.

Another cost-lowering property of broadcast technology is the fact that it enables travellers to enjoy ITS services abroad without having to pay expensive roaming fees.

Within the group of broadcast technologies, the usage of DVB-H and DVB-SH provides some additional advantages compared with its competitors:

Mobility – As DVB-H and DVB-SH are specifically developed for the delivery of data to handheld terminals, they provide a lot of error correction mechanisms for terminals that are moving at high speed. This is a major advantage over e.g. DVB-T.

High Bandwidth – As DVB-H and DVB-SH are initially developed for the delivery of mobile TV services, they provide a much bigger bandwidth (8 to 15 mbit per second for DVB-H) in comparison to other standards such as DAB (120 kbit per second) and RDS (1.1 kbit per second) (Chevul et al., 2005).

Industry adoption – As already mentioned in the previous sections, DVB-H has become the European standard for mobile television thus giving DVB-H a lead to other mobile television technologies such as T-DMB. This advantage is of course dependant on the region where the ITS service will be deployed but note that the deployment of DVB-H (and in the near future DVB-SH) is definitely not restricted to only Europe.

User adoption – When using DVB-H and DVB-SH technology for ITS services, user can also receive DVB-H/SH digital television broadcasts on their on-board equipment. This extra comfort service could be an important feature to attract new users, and could prove to be the catalyst that accelerates the adoption of on-board ITS equipment. The consumer interest in on-board television can already be observed today: portable DVD-players have become common consumer products, and some of the newest personal navigation devices can already receive DVB television broadcasts (e.g. Garmin nüvi).

Return channel integration – As DVB-IPDC is specifically designed for the convergence of DVB-H and DVB-SH with a bidirectional channel such as UMTS (through the integration of an IP layer), mobile terminals can still make use of a return channel. Since only the client uplink data will be transported over this return channel, the data cost will be much lower than when only using a unicast channel. This combination of technologies heavily relieves the bidirectional channel, having a positive influence on the scalability issues of wireless WAN solutions.

All the above makes it obvious that an ITS system based on IP Datacast over DVB-H/SH theoretically has many advantages, but as with all new technologies the question remains if the technology will be able to live up to the expectations in practice. Based on our experience within the IBBT MADUF project (MADUF, 2008), which was a trial DVB-H rollout in the city of Ghent, we are convinced that this will indeed be the case. In this trial we implemented our own middleware framework for the delivery of interactive services through DVB-H (Leroux et al., 2007). Measurements were also done concerning the performance of DVB-H for in-car usage (Plets et al., 2007). This trial made clear that DVB-IPDC is very suitable for the delivery of non-video data and that DVB-H has a good performance, even when using in a car at high speed.

3. Heterogeneous communication in mobile environments

In this section, we elaborate on the ISO TC204/WG16 CALM standard (Williams, 2004). This standard is the ISO approved framework for heterogeneous packet-switched communication in mobile environments, and supports user transparent continuous communications across various interfaces and communication media. It is, together with the DVB-H/SH broadcast technology, one of the key components of the ITS architecture presented in this book chapter.

The CALM architecture is depicted in Fig. 4. The two main elements are the CALM router, which provides the seamless connectivity, and the CALM host which runs the ITS applications with varying communication requirements. On both the CALM router and the host, different subcomponents can be distinguished:

- **CALM communication interface:** the CALM Communication Interface (CI) consists of a communication module and the necessary service access point for interfacing with the CALM networking layer (C-SAP)
- **CALM networking layer:** the CALM networking layer routes packets to the appropriate functional unit or program addressed. It also isolates the upper OSI layers from the different technologies that are making the connections. CALM supports multiple optional and complementary network protocols running independent of each other. Example protocols are standard IPv6 routing; CALM FAST, which is a non-IP protocol required for user applications with severe timing constraints and low-latency requirements (e.g. time-critical safety related applications); and geographic-aware routing, with or without map information. The CALM networking layer also provides a service access point for interaction with the CALM User Services / Applications (the T-SAP)
- **CALM management:** The CALM communications and station management comprises global functionality, and functionality grouped into three groups: Interface Management Entity (IME), Network Management Entity (NME) and CALM Management Entity (CME). Disregard of this grouping, the CALM management is one entity, and there are no observable or testable interfaces between IME, NME and CME. The role of the IME is to directly control the communication interfaces, and to allow access to a communication interface for the purpose of receiving and transmitting management data packets. The role of the NME is to directly control the network and transport layer protocols. The CME provides the decision-making process to the CALM mechanism. The CME collects the specification of the communication parameters enabled by each of the desired communications mechanisms and the requirements from each of the applications from the initialization process. It monitors the availability of lower level communications mechanisms and issues. Based on this information and on policies a decision on how to route data packets is made.
- **CALM service layer:** The CALM service layer shall provide an application programmer interface (API) to user applications, and it shall provide an A-SAP to the CME. Using the API, applications can easily define how their data should be exchanged with other CALM nodes (local broadcast, n-hop broadcast, directional communication, unicast to known address, ...), the level of importance of the data (for QoS classification), the delay constraints, etc.
- **CALM applications:** three kinds of applications can run on the CALM host: CALM FAST applications, CALM IP-based applications, and non-CALM aware applications. The first category has the ability to control the interaction with the CALM environment. Such applications can respond to CALM management entity requests for registration information or are able to request registration upon initialization. They get real-time access to pre-selected parameters of specific CALM communication interfaces in line with applicable regulations, and to the CALM networking layer in order to control the real-time behaviour of the communication link. This control functionality includes e.g. power settings, channel settings, beam pointing. These applications typically use the CALM FAST or Geo-routing networking protocols. CALM IP-based applications are

similar to CALM FAST applications, but they typically have less stringent timing constraints, and are more session oriented. Therefore they generally use the IPv6 networking protocol. Non-CALM aware applications operate with the assumption of the programmer that a normal UDP or TCP connection is being established for communication. Such applications operate without the ability to control any interaction with the CALM environment. The CALM management entity must hide all CALM environment peculiarities from these applications.

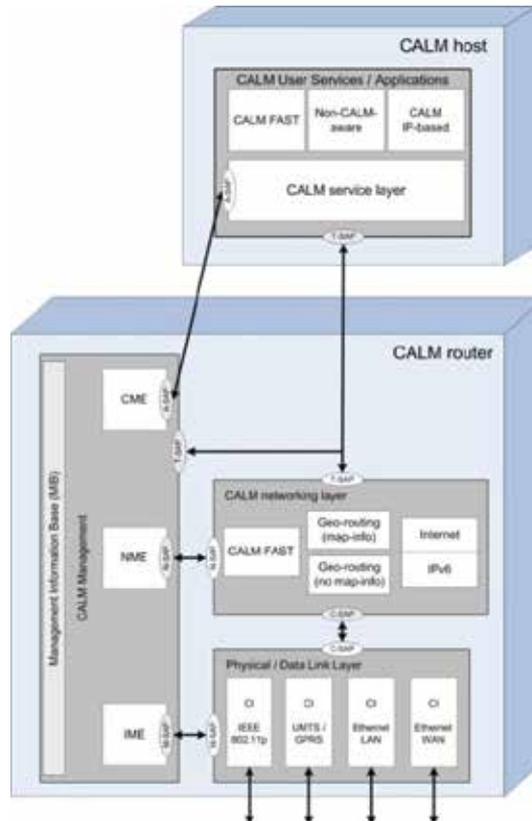


Fig. 4. The CALM architecture

4. Architecture

In this section, we present our ITS architecture based on IP Datacast over DVB-H/SH. First, the functional description of the architecture and its global communication aspects are described. Then the service architecture is detailed, and finally some implementation details of the architecture will be given.

4.1 Functional description

The core idea of the communication architecture (Fig. 5) is to use DVB technology to broadcast data to the vehicles. When DVB-H coverage is already available, this infrastructure can be reused, minimizing necessary investments. If this is not the case, rural areas can be covered by a single DVB-SH satellite, and DVB-SH repeater antennas can be

installed in urban areas to guarantee coverage. Communication between vehicles is provided by the IEEE 802.11p WAVE technology. Since the combination of DVB and WAVE technologies provides both communication from the central infrastructure to the vehicles, and between vehicles, most ITS applications are supported by this base architecture (collision avoidance, hazardous road warnings, traffic situation-aware navigation, etc.). If the user requires interactive applications (notification of emergency services, sending info to the traffic control centre, etc.) the base architecture can be expanded with a return channel, e.g. by using existing UMTS infrastructure.

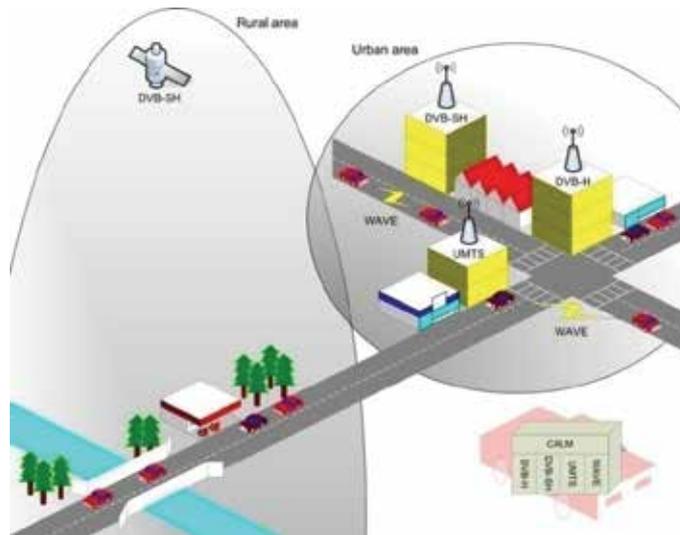


Fig. 5. Communication architecture

In this communication architecture, it is necessary to equip every vehicle with at least a WAVE module and a DVB-H or DVB-SH receiver. Optionally, an UMTS module can also be installed. To coordinate the collaboration of these different wireless interfaces, we rely on the ISO TC204/WG16 CALM standard (Williams, 2004) described in section 3. This standard is the ISO approved framework for heterogeneous packet-switched communication in mobile environments, and supports user transparent continuous communications across various interfaces and communication media. It already supports WAVE and UMTS technology, and due to its modular design, it can easily be expanded to support DVB-S/H media. The CALM standard is, together with the WAVE standard, one of the most important upcoming communication standards within the ITS domain. Since both technologies are incorporated in the architecture, it is compliant with current and future ITS trends and activities.

Since the proposed communication architecture is based on the usage of DVB-H/SH, it inherits the advantages of this technology (see section 2.4): the cost to end users is kept low, scalability issues do not arise, and users can enjoy extra comfort services such as mobile digital television.

4.2 Service architecture

The service architecture of the proposed ITS solution is depicted in Fig. 7. An important concept within this service architecture is the use of the Transport Protocol Expert Group

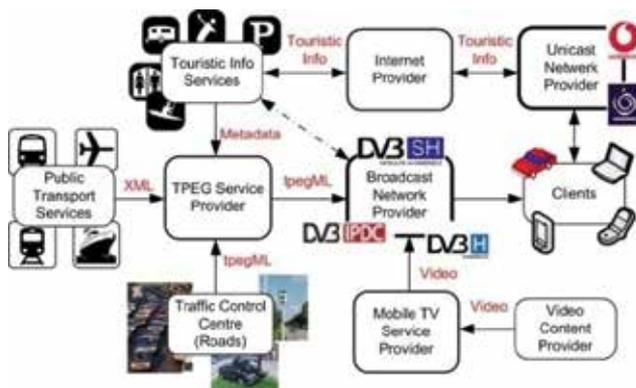


Fig. 7. Service architecture

providing these messages to the end user. Therefore, it will deliver the generated tpegML messages to the broadcast network provider. From the end users perspective, the TPEG service provider is the contact point for ITS information.

The Video Content Providers produces digital television programs and delivers them to the Mobile TV Service Provider. The Mobile TV Service Provider collects content from different Video Content Providers, and is responsible for providing this content to the clients. From the end user point of view, the Mobile TV Service Provider is the contact point for mobile digital television.

The Broadcast Network Provider is responsible for the DVB-H or DVB-SH network. It broadcasts digital television and ITS information to the end devices. It receives this content from the TPEG service provider and the Mobile TV Service Provider.

The clients are all the devices that can receive DVB-S/H broadcasts. The most obvious example is the in-vehicle infotainment system, but this can also be a PDA, a personal navigation device or a mobile phone.

The Unicast Network Provider is responsible for the wireless network that provides the (optional) two-way communication necessary for interactive applications. The Internet Provider is responsible for the internet access for clients connected to the unicast network. In most cases, the Internet Provider and the Unicast Network Provider will be the same company, both from a logical and service provider point of view they are two separate entities.

4.3 Implementation details

This subsection provides a more in depth explanation of how the required services can be provided over DVB-H/SH. Attention is given to the question how TPEG services can be integrated into the IPDC headend, and how the tpegML files can be delivered through FLUTE.

4.3.1 Integrating TPEG services in the IPDC headend

Fig. 8 details how the ITS services are integrated into a typical DVB-IPDC headend. In our setup, the Video Content Provider streams its multimedia data over SDI (Serial Digital Interface) to the encoders. These H.264 encoders also encapsulate all the data into RTP packets. Following the protocol stack as defined by DVB-IPDC, these RTP-streams are sent over UDP and then multicasted into the Multicast IP Network. At the TPEG Service Provider, a file server sends all the tpegML data to the Flute Server of the headend.

Both the encoders (through Session Description Protocol (SDP) (IETF, 1998) files) and the TPEG Service Provider (not standardized) sent metadata to the ESG server in order to make

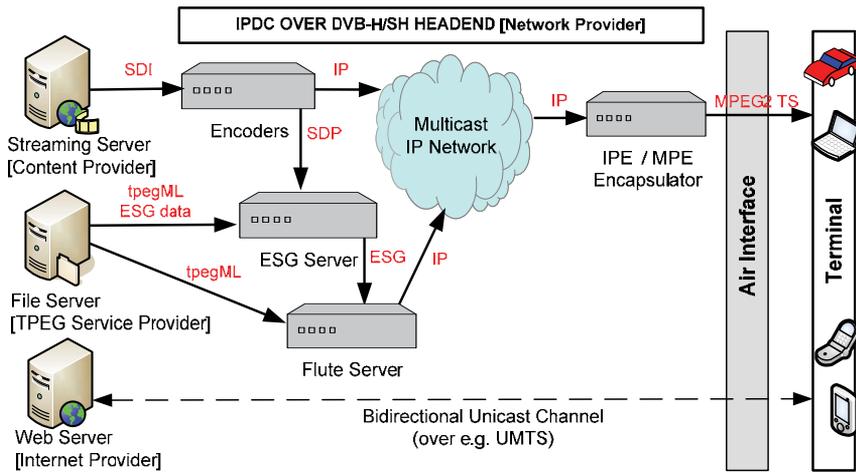


Fig. 8. Coupling of the TPEG Service Provider with the IPDC over DVB-H/SH headend

sure that the TPEG middleware and the decoders are able to find and correctly interpret all the broadcasted data. All the ESG data is then sent to the Flute server which will multicast all the data into the Multicast IP Network.

The IPE/MPE Encapsulator encapsulates all the multcasted IP packets into an MPEG2 TS (transport stream). This MPEG2 TS is then sent to the satellite or antennas, where the stream is finally modulated and sent over the air to the user's terminal. As a return channel, a bidirectional unicast channel such as e.g. UMTS may be used to acquire more information (such as local touristic information) through the Internet Provider. Note that for single frequency networks (SFN) an additional component, a so-called SFN adapter, should be placed after the IP/MPE encapsulators.

4.3.2 Delivery of tpegML through FLUTE

DVB-IPDC not only defines protocols for the delivery of audio and video streams, but it also specifies how binary files should be incorporated into the MPEG data streams and sent to the end users. This IPDC file delivery mechanism is based on the FLUTE protocol (IETF, 2004). FLUTE (File delivery over Unidirectional Transport) is a fully-specified protocol to transport files (any kind of discrete binary object), and uses special purpose objects – the File Delivery Table (FDT) Instances – to provide a running index of files and their essential reception parameters in-band of a FLUTE session. FLUTE is built on top of the Asynchronous Layered Coding (ALC) protocol instantiation (IETF, 2002) which provides reliable asynchronous delivery of content to an unlimited number of concurrent receivers from a single sender. FLUTE is carried over UDP/IP, and is independent of the IP version and the underlying link layers used.

There are 5 types of file delivery sessions that are specified on the basis of FLUTE. We will only detail the most advanced session type, more specifically the Dynamic file delivery carousel as this is the required method in our architecture for the delivery of ITS related data to all the users. A dynamic file delivery carousel is a possibly time-unbounded file delivery session in which a changing set of possibly changed, added or deleted files is delivered. The use of a carousel mechanism (of which teletext is a typical example) is necessary as in a broadcast scenario you don't know exactly when a user tunes in. As a carousel mechanism

continuously repeats or updates the traffic info, users who just started their car will still be able to receive all relevant traffic info, even if they missed the initial message.

The time that a random user has to wait for its traffic info will be dependent on the size of the carousel. As we want to support an unlimited number of TPEG services in our architecture, encompassing all these services into the same data carousel would invoke a round trip time of the data carousel that is much too high. Secondly, the use of one big carousel would ensue that the antenna has to be turned on for longer periods which in turn partly undoes one of the main advantages of DVB-H/SH, namely the reduced power consumption. Therefore we use one main FLUTE data carousel which continuously repeats all road related traffic info. After each such road related traffic block, exactly one other object is placed. This second object will be one of the public transport services or the touristic metadata. As already explained in section III.B, the touristic metadata only informs the user's terminal where to find specific information, related to the current location of the terminal.

Our FLUTE data carousel is illustrated in Fig. 9. Object 1 always contains the road-related data. As shown in Fig. 9, object 1 is continuously repeated while changes in this object (object 1') and the fact that the second object is continuously alternating are indicated by the File Delivery Table (object 0).

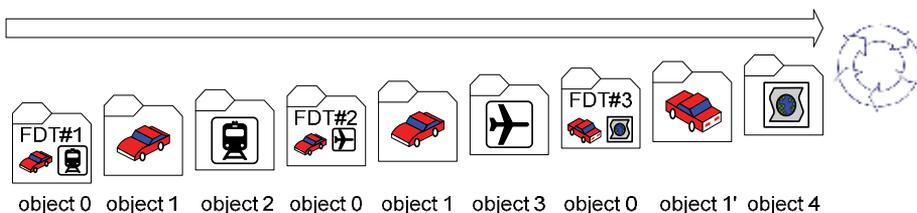


Fig. 9. Delivering ITS services via a dynamic FLUTE carousel

Note that DVB-IPDC only specifies that one FLUTE channel should be supported but that the use of several concurrent FLUTE channels may be supported by the terminals. For terminals that do not support multiple channels, it should be possible for them to receive enough data from the first channel named base FLUTE channel in order to declare the channel as complete. In our architecture the base FLUTE channel contains all the information that comes from the Traffic Control Centre and the changes in Bus, Rail and Air operator services. As such, all DVB-H terminals shall automatically be able to receive all the traffic related info. If the manufacturer finds it relevant to also support other services than the manufacturer may still incorporate the support of multiple FLUTE sessions into its devices. Terminals that do not support multiple channels shall ignore all but the base FLUTE channel.

5. Conclusion

In this paper we presented an ITS architecture that was based on the usage of the mobile broadcast technologies DVB-H and DVB-SH. It was explained why these technologies are very well suited for the delivery of ITS services and what the advantages are of using these technologies. The proposed architecture is complimentary with the available set of DVB-IPDC specifications and details were provided of how exactly ITS services should be integrated into a DVB-IPDC system.

6. Acknowledgment

The authors would like to thank the Flemish Interdisciplinary institute for Broadband technology (IBBT) for defining the MADUF and NextGenITS projects.

7. References

- Baldessari, R.; Bödekker, B.; Brakemeier, A. et al (2007). *CAR 2 CAR Communication Consortium Manifesto*, deliverable of C2C-CC project
- Bayly, M.; Fildes, B.; Regan, M. & Young K. (2007). Review of crash effectiveness of Intelligent Transport Systems, *Deliverable D4.1.1-D6.2*, TRACE project
- BBC (2009). *BBC Travel News - TPEG*, accessed online at <http://www.bbc.co.uk/travelnews/xml/> on July 13th 2009
- Chevul; Karlsson; Isaksson et al (2005). Measurements of Application-Perceived Throughput in DAB, GPRS, UMTS and WLAN Environments, *Proceedings of RVK'05*, Linköping, Sweden
- Cho, S.; Geon, K.; Jeong, Y.; Ah,n C.; Lee, S.I. & Lee, H. (2006). Real Time Traffic Information Service Using Terrestrial Digital Multimedia Broadcasting System. *IEEE transactions on broadcasting*, vol. 52, no 4, 2006
- Eriksen, A.; Olsen, E.; Evensen K. et al (2006). Reference Architecture, *Deliverable D.CVIS.3.1*, CVIS project
- European Broadcasting Union (2009). *TPEG- What is it all about?*, accessed online at <http://tech.ebu.ch/docs/other/TPEG-what-is-it.pdf> on June 25th 2009
- ETSI (2004). *EN 302304 v1.1.1:Digital Video Broadcasting (DVB), Transmission System for Handheld Terminals (DVB-H)*
- ETSI (2007a). *TS 102 585 V1.1.1:System Specifications for Satellite services to Handheld devices (SH) below 3 GHz - TS 102 585*
- ETSI(2007b). *TS 102 468 V1.1.1: Digital Video Broadcasting (DVB); IP Datacast over DVB-H: Set of Specifications for Phase 1*
- ETSI (2008). *EN 302 583 V1.1.0:Framing Structure, channel coding and modulation for Satellite Services to Handheld devices (SH) below 3 GHz*
- DVB Document A112-2r1. Digital Video Broadcasting (DVB); IP Datacast: Electronic Service Guide (ESG) Implementation Guidelines Part 2: IP Datacast over DVB-SH.
- Frötscher, A. (2008). *Co-operative Systems for Intelligent Road Safety*, presentation of COOPERS project, available on http://www.coopers-op.eu/uploads/media/COOPERS_Presentation_TRA_Ljubljana_2008.pdf
- IETF (1998). *RFC 2327: SDP, Session Description Protocol*
- IETF (2002). *RFC 3450: Asynchronous Layered Coding (ALC) Protocol Instantiation*
- IETF *RFC 3926: FLUTE- File Delivery over Unidirectional Transport*, 2004.
- Leroux, P.; Verstraete, V.; De Turck, F. & Demeester, P. (2007). Synchronized Interactive Services for Mobile Devices over IPDC/DVB-H and UMTS, *Proceedings of IEEE Broadband Convergence Networks*, Munchen
- MADUF, <https://projects.ibbt.be/maduf>, 2008.
- Plets, D.; Joseph, W.; Martens, L.; Deventer, E. & Gauderis, H. (2007). Evaluation and validation of the performance of a DVB-H network, *2007 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, Orlando, Florida, USA
- RACC Automobile Club (2007). Stakeholder utility, data privacy and usability analysis and recommendations for operational guarantees and system safeguards: Europe, *Deliverable D.DEPN.4.1*, CVIS project
- Williams, C.C. B. (2004). *The CALM handbook*, available on <http://www.tc204wg16.de/Public/The%20CALM%20Handbookv2-060215.pdf>

A New Waveform based on Linear Precoded Multicarrier Modulation for Future Digital Video Broadcasting Systems

Oudomsack Pierre Pasquero, Matthieu Crussière, Youssef Nasser,
Eddy Cholet and Jean-François Hélar
European University of Brittany (UEB)
Institute of Electronics and Telecommunications of Rennes (IETR)
INSA de Rennes, 20, avenue des Buttes de Coesmes, 35043 Rennes,
France

1. Introduction

Orthogonal Frequency Division Multiplexing (OFDM) has been perceived as one of the most effective transmission schemes for multipath propagation channels. It has been widely adopted in most of the digital video broadcasting (DVB) standards such as DVB-T in Europe, DMB-T in China, FLO in North America, ISDB-T in Japan. The major reason for this success lies in the capability of OFDM to split the single channel into multiple parallel intersymbol interference (ISI) free subchannels. It is easily carried out by implementing Inverse Fast Fourier Transform (IFFT) at the transmitter and Fast Fourier Transform (FFT) at the receiver [1]. Therefore, the distortion associated to each subchannel, also called subcarrier, can be easily compensated for by one coefficient equalization. For that purpose, the receiver needs to estimate the channel frequency response (CFR) for each subcarrier. In the DVB-T standard [2], one subcarrier over twelve is used as pilot for CFR estimation as illustrated in Fig. 1, i.e. symbols known by the receiver are transmitted on these subcarriers. Thus, the receiver is able to estimate the CFR on these pilot subcarriers and to obtain the CFR for any subcarrier using interpolating filtering techniques [3].

Nevertheless, OFDM systems are very sensitive to synchronization error such as carrier frequency offset (CFO) or sampling frequency offset (SFO) [4]. Indeed, when the carrier frequency or the sampling frequency of the transmitter and the receiver are not equal, the orthogonality between the different subcarriers is lost which can lead to strong intercarrier interference (ICI) effects [4]. This is why in addition to the scattered pilot subcarriers used for CFR estimation, continuous pilot subcarriers have been defined in the DVB-T standard [2] to estimate the CFO and the SFO [5]. Fig. 1 depicts the locations of the data subcarriers and the pilot subcarriers over the time and frequency grid as defined in the DVB-T standard. The originality of this work is to reduce the overhead part resulting from pilot insertion by using a joint CFR, CFO and SFO estimation approach based on a linear precoding function. Eventually, these pilots dramatically reduce the spectral efficiency and the useful bit rate of the system. The basic idea consists in using a two-dimensional (2D) linear

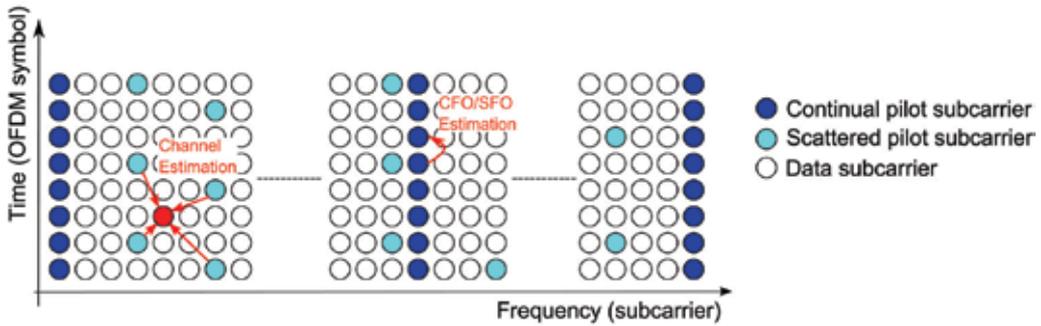


Fig. 1. Data and Pilot subcarriers locations over the time and frequency grid in the DVB-T standard

precoding matrix before the OFDM modulation, and to dedicate one of the precoding sequences to transmit a so-called *spread pilot* information [6]. It will be showed that these spread pilots can provide a diversity gain for some estimators. Moreover, the 2D linear precoding function improves the flexibility of the system compared to the DVB-T standard. This chapter is organized as follows. In section 2, we describe the proposed transmitter scheme based on spread pilots. The transmitted signal and the pilot symbols insertion technique are studied. In section 3, the proposed channel estimation principle based on spread pilot is described in perfect synchronization case. The analytical expression of the mean square error (MSE) of the estimator is derived. Some simulation results in term of bit error rate (BER) of the proposed CFR estimation are given and compared with those of the DVB-T standard. The proposed synchronization algorithms are presented in section 4. Two different stages of CFO and SFO estimations are considered. The first one is dedicated to the fine synchronization. The second one is used to estimate the residual CFO and SFO. Some simulation results in term of BER and mean square error (MSE) of the estimators are given and analyzed. Finally, we conclude this chapter by a general comparison between the proposed system based on spread pilots and the DVB-T standard in term of performance, complexity and flexibility.

2. Transmitter scheme

The transmitter structure on which spread pilot principles are based is depicted in Fig. 2. We consider an OFDM communication system using N subcarriers, N_u of which being active, with a guard interval size of v samples. In the first step, data bits are encoded, interleaved and converted to complex symbols $x_{t,s}[i]$, assumed to have zero mean and unit variance. These data symbols are then interleaved before being linearly precoded (LP) by a sequence c_i of L chips, with $0 \leq i \leq L = 2^n$ and $n \in \mathbb{N}$. The sequences used for the precoding function are the well-known Walsh-Hadamard (WH) codes [7] [8]. They have been chosen for their orthogonality property. The chips obtained are mapped over a subset of $L = L_t \times L_f$ subcarriers, with L_t and L_f defined as the time and frequency spreading factors respectively. The first L_t chips are allocated in the time direction. The next blocks of L_t chips are allocated identically on the adjacent subcarriers as illustrated in Fig. 3. Therefore, the 2D chip mapping follows a zigzag in time. Let us note that the way of applying the 2D chip mapping does not change significantly the system performance [9].

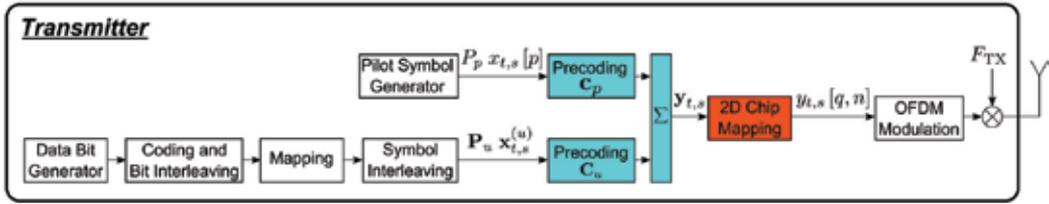


Fig. 2. 2D LP OFDM transmitter based on spread pilots

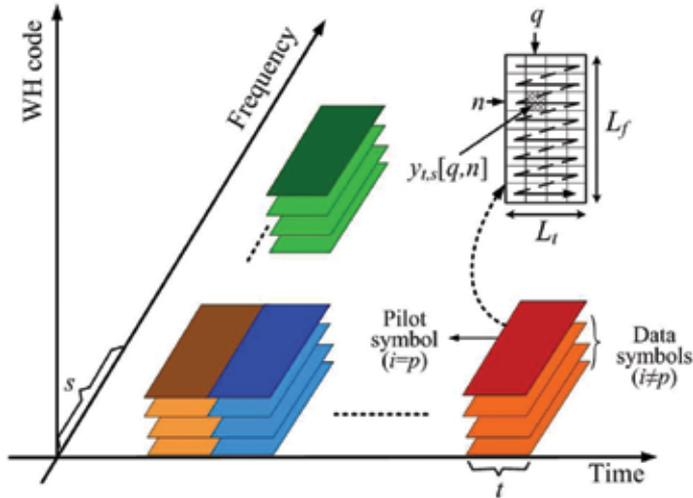


Fig. 3. 2D Spreading Scheme

Inspired by pilot embedded techniques [10], spreading the pilot symbols consists in transmitting low level pilot-sequences concurrently with the data. In order to reduce the cross-interferences between pilots and data, the idea is to select a pilot sequence which is orthogonal with the data sequences. This is obtained by allocating one of the WH orthogonal sequences c_p to the pilots symbols $x_{t,s} [p]$ on every subset of subcarriers. Contrary to the DVB-T system where the pilot symbols are transmitted by only a few subcarriers, in the proposed system each active subcarrier conveys a part of the spread pilot information. Consequently, numerous observation samples are available and the estimators can benefit from the frequency diversity of the channel over the whole bandwidth. Nevertheless, since each pilot symbol is superimposed to $(L - 1)$ data symbols, as illustrated in Fig. 3, a term of interference appears if the orthogonality is lost.

To derive the appropriate estimation algorithms in the sequel, we need to formalize the transmitted signal expression. Therefore, we define a frame as a set of L_t adjacent OFDM symbols, and a sub-band as a set of L_f adjacent subcarriers. In order to distinguish between the different subsets of subcarriers, let us denote t and s the indexes referring to the frame and the sub-band respectively, with $0 \leq s \leq S - 1$. Given these notations, we can express the signal transmitted on a subset of subcarriers (t, s) :

$$\mathbf{y}_{t,s} = \mathbf{C} \mathbf{P} \mathbf{x}_{t,s} \quad (1)$$

where $\mathbf{x}_{t,s} = [x_{t,s}[0] \dots x_{t,s}[i] \dots x_{t,s}[L-1]]^T$ is the complex symbol vector, $\mathbf{P} = \text{diag}\{\sqrt{P_0} \dots \sqrt{P_i} \dots \sqrt{P_{L-1}}\}$ is a diagonal matrix which elements are amplitude weighting factors associated to symbols, and $\mathbf{C} = [\mathbf{c}_0 \dots \mathbf{c}_i \dots \mathbf{c}_{L-1}]$ is the WH precoding matrix which i th column corresponds to i th precoding sequence $\mathbf{c}_i = [c_i[0, 0] \dots c_i[q, n] \dots c_i[L_t - 1, L_f - 1]]^T$. We assume normalized precoding sequences, i.e. $c_i[q, n] = \pm \frac{1}{\sqrt{L}}$. Note that power factor P_p can advantageously be used as a boost factor for pilot symbols in order to help CFR estimation and synchronization procedures. Since the applied 2D chip mapping follows a zigzag in time, $c_i[q, n]$ is the $(n \times L_t + q)$ th chip of the i th precoding sequence \mathbf{c}_i . Hence, the signal transmitted on the q th OFDM symbol and the n th subcarrier of the subset of subcarriers (t, s) writes:

$$y_{t,s}[q, n] = \sum_{i=0}^{L-1} \sqrt{P_i} x_{t,s}[i] c_i[n \times L_t + q] \quad (2)$$

3. Channel estimation

A. Principles

Let us define $\mathbf{H}_{t,s} = \text{diag}\{h_{t,s}[0, 0] \dots h_{t,s}[q, n] \dots h_{t,s}[L_t - 1, L_f - 1]\}$ as the $[L \times L]$ diagonal matrix of the channel frequency coefficients associated to a given subset of subcarriers (t, s) . In perfect synchronization case and considering that the guard interval can absorb all the interference due to previous symbols, after OFDM demodulation and 2D chip de-mapping the received signal simply writes:

$$\mathbf{z}_{t,s} = \mathbf{H}_{t,s} \mathbf{y}_{t,s} + \mathbf{w}_{t,s} \quad (3)$$

where $\mathbf{w}_{t,s} = [w_{t,s}[0, 0] \dots w_{t,s}[q, n] \dots w_{t,s}[L_t - 1, L_f - 1]]^T$ is the additive white Gaussian noise (AWGN) vector having zero mean and variance $\sigma_w^2 = E\{|w_{t,s}[q, n]|^2\}$.

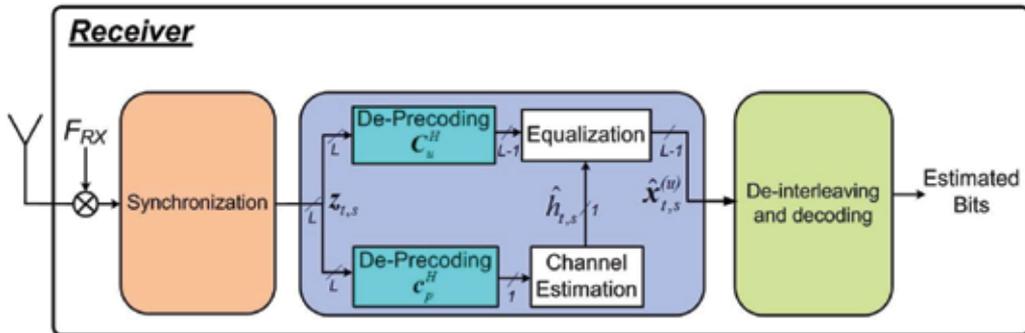


Fig. 4. Channel estimation scheme based on spread pilot

Fig. 4 depicts the receiver structure based on spread pilot CFR estimation. The basic idea of the proposed spread pilot CFR estimation algorithm is to estimate one average channel frequency coefficient $\hat{h}_{t,s}$ by subset of subcarriers (t, s) . It is obtained by deprecoding the received signal $\mathbf{z}_{t,s}$ by the pilot sequence \mathbf{c}_p^H and then dividing by the pilot symbol known by the receiver:

$$\begin{aligned}\hat{h}_{t,s} &= \frac{1}{\sqrt{P_p} x_{t,s}[p]} \mathbf{c}_p^H \mathbf{z}_{t,s} \\ &= \frac{1}{\sqrt{P_p} x_{t,s}[p]} \mathbf{c}_p^H (\mathbf{H}_{t,s} \mathbf{C} \mathbf{P} \mathbf{x}_{t,s} + \mathbf{w}_{t,s})\end{aligned}\quad (4)$$

Let us denote $\mathbf{C}_u = [\mathbf{c}_0 \dots \mathbf{c}_{i \neq p} \dots \mathbf{c}_{L-1}]$ the $[L \times (L-1)]$ useful data precoding matrix, $\mathbf{P}_u = \text{diag}\{\sqrt{P_0} \dots \sqrt{P_{i \neq p}} \dots \sqrt{P_{L-1}}\}$ the $[(L-1) \times (L-1)]$ diagonal matrix which entries are the amplitudes assigned to the data symbols, and $\mathbf{x}_{t,s}^{(u)} = [x_{t,s}[0] \dots x_{t,s}[i \neq p] \dots x_{t,s}[L-1]]^T$ the $[(L-1) \times 1]$ data symbols vector. Given these notations, (4) can be rewritten as:

$$\begin{aligned}\hat{h}_{t,s} &= \frac{1}{\sqrt{P_p} x_{t,s}[p]} \left(\mathbf{c}_p^H \mathbf{H}_{t,s} \mathbf{c}_p \sqrt{P_p} x_{t,s}[p] + \mathbf{c}_p^H \mathbf{H}_{t,s} \mathbf{C}_u \mathbf{P}_u \mathbf{x}_{t,s}^{(u)} + \mathbf{c}_p^H \mathbf{w}_{t,s} \right) \\ &= \frac{1}{L} \text{tr}\{\mathbf{H}_{t,s}\} + \frac{1}{\sqrt{P_p} x_{t,s}[p]} \left(\mathbf{c}_p^H \mathbf{H}_{t,s} \mathbf{C}_u \mathbf{P}_u \mathbf{x}_{t,s}^{(u)} \right) + \frac{1}{\sqrt{P_p} x_{t,s}[p]} \left(\mathbf{c}_p^H \mathbf{w}_{t,s} \right) \\ &= \bar{h}_{t,s} + \Xi_{\text{MCI}}(t, s) + \Xi_{\text{WGN}}(t, s)\end{aligned}\quad (5)$$

The first term $\bar{h}_{t,s}$ actually corresponds to the average CFR globally experienced by the subset of subcarriers (t, s) . The second term represents the multiple code interference (MCI). It results from the loss of orthogonality between the precoding sequences caused by the variance of the channel frequency coefficients over the subset of subcarriers (t, s) . One can actually check that if the channel coefficients are the same over the subset of subcarriers (t, s) which implies $\mathbf{H}_{t,s} = \bar{h}_{t,s} \mathbf{I}$, with \mathbf{I} the $[L \times L]$ identity matrix, the MCI term is null because of the orthogonality between the sequences \mathbf{c}_i . The last term is the noise sample obtained after despreading. Finally, the estimated channel frequency coefficient (5) is used to equalize the $(L-1)$ data symbols spread over the same subset of subcarriers.

B. Estimator analysis

In order to analyze the theoretical performance of the proposed estimator, we propose to derive its MSE expression under the assumption of a wide-sense stationary uncorrelated scattering (WSSUS) channel [11].

$$\begin{aligned}\text{MSE}\{\hat{h}_{t,s}\} &= \text{E}\left\{\left|\hat{h}_{t,s} - \bar{h}_{t,s}\right|^2\right\} \\ &= \text{E}\left\{|\Xi_{\text{MCI}}(t, s)|^2\right\} + \text{E}\left\{|\Xi_{\text{WGN}}(t, s)|^2\right\}\end{aligned}\quad (6)$$

First, let us compute the MCI variance:

$$\text{E}\left\{|\Xi_{\text{MCI}}(t, s)|^2\right\} = \frac{1}{P_p} \text{E}\left\{\mathbf{c}_p^H \mathbf{H}_{t,s} \mathbf{C}_u \mathbf{P}'_u \mathbf{C}_u^H \mathbf{H}_{t,s}^H \mathbf{c}_p\right\}\quad (7)$$

where $\mathbf{P}'_u = \mathbf{P}_u \mathbf{P}_u^H = \text{diag}\{P_0 \dots P_{i \neq p} \dots P_{L-1}\}$. Actually, (7) cannot be analyzed practically due to its complexity. Applying some properties of random matrix and free probability theories [12] [13] which are stated in Appendix, a new MCI variance formula can be derived:

$$\begin{aligned}
\mathbb{E} \left\{ |\Xi_{\text{MCI}}(t, s)|^2 \right\} &= \frac{1}{P_p} \mathbb{E} \left\{ \mathbf{c}_p^H \mathbf{H}_{t,s} (\mathbf{I} - \mathbf{c}_p \mathbf{c}_p^H) \mathbf{H}_{t,s}^H \mathbf{c}_p \right\} \\
&= \frac{1}{P_p} \mathbb{E} \left\{ \mathbf{c}_p^H \mathbf{H}_{t,s} \mathbf{H}_{t,s}^H \mathbf{c}_p - \mathbf{c}_p^H \mathbf{H}_{t,s} \mathbf{c}_p \mathbf{c}_p^H \mathbf{H}_{t,s}^H \mathbf{c}_p \right\} \\
&= \frac{1}{P_p} \mathbb{E} \left\{ \underbrace{\frac{1}{L} \text{tr} \{ \mathbf{H}_{t,s} \mathbf{H}_{t,s}^H \}}_A - \frac{1}{L^2} \text{tr} \{ \mathbf{H}_{t,s} \} \text{tr} \{ \mathbf{H}_{t,s}^H \}}_B \right\}
\end{aligned} \tag{8}$$

The expectation of A is the average power of the channel frequency coefficients on the subset of subcarriers (t, s) . Assuming that the channel frequency coefficients are normalized, its value is one:

$$\begin{aligned}
\mathbb{E} \left\{ \frac{1}{L} \text{tr} \{ \mathbf{H}_{t,s} \mathbf{H}_{t,s}^H \} \right\} &= \frac{1}{L} \sum_{q=0}^{L_t-1} \sum_{n=0}^{L_f-1} \mathbb{E} \left\{ |h_{t,s}[q, n]|^2 \right\} \\
&= 1
\end{aligned} \tag{9}$$

The expectation of B is a function of the autocorrelation of the channel $R_{HH}(\Delta q, \Delta n)$ which expression (43) is developed in Appendix. Indeed, it can be written:

$$\mathbb{E} \left\{ |\Xi_{\text{MCI}}|^2 \right\} = \frac{1}{P_p} \left(1 - \frac{1}{L^2} \sum_{q=0}^{L_t-1} \sum_{n=0}^{L_f-1} \sum_{q'=0}^{L_t-1} \sum_{n'=0}^{L_f-1} R_{HH}(\Delta q, \Delta n) \right) \tag{10}$$

Now, let us compute the noise variance:

$$\begin{aligned}
\mathbb{E} \left\{ |\Xi_{\text{WGN}}|^2 \right\} &= \frac{1}{P_p} \mathbb{E} \left\{ \mathbf{c}_p^H \mathbf{w}_{t,s} \mathbf{w}_{t,s}^H \mathbf{c}_p \right\} \\
&= \frac{1}{P_p} \frac{1}{L} \mathbb{E} \left\{ \text{tr} \{ \mathbf{w}_{t,s} \mathbf{w}_{t,s}^H \} \right\} \\
&= \frac{1}{P_p} \sigma_w^2
\end{aligned} \tag{11}$$

Finally, by combining the expressions of the MCI variance (10) and the noise variance (11), the MSE (6) writes:

$$\text{MSE} \left\{ \hat{h}_{t,s} \right\} = \frac{1}{P_p} \left(1 - \frac{1}{L^2} \sum_{q=0}^{L_t-1} \sum_{n=0}^{L_f-1} \sum_{q'=0}^{L_t-1} \sum_{n'=0}^{L_f-1} R_{HH}(\Delta q, \Delta n) + \sigma_w^2 \right) \tag{12}$$

The analytical expression of the MSE depends on the pilot power P_p , also called boost factor, the autocorrelation function of the channel $R_{HH}(\Delta q, \Delta n)$ and the noise variance σ_w^2 . The autocorrelation of the channel (43) is a function of both the coherence bandwidth and the coherence time. We can then expect that the proposed estimator will be all the more efficient than the channel coefficients will be highly correlated within each subset of subcarriers. One

can actually check that if the channel is flat over a subset of subcarriers, then the MCI (10) is null. Therefore, it is important to optimize the time and frequency spreading lengths L_t and L_f , according to the transmission scenario. It is clear from (12) that the greater the boost factor P_p , the better the CFR estimator performance. On the other hand, the greater the boost factor P_p , the lower the data symbol power and the harder the data symbol detection. Therefore, the boost factor P_p has to be optimized in term of BER.

C. Simulation results

In this section, we analyse the performance of the proposed 2D LP OFDM system based on spread pilot CFR estimation compared to the DVB-T system with perfect CFR estimation. Table I gives the simulation parameters. The time-invariant channel models used are the F1 and P1 models detailed in [2]. They are specified for fixed outdoor rooftop antenna reception. The F1 channel corresponds to a line-of-sight (LOS) transmission, contrary to the P1 channel model which corresponds to a non-LOS transmission. The COST207 Typical Urban 6 paths (TU6) channel model is used as mobile channel. We define parameter β as the relative Doppler frequency equal to the product between the maximum Doppler frequency and the total OFDM symbol duration T_{OFDM} .

| | |
|-------------------------------------|---|
| Bandwidth | 8 MHz |
| FFT size (N_{FFT}) | 2048 samples |
| Guard Interval size | 512 samples (64 μ s) |
| OFDM symbol duration (T_{OFDM}) | 280 μ s |
| Carrier frequency | 500 MHz |
| Data symbol constellations | QPSK - 16QAM - 64QAM |
| Pilot symbol constellation | BPSK |
| Convolutional code rate | $R_c = 1/2$ using (133, 171) _o |
| Time-invariant channel models | F1 and P1 |
| Mobile channel model | TU6 - 20km/h |
| Relative Doppler frequency | $\beta=0.003$ |

Table I Simulation parameters

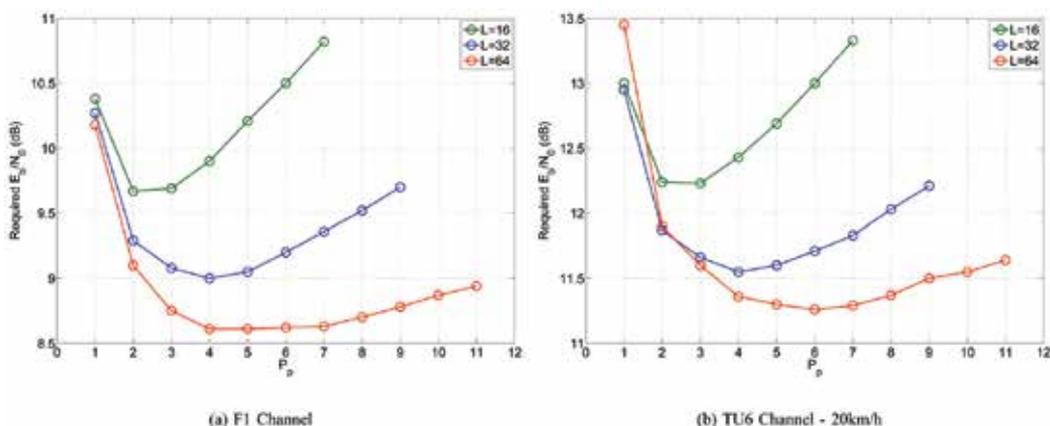


Fig. 5. Required $\frac{E_b}{N_0}$ to obtain a BER = 10^{-4} versus pilot power P_p - Spread Pilot Channel Estimation - 16QAM - $R_c = 1/2$

As it has been previously mentioned, it is important to optimize the boost factor P_p in function of the total spreading length L . For that purpose, Fig. 5(a) and Fig. 5(b) give the required $\frac{E_b}{N_0}$ (energy per bit to noise power spectral density ratio) to obtain a BER equal to 10^{-4} at the output of the Viterbi decoder for 16QAM data symbols under the F1 and the TU6 channel models respectively. It is noticeable that the boost factor values for which the required $\frac{E_b}{N_0}$ values reach a minimum, are similar for the time-invariant (F1) and the mobile (TU6) channels. Therefore, it is not necessary to adapt the boost factor in function of the channel characteristics. Moreover, for a given total spreading length, several values of P_p give similar performance in term of BER. Among these boost factor values, we will choose the largest one in order to obtain better estimators. In the following, the boost factor P_p will be equal to 3, 5 and 7 for a total spreading length of 16, 32 and 64 respectively.

The boost factor values optimized, we can analyze the CFR estimator performance. Fig. 6 depicts the estimator performance in term of MSE for QPSK data symbols, different mobile speeds and different spreading factor values L . The curves represent the MSE obtained with the analytical expression (12), and the markers those obtained by simulation. We note that the MSE measured by simulation are really close to those predicted with the MSE formula. This validates the analytical development made in the previous section. We note that beyond a given $\frac{E_b}{N_0}$, the MSE reaches a floor which is easily interpreted as being due to the MCI (12).

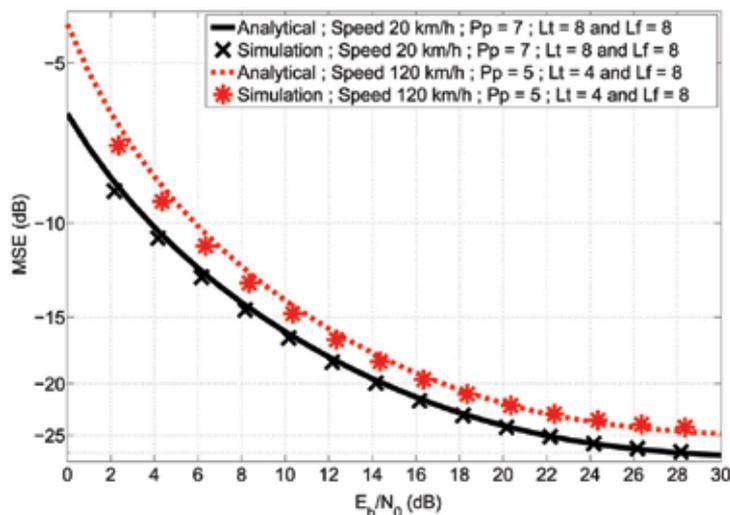


Fig. 6. MSE performance obtained with the analytical expression and by simulation - QPSK data symbols- Speeds: 20km/h and 120km/h - $\beta = 0.003$ and 0.018

Fig.7(a) and Fig.7(b) give the BER measured for 16QAM and 64QAM data symbols under the time-invariant channel models F1 and P1 respectively for both the DVB-T system and the proposed system. To quantify the loss due to CFR estimation error resulting from the MCI and the AWGN, we give the performance of the proposed system with perfect CFR estimation by subcarrier and perfect CFR estimation by subset of L subcarriers. According to (12), the BER degradation from perfect CFR estimation by subcarrier to perfect CFR estimation by subset of subcarriers is due to the MCI, and the degradation from perfect CFR estimation by subset of subcarriers to spread pilot CFR estimation is due to the AWGN. In

Fig. 7(a), the curves of the proposed system with perfect CFR estimation by subcarrier and those with perfect CFR estimation by subset of subcarriers overlap. Since the frequency spreading length is equal to 4, it means that the F1 channel model is very flat over four subcarriers. On the other hand, under P1 channel model, there is a degradation. This is explained by the P1 non-LOS characteristic which generates a higher selectivity in the frequency domain. Let us note for a BER equal to 10^{-4} , from perfect CFR estimation by subcarrier to spread pilot channel estimation, there is a loss of less than 1dB and 1.5dB under F1 and P1 channel models respectively. Since the F1 and P1 channel models do not vary in time, it is interesting to spread the symbols as much as possible in the time direction. For a time spreading length L_t larger or equal to 4 (which implies $L \geq 16 > 12$), a gain in term of spectral efficiency and useful bit rate is obtained compared to the DVB-T system. In most of the cases, for $L_t = 16$, the performance of the proposed system based on spread pilot CFR estimation slightly outperforms that of the DVB-T system with perfect CFR estimation.

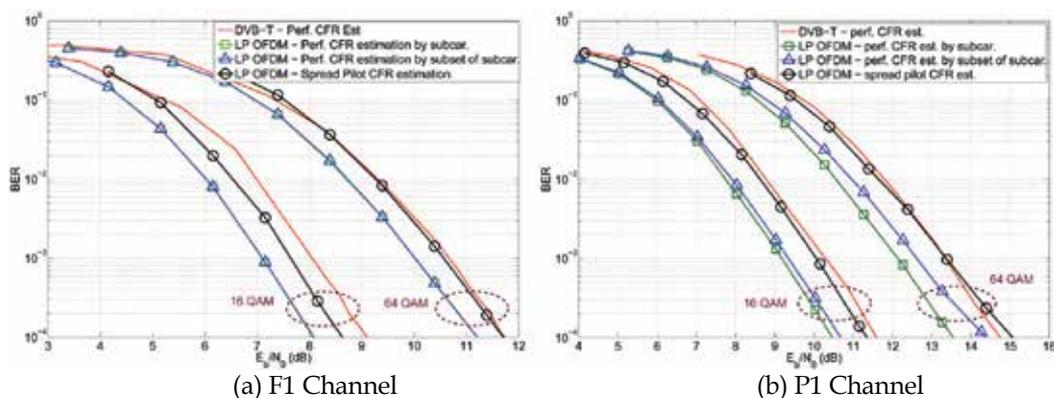


Fig. 7. Performance comparison between the DVB-T system with perfect CFR estimation and the proposed 2D LP OFDM under time-invariant channel models - $R_c = \frac{1}{2}$ - $L_f = 4$ and $L_t = 16$ - $P_p = 7$

To verify if the MCI caused by the channel time variation does not degrade to much the proposed system performance, Fig.8 gives the BER measured under the TU6 channel model with a mobile speed of 20 km/h. In this case, since the channel varies both in frequency and time domains, it is necessary to optimize the time and frequency spreading lengths L_t and L_f for a given total spreading length. Fig. 8(a) gives the BER of the proposed system based on spread pilot CFR estimation for a given $\frac{E_b}{N_0}$ with a total spreading length equal to 64 for QPSK and 16QAM data symbols. The values of L_t and L_f giving the lowest BER are 16 and 4 respectively. Using these spreading length values, we compare the proposed system to the DVB-T system with perfect CFR estimation in Fig. 8(b). Likewise with time-invariant channels, the performances of the DVB-T system with perfect CFR estimation and the proposed system with spread pilot CFR estimation are similar. Furthermore, for a BER equal to 10^{-4} , the loss from perfect CFR estimation by subcarrier to spread pilot CFR estimation is further less than 1 dB. Since the time spreading length L_t is equal to 16, it proves that our proposed CFR estimation is not sensible to low mobility scenarios in term of BER. Obviously, for high velocities, degradations of the proposed system performance would be notable. To resolve this weakness, it is possible to extend the system to the space dimension

using a space code block code (SCBC). Indeed, it is established in [14] that a SCBC system based on spread pilot CFR estimation is very robust to high mobility scenarios.

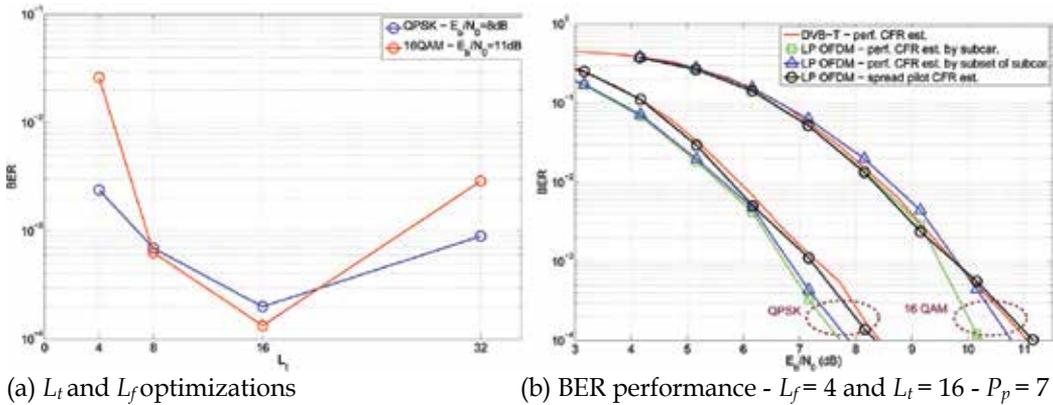


Fig. 8. Performance comparison between the DVB-T system with perfect channel estimation and the proposed 2D LP OFDM under the TU6 mobile channel model - 20 km/h - $\beta = 0.003$ - $L = 64$

4. Synchronization principles

In the 2K mode (corresponding to a 2048 FFT size) of the DVB-T standard, 45 continual pilot subcarriers are reserved to help the receiver to estimate the CFO and the SFO. This additional pilot information once again reduces the spectral efficiency and the useful bit rate of the system. In our system based on spread pilots, we propose to exploit the pilot symbols already used for CFR estimation to estimate the CFO and the SFO. Thus, we avoid a new reduction in the spectral efficiency.

Two stages of CFO and SFO estimation are proposed. The first one dedicated to fine CFO estimation is processed before despreading the pilot symbols. The estimated CFO is then used to synchronize the received signal in the time domain. Nevertheless, as detailed hereafter, residual CFO will still be present after this initial synchronization step. The aim of the second synchronization stage is indeed to estimate and compensate for the residual CFO. To further exploit pilot information, it is possible during this stage to also estimate SFO, thus obtaining a joint residual CFO and SFO estimation.

In the following, we define the CFO ζ and the SFO ξ as:

$$\zeta = (F_{TX} - F_{RX}) T_s \quad (13)$$

$$\xi = \frac{\Delta T}{T_s} \quad (14)$$

with F_{TX} and F_{RX} the carrier frequency of the transmitter and the receiver respectively, T_s the sampling period at the transmitter and $(T_s + \Delta T)$ at the receiver.

A. Fine CFO estimation

Since each active subcarrier conveys a part of the spread pilot symbol information, a gain in diversity can be obtained compared to the DVB-T system which uses only a few pilot subcarriers for synchronization issues. Therefore, we propose to estimate the CFO before

despreading the pilot symbols. Neglecting the SFO effects which are much less significant than CFO [5], the received symbol in the presence of CFO during the q th OFDM symbol on the n 'th subcarrier of the subset of subcarriers (t, s') corresponding to a component of vector $z_{t,s'}$ in equation (3) expresses as follows:

$$\begin{aligned} z_{t,s'}[q, n'] &= e^{j2\pi q(N+\nu)\zeta} \sum_{s=0}^{S-1} \sum_{n=0}^{L_f-1} y_{t,s}[q, n] h_{t,s}[q, n] \varphi(s', s, n', n) + w_{t,s'}[q, n'] \\ &= e^{j2\pi q(N+\nu)\zeta} y_{t,s'}[q, n'] h_{t,s'}[q, n'] \varphi(s', s', n', n') + \Xi_{\text{ICI}}(t, s', q, n') + w_{t,s'}[q, n'] \end{aligned} \quad (15)$$

where $\varphi(s', s, n', n)$ is an equivalent transfer function describing the attenuation and the phase rotation caused by the CFO. It is equal to:

$$\varphi(s', s, n', n) = \psi_N \left(\zeta + \frac{(s' - s)L_f + (n' - n)}{N} \right) \exp \left\{ j\pi(N-1) \left(\zeta + \frac{(s' - s)L_f + (n' - n)}{N} \right) \right\} \quad (16)$$

where $\psi_N(x)$ is the Dirichlet function defined by: $\psi_N(x) = \frac{\sin(\pi Nx)}{N \sin(\pi x)}$. The second term $\Xi_{\text{ICI}}(t, s', q, n')$ in (15) is the ICI coming from the other active subcarriers in the same OFDM symbol. It writes:

$$\Xi_{\text{ICI}}(t, s', q, n') = e^{j2\pi q(N+\nu)\zeta} \left(\underbrace{\sum_{\substack{n=0 \\ n \neq n'}}^{L_f-1} y_{t,s'}[q, n] h_{t,s'}[q, n] \varphi(s', s', n', n)}_{\text{ICI from the interested sub-band } s'} + \underbrace{\sum_{\substack{s=0 \\ s \neq s'}}^{S-1} \sum_{n=0}^{L_f-1} y_{t,s}[q, n] h_{t,s}[q, n] \varphi(s', s, n', n)}_{\text{ICI from the other sub-bands } s \neq s'} \right) \quad (17)$$

The phase rotation at the left-hand side of (15) is due to the CFO increment in time. It is clear that this phase rotation increases with the OFDM symbol index q . Thus, we will benefit from this increment to define the CFO estimation metric $\Gamma_t(\zeta)$ computed at the t -th frame from the pilot sequence by:

$$\begin{aligned} \Gamma_t(\zeta) &= \sum_{q=0}^{L_t-2} \sum_{s'=0}^{S-1} \sum_{n'=0}^{L_f-1} c_p[n'L_t + q] z_{t,s'}^*[q, n'] c_p^*[n'L_t + q + 1] z_{t,s'}[q + 1, n'] \\ &= \sum_{q=0}^{L_t-2} \sum_{s'=0}^{S-1} \sum_{n'=0}^{L_f-1} c_p[n'L_t + q] \left\{ e^{-j2\pi q(N+\nu)\zeta} y_{t,s'}^*[q, n'] h_{t,s'}^*[q, n'] \varphi^*(s', s', n', n') + \Xi_{\text{ICI}}^*(t, s', q, n') + w_{t,s'}^*[q, n'] \right\} \\ &\quad \times c_p^*[n'L_t + q + 1] \left\{ e^{j2\pi(q+1)(N+\nu)\zeta} y_{t,s'}[q + 1, n'] h_{t,s'}[q + 1, n'] \varphi(s', s', n', n') + \Xi_{\text{ICI}}(t, s', q + 1, n') + w_{t,s'}[q + 1, n'] \right\} \end{aligned} \quad (18)$$

Assuming the channel does not vary over two consecutive OFDM symbols, i. e. $h_{t,s'}[(q, q + 1), n'] = h_{t,s'}[q, n'] = h_{t,s'}[q + 1, n']$ and using (2), the estimation metric finally writes:

$$\begin{aligned} \Gamma_t(\zeta) &= \frac{P_p}{L} e^{j2\pi(N+\nu)\zeta} \sum_{s'=0}^{S-1} |x_{t,s'}[p]|^2 \sum_{q=0}^{L_t-2} \sum_{n'=0}^{L_f-1} |h_{t,s'}[(q, q + 1), n']|^2 + \Xi(t, s', q, n') \\ &= \frac{P_p}{L} S N_u (L_t - 1) e^{j2\pi(N+\nu)\zeta} + \sum_{s'=0}^{S-1} \sum_{n'=0}^{L_f-1} \sum_{q=0}^{L_t-2} \Xi(t, s', q, n') \end{aligned} \quad (19)$$

where $\Xi(t, s', q, n')$ results from the contributions of the interferences caused by the data chips superimposed to the pilot chips, the ICI due to CFO and the AWGN. In our study, we assumed that these interferences have Gaussian distribution with zero mean [15] [16]. Consequently, if the product $S \times L_f \times (L_t - 1)$ is large enough:

$$\sum_{s'=0}^{S-1} \sum_{n'=0}^{L_f-1} \sum_{q=0}^{L_t-2} \Xi(t, s', q, n') \rightarrow 0 \quad (20)$$

Using (18) and (20), it is straightforward to say that the CFO is the measure of the phase of $\Gamma_t(\zeta)$:

$$\hat{\zeta}_t^{(\text{fine})} = \frac{1}{j2\pi(N+\nu)} \arg \{ \Gamma_t(\zeta) \} \quad (21)$$

It should be noted that to avoid any phase ambiguity, it is necessary that:

$$|2\pi(N+\nu)\zeta| < \pi \quad (22)$$

This constraint determinates the maximum estimable CFO value at this stage. For instance, in the 2K mode, with a relative guard interval size equal of 1/4, the maximum estimable CFO is 40% of the intercarrier spacing $\frac{1}{NT_s}$. Let us remind that the widely used guard interval based coarse carrier frequency synchronization [17] [18] brings down the CFO to such values, which makes the proposed algorithm compatible with classical OFDM reception schemes.

As depicted in Fig. 9, the estimated CFO $\hat{\zeta}_t^{(\text{fine})}$ value is used to correct the signal in the time domain in order to mitigate the ICI. To evaluate the performance of the proposed fine CFO estimation, we give in Fig. 10 the residual CFO after this synchronization step in open loop (in the case when no CFO loop filter is used). In other words, we give the average error of the instantaneous estimated CFO $\hat{\zeta}_t^{(\text{fine})}$. According to BER simulations obtained for CFR estimation, the frequency spreading length L_f is set to 4. However, we reduce the time spreading length L_t to 8. The CFO value is set to 10% of the intercarrier spacing. It can be seen that the residual CFO falls down less than 2% whatever the SNR or the channel condition. On the other hand, we notice that under the TU6 channel model, the residual CFO values are very similar for any mobility scenario. From (18) and (19), we know that the phase rotation measurement used to estimate the CFO is carried out between two consecutive OFDM symbols under a constraint of flatness of the channel in the time domain over two OFDM symbols. This constraint is reasonable even in high mobility scenarios, which explains why the proposed fine CFO estimation method is not sensitive to velocity variations. Moreover, we remark that beyond a given $\frac{E_b}{N_0}$ value, residual CFO curves reach a plateau. Since the CFO estimation is processed before the deprecoding function, this error floor is due to the data chips interference. This motivates for the use of a second estimation step processed after deprecoding function.

B. Joint residual CFO and SFO estimation

In order to mitigate the data chips interference, we propose to add a second stage of CFO and SFO estimations after the deprecoding function. Thus, if the spreading lengths L_t and L_f

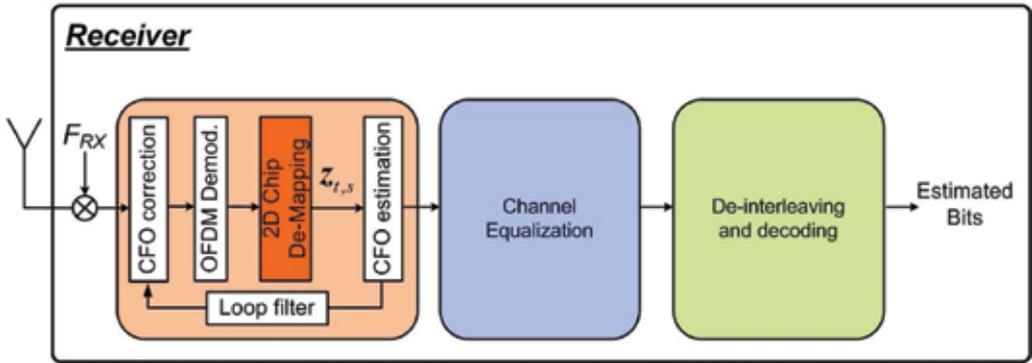


Fig. 9. Fine CFO estimation scheme before despreading function

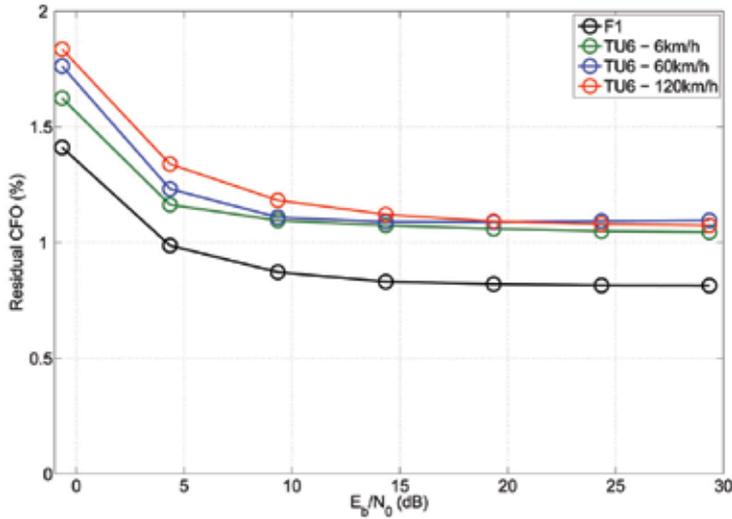


Fig. 10. Residual CFO after the 1st synchronization stage in open loop - SFO null - $L_f = 4$ and $L_t = 8$ - 16QAM data symbols

have been optimized such as the channel is flat enough over each subset of subcarriers, the interference from the data symbols should be strongly attenuated after the deprecoding function.

First of all, let us derive the expression of the received symbol on the n 'th subcarrier during the q th OFDM symbol of the subset of subcarriers (t, s) after the first stage of synchronization, before the deprecoding function:

$$z_{t,s'}[q, n'] = \sum_{s=0}^{S-1} \sum_{n=0}^{L_f-1} \exp \left\{ j2\pi q (N + v) \left(\zeta^{(res)} + \frac{sL_f + n}{N} \xi \right) \right\} y_{t,s}[q, n] h_{t,s}[q, n] \varphi(q, s', s, n', n) + w_{t,s'}[q, n'] \quad (23)$$

with $\varphi(q, s', s, n', n) = \psi_N \left(\zeta^{(res)} + \frac{sL_f + n}{N} \xi + \frac{(s'-s)L_f + (n'-n)}{N} \right) \exp \left\{ j\pi (N-1) \left(\zeta^{(res)} + \frac{sL_f + n}{N} \xi + \frac{(s'-s)L_f + (n'-n)}{N} \right) \right\}$. To simplify equation (23), we define $\phi_{t,s'}[q, n']$

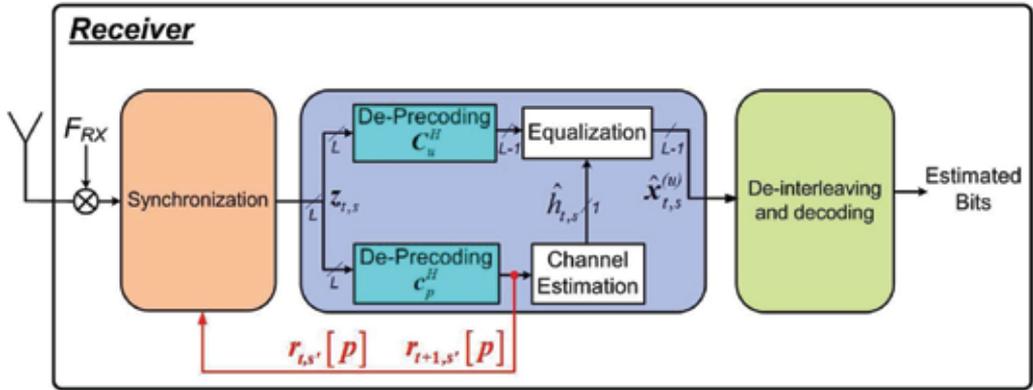


Fig. 11. Joint residual CFO and SFO estimation scheme after the deprecoding function as the function describing the total phase rotation and $\psi_{t,s'}[q,n']$ as the function describing the amplitude attenuation caused by CFO and SFO on the interested subcarrier.

$$\phi_{t,s'}[q,n'] = \exp \left\{ j2\pi q (N + \nu) \left(\zeta^{(\text{res})} + \frac{sL_f + n}{N} \xi \right) \right\} \exp \left\{ j\pi (N - 1) \left(\zeta^{(\text{res})} + \frac{sL_f + n}{N} \xi \right) \right\} \quad (24)$$

$$\psi_{t,s'}[q,n'] = \psi_N \left(\zeta^{(\text{res})} + \frac{sL_f + n}{N} \xi \right) \quad (25)$$

Hence, the expression of the received symbol on the interested subcarrier (23) can be simplified and re-expressed as:

$$z_{t,s'}[q,n'] = y_{t,s'}[q,n'] h_{t,s'}[q,n'] \phi_{t,s'}[q,n'] \psi_{t,s'}[q,n'] + \Xi_{\text{ICI}}(t,s',q,n') + w_{t,s'}[q,n'] \quad (26)$$

Now, let us develop the expression of the received pilot symbol on the subset of subcarriers (t,s') after deprecoding by pilot sequence \mathbf{c}_p^H :

$$r_{t,s'}[p] = \mathbf{c}_p^H [\Phi_{t,s'} \Psi_{t,s'} \mathbf{H}_{t,s'} \mathbf{C} \mathbf{P} \mathbf{x}_{t,s'} + \Xi_{\text{ICI}}(t,s') + \mathbf{w}_{t,s'}] \quad (27)$$

where $\Phi_{t,s'} = \text{diag} \{ \phi_{t,s'}[0,0] \dots \phi_{t,s'}[q,n'] \dots \phi_{t,s'}[L_t - 1, L_f - 1] \}$ and $\Psi_{t,s'} = \text{diag} \{ \psi_{t,s'}[0,0] \dots \psi_{t,s'}[q,n'] \dots \psi_{t,s'}[L_t - 1, L_f - 1] \}$ are the $[L \times L]$ diagonal matrices which components are the phase rotations and the attenuation factors respectively, due to CFO and SFO on the L received symbols over the subset of subcarriers (t,s') . To simplify equation (27), we define the equivalent channel matrix as $\mathbf{H}_{t,s'}^{(\text{eq})} = \Phi_{t,s'} \Psi_{t,s'} \mathbf{H}_{t,s'}$. Hence, the received pilot symbol can be rewritten as:

$$r_{t,s'}[p] = \mathbf{c}_p^H \left[\mathbf{H}_{t,s'}^{(\text{eq})} \mathbf{C} \mathbf{P} \mathbf{x}_{t,s'} + \Xi_{\text{ICI}}(t,s') + \mathbf{w}_{t,s'} \right] \quad (28)$$

Finally, using (4) and (5), the received pilot symbol over the subset of subcarriers (t,s') writes:

$$\begin{aligned} r_{t,s'}[p] &= \bar{h}_{t,s'}^{(\text{eq})} \sqrt{P_p} x_{t,s'}[p] + \Xi'_{\text{ICI}}(t,s') + \Xi_{\text{MCI}}(t,s') + \Xi_{\text{WGN}}(t,s') \\ &= \bar{\phi}_{t,s'} \bar{\psi}_{t,s'} \bar{h}_{t,s'} \sqrt{P_p} x_{t,s'}[p] + \Xi(t,s') \end{aligned} \quad (29)$$

where $\bar{\phi}_{t,s'}$ and $\bar{\psi}_{t,s'}$ are the average phase rotation and the average attenuation factor due to CFO and SFO over the subset of subcarriers (t, s') .

To jointly estimate the residual CFO and the SFO, we propose to measure the phase rotation between two consecutive frames t and $(t + 1)$. First of all, let us derive the expressions of the phase rotations $\bar{\phi}_{t,s'}$ and $\bar{\phi}_{t+1,s'}$ which are the average phase rotations associated to sub-band s' during frames t and $(t + 1)$ respectively:

$$\begin{aligned}\bar{\phi}_{t,s'} &= \frac{1}{L} \sum_{q=0}^{L_t-1} \sum_{n'=0}^{L_f-1} \phi_{t,s'}[q, n'] \\ &= \pi(N-1) \left[\zeta^{(\text{res})} + \frac{(2s'+1)L_f-1}{2N} \xi \right] + \pi(L_t-1)(N+\nu) \left[\zeta^{(\text{res})} + \frac{(s'+1)L_f-1}{N} \xi \right]\end{aligned}\quad (30)$$

$$\bar{\phi}_{t+1,s'} = \bar{\phi}_{t,s'} + 2\pi L_t(N+\nu) \left[\zeta^{(\text{res})} + \frac{(2s'+1)L_f-1}{2N} \xi \right] \quad (31)$$

By neglecting the interference term $\Xi(t, s')$, we define the CFO/SFO estimation metric on sub-band s' :

$$\begin{aligned}\Theta_{(t,t+1),s'} &= \arg \left\{ \left(\frac{r_{t+1,s'}[p]}{x_{t+1,s'}[p]} \right) \left(\frac{r_{t,s'}[p]}{x_{t,s'}[p]} \right)^* \right\} \\ &= (\bar{\phi}_{t+1,s'} + \arg \{ \bar{h}_{t+1,s'} \}) - (\bar{\phi}_{t,s'} + \arg \{ \bar{h}_{t,s'} \})\end{aligned}\quad (32)$$

Assuming that $\bar{h}_{t,s'} = \bar{h}_{t+1,s'}$, i.e. the channel does not vary during $(2 \times L_t)$ OFDM symbols, the effect of the CFR disappears and (32) becomes:

$$\begin{aligned}\Theta_{(t,t+1),s'} &= \bar{\phi}_{t+1,s'} - \bar{\phi}_{t,s'} \\ &= 2\pi L_t(N+\nu) \left[\zeta + \frac{(2s'+1)L_f-1}{2N} \xi \right] \\ &= 2\pi L_t(N+\nu) \left[\zeta + \frac{L_f-1}{2N} \xi + \frac{L_f}{N} \xi s' \right]\end{aligned}\quad (33)$$

Hence, using a least square estimator it is possible to estimate both the residual CFO and the SFO:

$$\hat{\xi} = \frac{\sum_{s'=0}^{S-1} (s' - \bar{s}') (\Theta'_{(t,t+1),s'} - \bar{\Theta}'_{(t,t+1)})}{\sum_{s'=0}^{S-1} (s' - \bar{s}')^2} \times \frac{N}{L_f} \quad (34)$$

$$\hat{\zeta}^{(\text{res})} = \bar{\Theta}'_{(t,t+1)} - \frac{L_f-1}{2N} \hat{\xi} \quad (35)$$

where $\Theta'_{(t,t+1),s'} = \Theta_{(t,t+1),s'} / (2\pi L_t(N+\nu))$, $\bar{\Theta}'_{(t,t+1)} = \frac{1}{S} \sum_{s'=0}^{S-1} \Theta'_{(t,t+1),s'}$ and $\bar{s}' = \frac{1}{S} \sum_{s'=0}^{S-1} s'$. Similarly to the previous CFO estimation stage, we notice that to avoid any phase ambiguity, it is necessary that:

$$2\pi L_t(N+\nu) \left[\zeta + \frac{L_f-1}{2N} \xi + \frac{L_f}{N} \xi s' \right] < \pi \quad (36)$$

Using the simulation parameters given in Table I, if the SFO is null, (36) implies a maximum estimable residual CFO value equal to 2.5% and 5% of the intersubcarrier spacing, for a time spreading length equal to 8 and 16 respectively. Since the maximum residual CFO at the output of the initial synchronization stage is lower than 2%, the proposed residual CFO estimation is well suitable.

To analyse the performance of the joint residual CFO and SFO estimation after deprecoding function, we give in Fig. 12 the final residual CFO and SFO measured at the output of the 2nd synchronization stage, in open loop, using the simulation parameters given in Table II. We set the CFO value to 2% of the intersubcarrier spacing which is the largest value resulting from the initial synchronization step. To highlight the influence of the time spreading length L_t and the mobile velocity on the estimators, the simulations have been carried out for two different L_t values and different mobility scenarios under the TU6 channel model. As expected, the proposed residual CFO estimation after deprecoding function allows the CFO to be advantageously reduced compared to the values obtained with the fine CFO estimation before deprecoding function. Indeed, for any mobility scenario and any $\frac{E_b}{N_0}$ value, the final residual CFO measured at the output of the 2nd synchronization stage is lower than 0.55% whereas it is higher than 1% after the initial synchronization step. Similarly to the residual CFO at the output of the initial synchronization stage, beyond a given $\frac{E_b}{N_0}$ value, the final residual CFO and SFO curves reach a floor due to the MCI. It appears that for moderate mobility scenarios (until 60km/h), the final residual CFO and SFO are similar for a time spreading length L_t both equal to 4 and 8. On the other hand, for high mobility scenarios, the higher the L_t value, the more significant the final residual CFO and SFO. It is explained by the fact that the constraint of flatness of the channel is all the more drastic than the L_t value gets higher, which translates into a higher sensitivity of our algorithm in high mobility scenarios.

Fig. 13 gives the global system performance in term of BER under the TU6 channel model. The CFO and SFO values are set to 10% and 100ppm respectively. The BER curves for QPSK and 16QAM data symbols have been simulated for a mobile velocity equal to 120km/h and 60km/h respectively. The DVB-T system performance is given as reference with perfect CFR estimation and perfect synchronization. To focus on the intrinsic performance of the proposed frequency synchronization method, perfect CFR estimation by subcarrier is

| | |
|--|---|
| Bandwidth | 8 MHz |
| FFT size N_{FFT} | 2048 samples |
| Guard Interval size | 512 samples (64 μ s) |
| OFDM symbol duration T_{OFDM} | 280 μ s |
| Carrier frequency | 500 MHz |
| Data symbol constellations | QPSK - 16QAM |
| Pilot symbol constellation | BPSK |
| Convolutional code rate | $R_c = 1/2$ using (133, 171) _o |
| Frequency spreading length | $L_f = 4$ |
| Channel model | TU6 |
| Mobile Speeds | 20km/h - 60km/h - 120km/h |
| Relative Doppler Frequencies β | 0.003 - 0.014 - 0.028 |
| Relative Carrier Frequency Offset | 10% |
| Relative Residual Carrier Frequency Offset $\zeta^{(res)}$ | 2% |
| Relative Sampling Frequency Offset ξ | 100 ppm |
| Loop filter gain for CFO estimation | 1/16 |
| Loop filter gain for SFO estimation | 1/64 |

Table II Simulation parameters

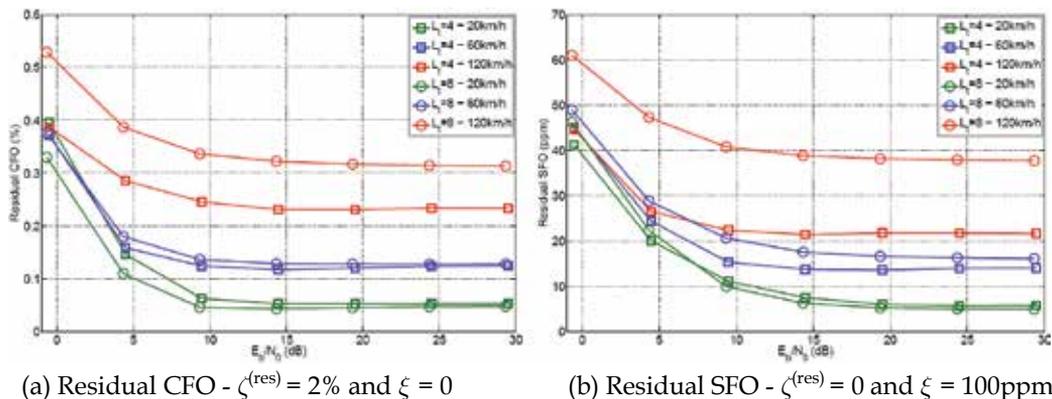


Fig. 12. Residual CFO and SFO after the 2nd synchronization stage - 16QAM data symbols - TU6 channel - $L_f = 4$

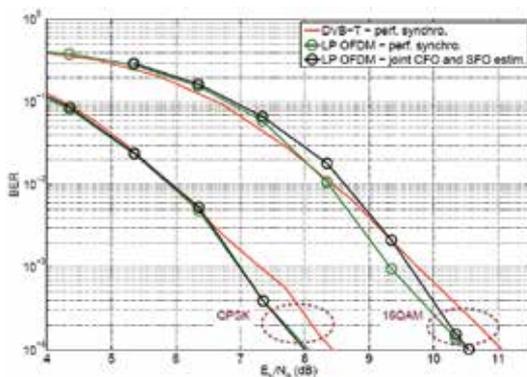


Fig. 13. BER performance under TU6 channel - Mobile Speed: 120km/h for QPSK and 60km/h for 16QAM - $R_c = 1/2$ - $\zeta = 1\%$ and $\xi = 100\text{ppm}$ - $L_f = 4$ and $L_t = 8$

assumed. The final estimated CFO and SFO values used to synchronize the received signal are those obtained at the output of the loop filters which gains values are set as in [5]. Similarly to CFR estimation simulations, the proposed system based on spread pilot synchronization slightly outperforms the DVB-T system with perfect synchronization. Moreover, the performance of the proposed synchronization method is very close to perfect synchronization case. It proves that the impact of such residual CFO and SFO values on BER performance are negligible and thus validates the benefit of the proposed CFO and SFO estimation techniques.

5. Conclusion

In this study, we proposed an efficient and very simple joint CFR, CFO and SFO estimation based on spread pilots for digital video broadcasting systems. The specificity in the LP OFDM waveform based on spread pilots is that all the active subcarriers convey a part of the pilot information, contrary to the classical OFDM systems in which only a few subcarriers are defined as pilot. It allows to not have to define different pilot symbols for each estimation algorithms. Thus, it avoids a reduction of the spectral efficiency and the useful bit rate of the system.

In contrast to classical existing OFDM systems, a deprecoding function is used instead of an interpolating filtering technique for CFR estimation. Therefore, the CFR estimation is highly simplified. Nevertheless, an interference term from data symbols appears in the estimators. This interference term called MCI is function of the autocorellation of the channel for the estimations carried out after despreading the pilot symbols. To avoid a significant degradation of the system performance, the time and frequency spreading lengths can be optimized depending on the channel characteristics. Let us note that an interference cancellation based on data decision directed could mitigate the MCI.

Two synchronization stages have been proposed. The first one processed before deprecoding function is dedicated to fine CFO estimation which brings down the CFO to less than 2% of the intersubcarrier spacing in open loop for any mobility scenario. The second one is applied after despreading the pilots symbols. It estimates both the residual CFO and the SFO. Although, it is more sensitive to high mobility scenarios, it improves the CFO estimation and diminuates the final residual CFO to a value lower than 0.55% in open loop. Finally, the simulations show that the proposed synchronization algorithm performance in term of BER is very close to perfect synchronization case in closed loop.

To conclude, the proposed system based on spread pilots is more flexible due to the possible adaptation of the time and frequency spreading lengths. It offers an improvement of the spectral efficiency and the useful bit rate which is stated in Table III. Eventually, taking into account the power loss due to pilot symbol insertion, it slightly outperforms the DVB-T system with perfect CFR estimation and perfect frequency synchronization.

| | |
|--|--|
| Bandwidth | 8 MHz |
| FFT size N_{FFT} | 2048 samples |
| Guard Interval size | 512 samples (64 μ s) |
| OFDM symbol duration T_{OFDM} | 280 μ s |
| Convolutional code rate | $R_c = 1/2$ using (133, 171) _o |
| Useful bit rates of DVB-T system | 4.98 Mbits/s for QPSK 9.95 Mbits/s for 16QAM 14.93 Mbits/s for 64QAM |
| Useful bit rates of 2D LP OFDM for QPSK | 5.33 Mbits/s for $L = 16$ 5.51 Mbits/s for $L = 32$ 5.60 Mbits/s for $L = 64$ |
| Useful bit rates of 2D LP OFDM for 16QAM | 10.66 Mbits/s for $L = 16$ 11.02 Mbits/s for $L = 32$ 11.20 Mbits/s for $L = 64$ |
| Useful bit rates of 2D LP OFDM for 64QAM | 15.99 Mbits/s for $L = 16$ 16.53 Mbits/s for $L = 32$ 16.80 Mbits/s for $L = 64$ |

Table III Useful bit rates of the DVB-T system of 2D LP OFDM system

6. Acknowledgement

This work was supported by the French national project "Mobile TV World" and the European project CELTIC B21C ("Broadcast for 21st Century") [19].

Appendix

In this section, a property from the random matrix and free probability theories is defined for the computation of the MCI variance (7). Furthermore, the computation of the autocorrelation function of the channel R_{HH} is carried out.

Random matrix and free probability theories property

Let \mathbf{C} be a Haar distributed unitary matrix [13] of size $[L \times L]$. $\mathbf{C} = (\mathbf{c}_p, \mathbf{C}_u)$ can be decomposed into a vector \mathbf{c}_p of size $[L \times 1]$ and a matrix \mathbf{C}_u of size $[L \times (L - 1)]$. Given these assumptions, it is proven in [20] that:

$$\mathbf{C}_u \mathbf{P}'_u \mathbf{C}_u^H \xrightarrow{L \rightarrow \infty} \alpha P_u (I - \mathbf{c}_p \mathbf{c}_p^H) \quad (37)$$

where $\alpha = 1$ is the system load and $P_u = 1$ is the power of the interfering users.

Autocorrelation function of the channel

The autocorrelation function of the channel writes:

$$R_{HH}(\Delta q, \Delta n) = E \{ H_{t,s}[q, n] H_{t,s}^*[q - \Delta q, n - \Delta n] \} \quad (38)$$

We can express the frequency channel coefficients $H_{t,s}[q, n]$ as a function of the channel impulse response (CIR):

$$H_{t,s}[q, n] = \sum_{k=0}^{N_{\text{FFT}}-1} \gamma_{t,q}[k] e^{-2j\pi \frac{(sL_f+n)}{N_{\text{FFT}}} k} \quad (39)$$

where $\gamma_{t,q}[k]$ is the complex amplitude of the k th sample of the CIR during the q th OFDM symbol of the m th frame, and N_{FFT} is the FFT size. Therefore, by injecting (39) in (38), the autocorrelation function of the channel can be rewritten as:

$$R_{HH}(\Delta q, \Delta n) = \frac{1}{N_{\text{FFT}}} \sum_{k=0}^{N_{\text{FFT}}-1} \sum_{k'=0}^{N_{\text{FFT}}-1} E \{ \gamma_{t,q}[k] \gamma_{t,q-\Delta q}^*[k'] \} e^{-2j\pi \frac{\Delta n}{N_{\text{FFT}}} k} \quad (40)$$

Since different taps of the CIR are uncorrelated, it comes:

$$R_{HH}(\Delta q, \Delta n) = \frac{1}{N_{\text{FFT}}} \sum_{k=0}^{N_{\text{FFT}}-1} E \{ \gamma_{t,q}[k] \gamma_{t,q-\Delta q}^*[k] \} e^{-2j\pi \frac{\Delta n}{N_{\text{FFT}}} k} \quad (41)$$

According to Jake's model [21], the correlation of the k th sample of the CIR is:

$$E \{ \gamma_{t,q}[k] \gamma_{t,q-\Delta q}^*[k] \} = \rho_k J_0(2\pi f_D \Delta q T_{\text{OFDM}}) \quad (42)$$

where ρ_k is the power of the k th sample of the CIR, $J_0(\cdot)$ the zeroth-order Bessel function of the first kind, f_D the maximum Doppler frequency and T_{OFDM} the total OFDM symbol duration. Finally, the autocorrelation function of the channel (41) can be expressed as:

$$R_{HH}(\Delta q, \Delta n) = \frac{1}{N_{\text{FFT}}} \sum_{k=0}^{N_{\text{FFT}}-1} \rho_k e^{-2j\pi \frac{\Delta n}{N_{\text{FFT}}} k} J_0(2\pi f_D \Delta q T_{\text{OFDM}}) \quad (43)$$

7. References

- [1] S. Weinstein and P. Ebert, "Data Transmission by Frequency-Division Multiplexing Using the Discrete Fourier Transform," *IEEE Trans. Commun.*, vol. 19, no. 5, pp. 628–634, Oct. 1971.
- [2] ETSI EN 300 744, "Digital Video Broadcasting (DVB) ; Framing structure channel coding and modulation for digital terrestrial television," Tech. Rep., Nov. 2004.

- [3] P. Höfer, S. Kaiser, and P. Robertson, "Two-dimensional pilot-symbol-aided channel estimation by Wiener filtering," *ICASSP-97*, vol. 3, pp. 184–1848, 1997.
- [4] H. Steendam and M. Moeneclaey, "Sensitivity of orthogonal frequency-division multiplexed systems to carrier and clock synchronization errors," *Signal Processing*, vol. 80, pp. 1217–1229, 2000.
- [5] S. A. Fechtel, "OFDM carrier and sampling frequency synchronization and its performance on stationary and mobile channels," *IEEE Trans. Consum. Electron.*, vol. 46, pp. 438–441, Aug. 2000.
- [6] O. P. Pasquero, M. Crussière, Y. Nasser, and J.-F. Hélar, "2D Linear Precoded OFDM for Future Mobile Digital Video Broadcasting," *Signal Processing Advances in Wireless Communications, 2008 IEEE 9th Workshop on*, 2008.
- [7] P. Shlichta, "Higher-dimensional hadamard matrices," *Information Theory, IEEE Transactions on*, vol. 25, no. 5, pp. 566–572, Sep 1979.
- [8] N. Kuroyanagi and L. Guo, "Proposal of an alternate m-sequence spread spectrum system," vol. 1, Oct 1993, pp. 41–46 vol.1.
- [9] N. Chapalain, D. Mottier, and D. Castelain, "Performance of uplink ss-mc-ma systems with frequency hopping and channel estimation based on spread pilots," *Personal, Indoor and Mobile Radio Communications, 2005. PIMRC 2005. IEEE 16th International Symposium on*, vol. 3, pp. 1515–1519 Vol. 3, Sept. 2005.
- [10] C. Ho, B. Farhang-Boroujeny, and F. Chin, "Added pilot semi-blind channel estimation scheme for ofdm in fading channels," *Global Telecommunications Conference, 2001. GLOBECOM '01. IEEE*, vol. 5, pp. 3075–3079 vol.5, 2001.
- [11] B. Molnar, I. Frigyes, Z. Bodnar, and Z. Herczku, "The wssus channel model: comments and a generalisation," *Global Telecommunications Conference, 1996. GLOBECOM '96. 'Communications: The Key to Global Prosperity*, pp. 158–162, Nov 1996.
- [12] J. Evans and D. Tse, "Large system performance of linear multiuser receivers in multipath fading channels," *Information Theory, IEEE Transactions on*, vol. 46, no. 6, pp. 2059–2078, Sep 2000.
- [13] M. Debbah, W. Hachem, P. Loubaton, and M. de Courville, "Mmse analysis of certain large isometric random precoded systems," *Information Theory, IEEE Transactions on*, vol. 49, no. 5, pp. 1293–1311, May 2003.
- [14] O. P. Pasquero, M. Crussière, Y. Nasser, and J.-F. Hélar, "Efficient space code block code mimo channel estimation for future mobile video broadcasting," *Wireless Communications and Networking Conference, 2009. WCNC 2009. IEEE*, pp. 1–6, April 2009.
- [15] T. Pollet, M. Van Bladel, and M. Moeneclaey, "BER Sensitivity of OFDM Systems to Carrier Frequency Offset and Wiener Phase Noise," *IEEE Trans. Commun.*, vol. 43, pp. 191–193, Feb/Mar/Apr 1995.
- [16] Y. Nasser, M. Noes, L. Ros, and G. Jourdain, "Sensitivity of OFDM-CDMA Systems to Carrier Frequency Offset," *Communications, ICC 2006. IEEE International Conference on*, vol. 10, pp. 4577–4582, June 2006.
- [17] H. Zhou, A. Malipatil, and Y.-F. Huang, "Synchronization issues in ofdm systems," Dec. 2006, pp. 988–991.
- [18] M. Speth, S. Fechtel, G. Fock, and H. Meyr, "Optimum receiver design for ofdm-based broadband transmission .ii. a case study," *Communications, IEEE Transactions on*, vol. 49, no. 4, pp. 571–578, Apr 2001.
- [19] <http://www.celticinitiative.org/Projects/B21C>.
- [20] J.-M. Chaufray, W. Hachem, and P. Loubaton, "Asymptotic analysis of optimum and suboptimum cdma downlink mmse receivers," *Information Theory, IEEE Transactions on*, vol. 50, no. 11, pp. 2620–2638, Nov. 2004.
- [21] W. Jakes, Ed., *Microwave Mobile Communications*. IEEE Press, 1994.

Performance Analysis of DVB-T/H OFDM Hierarchical Modulation in Impulse Noise Environment

Tamgnoue Valéry, Véronique Moeyaert, Sébastien Bette and Patrice Mégret
*University of Mons (UMONS), Faculty of Engineering,
Department of Electromagnetism and Telecommunications, Mons
Belgium*

1. Introduction

DVB-T\H (Digital Video Broadcasting - Terrestrial \ Handheld) are standards of the DVB consortium for digital terrestrial handheld broadcasting and are based on Coded Orthogonal Frequency Division Multiplex (COFDM) signals as described in the documents ETSI 300 744 and ETSI EN 302 304 in 2004. Orthogonal Frequency Division Multiplex (OFDM) modulation is based on a multi-carriers scheme, instead of the single carrier modulation scheme used classically in DVB-Cable. In OFDM, the bit stream to be transmitted is serially separated and modulated in parallel over several subcarriers which increases the robustness to multipath in wireless environment.

In the DVB standard, the recommended modulation scheme is multi-carriers QAM modulation but hierarchical modulation is also proposed as an optional transmission mode where two separate bit streams are modulated into a single bit stream. The first stream, called the "High Priority" (HP), is embedded within the second stream called the "Low Priority" (LP) stream. At the receiver side, equipments with good receiving conditions demodulate both streams, while those with bad receiving conditions only demodulate HP stream. Hierarchical modulation, therefore, gives more service opportunities for receivers in good operating conditions compared to receivers in bad receiving conditions that only decode basic services.

Hierarchical modulation has been included in many digital broadcasting standards such as MediaFLO, UMB (Ultra Mobile Broadband), T-DMB (Terrestrial Digital Multimedia Broadcasting), DVB-SH (Digital Video Broadcasting - Satellite Handheld). Nevertheless, in the new version of video broadcasting, DVB-T2 (Digital Video Broadcasting- Terrestrial second version), the hierarchical modulation has been replaced by Physical Layer Pipes (PLP). Notice that, in order to provide dedicated robustness, the PLP technologies allow different levels of modulation, coding and power transmitted per service. Thus, PLP differentiates the quality in service by service basis, while the hierarchical modulation works based on receivers conditions independently of the services (DVB Document A122, 2008).

Impulse noises combined with Gaussian noises are one of the major causes of errors in digital communications systems and thus in DVB-T\H networks (Biglieri, 2003). These two kinds of noises behave in a different way to corrupt digital communications. Gaussian noise,

also known as background noise, is permanently present through the network with a moderate power. On the contrary, impulse noise randomly appears as bursts of relative short duration and very large instantaneous power.

The primary purpose of this chapter is to investigate the behaviour of coupled OFDM/QAM and hierarchical modulation systems under impulse noise environment as it appears in DVB-T/H specifications. This impact is evaluated firstly by expressing Bit Error Rate (BER), when it is simulated, or Bit Error Probability (BEP) when it is calculated, of both HP bit stream and LP bit stream with impulsive impairment, and secondly by validating these expressions through simulation. The Signal to Noise Ratio (SNR) penalty induced by the presence of impulse noise in each stream is also analytically estimated and confirmed by computer simulations.

This chapter will first introduce the insights of OFDM and hierarchical modulation in DVB-T/H and will also give some examples of possible implementations. After, it will give a brief overview of impulsive noise characteristics. Then, it will describe the method to theoretically analyze the BER at the bit level in presence of impulse noise. At this level, an original method to obtain the SNR penalty on HP stream due to the introduction of LP stream is reported. And finally, the conclusion will bring some comments in next DVB-T/H technologies.

2. System model

DVB-T/H are technical standards that specify the framing structure, channel coding and modulation for terrestrial handheld television. These standards define the transmission of encoded audio and video signals using the COFDM modulation.

Many parameters, which are listed here, are settled to characterize DVB-T/H transmission stream:

- The code rate: it is the ratio of data rate of the useful bits to overall data rate (typical values are: $\frac{1}{2}$, $\frac{2}{3}$, $\frac{3}{4}$, $\frac{5}{6}$, $\frac{7}{8}$).
- The modulation order: non-hierarchical (4-QAM, 16-QAM, 64-QAM), hierarchical (4/16-QAM, 4/64-QAM).
- The guard interval: it is defined to guarantee the safety interval for the subsequent symbol (typical length of guard interval: $\frac{1}{4}$, $\frac{1}{8}$, $\frac{1}{16}$, $\frac{1}{32}$).
- The number of sub-carriers: 2k (2048), 4k (4096, only for DVB-H) and 8k (8192).
- The hierarchical uniformity parameter: 1 for uniform constellation and 2 or 4 for non-uniform constellation.

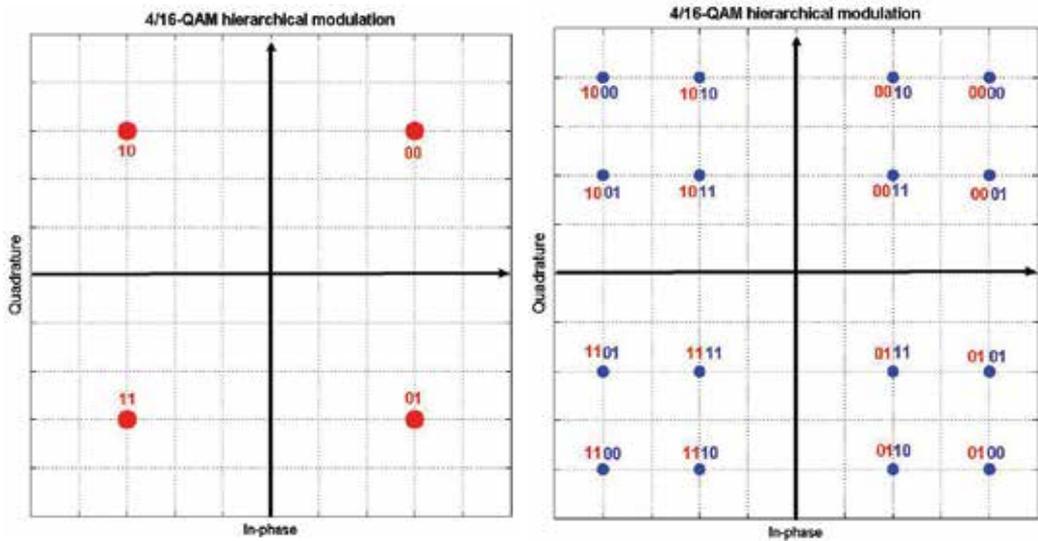
2.1 Introduction of hierarchical modulation

Hierarchical Modulation (HM) is defined to be used on quadrature modulations (QAM) and mainly consists of two separate bit streams that are modulated onto a single bit stream. The first stream, called the "High Priority" (HP), is modulated with basic quadrature modulation order (4-QAM) and the second stream, "Low Priority" (LP), is embedded on HP stream to define high order QAM modulation (M-QAM, where M is greater than 4).

The Hierarchical Modulation is designed such that HP stream is more robust against disturbances than LP stream. Therefore, only receiver with good channel condition (imposed by distance, fading, interferences, and impulse noise) can decode HP and LP streams while others can only use HP stream. This leads to differentiate service content depending of receiving conditions.

In DVB-T/H systems, Hierarchical Modulation (HM) is only defined on QAM (Quadrature Amplitude Modulation) and is denoted 4/M-QAM. Generally, square QAM is used in which M is 2^{2n} , and $2n$ is the number of bits per symbol. The 4/M-QAM is simply defined by its constellation diagram as depicted in fig. 1 for a 4/16-QAM. In this diagram, the HP stream is modulated with QPSK, equivalent to 4QAM, and thus specifies the quadrant in the entire constellation. The LP stream is modulated with (M/4)-QAM, and is embedded on primary QPSK. The entire constellation is mapped with Gray coding to allow only one bit variation between adjacent symbols.

For example, in fig. 1 for 4/16-QAM, on the entire $\log_2(16)=4$ bits ($i_1q_1i_2q_2$) of the 4/M-QAM symbol, the first two bits, i_1q_1 , are assigned the HP stream (fig.1a), and the next $\log_2(M)-2$, i_2q_2 , bits are allocated to LP stream (fig. 1b).



(a) HP stream (i_1q_1) modulated in QPSK

(b) Embedded LP (i_2q_2) stream into QPSK to give ($i_1q_1i_2q_2$)

Fig. 1. Illustration of construction of constellation diagram of 4/16-QAM modulation

Hence, hierarchical modulation can be viewed as HP stream transmitted with the fictitious symbol of QPSK defined by the first two bits (each fictitious symbol defining a quadrant), whereas the LP stream used the (M/4)-QAM modulation around these fictitious symbols.

Many parameters are used to characterize hierarchical constellation and are defined in fig. 2:

- $2d_1$ which represents the minimum distance between two fictitious symbols,
- $2d_2$ which represents the minimum distance between two neighboring symbols within one quadrant,
- $2d'_1$ which represents the minimum distance between two symbols in adjacent quadrants.

The ratio, α , of d'_1 and d_2 is the uniformity parameter and it defines the balance of power between HP stream and LP stream. As illustrated in fig. 3, the hierarchical QAM constellation is called:

- uniform constellation when $\alpha = 1$,
- non-uniform constellation when $\alpha \neq 1$.

Uniform constellation leads to equal power distribution between HP and LP while for non-uniform constellation, the power is unbalanced and is more in HP stream for growing α .

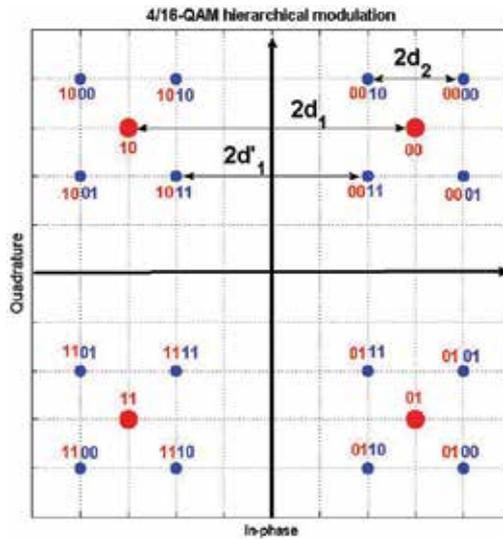


Fig. 2. Presentation of HM parameters in constellation diagram of 4/16-QAM modulation.

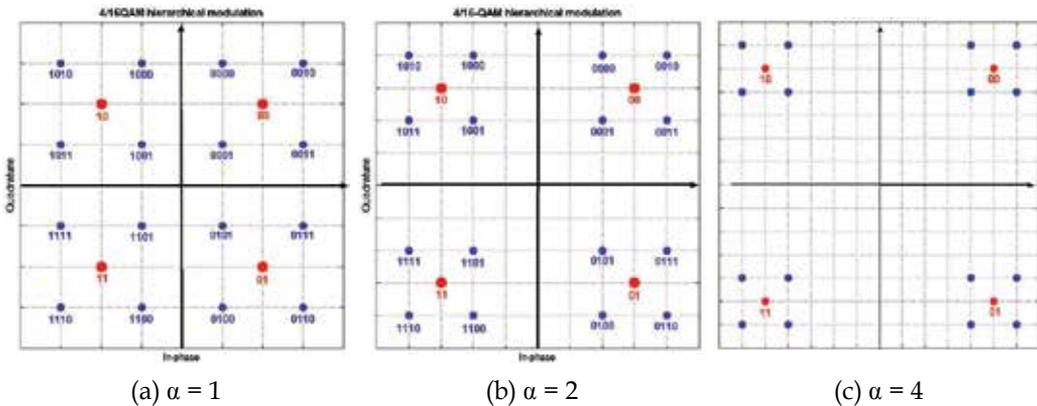


Fig. 3. Illustration of constellation diagram of 4/16-QAM modulation with various uniformity parameter ($\alpha = 1,2,4$).

In the square 4/M-QAM, i.e. where the constellation order is $M=2^{2n}$ which is the case of interest, the distances parameters are linked by this expression (Vitthaladevuni, 2001):

$$d_1 = d'_1 + \left(\frac{\sqrt{M}}{2} - 1 \right) d_2 \tag{1}$$

where M is the constellation order.

Concerning the energy distribution, it has been shown in (Vitthaladevuni, 2001) that, for square 4/M-QAM, the total average energy per symbol, E_s , is given by:

$$E_s = 2d_1^2 + \frac{2}{3} \left(\frac{M}{4} - 1 \right) d_2^2 \quad (2)$$

On this expression, the average energy per symbol of HP and LP streams, E_{hp} and E_{lp} , can be written as:

$$E_{hp} = 2d_1^2 \quad (3)$$

$$E_{lp} = \frac{2}{3} \left(\frac{M}{4} - 1 \right) d_2^2 \quad (4)$$

where E_{hp} represents the average energy of 4-QAM with constellation points separated by distance $2d_1$ (fictitious symbol in 4/M-QAM). Likewise, E_{lp} represents the average energy of M/4-QAM with constellation points separated by distance $2d_2$.

One important property that characterizes the hierarchical modulation is the penalty of modulation, also called modulation efficiency (Jiang, 2005, Wang, 2008) which is defined as the excess power that is needed in the HP stream to achieve the same Bit Error Probability (BEP) than in classical QPSK. In other words, it therefore quantifies the excess of power needed for receivers which are designed to deal with HP stream compared to the case where only classical QPSK is send.

2.2 Possible implementations of hierarchical modulation

Hierarchical modulation allows different kind of resources (bandwidth and bit rate) exploitation for receivers in various conditions. Therefore, it can be used to smoothly introduced new kind of services without needing new and expensive spectrum allocation. By this way, mobile reception or High Definition (HD) can complete and enhance the offer of the broadcaster (Schertz, 2003, Faria, 2006).

In the case of mobile reception scenario, the HP stream can be dedicated to broadcasting mobile programmes, while the LP stream can be devoted to broadcast traditional fixe or portable programmes. The HP stream will present sufficient robustness to achieve mobile channel communication. However, LP stream, weaker, will use less disturbed fixe or portable channel. Trade-off between mobile and fixe or portable channel will be assured by sizing uniformity parameters and error coding properties.

In the case of High Definition scenario, HP stream can be used for Standard Definition (SD) programmes, while LP stream can transmit HD programmes. Users with HD receivers will access broadcasted enhanced HD programmes and others, with no HD capabilities, will continue to enjoy traditional broadcasted SD programmes.

2.3 Introduction of OFDM modulation

OFDM is a parallel transmission scheme where the bit stream to be transmitted is serially separated and modulated with several sub-carriers. In practice, OFDM systems are implemented using the combination IFFT (Inverse Fast Fourier Transform) at the emitter side and FFT (Fast Fourier Transform) at receiver part (Proakis, 2001). Basically, the information bits are mapped into N baseband complex symbols c_k using quadrature modulation scheme as shown in fig. 4. The block of N complex baseband symbols, considered in the frequency domain, is changed by means of an IFFT that brings signal into

the time domain. The sequence of complex received symbols r_l , after sampling and assuming ideal channel, is given by:

$$r_l = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} c_k e^{j \frac{2\pi kl}{N}} + n_l \quad 0 \leq l \leq N-1 \quad (5)$$

where c_k is the M-QAM complex symbol of the n^{th} sub-carrier, n_l is the channel noise (jointly impulsive and Gaussian noises in this case), and N is the number of sub-carriers. The estimated baseband complex M-QAM symbol is recovered by performing a FFT that transforms the received signal in frequency domain, and it is given by:

$$\hat{c}_k = \frac{1}{\sqrt{N}} \sum_{l=0}^{N-1} r_l e^{-j \frac{2\pi kl}{N}} = c_k + \frac{1}{\sqrt{N}} \sum_{l=0}^{N-1} n_l e^{-j \frac{2\pi kl}{N}} = c_k + z_k \quad 0 \leq k \leq N-1 \quad (6)$$

where z_k denotes an additive noise term which is in fact the frequency conversion of n_l . From relation (2), it can be seen that the noise in the k^{th} QAM symbol depends on all noise samples present during the OFDM symbol. In fact, the noise, and particularly the impulsive noise, is spread over the N QAM symbols due to the FFT operation.

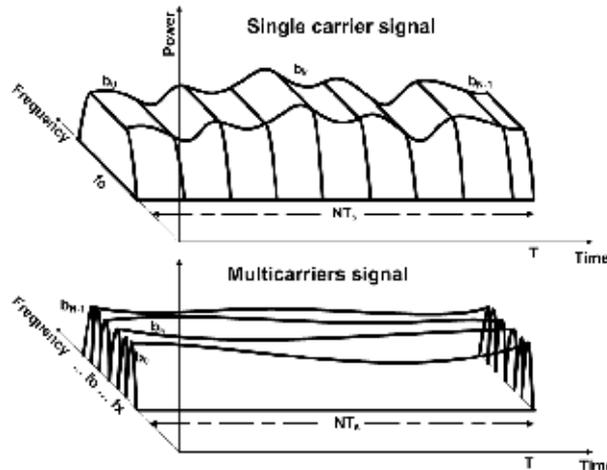


Fig. 4. Representation of OFDM modulation

2.4 Introduction of OFDM hierarchical modulation

OFDM hierarchical modulation is the combination of an OFDM system and some hierarchical QAM modulation as illustrated in fig. 5. It is constituted by a concatenation of a hierarchical QAM modulator and an OFDM modulator.

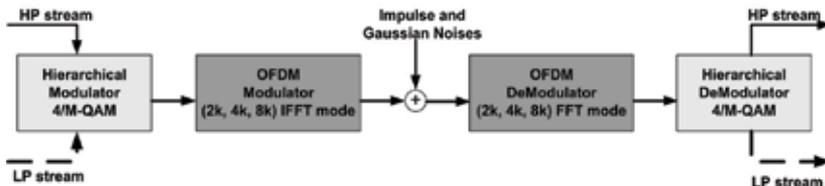


Fig. 5. Simulated block diagram of OFDM hierarchical QAM system

In DVB-{T, H} systems, OFDM modulation uses 2k (2048), 4k (4096), 8k (8192) subcarriers. Further, Hierarchical Modulation (HM) is defined based on QAM (Quadrature Amplitude Modulation) and it is denoted 4/M-QAM.

2.5 The Middleton class-A model

Impulse noise is basically defined with three statistical properties: the duration, the inter-arrival and the voltage amplitude. Middleton class-A is a complete canonical statistical model of joint impulse and Gaussian noise where the properties are defined by a compound Poisson process (Middleton, 1979, Berry, 1981). For this model, the in-phase and quadrature Probability Density Function (PDF) of voltage $f_n(x, y)$ is given by:

$$f_n(x, y) = e^{-A_A} \sum_{q=0}^{\infty} \frac{A_A^q}{q!} \cdot \frac{1}{2\pi\sigma_q^2} \exp\left(-\frac{x^2 + y^2}{2\sigma_q^2}\right) \quad (7)$$

where $\sigma_q^2 = \frac{(q/A_A + \Gamma)}{1 + \Gamma}$, $\Gamma = \frac{\sigma_g^2}{A_A\sigma_i^2}$, A_A is the impulsive index, Γ is the mean power ratio of the Gaussian component and the impulsive component, σ_g^2 and σ_i^2 are respectively the variances of Gaussian and impulse noises.

Specifically, A_A corresponds to the product of the average rate of impulse generation and the mean duration of typical impulses. Small values of A_A and Γ give predominance to impulse type of noise while large values of A_A and Γ lead to a Gaussian type of noise.

3. Bit error probability of OFDM hierarchical QAM calculation

3.1 Bit error probability of hierarchical 4/M-QAM calculation

In hierarchical modulation, the bit error probability must be determined at the bit level. In fact, a noise can lead to error in the LP bits and not disturb the HP bits. Bit error probability of hierarchical QAM HP and LP streams has been presented in scientific literature (Vitthaladevuni, 2001, 2004), and here the principles will be described.

For square 4/M-QAM ($M=2^{2n}$), the entire bit stream is $(i_1q_1i_2q_2\dots i_nq_n)$ in which the HP stream is (i_1q_1) and the LP stream is $(i_2q_2\dots i_nq_n)$. The BEP of HP bits is given by (Vitthaladevuni, 2001, 2004):

$$P_{hp}(M) = \frac{1}{2} [P(i_1, M) + P(q_1, M)] \quad (8)$$

where $P(i_1, M)$ and $P(q_1, M)$ represent respectively the bit error probability of the first in-phase and quadrature bits of 4/M-QAM hierarchical modulation.

The symmetry of the constellation diagram reduces equation (8) to:

$$P_{hp}(M) = P(q_1, M) = P(i_1, M) \quad (9)$$

On the other hand, the BEP of the LP bits is obtained by (Vitthaladevuni, 2001, 2004):

$$P_{lp}(M) = \frac{\sum_{k=2}^{(1/2)\log_2 M} P(i_k, M) + P(q_k, M)}{\log_2(M) - 2} \quad (10)$$

where $P(i_k, M)$ and $P(q_k, M)$ represents the BEP of k^{th} in-phase and quadrature LP bits of 4/M-QAM hierarchical modulation.

Using again the symmetry of the constellation diagram, this equation is rewritten as:

$$P_{lp}(M) = \frac{2 \sum_{k=2}^{(1/2)\log_2 M} P(q_k, M)}{\log_2(M) - 2} = \frac{2 \sum_{k=2}^{(1/2)\log_2 M} P(i_k, M)}{\log_2(M) - 2} \quad (11)$$

Equations (9) and (11) show that, due to the symmetry, the bit error probability of HP and LP streams only depends on only in-phase bits, or on only quadrature bits. Therefore, in-phase bits were chosen to make error probability calculation. The in-phase bits constellation diagram which corresponds to the reduced diagram is constructed and depicted in fig. 5 for 4/16-QAM. There, a symbol is defined by $i_1-i_2-i_3-\dots$, where the dashes represent the positions in quadrature axis.

In the next part, the development of error probability for the 4/16-QAM, which can be generalized to 4/64-QAM and widely to 4/M-QAM, will be described.

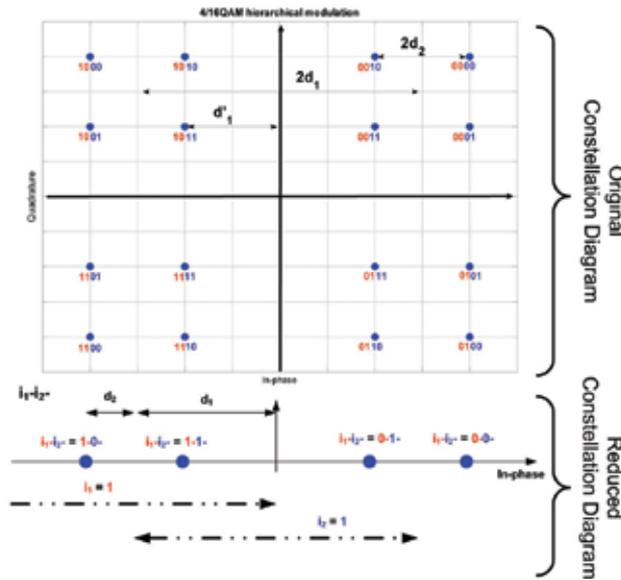


Fig. 6. Construction of in-phase reduced constellation diagram of 4/16-QAM

3.2 Case of 4/16-QAM hierarchical modulation

Before calculating the bit error probability of HP stream, we express here $P(i_1, 16)$:

$$P(i_1, 16) = P(i_1 = 1)P(\text{error} / i_1 = 1) + P(i_1 = 0)P(\text{error} / i_1 = 0) \quad (12)$$

where $P(i_1=x)$ is the probability that bit i_1 is equal to x , and $P(\text{error}/i_1=x)$ is the conditional probability to have an error if bit i_1 is equal to x .

Fig. 6 shows the reduced diagram of 4/16-QAM and two kinds of symbols with probability $\frac{1}{2}$ can be transmitted when $i_1=1$: 1-0- and 1-1-. Similarly, when $i_1=0$, the two kinds of symbols with probability $\frac{1}{2}$ can be transmitted: 0-0- and 0-1-. Derived from (12), the expression of $P(i_1,16)$ is thus given by:

$$P(i_1,16) = P(i_1 = 1) \left[\frac{1}{2} P(\text{error} / 1-0-) + \frac{1}{2} P(\text{error} / 1-1-) \right] + P(i_1 = 0) \left[\frac{1}{2} P(\text{error} / 0-0-) + \frac{1}{2} P(\text{error} / 0-1-) \right] \quad (13)$$

where $P(\text{error}/x-y-)$ is the conditional probability to have an error if the transmitted symbol is equal to $x-y-$.

Fig. 7a shows that when the bit i_1 is 1 and the emitted symbol is 1-0-, error appears when displacement on the reduced constellation diagram is greater than d_1+d_2 on the right. Likewise, fig. 7a shows that when the bit i_1 is 1 and the emitted symbols is 1-1-, error appears if the displacement is greater than d_1-d_2 on the right. Similar analysis when $i_1=0$ leads to express $P_{hp}(16)$ as:

$$P_{hp}(16) = P(i_1,16) = \frac{1}{2} \left[\frac{1}{2} \int_{d_1+d_2}^{+\infty} f(x) dx + \frac{1}{2} \int_{d_1-d_2}^{+\infty} f(x) dx \right] + \frac{1}{2} \left[\frac{1}{2} \int_{-\infty}^{-d_1-d_2} f(x) dx + \frac{1}{2} \int_{-\infty}^{-d_1+d_2} f(x) dx \right] \quad (14)$$

where $f(x)$ is the PDF of voltage amplitude of noise.

Thus, by using equation (7) to express the PDF of voltage amplitude noise $f_n(x)$, the bit error probability of HP stream can be written as:

$$P_{hp}(M) = \frac{1}{2} \left(\frac{1}{2} Y(d_1 + 2d_2) + \frac{1}{2} Y(d_1 - 2d_2) \right) \quad (15)$$

where $Y(d) = \int_{-\infty}^d f_n(x) dx$, and expressed by:

$$Y(d) = e^{-A_A} \sum_{k=0}^{+\infty} \frac{A_A^k}{k!} \cdot \frac{1}{2} \cdot \text{erfc} \left(\frac{d}{\sigma_k} \right)$$

In the other hand, the bit error probability of LP in-phase bits is obtained by calculating $P(i_2,16)$. It is given by:

$$P(i_2,16) = P(i_2 = 1) P(\text{error} / i_2 = 1) + P(i_2 = 0) P(\text{error} / i_2 = 0) \quad (16)$$

Based on reduced constellation diagram, it becomes:

$$P(i_2,16) = P(i_2 = 1) \left[\frac{1}{2} P(\text{error} / 1-1-) + \frac{1}{2} P(\text{error} / 0-1-) \right] + P(i_2 = 0) \left[\frac{1}{2} P(\text{error} / 1-0-) + \frac{1}{2} P(\text{error} / 0-0-) \right] \quad (17)$$

The fig. 7.b shows that when the bit i_2 is 1 and the emitted symbol is 1-1-, error appears when displacement on the reduced constellation diagram is greater than $2d_1-d_2$ on the right and d_2 on the left. Equally, fig. 7.b shows that when the bit i_2 is 1 and the emitted symbols is 0-1-, error appears if the displacement is greater than d_2 on the right and $2d_1-d_2$ on the left. Similar analysis when $i_2=0$ leads to express $P_{lp}(16)$ as:

$$P(i_2, M) = \frac{1}{2} \left[\frac{1}{2} \left(\int_{-\infty}^{-d_2} f(x) dx + \int_{2d_1-d_2}^{+\infty} f(x) dx \right) + \frac{1}{2} \left(\int_{-\infty}^{-2d_1+d_2} f(x) dx + \int_{d_2}^{+\infty} f(x) dx \right) \right] + \frac{1}{2} \left[\frac{1}{2} \int_{d_2}^{2d_1-d_2} f(x) dx + \frac{1}{2} \int_{-2d_1+d_2}^{-d_2} f(x) dx \right] \tag{18}$$

where $f(x)$ is the PDF of voltage amplitude of noise.

Thus, by using equation (7) to express the PDF of voltage amplitude noise $f_n(x)$, the bit error probability of LP stream can be written as:

$$P_{lp}(16) = \frac{1}{2} \left(Y(d_2) + \frac{1}{2} Y(2d_1 + d_2) - \frac{1}{2} Y(2d_1 + 3d_2) \right) \tag{19}$$

where $Y(d) = \int_{-\infty}^d f_n(x) dx$, and expressed by:

$$Y(d) = e^{-A_A} \sum_{k=0}^{+\infty} \frac{A_A^k}{k!} \cdot \frac{1}{2} \cdot \operatorname{erfc} \left(\frac{d}{\sigma_k} \right)$$

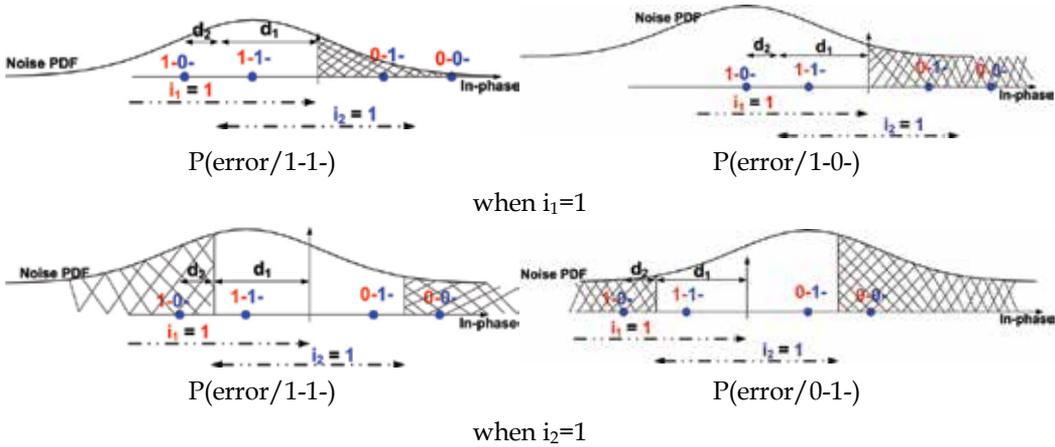


Fig. 7. Illustration of $P(\text{error}/x-y-)$ computation

Fig. 8a and 8b depict the error probability of HP and LP stream of hierarchical 4/16-QAM with α is equal to 1 and 4 and when the noise is given by Middleton class-A noise. Globally, the HP stream is less protected than the classical QPSK while the LP stream is less protected than the M-QAM with same constellation diagram.

When α is equal to 1, the HP and LP curves of error probability (red and green) are almost the same. They are around the curve of pure 16-QAM (blue), and far from the curve of pure 4-QAM (black). However, when α is equal to 4, the HP and LP curves are different. HP curve of error probability (blue) is close to the curve of pure 4-QAM (black).

This seems logic because when α grows, the energy of the constellation is more concentrated on HP stream and on base 4-QAM in the entire constellation. In fact, when α is increased, the energy of fictitious symbols grows (d_1 increases), expression (3). In contrary, the energy of refinement symbols diminishes (d_2 decreases), expression (4). Moreover, impulse noise induced a plateau in the error probability curve as it is the case of classical QAM (derived from Miyamoto, 1995).

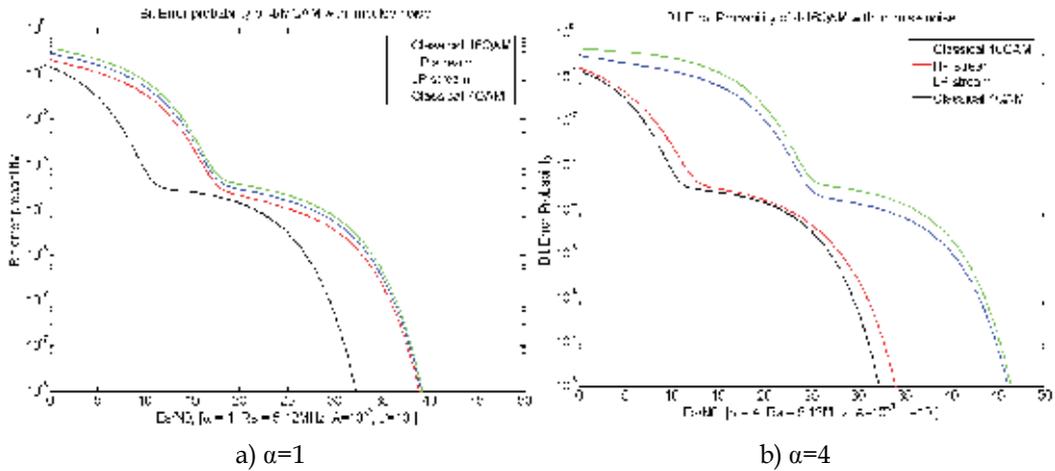


Fig. 8. Analytical curves of error probability of 4/16QAM with impulse noise ($A=10^{-3}$, $\Gamma=10$)

3.3 Analysis of modulation penalty

Hierarchical modulation allows transmission of two different streams on the same transmission channel: one hp stream based on 4-QAM, and one LP stream build above the 4-QAM. Therefore, HP 4-QAM stream differs from pure 4-QAM by a quantity called the penalty modulation.

One calculation method exists in the scientific literature (Jiang, 2005), but does not lead to exact value at high Signal to Noise Ratio (SNR) which is the case in practice.

A new calculation method has been suggested to compute penalty of modulation (Tamgnoue, 2007). For that, we define:

- SNR as the signal to noise ratio of the hierarchical modulation;
- MNR (Modulation Noise Ratio) as the signal to noise ratio of the HP stream considering the introduction of LP stream.

To estimate the MNR, the worst case symbols are selected in the constellation diagram. These symbols correspond to the nearest symbols to the middle of the constellation diagram as illustrated in fig. 9. Thus, SNR and MNR are respectively obtained as:

$$SNR = \frac{2d_1^2 + \frac{2}{3}\left(\frac{M}{4} - 1\right)d_2^2}{2\sigma_N^2} \quad (20)$$

and

$$MNR = \frac{2\alpha^2 d_2^2}{2\sigma_N^2} \quad (21)$$

Therefore, the penalty denoted P_{mnr} is given by (Tamgnoue, 2007):

$$P_{mnr} = \frac{SNR}{MNR} = \left(1 + \frac{(\sqrt{M}/2) - 1}{\alpha}\right)^2 + \frac{1}{3} \frac{\left(\frac{M}{4} - 1\right)}{\alpha^2} \quad (22)$$

This penalty is function of M and α , whereas it does not depend on SNR. It grows with M and it is in inverse proportion to α .

The fig. 10.a and fig. 10.b depict the calculated penalty in term of error probability. In these figures, the calculated penalty is in green, and the modulation penalty estimated from curves of error probability is depicted in red. Globally, the penalty does not changes significantly and is constant when probability error varies.

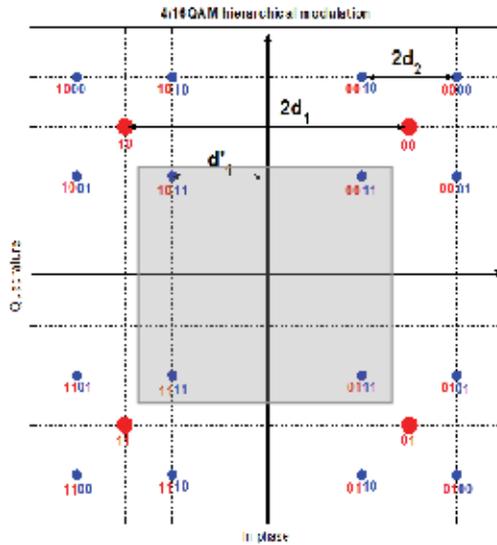


Fig. 9. Penalty analysis in constellation diagram

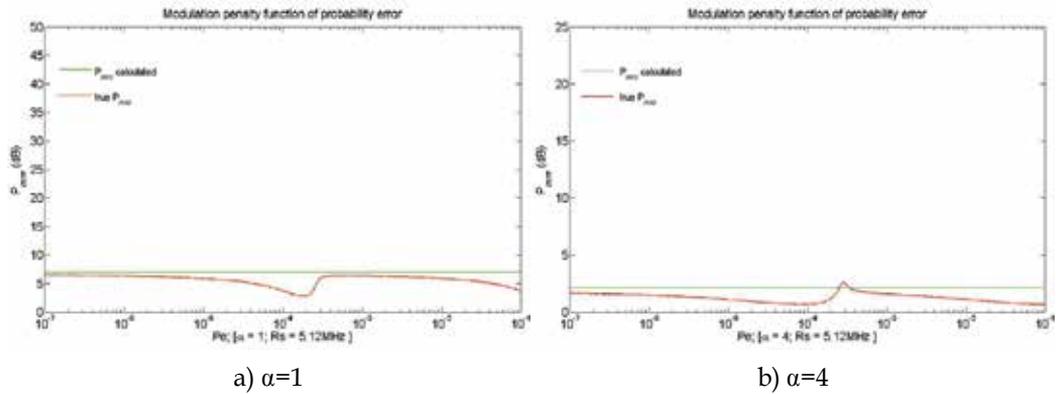


Fig. 10. Analytical curves of modulation penalty of 4/16QAM with impulse noise ($A=10^{-3}$, $\Gamma=10$)

3.4 Bit error probability of OFDM hierarchical 4/M-QAM calculation

At the receiver side, the noise in hierarchical symbol is given by:

$$z_k = \frac{1}{\sqrt{N}} \sum_{l=0}^{N-1} n_l e^{-j \frac{2\pi k l}{N}} \quad 0 \leq k \leq N-1 \quad (23)$$

where n_i is the noise in the transmission channel.

Therefore, error probability OFDM hierarchical 4/M-QAM is the same as 4/M-QAM where the noise in presence is given by z_k . PDF of z_k is thus needed.

Like presented by Huynh (Huynh, 1998), lets define:

$$u_k = \frac{1}{\sqrt{N}} n_l e^{-j \frac{2\pi k l}{N}} \quad (24)$$

Given the circular symmetry of the PDF of n_l , the PDF of u_k is given by:

$$f_u(x, y) = e^{-A_A} \sum_{q=0}^{\infty} \frac{A_A^q N}{q!} \cdot \frac{1}{2\pi\sigma_q^2} \exp\left(-\frac{(x^2 + y^2)N}{2\sigma_q^2}\right) \quad (25)$$

Following the approach discussed by Ghosh (Ghosh, 1996), it is supposed that u_k are independent random variables. Using the characteristic function method, the PDF of z_k is thus given by:

$$f_z(x, y) = e^{-A_A N} \sum_{q=0}^{\infty} \frac{(A_A N)^q}{q!} \cdot \frac{1}{2\pi\sigma_{z^2 q}^2} \exp\left(-\frac{x^2 + y^2}{2\sigma_{z^2 q}^2}\right) \quad (26)$$

where $\sigma_{z^2 q}^2 = \frac{q/(A_A N) + \Gamma}{1 + \Gamma}$ and $\Gamma = \frac{\sigma_g^2}{(A_A N) \frac{\sigma_i^2}{N}}$.

It is thus similar to general Middleton class-A equation when A_A becomes $A_A N$ and σ_i^2 becomes σ_i^2/N .

However, the hidden hypothesis surrounding this expression is that only when one pulse of duration equal to one QAM symbol impedes the entire OFDM symbol. It means that this pulse has been spread in all the N subcarriers, so duration grows N time and power decreases by a factor N .

In the general case, pulse will have duration of N_i in term on number of QAM symbols and an OFDM symbol will contain X_{imp} pulses. Therefore, PDF of z_k , $f_z(x, y)$, will become (Tamgnoue, 2007):

$$f_z(x, y) = e^{-A_A} \sum_{q=0}^{\infty} \frac{A_z^q}{q!} \cdot \frac{1}{2\pi\sigma_{zq}^2} \exp\left(-\frac{x^2 + y^2}{2\sigma_{zq}^2}\right) \quad (27)$$

$$A_z = \frac{N}{N + D_{imp}/A_A} \quad (28)$$

$$\sigma_{zq}^2 = \frac{(q/A_z + \Gamma_z)}{1 + \Gamma_z} \quad (29)$$

$$\sigma_{zg}^2 = \sigma_g^2 + \frac{(X_{imp} - 1)A_A\sigma_i^2}{X_{imp}} \quad (30)$$

In the same figures, when α is equal to 1, the HP and LP curves of error probability of OFDM 4/16-QAM are almost the same. However, when α is equal to 4, the HP and LP curves are different. HP curve of error probability (blue) is close to the curve of pure OFDM with 4-QAM (black).

4. Simulation results

4.1 Impulse noise simulation

For analysis purpose, DVB has defined impulse noise model which is based on gated Gaussian noise (ETSI TR 102 401, 2005). In this model of impulse noise, an impulsive event comprises a train of bursts and each burst contains many impulses. The impulsive events are defined by their patterns which are therefore characterized by: burst duration, burst inter-arrival time, pulse duration, impulse inter-arrival time. This pattern is used as a gate and applied to Additive White Gaussian Noise (AWGN) to obtain gated Gaussian noise. By this way, 6 patterns have been defined and used to resilience evaluation of impulse noise.

But this model does not fit directly to Middleton class-A model which is a statistical analytical model. Nevertheless, to transform this model according to analytical Middleton model, the pattern parameters have been taken as statistical parameters and some links have been defined between them. As illustrated in fig. 12, the pattern has been built in such a way that the mean values of the duration and inter-arrival time are linked with the constraint that their ratio equals the Middleton impulsive index A_A . The duration has been given by a Rayleigh distribution while the inter-arrival has been given by the Poisson distribution. The amplitude has been statistically obtained with Rayleigh distribution. On the top obtained impulse noise, AWGN has been added to simulate the Gaussian contribution with the constraint that the ratio of mean Gaussian noise power and impulse noise power is equal to the mean power ratio Γ .

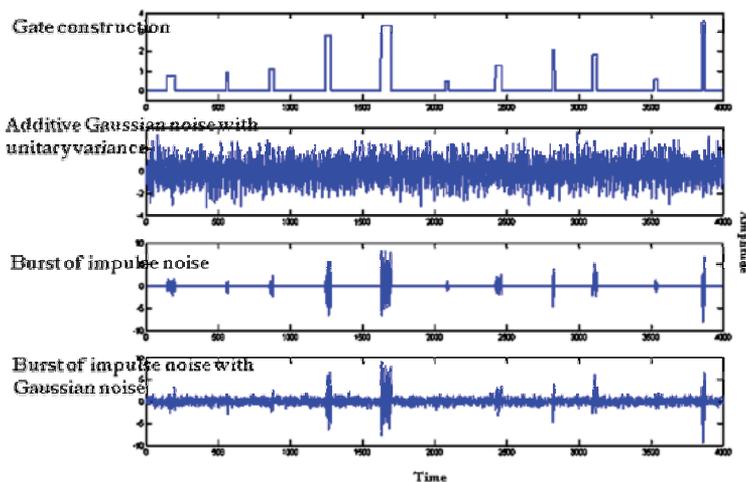


Fig. 12. Simulation of joint Gaussian and impulse noise.

4.2 Results analysis

The fig. 13.a and fig. 13.b depict some analytical and simulated bit error probability of HP and LP stream of OFDM hierarchical 4/16-QAM in presence of impulse noise. The values of

the parameters of impulse noise are given by the impulse noise test pattern #3 specified in validation Task Force Report (ETSI TR 102 401, 2005, Lago-Fernández, 2004). In this figure, we illustrate simulation points by markers, analytical curves of BEP with both impulse and Gaussian noise by solid lines and analytical curves of BEP with Gaussian noise by dashed lines.

We observe that the simulated results are very close to analytical curves. Generally, HP stream is more robust than classical OFDM QAM which is also more robust than LP stream. The parameter $\alpha = 2$ leads to a good compromise between the strength of HP stream and the weakness of LP stream.

Compared to the case where the noise is purely Gaussian, we observe a right shift in the BEP curve. This shift corresponds to the penalty induced by impulse noise. For the case of impulse noise test pattern #3, in 4k OFDM mode, the impulse noise penalty is the same for the two streams and is equal to 8.3 dB.

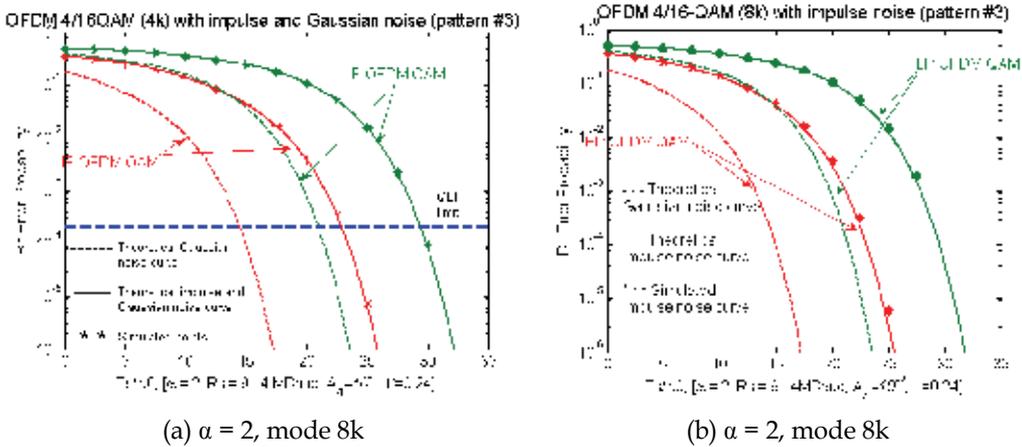


Fig. 13. Simulation of HP and LP bit error probability of 2k OFDM 4/16-QAM in presence of both impulse and Gaussian noise given by pattern #3

5. Conclusions

Hierarchical modulation is a modulation technique allowing to transmit two completely separate streams modulated in the same transmitter and on the same channel. On this, a High Priority (HP) streams are embedded within a Low Priority (LP) streams. Hierarchical modulation has been defined in many standards.

In the frame of DVT-T\H, this paper have analyzed the bit error probability of hierarchical OFDM QAM modulation in presence of both Gaussian and impulse noises. It has proved that the curve of error probability depends on the parameters of hierarchical QAM, on the number of subcarriers, and on all the noise properties (not only on the PDF of voltage). Furthermore, the power penalty induced by hierarchical modulation has been defined and analytical formula to tackle this penalty has been provided. The formulas we have derived can be used when installing a new OFDM hierarchical QAM communications. Indeed, according to the desired BEP and the impulse noise properties, transmission parameters,

like the number of subcarriers, the modulation order and the signal power, can be obtained efficiently to make right use of system resources.

Analysis of bit error probability performance has show that mix of OFDM and hierarchical modulation used in DVB-T\H systems present real differentiate performance for HP streams and LP streams. It improves the spectrum and the resources utilisation and it gives features to deal with impulse noise. It offers many opportunities for operators to delivering enhanced services. For instance, operators can launch different services for different kind of receiver (fixe or mobile). Many trials are been taken around the world demonstrating the capabilities of hierarchical modulation.

6. References

- F. Abdelkefi, P. Duhamel, F. Alberge, " Impulsive noise cancellation in multicarrier transmission", *IEEE Trans. Commun.*, vol. 53, No. 1, pp. 94-106, Jan. 2005.
- L. Berry, "Understanding Middleton's Canonical Formula for Class A Noise", *IEEE Trans. EMC*, vol. EMC-23, No. 4, pp. 337-344, Nov. 1981.
- E. Biglieri, "Coding and modulation for a horrible channel," *IEEE Commun. Mag.*, vol. 41, no. 5, pp. 92-98, May 2003.
- DVB Document A122: "Framing structure, channel coding and modulation for a second generation digital terrestrial television broadcasting system (DVB-T2)", June 2008.
- ETSI, EN 300 744, V1.5.1, Digital Video Broadcasting (DVB): framing structure, channel coding and modulation for digital terrestrial television, Nov. 2004.
- ETSI TR 102 401 v1.1.1, "Digital Video Broadcasting (DVB); Transmission to Handheld Terminals (DVB-H); Validation Task Force Report", May 2005.
- ETSI EN 302 304 v1.1.1, "Digital Video Broadcasting (DVB): Transmission System for Handheld Terminals (DVB-H)", Nov. 2004.
- ETSI 300 744 v1.5.1, "Digital Video Broadcasting (DVB): Framing structure, channel coding and modulation for digital terrestrial television", Nov. 2004.
- G. Faria, J.A. Henriksson, E. Stare, P. Talmola, "DVB-H: Digital Broadcast Services to Handheld Devices", *Proceedings of the IEEE*, Vol. 94, NO. 1, pp. 194 - 209, Jan. 2006.
- M. Ghosh, "Analysis of the effect of impulse noise on Multicarrier and Single carrier QAM systems", *IEEE Trans. Commun.* Vol.44, pp. 145-147, Feb. 1996.
- H. T. Huynh, P. Fortier, G. Delisle, "Influence of a class of man-made noise on QAM Multicarrier systems", *Seventh Communication Theory Mini-conference (Globecom '98)*, Sydney, Australie, pp. 231-236, 8-12 Nov 1998.
- H. Jiang and P. A. Wilford, "A hierarchical modulation for upgrading digital broadcast systems", *IEEE Trans. Broadcasting*, vol. 51, no2, pp. 223-229 , 2005.
- José Lago-Fernández and John Salter, "Modeling impulsive interference in DVB-T", *EBU technical review*, July 2004.
- Middleton, "Canonical Non-Gaussian Noise Models: Their Implications for Measurement and for Prediction of Receiver Performance", *IEEE Trans. EMC*, vol. EMC-21, No. 3, pp. 209-220, Aug. 1979.
- Miyamoto S., M. Katayama, "Performance analysis of QAM system under Class A impulsive noise environment", *IEEE Trans. EMC*, vol. 37, No.2, May 1995.
- JG Proakis, *Digital Communications*, 4th ed., McGraw-Hill, New York, 2001.

- Alexander Schertz and Chris Weck, "Hierarchical modulation – the transmission of two independent DVB-T multiplexes on a single TV frequency", EBU technical review, april 2003.
- H.A. Suraweera, Armstrong J., "Noise bucket effect for impulse noise in OFDM", *Electronics Letters*, vol. 40-10, pp. 1156- 1157, Sept. 2004.
- Tamnoue V., Moeyaert V., Bette S., Mégret P., "Performance analysis of DVB-H OFDM hierarchical modulation in impulse noise environment", 14th IEEE Symposium on Communications and Vehicular Technology in the BENELUX, Delft (The Netherlands), 15/11/2007.
- Tamnoue V., Moeyaert V., Mpongo H., Bette S., Mégret P., "Performance analysis of hierarchical modulation in presence of impulse noise", published in Proceedings (on CD-ROM) of BroadBand Europe 2007, Anvers (BE), 03/12-06/12, 2007.
- Vitthaladevuni P. K., Alouini M-S., "BER Computation of 4/M-QAM Hierarchical Constellations", *IEEE Transactions on Broadcasting*, Vol. 47, No. 3, pp. 228-239, 2001.
- Vitthaladevuni P. K., Alouini M-S., "A closed-form expression for the exact BER of generalized PAM and QAM constellations," *IEEE Transactions on Communications*, vol. 52, no. 5, pp. 698–700, 2004.
- Wang Shu, Kwon Soonyil, Yi, "On enhancing hierarchical modulation", 2008 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (USA-Las Vegas), pp. 1-6, March 31 2008-April 2 2008.
- Zhidkov S. V., "Performance analysis and optimization of OFDM receiver with blanking nonlinearity in impulsive noise environment", *IEEE Transactions on Vehicular Technology*, Vol. 55, No. 1, pp. 234-242, Jan. 2006.

IP Datacast and the Cost Effectiveness of its File Repair Mechanism

Bernhard Hechenleitner
Salzburg University of Applied Sciences
Austria

1. Introduction

Internet protocol datacast (IPDC) (ETSI, 2006a) is a framework for the distribution of digital data services to wireless, mobile handsets via a broadcast infrastructure based on digital video broadcasting for handhelds (DVB-H). The technical extensions of DVB-H (ETSI, 2004; Faria et al., 2006) to the widely used terrestrial DVB transmission system called DVB-T (ETSI, 2009a; Reimers, 2005) allow for specific requirements of rather small mobile devices, like a smaller screen size compared to television sets, or their dependency on accumulators for electricity and therefore their special need for low power consumption. As a broadcasting infrastructure does not automatically offer the possibility of data transmission in the “upstream” path from a mobile device to the sender, in the context of IPDC this sole one-way system is enriched with upstream paths using cellular networks or wireless local area networks (WLAN) – two types of networks which are typically used by mobile devices. Using these upstream paths for interactivity or signalling purposes, handsets are given the possibility to inform the sender about lost data segments allowing it to retransmit portions of data already sent.

This chapter will give an introduction into the digital video broadcasting (DVB) technology, especially into the DVB transmission system useful for mobile devices called DVB-H. It will describe the architecture of IP datacast, which is based on DVB-H, and the protocols used by this technology. One means of using IPDC is the broadcasting of binary data to a number of receivers using IP multicast protocols. In order to achieve correct reception of all of the sent binary data, a signalling mechanism has been specified, which enables receivers to inform the sender about lost or irreparable data fragments. This so-called file repair mechanism allows the setting of different parameters which affect the amount of data which can be corrected at the receiver and therefore have impact on the retransmission costs. The main topics of the research presented in this chapter are studies on the parameterization of the IPDC file repair mechanism and the effects of retransmissions on financial repair costs. In order to accomplish these studies, a simulation model has been designed and implemented which simulates an IP datacast network including a state-of-the-art error model for the wireless transmission channel and a versatily parametrizable implementation of the file repair mechanism.

2. Digital video broadcasting

The systems, procedures and protocols necessary for the distribution of digital television and other media services are contained in the general term “digital video broadcasting”. The specifications describing all of these components are worked out by the Digital Video Broadcasting Project, which is a consortium of over 270 broadcasting network operators, equipment manufacturers, regulatory bodies and others (DVB Project Office, 2009a). The predecessor of the DVB Project was the European Launching Group (ELG), which constituted in 1991 with the goal to introduce an open and interoperable digital television system in Europe. The DVB Project consists of several organizational parts. One of them is the Commercial Module, which is responsible for determining market requirements. Another one, the Technical Module, drafts the technical specifications necessary for implementing these requirements. Once a specification is finished by the corresponding working groups within the Technical Module, it has to be approved by the Steering Board. If this approval is successful, the specification is handed over to the European Telecommunications Standards Institute (ETSI) for formal standardisation. Although the work on DVB started in Europe, it has become a very successful solution worldwide, with more than 230 million DVB receivers (DVB Project Office, 2009a).

2.1 DVB transmission systems

The set of specifications of the DVB Project contains a multitude of technical solutions for the distribution of DVB services, some of them are shortly described in the following paragraphs.

DVB-S/DVB-S2

DVB-Satellite (DVB-S) and DVB-Satellite 2nd generation (DVB-S2) are digital satellite transmission systems. The first digital satellite TV services using DVB-S started in 1994. Currently, more than 100 million receivers are deployed conforming to this technology (DVB Project Office, 2008a). Due to the progress in many related technical areas like channel coding, modulation, error correction and media compression, the second generation digital satellite system DVB-S2 was published in 2005, taking advantages of these improvements thus creating the basis for commercially viable high-definition television (HDTV) services. As it is expected that both systems will coexist for several years, DVB-S2 was created with backwards compatibility in mind. Through the use of hierarchical modulation, legacy DVB-S receivers will continue to operate while additional capacity and services will be delivered to the second generation receivers (DVB Project Office, 2008a).

DVB-C/DVB-C2

DVB-Cable (DVB-C) and DVB-Cable 2nd generation (DVB-C2) are the first and second generation digital cable transmission systems specified by the DVB Project. Generally, the two systems make use of the wired infrastructure of cable TV providers, which mainly consists of hybrid fibre-coaxial (HFC) systems. DVB-C was published by ETSI in 1994 and the second generation DVB-C2 is expected to be published in 2009. Comparable to the satellite system, DVB-C2 among other things makes use of improved error correction and modulation schemes. As an example, data rates of up to about 80 Mbps per 8 MHz channel can be achieved when using 4096-quadrature amplitude modulation (DVB Project Office, 2009b).

DVB-T/DVB-T2

Concerning the terrestrial broadcasting of digital TV services, the corresponding first and second generation transmission systems developed by the DVB Project are called DVB-Terrestrial (DVB-T) and DVB-Terrestrial 2nd generation (DVB-T2). The first version of the specifications was published in 1997, and the first TV service using DVB-T was launched in the United Kingdom in 1998. In the meantime, more than 90 million receivers have been sold in more than 35 countries (DVB Project Office, 2009c). The second generation system provides improved modulation and error correction techniques and is expected to be published in the second quarter of 2009 (DVB Project Office, 2009d).

DVB-H

The approach of analogue switch-off (ASO) in many countries worldwide leads to the installation of digital terrestrial television (DTT) infrastructures. In Europe, as in many other parts of the world like in Australia, Asia and South America, this DTT infrastructure is based on DVB-T, mainly focused on fixed receivers. To make this system also available for mobile end systems, extensions have been specified under the designation DVB-Handheld (DVB-H). These typically small handheld devices have special characteristics like comparatively small screen sizes and special requirements like low power consumption. To meet these features and needs, appropriate extensions to the DVB-T specifications were needed (see section 2.2). Concerning DVB-H services, the main focus was on video streaming including mobile TV and file downloads (DVB Project Office, 2009e). A huge number of DVB-H trials have taken place all over the world, and full service has been launched in 14 countries (DVB Project, 2009a).

DVB-IPDC

A set of necessary components has to be built on top of the broadcasting infrastructure in order to make mobile TV services successful. Such components include e.g. an electronic service guide (ESG) for the announcement of the services offered or a framework for the purchase of specific services. These complementary building blocks have been specified within DVB-IP datacast (DVB-IPDC), which is intended to work with the DVB-H physical layer (DVB Project Office, 2008b). As a general set of system layer specifications, DVB-IPDC can be implemented for any other Internet protocol (IP) (Postel, 1981) capable system. A special property of mobile TV systems is the expected availability of an additional bidirectional connection based on 3G cellular or WLAN networks. This provides the possibility of individualization of certain services.

DVB-SH

To extend the coverage area of DVB-H based mobile TV systems, the hybrid satellite/terrestrial DVB-Satellite services to handheld (DVB-SH) transmission system has been specified. The main distribution is achieved by using satellite transmission, with terrestrial gap fillers servicing areas with poor satellite reception characteristics. DVB-SH is an extension to the DVB-H infrastructure and can be used as a transmission layer for IPDC based services. In April 2009, a satellite for the construction of a DVB-SH network targeting several European markets was launched (DVB Project Office, 2009f).

2.2 Digital video broadcasting – handhelds

As DVB-H is generally based on DVB-T, this section first of all describes the transmission system of DVB-T, followed by the extensions that constitute DVB-H.

DVB-T Transmission System

The basic transmission system that is used by both DVB-T and DVB-H was specified by the Moving Picture Experts Group (MPEG). It is called the MPEG-2 transport system and was published by the International Organisation for Standardisation (ISO) as an international standard (ISO/IEC, 2007). The basic building blocks of this system are shown in figure 1. The media encoder applies specific operations on the audio and video streams to reduce the necessary capacity for transmitting them. These are in general lossy encoders based on MPEG specifications. For example, using MPEG-2 video encoding, a standard definition television (SDTV) video signal can be compressed to about 4 – 6 Mbps instead of 166 Mbps without compression (Reimers, 2005). The compressed media streams are called elementary streams (ES) and are input to the packetizer, which divides each stream into packets of variable length up to 64 kB, depending on the content of the stream. A compressed media stream packetized in this way is called packetized elementary stream (PES). As these relatively large PES data segments are not suitable for broadcasting a number of audio and video programs within one data signal, they are further broken down into smaller data chunks called TS packets by the transport stream (TS) multiplexer. Each TS packet has a fixed length of 188 bytes consisting of a 4 bytes header and 184 bytes payload, which carries the particular parts of the PES packets. A single service program can consist of multiple audio, video and other data streams represented as streams of TS packets, and the combination of several TV programs plus additional signalling data is multiplexed by the TS multiplexer into a complete data stream called MPEG-2 transport stream.

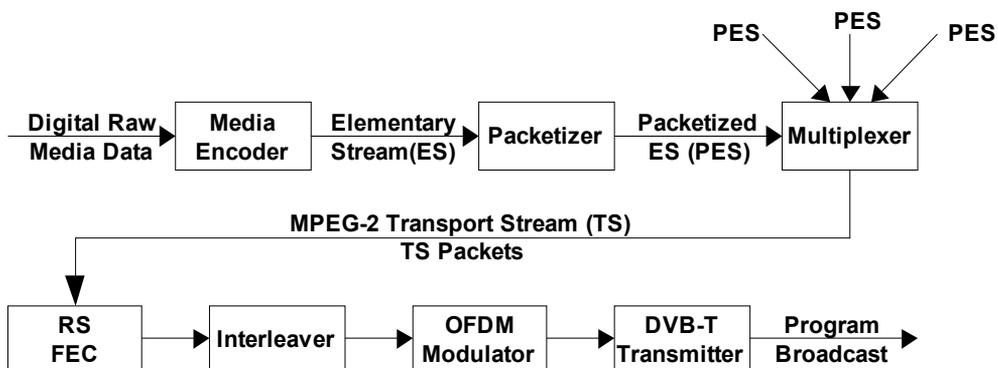


Fig. 1. Basic building blocks of the MPEG-2 transport system.

In the course of media encoding and compression, redundancy is removed from a media stream and the resulting data stream is rather fragile with respect to disruptions during the broadcast. Therefore, some redundancy has to be added systematically to be able to react on transmission errors and correct the received data stream. In DVB, primary error protection is achieved by using the Reed-Solomon (RS) forward error correction (FEC) block code. For each 188 bytes long TS packet, 16 bytes of checksum (or parity data) computed by the used RS code is added, resulting in a total length of 204 bytes. This FEC code allows the repairing of up to 8 bytes per TS packet. Packets containing more than 8 errors can not be corrected but still reliably detected and flagged as erroneous (Reimers, 2005).

Concerning DVB-T, the resulting stream of error protected TS packets is interleaved to counteract error bursts, and in the convolutional encoder additional error protection is added. After a further subdivision and interleaving of the data stream depending on the type of modulation, which can be quadrature phase-shift keying (QPSK), 16-quadrature amplitude modulation (QAM), or 64-QAM, the resulting signal is passed on to the orthogonal frequency-division multiplexing (OFDM) modulator. DVB-T supports the two OFDM transmission modes 2K and 8K, which make use of 2048 and 8192 subcarriers respectively mapped on a channel with a bandwidth of 8, 7 or 6 MHz. While the 2K mode provides larger subcarrier spacing (about 4 kHz compared to about 1 kHz using 8K mode), the symbol length is shorter (about 250 μ s versus 1 ms). For mobile devices, the 2K mode is less susceptible to the Doppler effects due to its larger carrier spacing. On the other hand, the 8K mode allows greater transmitter spacing because of longer guard intervals and is therefore the preferable mode for single frequency networks (SFN). More specific technical details of the modulation and coding techniques of DVB-T can be found in (Reimers, 2005).

Extensions of DVB-T

DVB-H was specified as a set of extensions to DVB-T focussing on small, mobile receivers. Due to limited energy supply and more difficult signal reception conditions, two main extensions to DVB-T were defined: an access scheme called time slicing and an additional FEC mechanism. With time slicing, which is mandatory in DVB-H, the DVB-H data streams are not sent continuously but in bursts. This allows the receiver to power off between these bursts, thus saving energy. Depending on the bit rate of a DVB-H stream, these power savings can be up to about 90% (ETSI, 2009b). The second extension, which is an optional extension, is called multiprotocol encapsulation forward error correction (MPE-FEC) and is described in more detail in the next paragraph. Both extensions work perfectly upon the existing DVB-T physical layer and therefore DVB-H is totally backwards compatible to DVB-T (Faria et al., 2006). Nonetheless, some extensions to the physical layer were specified as well, among these the new 4K transmission mode, which is a complement to the existing 2K and 8K modes and is a compromise between mobility and SFN cell size.

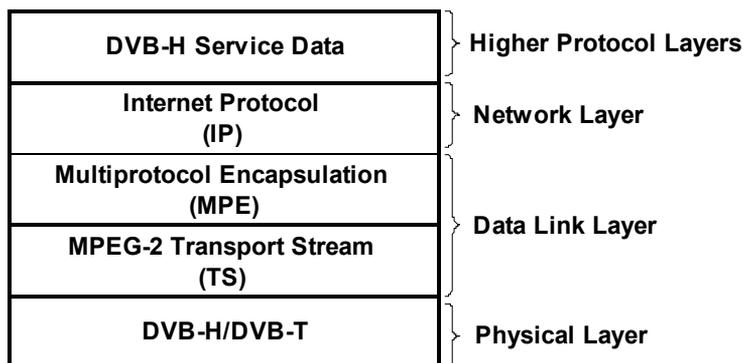


Fig. 2. DVB-H protocol stack adapted from (ETSI, 2009b).

All DVB-H payloads are carried within IP packets or other network layer protocol data units which are further encapsulated using multiprotocol encapsulation (MPE). Each IP packet is encapsulated into one MPE section and a stream of MPE sections build an elementary stream, which is further processed as described above and which is therefore finally

mapped to a stream of TS packets that can be multiplexed into an MPEG-2 transport stream (see figure 2).

Multiprotocol Encapsulation Forward Error Correction

All DVB-H user data is transported in IP packets, each of them encapsulated in one MPE section. Each MPE section consists of a 12 bytes header, an IP packet as payload (max. 4080 bytes) and 4 bytes checksum. To further protect the IP data, an additional FEC code may be applied on the data. By using multiprotocol encapsulation forward error correction (MPE-FEC), some parity data is calculated and transmitted to the receivers which in turn can use this redundancy to correct corrupt data segments. The FEC code used by MPE-FEC is the Reed-Solomon (RS) block code RS(255,191). For a more detailed explanation of this FEC code see (ETSI, 2009b) and (Reimers, 2005).

An MPE-FEC frame is a logical data structure containing two data tables: the application data table (ADT) which consists of 191 columns and the Reed-Solomon (RS) data table which consists of 64 columns containing the parity bytes calculated by the FEC code (see figure 3). The number of lines can be 256, 512, 768 or 1024. As each cell in the two tables contains one byte, the size of a MPE-FEC frame either is 47.75 kB, 95.5 kB, 143.25 kB, or 191 kB.

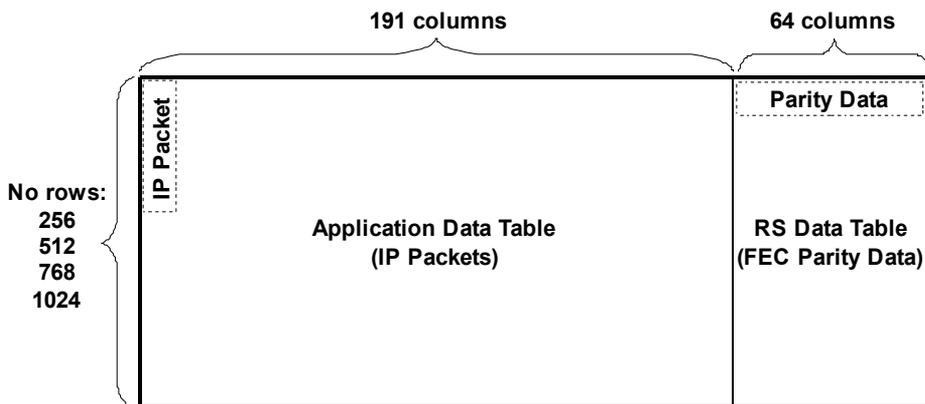


Fig. 3. MPE-FEC frame.

The IP packets plus optional padding are inserted into the ADT column by column. The RS code is applied on this data to calculate 64 parity bytes per ADT line, contained in the RS data table of the MPE-FEC frame. Concerning the data transmission, each IP packet of the MPE-FEC frame is transported within an MPE section, followed by the parity data. Each RS data table column is transported within an MPE-FEC section. The corresponding protocol stack is shown in figure 4.

2.3 IP datacast

IP datacast is a set of specifications for mobile TV systems based on IP capable systems. IPDC is specifically defined as a complement to DVB-H transmission systems, in this context called DVB-IPDC or IPDC over DVB-H. In the following, IPDC is taken short for IPDC over DVB-H. The basic building blocks of IPDC, which are described in this section, generally define how any types of digital content and services are described, delivered and protected.

Building Blocks of IPDC

An inherent part of the IPDC architecture is the possible combination of a unidirectional broadcast path realized by DVB-H and a bidirectional unicast interactivity path realized by e.g. cellular or WLAN data connections. In this way, an IPDC receiver, which is mainly built into a mobile phone, cannot only receive broadcast services but can make use of individual services or signalling through the interactivity path.

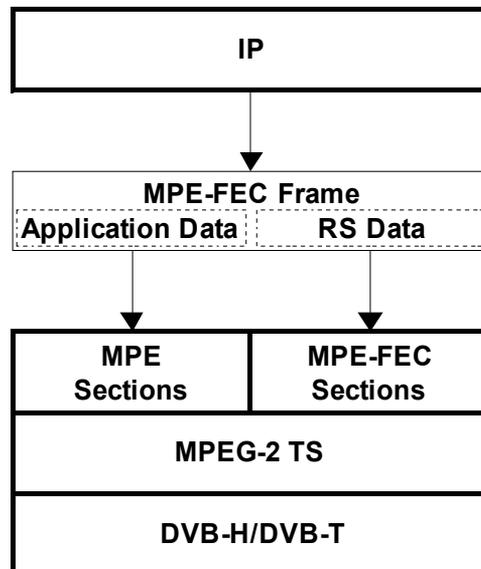


Fig. 4. DVB-H protocol stack using MPE-FEC.

The most important building blocks of IPDC are (DVB Project Office, 2008b):

- Electronic service guide (ESG): the ESG describes the services offered by an IPDC system as structured information, based on the extensible markup language (XML).
- Service purchase and protection (SPP): the SPP framework provides protocols and mechanisms for rights management and the encryption of digital content.
- Content delivery protocols (CDP): CDP is the set of protocols used by IPDC for media streaming and file delivery services.

Content Delivery Protocols

The set of protocols used for the delivery of the various mobile TV services, including media streaming and general file delivery, is called content delivery protocols (CDP) and is specified in (DVB Project, 2009b). As shown in figure 5, IPDC uses a complex stack of protocols, which can be divided in two subgroups: IP multicast based protocols for unidirectional delivery and IP unicast based protocols for interactive services, service additions or additional signalling purposes. Audio and video real-time media streams are delivered using the real-time transport protocol (RTP) and the RTP control protocol (RTCP) (Schulzrinne et al., 2003). The delivery of arbitrary binary data objects (like applications, ring tones and the like) is conducted via the file delivery over unidirectional transport (FLUTE) protocol (Paila et al., 2004), which is based on asynchronous layered coding (ALC), one of the Internet Engineering Task Force's (IETF) base protocols for massively scalable reliable

multicast distribution (Luby et al., 2002). The number and types of binary data objects to be distributed within a FLUTE session as well as other metadata are described by the XML-based file delivery table (FDT), which itself is distributed using FLUTE. Summing up, the most important protocol stacks are: RTP/UDP/IP multicast for real-time streaming of audio and video, FLUTE/ALC/UDP/IP multicast for the distribution of binary objects and HTTP/TCP/IP unicast for interactivity services.

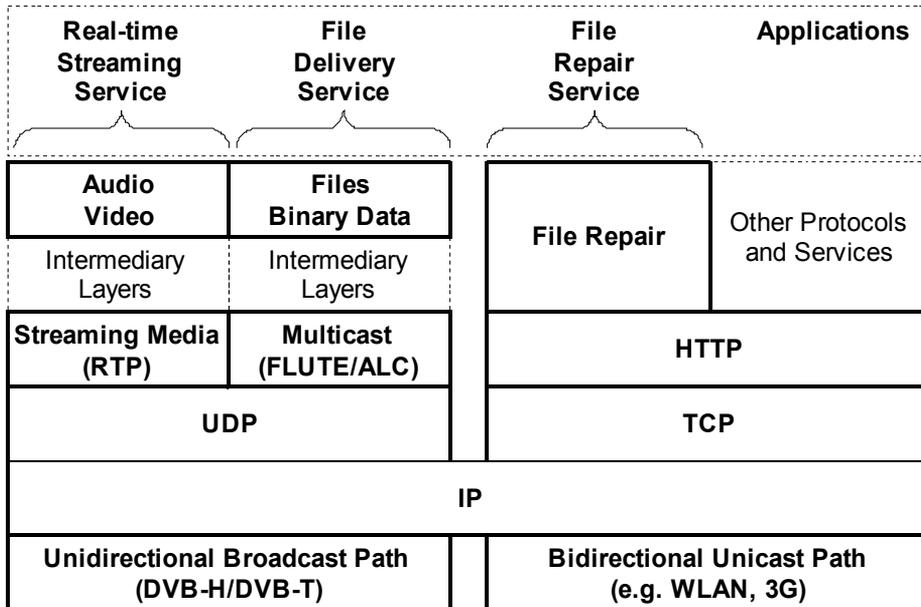


Fig. 5. IPDC content delivery protocols.

Blocking and Encapsulation

For the delivery of binary data objects using FLUTE/ALC/UDP/IP multicast, the procedure for splitting and encapsulating a binary data object for transmission via an IPDC/DVB-H system is basically as follows (ETSI, 2004; ETSI, 2006b). First, the binary object is split into source blocks and encoding symbols by a blocking algorithm as shown in figure 6. The algorithm used depends on the FEC mechanism used at the application layer. For the research done within this work (see section 4), no special application layer FEC (AL-FEC) mechanism was used¹. Therefore, the binary object is simply split into several source blocks, with each source block consisting of several encoding symbols. The number of encoding symbols per block depends on the chosen encoding symbol size. Next, each symbol produced by the blocking algorithm is encapsulated in a FLUTE/UDP/IP packet and handed over to the MPE and MPEG-2 TS layers. Finally, if additional MPE-FEC is enabled, the procedures described in section 2.2 are applied.

¹ More specifically, this scheme is called the “compact no-code FEC scheme” or “null-FEC” (Watson, 2009).

3. IPDC file repair mechanism

IPDC is a general system for the distribution of arbitrary digital services to as many receivers as are within the coverage area. One specific service is the transmission of binary data objects, such as individual songs, which are to be received error-free. In this context the file repair mechanism, which is part of the IPDC specification, allows for the signalling of transmission errors from individual receivers to the sender. This section describes the technical details and procedures of this mechanism as well as the formulas for the calculation of financial retransmission costs due to multiple transmissions of data segments.

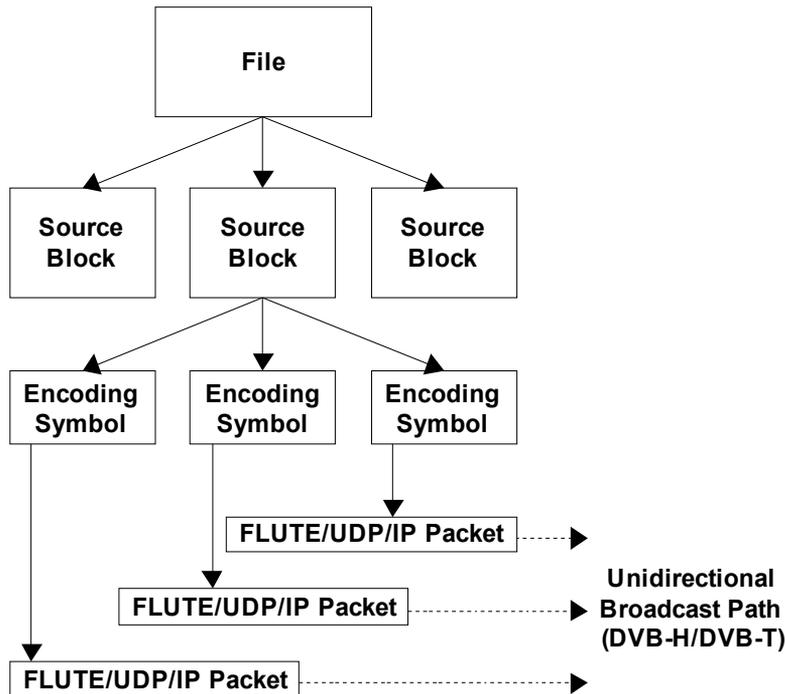


Fig. 6. Blocking and encapsulation.

3.1 File repair procedure

The IPDC file repair mechanism, specified in (DVB Project, 2009b), is part of IPDC's so-called associated delivery procedures (ADP). ADPs specify procedures for returning reception information (called "reception reporting procedures") and procedures for carrying out file repairs (called "post-repair procedures"). The former procedures are used for signalling the complete reception of one or more files as well as reporting some statistics about a streaming session. The latter ones are shortly called "file repair mechanism" and their basic purpose is the correction of lost or unrecoverable data segments of a file delivery or FLUTE session. One or more file repair servers accept file repair request messages from the receivers and either answer these requests directly via unicast 3G or WLAN connections (called "point-to-point" or ptp repair) or indirectly by repeating the requested segments via the IPDC/DVB-H broadcast infrastructure (called "point-to-multipoint" or ptm repair), thus delivering the requested repair data to all receivers once again.

The simplified basic procedure of the file repair mechanism is as follows (DVB Project, 2009b). First, a receiver identifies lost or unrecoverably erroneous data fragments or encoding symbols, respectively. Then, after the file transmission is finished, the receiver calculates a random time value called “back-off time”. After the back-off time has expired, the receiver sends its file repair requests to the file repair server. Finally, the file repair server replies with a repair response message, which either contains the requested encoding symbols within a ptp repair session or the session description for the ptm repair session.

Concerning the transmission of file repair requests a time window is defined, which begins after the data session has ended. During this window the receivers send their requests at random times, represented by the individual back-off times of the single receivers. All file repair request messages should be transmitted within an HTTP 1.1 GET request (Fielding et al., 1999) to the file repair server via the bidirectional unicast interactivity path. If more than one GET request is necessary, they should be sent without intermediary waiting times. Erroneous data fragments are specified by their source block numbers (SBN) and encoding symbol IDs (ESI). In case of a ptp file repair session, all requested fragments are sent to the receivers via HTTP responses using the interactivity path. In case of a ptm file repair session, all requested fragments are sent via a FLUTE file delivery session based on IPDC/DVB-H.

Example of a File Repair Request Message

In the following example taken from (DVB Project, 2009b) and shown in figure 7, an HTTP GET message to the file repair service “ipdc_file_repair_service” at the server with the fully qualified domain name (FQDN) “www.repairserver.com”² is shown. This message represents a file repair request message concerning the binary data object “latest.3gp”, which is a downloadable news video file hosted at the server with the FQDN “www.example.com”. In this example, the receiver was not able to correctly receive two packets with SBN = 5, ESI = 12 and SBN = 20, ESI = 27. The corresponding MD5 message-digest algorithm (Rivest, 1992) value of the file is used to identify a specific version of the file.

```
GET /ipdc_file_repair_service?fileURI=www.example.com/news/latest.3gp
&Content-MD5=ODZiYTU1OTFkZGY2NWY5ODh==
&SBN=5;ESI=12&SBN=20;ESI=27 HTTP/1.1
```

Fig. 7. Example of an IPDC file repair request message.

Example of a ptp Repair Response Message

In the following example adapted from (DVB Project, 2009c) and shown in figure 8, an HTTP response message from a file repair server is shown. The repair server uses a ptp repair strategy by directly answering the file repair request within the same HTTP session as the request. This HTTP response contains the two requested symbols in two separate groups.

² Therefore, the uniform resource locator (URL) of the service is “http://www.repairserver.com/ipdc_file_repair_service”.

| HTTP Header | | | |
|---|---|--|---------------------|
| HTTP/1.1 200 OK | | | |
| Content-type: application/simpleSymbolContainer | | | |
| Group 1 | 1 | (5 , 12) | ABCDE... |
| Group 2 | 1 | (20 , 27) | ABCDE... |
| Group N | x | (y , z) | ABCDE... |
| | No of Symbols in this Group (2 bytes) | FEC Payload ID (SBN , ESI) (4 bytes) | Encoding Symbols |

Fig. 8. Example of an IPDC ptp file repair response message.

3.2 Cost calculation

Due to the two file repair strategies, different file repair costs may result. Costs may be declared regarding the transmission volume or bytes that result from retransmitted data segments or regarding financial costs that result from transmitting the repair data via the broadcast or interactive paths, respectively. After a file transmission is finished, a file repair server can calculate the expected ptp and ptm repair costs and therefore decide which file repair strategy to choose in order to minimize the repair costs. In case of a ptp repair session, only one repair round is normally sufficient for the receiver to correct all erroneous data segments. For ptm sessions, further rounds could be necessary until all receivers have received complete and error-free data. The cost estimation is done during the time window for file repair requests, which consists of a predefined, fixed value and a random part, the maximum back-off time. At a fraction of the maximum random part of the repair request window, defined by the parameter α , the file repair server executes the calculation of the expected costs, based on the repair requests received until this point in time. After the time window for file repair requests has expired, the actual cost of one round of a file repair session can be calculated.

Projected Financial Repair Costs

The projected financial repair costs of a ptp repair session are defined by formula (1), which is specified in (DVB Project, 2009b).

The following parameters are used:

- c_u defines the financial cost of the transmission of a single byte via the used ptp network
- N_{sym} specifies the expected number of requested symbols to repeat
- s_{sym} defines the average size of an encoding symbol in bytes
- N_{req} defines the expected number of repair request messages
- s_{req} specifies the average size of a repair request message in bytes

The expected numbers of repair requests (N_{req}) and requested symbols (N_{sym}) are calculated by dividing the numbers of repair requests (n_{req}) and requested symbols (n_{sym}) the repair server accepted until the time of the expected cost calculation (t) by α . This calculation is based on the assumption that the repair requests are uniformly randomly distributed over time.

$$C_{ptp(expected)} = c_u \cdot N_{sym} \cdot S_{sym} + c_u \cdot N_{req} \cdot S_{req} \quad (1)$$

Concerning a ptp repair session, it is assumed that the repair server conducts ptp repair until time t and then switches over to ptm repair mode. A service operator should further assume that the ptm repair session contains the whole file to achieve complete reception. Therefore, formula (2) taken from (DVB Project, 2009b) can be used for the calculation of the projected ptm financial repair costs.

The following additional parameters are used:

- c_m defines the cost of the transmission of a single byte via the used ptm network
- S defines the size of the whole file in bytes
- s_{an} specifies the size of a ptm repair session announcement in bytes

$$C_{ptm(expected)} = c_m \cdot (S + s_{an}) + c_u \cdot N_{req} \cdot S_{req} + c_u \cdot n_{sym} \cdot S_{sym} \quad (2)$$

Actual Financial Repair Costs

When the time window for file repair requests has expired, the total cost for one round of a file repair session can simply be calculated by using formulas (1) and (2) and replacing the estimated values for the number of repair requests (N_{req}) and the number of requested symbols (N_{sym}) therein by their actual values n_{req} and n_{sym} respectively.

4. Simulations of the IPDC file repair mechanism

The study shown in this section deals with a simulation framework developed to examine the IPDC file repair mechanism. More specifically, it is used to examine resulting financial ptp and ptm repair costs in different simulation scenarios. The research shown focuses on the broadcast transmission of time-uncritical binary data objects, such as applications, images, ring tones, songs, or other data using an existing IPDC/DVB-H infrastructure. These data objects need to be delivered lossless, therefore the transmission relies on the IPDC/DVB-H file repair mechanism. For the examination of different aspects of this file repair mechanism, a simulation framework has been developed. The framework is based on the open source simulation tool OMNeT++ (Varga, 2009) and implements current error models for the wireless transmission channel as well as additional forward error correction measures (MPE-FEC, see section 2.2). It supports the simulation of both ptp and ptm file repair sessions, which are basically depicted in section 3.1. The calculations of expected and actual financial file repair costs are based on the formulas specified in (DVB Project, 2009b) and described in section 3.2.

The most important research questions of the study concerning the transmission of a single file to a varying number of receivers were:

- What are the financial file repair costs of both ptp and ptm repair sessions?
- What are the effects of different MPE-FEC frame sizes on the repair costs?
- What are the influences of different encoding symbol sizes on the repair costs?

In the context of this study, the costs for the first rounds of file repair sessions are compared. For ptp repair sessions, one round should be sufficient because each receiver gets its individual repair data from the repair server via an individual bidirectional connection. For ptm sessions, further rounds could be necessary until all receivers have received complete and error-free data.

4.1 Simulation topology

Figure 9 shows the general topology used for conducting the simulations. The component *Sender* represents the IPDC/DVB-H sender, which broadcasts binary data objects to the receivers. The broadcast transmission is represented by the component *PacketBroadcast*. This component simply replicates a packet sent by the sender for each receiver. The component *Receiver* incorporates the error model for the transmission channel and the optional MPE FEC error correction (see section 2.2). In this example, two receivers send file repair requests to the component *FileRepairServer* which receives file repair requests from the receivers and conducts cost calculations (see section 3.2).

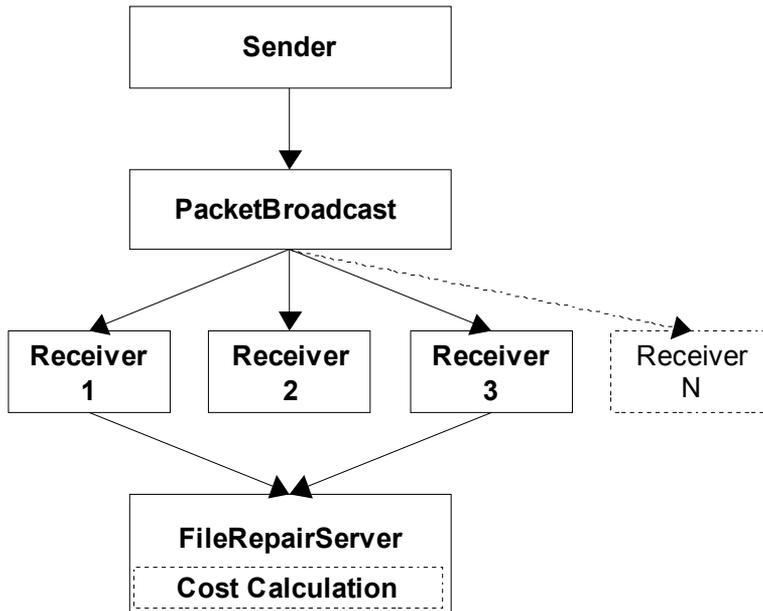


Fig. 9. Simulation topology.

The following parameters can be set for the component *Sender*: size of the data object to be transmitted (file size), size of an encoding symbol (each encoding symbol is the payload of a FLUTE packet), number of encoding symbols per source block, sum of additional protocol overheads (FLUTE/ALC, UDP, IP), and MPEG-2 TS data rate.

The component *Receiver* allows the configuration of the following parameters: file size, encoding symbol length, number of encoding symbols per source block, number of rows of the MPE-FEC frame (optional), DVB-T transmission mode (2K, 4K or 8K), minimal distance from sender in km, offset time and random time period for sending file repair requests after the transmission is finished.

Figure 10 shows the most important parts of the component *Receiver* in more detail. Each encoding symbol is transported within a FLUTE packet, which itself is mapped to TS packets of the used MPEG-2 TS. The error model described below is applied to the received stream of TS packets. Depending on the states of the error model, this results in no or several erroneous TS packets. If MPE-FEC is used, some or all of these erroneous TS packets can be recovered. Unrecoverable encoding symbols are listed in the table of erroneous symbols and can be requested at the file repair server after the transmission of the file is finished.

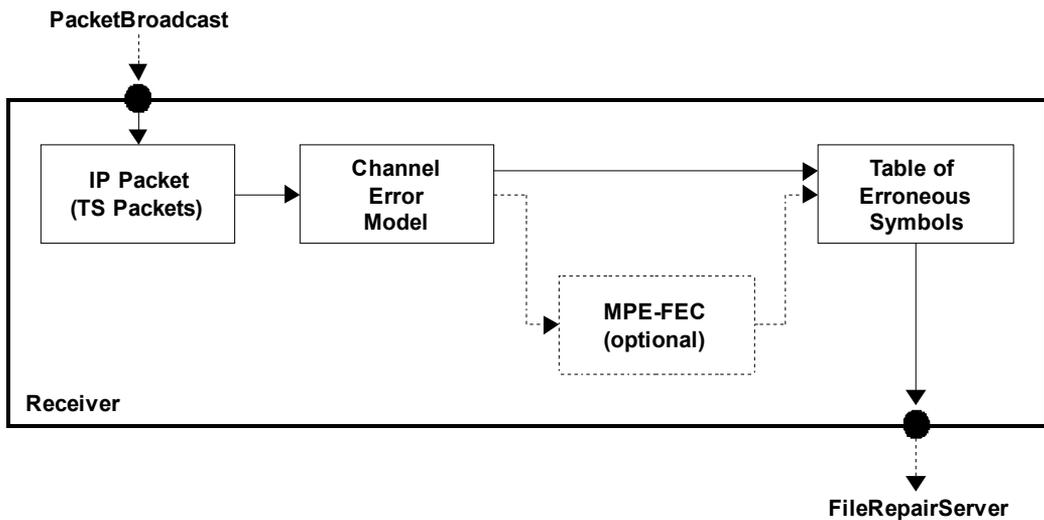


Fig. 10. Details of the component Receiver.

The following parameters can be set for the component *FileRepairServer*: number of receivers, cost for the transmission of a single byte via the ptp network, cost for the transmission of a single byte via the ptm network, offset time and maximum back-off time for receiving file repair requests after the transmission is finished, fraction of the maximum back-off time for the calculation of estimated repair costs (α), file size, average symbol size, average size of a file repair request, size of a ptm repair session announcement, estimated success rate for a receiver using a ptm repair session, and fraction of receivers without a ptp connection.

Error Model of the Transmission Channel

To simulate erroneous data segments, an error model had to be implemented into the transmission system of the simulation framework. The error model used is based on the so-called four-state run length model described in (Poikonen & Paavola, 2006) and (Poikonen & Paavola, 2005). This model operates on the resulting MPEG-2 transport stream of the DVB-T data transmission (see section 2.2) and produces streams of erroneous TS packets by using the four states *Good (short)*, *Bad (short)*, *Good (long)* and *Bad (long)*. If the model is in the states *Bad (short)* or *Bad (long)*, erroneous TS packets are generated, and if the model operates in the other two states, correct TS packets are produced.

According to the state diagram of this model (Poikonen & Paavola, 2006), the probabilities for remaining in a state denoted as *long* are very high. Therefore, whenever the error model switches over to one of these states, it will remain there for a rather long period of time compared to the states denoted as *short*, leading to long sequences of erroneous or error-free TS packets. With these settings, the model has shown to produce a good approximation to error streams measured using the COST 207 Six-tap Typical Urban (TU6) multi-path channel model (COST, 1989). See (Poikonen & Paavola, 2006) for a comparison.

Modified ptm Repair Cost Formula

For the simulations done within this study, a modified formula (3) for the calculation of the ptm repair costs (see section 3.2) was used. This formula focuses on the characteristic that

only distinctive erroneous symbols have to be retransmitted in a ptm repair session instead of the whole file as specified in the original formula for ptm sessions in (DVB Project, 2009b). Therefore, the parameter n_{dsym} denotes the number of distinctive requested symbols. For the calculation of the actual financial ptm repair costs, the estimated value for the number of repair requests (N_{req}) is replaced by the actual value denoted by n_{req} .

$$C_{ptm(expected)} = c_m \cdot s_{an} + c_u \cdot N_{req} \cdot s_{req} + c_m \cdot n_{dsym} \cdot s_{sym} \quad (3)$$

Simulation Parameters

For all simulations done within this study, the size of the transmitted object (file size) was 4 MB. This is for example the typical size of an MPEG-1 audio layer 3 (MP3) encoded sound file. MPE-FEC either was disabled or enabled, using 256, 512 or 1024 rows for the corresponding MPE-FEC frames. With these settings, it was possible to compare the effectiveness of different MPE-FEC settings with regard to the resulting number of repair requests and therefore with regard to the resulting transmission overheads and repair costs. The size of an encoding symbol either was 100 bytes, 500 bytes or 1400 bytes. This allowed for the comparison of the effects of very small, medium-sized, and large FLUTE packets on the effective repair costs. The cost for the transmission of a single byte via the ptm network was based on the pricing published by an Austrian DVB-H provider which bills EUR 39700 excluding 20% value added tax (VAT) per 100 kbps stream per year³ for nationwide coverage. This is equivalent to EUR $1.225 \cdot 10^{-7}$ per byte. The cost for the transmission of a single byte via the ptp network was based on the pricing published by an Austrian mobile network operator that bills EUR 10 including 20% VAT per 1 GB per month⁴ for nationwide coverage. This equates to EUR $1 \cdot 10^{-8}$ per byte. After one-third of the time window for repair requests, the calculation of the projected repair costs was done by the file repair server. Other parameters, e.g. concerning values for the DVB transmission mode and TS data rate were set to typical, technical parameters of current DVB-H trial services (DVB Project, 2009a) or according to examples in (DVB Project, 2009b) and (DVB Project, 2009c).

4.2 Simulation results

Figure 11 shows the financial repair costs for a transmitted file of size 4 MB without MPE-FEC for up to 1000 receivers. Whereas the costs for ptp sessions rise almost linearly with the number of receivers, the costs for ptm sessions quickly converge to a maximum. This maximum represents the correction of the whole file. These different behaviors are due to the fact that in a ptp repair session each requested symbol has to be transmitted to the requester individually, whereas in a ptm session only distinctive requested symbols are repeated, independent of the number of different requests for one and the same symbol.

Figure 12 shows the file repair efficiencies of the ptp and ptm repair sessions for up to 5000 receivers. The file repair efficiency is defined in (DVB Project, 2009b) as the number of receivers with successful recovery divided by the cost of the transmission of repair data. For ptp sessions, the number of receivers with successful reception is given by the total number of receivers, because it is assumed that all receivers will be successful. For ptm sessions, it is

³ <http://www.ors.at/bekanntmachungen/301.pdf> (18.06.2009)

⁴ <http://www.a1.net/privat/breitband> (18.06.2009)

assumed that there will be a fraction of the receivers that are still not able to recover the file after the first round of the ptm repair session. In the simulations done, it is assumed that 90 % of the receivers will be able to recover the file after the first round of the ptm session. Due to the fact that the ptm repair costs are independent of the number of receivers, ptm sessions result in higher efficiencies, especially when there are many receivers.

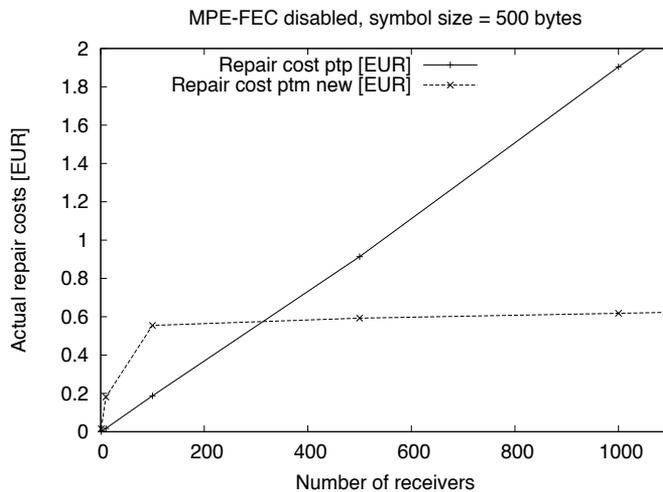


Fig. 11. File repair costs without MPE-FEC.

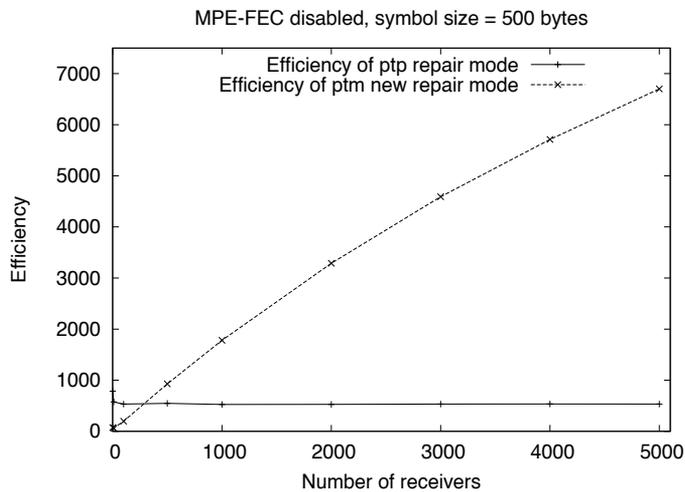


Fig. 12. Efficiencies of the different file repair methods.

Figure 13 compares the expected and actual file repair costs for ptp and ptm repair sessions without MPE-FEC. It can be seen that the expected costs are very close to the final costs, although the calculation of the expected costs was done at only one-third of the repair request window ($\alpha = 0.33$).

The results of the simulations which explore the impact of MPE-FEC for additional forward error correction are shown in figures 14 and 15. Figure 14 compares the resulting amount of necessary repair data for ptp repair sessions depending on the used MPE-FEC frame size. As can be seen clearly, the bigger the MPE-FEC frame size, the lower the amount of erroneous data. This is due to the fact that the probability of being able to recover a single IP packet is higher when the MPE-FEC consists of more rows, because the shorter the frame the higher the probability that an IP packet spans several columns.

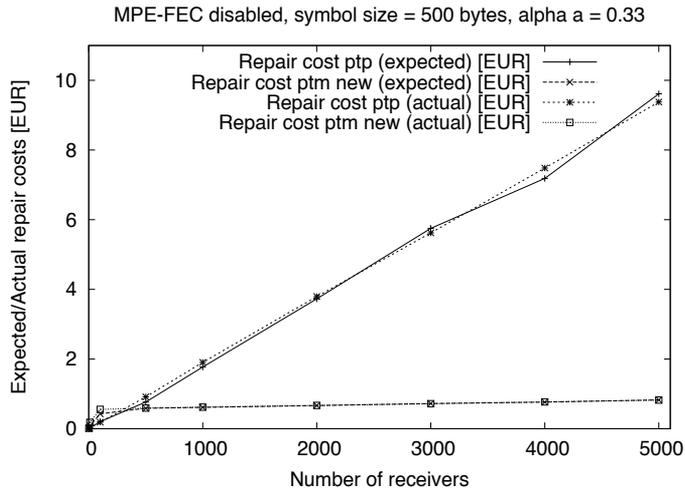


Fig. 13. Expected vs. actual file repair costs.

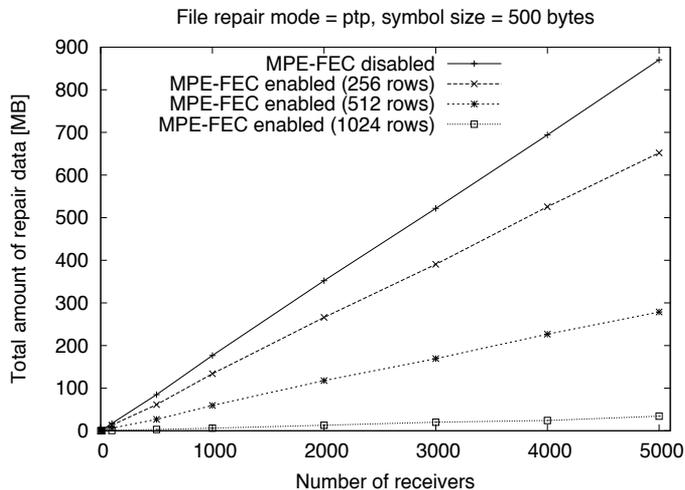


Fig. 14. Effects of MPE-FEC on ptp repair.

Concerning ptm sessions, the amount of repair data converges to the file size including protocol overheads. The bigger the used MPE-FEC frames, the more slowly this convergence happens (see figure 15).

Figure 16 compares the ptp and ptm repair costs for the two cases of disabled and enabled MPE-FEC. As can be seen, ptp can outperform ptm, when MPE-FEC is used with the maximum MPE-FEC frame size due to ptp's lower transmission cost per byte.

Figures 17 and 18 show the comparison of ptp and ptm repair costs using maximum MPE-FEC frame size and different symbol sizes. For ptp repairs, the usage of a symbol size of 500 bytes yields better results than using a bigger symbol size of 1400 bytes or a very small symbol size of only 100 bytes.

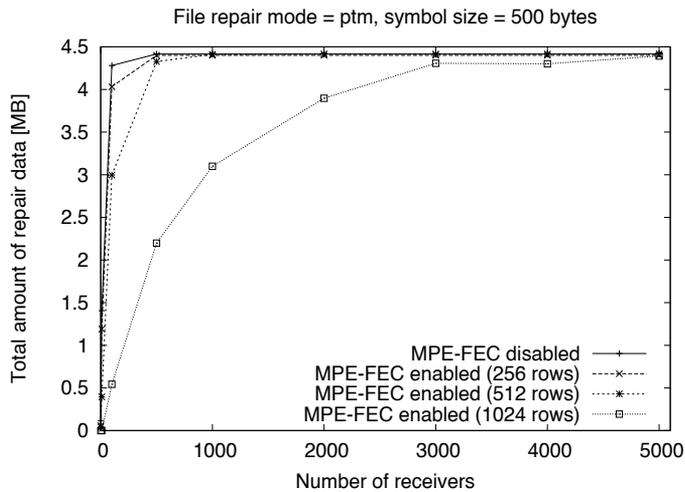


Fig. 15. Effects of MPE-FEC on ptm repair.

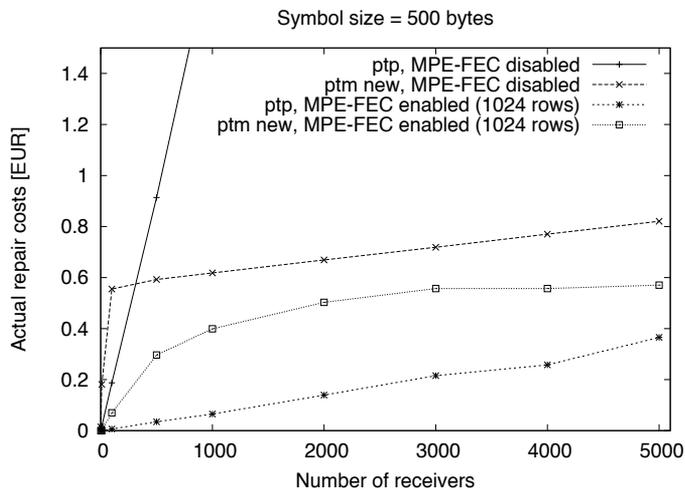


Fig. 16. File repair costs with/without MPE-FEC.

For large symbols, the probability of spanning several MPE-FEC frame columns is higher than for shorter symbols. As the number of bytes per row which can be corrected is limited by the used RS code, a symbol contained in only one column can more likely be corrected

than a symbol distributed over several columns. Very small symbols have a misbalanced ratio between protocol headers and payload. Each symbol is prefixed with 52 bytes of FLUTE/UDP/IP encapsulation overheads. Therefore, very small symbol sizes intrinsically result in increased repair costs per user data.

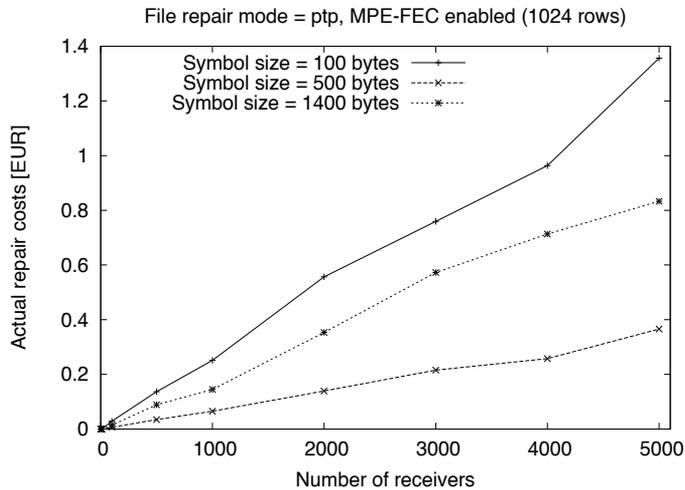


Fig. 17. Effects of encoding symbol size on ptp repair costs.

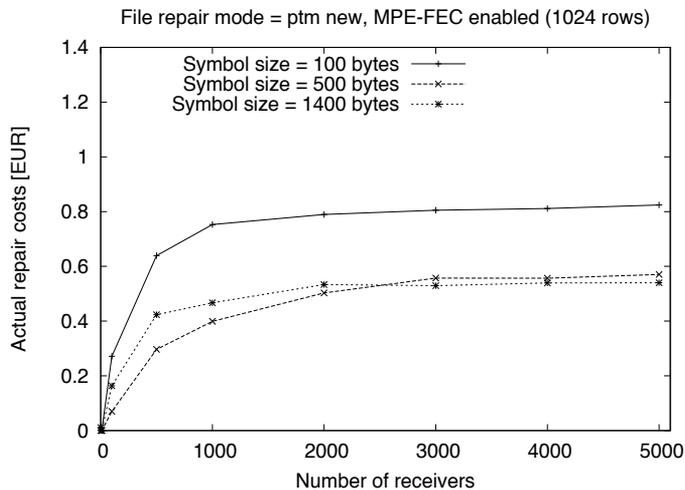


Fig. 18. Effects of encoding symbol size on ptm repair costs.

The results depicted in figure 18 dealing with ptm repairs show once again the negative effects of very small symbol sizes. As the amount of repair data for this repair mode converges to the file size, the symbol size with the most efficient payload/header ratio produces the lowest repair costs if a large number of receivers is considered. Summing up the results of both figures, the usage of medium-sized symbols is recommended.

5. Conclusions

IP datacast over DVB-H is one possible technical system to deliver mobile multimedia services, which are gaining more and more importance. This chapter gave an overview of the different digital video broadcasting transmission systems, including DVB-H and IP datacast. An IPDC/DVB-H infrastructure can provide different services, such as real-time streaming services for mobile TV or file delivery services for the distribution of arbitrary files to all receivers within the service area. Concerning the latter services, a file repair mechanism was specified for enabling the retransmission of data segments which were received erroneously at the receivers. In order to understand the technical details of the file repair mechanism and its technical context, the relevant building blocks and protocols were described in detail within the chapter.

A considerable part of the chapter deals with a simulation study to examine the financial costs of file repair sessions based on IPDC's file repair mechanism. This examination was done using a simulation framework, which incorporates the basic IPDC/DVB-H transmission and file repair procedures as well as a current error model concerning the DVB-H transmission channel. Simulations concerning the transmission of a file to a variable number of receivers with varying values for several critical parameters such as the use of MPE-FEC for error correction, the size of MPE-FEC frames, or the size of encoding symbols were executed. Three of the most important results of the simulations are as follows. First, the file repair costs of ptp repair sessions rise (almost) linearly with the number of receivers, whereas the file repair costs of ptm repair sessions rapidly converge to the costs of retransmitting the whole file. For a high number of receivers and without MPE-FEC for additional error correction, ptp repair costs are considerably higher than ptm repair costs. Second, MPE-FEC can drastically reduce the amount of repair data necessary for data recovery, especially concerning ptp repair sessions. The bigger the used MPE-FEC frame, the lower the rising of the file repair costs, especially concerning ptp repair sessions. The biggest MPE-FEC frame size (1024 rows) provides the best results. Depending on the number of receivers, the ptp repair mode can outperform the ptm repair mode, if MPE-FEC is used. Third, the size of the encoding symbols has a strong impact on the repair costs. Neither very big symbols nor very small symbols lead to optimal results. It is recommended to use medium-sized symbols of about 500 bytes.

6. References

- COST (1989). *Digital land mobile radio communications*, Commission of the European Communities
- DVB Project (2009a). *DVB-H: Global Mobile TV*, online, accessed 18th June 2009. Available at <http://dvb-h.org/>
- DVB Project (2009b). *IP Datacast over DVB-H: Content Delivery Protocols (CDP)*, DVB BlueBook A101 Rev.1. Available at <http://dvb-h.org/>
- DVB Project (2009c). *Digital Video Broadcasting (DVB); IP Datacast over DVB-H: Content Delivery Protocols (CDP) Implementation Guidelines*, DVB BlueBook A113 Rev.1. Available at <http://dvb-h.org/>

- DVB Project Office (2008a). *2nd Generation Satellite – DVB-S2 (DVB Fact Sheet)*. Available at http://www.dvb.org/technology/fact_sheets/
- DVB Project Office (2008b). *Internet Protocol Datacast – DVB-IPDC (DVB Fact Sheet)*. Available at http://www.dvb.org/technology/fact_sheets/
- DVB Project Office (2009a). *Introduction to the DVB Project (DVB Fact Sheet)*. Available at http://www.dvb.org/technology/fact_sheets/
- DVB Project Office (2009b). *2nd Generation Cable – DVB-C2 (DVB Fact Sheet)*. Available at http://www.dvb.org/technology/fact_sheets/
- DVB Project Office (2009c). *Digital Terrestrial Television – DVB-T (DVB Fact Sheet)*. Available at http://www.dvb.org/technology/fact_sheets/
- DVB Project Office (2009d). *2nd Generation Terrestrial – DVB-T2 (DVB Fact Sheet)*. Available at http://www.dvb.org/technology/fact_sheets/
- DVB Project Office (2009e). *Broadcasting to Handhelds – DVB-H (DVB Fact Sheet)*. Available at http://www.dvb.org/technology/fact_sheets/
- DVB Project Office (2009f). *Satellite Services to Handhelds – DVB-SH (DVB Fact Sheet)*. Available at http://www.dvb.org/technology/fact_sheets/
- ETSI (2004). *Digital Video Broadcasting (DVB); Transmission System for Handheld Terminals (DVB-H)*, ETSI EN 302 304 V1.1.1
- ETSI (2006a). *Digital Video Broadcasting (DVB); IP Datacast over DVB-H: Architecture*, ETSI TR 102 469 V1.1.1
- ETSI (2006b). *Digital Video Broadcasting (DVB); IP Datacast over DVB-H: Content Delivery Protocols*, ETSI TS 102 472 V1.2.1
- ETSI (2009a). *Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for digital terrestrial television*, ETSI EN 300 744 V1.6.1
- ETSI (2009b). *Digital Video Broadcasting (DVB); DVB-H Implementation Guidelines*, ETSI TR 102 377 V1.3.1
- Faria, G. et al. (2006). DVB-H: Digital Broadcast Services to Handheld Devices. *Proc. IEEE*, Vol. 94, No. 1, pp. 194-209
- Fielding, R. et al. (1999). *Hypertext Transfer Protocol -- HTTP/1.1*, IETF RFC 2616, The Internet Society
- ISO/IEC (2007). *Information technology -- Generic coding of moving pictures and associated audio information: Systems*, ISO/IEC 13818-1:2007
- Luby, M. et al. (2002). *Asynchron Layered Coding (ALC) Protocol Instantiation*, IETF RFC 3450, The Internet Society
- Paila, T. et al. (2004). *FLUTE - File Delivery over Unidirectional Transport*, IETF RFC 3926, The Internet Society
- Poikonen, J. & Paavola, J. (2005). *Comparison of Finite-State Models for Simulating the DVB-H Link Layer Performance*, Technical Report, University of Turku
- Poikonen, J. & Paavola, J. (2006). Error Models for the Transport Stream Packet Channel in the DVB-H Link Layer. In: *Proceedings of the IEEE International Conference on Communications 2006 (ICC '06)*, Vol. 4, pp. 1861-1866
- Postel, J. (1981). *Internet Protocol*, IETF RFC 791, USC/Information Sciences Institute

-
- Reimers, U. (2005). *DVB – The Family of International Standards for Digital Video Broadcasting*, Springer, Berlin Heidelberg New York
- Rivest, R. (1992). *The MD5 Message-Digest Algorithm*, IETF RFC 1321, MIT Laboratory for Computer Science and RSA Data Security, Inc.
- Schulzrinne, H. et al. (2003). *RTP: A Transport Protocol for Real-Time Applications*, IETF RFC 3550, The Internet Society
- Varga, A. (2009). *OMNeT++ Community Site*, online, accessed 18th June 2009. Available at: <http://www.omnetpp.org/>
- Watson, M. (2009). *Basic Forward Error Correction (FEC) Schemes*, IETF RFC 5445, IETF Trust

DVB-T2: New Signal Processing Algorithms for a Challenging Digital Video Broadcasting Standard

Mikel Mendicute, Iker Sobrón, Lorena Martínez and Pello Ochandiano
Signal Theory and Communications Area
Mondragon Goi Eskola Politeknikoa
University of Mondragon,
Spain

1. Introduction

Digital Video Broadcasting-Terrestrial (DVB-T) is the most widely deployed digital terrestrial television system worldwide with services on air in over thirty countries. In order to increase its spectral efficiency and to enable new services the DVB consortium has developed a new standard named DVB-T2. A nearly definitive specification has already been published as a BlueBook as well as an implementation guideline, where the structure and main technical novelties of the standard have been defined. The imminent publication of the final DVB-T2 standard will give rise to the deployment of new networks and commercial products.

The differences between the original DVB-T and the new DVB-T2 standards are many and important. The latest coding, interleaving and modulation techniques have been included in this large and flexible specification to provide capacity and robustness in the terrestrial transmission environment to fixed, portable and mobile terminals. Multiple-input multiple-output (MIMO) techniques, low-density parity-check codes (LDPC), rotated constellations, new pilot patterns or large interleaving schemes are the most remarkable signal processing algorithms that have been included to overcome the limitations of the much simpler DVB-T broadcasting standard.

This chapter focuses on the mentioned new algorithms and the opportunities that arise from a signal processing perspective. New transmission and reception techniques are proposed which can be used to enhance the performance of DVB-T2, such as iterative demapping and decoding, new antenna diversity schemes or more efficient channel estimation algorithms. Furthermore, the performance of the new standard is analyzed and evaluated through simulations focusing on the aforementioned algorithms. The behaviour of the standard is specially studied in single-frequency networks (SFN), where the vulnerability of the former standard is prohibitive when destructive interferences arise.

The chapter first describes the main architecture and limitations of the original DVB-T specification. The physical layer of the new DVB-T2 standard is then defined, emphasizing the main differences in comparison to its predecessor. The next section of the chapter proposes and analyzes iterative demapping and decoding techniques at reception which can profit from the benefits of the new LDPC codes. Multi-antenna transmission and reception is

next studied, evaluating the benefits of the antenna diversity schemes proposed by the standard on the performance of the system. Channel estimation issues are analyzed in the following section, presenting a bidimensional estimation algorithm which is specially interesting due to the mobility requirements of the new standard and to the plethora of pilot patterns that have been defined. Last, two relevant issues of the new standard are analyzed and evaluated through simulations: the rotated constellation-based transmission and the performance in SFN scenarios. Provided results show the behavior of the new DVB-T2 standard and the improvement achievable by applying the new signal processing algorithms proposed throughout this chapter.

2. DVB-T and its limitations

The DVB-T terrestrial digital video broadcasting standard (ETSI, 1997) is replacing the former analogue systems in many countries around the world. The benefits of digital coding and transmission techniques allow perfect signal recovery in all the serviced areas avoiding the effects of the wireless channel and noise. Considering the physical level of the communications, the digital data sequences, which contain MPEG video, audio and other information streams, are transmitted using coded orthogonal frequency division multiplexing (COFDM) modulation. The information bits are coded, interleaved, mapped to a quadrature amplitude modulation (QAM) constellation and grouped into blocks. All the symbols in a block are transmitted simultaneously at different frequency subcarriers using an inverse fast Fourier transform (IFFT) operation. The number of IFFT points, which can be either 2048 (2K) or 8192 (8K), determines the transmission mode and the number of the available subcarriers in the transmission bandwidth. Some of these subcarriers are not used to allow for guard frequency bands whereas others are reserved for pilot symbols, which are necessary to acquire the channel information required for signal recovery.

Fig. 1 shows the main diagram of a DVB-T transmitter. As it can be seen, the data bit stream is scrambled, processed by an outer Reed-Solomon (RS) coder, an interleaver and an inner convolutional coder. The first coding stage removes possible error floors at high signal-to-noise (SNR) values, whereas the second reduces the bit error rate (BER) at the receiver by including more redundant information depending on the selected coding rate (CR), which can range from 1/2 to 5/6. The coded information bits are interleaved again in order to allocate consecutive bits to different subcarriers. The resulting information bits are then arranged by blocks, mapped and modulated using OFDM, which involves an IFFT operation and the addition of a cyclic prefix to enable a guard interval (GI) that avoids interference between consecutive blocks. The use of coding and interleaving processes over OFDM provides an efficient and robust transmission method in multipath scenarios enabling time and frequency diversity.

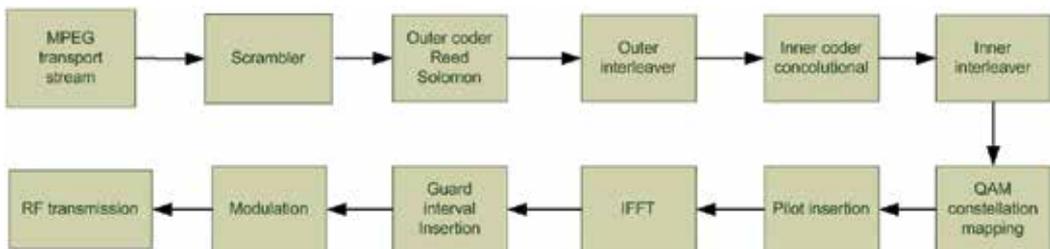


Fig 1. Elementary transmission chain of DVB-T.

Despite the many benefits achieved by the deployment of the DVB-T network, its limitations became clear from the beginning. First, the number and bit rates of the transmitted channels are limited in comparison with new wireless transmission techniques. A new standard was soon required to broadcast more channels and high-definition television (HDTV) using the same frequency spectrum. Second, a new information system was required to allow more interaction with the user. Third, the DVB-T standard, which had been designed for fixed scenarios, had a very bad performance in mobile or portable environments, so it could not be properly implemented in scenarios such as moving vehicles. Last but not least, the deployment of the DVB-T network has been and still is a true nightmare in SFN scenarios, where interferences between repeaters, which transmit the same information on the same frequency bands, may destroy the received signal avoiding its reception in areas with good reception levels.

Considering the new advances in signal processing, modulation and coding, the DVB consortium has published a draft standard named DVB-T2 aiming to extend the capabilities of the aforementioned DVB-T standard.

3. The new DVB-T2 standard

Based on recent research results and a set of commercial requirements, the DVB consortium concluded that there were suitable technologies which could provide increased capacity and robustness in the terrestrial environment, mainly for HDTV transmission. Therefore, a new standard named DVB-T2 has been designed primarily for fixed receptors, although it must allow for some mobility, with the same spectrum characteristics as DVB-T. Fig. 2 shows the main stages of a DVB-T2 transmitter, where dashed lines represent optional stages.

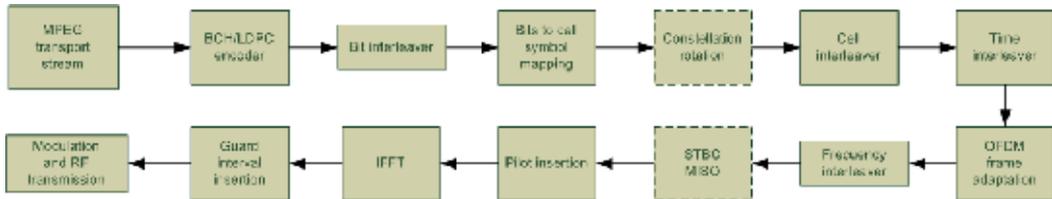


Fig. 2. Elementary transmission chain of DVB-T2.

The first remarkable novelty lies on the error correction strategy, since DVB-T2 uses the same channel codes that were designed for DVB-S2. The coding algorithms, based on the combination of LDPC and Bose-Chaudhuri-Hocquenghem (BCH) codes, offer excellent performance resulting in a very robust signal reception. LDPC-based forward error correction (FEC) techniques can offer a significant improvement compared with the convolutional error correcting scheme used in DVB-T.

Regarding the modulation, DVB-T2 uses the same OFDM technique as DVB-T. Maintaining the 2K and 8K modes, the new standard has introduced longer symbols with 16K and 32K carriers in order to increase the length of the guard interval without decreasing the spectral efficiency of the system. The new specification offers a large set of modulation parameters by combining different numbers of carriers and guard interval lengths, making it a very flexible standard as it is shown in Table 1. Furthermore, the highest constellation size has been increased to 256 symbols (256QAM).

As it will be extended in Section 6, another interesting innovation is the introduction of 8 different scattered pilot patterns, whose election depends on the parameters of the current

transmission. Thus, thanks to all the configurable parameters of the new standard, the modulation can be adapted to the characteristics of the actual transmission, making the most of the spectral efficiency. As it can be seen in Fig. 2, an important innovative feature proposed by the DVB-T2 specification is the use of three cascaded forms of interleaving, which are the following: bit interleaver, time interleaver and frequency interleaver. The aim of all these interleaving stages is to avoid error bursts, giving rise to a random pattern of errors within each LDPC FEC frame.

| | DVB-T | DVB-T2 |
|------------------|---|---|
| FEC | Convolutional + Reed-Solomon 1/2, 2/3, 3/4, 5/6, 7/8 | LDPC + BCH 1/2, 3/5, 2/3, 3/4, 4/5, 5/6 |
| Modes | QPSK, 16QAM, 64QAM | QPSK, 16QAM, 64QAM, 256QAM |
| Guard intervals | 1/4, 1/8, 1/16, 1/32 | 1/4, 19/256, 1/8, 19/128, 1/16, 1/32, 1/128 |
| FFT size | 2K, 8K | 1K, 2K, 4K, 8K, 16K, 32K |
| Scattered pilots | 8% of total | 1%, 2%, 4% and 8% of total |
| Continual pilots | 2.6% of total | 0.35% of total |

Table 1. Available modes in DVB-T and DVB-T2.

On the other hand, a new technique called rotated constellations and Q-delay (RQD) is provided as an option, which comes to offer additional robustness and diversity in challenging terrestrial broadcasting scenarios. Furthermore, a mechanism has been introduced to separately adjust the robustness of each delivered service within a channel in order to meet the required reception conditions (in-door antenna/roof-top antenna, etc.). DVB-T2 also specifies a transmitter diversity method, known as Alamouti coding, which improves coverage in small scale single-frequency networks.

Finally, the DVB-T2 standard takes into account one of the main drawbacks of OFDM, the peak to average power ratio (PAPR) of the signal and its effects on the transmitter equipments. High power peaks are usually generated by OFDM transmission leading to distortions at the amplifiers, thus minimizing their efficiency. Two techniques have been included in the standard to limit the PAPR without degrading the transmitted signal: carrier reservation and active constellation extension. The first reserves some subcarriers that can be used to correct the PAPR level of the transmitted signal whereas the latter achieves the same effects modifying the QAM constellation without degrading the signal recovery at reception.

Fig. 3 shows the comparative performance of DVB-T and DVB-T2 for similar communication parameters. The BER at the output of the inner decoder has been considered in all the simulation results provided in this chapter. In order to allow a fair comparison of both standards, a quasi error free (QEF) of $BER=2 \cdot 10^{-4}$ and $BER=10^{-7}$ must be considered for DVB-T and DVB-T2 after convolutional and LDPC decoders, respectively. If these QEF reference values are analyzed, a gain of 6 dB can be established between the two standards in an additive white Gaussian noise (AWGN) channel model and nearly 4 dB in a Rayleigh channel. The code rates have been selected in order to approach equivalent systems.

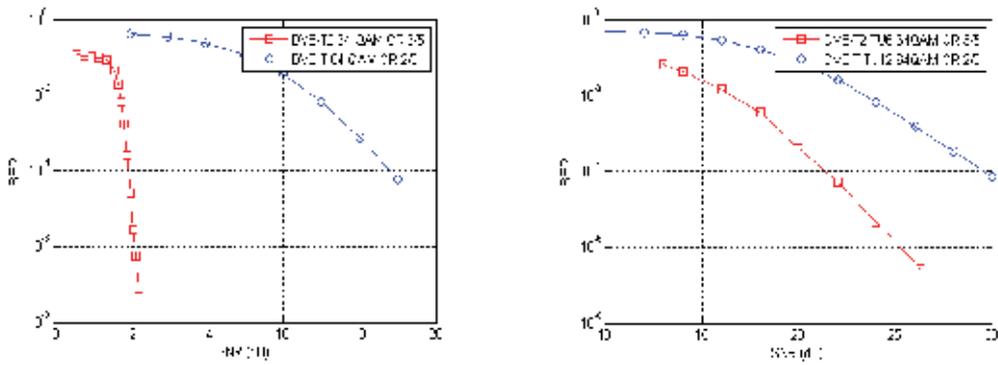


Fig. 3. BER performance of DVB-T and DVB-T2 systems in AWGN (a) and Rayleigh (b) channels.

All the DVB-T2 simulation results presented in this chapter have been obtained using the following transmission parameters: FEC frame length of 16200 symbols; 2K OFDM mode and a guard interval of 1/4.

4. Iterative demapping and decoding of LDPC codes

As it has been stated, one of the major innovations of DVB-T2 lies on the selected channel coding techniques. The coding schemes used in first-generation digital television standards (Reed Solomon and a convolutional code for outer and inner coding, respectively) have been replaced by LDPC and BCH codes in the second generation of the digital television standards published to date, such as DVB-S2 and DVB-T2. The main advantage of LDPC codes is that they provide a performance which approaches the channel capacity for many different scenarios, as well as the linear algorithms that can be used for decoding. Actually, the efficiency improvement provided by DVB-T2 in comparison with DVB-T is mainly based on these new coding and interleaving schemes.

4.1 Basics of BICM schemes and SISO processing

LDPC codes are commonly decoded by a soft-input soft-output (SISO) algorithm which iteratively computes the distributions of variables in graph-based models. It has been published under different names and models, such as the sum-product algorithm (SPA), the belief propagation algorithm (BPA) or the message-passing algorithm (MPA). The decoding of the information bits is based on the computation of the a posteriori probability (APP) of a given bit in the transmitted codeword $c = [c_0 c_1 \dots c_{n-1}]$ subject to the received symbol vector $y = [y_0 y_1 \dots y_{n-1}]$. Therefore, the APP ratio must be computed. A numerically more stable version called log-likelihood ratio (LLR) is commonly used as defined in the following equation:

$$a_l = \ln \left[\frac{P(c_l = 1)}{P(c_l = 0)} \right] \tag{1}$$

where $P(c_l=x)$ denotes the probability of codeword bit c_l having the value x .

DVB-T2 implements a bit-interleaved LDPC coded modulation (BILCM) scheme, which has been employed in many broadcasting and communication systems. BILCM is a special case of a more general architecture named bit-interleaved coded modulation (BICM). It was proposed by Zehavi (Zehavi, 1992) and consists of coding, bit-wise interleaving and constellation mapping. Several studies have shown that BICM presents an excellent performance under fading channels (Li et al., 1998).

The capacity of BICM schemes depends on several design parameters. An information theory point of view is given in (Caire et al., 1998) for input signals constrained by a specific complex constellation χ . Considering the simplest discrete-time memoryless complex AWGN channel modelled as $y = x + n$, where y , x and n denote the output value, the input sample and the complex Gaussian noise sample with zero mean and variance $N_0/2$ for each real and imaginary part respectively, and being N_0 the noise spectral power density. The channel capacity can be evaluated as follows for an m -order modulation in case of a coded modulation (CM) system.

$$C = m - E_{x,y} \left[\log_2 \frac{\sum_{z \in \chi} p(y|z)}{p(y|x)} \right] \quad (2)$$

However, in case of applying a BICM system, the channel capacity is always lower because each bit level is demapped independently.

$$C' = m - \sum_{i=1}^m E_{b,y} \left[\log_2 \frac{\sum_{z \in \chi} p(y|z)}{\sum_{z \in \chi_i^{(b)}} p(y|z)} \right] \quad (3)$$

where $\chi_i^{(b)}$ denotes the subset of χ whose corresponding i -th bit value is $b \in \{0,1\}$. Therefore, BICM is a sub-optimal scheme since $C \geq C'$.

4.2 Iterative demapping and decoding of LDPC codes for DVB-T2 receivers

This section describes the application of novel iterative demapping and decoding algorithms over BILCM for DVB-T2 receivers. This iterative receiver scheme, named BILCM with iterative demapping (BILCM-ID) was firstly described in (Li et al., 1998) and (Li et al., 2002), where it has been shown that BICM schemes are sub-optimal from an information theoretical point of view. Nevertheless, BICM-ID schemes are optimal because, although the bit-levels are not demapped independently, they are fed back to assist demapping other bits within the same symbol.

The BILCM-ID model is represented by the block diagram of Fig. 4. Soft information given by SISO blocks, such as the LDPC decoder or the soft demapper, is usually fed back from one block to another. In the research described in this paper, the demapping process is fed with soft values from the SISO LDPC decoder, exchanging information iteratively between the two blocks. The soft demapper uses the extrinsic information generated by the LDPC decoder as a priori information for the demapping process.

In general, the complex received signal at symbol index j , r_j , can be expressed as $r_j = h_j s_j + n_j$. The demapping stage consists of two stages. First, the soft demapper computes m a posteriori probabilities (one for each point of the modulated constellation) for every symbol received from the channel as it is shown in the following equation:

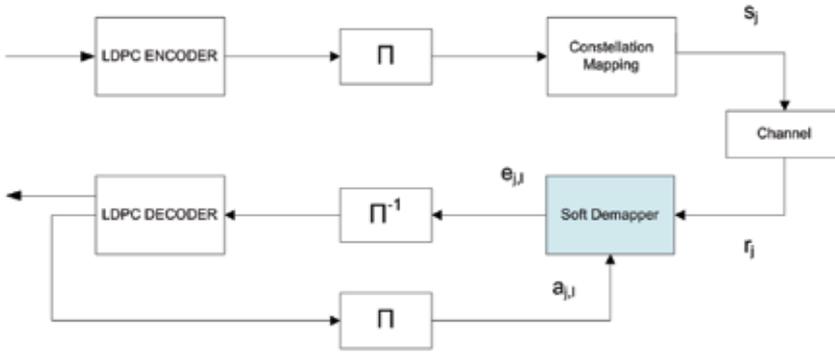


Fig. 4. Transmitter and receiver modules of a BILCM-ID system.

$$f(y | s_j) = \exp \left\{ \frac{-|r_j - h_j s_j|^2}{N_0} \right\} \quad (4)$$

In the second step, extrinsic LLRs $e_{j,l}$, where $l = [1, \dots, m]$, are calculated as follows:

$$e_{j,l} = \frac{\sum_{s_j \in \chi^1} \exp \left\{ \frac{-|r_j - h_j s_j|^2}{N_0} + \sum_{i=1, i \neq l}^m s_{j,i} a_{j,i} \right\}}{\sum_{s_j \in \chi^0} \exp \left\{ \frac{-|r_j - h_j s_j|^2}{N_0} + \sum_{i=1, i \neq l}^m s_{j,i} a_{j,i} \right\}} \quad (5)$$

where χ^t denotes the signals $s_j \in \chi$ whose l^{th} bit has the value $t \in \{0,1\}$. As it is shown in Equation (5), the use of a priori information tries to enhance the reliability of symbol probabilities.

4.3 Simulation results

The simulation-based BER performance of iterative demapping in DVB-T2 receivers is analyzed in this section for different modulation orders, code rates and channel models. Bit-interleaving is restricted within one LDPC FEC codeword as defined by the standard. As can be seen in Fig. 5a and 5b, iterative processing always provides a gain in comparison to a single demapping and decoding stage. Nevertheless, it can be seen that the performance of BILCM-ID systems has a strong dependence on the code rate and the modulation order. The rationale behind this dependence is that the capacity gap between CM and BICM systems decreases as the coding rate grows, whereas it grows with the constellation order. Furthermore, several studies have proved that Gray mapping makes such gap negligible at high coding rates.

Fig 5a depicts the simulation results for code rate 1/2 using 16QAM and 64QAM constellations over an AWGN channel. On the other hand, Fig. 5b shows results for different code rates and a 64QAM constellation over a Typical Urban 6 (TU6) channel (COST207, 1989). Blue lines correspond to standard reception without iterative demapping, whereas red lines represent 3 demapping-decoding iterations. For all the simulation results provided

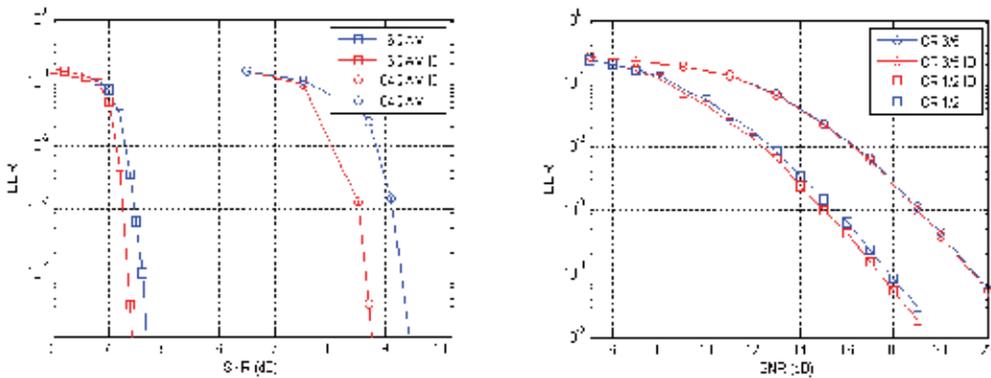


Fig. 5. BER versus SNR performance curves for iterative and non-iterative demapping and decoding in AWGN (a) and TU6 (b) channels.

in this chapter, the LDPC decoder runs a maximum of 50 internal iterations for each detection stage.

Simulation results confirm the values expected from the information theoretical analysis described previously: the gain provided by iterative demapping is insignificant for high coding rates and increases as the constellation order grows. It can be seen in Fig. 5a that there is a gain of 0.25 dB with 16QAM and 0.6 dB with 64QAM at a BER of 10^{-4} .

As has been said, the feedback to the demapper is usually performed when the decoding process has finished after 50 iterations. However, the number of decoder iterations in each demapping stage can be modified in order to offer the best error correcting performance and keep the same maximum number of overall iterations (50 decoder iterations \times 3 demapping stages). This new design is called irregular iterative demapping (ID-I) and is specially interesting to design efficient iterative demapping receivers.

Fig. 6a describes the performance of the LDPC decoder for 3 demapping iterations (64QAM over AWGN channel) at a specific SNR value of 8.75 dB, both for the regular case and the irregular one. The implemented irregular demapping approach performs 25, 75 and 50 decoding iterations at the first, second and third demapping stages, respectively. Regarding the regular case, it can be seen that the LDPC decoder converges at the first demapping

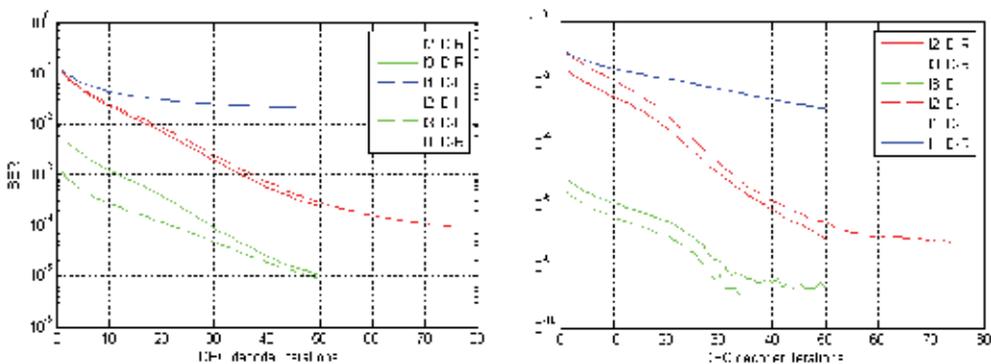


Fig. 6. BER versus LDPC decoder iterations for regular and irregular iterative demapping strategies at SNR value of 8.75 dB (a) and at SNR value of 9.1 dB (b).

iteration, which means that the BER performance will not improve if more decoder iterations are carried out. Moreover, the BER gain reached at the first demapping iteration is not too high in comparison with the second one which has not converged at the end of the decoding process, which means it could be extended up to a point close to the convergence value.

On the other hand, the irregular approach performs more decoder iterations in the second demapping stage and consequently the error correcting performance can be improved at the end of the decoding process. Fig. 6b demonstrates that irregular iterative demapping gain is even larger in higher SNR values. For example, at a BER value of about 10^{-9} , the third demapping iteration has converged in the 35th decoder iteration for the regular case, whereas the irregular one has not.

5. MISO transmission and receiver diversity

MIMO wireless communication systems are based on signal processing with multiple antennas at both transmitter and receiver side. Theoretical works such as (Foschini & Gans, 1998) and (Telatar, 1999) have shown that the use of multiple antennas can increase the limits of the channel capacity. Wireless telecommunications systems such as WLAN 802.11n or WMAN 802.16e have included MIMO techniques in their newest specifications. However, the new DVB-T2 standard only proposes the use of several antennas at one side of the transmission. These subsets of MIMO systems are called multiple-input single-output (MISO) and single-input multiple-output (SIMO) schemes. The first, which corresponds to multiple transmit and only one receive antenna, offers transmit diversity, whereas the latter, which includes multiple receive antennas, offers receive diversity.

5.1 The DVB-T2 MISO transmission scheme

The DVB-T2 standard describes a transmit diversity method with two antennas based on a modified Alamouti coding scheme (Alamouti, 1998). The coding algorithm is generically called space–frequency block coding (SFBC) since the Alamouti scheme is used in spatial and frequency domain as is depicted in Fig. 7. As can be seen, the Alamouti SFBC approach processes the symbols pairwise, sending the original values ($[a_0, b_0]$ for the first symbol pair) at one of the antennas and modified values ($[-b_0^*, a_0^*]$) at the other one, thus increasing the transmit diversity while keeping the symbol rate.

The received complex values for the first pair of MISO cells are given by:

$$\begin{aligned} R_1 &= H_1 a_0 - H_2 b_0^* + N_1 \\ R_2 &= H_1 b_0 + H_2 a_0^* + N_2 \end{aligned} \quad (6)$$

where H_1 and H_2 denote the channel transfer gains from transmitters 1 and 2 to the receiver, while N_1 and N_2 are the noise samples. These equations can be represented in matrix notation as follows:

$$\begin{pmatrix} R_1 \\ R_2^* \end{pmatrix} = \begin{pmatrix} H_1 & -H_2 \\ H_2^* & H_1^* \end{pmatrix} \begin{pmatrix} a_0 \\ b_0^* \end{pmatrix} + \begin{pmatrix} N_1 \\ N_2^* \end{pmatrix} \quad (7)$$

which makes the decoding process simpler at the receiver. This method aims to improve the coverage and robustness of the reception in SFN networks, so that the transmitters of two

SFN cells form a “distributed” MISO system, providing space and frequency diversity. Fig. 8 shows the BER performance of the MISO DVB-T2 system in comparison to the SISO system with a TU6 channel model of six paths (COST207, 1989). This channel corresponds to a multipath propagation scenario with Rayleigh fading (Patzold, 2002). The diversity gain for a 64QAM mode with LDPC code rate 3/5 is around 5 dB just above the QEF value of 10^{-7} after LDPC decoder as is specified in the DVB-T2 implementation guidelines (DVB, 2009).

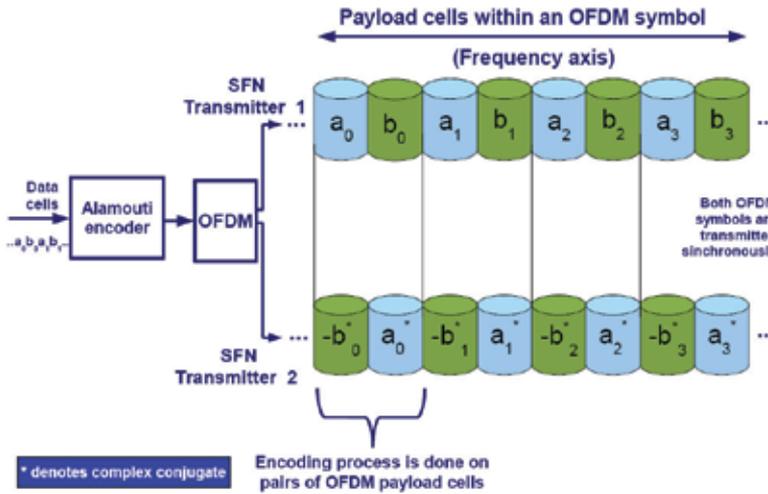


Fig. 7. MISO encoding in DVB-T2 systems.

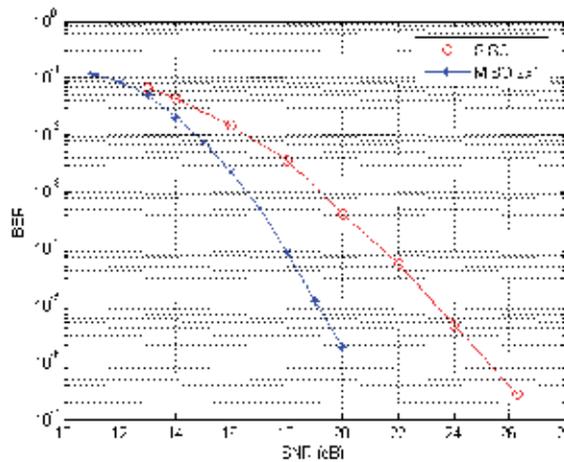


Fig. 8. Performance of MISO transmission for a 64QAM constellation and CR 3/5 over TU6 channel in DVB-T2.

5.2 Effects of receive diversity

The DVB-T2 standard only includes requirements for transmission, so the signal processing at the receiver can be freely modified to improve the performance of the overall system.

There are several receiver techniques that can be applied, such as iterative and non iterative algorithms of equalization and decoding (Proakis, 1995) or receive diversity techniques (Jakes, 1974). This section will focus on a simple but effective receive diversity technique called maximal ratio combining (MRC) which performs a weighted linear combination of the input signals. The signal recovery process can be represented by the following equation:

$$R_k = \sum_{i=1}^N H_{k,i}^* R_{k,i}, \tag{8}$$

where $R_{k,i}$ and $H_{k,i}^*$ are the received signal and the complex conjugate of the channel transfer function between the transmitter and the i -th receiver antenna for subcarrier index k . This receiver diversity method maximizes the output SNR and has been widely studied for DVB-T in portable and mobile scenarios (Levy, 2004). Since DVB-T2 is targeted at fixed, portable and mobile scenarios, MRC results a suitable technique to improve the reception quality. Furthermore, it can also be combined with the aforementioned MISO transmission scheme specified in DVB-T2, hence forming a 2x2 setup.

Fig. 9 depicts the performance of such a 2x2 transmitter and receiver diversity system in a TU6 channel with the former configuration. An improvement of 6 dB can be observed at the QEF level in comparison to the 2x1 MISO system, which is due to antenna array and diversity gains at the receiver. Nevertheless, it involves a greater cost than MISO as part of the receiver chain must be replicated, which can be expensive for consumer products. Consequently, the receive diversity may be targeted to specific equipments, such as mobile or portable receivers and problematic fixed locations.

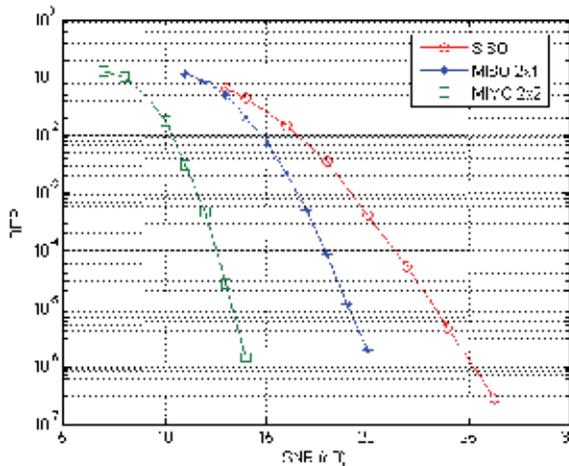


Fig. 9. Performance of the MIMO transmission in DVB-T2.

6. Channel estimation and tracking

As has been detailed in previous sections, one of the main innovative aspects of DVB-T2 is the plethora of pilot types and patterns provided, which make channel estimation more flexible. This section describes the pilot structure in more detail and proposes the application of an effective and well-known channel estimation algorithm which can profit from the flexibility and the information provided by all the available pilot subcarriers.

6.1 Pilot symbol locations and sets in DVB-T2

As it has been summarized in Sections 2 and 3, an OFDM frame is composed of a block of symbols whose length depends on the selected transmission mode. For the case of DVB-T, two modes are defined, 2K and 8K, which use 1705 and 6817 of the available subcarriers, respectively. However, only 1512 and 6048 of the subcarriers correspond with information symbols. The rest of subcarriers, named pilots, are commonly used for synchronization and channel estimation. These pilot subcarriers are divided into three groups: continual (which are available at every symbol), scattered (whose location varies for different set of symbols) and transmission parameter signaling (TPS), which are used to acquire information about the format of the signal.

DVB-T2 introduces some remarkable differences in this respect, as it incorporates new reference signals. Several pilot groups have been defined: continual, scattered, edge, P2 and frame-closing pilot cells. The location and amplitudes of the pilots differ with respect to DVB-T and various sets of pilots are applied depending on the transmission parameters. There are eight different scattered pilot patterns (named from PP1 to PP8). Therefore, depending on the FFT size, the length of the guard interval, pilot pattern and the antenna settings, the number of pilot and overall used subcarriers can vary. Furthermore, the continual pilot locations are taken from one or more sets depending on the FFT mode and the pilot positions belonging to each set depend on the scattered pilot pattern in use. Fig. 10 shows the locations of scattered and other pilots for a sample pilot pattern named PP1.

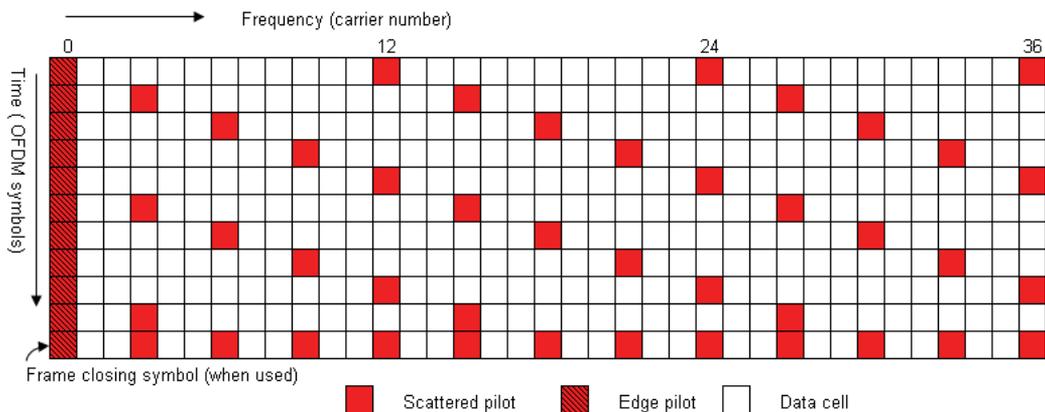


Fig. 10. Sample scattered pilot pattern (PP1) defined by the DVB-T2 standard for SISO transmission.

The edge pilots are inserted in order to allow frequency interpolation up to the edge of the spectrum. P2 pilots are inserted in P2 symbols, which may carry data and are used to transmit signaling information. As can be stated from the previous enumeration, the channel estimator must be a very flexible structure in order to allow for all the possible options and to profit from the information provided by all the pilot symbols. Channel estimation can be divided into two stages: estimation at the pilot subcarriers and interpolation of the intermediate subcarrier gains, both in the frequency (rest of subcarriers) and the time domain (surrounding symbols). Bidimensional Wiener filter-based approaches, such as the one that is described in this section, can perform both stages at once.

6.2 Two dimensional channel estimation in DVB-T2

An efficient and adaptive estimate of the channel can be obtained if the channel estimator processes the information in both time and frequency domain, profiting from the coherence time and bandwidth of the channel, respectively. Such an algorithm based on pilot subcarriers and Wiener filtering is studied in (Necker & Stüber, 2004) for an OFDM system, showing a good trade-off between low complexity and speed of convergence. The structure of a two-dimensional estimator is too complex for practical implementation but it is possible to replace a 2D-filter by two 1D-filters, one running over the OFDM symbols (time domain) and the other over the subcarriers of an OFDM symbol (frequency domain), to simplify the implementation.

Given the Wiener design criteria outlined in (Hoeher et al., 1997), the filter coefficients are calculated through the cross-correlation of each of the carriers with the pilots and through the autocorrelation of the pilot carriers. This way, filter weights are generated from the statistics of the channel and the position of the pilots. Fig. 11 shows the performance of this channel estimation algorithm in a DVB-T2 scenario. Fig. 11a shows the value of the channel estimate using several different channel estimation and interpolation algorithms.

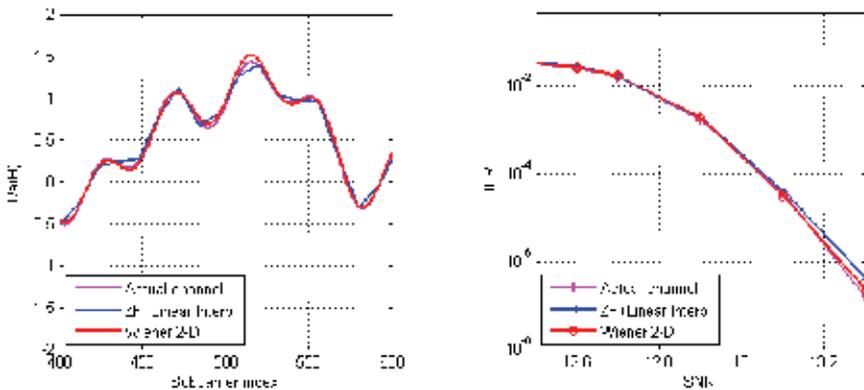


Fig. 11. Simulation results showing the performance of different channel estimation and interpolation schemes: channel value (a) and BER curve (b).

As can be seen in Fig. 11, the Wiener-based bidimensional channel estimator is the one that gets closest to the actual channel, whereas the simplest algorithm, based on zero-forcing and linear interpolation in the frequency-domain, shows to be the worst with a very poor performance. On the other hand, Fig. 11b shows the BER curves of the different estimation approaches for a static TU6 channel. The differences in these results depend on the performance of the channel interpolation method.

7. Effects of constellation rotation on DVB-T2 reception

The constellation rotation or RQD operation specified in DVB-T2 aims to increase the diversity order of the DVB-T2 BICM scheme. This technique is comprised of two stages: correlating the in-phase (I) and quadrature (Q) components of the transmitted signal through the rotation of the QAM constellation and making these components fade independently by means of a cyclic delay of the Q component. The rotation of the

constellation allows casting every constellation symbol over I and Q axis independently in such a way that the I component contains intrinsic information of the Q component and vice-versa. The cyclic delay of the Q component makes the I and Q components fade independently using a simplified approach of the signal space diversity (SSD) (Al-Semari & Fuja, 1997). The insertion of a simple time delay for one of the two components avoids the loss of I and Q information due to a same fading event.

Fig. 12 shows the aforementioned process. Symbols are first mapped onto the constellation (white points) and then rotated (black points) to correlate the I and Q components. The resulting symbols are fed to a FEC block, after which the imaginary part is delayed one cell.

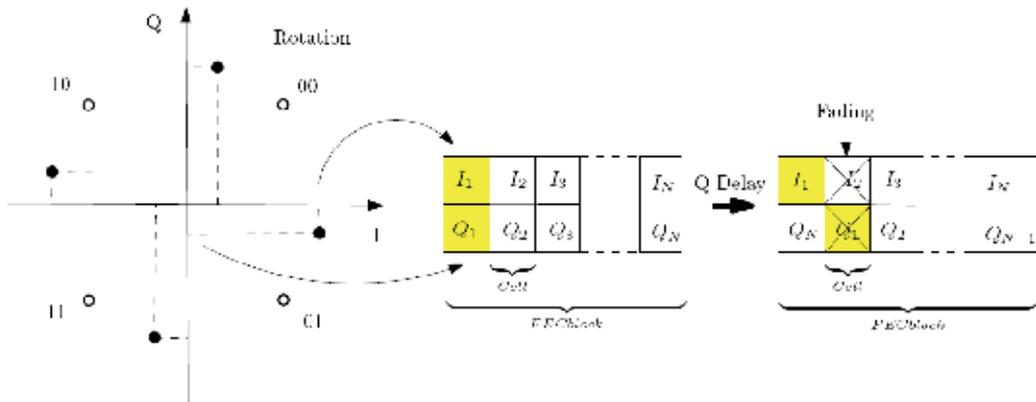


Fig. 12. Diagram of the rotated constellation and the Q delay process.

7.1 Analysis of RQD in the DVB-T2 BICM scheme over flat fading Rayleigh channel with erasures

This section only considers the BICM module of DVB-T2 to study the effects of RQD on the performance of the overall system. In order to model a simple transmission with fading events, the approach to a flat Rayleigh channel with erasures (RME) based on (Nour & Douillard, 2008) is assumed. The channel samples are considered uncorrelated due to all the interleaving modules of the DVB-T2 system. Hence, the equivalent received symbol Y at the discrete instant t is given by:

$$Y(t) = H(t)E(t)X(t) + N(t) \tag{9}$$

where $X(t)$ is the complex discrete transmit symbol, $H(t)$ is a Rayleigh distributed fading channel coefficient with $E[H(t)]=1$, $E(t)$ is a uniform random process which takes the value of zero with probability P_E and $N(t)$ is the AWGN sample at time index t . The block diagram of Fig. 13 shows the simplified DVB-T2 BICM system over the RME channel

Fig. 14 presents the BER results of the DVB-T2 BICM system over a RME channel with 20% of erasures or lost subcarriers for two different code rates: 3/4 and 2/3. For the first case, the LDPC cannot recover the lost information due to erasures and the performance of the system without RQD tends to an error floor above the QEF limit. However, the diversity added by RQD makes the LDPC correction possible providing a considerable improvement. On the other hand, when the error-correcting capacity of the code is greater, the improvement provided by RQD becomes smaller, as is shown in Fig. 14 for the case of a

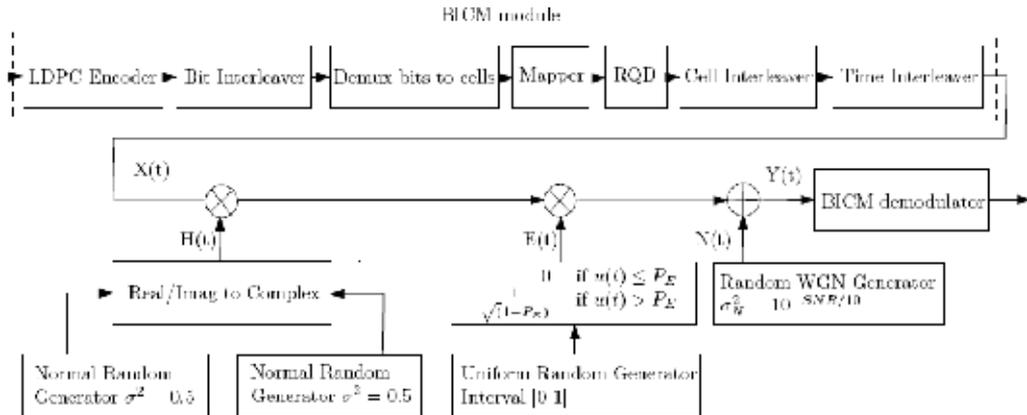


Fig. 13. Equivalent DVB-T2 BICM system over flat fading Rayleigh channel with erasures.

code rate of 2/3. Therefore, the gain of the system with RQD transmission results very significant over this kind of channel models for high coding rates and low modulation orders.

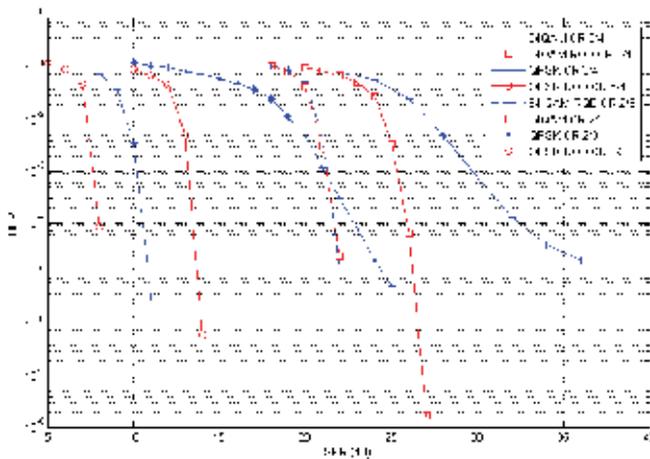


Fig. 14. Performance of RQD systems over flat Rayleigh channel with erasures of 20 %.

7.2 Analysis of RQD in the DVB-T2 system over Rayleigh fading channels

The complete DVB-T2 system model has now been used to analyze the behavior of the RQD technique. In this case, the channels are frequency-selective and hence every carrier of the OFDM symbol suffers a different fading event. Fig. 15 shows the results for Rayleigh Portable 1 (ETSI, 1997) and TU6 (COST207, 1989) channels. It can be seen that the gain of RQD decreases significantly in comparison to the results displayed in the previous section. This gain depends on the order of modulation and the code rate, decreasing when the latter is robust enough to recover the signal without RQD and increasing with the modulation order.

Fig. 15a depicts the RQD gain as a function of the modulation order considering a fixed code rate of 3/4. QPSK mode obtains a gain of 0.7 dB, which is reduced to 0.3 dB in 16QAM and

negligible for 64QAM. When the code rate is low, the RQD method improves the reception thanks to the increased diversity level. Consequently, the RQD method can be taken into account for high code rates where the redundancy rate of the coding is similar or even less than the loss of information. This is shown in Fig. 15b where the system with 16QAM modulation and code rate 5/6 offers a gain of 0.5 dB, whereas the gain disappears for the same modulation with code rate 3/5. A very interesting conclusion that can be drawn from the results depicted in Fig. 15 is that the RME channel is not a very realistic channel when modeling realistic hard transmission scenarios like a TU6 channel.

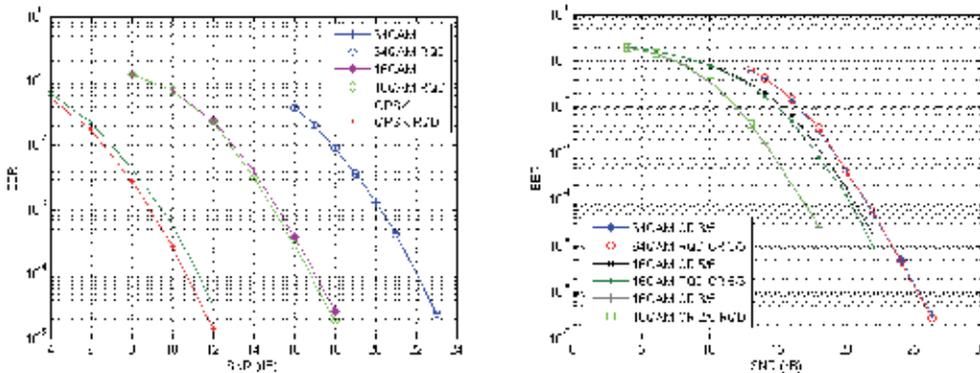


Fig. 15. Comparison of DVB-T2 systems in Rayleigh P1 channel with code rate 3/4 (a) and TU6 channel with code rates of 3/5 and 5/6 (b).

8. Performance in single frequency networks

DVB-T2 and its predecessor DVB-T are based on OFDM which is robust against intersymbol interference and fading caused by multipath propagation. These characteristics facilitate the deployment of SFN networks. Unlike multiple frequency networks (MFN), SFN involves several transmitters broadcasting synchronously the same program at the same frequency. The coverage area of every transmitter is called cell and the network can be composed of two or more cells which are deployed to cover wide areas with a unique frequency band.

The main advantage of this deployment strategy is the efficient use of the television spectrum, allowing a higher number of TV programs (Penttinen, 2008). On the other hand, the addition of two identical delayed signals is not always constructive and can lead to severe destructive interference in specific locations, making the signal unrecoverable. This section of the chapter evaluates the performance of DVB-T and DVB-T2 systems over SFN scenarios focusing on the benefits provided by the new standard.

8.1 Self-interference analysis

The simultaneous transmission from several sources can be considered as multipath propagation since the received echoes of the same signal form constructive or destructive interference which results in fading. This is known as self-interference and a correct SFN deployment is essential to reduce its degrading effects on the service area. The guard interval of the OFDM technique avoids ISI allowing operating in multipath fading

environments. Nevertheless, although all the received signals of the SFN transmission are within the guard interval, their superposition may be destructive and the self-interference can degrade considerably the quality at reception. This effect can be avoided by means of directive antennas, but it is occasionally impossible due to environmental constraints. In that case, the performance of the receivers is reduced in locations with good signal reception level.

The SFN network gain is defined as the positive power contribution due to constructive superposition of all the received signals within the guard interval in comparison to the necessary power in order to cover the same area with a MFN network (Santella et al., 2004). However, as it has been explained previously, this contribution also generates destructive interference despite being inside the guard interval. The following simple constructive and destructive wave interference problem can justify this effect.

Using a phasor diagram as the one depicted in Fig. 16, the resulting waveform from the combination of two coherent sources can be expressed as the real part of the vector sum of the individual phasors with magnitude A_1 and A_2 and phases φ_1 and φ_2 respectively, which can be written as:

$$\Psi_T(t) = A_T \cos(\omega t + \varphi_T) \tag{10}$$

where the resultant waveform amplitude is:

$$A_T = \sqrt{A_1^2 + A_2^2 + 2A_1A_2 \cos(\varphi_1 - \varphi_2)} \tag{11}$$

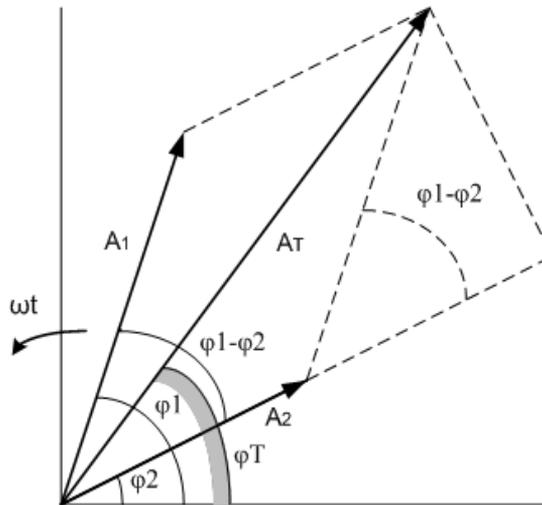


Fig. 16. Phasor diagram.

If T_0 is the resulting delay due to the phase differences, Equation (11) can be written as:

$$A_T = \sqrt{A_1^2 + A_2^2 + 2A_1A_2 \cos(2\pi fT_0)} \tag{12}$$

where f is the carrier frequency.

Equation (12) states that the resultant waveform amplitude will never be zero if received signal powers are different. However, if the power levels are equal, which involves that $A_1=A_2=A$, Equation (12) can be simplified after some operations as follows:

$$A_T = 2A \cos(\Pi f T_0) \quad (13)$$

In this case, the resultant waveform amplitude is zero periodically, meaning that several frequency subcarriers are lost. The number of nulls within the OFDM spectrum depends on the delay between transmitters since the distance between nulls is $\Delta f=1/T_0$. Depending on the number of lost subcarriers and the code rate, this effect can avoid the QEF reception of the television signal.

8.2 Performance of DVB-T/T2 systems in self-interference scenarios

Simulation results are provided to analyze and compare the behavior of DVB-T and DVB-T2 systems in SFN networks due to the aforementioned self-interference scenario. The results, presented as BER over SNR curves, are achieved by means of Montecarlo simulations. The QEF limit after the convolutional decoder in DVB-T and after LDPC decoder in DVB-T2 is considered as an evaluation reference.

A DVB-T scenario is first considered where the receivers are at line of sight of two transmitters, being one of them delayed with different power levels. The synchronization to one transmitter is perfect, the direct component is only considered and an AWGN channel is assumed for simplicity. The Spanish DVB-T broadcasting options have been simulated as defined in the following table:

| FFT Size | Mode | Code Rate | GI |
|--------------------|-------|-----------|-----|
| 8K (8192 Carriers) | 64QAM | 2/3 | 1/4 |

Table 2. Options of DVB-T simulations.

Fig. 17a and Fig. 17b show the behavior of the DVB-T system with echoes delayed for a half and a complete guard interval, respectively. Both situations provide an error floor in the performance for the 0 dB (same power) echo case. However, this keeps on appearing for -1 dB and -2 dB in Fig. 17b. An echo delayed the guard interval length generates the most severe fading without introducing ISI and the number of cancelled carriers is twice as much as in the case of Fig. 17a. Furthermore, it can be seen that 0 dB and -1 dB echoes generate error floors above the QEF limit.

The behavior of DVB-T2 in SFN networks has been evaluated with similar parameters as for DVB-T, as defined in the Table 3.

| FFT Size | Mode | LDPC Block Length | LDPC Code Rate | GI | Pilot Pattern |
|--------------------|-------|-------------------|----------------|-----|---------------|
| 2K (2048 Carriers) | 64QAM | 16200 | 3/5 | 1/4 | PP1 |

Table 3. Options of DVB-T2 simulations.

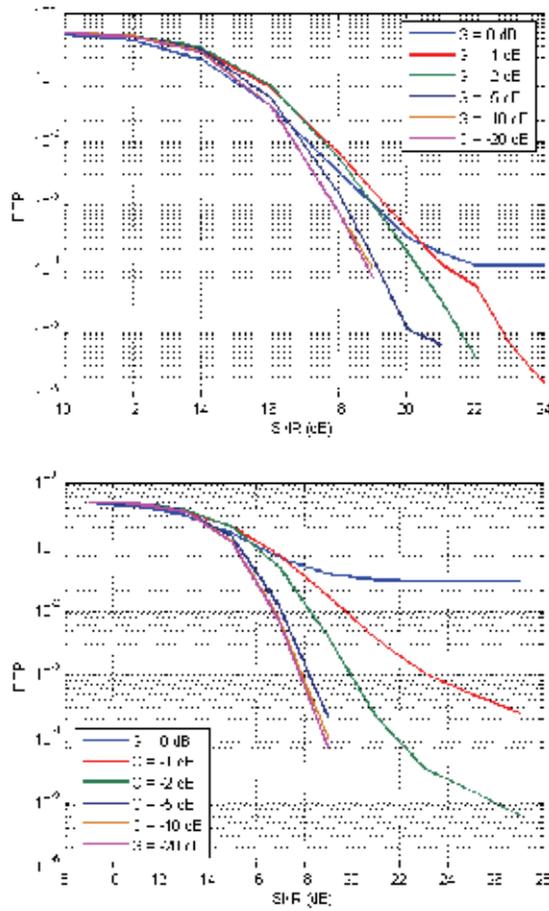


Fig. 17. BER performance of DVB-T system with an echo delayed GI/2 (a) and GI (b).

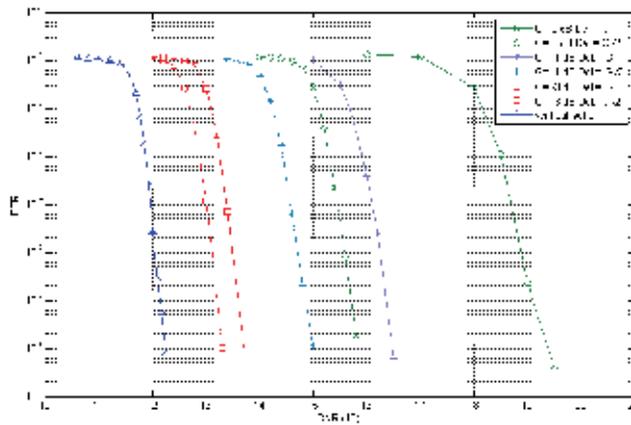


Fig. 18. DVB-T2 performance for echoes delayed GI/2.

The FFT size is shorter in the case of DVB-T2 in order to reduce the simulation time, which does not affect the performance if delays are proportional. The optional RQD technique has been included since it adds diversity to the system.

Fig. 18 shows the simulation results with echoes of variable power delayed GI/2 and GI. One can see that the performance gain of the DVB-T2 system is around 6.5 dB greater than that of DVB-T and that it is more robust against self-interference. When the system is affected by an echo delayed GI, DVB-T2 achieves good reception where DVB-T does not reach QEF. As can be seen, all the simulated scenarios, which are as severe as in the previous DVB-T simulations, achieve the QEF limit of 10^{-7} .

9. Summary and conclusions

This chapter has described the most interesting aspects of the new DVB-T2 digital video broadcasting standard. New research approaches and results have been shown regarding signal processing improvements over DVB-T2. A novel iterative demapping and decoding strategy has been shown which can be used with the LDPC codes provided by the new standard. Furthermore, results showing the effects of the number of demapping and LDPC iterations on the overall performance of the receiver have been provided. It has been proved that iterative demapping and detection, both regular and irregular, can improve considerably the performance of a DVB-T2 receiver, being specially suitable for problematic or low-level locations.

Multiantenna schemes and channel estimation algorithms have also been analyzed for the new standard, showing the importance of this issues on its behavior. The improvement achieved by the Alamouti MISO transmission scheme has been evaluated in conjunction with an MRC diversity scheme. This rather simple diversity approaches have shown a high performance improvement in DVB-T2 scenarios.

Two special features of DVB-T2 have been specially considered: the effects of the newly introduced rotated constellation transmission scheme and the performance of DVB-T2 in SFN scenarios. Regarding the rotated constellation approach its effects have been proven to be determinant in RME channels, whereas they introduce slight improvements in realistic frequency-selective channel like TU6. Considering the performance of the standard in a SFN interference scenario, DVB-T2 has shown to be a much more robust standard, allowing perfect reception of the signals in interference regions where DVB-T did not work.

10. Acknowledgements

The authors would like to thank the Spanish Government for the funding received through the research projects FURIA and FURIA2, as well as to all the partners of the FURIA consortium.

11. References

- Alamouti, S. (1998). A simple transmit diversity technique for wireless communications. *IEEE Journal on Selected Areas in Communications*, Vol. 16, No. 8, Oct. 1998, pp 1451-1458, ISSN: 0733-8716.

- Al-Semari, S. & Fuja, T. (1997). I-Q TCM: reliable communication over the Rayleigh fading channel close to the cutoff rate. *IEEE Transactions on Information Theory*, Vol. 43, No. 1, Jan. 1997, pp. 250-262.
- Caire, G.; Taricco, G. & Biglieri, E. (1998). Bit-interleaved coded modulation. *IEEE Transactions on Information Theory*, vol. 44, no. 3, May 1998, pp. 927-946.
- COST207 (1989). Digital land mobile radio communications (Final report). Technical report, Commission of the European Communities, Directorate General Telecommunications, Information Industries and Innovation, 1989.
- DVB (2008). Framing structure, channel coding and modulation for a second generation digital terrestrial television broadcasting system (DVB-T2). *DVB Document A122*, Jun. 2008.
- DVB (2009). Implementation guidelines for a second generation digital terrestrial television broadcasting system (DVB-T2). *DVB Document A133*, Feb. 2009.
- ETSI (1997). Digital video broadcasting (DVB); framing structure, channel coding and modulation for digital terrestrial television (DVB-T). *ETS EN 300 744*. Mar. 1997
- Foschini, G. J. & Gans, M. J. (1998). On limits of wireless communications in a fading environment when using multiple antennas. *Wireless Personal Communications*, Vol. 6, No. 3, Mar. 1998, pp. 311-335, ISSN: 0929-6212
- Hoher, P ; Kaiser, S; Robertson, P. Two-dimensional pilot-symbol-aided channel estimation by wiener filtering. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 1845-1848, ISBN: 0-8186-7919-0, Munich, Germany, Apr. 1997.
- Jakes, W. C. (1974). *Microwave Mobile Communications*. Wiley-IEEE Press. ISBN 0 7803 1069 1. New York
- Levy, Y. (2004). DVB-T - A fresh look at single and diversity receivers for mobile and portable reception. *EBU Technical Review*, No. 298, Apr. 2004.
- Li, X.; Chindapol, A. & Ritcey, J.A. (1998). Bit-interleaved coded modulation with iterative decoding using soft feedback. *IEEE Electronic Letters*, vol. 34, no. 10, May 1998, pp. 942-943.
- Li, X.; Chindapol, A. & Ritcey, J.A. (2002). Bit-interleaved coded modulation with iterative decoding and 8PSK signalling. *IEEE Transactions on Communications*, vol. 50, no. 8, Aug. 2002, pp. 1250-1257.
- Necker, M ; Stüber, G. Totally blind channel estimation for OFDM on fast varying mobile radio channels. *IEEE Transactions on Wireless Communications*, vol. 3, no. 5, Sept. 2004, pp. 1514-1525. ISSN: 1536-1276.
- Nour, C. A. & Douillard, C. (2008). Rotated QAM constellations to improve BICM performance for DVB-T2. *IEEE 10th International Symposium on Spread Spectrum Techniques and Applications*, pp. 354-359, ISBN: 978-1-4244-2203-6, Bologna, Italy, Aug. 2008.
- Patzold, M. (2002). *Mobile Fading Channel*. John Wiley & Sons. ISBN 0471495492. Chichester, England.
- Penttinen, J. T. J. (2008). The SFN gain in non-interfered and interfered DVB-H networks. *The Fourth International Conference on Wireless and Mobile Communications*, pp.294-299, ISBN: 978-0-7695-3274-5, Athens, Greece, Jul. 2008.

-
- Proakis, J. G. (1995). *Digital Communications*. McGraw-Hill. ISBN 0072321113. New York, USA.
- Santella, G.; De Martino, R. & Ricchiuti, M. (2004). Single frequency network (SFN) planning for digital terrestrial television and radio broadcast services: the Italian frequency plan for T-DAB. *IEEE 59th Vehicular Technology Conference*, Vol. 4, May 2004, pp. 2307-2311, ISSN: 1550-2252.
- Telatar, E. (1999). Capacity of multi-antenna Gaussian channels. *European Transactions on Communications*, Vol. 10, No. 6, Nov./Dec. 1999, pp. 585-595.
- Zehavi, E. (1992). 8-PSK trellis codes for Rayleigh channel. *IEEE Transactions on Communications*, vol. 40, no. 5, May 1992, pp. 873-874.

Passive Radar using COFDM (DAB or DVB-T) Broadcasters as Opportunistic Illuminators

Poullin Dominique
ONERA
France

1. Introduction

This chapter is not dedicated to improve DVB-T (Digital Video Broadcasters-Terrestrial) reception in critical broadcasting conditions. Our purpose is to explain and illustrate the potential benefits related to the COFDM (Coded Orthogonal Frequency Division Multiplex) waveform for passive radar application. As we'll describe, most of the benefits related to COFDM modulation (with guard interval) for communication purpose, could be derived as advantages for passive radar application. The radar situation considered is the following: the receiver is a fixed terrestrial one using COFDM civilian transmitters as illuminators of opportunity for detecting and tracking flying targets. The opportunity COFDM broadcasters could be either DAB as well as DVB-T ones even in SFN (Single Frequency Network) mode for which all the broadcasters are transmitting exactly the same signal. Such application is known in the literature as PCL (Passive Coherent Location) application [Howland et al 2005], [Baker & Griffiths 2005].

This chapter will be divided into three main parts. The first ones have to be considered as simple and short overviews on COFDM modulation and on radar basis. These paragraphs will introduce our notations and should be sufficient in order to fully understand this chapter. If not, it is still possible to consider a „classical“ radar book as well as some articles on COFDM like [Alard et al 1987]. More specifically, the COFDM description will outline the properties that will be used in radar detection processing and the radar basis will schematically illustrate the compulsory rejection of the „zero-Doppler“ paths received directly from the transmitter or after some reflection on the ground.

Then the most important part will detail and compare two cancellation filters adapted to COFDM waveform. These two filters could be applied against multipaths (reflection on ground elements) as well as against multiple transmitters in SFN mode. In this document, no difference will be done between SFN transmitters contributions and reflections on fixed obstacles : all these zero-Doppler paths will be considered as clutter or propagation channel. Obviously, these filters will be efficient also in a simple MFN (Multiple Frequency Network) configuration. Most of the results presented below concerns experimental data, nevertheless some simulations will also be used for dealing with some specific parameters.

2. Principle of COFDM modulation

As mentioned in the introduction, the purpose of this paragraph is just to briefly describe the principle and the main characteristics of the COFDM modulation in order to explain its

advantages even for radar application. For further details, it's better to analyse the reference [Alard et al 1987], however for radar understanding this short description should be sufficient.

2.1 Basis principle

In a COFDM system of transmission, the information is carried by a large number of equally spaced sinusoids, all these sub-carriers (sinusoids) being transmitted simultaneously.

These equidistant sub-carriers constitute a "white" spectrum with a frequency step inversely proportional to the symbol duration.

By considering these sub-carriers:

$$f_k = f_0 + k/T_s \quad (1)$$

with T_s corresponding to symbol duration.

It becomes easy to define a basis of elementary signals taking into account the transmission of these sinusoids over distinct finite duration intervals T_s :

$$\psi_{j,k}(t) = g_k(t - jT_s) \text{ with } \begin{cases} 0 \leq t < T_s & : g_k(t) = e^{2i\pi f_k t} \\ \text{elsewhere} & : g_k(t) = 0 \end{cases} \quad (2)$$

All these signals are verifying the orthogonality conditions:

$$j \neq j' \text{ or } k \neq k' : \int_{-\infty}^{+\infty} \psi_{j,k}(t) \psi_{j',k'}^*(t) dt = 0 \quad \text{and} \quad \int_{-\infty}^{+\infty} \|\psi_{j,k}\|^2 dt = T_s \quad (3)$$

By considering the complex elements $\{C_{j,k}\}$ belonging to a finite alphabet (QPSK, 16 QAM,...) and representing the transmitted data signal, the corresponding signal can be written:

$$x(t) = \sum_{j=-\infty}^{+\infty} \sum_{k=0}^{N-1} C_{j,k} \psi_{j,k}(t) \quad (4)$$

So the decoding rule of these elements is given by:

$$C_{j,k} = \frac{1}{T_s} \int_{-\infty}^{+\infty} x(t) \psi_{j,k}^*(t) dt \quad (5)$$

Remark:

From a practical point of view this decomposition of the received signal on the basis of the elementary signals $\psi_{j,k}(t)$ could be easily achieved using the Fourier Transform over appropriate time duration T_s .

2.2 Guard interval use

In an environment congested with multipaths (reflections between transmitter and receiver), the orthogonality properties of the received signals $\psi_{j,k}(t)$ are no longer satisfied.

In order to avoid this limitation, the solution currently used, especially for DAB and DVB, consists in the transmission of elementary signals $\psi_{j,k}(t)$ over a duration T_s' longer than T_s . The difference between these durations is called guard interval. The purpose of this guard interval is to absorb the troubles related to the inter-symbols interferences caused by the propagation channel. This absorption property needs the use of a guard interval longer than the propagation channel length. Then, we just have to "wait for" all the contributions of the different reflectors in order to study and decode the signal on a duration restricted to useful duration T_s .

The transmitted signal could be written:

$$x(t) = \sum_{j=-\infty}^{+\infty} \sum_{k=0}^{N-1} C_{j,k} \psi'_{j,k}(t) \quad (6)$$

$$\text{with } \psi'_{j,k}(t) = g'_k(t - jT_s') \text{ with } \begin{cases} -\Delta \leq t < T_s & : g'_k(t) = e^{2i\pi f_k t} \\ \text{elsewhere} & : g'_k(t) = 0 \end{cases} \quad (7)$$

Nevertheless the decoding rule of these elements is still given by:

$$C_{j,k} = \frac{1}{T_s} \int_{-\infty}^{+\infty} x(t) \psi_{j,k}^*(t) dt \quad (8)$$

with $\psi_{j,k}(t)$ always defined on useful duration T_s while signal is now specified (and transmitted) using elementary signals $\psi'_{j,k}(t)$ defined on symbol duration $T_s' = T_s + \Delta$.

This decoding rule means that even when signals are transmitted over a duration $T_s' = T_s + \Delta$, the duration used, in reception for decoding will be restricted to T_s . Such a "cut" leads to losses equal to $10 \log T_s' / T_s$ but allows easy decoding without critical hypothesis concerning the propagation channel. In practice, this truncation doesn't lead to losses higher than 1 dB (the maximum guard interval Δ is generally equal to a quarter of the useful duration T_s).

The guard interval principle could be illustrated by the figure 1.

The previous figure illustrates the main advantage of guard interval truncation: by "waiting" for all the fixed contributors, it's easy to avoid signal analysis over transitory (and unstationary) time durations.

Considering the parts of signal used for decoding (so after synchronisation on the end of the guard interval related to the first path received), the received signal in an environment containing clutter reflectors could be written as:

$$jT_s' \leq t < jT_s' + T_s \quad : \quad y(t) = \sum_{k=0}^{N-1} H_{j,k} C_{j,k} \psi_{j,k}(t) \quad (9)$$

The propagation channel for the symbol j after the guard interval could be "summarized" with only one complex coefficient per transmitted frequency ($H_{j,k}$) as, during this portion of studied time, all the reflectors were illuminated by the signal $C_{j,k} \psi_{j,k}(t)$ alone.

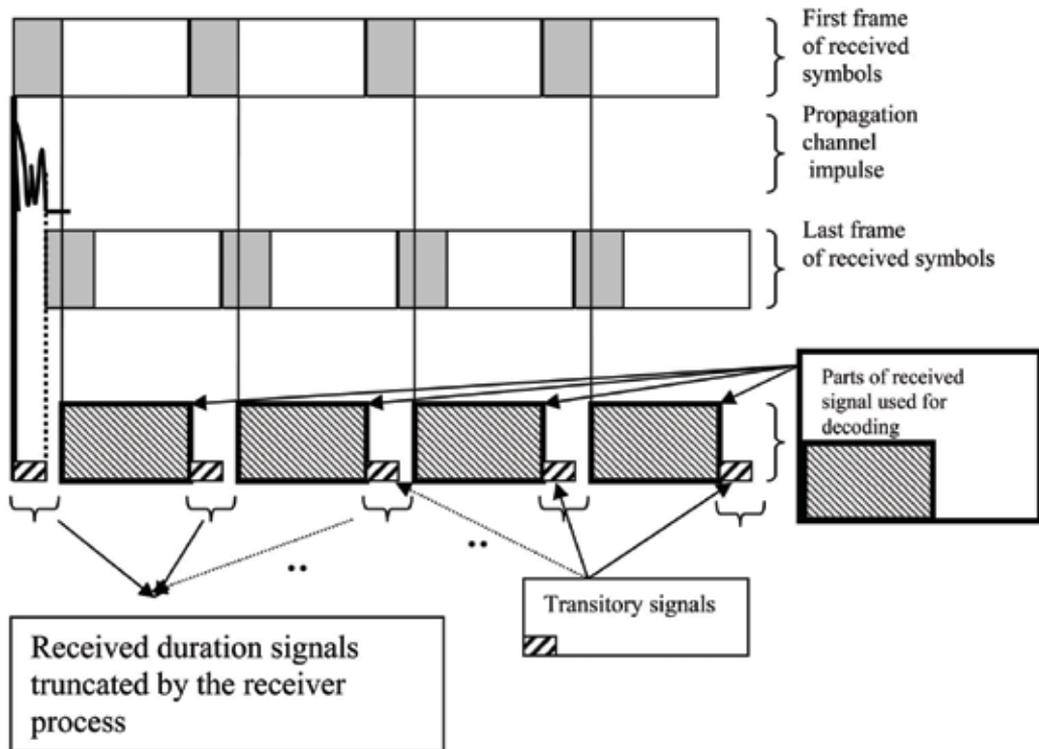


Fig. 1. Guard interval principle

Remark:

COFDM Waveform (with guard interval principle) can support superposition of different paths "without troubles". Such a property also allows a particular mode in a multiple transmitters configuration: all the transmitters can use simultaneously the same code and the same carrier frequency. This specific mode is called SFN (Single Frequency Network). In the rest of this chapter, there will be no difference considered between a multipath or a SFN transmitter. Furthermore, the propagation channel considered will include all the coherent paths, that means multipath on ground clutter as well as SFN transmitters.

2.3 Demodulation

The purpose of this paragraph is not to explain the demodulation principle well described in the DVB norm or in articles [Alard et al 1987 for example] for differential decoding when phase modulation is used.

Whatever considering optimal demodulation or differential one for phase codes, the decoding principle is based on estimating the transmitted codes using the received signal:

$$Y_{j,k} = H_{j,k}C_{j,k} + N_{j,k} \quad (10)$$

where $N_{j,k}$ represents a gaussian noise:

The knowledge of the channel impulse response $H_{j,k}$ and of the noise standard deviation $\sigma_{j,k}^2$ can be used for the coherent demodulation. This optimal demodulation consists in maximising over the $C_{j,k}$ the following relation:

$$\sum_j \sum_k \operatorname{Re} \left(Y_{j,k} H_{j,k}^* C_{j,k}^* / \sigma_{j,k}^2 \right) \quad (11)$$

In order to simplify this demodulation, it's possible to perform differential demodulation instead of coherent demodulation for QPSK codes. This differential demodulation assumes propagation channel stationarity and consists in estimating the channel response from the previous symbol:

$$H_{j,k} \cong \frac{Y_{j-1,k}}{C_{j-1,k}} \quad (12)$$

This differential demodulation is particularly interesting for its simplicity. The 3 dB losses due to this assumption have to be compared to the practical difficulties encountered for the coherent demodulation implementation.

As a small comment, the differential demodulation doesn't estimate directly the elements of code $C_{j,k}$ but only the transitions between $C_{j-1,k}$ and $C_{j,k}$. However, for phase codes, like BPSK (or QPSK) the transition codes remains phase codes with two (or four) states of phase. In practice, such a differential demodulation just consists in Fourier transforms and some differential phase estimations (according to four possible states).

The most important conclusion dealing with these two possible demodulation principles is the following: using the received signal, it is possible to obtain and reconstruct an ideal vision of the transmitted one. In communication domain, this ideal signal is used for estimating the information broadcasted while for radar application this ideal signal will be used as a reference for correlation and could be also used for some cancellation process. For these radar applications, it is important to notice that this reference is a signal based on an ideal model. Furthermore, the decision achieved during the demodulation process has eliminated any target (mobile) contribution in this reference signal.

2.4 Synthesis

The COFDM signal has interesting properties for radar application such as:

- it is used for DAB and DVB European standard providing powerful transmitters of opportunity.
- the spectrum is a white spectrum of 1.5MHz bandwidth (1536 orthogonal sub-carriers of 1kHz bandwidth each) for DAB and 7.5 MHz for DVB-T
- the transmitted signal is easy to decode and reconstruct
- this modulation has interesting properties in presence of clutter : it is easy to consider and analyze only some parts of received signal without any transitory response due to multipaths effects.

3. Radar detection principle

3.1 Introduction

The principle of radar detection using DAB or DVB-T opportunistic transmitters will be classically based on the correlation of the received signals with a reference (match filter).

In the case of a transmitter using COFDM modulation, the estimation of the transmitted signal (reference) is easy to implement in order to ensure capabilities of range separation and estimation.

However, as the transmitted signal is continuous, we have to take a particular care of the ambiguity function side lobes for such a modulation. Firstly, we'll just verify that these side lobes related to the direct path (path between the transmitter and the receiver) are too high in order to allow efficient target detection and then we'll describe an adaptive filter whose purpose is to cancel all the main zero-Doppler path contributions and ensure efficient detection for mobile targets.

For limiting some specific correlation side lobes observable with the DVB-T signals, it is possible to consider the following article [Saini & Cherniakov 2005]: their analysis lead to a strong influence of the boosted pilot sub-carriers. The main suggestion of this article is to limit this influence by weighting these specific sub-carriers proportionally to the inverse of the „boosted level“ of 4 over 3.

3.2 Radar equation example

In a first approach, that means excepting the specific boosted sub-carriers mentioned above for DVB-T, the COFDM modulation ambiguity side lobes can be considered as quite uniform (in range-Doppler domain) with a level, below the level of direct path, given by the following figure:

$$-10\log_{10}(MN) \quad (13)$$

where M designs the number of symbols (considered for correlation) and N the number of sinusoids broadcasted.

The next figure presents the exact ambiguity function (left part of the figure) for a COFDM signal with 100 symbols and 150 sinusoids per symbol, we can observe that the secondary lobes are roughly - 42 dB below the main path (except for low Doppler and range lower than the guard interval: here 75 kilometres). Under some assumptions (right part of the figure: Doppler rotation neglected inside one symbol)), we can consider, in some restricted range-Doppler domain (especially for range lower than the guard interval), a lower level of side-lobes. However, this improvement, related to an "optimal" use of the sub-carriers orthogonality, remains not enough efficient in an "operational" context so we'll don't discuss such considerations in this paper.

We'll just end this COFDM ambiguity function considerations by the following expression ($\phi(\tau, \nu)$ represents the (range, Doppler) ambiguity function).

$$\phi_{left}(\tau, \nu) = \int_{T_{integration}} s_{received}(t) s_{reference}^*(t - \tau) e^{-i2\pi\nu t} dt \quad (14)$$

$$\phi_{right}(\tau, \nu) = \sum_{j=0}^{J-1} e^{-i2\pi\nu \left(j + \frac{1}{2}\right) T_s} \int_{jT_s + \Delta}^{(j+1)T_s} s_{received}(t) s_{reference}^*(t - \tau) dt \quad (15)$$

where the coherent integration time $T_{integration}$ is equal to $T_{integration} = J(T_s) = J(T_s + \Delta)$

The signal of reference is obtained using differential decoding principle.

The two previous expressions illustrate that the "right" correlation is equal to the "left" one under the assumption that Doppler influence is negligible inside each symbol duration. Furthermore, equation (15) illustrates that range correlations are just estimated over useful signal durations for which all the sub-carriers are orthogonal until the effective temporal support (function of the delay) remains exactly equal to useful duration T_s .

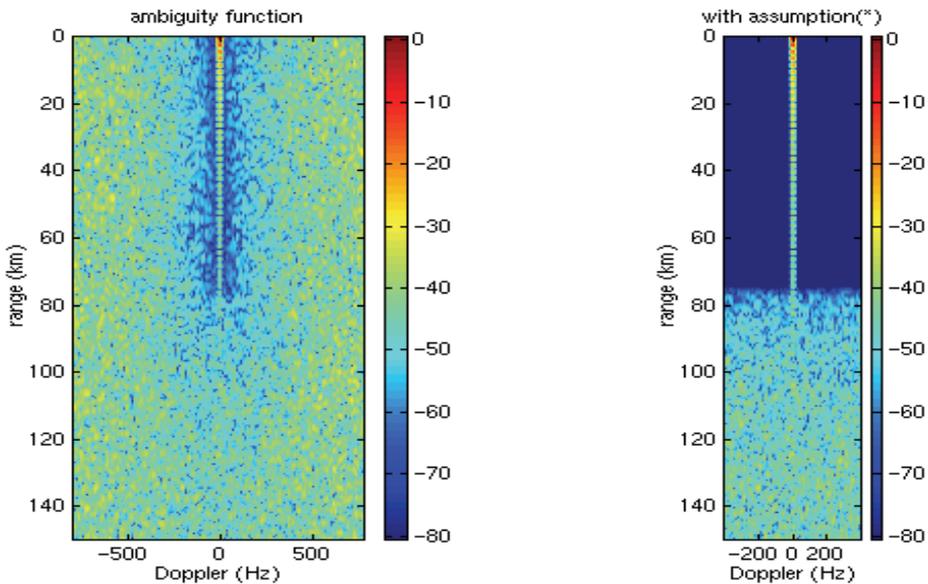


Fig. 2. COFDM ambiguity side lobes

(*) The Doppler rotation inside one symbol is neglected (right figure)

This property implies the lower level of side-lobes (visible on previous figure) for delays lower than guard interval length as using expression (15) there are no sub-carriers interferences in this range domain.

As our main purpose is to focus on the adaptive filter and not on radar equation parameters (coherent integration time, antenna gain and diagram,...), we'll don't discuss more in details on these radar equation parameters. We'll just consider: "as DAB or DVB-T waveforms are continuous, the received level of main path is always high and the isolation provided by side-lobes is not sufficient in order to allow detection."

As the side-lobes isolation (eq 13) is equal to the correlation gain (product between bandwidth and coherent integration time): when we receive a direct path with a positive signal to noise ratio (in the bandwidth of the signal), such a received signal allows reference estimation but its side-lobes will hide targets as these side lobes will have the same positive signal to noise ratio after compression (whatever coherent integration time we consider).

This phenomenon is schematically represented on next figure. Finally, observing this schematic radar equation, it's obvious that an efficient zero-Doppler cancellation filter is required as the targets are generally hidden by zero-Doppler paths side lobes.

3.3 Synthesis

This short description on radar principle had the only objective to prove the compulsory cancellation of the zero-Doppler paths in order to allow mobile target detection.

Only short overview on the correlation hypothesis and adjustments (for example for the boosted DVB-T pilots carriers) were given in order to be able to focus on the cancellation filter in the next part.

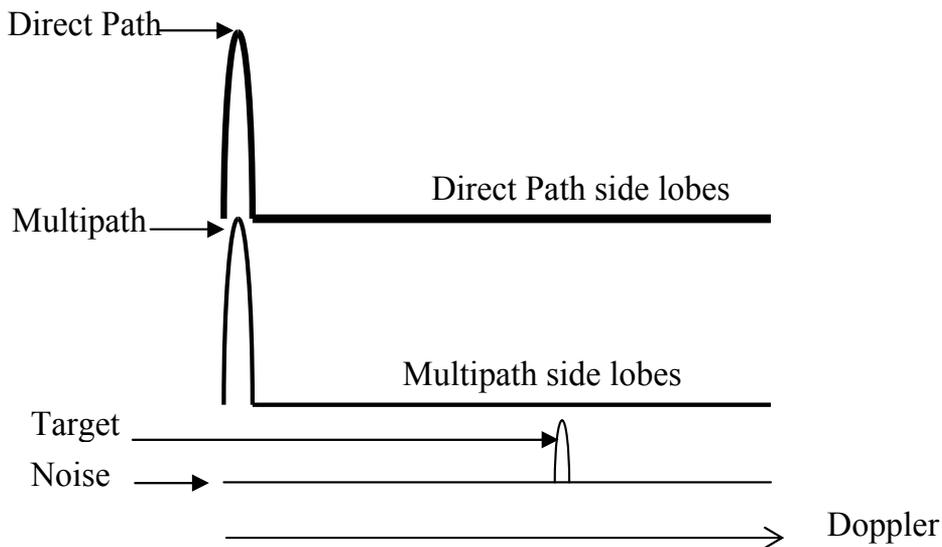


Fig. 3. Schematic radar equation (target hidden by side-lobes).

4. Detection principle

The purpose here is to present two approaches for the adaptive cancellation filter after a schematic description of the whole detection process.

The detection principle is divided into four main tasks described below:

- the first part consists in the transmitter parameters analysis (like carrier frequency, sampling frequency) and a "truncation" of the received signal in order to process only on stationary data
- the second part consists in estimating (by decoding) the reference signal that will be used for correlation
- the third part is more a diagnostic branch in order to allow a finest synchronisation for the direct path and consequently for the target echoes delays. This branch is also used for the propagation channel characterisation.
- The fourth part is related to the target detection and parameters estimation (Bistatic Doppler, bistatic range and azimuth).

This part dedicated to the target detection will be described in details in the following paragraphs.

5. Adaptive filter

5.1 Introduction

Before analyzing the filter itself, it seems important to remind the following elements.

- COFDM waveform allows the specific mode called SFN for which all the transmitters in a given area are broadcasting the same signal.
- From a global point of view, the level of COFDM side lobes is lower than the main path from the product (Bandwidth \times integration time). As this product is also equal to the coherent gain over the integration time, a path with a positive signal to noise ratio in the

bandwidth of the signal (so typically most of the SFN transmitters direct paths) will have side lobes with the same positive signal to noise ratio after coherent integration. Such considerations imply that the adaptive filter has to cancel efficiently all zero-Doppler contributors and not only the direct path. The two following filters considered here are fully adapted to the COFDM modulation and requires only a small array elements for the receiving system despite some other solutions sometimes developed [Coleman & Yardley 2008]. Furthermore, all the antennas (and related receivers) are used for the target analysis and detection: no additional hardware complexity and cost is added due to the zero-Doppler cancellation filters.

5.2 Adaptive filter principles

5.2.1 Cancellation filter using a receiving array

This first cancellation filter considers a small receiving array constituted by a set of typically four or eight receiving antennas: all these antennas will be used for the target analysis [Poullin 2001a]

Considering the signals over the different antennas of the receiver system, the zero-Doppler received signals for antenna i and symbol j (index k corresponds to the frequency) can be expressed as follows:

$$S_j^i = \sum_k H_{j,k}^i C_k^j \exp(j 2\pi k \frac{t_j}{T_s}) + N_{j,k}^i \quad (16)$$

$$\text{for } t_j \in [jT_s' + T_0 + L, (j+1)T_s' + T_0]$$

with: $H_{j,k}^i$: complex coefficient characterizing the propagation channel for symbol j , antenna i and frequency k . We'll see an explicit expression of such a coefficient some lines below, this expression will consider a specific simple configuration.

T_s' : is the transmitted duration (per symbol)

T_0 : corresponds to the first path time of arrival

L : designs the propagation channel length (delay between first path and last significant one including multipaths (echoes on the ground) as well as SFN paths).

$N_{j,k}^i$ designs the contribution of the noise (symbol j , antenna i and frequency k)

If the propagation channel length is lower than the guard interval, the previous expression will be valid for a duration longer than the useful one $T_s = T_s' - \Delta$. So it will be possible to consider this expression over durations T_s for which all sub-carriers are orthogonal between each other.

Generally, we could consider stationary propagation channel over the whole duration of analysis (coherent integration time for radar) and so replace expression $H_{j,k}^i$ by H_k^i

Finally considering the received signals over:

- the appropriate signal durations: for each transmitted symbol over T_s' , we just keep signal over useful duration T_s . (defined by the first path received and the guard interval).
- the appropriate frequencies: over that specific durations, the composite received signals always verify the sub-carrier orthogonality conditions even in multipath (and SFN) configuration.
- the receiver antenna array.

It's possible to synthesise the propagation channel response over the receiver array with a set of vectors

$$\left\{ (H_k^1, \dots, H_k^i, \dots, H_k^N) / k = 1, \dots, K : \text{frequency}, i = 1, \dots, N : \text{number of antennas} \right\} \quad (17)$$

where N is the number of elements in the receiver system.

So for each frequency k, it is possible to cancel the "directional vector" $\mathbf{H}_k = (H_k^1, \dots, H_k^i, \dots, H_k^N)^t$ using classical adaptive angular method based on covariance matrix as it can be seen below:

Considering for each frequency k the covariance matrix (with size related to the number of antenna) given by

$$R_k = E\left(\mathbf{H}_k \mathbf{H}_k^H C_k^j C_k^{j*} + \sigma_k^2 I\right) \quad (18)$$

$$\text{So } R_k = \mathbf{H}_k \mathbf{H}_k^H + \sigma_k^2 I \quad (19)$$

Consequently, when we'll apply the weightings related to the inverse of R_k for each frequency k, it appears weighting coefficients related to:

$$R_k^{-1} \approx \frac{1}{\sigma_k^2} \left(I - \frac{\mathbf{H}_k \mathbf{H}_k^H}{\mathbf{H}_k^H \mathbf{H}_k} \right) \quad (20)$$

which is the orthogonal projector to $\mathbf{H}_k = (H_k^1, \dots, H_k^i, \dots, H_k^N)^t$: propagation channel response vector at the sub-carrier k.

Remark:

This remark is just to give an explicit expression of a typical propagation channel response H_k^i (k: frequency, i antenna) in the particular case of two receiver antennas with a main path in the normal direction and a multipath characterized by its angle of arrival (θ). The normal path received on the first antenna is considered as reference. Under these hypothesis, the propagation channel responses could be written as:

$$\begin{aligned} H_k^1 &= (1 + \alpha \exp(j\phi) \exp(-2\pi j f_k \tau)) \\ H_k^2 &= (1 + \alpha \exp(j\phi) \exp(-2\pi j f_k \tau) \exp(j 2\pi d_{12} \sin(\theta) / \lambda)) \end{aligned} \quad (21)$$

where d_{12} designs the distance between the two antennas and λ is the wavelength $\alpha \exp(j\phi)$ represents the difference of reflectivity between main and multi-path (and τ is the delay between main path and multipath referred to antenna 1).

It is quite clear that H_k^1 and H_k^2 will quickly fluctuate according to frequency k due to the term $\exp(-j 2\pi f_k \tau)$. Furthermore, for a given frequency the term $j 2\pi d_{12} \sin(\theta) / \lambda$ implies different combinations of the two paths for the antenna.

5.2.1.1 *Example of cancellation efficiency on experimental data*

The filter implemented in order to cancel the zero-Doppler contributions was using four real antennas and the adaptive angular cancellation for each transmitted frequency as described previously. The transmitter was a DAB one and the correlation outputs in range-

Doppler already illustrate the cancellation capabilities that could be read on the cut along the Doppler axis for which the zero level reference corresponds to the receiver noise.

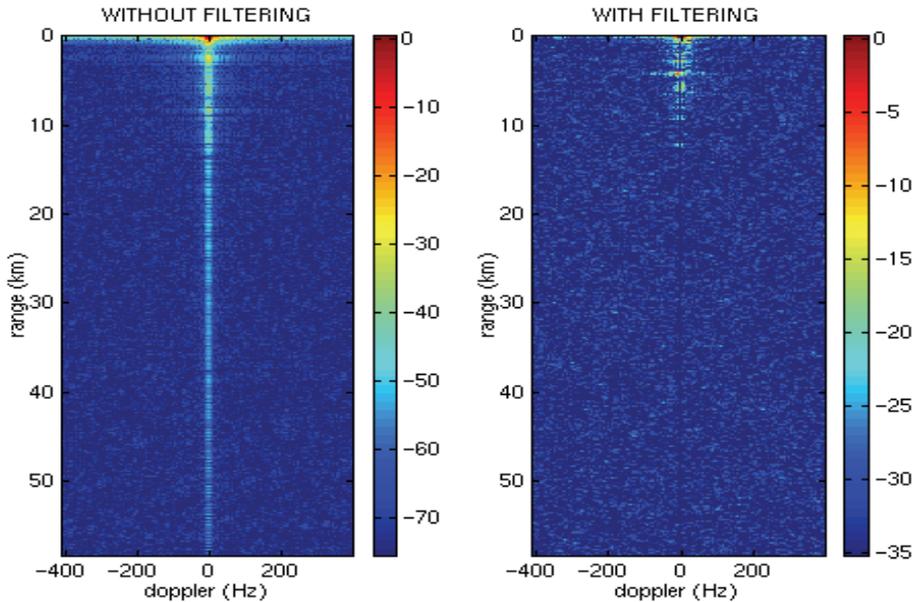


Fig. 4. Correlation output without cancellation filter (left) and with cancellation filter (right)

On the next figure, it could be seen that the level of the main path before cancellation had a signal to noise ratio (in the bandwidth of 10 Hz (related to the 100 milliseconds of integration) of 110 dB and the corresponding side lobes (for range lower than the guard interval which is equal to 75 kilometres) were still 35 dB above the noise level. After cancellation this residual spurious was only 7 dB above noise level. So the residual level of spurious could be considered as -103 dB below the main path

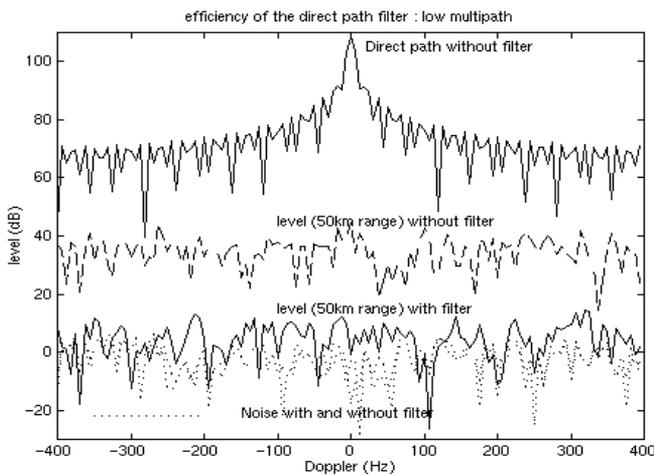


Fig. 5. Comparison of correlation outputs with and without cancellation filter .

5.2.1.2 Specific case of filter efficiency

The next figure corresponds to a target crossing the zero-Doppler axis. The trial configuration corresponds to the VHF-DAB transmitter analyzed just previously and six “snapshots” delayed from 1.5 seconds each are presented

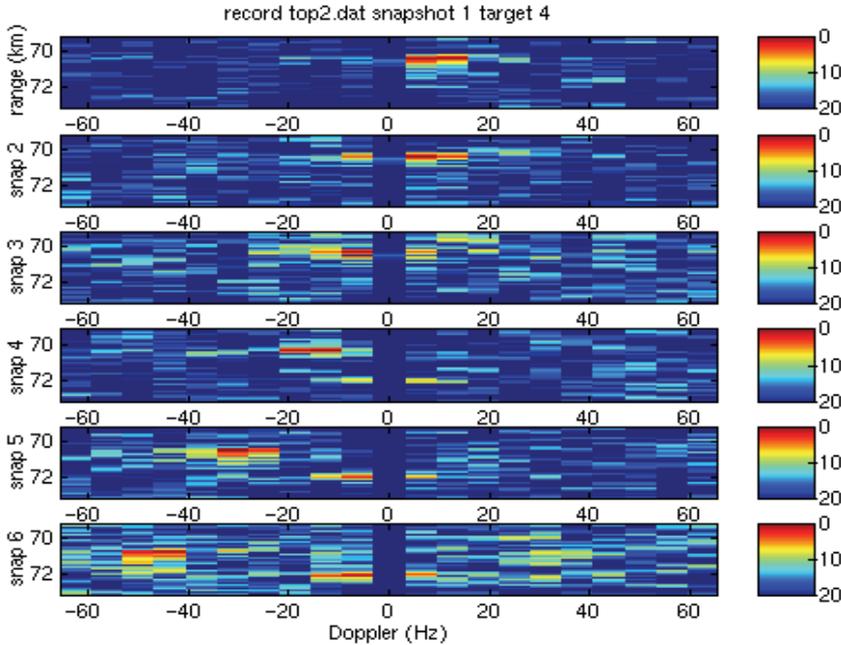


Fig. 6. Example of Results

This example of detection clearly shows the efficiency of the filter against multipath (reflector with null Doppler): when the target crosses the zero Doppler axis it is considered as an element of the clutter, so such a target is (during that time) fully coherent with clutter and main path and its contribution is integrated into the filter coefficients estimation, so such “mixed” coefficients reject both clutter and zero-Doppler target.

Such a result implies that even low multipath (clutter element whose signal to noise ratio is much lower than zero (dB) for the filter coefficient learning phase) could be filtered using such adaptive technique as “they are carried by the main path”. From a schematic point of view, the filter detects a level of interference related to $(A + \Delta A)^2$ even if ΔA^2 is negligible with respect to the noise level (A correspond to the main path level and ΔA to the multipath).

5.2.1.3 Example of target detections after filtering in SFN mode

The first figure represents the output of the correlation filter (match filter) without zero-Doppler cancellation filter. Such a process allows the analysis of the main fixed echoes generally corresponding to the main transmitters in a SFN configuration.

The different transmitters are identified using “a priori” knowledge of the multi-static configuration while the multipath was located and identified using several receiver

locations and triangulation. During that experiment, receiver noise level was high: the receiving system used was an existing “generic” one and not a specific receiver defined for passive DAB application. The figure 9 is normalised according to this high receiver noise. These results were obtained in a DAB-SFN configuration with numerous broadcasters and a two receiver antennas.

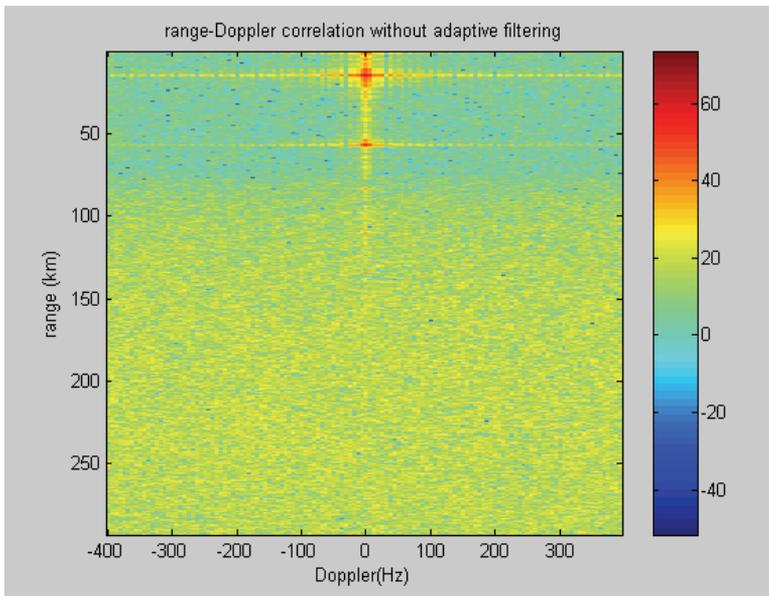


Fig. 7. Range Doppler correlation without zero-Doppler cancellation filter

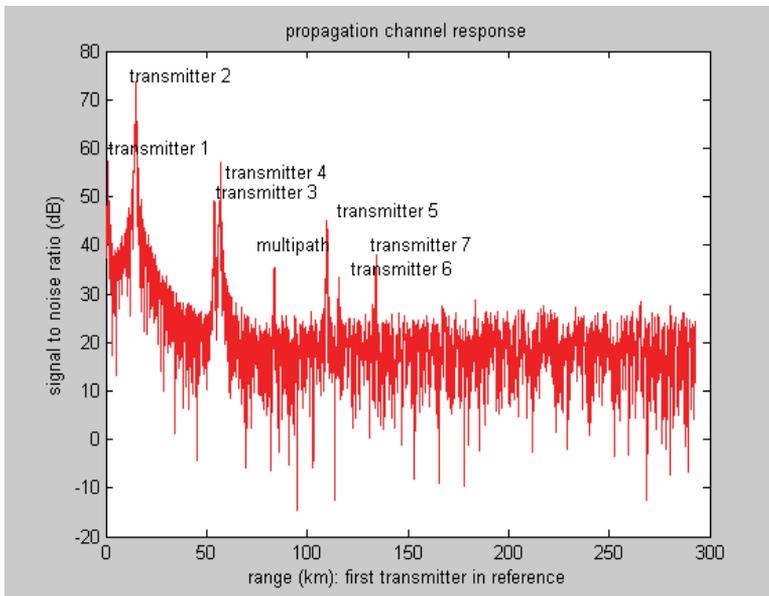


Fig. 8. Propagation channel response (analysis of correlation at zero Doppler: no filtering)

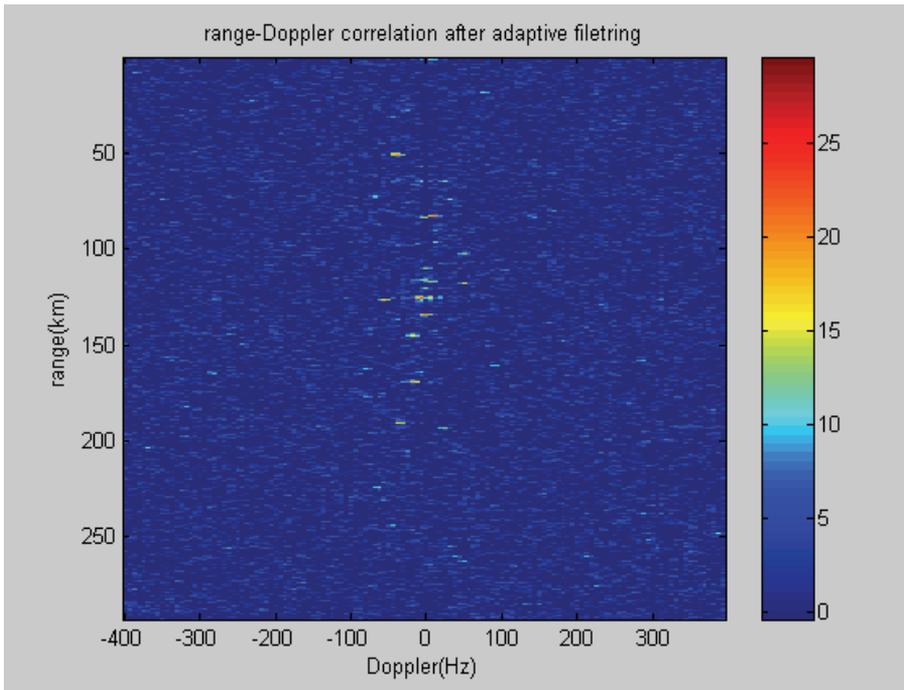


Fig. 9. Examples of mobile (non zero Dopplers) target detections after clutter cancellation

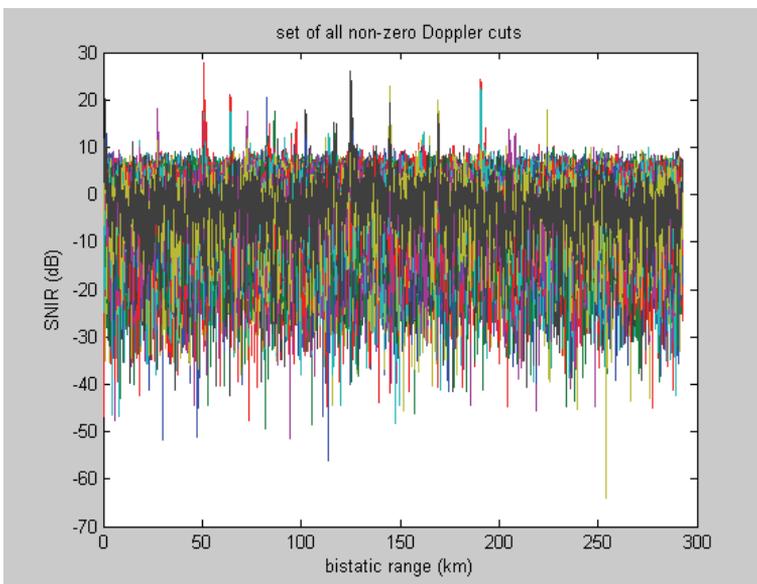


Fig. 10. Non-zero-Doppler cuts (of the range-Doppler correlation) after adaptive filtering

This figure corresponds to the previous multistatic situation with at least seven transmitters clearly identified on the propagation channel response. It becomes obvious that many mobile targets could be detected after zero-Doppler cancellation adapted to the COFDM-SFN configuration even using only two receiving antenna.

The superposition of all non-zero Doppler cuts is represented in order to give a “clear idea” of the detected targets (2-D images like the upper one with high number of pixels aren’t always suitable for such a purpose). Furthermore, these cuts illustrate that after adaptive filtering the floor level corresponds to the (high) level of receiver noise.

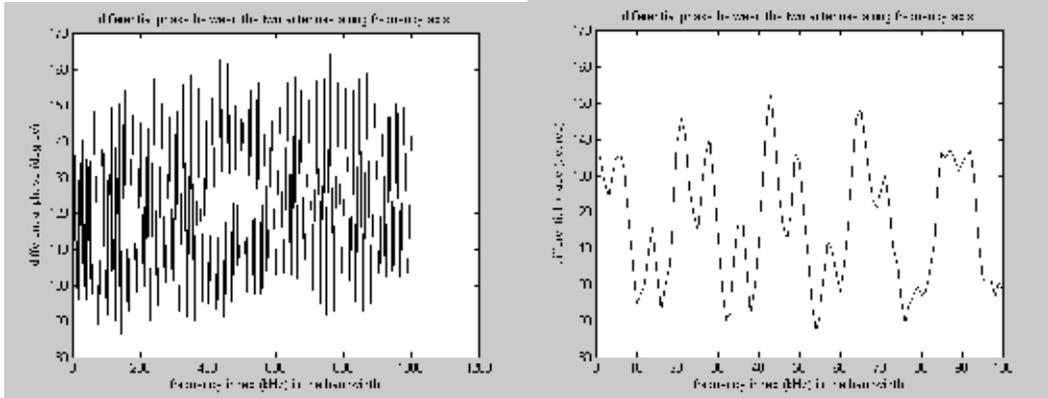


Fig. 11. Differential phase fluctuation between the two antennas along frequency axis.

Between the two distinct antenna, the important differential phase fluctuation along frequency axis in SFN mode (or high multipath configuration) is clearly illustrated on the previous figure.

These results show that, with COFDM modulation, it is possible to filter many SFN transmitters (or multipaths) with a small antenna array receiver. Nevertheless, the following principle: “bigger is your array, better are your results in terms of stability and narrow corrupted domain” remains true.

5.2.1.4 Synthesis:

In order to filter the clutter contributions (or the SFN transmitters), it is possible to consider the following algorithm described above involving time, frequency and angular domains:

- time domain: (cut of received guard intervals)
this truncation ensure stationary durations for signal analysis with no time codes superposition
- frequency domain (analysis over the transmitted sub-carriers)
The Fourier transform over the selected useful durations ensure signal analysis over stationary frequencies: no frequency codes superposition.
- “angular” domain: (adaptive beamforming for each frequency)
the adaptive filter (for each transmitted sub-carrier) ensures clutter rejection. On that specific durations and frequencies, as all the clutter contributors are fully coherent, only one degree of freedom is necessary for adaptive cancellation of all the fixed echoes.

This filter has been successfully tested on real DAB signals and is currently tested using DVB-T broadcasters for which preliminary results seem encouraging.

This “angular” filtering applied for each transmitted sub-carrier will:

- lower all the zero-Doppler contributors as the set of H_k coefficients summarises all the clutter contributions.
- Be theoretically able to lower multiple transmitters using two antennas as only one degree of freedom is required for cancelling the zero-Doppler paths as long as the propagation channel length remains lower than the guard interval.

- Orthogonalise the received signals to a composite vector that doesn't correspond to a particular direction (see explicit expressions of H_k coefficient in equation (21)). This phenomenon is due to the full coherency of all the clutter contributors over that selected time durations and that frequency sub-carriers. This particularity also implies that only one degree of freedom (per frequency) is used for all clutter cancellation.
- Have to be applied for each transmitted frequency as the composite propagation channel vector fluctuates quickly in frequency domain. This fluctuation could be deduced from the explicit expression of H_k (equation 21) coefficient. Furthermore, this fluctuation was illustrated on an experimental example on figure 11.

The next paragraph will present another cancellation filter that will be less efficient in most of the situations. Nevertheless, its interest relies in the following capabilities: it requires only one real antenna in order to lower the different SFN contributions and it could be more efficient than the previous filter when the target is close to the composite directional vector of the zero-Doppler contributors.

5.2.2 Cancellation using a single antenna

This other zero-Doppler path cancellation filter could be obtained according two different approaches:

- An "angular" approach derived from the previous method but with the following adaptation: the cancellation is no longer achieved between several real antenna of the receiving array but between each real antenna and a sort of fictive one receiving the signal of reference obtained after decoding.
- A "temporal approach" using the classical Wiener Filter adapted to COFDM waveform. We'll detail simply the "temporal" approach

Considering the decoded signal and a signal received on a real antenna it is possible to consider the Wiener filter:

$$z(t) = s_{received}(t) - \sum_{\tau=1}^L \mathcal{W}(\tau)ref(t-\tau) \quad (22)$$

$$\text{with } \min_w \left(\left| s_{reçu}(t) - \sum_{\tau} w(\tau)ref(t-\tau) \right|^2 \right)$$

Under that formulation, there is no specificity due to COFDM waveform and the similarity with the previous cancellation filter is not obvious.

So let us consider the spectral domain and the assumption that the length of the propagation channel and the corresponding Wiener filter length (here L) are lower than the guard interval.

So under that assumption, it is possible to consider the following expression

$$\sum_{\tau} \mathcal{W}(\tau)ref(t-\tau) = \sum_{k=1}^K G_k C_k e^{j2\pi \frac{k}{T_u} t} \quad \text{with } L_{channel} < \Delta \quad (23)$$

And as the signal received on one of the real antenna is:

$$S_{antenna1}(t_j) = \sum_{k=1}^K H_k^{antenna1} C_k^j e^{j2\pi \frac{k}{T_u} t_j} + \text{target}(t) + b(t) \quad (24)$$

It becomes clear that these two signals could be used to cancel the zero-Doppler paths using the same kind of algorithm than previously but with the important following modification:

- The previous adaptive angular filter was using only real antenna. All these antenna were containing the moving targets contribution as well as some “common” imperfections on the received signal.
- The new suggested filter is involving one real antenna with the target contributions and the signal of reference which is an ideal one. Furthermore, this ideal signal doesn't contain the targets contributions.

5.2.3 Comparison of the two filters

5.2.3.1 Introduction

The previous comments dealing with the main differences in the signals used at the input of the two zero-Doppler cancellation filters allow us to have the following observations:

- The filter involving only real antenna:
 - Requires several receiving antenna to cancel the zero-Doppler paths. Nevertheless a few antenna system could be sufficient as only one degree of freedom is required for cancelling all the zero-Doppler paths below the guard interval.
 - will be more robust to the imperfections as all the antennas suffer from that nuisances and the cancellation filter will be able (at least in a first order) to deal with most of these troubles
 - will have potential influence (and losses) on the targets contributions.
- The filter involving each real antenna and the signal of reference
 - Could be implemented using only one real receiving antenna
 - Will be more sensitive to all the defaults affecting the received signal according to the ideal model used for estimating the reference
 - Will have no influence on the targets contributions as these targets are no longer in the signal of reference obtained after decoding the received one.

In other words, the limitation of these two filters are not identical and it is quite obvious, that the filter involving only real antenna will be more efficient in terms of zero-Doppler cancellation. It is also evident that the consequences on some targets will be higher than with the filter involving real antenna mixed with the reference signal.

5.2.3.2 Example of comparison on experimental data

On the left figure, the adaptive filter was able to cancel efficiently the zero-Doppler paths but the losses on the target were too high due to the vicinity between the target directional vector and the one characterizing the zero-Doppler paths.

On the right figure, the adaptive filter involving one real antenna and the reference signal was also able to cancel efficiently the zero-Doppler path without destructive effect on the target. The small image of the target called ghost (fantôme) was due to a required correction in order to adapt the reference signal model to some imperfections occurring in the receiver system.

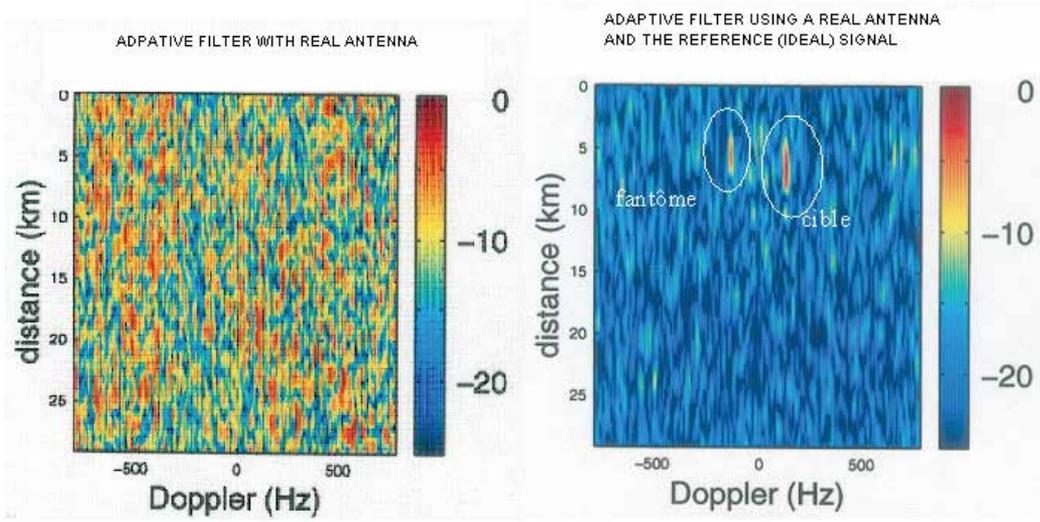


Fig. 12. Example for which the losses on the targets were too high with the adaptive filter involving only real antennas

Nevertheless, generally the method using the reference signal could not be as efficient (in terms of cancellation) as the one involving only real antenna.

5.2.3.3 Example of limitations due to carrier frequency errors.

This short paragraph is just to illustrate the higher sensitivity of the cancellation filter involving the ideal reference signal to one of the possible misfits between the received signal and this “ideal reference”. This illustration will consider a non-corrected error of frequency between the receiver and the transmitter which corresponds to an error between the received signal and the ideal reconstructed reference.

In such a situation, it is possible to consider that this frequency error will lead, for the filter using the reference signal, to an additional interference due to the superposition of the different sinus cardinal functions:

The other filter, involving only real antenna, is less sensitive to such an error as it is the same error on all the antenna used for cancellation.

$$RSI_j = \frac{(H_j C_j)^2}{\left(\sum_{\substack{k=1 \\ k \neq j}}^K \left(H_k C_k \frac{\sin((k-j)\pi + \pi\Delta\nu T)}{(k-j)\pi + \pi\Delta\nu T} \right) e^{j(\pi(k-j) + \pi\Delta\nu T)} \right)^2} \tag{25}$$

If we consider small errors on this frequency carrier, the expression above could be simplified using:

$$I_j = \left(\sum_{\substack{k=1 \\ k \neq j}}^K \left(H_k C_k \frac{\sin((k-j)\pi + \pi\Delta\nu T)}{(k-j)\pi + \pi\Delta\nu T} \right) e^{j(\pi(k-j) + \pi\Delta\nu T)} \right)^2 \tag{26}$$

$$I_j \approx H^2 \left(\sum_{\substack{k=1 \\ k \neq j}}^K \left(C_k \frac{(-1)^{k-j} \Delta\nu T}{(k-j)} \right) \right)^2$$

considering the average power of that perturbation:

$$H^2 C^2 \Delta\nu^2 T^2 \frac{\pi^2}{6} \leq E[I_j] \leq 2H^2 C^2 \Delta\nu^2 T^2 \frac{\pi^2}{6} \tag{27}$$

Finally

$$\frac{1}{2\Delta\nu^2 T^2 \frac{\pi^2}{6}} \leq RSI_j \leq \frac{1}{\Delta\nu^2 T^2 \frac{\pi^2}{6}} \tag{28}$$

The following figure illustrates the influence of an error of 80 Hertz on simulated data. According to the level of the main path (80 dB including the gain of 50 dB for coherent integration time) considered and the useful duration time of 1 millisecond, the troubles due to a misfit between the transmitter frequency and the receiver one become to occur at 20 Hz.

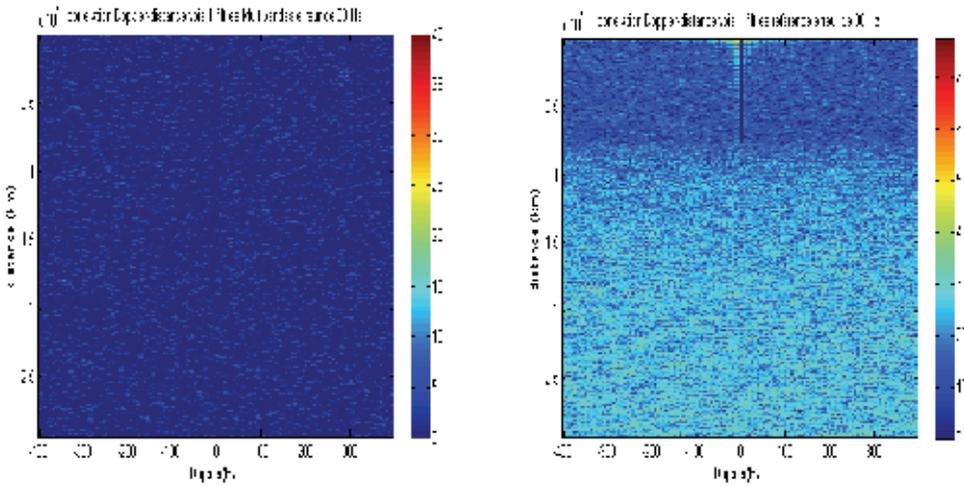


Fig. 13. Correlation output for the two cancellation filter described considered a 80 Hz error between transmitter and receiver (filter with real antenna only: left, filter with real antenna and ideal signal: right)

As illustrated on the following figures, the influence of such an error becomes to occur (according to our simulation parameters) at 20 Hz

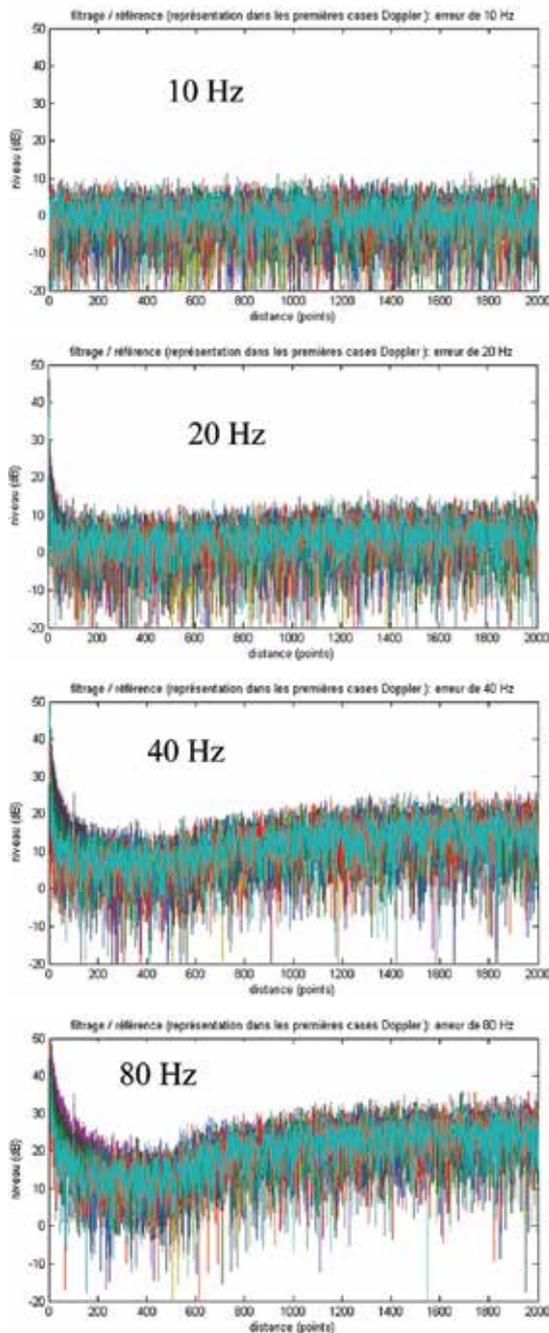


Fig. 14. Analysis of the frequency errors over correlation cut for the cancellation filter involving real antenna and ideal signal.

Of course, it is still possible to define and correct such an error, this example was just an illustration of the higher sensitivity of the second filter to the misfits. Nevertheless, this higher sensitivity will remain even for other kind of interferences that couldn't be corrected as easily as the frequency error...

6. Acknowledgement

We'd like to thanks French MoD (former DGA / DRET and DGA/UM AERO) for his financial support and interest with special thanks to Michel Granger who initiated these works.

7. Conclusion

The COFDM waveform has a great robustness against propagation effects as according to some basic operations (synchronisation and truncation), under the hypothesis of a propagation channel length lower than the guard interval, it is still possible to analyse the received signals over the orthogonal basis of the transmitted sub-carriers even in an environment with numerous reflections.

Using such COFDM civilian broadcasters like DAB or DVB-T as opportunity transmitters for radar application leads to implement a compulsory efficient cancellation filter in order to remove all the main fixed (zero-Doppler) contributors and their corresponding multipaths. Such application is known as Passive Coherent Location: PCL.

Two specific cancellation filters were described in this chapter and illustrated on real data. Their main characteristics are the following:

- The two filters are using the properties of the COFDM modulation in order to "optimise" their efficiencies
- Under the assumption of a propagation channel length lower than the guard interval, only few antenna are necessary in order to lower all the fixed contributors as only one degree of freedom is required for such a cancellation.
- The first method requires a small receiving array (typically 4 or 8 antenna) while the second method could be applied even with one antenna but it implies a higher sensitivity to errors and misfits between the receiving signal and the ideal reconstructed reference
- In practice, these two methods can be complementary as the first one is more efficient for cancelling zero-Doppler paths but it could lower also the targets while the second one is less efficient (due to its higher sensitivity) from the cancellation consideration but it has no destructive effects on the targets.

8. References

- Paul E.Howland, D Maksimiuk and G Reitsma 'FM radio based bistatic radar' IEE Proceedings Radar Sonar and Navigation. Special issue: Passive Radar system Volume 152 Number 3 june 2005 pages 107-115.
- CJ Baker, H D Griffiths and I. Papoutsis 'Passive coherent location radar systems: Part 2: Waveform properties' IEE Proceedings Radar Sonar and Navigation. Special issue: Passive Radar system Volume 152 Number 3 june 2005 pages 160-169.

-
- M.Alard, R.Halbert, R.Lassalle: Principles of modulation and channel coding for digital broadcasting for mobile receivers. EBU review N° 224, August 1987, pp3-25
- R Saini, M.Cherniakov 'DTV signal ambiguity function analysis for radar application' Proceedings Radar Sonar and Navigation. Special issue: Passive Radar system Volume 152 Number 3 June 2005 pages 133-142.
- C Coleman, H Yardley ' Passive bistatic radar based on target illuminations by digital audio broadcasting' IET Radar Sonar and Navigation Volume 2 issue 5 October 2008, pages 366-375
- D Poullin Patent 2 834 072 'Réjection de fouillis dans un récepteur radar passif de signaux OFDM a réseau d'antennes' 26/12/2001
- D Poullin Patent 2 820 507 'Réjection de fouillis dans un récepteur radar passif de signaux OFDM' 07/02/2001

Reliable and Repeatable Power Measurements in DVB-T Systems

Leopoldo Angrisani¹, Domenico Capriglione²,
Luigi Ferrigno² and Gianfranco Miele²

¹*Dept. of Computer Science and Control Systems, University of Naples Federico II
via Claudio 21, 80125 Napoli,*

²*Dept. of Automation, Electromagnetism, Information Engineering and Industrial
Mathematics, University of Cassino,
via G. Di Biasio, 43 03043 Cassino (Fr),
Italy*

1. Introduction

Development and diffusion of digital video broadcasting (DVB) standards have revolutionized the television transmission; whether via satellite (DVB-S), via cable (DVB-C), or terrestrial (DVB-T), the number of services it can offer is able to satisfy the expectation of more demanding customers (ETSI, 2004), (Fischer, 2004). Since many countries in the world suffer from poor coverage of satellite and cable TV, DVB-T is playing a more significant role with respect to the other standards. DVB-T broadcasting networks are, in fact, growing very rapidly. A consequent and pressing need of performance assessment and large scale monitoring of DVB-T systems and apparatuses is thus posed. To reach this goal, a new set of measurements is required and a large number of parameters has to be taken into account, especially due to the complexity characterizing the DVB-T modulation process.

European Telecommunications Standards Institute (ETSI) specifies the parameters and quantities to be measured, and recommends the procedures to be adopted as well as test beds and laboratory equipments to be arranged (ETSI, 2004-2). Power measurement is, in particular, of primary concern: radiofrequency (RF) and intermediate frequency (IF) signal power, noise power, RF and IF power spectrum, should be measured as accurately as possible. Many advantages are connected with this practice, such as better optimization of transmitted power level, thus avoiding waste of energy and reducing the probability of interference with other systems that operate in the same coverage area, and reliable estimation of radiated emissions for verifying compliance limits applied in the regions of interest. Moreover, ETSI suggests the type of instrument to be used for power measurement, such as spectrum analyzer or power meter equipped with a proper sensor and a band-pass filter suitably tuned to the DVB-T frequency band. The former has to be equipped with a specific personality addressed to the integration of the input signal power spectrum on a certain frequency range (channel power measurement), the latter allows only peak and average power to be measured.

Several types of spectrum analyzer and power meter are available on the market. Most of them are general-purpose instruments, and not specifically designed to analyze DVB-T

signals. They exhibit relevant accuracy and repeatability problems in the presence of noise-like signals characterized by high peak to average power ratio (PAR), like DVB-T signals. In addition, they are not suited for large scale monitoring of DVB-T networks, where small size, light weight and low cost are critical constraints.

To give an answer to the cited needs, the scientific community has focused the attention on the definition and implementation of new digital signal processing (DSP) based methods for power measurement in DVB-T systems (Angrisani et al., 2006), (Angrisani et al., 2007), (Angrisani et al., 2008), (Angrisani et al., 2009). In particular, the methods based on power spectral density (PSD) estimators have seemed to be the most appropriate. They exploit straightforward measurement algorithms working on the achieved PSD to provide the desired value of the parameter or quantity of interest. Both non-parametric and parametric estimation algorithms have been considered. An overview of their performance in terms of metrological features, computational burden and memory needs if implemented on a real DSP hardware architecture is given hereinafter.

2. Power measurement in DVB-T systems

For assessing the performance of DVB-T systems and apparatuses, a new set of measurements is required. Many parameters and quantities have, in fact, to be evaluated, pointed out by ETSI in the ETSI TR 101 290 technical report (ETSI, 2004-2), called Digital Video Broadcasting Measurements (DVB-M). ETSI also recommends the procedures to be adopted for arranging test-beds or measurement systems.

A list of the measurement parameters and quantities defined for the DVB-T OFDM environment is shown in Table 1, and full referenced in (ETSI, 2004-2). All of them are keys for evaluating the correct operation of DVB-T systems and apparatuses, and each of them is addressed to a specific purpose. The technical report describes this purpose, where the parameter or the quantity has to be evaluated and in which manner. For the sake of clarity, it reports a schematic block diagram of a DVB-T transmitter and receiver, in which all the measurement interfaces are marked with a letter.

As it can clearly be noted from Table 1, power measurement is of great concern. RF and intermediate frequency (IF) signal power, noise power as well as RF and IF power spectrum are, in fact, relevant quantities to be measured as accurately as possible.

There are several RF power measurement instruments available in the market. They can be divided in two main categories: power meters and spectrum analyzers. Even though suggested by (ETSI, 2004-2), all of them suffer from a number of problems when measuring the power of a noise-like signal with a high PAR, as the DVB-T signal. The problems may dramatically worsen if the measurement is carried out in the field and with the aim of a large scale monitoring.

With regard to power meters, they are typically wideband instruments, and as such they must be connected to one or more calibrated band-pass filters centered at the central frequency of the DVB-T signals to be measured and with an appropriate bandwidth. Moreover, their metrological performance strongly depends on the power sensor they rely on. Several power sensors designed to measure different parameters and characterized by different frequency ranges are available on the market. Even though the choice is wide, not all power sensors are suitable to operate with signals characterized by a high PAR, as explained in (Agilent, 2003).

| <i>Measurement parameter</i> | <i>T</i> | <i>N</i> | <i>R</i> |
|--|----------|----------|----------|
| RF frequency accuracy (precision) | X | | |
| Selectivity | | | X |
| AFC capture range | | | X |
| Phase noise of local oscillators | X | | X |
| RF/IF signal power | X | | X |
| Noise power | | | X |
| RF and IF spectrum | X | | |
| Receiver sensitivity/ dynamic range for a Gaussian channel | | | X |
| Equivalent Noise Degradation (END) | | | X |
| Linearity characterization (shoulder attenuation) | X | | |
| Power efficiency | X | | |
| Coherent interferer | | | X |
| BER vs. C/N ratio by variation of transmitter power | X | | X |
| BER vs. C/N ratio by variation of Gaussian noise power | X | | X |
| BER before Viterbi (inner) decoder | | | X |
| BER before RS (outer) decoder | | | X |
| BER after RS (outer) decoder | | | X |
| I/Q analysis | X | | X |
| Overall signal delay | X | | |
| SFN synchronization | | X | |
| Channel characteristics | | X | |

Table 1. DVB-T measurement parameters and their applicability

Differently from power meters, spectrum analyzers are narrowband instruments, and they are characterized by a more complex architecture. They allow different measurements on different RF signals. Their performance depends on several parameters like the resolution bandwidth (RBW), video bandwidth (VBW), detectors, etc. In particular, the detectors play a very important role because they can emphasize some signal characteristics giving unreliable measurement results. This is especially true when the signals involved are noise-like, as the DVB-T signal. To mitigate this problem, some suggestions described in (Agilent, 2003-2) can be followed.

In many cases, power meters and spectrum analyzers are expressly designed to be used only in laboratories; their performance drastically reduces when used in other environments, especially in the field. But, the fundamental problem that can limit their use is their cost. The total financial investment turns to be prohibitive for any interested company if a great number of instruments is needed, as when a large scale monitoring of DVB-T systems and apparatuses has to be pursued.

3. Nonparametric estimation for power measurement in DVB-T systems

In this chapter the most widely used correlation and spectrum estimation methods belonging to the nonparametric techniques, as well as their properties, are presented. They

do not assume a particular functional form, but allow the form of the estimator to be determined entirely by the data. These methods are based on the discrete-time Fourier transform of either the signal segment (direct approach) or its autocorrelation sequence (indirect approach). Since the choice of an inappropriate signal model will lead to erroneous results, the successful application of parametric techniques, without sufficient a priori information, is very difficult in practice. In the following two major nonparametric algorithms for PSD estimation have been taken into account (Angrisani L. et al., 2003). The first is based on the Welch method of averaged periodograms, which is also known as the WOSA estimator; the second applies wavelet thresholding techniques to the logarithm of the multitaper estimator.

3.1 WOSA Estimator

The WOSA estimator is computationally one of the most efficient methods of PSD estimation, particularly for long data records (Jokinen H. et al., 2000). This method is based on the division of the acquired signal $x(n)$ into smaller units called segments, which may overlap or be disjoint. The samples in a segment are weighted through a window function to reduce undesirable effects related to spectral leakage. For each segment, a periodogram is calculated.

$$S_x^i(f) = \frac{T_s}{N_s U} \left| \sum_{n=0}^{N_s-1} x^i(n) \omega(n) e^{-j2\pi f n T_s} \right|^2 \quad (1)$$

Variable f stands for frequency, $x^i(n)$ are the samples of the i -th segment, $\omega(n)$ accounts for the window coefficients, N_s denotes the number of samples in a segment, U is a coefficient given by

$$U = \frac{1}{N_s} \sum_{n=0}^{N_s-1} \omega^2(n) \quad (2)$$

and is used to remove the window effect from the total signal power, and T_s represents the sampling period. The PSD estimate $S_x(f)$ is then computed by averaging the periodogram estimates

$$S_x(f) = \frac{1}{K} \sum_{i=0}^{K-1} S_x^i(f) \quad (3)$$

where K represents the number of segments and is given by

$$K = \frac{N - N_s}{N_s - N_p} + 1 \quad (4)$$

where N stands for the total number of acquired samples, and N_p is the number of the overlapped samples between two successive segments. Overlap ratio r is defined as the percentage of ratio between the number of the overlapped samples and the number of samples in a segment, i.e.,

$$r = 100 \frac{N_p}{N_s} \% \quad (5)$$

It is worth noting that proper use of the WOSA estimator imposes the optimal choice of two parameters: 1) window function $\omega(\cdot)$ and 2) overlap ratio r . The periodogram in (2) can be easily evaluated over a grid of equally spaced frequencies through a standard fast Fourier transform (FFT) algorithm (Welch P. D. 1967).

3.2 Multitaper estimation and wavelet thresholding

The idea is to calculate a certain number H of PSD estimates, each using a different window function, which is also called data taper and applied to the whole acquired signal, and then to average them together (Moulin P., 1994). If all data tapers are orthogonal, the resulting multitaper estimator can exhibit good performance, in terms of reduced bias and variance, particularly for signals characterized by a high dynamic range and/or rapid variations, such as those that are peculiar to DVB-T systems.

The multitaper estimator has the following form:

$$S_x(f) = \frac{1}{H} \sum_{i=0}^{H-1} S_x^i(f) \quad (6)$$

where the terms $S_x^i(f)$ called eigenspectra are given by

$$S_x^i(f) = \left| \sum_{n=0}^{N-1} x(n) h_i(n) e^{-j2\pi f n T_s} \right|^2 \quad (7)$$

where $\{h_i(n) : n = 0, \dots, N-1; i=1, \dots, H\}$ denotes a set of orthonormal data tapers. A convenient set of easily computable orthonormal data tapers is the set of sine tapers, the i th of which is

$$h_i(n) = \left(\frac{2}{N+1} \right)^{1/2} \sin \left(\frac{(i+1)\pi n}{N+1} \right). \quad (8)$$

A standard FFT algorithm proves to be appropriated in evaluating the eigenspectra over a grid of equally spaced frequencies (Walden et al., 1998).

Provided that H is equal to or greater than 5, it can be demonstrated that random variable $\eta(f)$, as given by

$$\eta(f) = \log \frac{S_x(f)}{S(f)} - \psi(H) + \log H \quad (9)$$

has Gaussian distribution with zero mean and variance σ_η^2 equal to $\psi'(H)$; $S(f)$ represents the true PSD, and $\psi(\cdot)$ and $\psi'(\cdot)$ denote the digamma and trigamma functions, respectively (Moulin P., 1994). If we let

$$Y(f) = \log S_x(f) - \psi(H) + \log H \quad (10)$$

we have

$$Y(f) = \log S(f) + \eta(f) \quad (11)$$

i.e., the logarithm of the multitaper estimator, plus a known constant, can be written as the true log spectrum plus approximately Gaussian noise with zero mean value and known variance σ_η^2 .

These conditions make wavelet thresholding techniques particularly suitable to remove noise and, thus, to produce a smooth estimate of the logarithm of the PSD. In particular, after evaluating the discrete wavelet transform (DWT) of $Y(f)$ computed according to (10), the resulting wavelet coefficients, which are also Gaussian distributed, can be subjected to a thresholding procedure, and the aforementioned smooth estimate can be obtained by applying the inverse DWT to the thresholded coefficients (Walden et al., 1998). A soft threshold function $\delta(\alpha, T)$ is suggested, and it is defined by

$$\delta(\alpha, T) = \text{sgn}(\alpha) \begin{cases} |\alpha| - T, & \text{if } |\alpha| > T \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

where α denotes the generic wavelet coefficient, and T is the threshold level. In (Donoho D. L. & Johnstone I. M., 1994), Donoho and Johnstone demonstrated that, in the presence of Gaussian noise with zero mean value and variance σ_η^2 , the optimal value of T is

$$T = \sigma_\eta \sqrt{2 \times \log N} \quad (13)$$

where N , which is the number of samples, must be of power of two.

In addition, in this case, the right choice of two parameters, i.e., the number of data tapers H and the mother wavelet $\zeta(\cdot)$ for DWT and inverse DWT evaluation, has to be made to gain a sound spectral estimation.

3.3 Performance optimization and assessment

To optimally choose window function $\omega(\cdot)$ and overlap ratio r for the WOSA estimator and the number of data tapers H and mother wavelet $\zeta(\cdot)$ for the multitaper estimator, a suitable simulation stage has been designed. Regarding r , all values ranging from 0% up to 90%, with a step of 10%, have been considered. As for $\omega(\cdot)$, a large set of functions, which differ from one another in relevant spectral characteristics, has been arranged; the set includes most windows defined in (Reljin I. et al., 1998), such as Hanning, Blackman, MS-3FT, MS-4FT, FD-3FT, and FD-4FT, and the new window proposed in (Jokinen H. et al., 2000), which is referred to as Ollila. Concerning H , the considered values range from 5 up to 50, with a step of 5. In addition, various mother wavelets characterized by different vanishing moments (db3, db8, sym3, sym8, coif1, coif5, bior2.2, and bior2.8) have been enlisted (Daubechies I., 1992).

A number of numerical tests have, in particular, been executed in the Matlab 7 environment, with the aim of minimizing the following figures of merit:

1. experimental standard deviation characterizing both total (σ_T) and channel (σ_C) power measurement results;
2. difference between the mean value of the results provided by the method and the imposed value, which is considered as reference, for both total (Δ_T) and channel (Δ_C) power.

The channel power is obtained by integrating the PSD over the frequency interval that is centered at the tune frequency and as wide as the nominal spacing of the channel itself

(ETSI, 2004). Instead the total power is evaluated integrating the PSD over the whole frequency span analyzed from zero up to half of the adopted sample rate f_s ($f_s=1/T_s$). DVB-T reference signals have first been generated. To this aim, the analytical expression for the PSD of a DVB-T signal given by

$$S_x(f) = \sum_{k=-(K-1)/2}^{(K-1)/2} \left[\frac{\sin(\pi(f-f_k)(\Delta+T_u))}{\pi(f-f_k)(\Delta+T_u)} \right]^2 \quad f_k = f_c + \frac{k}{T_u} \quad (14)$$

has been considered, where f_c is the RF signal central frequency, K is the number of transmitted carriers, Δ is the duration of the guard interval, and T_u is the time duration of the useful part of a DVB-T symbol (the useful part does not include the guard interval) (ETSI, 2004). Moreover, the approximate method in the frequency domain presented in (Percival D. B., 1992) has been adopted. It assures accurate time-domain realizations of a zero-mean Gaussian process, which is characterized by a known PSD.

The following DVB-T transmission settings have been imposed: 8K transmission mode ($K=6817$ and $T_u=896 \mu\text{s}$) and $1/4$ ($\Delta=224 \mu\text{s}$) and $1/32$ ($\Delta=28 \mu\text{s}$) guard intervals. In addition, three values of the oversampling factor (considered as the ratio between the sample rate and the RF signal central frequency) have been simulated, and the hypothesis of the acquired records covering one DVB-T symbol has been held. For each transmission setting and oversampling factor value, 50 different realizations (test signals) have been produced.

The obtained results are given in Tables 2 and 3 for the multitaper and WOSA estimators, respectively. Each pair of round brackets describes the couple ($\zeta(\cdot) - H$ or $\omega(\cdot) - r$) that minimizes the related figure of merit. The last row of both tables quantifies the computation burden in terms of mean processing time on a common Pentium IV computer.

From the analysis of the results, some considerations can be drawn.

- Both estimators have assured good repeatability; the experimental standard deviation is always lower than 0.20%.
- Repeatability improves upon the widening of the guard interval, and the oversampling factor seems to have no influence.
- The WOSA estimator exhibits better performance in terms of Δ_T and Δ_C .
- Measurement time peculiar to the multitaper estimator is much longer than that taken by the WOSA estimator.

The WOSA estimator has given a better trade-off between metrological performance and measurement time, thus confirming the outcomes presented in (Angrisani L. et al., 2006). This is the reason the multitaper estimator has no longer been considered in the subsequent stages of the work.

To fix the minimum hardware requirements of the data acquisition system (DAS) to be adopted in the experiments on emulated and actual DVB-T signals described in the succeeding sections, further tests have been carried out. The sensitivity of the proposed method to the effective number of bits (ENOB) and acquired record length has been assessed. The obtained results are given in Figs. 1 and 2; they refer to a guard interval equal to $224 \mu\text{s}$. In particular, Fig. 1 shows the values of σ_C [Fig. 1(a)], Δ_C [Fig. 1(b)], and Δ_T [Fig. 1(c)] versus ENOB for three values of the oversampling factor; Fig. 1(d) presents the estimated PSD for the considered values of ENOB. With regard to σ_T , values very similar to those characterizing σ_C have been experienced. Fig. 2 shows the values of σ_T [Fig. 2(a)] and Δ_T [Fig. 2(b)] versus the acquired record length for the same values of the oversampling factor. With regard to σ_C and Δ_C , values very similar to those characterizing σ_T and Δ_T , respectively, have been experienced.

| Figure of merit | Guard interval [μs] | Oversampling factor | | |
|-------------------------|-------------------------------------|------------------------|------------------------|------------------------|
| | | ~ 3 | ~ 6 | ~ 12 |
| σ_T [%] | 28 | 0,148 (sym8,10) | 0,176 (db3,25) | 0,151 (db3,25) |
| | 224 | 0,129 (coif1,50) | 0,097 (coif1,50) | 0,117 (db3,25) |
| σ_C [%] | 28 | 0,148 (sym8,10) | 0,176 (db3,25) | 0,151 (db3,25) |
| | 224 | 0,129 (coif1,50) | 0,097 (coif1,50) | 0,117 (db3,25) |
| Δ_T [%] | 28 | 0,1104 (bior2.8,50) | 0,1252 (bior2.8,50) | 0,6335 (bior2.8,50) |
| | 224 | 0,1509 (bior2.8,50) | 0,1050 (bior2.8,50) | 0,1237 (bior2.8,50) |
| Δ_C [%] | 28 | 0,1105 (bior2.8,50) | 0,1253 (bior2.8,50) | 0,6336 (bior2.8,50) |
| | 224 | 0,1509 (bior2.8,50) | 0,1050 (bior2.8,50) | 0,1238 (bior2.8,50) |
| Measurement time [s] | 28 | 12,55 | 59,46 | 280,53 |
| | 224 | 53,61 | 280,58 | 1168,20 |

Table 2. Results obtained in the simulation stage: multitaper estimator is involved.

| Figure of merit | Guard interval [μs] | Oversampling factor | | |
|-------------------------|-------------------------------------|-------------------------|-------------------------|----------------------|
| | | ~ 3 | ~ 6 | ~ 12 |
| σ_T [%] | 28 | 0,149 (Ollila,70) | 0,181 (Ollila,50) | 0,148 (Ollila,30) |
| | 224 | 0,130 (blackman,60) | 0,098 (hanning,50) | 0,092 (Ollila,50) |
| σ_C [%] | 28 | 0,149 (Ollila,70) | 0,181 (Ollila,50) | 0,148 (Ollila,30) |
| | 224 | 0,130 (blackman,60) | 0,098 (hanning,50) | 0,092 (Ollila,50) |
| Δ_T [%] | 28 | 0,0015 (FD3FT,30) | 5,4091e-4 (FD3FT,30) | 0,0017 (FD4FT,60) |
| | 224 | 0,0068 (blackman,40) | 0,0028 (MS3FT,10) | 0,0020 (MS4FT,60) |
| Δ_C [%] | 28 | 0,0015 (FD3FT,30) | 6,0611e-4 (FD3FT,30) | 0,0012 (FD4FT,60) |
| | 224 | 0,0068 (blackman,40) | 0,0017 (MS3FT,10) | 0,0020 (MS4FT,60) |
| Measurement time [s] | 28 | 0,032 | 0,052 | 0,096 |
| | 224 | 0,053 | 0,094 | 0,177 |

Table 3. Results obtained in the simulation stage: WOSA estimator is involved.

Looking at Fig. 1, it is possible to establish that 1) an ENOB equal to or greater than six grants an experimental standard deviation in both total (σ_T) and channel (σ_C) power measurements of less than 0.15%, and 2) Δ_C does not seem to be affected by vertical quantization, as, on the contrary, Δ_T does. Furthermore, Fig. 2 clearly evidences that σ_T improves upon the widening of the record length, whereas satisfying values of Δ_T can be achieved if the record lengths covering greater than one half of the DVB-T symbol are considered.

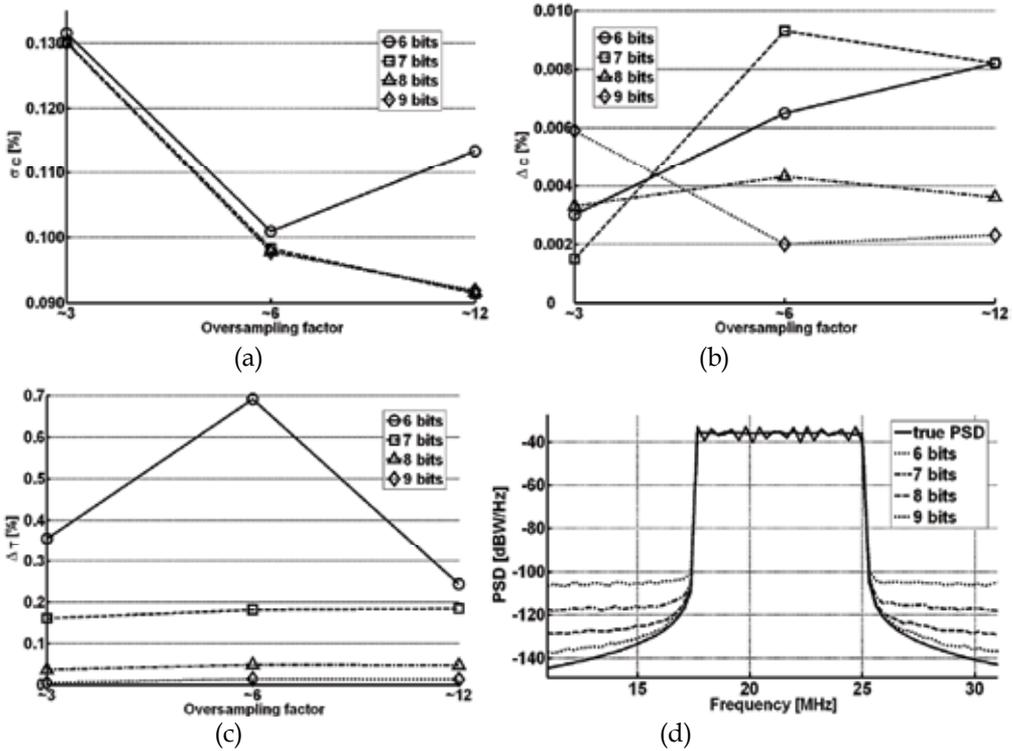


Fig. 1. Simulation stage: a) σ_C , b) Δ_C , and c) Δ_T versus ENOB for three values of the oversampling factor; d) estimated PSD for the considered values of ENOB.

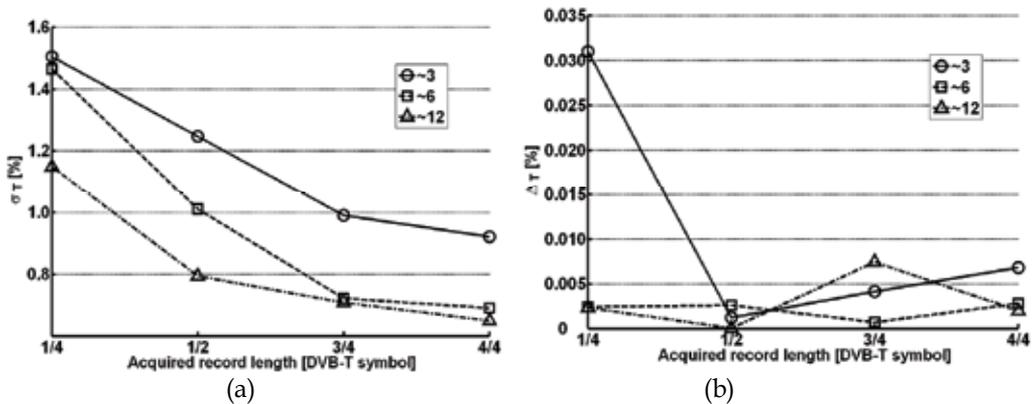


Fig. 2. Simulation stage: a) σ_T and b) Δ_T versus acquired record length for three values of the oversampling factor.

These considerations match well with the typical characteristics of the data acquisition systems available on the market today. High values of the sample rate, required to optimally acquire RF or IF DVB-T signals, are often associated with ENOB not lower than 6 bits. Further an emulation stage has been designed and applied, with the aim of assessing the performance of the proposed method in the presence of a real DAS and comparing it with that assured by competitive measurement solutions that are already available on the market. Stemming from past experience documented in (Angrisani L. et al., 2006), a suitable measurement station, which is sketched in Fig. 3, has been set up. It has included the following: 1) a processing and control unit, i.e., a personal computer, on which the measurement algorithm has run; 2) an RF signal generator equipped with DVB-T personalities Agilent Technologies E4438C (with an output frequency range of 250 kHz–6 GHz); 3) a traditional spectrum analyzer [express spectrum analyzer (ESA)] Agilent Technologies E4402B (with an input frequency range of 9 kHz–3 GHz); 4) a VSA Agilent Technologies E4406A (with an input frequency range of 7 MHz–4 GHz); 5) a real-time spectrum analyzer (RSA) Tektronix RSA3408A (with an input frequency range of dc–8 GHz); 6) an RF power meter (PM) Agilent Technologies N1911A equipped with two probes N1921A (with an input frequency range of 50 MHz–18 GHz) and E9304A (with an input frequency range of 6 kHz–6 GHz); and 7) a DAS LeCroy SDA6000A (with 6-GHz bandwidth and 20-GS/s maximum sample rate). They are all interconnected through an IEEE-488 interface bus. The function generator has provided 8-MHz-bandwidth DVB-T test signals characterized by an RF central frequency equal to 610 MHz, a nominal total power of -20 dBm, and a 64-state quadrature amplitude modulation (QAM) scheme. Moreover, the same transmission settings considered in the previous stage have been imposed.

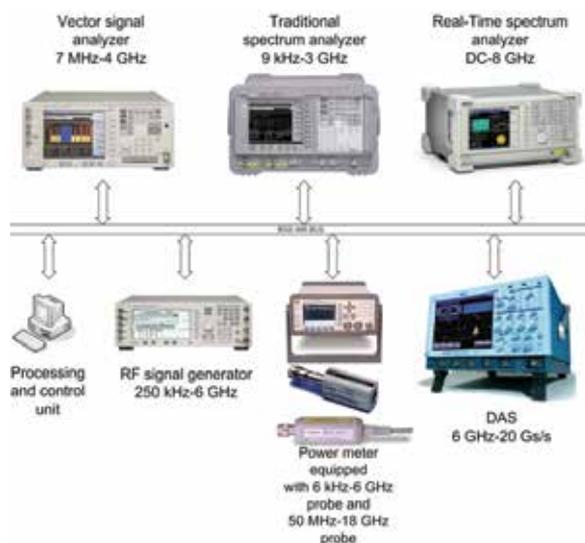


Fig. 3. Measurement station for performance assessment.

A preliminary characterization of cables and connectors utilized in the measurement station has been carried out through the vector network analyzer ANRITSU 37347C (with an input frequency range of 40 MHz–20 GHz), which is equipped with a 3650 SMA 3.5-mm calibration kit (Anritsu, 2003). The mean value and experimental standard deviation of 100 attenuation measures obtained in the interval of 606–614 MHz are given in Table 4.

| | Mean attenuation | Experimental standard deviation |
|--------------------|------------------|---------------------------------|
| Power meter | 0.829150 | 0.000039 |
| Spectrum analyzers | 0.834860 | 0.000019 |
| Oscilloscope | 0.834140 | 0.000014 |

Table 4. Characterization results of cables and connectors utilized in the measurement station of Fig. 3.

Different operative conditions of the DAS, in terms of vertical resolution (7 and 8 bits nominal) and observation period (1/4, 1/2, 3/4, and 1 DVB-T symbol), have been considered. For each operative condition and transmission setting, 50 sample records have been acquired and analyzed through the proposed method. Examining the obtained results given in Table 5 and Fig. 4, it can be noted that two conditions hold.

1. Higher sampling factors do not seem to affect the method's metrological performance; the same is true if vertical resolution is considered.
2. Performance enhancement can be noticed both in the presence of acquired records covering increasingly longer observation periods.

Successively, 50 repeated measurements of total and channel power have been executed by means of PM and spectrum analyzers (ESA, VSA, and RSA), respectively. Table 6 accounts for the results provided by the PM, whereas Table 7 enlists those that are peculiar to the analyzers. As an example, Fig. 5 sketches a typical PSD estimated by the proposed method [Fig. 5(a)], ESA [Fig. 5(b)], VSA [Fig. 5(c)], and RSA [Fig. 5(d)].

With regard to total power, three considerations can be drawn.

1. Results furnished by the PM are different for the two probes adopted.
2. Experimental standard deviation peculiar to the PM is slightly better than that assured by the proposed method.
3. PM outcomes concur with the total power measurement results of the proposed method; a confidence level equal to 99% is considered (Agilent, 2005).

As for the channel power, it is worth stressing that two conditions hold.

1. The proposed method exhibits satisfying repeatability. The related experimental standard deviation is better than that characterizing ESA, VSA, and RSA results.

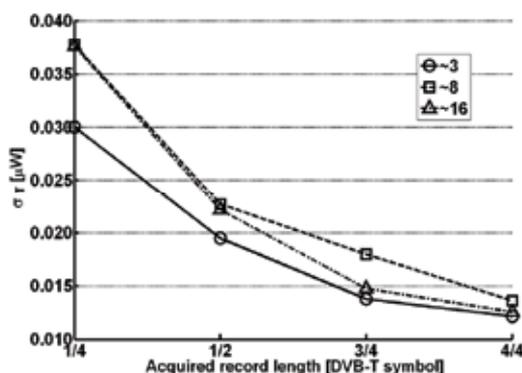


Fig. 4. Emulation stage: σ_T versus acquired record length for three values of the oversampling factor.

| 8k transmission mode, 64-QAM modulation scheme, 8 bit vertical resolution | | | | |
|---|---------------------------|---------------------|----------|-----------|
| | | Oversampling factor | | |
| Figure of Merit | Guard Interval [μ s] | ~ 3 | ~ 8 | ~ 16 |
| σ_T [μ W] | 28 | 0.012 | 0.014 | 0.013 |
| | 224 | 0.0094 | 0.017 | 0.011 |
| σ_C [μ W] | 28 | 0.012 | 0.014 | 0.013 |
| | 224 | 0.0094 | 0.017 | 0.011 |
| P_T [μ W] | 28 | 9.931 | 10.024 | 9.937 |
| | 224 | 10.142 | 10.163 | 10.144 |
| P_C [μ W] | 28 | 9.890 | 9.989 | 9.895 |
| | 224 | 10.1030 | 10.125 | 10.105 |

| 8k transmission mode, 64-QAM modulation scheme, 7 bit vertical resolution | | | | |
|---|---------------------------|---------------------|----------|-----------|
| | | Oversampling factor | | |
| Figure of Merit | Guard Interval [μ s] | ~ 3 | ~ 8 | ~ 16 |
| σ_T [μ W] | 28 | 0.011 | 0.034 | 0.014 |
| | 224 | 0.011 | 0.029 | 0.015 |
| σ_C [μ W] | 28 | 0.011 | 0.032 | 0.014 |
| | 224 | 0.011 | 0.028 | 0.016 |
| P_T [μ W] | 28 | 10.148 | 10.162 | 10.157 |
| | 224 | 10.079 | 10.098 | 10.097 |
| P_C [μ W] | 28 | 9.971 | 9.985 | 9.980 |
| | 224 | 9.899 | 9.919 | 9.916 |

Table 5. Total and channel power measures provided by the proposed method. The acquired record covers a single DVB-T symbol.

| Transmission Settings 8k, 64-QAM, 610 MHz central frequency | | | |
|---|---------------------------|---------------------|--------------------------|
| PM | Guard Interval [μ s] | P_{PM} [μ W] | σ_{PM} [μ W] |
| N1921A PROBE | 28 | 9.9444 | 0.0018 |
| | 224 | 9.9630 | 0.0020 |
| E9304A PROBE | 28 | 8.0402 | 0.0060 |
| | 224 | 7.95910 | 0.00086 |

Table 6. Mean values (PPM) and experimental standard deviations (σ_{PM}) of total power measures provided by the PM equipped with N1921A and E9304A probes.

2. ESA, VSA, and RSA outcomes concur with the channel power measurement results of the proposed method; a confidence level equal to 99% is considered (Agilent, 2004), (Agilent, 2001), (Tektronix, 2006).

Finally, a number of experiments on real DVB-T signals have been carried out through the optimized method. The signals have been radiated by two MEDIASET DVB-T multiplexers operating on the UHF 38 (610-MHz RF central frequency) and UHF 55 (746-MHz RF central frequency) channels, respectively.

A simplified measurement station, as sketched in Fig. 6, has been adopted. With respect to that used in the emulation stage, the function generator has been replaced by a suitable amplified antenna, the VSA and RSA have been removed, and a power splitter has been added. Cables, connectors, and a power splitter have been characterized through the

| Transmission Settings: 8k, 64-QAM, 610 MHz central frequency | | | | | |
|--|--------------------|---------------------------|---------------------|--------------------------|-------|
| Instrument | RBW [kHz] | Guard Interval [μ s] | P_{SA} [μ W] | σ_{SA} [μ W] | |
| ESA | 100 | 28 | 10.322 | 0.074 | |
| | 100 | 224 | 10.656 | 0.080 | |
| | 30 | 28 | 10.376 | 0.068 | |
| | 30 | 224 | 10.142 | 0.070 | |
| VSA | 0.871 | 28 | 10.506 | 0.036 | |
| | 0.871 | 224 | 10.218 | 0.023 | |
| | 30 | 28 | 10.162 | 0.099 | |
| | 30 | 224 | 9.52 | 0.12 | |
| RSA | SPECTRUM ANALYZERS | 50 | 28 | 9.311 | 0.044 |
| | | 50 | 224 | 9.318 | 0.042 |
| | | 30 | 28 | 9.158 | 0.041 |
| | | 30 | 224 | 9.042 | 0.044 |
| | REAL TIME MODE | | 28 | 9.177 | 0.097 |
| | | | 224 | 9.088 | 0.081 |

Table 7. Mean values (P_{SA}) and experimental standard deviations (σ_{SA}) of channel power measures provided by ESA, VSA and RSA; different settings of their resolution bandwidth have been considered.

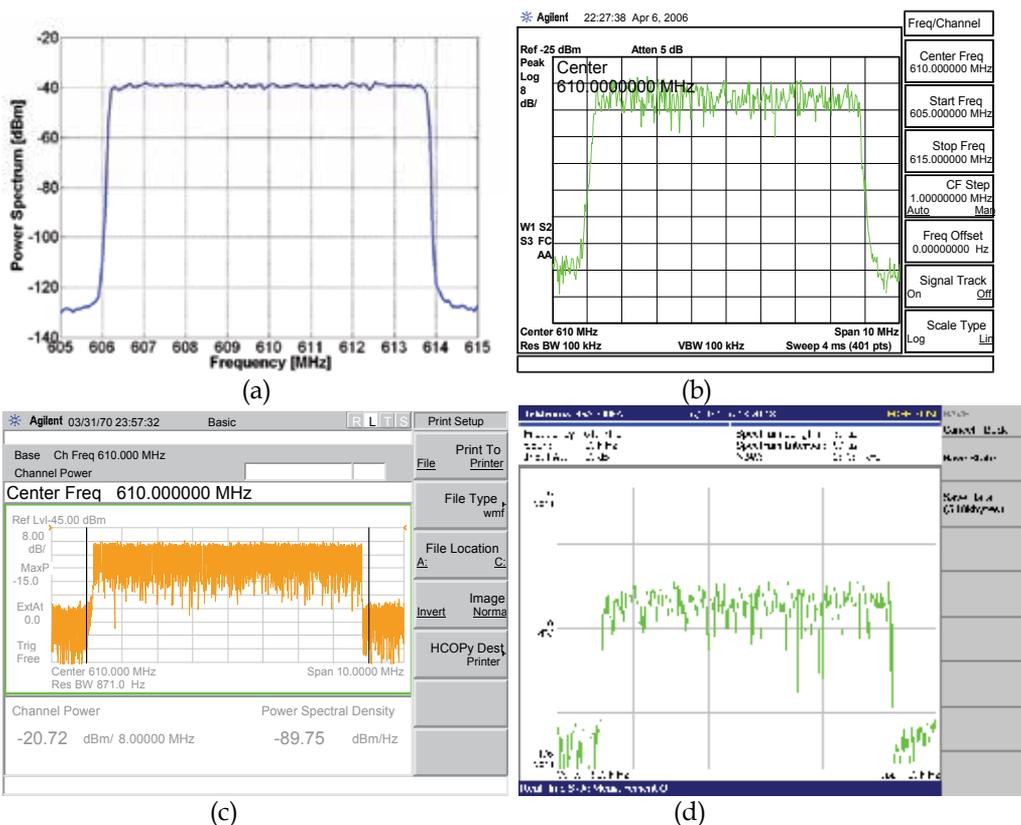


Fig. 5. Power spectrum of an emulated DVB T signal estimated by a) the proposed method, b) ESA, c) VSA and d) RSA.

aforementioned vector network analyzer. The mean value and experimental standard deviation of 100 attenuation measures obtained in the UHF 38 and UHF 55 channels are given in Table 8.

As an example, Fig. 7(a) and (b) shows the power spectrum of a DVB-T signal, which is radiated by the MEDIASET multiplexer operating on UHF 55, as estimated by the proposed method and ESA, respectively. Channel power measurement results are summarized in Table 9; good agreement can be appreciated, confirming the efficacy of the proposal (Angrisani L. et al., 2008).

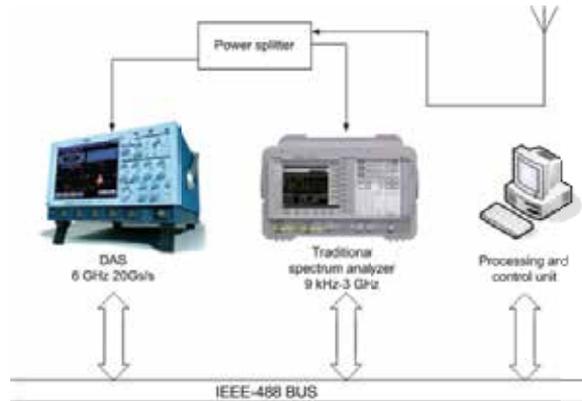


Fig. 6. Measurement station for the experiments on real DVB-T signals.

| | UHF Channel | Mean attenuation [dB] | Experimental standard deviation [dB] |
|-------------------------------|-------------|-----------------------|--------------------------------------|
| DAS | 38 | -4.703 | 0.032 |
| | 55 | -5.403 | 0.042 |
| Traditional Spectrum Analyzer | 38 | -19.393 | 0.021 |
| | 55 | -19.3886 | 0.0086 |

Table 8. Characterization results of cables and connectors utilized in the measurement station of Fig. 6.

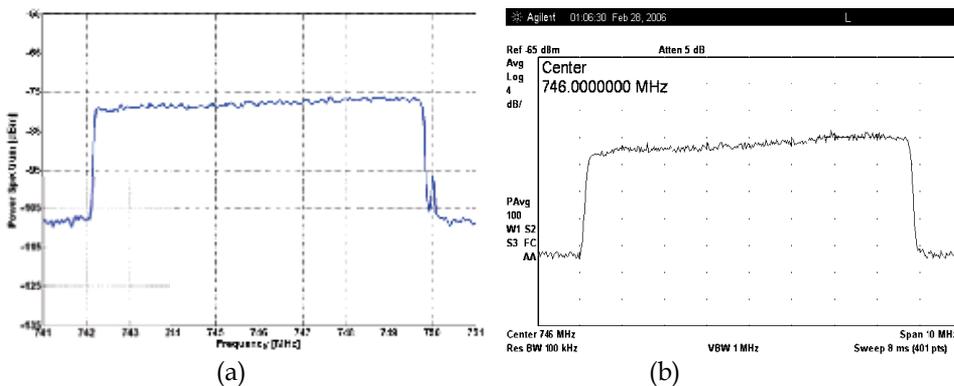


Fig. 7. Power spectrum of a real DVB T signal measured by the a) proposed method and b) ESA.

| 8k transmission mode, 64-QAM, 28μs guard interval | | |
|---|---------------------------|---------------------------|
| | UHF Channel 38 610 MHz | UHF Channel 55 746 MHz |
| Proposed Method | 90.94 nW | 93.07 nW |
| Traditional Spectrum Analyzer | 94.06 nW | 93.23 nW |

Table 9. Experimental results.

4. Parametric estimation for power measurement in DVB-T systems

Parametric estimation methods suppose that the analyzed signal is the output of a model, which is represented as a linear system driven by a noise sequence ϵ_n . They evaluate the PSD of the signal by estimating the parameters (coefficients) of the linear system that hypothetically “generates” the signal. Among the various methods, autoregressive (AR) approaches are widespread. The computational burden related to AR approaches is, in fact, significantly less than that required to implement moving average (MA) or autoregressive moving average (ARMA) parameter estimation algorithms (Marple, 1980).

A stationary autoregressive process of order p , i.e., AR(p), satisfies

$$x_n = -\sum_{m=1}^p a_{p,m} x_{n-m} + \epsilon_n \tag{15}$$

where $a_{p,1}, a_{p,2}, \dots, a_{p,p}$ are fixed coefficients, and $\{\epsilon_n\}$ is a white noise process with variance σ_p^2 . The PSD of the stationary process described by AR(p) is totally described by the model parameters and the variance of the white noise process. It is given by

$$S(f) = \frac{\sigma_p^2 T_s}{\left| 1 + \sum_{m=1}^p a_{p,m} e^{-j2\pi m f T_s} \right|^2} \quad |f| \leq f_N \tag{16}$$

where $T_s = 1/f_s$ is the sampling interval, and $f_N = 1/(2T_s)$ is the Nyquist frequency. Consequently, with known p , it is necessary to properly estimate the $p+1$ parameters $a_{p,1}, a_{p,2}, \dots, a_{p,p}$ and σ_p^2 . To reach this goal, the relationship between the AR parameters and the autocorrelation sequence (known or estimated) of x_n has to be fixed, as described here.

4.1 Yule–Walker equations

Achieving the expectations on the product $x_n x_{n-k}^*$, the autocorrelation sequence is evaluated as

$$R_{xx}(k) = E[x_n x_{n-k}^*] = -\sum_{m=1}^p a_{p,m} R_{xx}(k-m) + E[\epsilon_n x_{n-k}^*]. \tag{17}$$

The plausible fact that $E[\epsilon_n x_{n-k}^*] = 0$, for $k > 0$, implies that

$$\begin{aligned} E[\epsilon_n x_n^*] &= E\left[\epsilon_n \left(-\sum_{m=1}^p a_{p,m}^* x_{n-m} + \epsilon_n\right)\right] = \\ &= -\sum_{m=1}^p a_{p,m}^* E[\epsilon_n x_{n-m}] + \sigma_p^2 = \sigma_p^2 \end{aligned} \tag{18}$$

Hence, the evaluation of (18) for $k=0,1,\dots,p$ makes it possible to obtain the so-called augmented Yule-Walker equations

$$\begin{bmatrix} R_{xx}(0) & R_{xx}(-1) & L & R_{xx}(-p) \\ R_{xx}(1) & R_{xx}(0) & L & R_{xx}(-p+1) \\ M & M & O & M \\ R_{xx}(p) & R_{xx}(p-1) & L & R_{xx}(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_{p,1} \\ M \\ a_{p,p} \end{bmatrix} = \begin{bmatrix} \sigma_p^2 \\ 0 \\ M \\ 0 \end{bmatrix} \quad (19)$$

If we have no stationary process $\{x_n\}$ but we are in the presence of a time series that is a realization of a portion x_1, x_2, \dots, x_N of any discrete-parameter stationary process, replacing $R_{xx}(k)$ with

$$\hat{R}_{xx}(k) = \frac{1}{N} \sum_{i=0}^{N-k} x_i x_{i+k}^* \quad \text{for } k=0, \dots, p \quad (20)$$

it is possible to solve system (19) by inversion.

4.2 Levinson-Durbin algorithm

To avoid the matrix inversion, which is a time-consuming task and is performed using Gaussian elimination, that requires operations of order p^3 , which are denoted as $o(p^3)$, the system (19) can be solved through Levinson-Durbin recursions (Kay & Marple, 1981), (Marple, 1980), which require only $o(p^2)$ operations. The algorithm proceeds with recursively computing the AR parameters for order k from the AR parameters previously determined for order $k-1$.

In particular, the recursive algorithm is initialized by

$$a_{1,1} = -\frac{R_{xx}(1)}{R_{xx}(0)} \quad (21)$$

$$\sigma_1^2 = (1 - |a_{1,1}|^2) R_{xx}(0) \quad (22)$$

and the recursion for $k = 2, 3, \dots, p$ is given by

$$a_{k,k} = -\frac{R_{xx}(k) + \sum_{m=1}^{k-1} a_{k-1,m} R_{xx}(k-m)}{\sigma_{k-1}^2} \quad (23)$$

$$a_{k,m} = a_{k-1,m} + a_{k,k} a_{k-1,k-m}^*, \quad 1 \leq m \leq k-1 \quad (24)$$

$$\sigma_k^2 = \sigma_{k-1}^2 (1 - |a_{k,k}|^2) \quad (25)$$

where $a_{k,k}$ is the reflection coefficient (Kay & Marple, 1981). This algorithm is useful when the correct model order is not known a priori since (21)-(25) can be used to successfully generate higher order models until the modeling error σ_k^2 is reduced to a desired value.

4.3 Forward linear prediction algorithm

In the literature, several least-squares estimation procedures that directly operate on the data to yield better AR parameter estimates can be found. These techniques often produce better AR spectra than that obtained with the Yule-Walker approach.

Assume that the sequence x_0, \dots, x_{N-1} is used to find the p -th-order AR parameter estimates. The forward linear predictor is (Makhoul, 1975)

$$\hat{x}_n = -\sum_{k=1}^p a_{p,k} x_{n-k} \tag{26}$$

It is possible now to define the forward linear prediction error

$$e_p(n) = x_n - \hat{x}_n = \sum_{k=0}^p a_{p,k} x_{n-k} \quad \text{for } p \leq n \leq N-1 \tag{27}$$

where $a_{p,0}=1$. Therefore, $e_p(n)$, for $n=p$ to $n=N-1$, can be obtained by

$$\begin{bmatrix} e_p(p) \\ \vdots \\ e_p(N-1) \end{bmatrix} = \begin{bmatrix} x_p & L & x_0 \\ \vdots & & \\ x_{N-1} & L & x_{N-p-1} \end{bmatrix} \begin{bmatrix} 1 \\ a_{p,1} \\ \vdots \\ a_{p,p} \end{bmatrix} \tag{28}$$

where X_p is an $(N-p) \times (p+1)$ Toeplitz matrix.

The approach followed to estimate $a_{p,k}$ consists of minimizing a sum of $e_p(n)$ called prediction error energy, i.e.,

$$SS_p = \sum_{n=p}^{N-1} |e_p(n)|^2 = \sum_{n=p}^{N-1} \left| \sum_{k=0}^p a_{p,k} x_{n-k} \right|^2 = E^H E \tag{29}$$

Using an alternative description of the $N-p$ error equation (28) such as

$$E = \begin{bmatrix} \overbrace{y}^{x_p} \\ \vdots \\ \overbrace{X}^{x_p} \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ a \end{bmatrix} \tag{30}$$

where $y = [x_p, \dots, x_{N-1}]^T$, $a = [a_{p,1}, \dots, a_{p,p}]^T$, and $X = \begin{bmatrix} x_{p-1} & L & x_0 \\ \vdots & & \\ x_{N-2} & L & x_{N-p-1} \end{bmatrix}$ the prediction error

energy (29) may be expressed as

$$SS_p = E^H E = y^H y + y^H X a + a^H X^H y + a^H X^H X a \tag{31}$$

To minimize SS_p , this term must be set to zero (Marple, 1987), i.e.,

$$X^H y + X^H X a = 0_p \tag{32}$$

where 0_p is the all-zeros vector, obtaining

$$SS_{p,\min} = y^H y + y^H X a. \tag{33}$$

Equations (32) and (33) may be combined into a single set of

$$\begin{aligned} \begin{bmatrix} y^H y & y^H X \\ X^H y & X^H X \end{bmatrix} \begin{bmatrix} 1 \\ a \end{bmatrix} &= \begin{bmatrix} y & X \end{bmatrix}^H \begin{bmatrix} y & X \end{bmatrix} \begin{bmatrix} 1 \\ a \end{bmatrix} = \\ &= (X_p)^H X_p \begin{bmatrix} 1 \\ a \end{bmatrix} = \begin{bmatrix} SS_{p,\min} \\ 0_p \end{bmatrix} \end{aligned} \tag{34}$$

These equations form the normal equations of the least squares analysis. This method is called the covariance method (Makhoul, 1975). Due to the particular properties of $(X_p)^H X_p$, it is possible to develop a fast algorithm that is similar to that of the Levinson algorithm. The original fast algorithm for solving the covariance normal equations was developed by Morf et al. (Morf et al, 1977), and further computational reduction was studied by Marple and reported in (Marple, 1987), producing an algorithm that requires $o(p^2)$ operations.

4.4 Burg algorithm

This is the most popular approach for AR parameter estimation with N data samples and was introduced by Burg in 1967 (Burg, 1967). It may be viewed as a constrained least-squares minimization.

The approach followed to estimate $a_{k,k}$ consists of minimizing a sum of forward and backward linear prediction error energies, i.e.,

$$SS_p = \sum_{n=p}^{N-1} \left[|e_p(n)|^2 + |b_p(n)|^2 \right] \tag{35}$$

where $e_p(n)$ is defined by (27), and $b_p(n)$ is the backward linear prediction error, which is given by

$$b_p(n) = \sum_{k=0}^p a_{p,k}^* x_{n-p+k} \quad \text{for } p \leq n \leq N-1. \tag{36}$$

Note that $a_{p,0}$ is defined as unity.

Substitution of (24) into (27) and (36) yields the following recursive relationship between the forward and backward prediction errors:

$$e_p(n) = e_{p-1}(n) + a_{p,p} b_{p-1}(n-1) \quad \text{for } p \leq n \leq N-1 \tag{37}$$

$$b_p(n) = b_{p-1}(n-1) + a_{p,p}^* e_{p-1}(n) \quad \text{for } p \leq n \leq N-1 \tag{38}$$

and substituting (37) and (38) into (35), SS_p can be written as

$$SS_p = \Gamma_p + 2a_{p,p} \Lambda_p + \Gamma_p a_{p,p}^2 \tag{39}$$

whose coefficients are

$$\Gamma_p = \sum_{n=p}^{N-1} \left[|e_{p-1}(n)|^2 + |b_{p-1}(n-1)|^2 \right] \tag{40}$$

$$\Lambda_p = 2 \sum_{n=p}^{N-1} e_{p-1}(n) b_{p-1}^*(n-1). \tag{41}$$

The value of $a_{p,p}$ that minimizes SS_p can easily be calculated by setting the derivative to zero and obtaining

$$a_{p,p} = -\frac{\Lambda_p}{\Gamma_p}. \tag{42}$$

The routine implemented to estimate the AR coefficients is shown in Fig. 2. It needs an initializing step, in which the starting value of the observed forward and backward prediction errors and the innovation variance are chosen using the following relations:

$$e_0(n) = b_0(n) = x_n \tag{43}$$

$$\sigma_0^2 = \frac{1}{N} \sum_{n=1}^N |x_n|^2. \tag{44}$$

The Burg algorithm requires a number of operations proportional to p^2 .

4.5 Forward and backward linear prediction algorithm

This approach, which was independently proposed by Ulrych and Clayton (Ulrych T. J. & Clayton R. W., 1976). and Nuttall (Nuttall A. H., 1976), is a least-squares procedure for forward and backward predictions, in which the Levinson constraint imposed by Burg is removed.

Noting that (27) and (36) can be summarized by

$$\Delta = \begin{bmatrix} E \\ B' \end{bmatrix} = \begin{bmatrix} X_p \\ X_p^* J \end{bmatrix} \begin{bmatrix} 1 \\ a \end{bmatrix} \tag{45}$$

where $B = [b_p(p), \dots, b_p(N-1)]^T$, J is an $(p + 1) \times (p + 1)$ reflection matrix, and $X_p^* J$ is a Hankel matrix of conjugated data elements, it is possible to rewrite (35) as

$$SS_p = \Delta^H \Delta = E^H E + B^H B. \tag{46}$$

The preceding equation can be minimized with the same procedure used for the covariance method, leading to the set of normal equations

$$\overbrace{\begin{bmatrix} X_p \\ X_p^* J \end{bmatrix}^H \begin{bmatrix} X_p \\ X_p^* J \end{bmatrix}}^{R_p} \begin{bmatrix} 1 \\ a \end{bmatrix} = \begin{bmatrix} SS_{p,\min} \\ 0_p \end{bmatrix}. \tag{47}$$

Because the summation range in (35) is identical to that of the covariance method, this least-squares approach is called the modified covariance method. The system (47) can be solved by a matrix inversion that requires a number of operations proportional to p^3 , which is one order of magnitude greater than Burg's solution.

Due to the characteristic of the actual structure of the matrix R_p , Marple (Marple, 1980), (Marple, 1987) suggested an algorithm requiring a number of computations proportional to p^2 .

4.6 Performance optimization and assessment

The performance of parametric power spectrum estimation methods depends on the model order p . To regulate this parameter to operate with success on DVB-T systems, a suitable simulation stage has been designed and set-up. Once the optimal value of p has been found, a first comparison with the optimized Welch method has been made. Successively further investigation has been carried out in simulation environment, with the aim of evaluating the performance of parametric spectrum estimation methods when they are applied to signals characterized by different quantization levels. Afterwards an emulation stage has been designed and applied with the aim of:

- assessing the performance of the proposed method in the presence of a real DAS;
- comparing it to that assured by competitive measurement solutions already available on the market;
- comparing it to that assured by the optimized Welch method.

Moreover a number of experiments on real DVB-T signals have been carried out through the optimized method, in order to make a comparison with the results obtained in the previous stages. At last the suitability of these methods to be implemented in a low cost DSP platform has been investigated.

As said in the previous paragraph, the performance of PSD AR estimators depends on the polynomial order p . To optimally choose this parameter, a suitable simulation stage has been designed. A number of numerical tests have been executed in the Matlab 7 environment, with the aim of minimizing the same figures of merit defined in the previous section. These tests have been carried out by adopting the same reference signals defined above.

With special regard to AR estimation algorithms, taking into account that higher values of p may introduce spurious details in the estimated spectrum and lower values of p may drive to a highly smoothed spectral estimate (Kay & Marple, 1981), a dual stage optimization procedure has been applied. In the first stage, a rough optimization has been pursued; in particular, a suitable operative range for p has been fixed. The second stage has finely tuned the value of p , within the range previously determined, through the minimization of σ_C and Δ_C .

4.6.1 Rough optimization

Suitable figures of merit, which are addressed to highlight the goodness of the PSD estimates, have been considered. Much attention has been paid to the final prediction error, Akaike's information criterion, and the root mean square error (RMSE); details can be found in (Kay & Marple, 1981) and (Angrisani L. et al., 2003).

Concerning p , two different and consecutive sets have been organized: $\Sigma_1 = \{p \mid 10 \leq p \leq 100\}$ and $\Sigma_2 = \{p \mid 100 < p \leq 5000\}$. In Σ_1 , an analysis step of 10 has been adopted, whereas a step of 100 has been considered for Σ_2 .

All tests have highlighted quite the same behavior of the three figures of merit; they have reached their minima in strictly overlapping p ranges. For the sake of brevity, Fig. 8 shows only the minimum value of RMSE [Fig. 8(a)] and the corresponding value of p [Fig. 8(b)]

versus the observation period expressed as a fraction of the time interval associated with one DVB-T symbol; a guard interval of 224 μ s and an oversampling factor of 3 have, in particular, been considered. Very similar outcomes have been attained with a guard interval of 28 μ s and two oversampling factors, which are equal to 6 and 12.

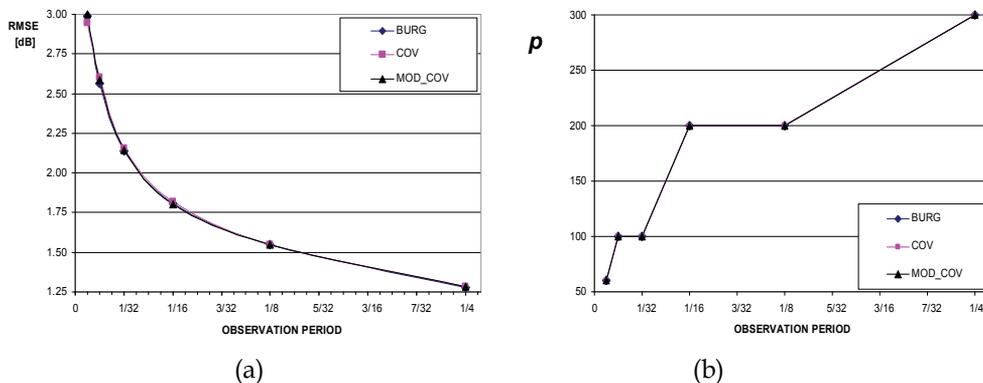


Fig. 8. a) Minimum values of RMSE and (b) corresponding values of p versus the observation period, which is expressed as a fraction of the time interval associated with one DVB-T symbol.

From the analysis of the results, some considerations have emerged.

1. The covariance, Burg, and modified covariance estimators reach the lowest RMSE for very similar values of the polynomial order p .
2. RMSE values related to the covariance, Burg, and modified covariance algorithms concur, showing comparable performance in PSD estimation.
3. The values of p that minimize RMSE are significantly high for observation periods that are longer than 1/128 of the time interval associated with one DVB-T symbol.

To fix an operative range of p of practical use, it has been assumed that RMSE values lower than 3 dB assure acceptable performance in channel power measurement (Fig. 9). A threshold of 3 dB has been applied to the results already obtained, thus achieving a strong reduction in the values of p of interest, with a consequent benefit to the computational burden.

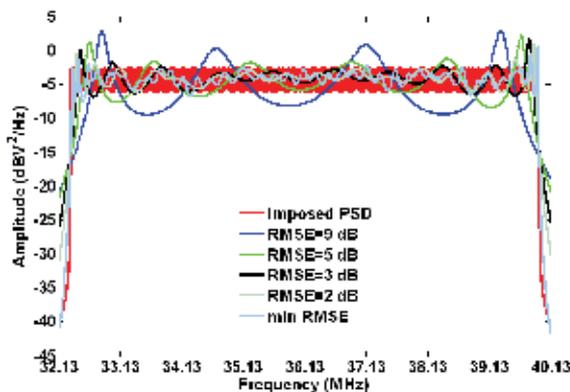


Fig. 9. Estimated PSD versus RMSE.

4.6.2 Fine Optimization

The stage has aimed at fixing the optimal value of p within the operative range established before and comparing the performance granted by the so-optimized covariance, Burg, and modified covariance estimator-based measurement algorithms to that assured by the WOSA-estimator-based algorithm. To reach this goal two figures of merit, σ_C and Δ_C , already defined in paragraph 3, have been minimized.

The obtained values of Δ_C and σ_C and the polynomial order p versus the observation period, which is expressed as a fraction of the time interval associated with one DVB-T symbol, are shown in Fig. 10. An oversampling factor of 3 and a guard interval of 224 μs have been considered. Very similar results have been experienced with a guard interval of 28 μs and two oversampling factors, which are equal to 6 and 12.

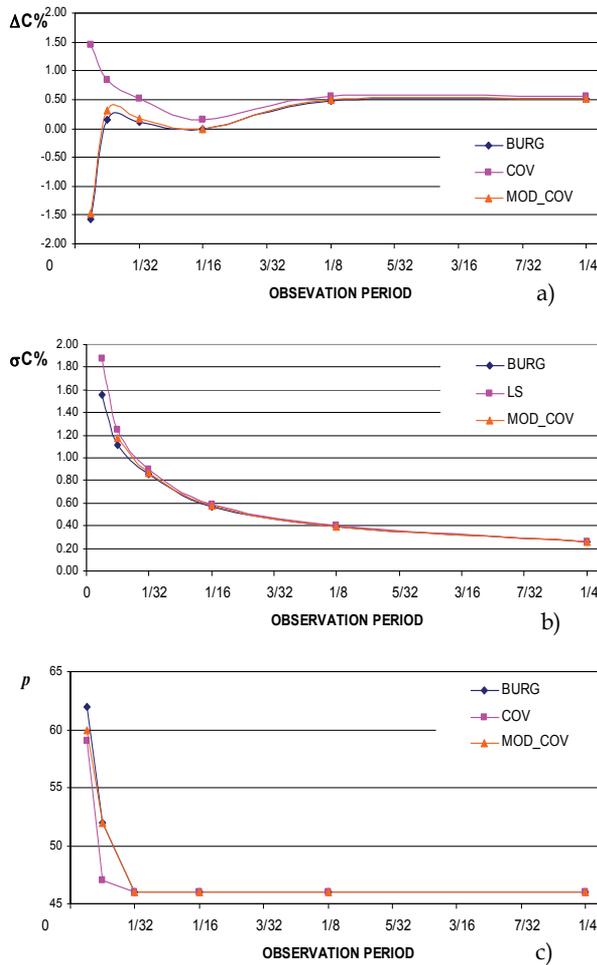


Fig. 10. a) Δ_C , b) σ_C and c) polynomial order p versus the observation period for the considered AR estimator-based measurement algorithms. An oversampling factor equal to 3 and a guard interval equal to 224 μs have been considered.

It is possible to state that the considered AR algorithms grant a very similar performance for both σ_C and Δ_C and that the optimum polynomial order p is equal to 46. In addition, the oversampling factor seems to have no influence; its lowest value (3) is advisable for reducing memory needs.

To fix the minimum hardware requirements of the DAS to be adopted in the experiments on emulated and actual DVB-T signals, the results of which are described in the following, further tests have been carried out. Table 10 gives the estimated σ_C versus the analyzed values of the effective number of bits (ENOB). Observation periods ranging from $1/128$ up to $1/4$ of the time interval associated with one DVB-T symbol have been considered. σ_C does not seem to be affected by vertical quantization, and the Burg algorithm seems to be more stable if short observation periods are involved.

| Estimators | Observation period | ENOB | | | |
|---------------------|--------------------|------|------|------|------|
| | | 6 | 7 | 8 | 9 |
| BURG | $1/128$ | 1.6 | 1.6 | 1.5 | 1.5 |
| | $1/64$ | 1.1 | 1.1 | 1.1 | 1.1 |
| | $1/32$ | 0.87 | 0.86 | 0.86 | 0.86 |
| | $1/16$ | 0.57 | 0.56 | 0.57 | 0.57 |
| | $1/8$ | 0.39 | 0.39 | 0.39 | 0.39 |
| | $1/4$ | 0.26 | 0.26 | 0.26 | 0.26 |
| COVARIANCE | $1/128$ | 1.8 | 1.9 | 2.1 | 1.9 |
| | $1/64$ | 1.3 | 1.3 | 1.3 | 1.2 |
| | $1/32$ | 0.90 | 0.90 | 0.90 | 0.90 |
| | $1/16$ | 0.59 | 0.58 | 0.59 | 0.59 |
| | $1/8$ | 0.40 | 0.40 | 0.40 | 0.40 |
| | $1/4$ | 0.26 | 0.26 | 0.26 | 0.26 |
| MODIFIED COVARIANCE | $1/128$ | 1.7 | 1.7 | 1.7 | 1.7 |
| | $1/64$ | 1.2 | 1.2 | 1.2 | 1.2 |
| | $1/32$ | 0.87 | 0.87 | 0.87 | 0.86 |
| | $1/16$ | 0.58 | 0.58 | 0.58 | 0.58 |
| | $1/8$ | 0.40 | 0.40 | 0.40 | 0.40 |
| | $1/4$ | 0.26 | 0.26 | 0.26 | 0.26 |

Table 10. $\sigma_C\%$ versus ENOB for different observation periods.

Computational burden, in terms of the mean processing time on a common Pentium IV computer, has also been quantified.

The results are given in Table 11. It is possible to note that the measurement time peculiar to the Burg-estimator-based measurement algorithm is lower than that taken by the covariance and modified-covariance-estimator-based algorithms for short observation periods.

The Burg-estimator-based measurement algorithm has shown the best tradeoff between metrological performance and measurement time. This is the reason the covariance- and modified covariance estimator-based algorithms have no longer been considered in the subsequent stages of the work.

| 8k transmission mode and $\Delta=224 \mu\text{s}$ | | | | | | |
|---|--------------------|-------|-------|-------|-------|-------|
| | Observation period | | | | | |
| Estimator | 1/4 | 1/8 | 1/16 | 1/32 | 1/64 | 1/128 |
| Burg | 0.062 | 0.036 | 0.020 | 0.010 | 0.008 | 0.005 |
| Covariance | 0.044 | 0.028 | 0.018 | 0.015 | 0.008 | 0.015 |
| Modified Covariance | 0.052 | 0.035 | 0.024 | 0.021 | 0.022 | 0.023 |
| 8k transmission mode and $\Delta=28 \mu\text{s}$ | | | | | | |
| | Observation period | | | | | |
| Estimator | 1/4 | 1/8 | 1/16 | 1/32 | 1/64 | 1/128 |
| Burg | 0.055 | 0.033 | 0.016 | 0.008 | 0.006 | N/A |
| Covariance | 0.039 | 0.027 | 0.017 | 0.015 | 0.014 | N/A |
| Modified Covariance | 0.046 | 0.040 | 0.022 | 0.019 | 0.020 | N/A |

Table 11. Computation time in ms versus the observation period.

An emulation stage has been designed and executed with the aim of assessing the performance of the optimized Burg estimator-based measurement algorithm in the presence of a real DAS and comparing the obtained results to those furnished by the optimized WOSA algorithm. Moreover, all results have been compared to those assured by competitive measurement solutions already available on the market.

Thanks to the experience described in the paragraph 3, a suitable measurement station, sketched in Fig. 11, has been designed and setup.

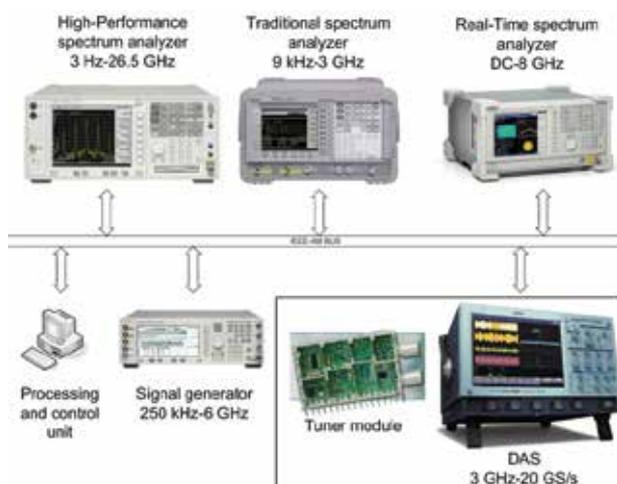


Fig. 11. Measurement station.

It has included:

- a control unit, namely a personal computer (PC);
- a RF signal generator Agilent Technologies E4438C (250 kHz-6 GHz output frequency range), equipped with DVB-T personalities;
- an express spectrum analyzer (ESA) Agilent Technologies E4402B (9 kHz-3 GHz input frequency range);

- a high performance spectrum analyzer (PSA) Agilent Technologies E4440A (3 Hz-26.5 GHz input frequency range);
- a real-time spectrum analyzer (RSA) Tektronix RSA3408A (DC-8 GHz input frequency range);
- a DAS LeCroy WavePro 7300A, (3 GHz bandwidth, 20 GS/s maximum sample rate) coupled to the tuner module for digital terrestrial application described in paragraph 3.

All instruments have been interconnected through an IEEE-488 standard interface bus. The signal generator has provided 8 MHz bandwidth, DVB-T test signals, characterized by a RF center frequency equal to 610 MHz, a nominal total power of -10 dBm and a 64-QAM modulation scheme. Moreover, the same transmission settings considered in the previous stage have been imposed.

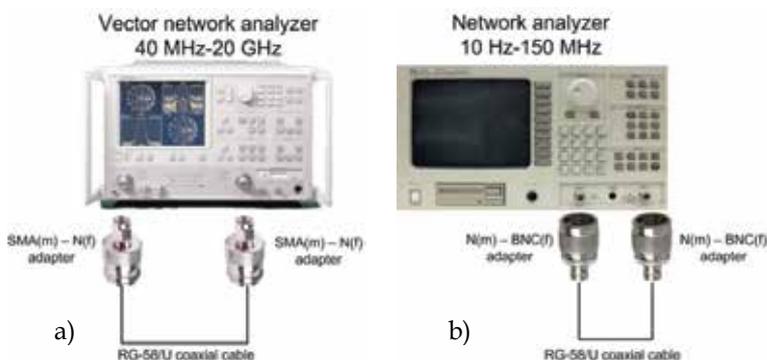


Fig. 12. Measurement bench for the characterization of cables and connectors at a) RF, and b) IF.

A preliminary characterization of cables and connectors utilized in the measurement station has been carried out through the vector network analyzer ANRITSU 37347C (40 MHz-20 GHz input frequency range), equipped with 3650 SMA 3.5 mm calibration kit (Anritsu, 2003), and the spectrum/network analyzer HP 3589A (10 Hz--150 MHz input frequency range) (Agilent, 1991), respectively for RF and IF frequencies. Also the tuner has been characterized.

Different operative conditions of the DAS, in terms of vertical resolution (7- and 8-bit nominal) and observation period (from 1/128 up to 1/4 of the time interval associated with one DVB-T symbol), have been considered; the oversampling factor has been chosen to be equal to 3. For each of them, 100 sample records have been acquired and analyzed both through the Burg- and WOSA-estimator-based measurement algorithms.

The obtained results, which are given in Tables 12-14, have highlighted five conditions.

1. The channel power measures provided by the Burg estimator- based measurement algorithm concur with those furnished by the WOSA-estimator-based algorithm.
2. The channel power measures are influenced by the DAS vertical resolution for both the Burg- and WOSA estimator- based measurement algorithms.
3. Both algorithms exhibit satisfying and comparable repeatability, which is not affected by the DAS vertical resolution and observation period.
- 4) ESA and PSA outcomes concur with the channel power measurement results of the Burg- and WOSA estimator-based measurement algorithms when a DAS resolution of 8 bits is adopted; a confidence level of 95% is considered.

5. The outcomes of the RSA operating both in normal conditions and as a spectrum analyzer seem to concur with the channel power measurement results of the Burg- and WOSA estimator-based algorithms only for a DAS resolution of 7 bits; a confidence level of 99% is considered.

| 8 bit DAS resolution | | | | | | | |
|--------------------------|---------------------------|--------------------|--------|--------|--------|--------|--------|
| | | Observation period | | | | | |
| Figure of merit | Guard Interval [μ s] | 1/4 | 1/8 | 1/16 | 1/32 | 1/64 | 1/128 |
| P_C [μ W] | 28 | 106.41 | 106.39 | 106.38 | 106.52 | 105.72 | N/A |
| | 224 | 104.87 | 104.83 | 104.71 | 104.68 | 104.08 | 102.57 |
| σ_{PC} [μ W] | 28 | 0.75 | 0.75 | 0.76 | 0.77 | 0.78 | N/A |
| | 224 | 0.74 | 0.74 | 0.74 | 0.75 | 0.75 | 0.77 |
| 7 bit DAS resolution | | | | | | | |
| | | Observation period | | | | | |
| Figure of merit | Guard Interval [μ s] | 1/4 | 1/8 | 1/16 | 1/32 | 1/64 | 1/128 |
| P_C [μ W] | 28 | 97.55 | 97.43 | 97.31 | 97.41 | 96.52 | N/A |
| | 224 | 97.35 | 97.33 | 97.12 | 97.31 | 96.59 | 94.92 |
| σ_{PC} [μ W] | 28 | 0.69 | 0.69 | 0.69 | 0.70 | 0.71 | N/A |
| | 224 | 0.71 | 0.71 | 0.71 | 0.72 | 0.72 | 0.73 |

Table 12. Mean values (P_C) and experimental standard deviations (σ_{PC}) of channel power measures provided by the WOSA estimator-based measurement algorithm. DVB-T settings: 8k transmission mode, 64-QAM modulation scheme.

| 8 bit DAS resolution | | | | | | | |
|--------------------------|---------------------------|--------------------|--------|--------|--------|--------|--------|
| | | Observation Period | | | | | |
| Figure of merit | Guard Interval [μ s] | 1/4 | 1/8 | 1/16 | 1/32 | 1/64 | 1/128 |
| P_C [μ W] | 28 | 105.65 | 105.69 | 105.83 | 105.90 | 106.13 | N/A |
| | 224 | 104.30 | 104.30 | 104.29 | 104.27 | 104.53 | 104.96 |
| σ_{PC} [μ W] | 28 | 0.74 | 0.75 | 0.75 | 0.76 | 0.78 | N/A |
| | 224 | 0.74 | 0.74 | 0.74 | 0.74 | 0.75 | 0.78 |
| 7 bit DAS resolution | | | | | | | |
| | | Observation Period | | | | | |
| Figure of merit | Guard Interval [μ s] | 1/4 | 1/8 | 1/16 | 1/32 | 1/64 | 1/128 |
| P_C [μ W] | 28 | 96.86 | 96.76 | 96.74 | 96.93 | 96.99 | N/A |
| | 224 | 96.84 | 96.83 | 96.82 | 96.78 | 97.01 | 97.46 |
| σ_{PC} [μ W] | 28 | 0.68 | 0.68 | 0.69 | 0.69 | 0.71 | N/A |
| | 224 | 0.71 | 0.71 | 0.71 | 0.72 | 0.72 | 0.74 |

Table 13. Mean values (P_C) and experimental standard deviations (σ_{PC}) of channel power measures provided by the Burg estimator-based measurement algorithm. DVB-T settings: 8k transmission mode, 64-QAM modulation scheme.

| Instrument | Guard Interval [μ s] | P_c [μ W] | σ_{PC} [μ W] |
|------------|------------------------------|---------------------|-----------------------------|
| ESA | 28 | 106.63 | 0.59 |
| | 224 | 106.24 | 0.63 |
| PSA | 28 | 106.60 | 0.64 |
| | 224 | 106.77 | 0.54 |
| RSA | 28 | 92.68 | 0.66 |
| | 224 | 93.27 | 0.61 |
| RSA-SA | 28 | 93.40 | 0.30 |
| | 224 | 93.69 | 0.31 |

Table 14. Mean values (P_c) and experimental standard deviations (σ_{PC}) of channel power measures provided by ESA, VSA, RSA and RSA operating as spectrum analyzer. DVB-T settings: 8k transmission mode, 64-QAM modulation scheme.

A number of experiments on real DVB-T signals have been carried out through the optimized algorithm. The signals have been radiated by one MEDIASET DVB-T multiplexer operating on the UHF 38 (610-MHz RF central frequency) channel.

A simplified measurement station has been adopted. With respect to that used in the emulation stage, the function generator has been replaced by a suitable amplified antenna, the PSA and RSA have been removed, and a power splitter has been added (Angrisani L. et al., 2008). The cables, connectors, and power splitter have been characterized through the aforementioned network analyzers.

The channel power measurement results are summarized in Table 15, and a good agreement can be appreciated.

| 8k transmission mode, 64-QAM, 28 μ s guard interval UHF Channel 38 (610 MHz) | |
|--|---------------------------|
| | Measured Power [μ W] |
| WOSA estimator-based measurement algorithm | 26.75 |
| Burg estimator-based measurement algorithm | 26.92 |
| ESA | 26.42 |

Table 15. Experimental results.

5. Implementation issues in DSP-based meters

In order to evaluate the suitability of these methods to be implemented on real cost effective DSP platform two figures of merit have been taken into account:

- memory requirement, intended as the maximum number of samples to be preserved in the hardware memory;
- computational burden, defined as the number of operations (real additions and real multiplications) to be performed for gaining the desired PSD.

5.1 WOSA estimator

It can be demonstrated that an optimized implementation that is able to reduce the memory requirements would have two requirements.

1. A meter memory that is able to preserve, for the whole measurement time, 2M real samples, which are related to the acquired and overlapped buffers, and 2M complex

samples ($4M$ real samples) for the current and averaged FFT. To prevent additional memory requirements, a computational time to perform the current FFT shorter than $M*(1-r)*T_s$ is desirable, with T_s being the sampling interval.

2. $M*\log_2(M)$ additions and $M*\log_2(M/2)$ multiplications for each FFT calculation performed on M real samples. It is worth stressing that, to achieve a satisfying frequency resolution in PSD estimation, both K and M should sufficiently be high, with a consequent increase in the memory requirement and computational burden.

As an example, let us consider a DVB-T signal with a center frequency of 36.13 MHz, sampled at 100 MS/s. To achieve a good frequency resolution, i.e., 24 kHz, and good metrological performance in the WOSA estimation (with an overlap ratio of 90% [6], [7]), each FFT has to be calculated on 4096 samples, thus requiring 49152 additions and 45056 multiplications. The storage capability of the meter has to allow at least 24576 real samples to be preserved.

The reduction of memory need and computational burden is possible only if the computational time is lower than $40.96 \mu\text{s}$ (i.e., 4096 samples at 100 MS/s). This is a pressing condition that typically requires the use of expensive multicore platforms.

5.2 Burg estimator

Starting from what Kay and Marple have presented in (Kay & Marple, 1981) and considering the same acquired sequence previously described, it is possible to demonstrate that the minimum number of samples to be stored is $2N+p+2$, where p is the selected polynomial order, and the $3Np-p^2-2N-p$ real additions and $3Np-p^2-N+3p$ real multiplications are required for PSD estimation. As for the computational time, the estimation of the current PSD has to take a time interval that is not greater than $N*T_s$, because the whole acquired sequence is involved if real-time operations are pursued.

For $N>345$, both the computational burden and the required memory depth are higher than those peculiar to the WOSA estimator. To improve this aspect, an optimized implementation called "sequential estimation" that is able to update the PSD estimate whenever a new sample is available can be adopted.

5.3 Sequential Burg estimator

Let us consider (42) for $p=k$, and denote $a_{k,k}$ as K_k .

Making the time dependence explicit, the following relation is obtained:

$$K_k(N) = \frac{2 \sum_{s=k}^N e_{k-1}(s) b_{k-1}^*(s-1)}{\sum_{s=k}^N \left[|e_{k-1}(s)|^2 + |b_{k-1}(s-1)|^2 \right]} \quad (47)$$

where N is the time index. A time-update recursive formulation for (47) is given by

$$K_k(N+1) = K_k(N) + \frac{\left[K_k(N) \left(|e_{k-1}(N)|^2 + |b_{k-1}(N-1)|^2 \right) + 2e_k(N) b_k^*(N-1) \right]}{\sum_{s=k}^N \left[|e_k(s)|^2 + |b_k(s-1)|^2 \right]} \quad (48)$$

Equation (48), combined with (24) and (25), for $k=1,\dots,p$ and initial conditions $e_0(N)=b_0(N)=x_N$, suggests a sequential time-update algorithm for the reflection coefficients. After updating the reflection coefficient K_k , the $k + 1$ parameters $a_{k,1}, a_{k,2}, \dots, a_{k,k}$ and σ_k^2 can be calculated using the Levinson–Durbin recursions. The order of complexity involved is $o(p^2)$, which could significantly worsen the overall computational burden if all coefficients and parameters have to be updated whenever a new sample is available.

Significant computational saving is granted by updating the reflection coefficients $K_k(N)$ at each new sample and all the other parameters after a suitable time interval. More specifically, $9p$ multiplications and $7p$ additions are required to update the reflection coefficients, whereas p^2 multiplications and p^2 additions are needed for all the other parameters.

As for the memory requirement, the minimum number of samples that the sequential implementation requires to be stored is equal to $7p$. The result accounts for p reflection coefficients k_k ; p polynomial coefficients $a_{p,m}$; p forward and p backward prediction errors e_k and b_k , respectively; p coefficients Λ_k and p coefficients Γ_k ; and p estimates of the noise variance σ_k^2 . The obtained value is significantly lower than that required by the non-sequential implementation.

According to the example previously given and considering a value of p equal to 46 (the optimum value previously found), it is possible to assert that 280 samples should be stored in the meter memory, and a computational burden of 360 multiplications and 280 additions should be required to estimate the reflection coefficients.

Hence, it is possible to state that the sequential version of the Burg estimator exhibits better performance than that peculiar to the WOSA estimator and is entitled to be the core of a cost effective DSP-based meter.

6. A cost effective DSP-based DVB-T power meter

In the following the development of a new cost-effective instrument for power measurement in DVB-T systems is proposed. It is based on the improved measurement method sketched in Fig. 13.

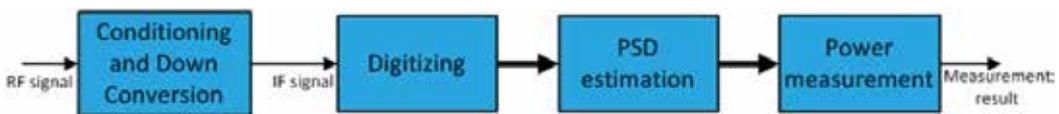


Fig. 13. Simplified block diagram of the proposed measurement method.

After a suitable conditioning and downconversion section, the input signal is digitized and its acquired samples processed in order to estimate its PSD. A proper measurement algorithm, operating on the achieved PSD, finally provides the desired power values.

As for the PSD estimation, the use of Autoregressive (AR) parametric estimators has shown the best trade-off between metrological performance and memory requirements with respect to nonparametric approaches (Angrisani et al., 2009). In addition, as described in (Angrisani et al., 2008-2), efficient implementations of AR parametric PSD estimators have been considered to drastically reduce the measurement time and to limit the memory needs over long observation intervals. A sequential Burg based estimation algorithm has been adopted because it is able to update estimates on data sample by data sample, (Kay & Marple, 1981),

(Marple, 1987) and warrants the best trade-off between computational burden and accuracy, as well as negligible bias and good repeatability.

The core of the proposed instrument is the PSD estimation section that has been implemented on a suitable Field Programmable Gate Array (FPGA) platform. These kinds of digital signal processors are particularly suited for algorithms, such as the sequential implementation of the Burg algorithm, which can exploit the massive parallelism offered by their architecture.

6.1 The hardware

A cost-effective hardware characterizes the meter. It consists of the following sections: (a) the tuner, (b) the analog-to-digital conversion, and (c) the FPGA-based computing platform. A simplified block diagram is depicted in Fig. 14.

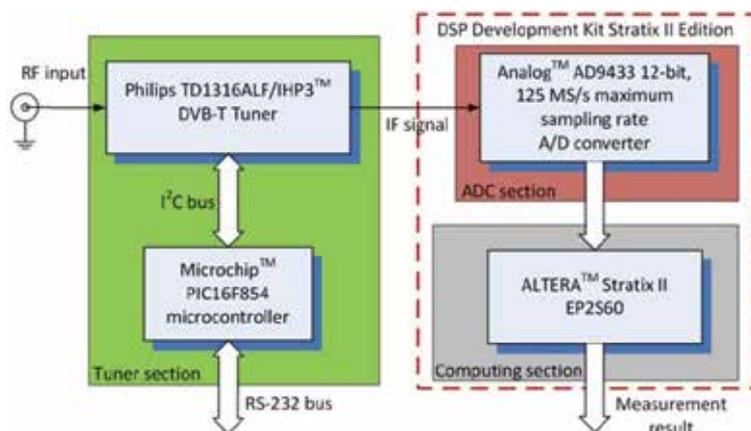


Fig. 14. Simplified block diagram of the proposed DVB-T power meter.

(a) The tuner section down-converts the incoming DVB-T signal, detected by a suitable antenna, to an intermediate frequency (IF) equal to 36.13 MHz. The task is performed by a Philips TD1316ALF/IHP3™ device, which is a single conversion tuner for digital terrestrial applications (NXP, 2006). It is provided with two IF outputs: a narrow-band one, equipped with a surface acoustic wave (SAW) filter and a gain controllable IF-amplifier, and a wideband output without any filter. Both the output circuits are regulated by an internal gain control loop with selectable takeover point settings via I²C bus. An external gain control is also possible if the internal loop is disabled. As far as the narrow-band IF output is concerned, it is possible to select the bandwidth of the SAW filter among 7 MHz and 8 MHz via I²C bus.

All these settings have been controlled and set up by a Microchip™ PIC16F854 microcontroller. It also provides a bus interface conversion between the serial I²C bus of the tuner and a common RS-232 one, allowing a simple connection with PC based environments. In this way, it is possible to set up the DVB-T channel, the SAW filter bandwidth, and the IF amplifier gain.

(b) The analog to digital conversion section is constituted by an Analog™ AD9433. This is a 12-bit monolithic sampling ADC that operates with conversion rates up to 125 MS/s. It is optimized for outstanding dynamic performance in wideband and high IF carrier systems (Analog, 2001).

(c) The computing platform is based on a FPGA chip. In particular the ALTERA™ Stratix II EP2S60 device mounted on the DSP Development Kit Stratix II Edition is considered (Altera, 2007). The chip is a fixed point FPGA that works with operative frequencies from tens of kHz to 400 MHz. This is obtained by using suitable Phase Locked Loops (PLLs) circuits. Other important features of the considered device are the 24176 Adaptive Logic Modules (ALMs), 48352 Adaptive Look-Up Tables (ALUTs), 36 Digital Signal Processing blocks (corresponding to 144 full-precision 18x18-bit multipliers) and 2544192 RAM bits. The typical cost of the considered FPGA chip is of about \$300.

6.2 The firmware

As far as the firmware of the computing platform is concerned, the sequential version of the Burg estimator proposed in (Angrisani et al., 2008-2) and summarized in the block diagram sketched in Fig. 15 is implemented. The firmware operates as follows. After a preliminary initialization phase, not reported in the block diagram, every time a sample is acquired a p -order cycle (i.e. a cycle repeated p times) is started. In each iteration, the estimation of the reflection coefficients (k_i), the σ_i , and the update of the prediction errors is performed. At this

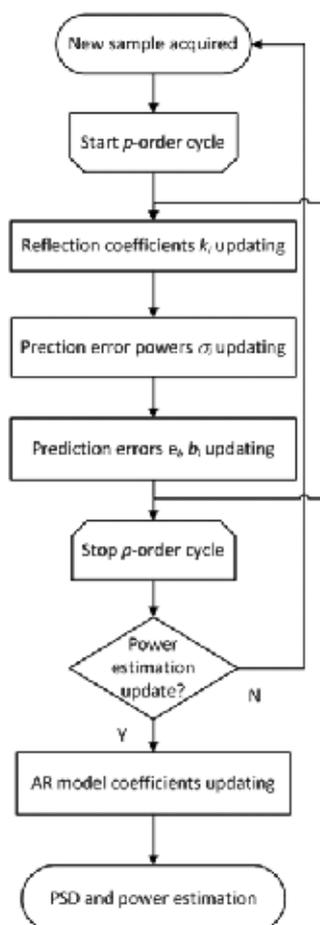


Fig. 15. Block diagram of the implemented FPGA firmware.

stage the user can select if to update the measured power and PSD or not. By updating the desired values of the new coefficients of the p -order model, the estimation of the PSD and channel power is performed, otherwise the p -order cycle begins when a new sample is made available from the ADC.

An order of p equal to 40 has been chosen, and a cascade of 40 sequential blocks implementing the p iterations of the cycle has been realized. The operative frequency of the FPGA device has been set equal to the sampling frequency (100 MS/s). All the implemented blocks warrant a calculation time lower than the sample period (10 ns), thus allowing the real time operation of the instrument. It is worth noting that the firmware architecture realizes a pipeline cascade of computing blocks. This operating way is allowed by the massive parallelism offered by the FPGA architecture. In particular, after a starting latency of 400 ns (i.e. 40 blocks for a sampling time of 10ns) the first measurement result is available. From this time instant on, measurement results are updated each 10 ns.

As far as the hardware resources are concerned, each one of the 40 blocks requires about 4525 ALUTs (Adaptive Look-Up Tables) and 2431 ALMs (Adaptive Logic Modules), thus resulting in a total of 18100 ALUTs and 97240 ALMs for the whole p -order cycle. These values impose the use of a multi-FPGA platform including 4 Stratix II EP2S60 FPGA chips operating in cascade arrangement.

8. References

- Agilent (1991). *HP3589A Operator's Guide*. Hewlett-Packard Company. P/N 03589--90021. Santa Clara (CA), USA.
- Agilent (2001). *E4406A VSA Series Transmitter Tester User's Guide*. Agilent Technol. P/N E4406-90177. Santa Clara (CA), USA.
- Agilent (2003). *AN 1449--2: Fundamentals of RF and Microwave Power Measurements*. Agilent Technologies Inc. EN 5988--9214, Santa Clara (CA), USA.
- Agilent (2003-2). *AN 1303: Spectrum Analyzer Measurements and Noise*. Agilent Technologies Inc. EN 5966--4008, Santa Clara (CA), USA.
- Agilent (2004). *ESA-E Series Spectrum Analyzers Specification Guide*. Agilent Technol. P/N E4401-90472. Santa Clara (CA) USA.
- Agilent (2005). *N1911A and N1912A P-Series Power Meters User's Guide*. Agilent Technol. P/N N1912-90002. Santa Clara (CA) USA.
- Altera (2007). *Stratix II Device Family Handbook*, Altera Corporation. San Jose (CA), USA.
- Analog (2001). 12-bit, 105/125 MSps IF sampling A/D converter AD9433, Analog Devices Inc. Norwood (MA), USA.
- Angrisani L.; D'Apuzzo M. & D'Arco M. (2003). A new method for power measurements in digital wireless communication systems. *IEEE Trans. Instrum. Meas.*, Vol. 52, No. 4, (Aug. 2003) pp. 1097-1106.
- Angrisani L.; Capriglione D.; Ferrigno L. & Miele G. (2006). Reliable and repeatable power measurements in DVB-T systems, *Proceedings IMTC 2006*, pp. 1867-1872, Sorrento, Italy, Apr. 24-27, 2006.
- Angrisani L.; Capriglione D.; Ferrigno L. & Miele G. (2007). Power measurement in DVB-T systems: on the suitability of parametric spectral estimation in DSP-based meters., *Proceedings IMTC 2007*, pp. 1-6, Warsaw, Poland, May 1-3, 2007.

- Angrisani L.; Capriglione D.; Ferrigno L. & Miele G. (2008). Power measurements in DVB-T systems: New proposal for enhancing reliability and repeatability. *IEEE Trans. Instrum. Meas.*, Vol. 57, No. 10, (Oct. 2008) pp. 2108-2117.
- Angrisani L.; Capriglione D.; Ferrigno L. & Miele G. (2008-2) Sequential parametric spectral estimation for power measurements in DVB-T systems, *Proceedings I²MTC 2008*, pp. 314-319, Victoria (BC), Canada, May 2008.
- Angrisani L.; Capriglione D.; Ferrigno L. & Miele G. (2009). Power measurement in DVB-T systems: On the suitability of parametric spectral estimation in DSP-based meters. *IEEE Trans. Instrum. Meas.*, Vol. 58, No. 1, (Jan. 2009) pp. 76-86.
- Anritsu (2003). Vector Network Analyzers Technical Data Sheet ANRITSU 37100C/37200C/37300C, Rev. C.
- Burg J. P. (1967). Maximum Entropy Spectral Analysis, *Proceedings of 37th Meeting Soc. Explor. Geophys.*, Oklahoma City (OK), USA, Oct. 31, 1967.
- Daubechies I. (1992). *Ten Lectures on Wavelets*. SIAM, Philadelphia (PA), USA.
- Donoho D. L. & Johnstone I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, Vol. 81, No. 3, (Aug. 1994) pp. 425-455.
- ETSI (2004). EN 300 744: "Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for digital terrestrial television (V1.5.1)". ETSI Std, Sophia Antipolis, France.
- ETSI (2004-2). TR 101 190: "Digital Video Broadcasting (DVB); Implementation guidelines for DVB terrestrial services; Transmission aspects (V1.2.1)". ETSI Std, Sophia Antipolis, France.
- Fischer (2004). *Digital Television - A Practical Guide for Engineers*. Springer-Verlag, ISBN 3540011552, Heidelberg, Germany.
- Jokinen H.; Ollila J. & Aumala O. (2004). On windowing effects in estimating averaged periodograms of noisy signals. *Measurement*, Vol. 28, No. 3, (Oct. 2000) pp. 197-207.
- Kay S. M. & Marple S. L. (1981). Spectrum analysis - A modern perspective. *Proc. IEEE*, Vol. 69, No. 11, (Nov. 1981) pp. 1380-1419.
- Makhoul J. (1975). Linear prediction: A tutorial review. *Proc. IEEE*, Vol. 63, No. 4, (Apr. 1975) pp. 561-580.
- Marple L. (1980). A new autoregressive spectrum analysis algorithm. *IEEE Trans. Acoust., Speech, Signal Process.*, Vol. ASSP-28, No. 4, (Aug. 1980) pp. 441-454.
- Marple S. L. (1987). *Digital Spectral Analysis With Applications*. Prentice-Hall, Englewood Cliffs (NJ), USA..
- Morf M.; Dickinson B.; Kailath T. & Vieira A. (1977). Efficient solution of covariance equations for linear prediction. *IEEE Trans. Acoust., Speech, Signal Process.*, Vol. ASSP-25, No. 5, (Oct. 1977) pp. 429-433.
- Moulin P. (1994). Wavelet thresholding techniques for power spectrum estimation. *IEEE Trans. Signal Process.*, Vol. 42, No. 11, (Nov. 1994) pp. 3126-3136.
- NXP (2006). TD1300A(L)F mk3 Tuner modules for analog and digital terrestrial (OFDM) applications, NXP Semiconductors. Eindhoven, The Netherlands.
- Nuttall A. H. (1976). Spectral analysis of a univariate process with bad data points, via maximum entropy and linear predictive techniques. *Naval Undersea Syst. Cent.*, New London (CT), USA, Tech. Rep. 5303.

- Percival D. B. (1992). Simulating Gaussian random processes with specified spectra. *Comput. Sci. Stat.*, Vol. 24, (1992) pp. 534–538.
- Reljin I.; Reljin B.; Papic V. & Kostic P. (1998). New window functions generated by means of time convolution – Spectral leakage error, *Proceedings of 9th MELECON*, pp. 878–881, Tel-Aviv, Israel, May 18–20, 1998.
- Tektronix (2006). RSA3408A 8 GHz Real-Time Spectrum Analyzer User Manual, Tektronix Inc. 071-1617-01. Beaverton (OR), USA.
- Ulrych T. J. & Clayton R. W. (1976). Time series modelling and maximum entropy. *Phys. Earth Planet. Inter.*, Vol. 12, No. 2/3, (Aug. 1976) pp. 188–200.
- Walden A. T.; Percival D. B. & McCoy E. (1998). Spectrum estimation by wavelet thresholding of multitaper estimators. *IEEE Trans. Signal Process.*, Vol. 46, No. 12, (Dec. 1998) pp. 3153–3165.
- Welch P. D. (1967). The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms. *IEEE Transactions on Audio and Electroacoustics*, Vol. AU-15, No. 2, (1967) pp. 70–73.

MidField: An Adaptive Middleware System for Multipoint Digital Video Communication

Koji Hashimoto and Yoshitaka Shibata
Iwate Prefectural University
Japan

1. Introduction

As broadband networks become more common, using high quality video streams such as DV (Digital Video) or HDV (High-Definition Video) has become increasingly popular in multipoint communications. Today's video communication systems which support high quality digital video formats on broadband networks are used for a variety of purposes (Bak, 2006; Gharai, 2006; Bodecek & Novotny, 2007). Although those systems have many useful functions for multipoint communications, it requires much bandwidth for video streams. A unidirectional DV stream requires about 28.8 Mbps. HDV 720p and HDV 1080i require about 19 Mbps and 25 Mbps respectively. Therefore, when we use high quality digital video communication systems, in many cases we should consider whether each location has enough communication environments to process the video streams or not. Of course, we can't always use effective format transcoding functions in multipoint communications. As the result, most video formats that are used by current multipoint communication systems are low resolution formats than DV and HDV even if we can use DV/HDV cameras at each end point.

If suitable format transcoding functions as defined by RTP (Schulzrinne, 2004) or stream integration functions for large scale video communications (Gharai, 2002) are available on our communication environments, it may be able to transcode the stream format into another suitable stream format or integrate several streams to a mixed stream. If the required communication environments are permanent and the total number of participants is always limited, we may be able to prepare the required transcoding and integration functions into suitable intermediate nodes in advance. However, we can't always use enough bandwidth and CPU power, and our communication environments aren't always permanent, therefore we should consider on demand transcoding functions with relocatable mechanisms.

In order to design and implement relocatable transcoding functions for multipoint digital video communications, employing an effective middleware (Ferreira, 2003; Jameela, 2003) on programmable networks (Campbell, 1999) is an effective way. As one of the methods for executable program modules to migrate to anywhere, many mobile agent systems (Guedes, 1997; Ohsuga, 1997; Dong & Seung, 2001; Antonio, 2002; Guo, 2004) also have been proposed. We consider that mobile agents which perform given tasks on suitable nodes are one of important base techniques for constructing flexible and interactive multipoint digital video communication environments.

To perform various streaming functions in suitable intermediate nodes according to each user's communication environment, the executable and relocatable program modules should perform the required functions for audio-video streams in realtime. We are developing a new middleware system (Hashimoto & Shibata, 2008) for multipoint audio-video communications which can use suitable formats that include high quality digital video formats such as DV and HDV. The system runs on a mobile agent subsystem and relocatable transcoding program modules as the mobile agent process audio-video streams. The system has also resource monitoring/management functions and video session management functions.

The remainder of this chapter is organized as follows. The next Section presents MidField System architecture and describes flexible and interactive communication sessions which are constructed by MidField System entities. Section 3 presents a communication protocol and packet transmission mechanisms for streaming modules, and Section 4 illustrates inside of the streaming module. Then the current implementation and use cases are described in Section 5 and 6. Finally, Section 7 concludes this chapter.

2. MidField system

Figure 1 shows the system architecture, which is between the application and the transport layer. MidField system consists of three layers and four vertical planes.

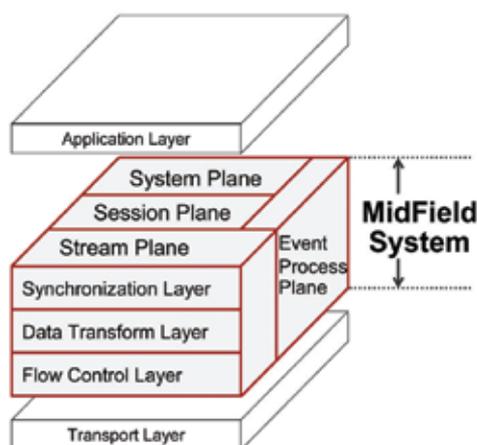


Fig. 1. MidField System Architecture

The stream plane consists of three layers, which are synchronization, data transform and media flow control layers. This plane performs inter and intra media synchronization, transcodes audio-video formats as necessary. In the flow control layer, packetization and depacketization of continuous audio-video stream data are performed and audio-video packets are sent and received among other MidField Systems.

On multipoint communications, the session management is an important function. The session plane manages multipoint session information (e.g. the number of users, available audio-video formats in the session and available computer/network resources, etc.). We defined MidField Session as a multipoint communication session in MidField System. By using several IP multicast sessions in a MidField Session, MidField System can connect remote users according to users' communication environments.

Local system resources such as CPU and required bandwidth are managed in the system plane. This plane monitors incoming and outgoing network traffic and CPU utilization rate

in a local node, and performs admission tests to check whether the required stream processing is possible on the node.

The system plane includes a simple mobile agent platform. Agents that are executable software components in MidField System can use an inter-agent message protocol and mobile facilities. Because each functional entity is based on the agent basically, each entity can exchange messages easily. Of course, although not all entities require mobile facilities, mobility of agents is able to connect users' various communication environments effectively and dynamically.

MidField System is based on an event driven model. Each function of these planes and layers is performed by various events, which are associated with media processing tasks and inter-agent messages. The event process plane not only handles various events on several threads but also manages message connections with remote MidField Systems.

2.1 MidField session

If our computer network environment for interactive audio-video communications has enough resources to use DV or HDV streams, of course we can communicate with remote users by DV or HDV streams. However we can't always use required enough resources for high quality digital video streams. MidField System can establish a communication session according to each user's communication environment by using IP multicast sessions.

Figure 2 shows an example of MidField Session, which includes three users and uses three video streams at each end point. A MidField Session consists of at least one multicast session and provides peer-to-peer communications to users.

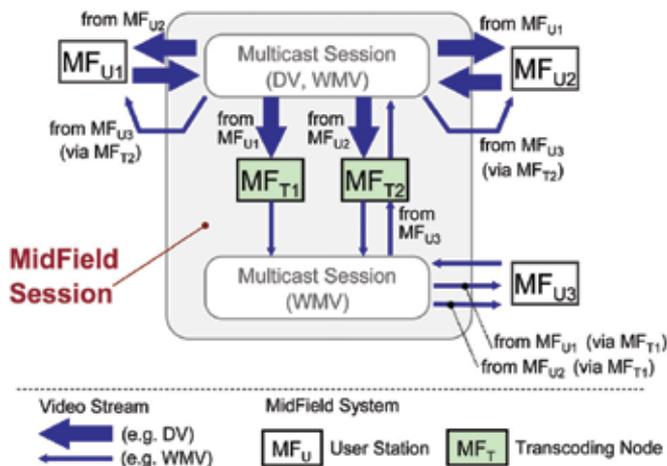


Fig. 2. MidField Session

In Figure 2, the user station MF_{U1}, MF_{U2} and MF_{U3} join a MidField Session. Now, MF_{U1} and MF_{U2} have enough communication resources to process DV streams that are required about 28.8 Mbps per stream. On the other hand, the MF_{U3} does not have enough communication resources to process DV streams because of lack of enough bandwidth or computing power to process DV streams.

In this case, to connect these users MidField System tries to use IP multicast sessions and transcoding nodes. If the MF_{U3} is able to process another video stream format such as Windows Media Video (WMV), the system will prepare transcoding nodes for transcoding DV streams into WMV streams.

In Figure 2, there are two transcoding nodes (MF_{T1} and MF_{T2}). At first, MF_{U1} and MF_{U2} join the DV multicast session, then MF_{U3} joins the WMV multicast session. Then, for connecting three users, MF_{U1} and MF_{U2} send program modules that are implemented as mobile agents into MF_{T1} and MF_{T2} . On these transcoding nodes, the program modules will start to transcode or relay streams.

As the result, three users will be able to communicate with each other by using suitable video stream formats according to each user's communication environment.

2.2 MidField stream

To establish a multipoint communication session, the system must have adaptive streaming modules, which need to support several audio-video stream formats, to be able to transcode stream formats and to perform on suitable node anywhere.

We have considered which formats should be supported in our system. In order to realize high quality digital video communications, DV and HDV formats should be supported. We can use DV and HDV cameras as input devices for our PCs commonly. However, not all users can use high specification PCs and enough broadband networks always, therefore, for connecting various communication environments the other stream formats also should be supported. At the beginning of the system design, we had employed MPEG4 video format for users who don't have enough communication environments to process DV/HDV streams. Although MPEG4 was one of suitable formats for multipoint communications, but it is one of choice. Now our system mainly supports DV, HDV and ASF (Advanced Stream Format) that includes WMV (Windows Media Video) and WMA (Windows Media Audio) formats.

Figure 3 illustrates the outline of a streaming module, which processes audio-video streams in the stream plane. A Stream Agent is a part of an end-to-end audio/video stream.

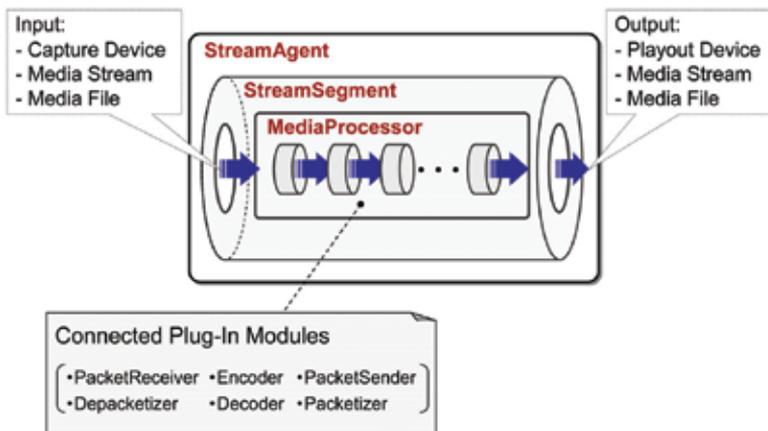


Fig. 3. Stream Agent

Stream Agent consists of mainly three parts and the implementation is based on mobile agent technique, therefore, it can exchange control messages with the other Stream Agent(s). In addition, Stream Agent can migrate from a local MidField System to another MidField System running on a remote node. By implementing Stream Agent as a mobile agent, our system is able to process the required streaming on any remote nodes where a MidField System is running.

The implementation of a simple mobile agent platform in MidField System is written in Java programming language. Stream Agent also is implemented in Java. At the beginning of the system implementation, we had been trying to use only Java for the whole system, because we considered that Java would become one of the most effective programming languages to implement the mobile agent platform.

Although we had implemented Stream Agent by using Java Media Framework API (JMF), unfortunately the implementation could not get enough performance for practical multipoint communications. However, because a transcoding module with mobility is required for connecting IP multicast sessions in a MidField Session, we had replaced a part of streaming modules with native program modules. As native program modules for audio-video streaming, currently, we employ DirectShow of Microsoft. Of course, these native modules don't have mobility. Therefore, the native modules don't migrate even if a Stream Agent migrates. However, a Stream Agent can migrate with I/O information for a stream segment and then on the destination node the Stream Agent can handle stream data by using native resources of the destination node. For the above reason, Stream Segment and Media Processor in Figure 3 are implemented in C++ programming language.

Stream Segment is mediator for Stream Agent and Media Processor. Because Stream Agent is implemented in Java, in order to issue stream control commands to native program modules, Stream Segment uses Java Native Interface (JNI). In fact, there are two instances of Stream Segment in Java and C++ context respectively.

Media Processor processes audio-video data by using some suitable plug-in modules. A Media Processor has at least one input source and creates at least one output source. When a Media Processor gets a set of input(s) and output(s), the Media Processor tries to connect suitable plug-in modules according to the input and output requirements and the audio-video data move from one plug-in module to the connected next plug-in module. This mechanism is able to be implemented in JMF, DirectShow, and the other existing libraries. Besides, each plug-in module has the own interface for controlling data process from application programs.

2.3 Software module configuration of MidField system

MidField System is a middleware system for flexible and interactive communication environment, which provides various kinds of audio-video communication functions as API to application programs. The software module configuration is shown in Figure 4. The system has three interfaces for Stream, Session and System Plane. Stream Agent, Session Agent and System Agent are agent-based software modules for each plane.

As written above, Stream Agent performs audio-video streaming and Session Agent manages audio-video session information. System Agent monitors local resources and performs admission tests when users join MidField Session.

Here, Agent Place manages all active agents in a local system and performs creation, migration and termination of agents. Agent Place uses Agent Loader(s) and Agent Server for migration functions. Agent Place also is a kind of agent.

These agents can exchange messages with other agents that are active in local or remote systems. Connection Manager in the event process plane manages Connected Socket objects for inter-agent message delivery functions. System Event Manager has mechanisms to process various events that are created by each agent. If an agent creates some events, the agent posts the events to System Event Manager and the events are stored into Event Queue. Each stored event is processed by one of Event Processors that have an idle thread.

2.4 Adaptive communication session

One of the important functions of MidField System is to support adaptive communication sessions. Software modules in Figure 4 cooperate to establish a MidField Session. Figure 5 shows an example of MidField Session.

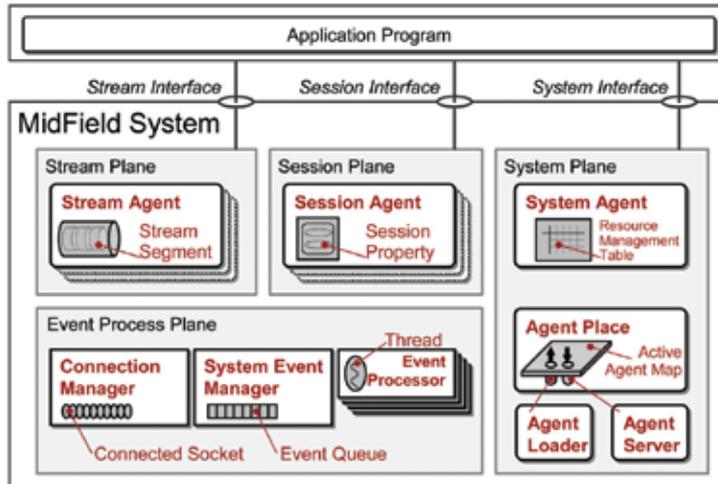


Fig. 4. Software Module Configuration of MidField System

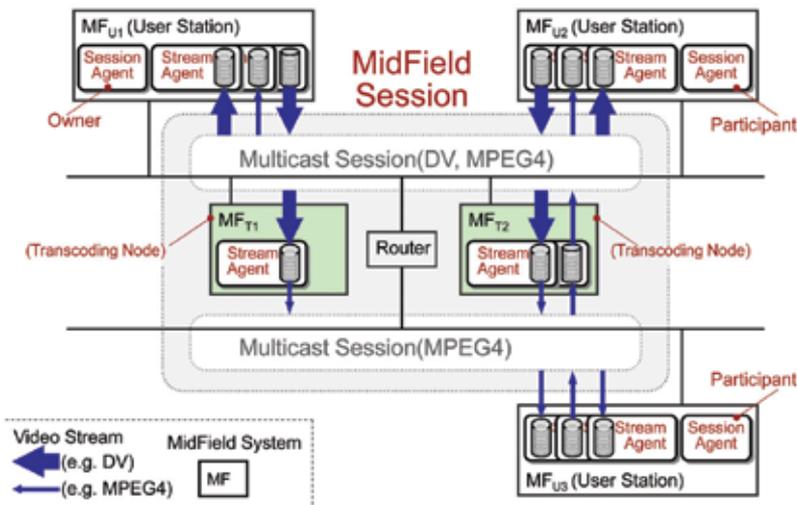


Fig. 5. An Example of MidField Session

In Figure 5, five MidField Systems are connected to computer networks. Three of them are User Stations (MF_{U1} - MF_{U3}), and remain two are Transcoding Nodes (MF_{T1} , MF_{T2}). At first, a Session Agent in MF_{U1} opens a new MidField Session as session owner. This MidField Session consists of two IP multicast sessions that are for DV streams and for MPEG4 streams respectively.

Now, a Session Agent in MF_{U2} joins the MidField Session as a participant, and begins to communicate with MF_{U1} on the multicast session for DV streams. MF_{U1} and MF_{U2} have

enough bandwidth and CPU power to process DV multicast streams. Then, MF_{U3} tries to join the MidField Session but unfortunately MF_{U3} doesn't have enough resources for DV streams. MF_{U3} requests MF_{U1} 's Session Agent to use the multicast session for MPEG4 streams.

In this case, the Stream Agents in MF_{U1} and MF_{U2} try to extend their DV streams to the MPEG4 multicast session. Selecting suitable transcoding nodes, after that Stream Agents for transcoding migrate to the selected transcoding node. Then each Stream Agent in transcoding node receives a DV stream and transcodes it into a MPEG4 stream. MF_{U3} is able to receive the streams from MF_{U1} and MF_{U2} as MPEG4 video streams.

Finally the Stream Agent in MF_{U3} starts to send a MPEG4 stream to the MPEG4 multicast session, after that the Stream Agent in MF_{U2} will receive the MPEG4 stream and relay it to the multicast session for DV streams. Thus MF_{U1} , MF_{U2} and MF_{U3} are able to communicate with each other by using a suitable IP multicast session.

We had developed a version of MidField System that includes an application that enables multipoint audio-video communications and experimented with 8 PCs in 2004. Figure 6 shows an image of the experimental communication. In this image, the upper 4 PCs use a multicast session for DV streams and the under left 2 PCs use another multicast session for MPEG4 streams. The under right 2 PCs are running as transcoding node.



Fig. 6. Experimental Communication (6 User Stations and 2 Transcoding Nodes)

MidField System had realized a dynamic session configuration method according to users' communication environments. It is one of results that the dynamic session configuration by integrating IP multicast sessions was introduced to multipoint communications. However, in many cases our communication environments are restricted to use IP multicast functions. As the result, unfortunately when we communicate with remote users on computer networks, we almost can't use IP multicast.

Currently, we are trying to design and develop a new adaptive session configuration method, which is based on the dynamic session configuration.

On the other hand, of course, in order to adapt to various communication environments, it is required for the system to have adaptive and flexible streaming functions. The following sections illustrate the design and implementation of Stream Agent, which has audio-video streaming functions for supporting adaptive communications.

3. Inter stream agent communication

In MidField System, an end-to-end audio/video stream consists of a chain of Stream Agents. Since a Stream Agent can handle several inputs and outputs, it is easy to relay audio-video streams from each source node. Figure 7 shows the outline of Stream Agents with direction of media streams (audio/video streams). In this figure, the relay stream agent $stm(n)$ receives three input media streams from the upper Stream Agents $stm(n-1)_0 - stm(n-1)_2$, and sends three output media streams to the lower Stream Agents $stm(n+1)_0 - stm(n+1)_2$. If a Stream Agent has several inputs, the Stream Agent performs mixer functions for these inputs, the mixed stream is sent to lower Stream Agents.

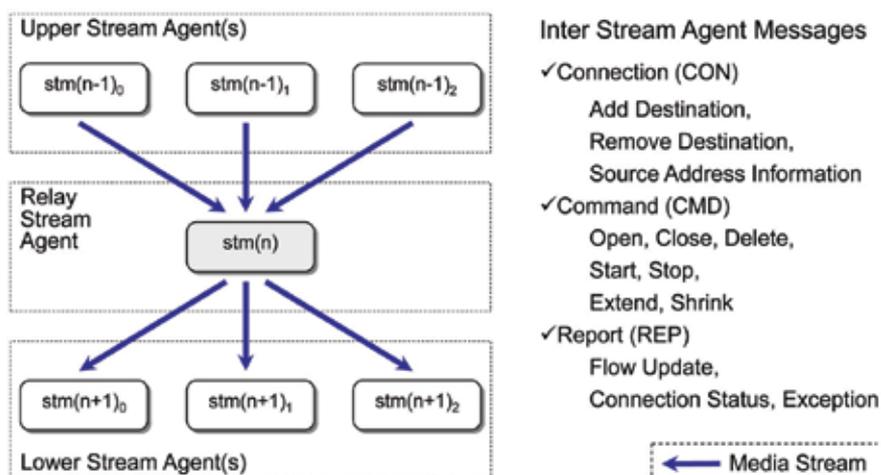


Fig. 7. Stream Agent and Messages

In order to realize adaptive connections, Stream Agents exchange messages shown in Figure 7. The connection (CON) messages are used for media stream connections. Once a media stream connection is established, it is controlled by the command (CMD) messages. In addition, to report flow update, connection status and exception to uppers, the report (REP) messages are exchanged.

3.1 Message flow of inter stream agent

Figure 8 (a) shows message flow for addition and removal of a media stream. At first, a NEW event is created in a lower node to receive a media stream from an upper Stream Agent $stm(n-1)$, and $stm(n)$ in the lower node gets the NEW event. Then $stm(n)$ sends a Flow Update message to $stm(n-1)$. In this figure, since $stm(n-1)$ is connected to an upper Stream Agent, $stm(n-1)$ also sends the Flow Update message to $stm(n-2)$. After the Stream Agents receive Flow Update messages from lower Stream Agents, each upper Stream Agent updates internal data structures for media stream flow.

Next, $stm(n)$ gets an OPEN event in the lower node and sends an Add Destination message to $stm(n-1)$. After that, $stm(n-1)$ replies to $stm(n)$. The response message includes a source address and a port number of the media stream. Establishing media stream connection between $stm(n-1)$ and $stm(n)$, $stm(n-1)$ starts to send media packets to $stm(n)$.

A Remove Destination message for a CLOSE event is used to remove the media stream connection. Finally, after a DELETE event in the lower node, $stm(n)$ sends a Flow Update message to the upper Stream Agent and updates the flow's data structure.

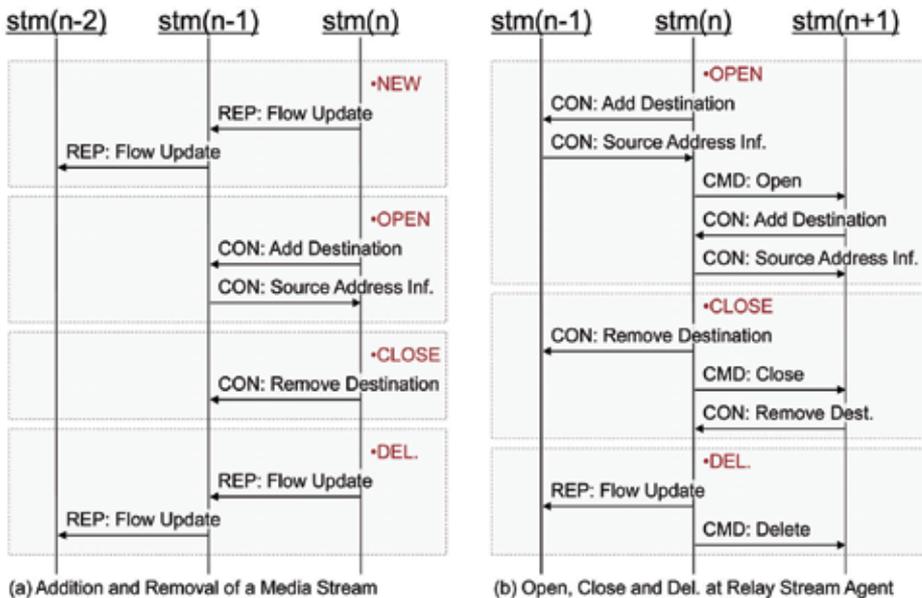


Fig. 8. Inter Stream Agent Message Flow #1

If $stm(n)$ is already connected to lower Stream Agents, the lowers receive command messages. Figure 8 (b) shows message flow in the case of generating OPEN, CLOSE and DELETE events in a relay node. An Open message which is sent from $stm(n)$ to $stm(n+1)$ is handled on $stm(n+1)$ as the same as a new created OPEN event. CLOSE and DELETE events also are handled as the same as OPEN event. Thus, MidField System is able to change flows of media streams adaptively.

As Figure 9 (a) shows, all command messages such as START and STOP are sent from upper Stream Agents to lower Stream Agents, and all report messages are sent from lower Stream Agents to upper Stream Agents.

Figure 9 (b) and (c) show a stream extension process briefly. As previously stated, Stream Agent has mobility in the capacity of mobile agent, which is used to extend a media stream. A Stream Agent starts to perform extension process when the Stream Agent gets an EXTEND event. In Figure 9 (b), $stm(n)$ has already I/O parameters, and after getting an EXTEND event, $stm(n)$ creates a clone of itself and the clone migrates to the other node. The clone $stm(n+1)$ is able to get NEW and OPEN events in the node in which the clone has immigrated, and start to receive the media stream from $stm(n)$. Of course $stm(n+1)$ can relay the receiving media stream to other Stream Agents.

This extension process can be also started from an upper Stream Agent. In Figure 9 (c), $stm(n)$ sends an Extend message to $stm(n+1)$ and the $stm(n+1)$ performs extension process.

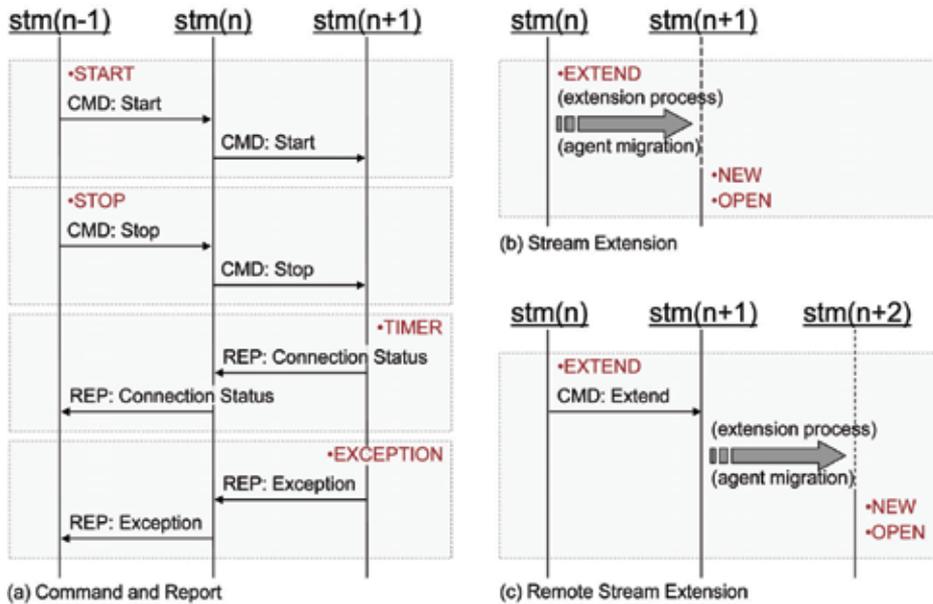


Fig. 9. Inter Stream Agent Message Flow #2

3.2 Stream extension process

In order to integrate various audio-video communication environments, to relay media streams adaptively is the one of important functions in audio-video communication systems. MidField System performs adaptive relay processes of media streams by mobility of Stream Agent. If it is clear which node is a suitable node to perform required relay functions, of course MidField System will employ the node as a relay or transcoding node. However, in case that there are several nodes which are able to employ as a relay or transcoding node, the system needs a node selection function.

MidField System's stream extension process realizes node selection and load balancing functions. Figure 10 shows the stream extension process.

At the beginning of a stream extension process, the system sends resource status requests to other MidField Systems which are running on other remote nodes. After receiving the request, those remote systems reply with a report of current resource status that includes available bandwidth (incoming/outgoing) and CPU utilization rate.

Replied resource status reports are added to a temporary remote resource status list, and after a timeout that has specified in advance, an available relay node list is created. Each element of the list includes resource status of relay nodes which have enough network bandwidth and CPU power in order to relay or transcode the required media stream. Then, the relay stream agent tries to migrate to a relay node which is included in the list.

If the number of timeouts is less than the constant value N and there are no available relay nodes, the system sends resource status requests to other MidField Systems again. If the number of timeouts is over the N and there are no available relay nodes, then the whole system for this communication environment can't allocate the required resources for the stream extension and this extension process is failure.

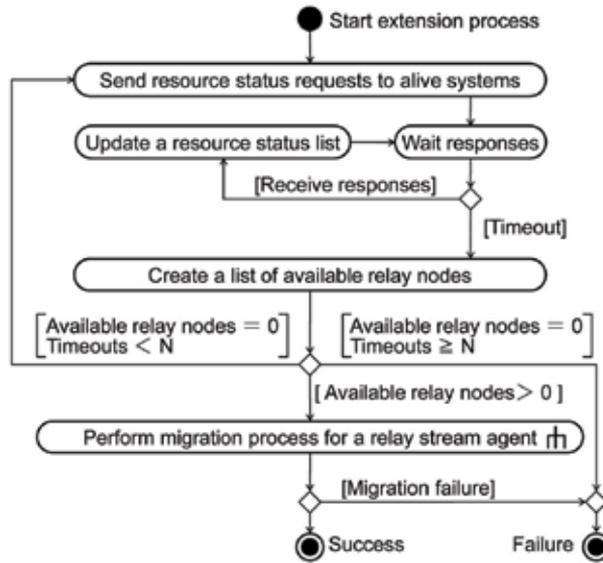


Fig. 10. Extension of Media Stream

The migration process for a relay stream agent to migrate to another node is shown in Figure 11. At first, a migration request is sent to the relay node. The request includes parameters for input and output associated with this relay stream. In the relay node, admission tests are performed for specified input and output parameters. If the required resources to relay or transcode the media stream are able to allocate, the resources are allocated in the relay node. Finally, the relay stream agent migrates to the relay node and starts to relay or transcode the media stream.

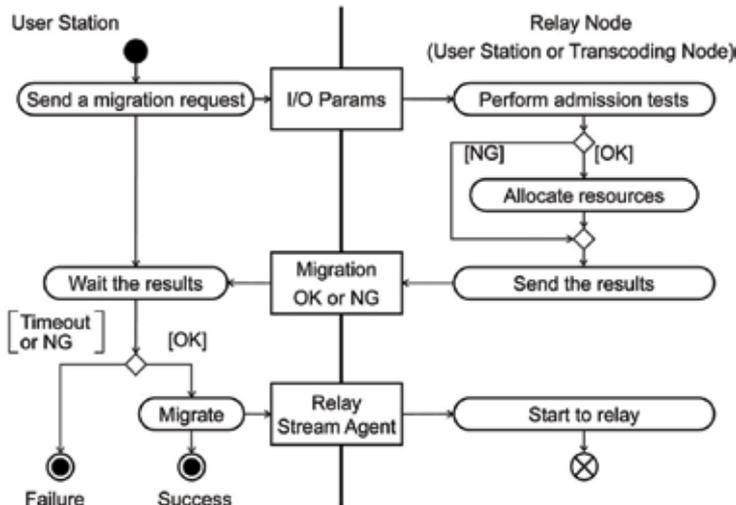


Fig. 11. Migration of Relay Stream Agent

As shown in Figure 9 (b) and (c), these processes for a stream extension can be performed in not only a local node but also another remote node. By considering an end-to-end media

stream as an adaptive object, MidField System has realized that as one of functions to establish a flexible and interactive communication environment.

3.3 Multithreaded stream packet transmission

As each element in an end-to-end media stream of MidField System, Stream Agent is required to handle several inputs and outputs dynamically. As stated in the previous section, all stream data in MidField System are processed by DirectShow, but DirectShow doesn't provide any packet transmission filters basically. For realising multipoint communications, we have newly designed and implemented packet transmission mechanisms, which work as filters of DirectShow. In order to adapt with the modular architecture of DirectShow, a part of the modules for packet transmission is implemented as filters of DirectShow.

Here, Packetizer is a filter that fragments each audio-video sample data which comes from an upper filter into transmittable data packets, and Depacketizer is a filter that assembles each data packet which receives from the source node into a sample data.

(1) For sending packets

Figure 12 illustrates the inside of a Media Processor in sender side, which includes a Packetizer that is connected to several Packet Senders for each destination. A pair of Packetizer and Packet Sender is associated with a packet queue, and at the both sides of packet queue, packet data are processed on respective threads.

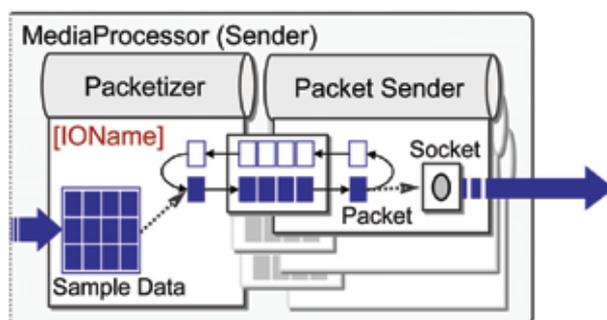


Fig. 12. Packetizer and Packet Sender(s)

In multipoint audio-video communications, to use IP multicast is effective clearly. However, on actual computer networks such as just the Internet, IP multicast is almost not available, therefore, audio-video communication systems should be able to adapt with the network environments. In MidField System, Packet Sender supports both IP unicast (MFSP/TCP and MFSP/UDP) and multicast (MFSP/UDP), in addition, a Packetizer is able to add and remove Packet Sender(s) dynamically depending on the number of destinations. Thus, Packet Sender supports not only IP multicast but also application level multicast (ALM) functions.

Since MidField System supposes a source node has several different source streams simultaneously, each stream must have a unique ID. The IOName in Figure 12 shows the unique ID, and on the destination nodes, each receiver uses the same IOName that is created in the source node.

(2) For receiving packets

Figure 13 illustrates the inside of a Media Processor in receiver side, which includes a Packet Receiver that is connected to several Depacketizer(s) for each lower filter. Like the sender side, a pair of Depacketizer and Packet Receiver is associated with a packet queue and packet data are processed on respective threads at the both sides of the packet queue.

As stated above, a stream has unique ID between a pair of sender and receiver node. By using this unique ID in a receiver, receiving redundant streams can be eliminated. For example, suppose that a receiver node is receiving a DV stream from a sender node and relaying the DV stream to another node, then next the receiver node starts to transcode the DV stream into a WMV+PCM stream newly. In this case, if the receiver node is receiving a DV stream of the target IOName, the receiver doesn't receive the same DV stream newly. The receiver node can create a new Depacketizer in a Media Processor and connect to the Packet Receiver that is receiving the target DV stream. Thus, the receiver node is able to start transcoding the DV stream newly without receiving a redundant DV stream.

The packet transmission mechanism in MidField System adapts media streams for various requirements. The mechanism is able to distribute streams dynamically and effectively. The combination between DirectShow's filters and this transmission mechanism will support a wide variety of multipoint audio-video communications.

Our current implementation has supported IPv4 and IPv6. In addition if users can use a relay node which has a global IP address, MidField System can communicate by multipoint peer-to-peer media streams even if each end user's node has a private IP address.

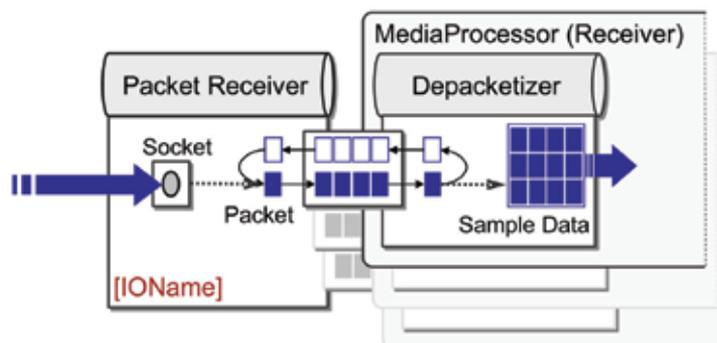


Fig. 13. Packet Receiver and Depacketizer(s)

4. Inside of Media Processor

Configurations of Media Processor are concerned with adaptive connectivity. In order to process a variety of inputs and outputs effectively, Media Processor's implementation employs DirectShow which has a modular architecture that can connect several software components called filters.

DirectShow provides several components that can be used to build filter graphs which are a set of connected filters. Suitable filters for a required streaming process are connected automatically by using the intelligent connection mechanism of DirectShow.

However, since each streaming process for DV, HDV, WMV, etc., has various special characteristics, not all filters are connected well automatically always.

4.1 Supported patterns of filter graphs

To satisfy the requirements for implementing adaptive streaming modules, we have designed eight patterns of filter graphs including packet transmission filters, which are shown in Figure 14.

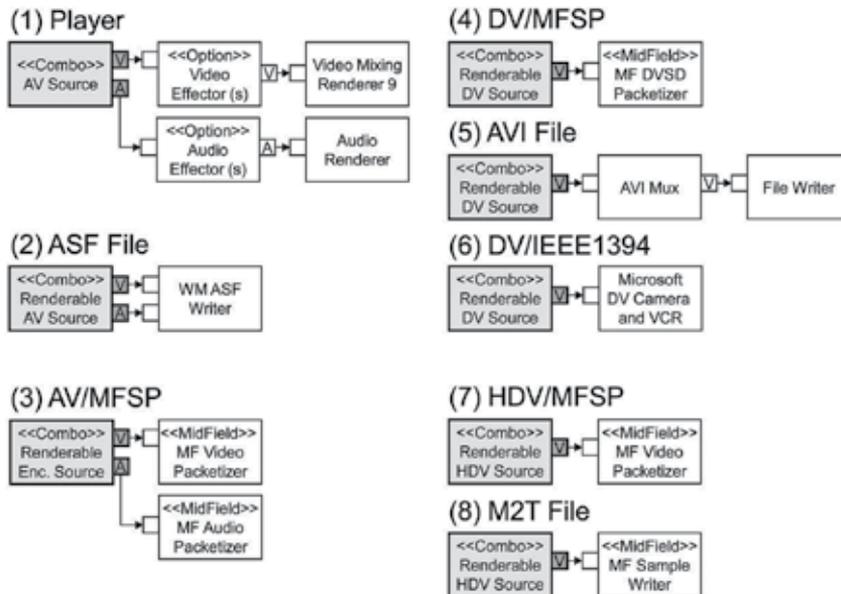


Fig. 14. Supported Filter Graphs in Media Processor

As the diagrams show, each filter is connected to one or more other filters, and the connection points are called pins. All filters use pins to move audio-video data from one filter to the next. The arrows in the diagrams show the direction in which the data travels.

Here, in the diagrams, <<Combo>> means the filter includes more several connected filters. For example, AV Source filter in Figure 14 is one of connected filter patterns shown in Figure 15. <<Option>> filter is inserted as necessary. In Figure 14 (1), Video Effector(s) and Audio Effector(s) are optional filters. Then <<MidField>> means the filter is implemented newly in MidField System for realising Stream Agent. There are several <<MidField>> filters, which appear also in following diagrams.

In Figure 14, the pattern (1) is a player that renders various AV sources shown in Figure 15. The pattern (2) creates an advanced systems format (ASF) file from a renderable AV source. Here, ASF is a container format for Windows Media Audio (WMA) and WMV contents. Then, the pattern (3) transmits outputs of Renderable Encoded Source filter shown in Figure 16 to destination(s). In order to send and receive media packets, we have designed MidField Streaming Protocol (MFSP), which is a simple stream transmission protocol that designed by referring to RTP. The MFSP is able to be used on TCP or UDP.

Next, in the pattern (4), outputs of Renderable DV Source are transmitted to destination(s) by using MFSP. The pattern (5) creates an AVI file from Renderable DV Source and the pattern (6) transmits a DV stream to an IEEE1394 device.

In order to process HDV stream, the pattern (7) transmits packets for a HDV stream to destination(s) and the pattern (8) creates a M2T (MPEG2-TS format) file from Renderable HDV Source.

4.2 Source filters and renderable sources

In Figure 15, there are six patterns of AV source filters, and each pattern creates outputs of audio or video sample data for a connected lower filter continuously. Basically, the output audio format is PCM and the output video format is RGB.

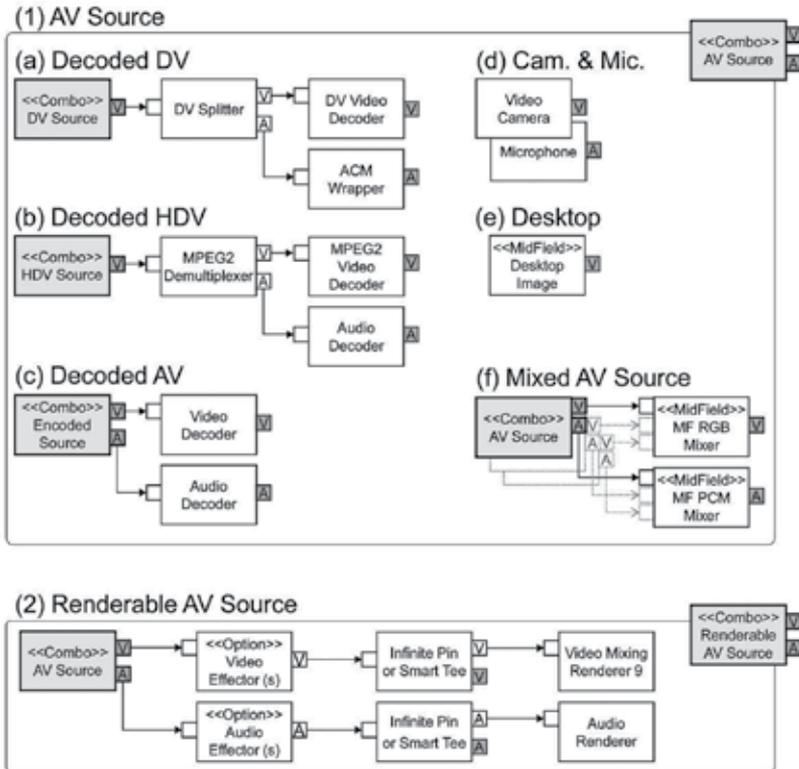


Fig. 15. AV Source Filters

Combo filters appear again in Figure 15 (a), (b) and (c). These combo filters are DV Source, HDV Source and Encoded Source. Decoded audio-video samples in these filters are available to use as an AV Source filter. In case of using microphones and video cameras such as USB video camera, the pattern (d) is applied to process these input devices. The pattern (e) captures and outputs continuously desktop images in RGB video format.

In addition, we have implemented audio-video mixer filters shown as the pattern (f). These filters take several input streams and mix the inputs. MF PCM Mixer adds each input PCM sample simply. MF RGB Mixer mixes each input RGB sample in spatial.

Although these AV Sources create output samples continuously only, if rendering audiovideo samples is required, Renderable AV Source filter is available. Renderable AV Source includes an AV Source, and can insert Audio/Video Effector(s) as necessary. The effected audio/video samples are shared in two directions for rendering by the infinite pin filter or the Smart Tee filter that are included in DirectShow. One of the outputs is connected to a renderer filter and another is just an output of Renderable AV Source.

Next, Figure 16 shows connected filters for encoded sources. MF Video Depacketizer and MF Audio Depacketizer filters receive media packets continuously. File Reader is used to read encoded media files stored in a local system. The same as AV Source in Figure 15, if

rendering encoded media samples is required, Renderable Encoded Source filter is available. Furthermore, in Renderable Encoded Source, type (b) supports re-encoding, where the source filter is a Renderable AV Source. Of course, Renderable AV Source is including the all patterns of Figure 15.

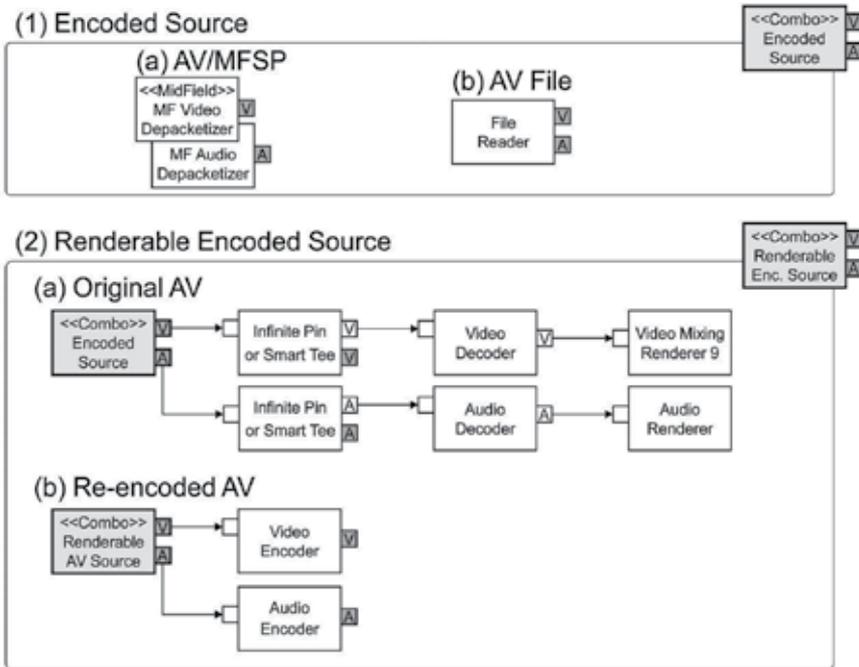


Fig. 16. Encoded Source Filters

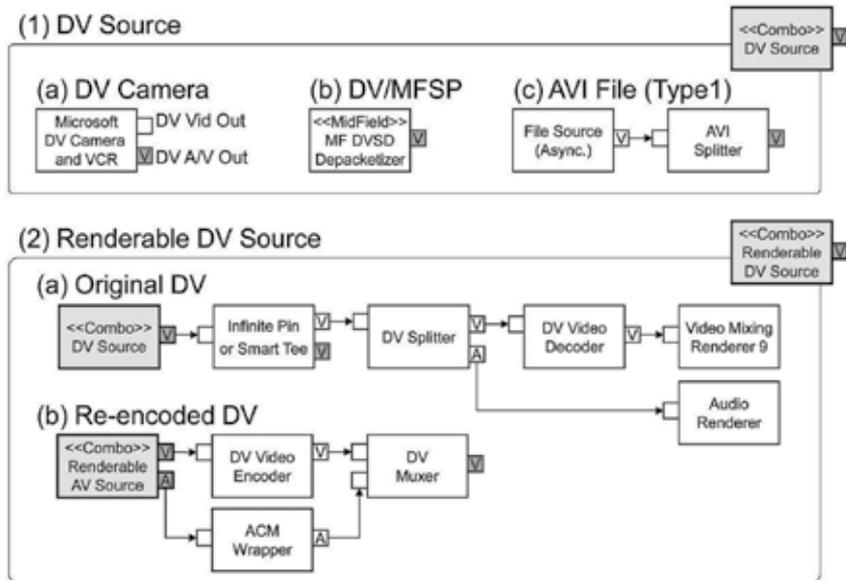


Fig. 17. DV Source Filters

Then, Figure 17 shows DV Source filters. As DV Source, our implementation supports IEEE1394 DV Camera, DV/MFSP and AVI File. The same as Encoded Source filter, both rendering and encoding of DV samples also are supported. In our system, supported filters for HDV are shown in Figure 18. The same as DV Source, our implementation supports IEEE1394 HDV camera, HDV/MFSP and MPEG2-TS File. Although rendering is available, unfortunately HDV encoding is not available.

Media Processor can select and connect these filters according to a set of input and output requirements. Then, a complete filter graph is created for handling internal media samples.

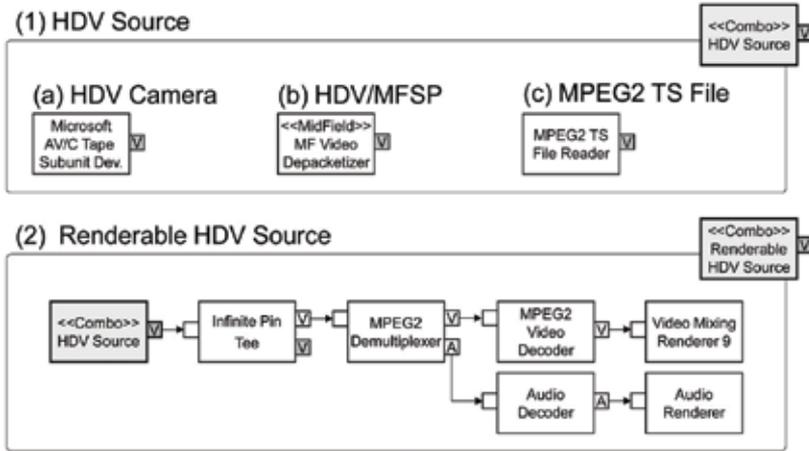


Fig. 18. HDV Source Filters

5. Current implementation

MidField System as a middleware provides both Java and C++ API for application programs. In the current implementation, Java API provides full functions that include a mobile agent platform of MidField System, and C++ API provides stream functions only. Although MidField System is a middleware system, we have implemented not only the middleware functions but also an application program for audio-video communications. The application program is just a multipoint audio-video communication system.



(a) Tiled View



(b) Picture-in-Picture

Fig. 19. Example Desktop Images of Stream Viewer

The application has a stream viewer that can customise for multipoint communication. Figure 19 (a) is a sample image of its tiled view, in this image the application is sending one video stream and receiving three video streams. Those streams are rendered in panel panes of the stream viewer respectively. Figure 19 (b) is another sample image. The application uses the video mixing function of MidField stream [Figure 15 (1) (f)] and each video stream is rendered by picture-in-picture in a panel pane.

In order to improve the performance of rendering video streams, all video images are drawn on a Java component's native window handler directly from a DirectShow's video rendering filter. Therefore, even if an application program uses the Java API of MidField System, the application is able to get high performance for rendering video images.

| | | | |
|------------------------------|-----------------------------------|------------------------------|----------------------|
| Video Input | | Audio Input | |
| Capture Device | | Capture Device | |
| Interface | IEEE1394, USB, VIDEO | Interface | Mic., Line, IEEE1394 |
| Format | DV, HDV, RGB, YUV | Format | PCM |
| Incoming Media Stream | | Incoming Media Stream | |
| Protocol | MFSP/UDP, MFSP/TCP | Protocol | MFSP/UDP, MFSP/TCP |
| Format | DV, HDV, WMV, etc. | Format | PCM, etc. |
| Media File | | Media File | |
| Format | DV(AVI), HDV(MPEG2-TS), WMV, etc. | Format | WMA, etc. |
| Video Output | | Audio Output | |
| Playback | | Playback | |
| MidField Stream Viewer | | MidField Stream Viewer | |
| Outgoing Media Stream | | Outgoing Media Stream | |
| Protocol | MFSP/UDP, MFSP/TCP | Protocol | MFSP/UDP, MFSP/TCP |
| Format | DV, HDV, WMV, etc. | Format | PCM, etc. |
| Media File | | Media File | |
| Format | DV(AVI), HDV(MPEG2-TS), WMV | Format | WMA |

Table 1. Supported I/O in Current Implementation

Table 1 shows audio and video I/O that current implementation of MidField System supports. Although a DV stream consumes about 28.8Mbps and a HDV stream consumes about 25Mbps network bandwidth, since we have supported WMV (Windows Media Video) and WMA (Windows Media Audio) as main codecs except for DV and HDV, MidField System can use WMV streams according to users' communication environments. Of course, MidField System can also employ external codecs.

Supporting these formats, by realizing adaptive connection mechanisms and relocatable transcoding functions, MidField System supports flexible and interactive communications according to users' resource environments.

6. Use cases

Since 2003, we have been using MidField System for multipoint communications. So far, the system has supported over thirty various remote communications.

The left of Figure 20 is a photograph of a symposium on disaster prevention (15 March 2005). The symposium was held on JGN2 (<http://www.jgn.nict.go.jp/english/>) that is an advanced testbed network for R&D. In the symposium, MidField System connected five remote locations in Japan. Each location sent a DV stream to a centre location and the five

DV streams were mixed for distribution, then the mixed DV stream were distributed to each location from the centre location.

We empirically know audio and video streams should be handled as separate streams in practical communication session. Although a DV is an interleaved audio-video stream, each location used not only a DV stream but also another audio stream. In this symposium, audio streams of each location also were mixed at the centre location. An audio stream that was received at each location contained audio data of the other locations only.

MidField System had supported one of distance learning projects from 2004 to 2008. The right of Figure 21 is a photograph of the distance learning class held on two locations (30 August 2007). In this distance learning, a large hydrogen ion accelerator at Tohoku University was controlled interactively from high school students in a remote location.

The location where the large hydrogen ion accelerator was in is different from the place of a lecturer, and therefore, several DV streams and WMV + PCM streams were used at the same time according to communication environments.



Fig. 20. Symposium on Disaster Prevention and Distance Learning

Since 2005, HDV streams also have been used for live video distribution. By using HDV streams we can get very high quality video images, but unfortunately HDV that employs MPEG2 codec has a delay than DV, therefore the HDV streams are not suitable for communications which require interactivity such as the above distance learning.

We are mainly using HDV streams as unidirectional multipoint live streams. The left of Figure 21 is an example of HDV distribution, which relayed Morioka Sansa Dance Festival in Morioka City of Iwate Prefecture, Japan.



Fig. 21. HDV and DV Distribution

The right of Figure 21 is a photograph of an international symposium that was held in University of Oulu, Finland (13 December 2007). MidField System was used as a unidirectional live video distribution system and a live DV stream was relayed from University of Oulu to Tohoku University.

The current implementation of MidField System can handle MIDI (Musical Instrument Digital Interface) messages. Captured MIDI messages from a musical instrument such as piano and electone are packetized and transmitted to remote musical instruments, and MidField System can make a sound on the remote instruments.

Combination of audio-video streams and MIDI streams may realize a new method of remote lesson. As shown in Figure 22, a remote lesson project is being carried out now, and if the special piano (right of Figure 22) is available the local piano's keys move up and down just like the remote piano.

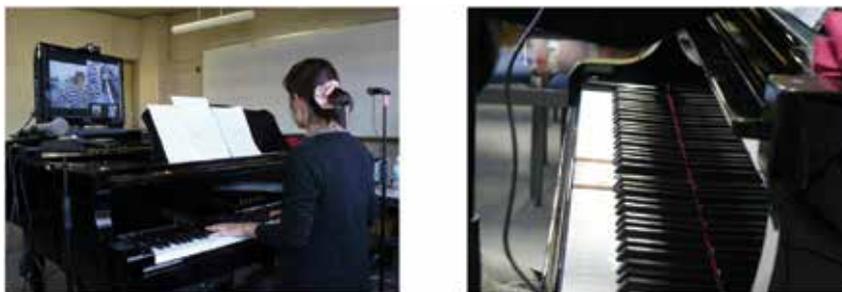


Fig. 22. Remote Lesson of Piano

Figure 23 is an example of using external filters in Stream Agent. The left of photographs is a state of experimental communications for a remote mental health care project. In this project, MidField System uses an omni-directional camera as shown in the right of photographs and captures a view of 360 degree angle of a round table. The captured ring shaped video images are converted to panorama images in realtime (Ookuzu, 2008). The system realizes this panorama communication by Stream Agent which inserted conversion filters as Video Effector in Figure 15 (2).

Using DirectShow's filter architecture in implementation of Stream Agent has enabled easy functional extensions by external software modules.



Fig. 23. Using External Filters on MidField System

Moreover, by using MidField System as a middleware, a new surveillance system has been developed that name is Telegnosis System (Sato, 2009). The system has a combination of omni-directional camera and networked Pan-Tilt-Zoom (PTZ) camera, which can detect moving objects in wide area with 360 degree by the omni-directional camera image, and extract the position of the moving object. Then the relative pan and tilt angle for the PTZ camera are calculated from the current position of the extracted moving object, and the PTZ camera can focus on the object automatically.

The left photograph of Figure 24 shows the omni-directional camera and the networked PTZ camera, which are installed under the ceiling. In the right photograph, a moving object is detected by the omni-directional camera image and the object is focused by the PTZ camera.



Fig. 24. Telegnosis System

7. Conclusions

MidField System is a **M**iddleware for **F**lexible and **i**nteractive **e**nvironment in **l**ong **d**istance. We can communicate with each other on computer networks by using the system. Supported DV and HDV video formats enable us to get high quality digital video images for interactive communications, and the system is able to construct audio-video communication sessions according to users' communication environments.

We are always improving the system through many practical audio-video communications. If we can use a HDMI (High-Definition Multimedia Interface) capture card that is able to use as a DirectShow's source filter, the latest implementation enables us to communicate each other by full HD video streams. In our ordinary communication, to use high quality digital video streams has become easier than before.

However to use the video streams in multipoint communications is not always easy. Even if we have communication environments enough to use high quality digital video streams, the preparations for audio-video communications are often hard works. To modulate levels of each location's microphones and speakers takes much time always. Trouble shootings for burst packet loss cause often us great distress. For settings of communication nodes on each location, we often need cell phones. Through practical communications, we have many problems to be solved in multipoint digital video communications.

In this chapter, the idea for integrating IP multicast sessions and the implementation have been described. The current implementation of MidField System has many useful functions for audio-video streaming as a multipoint communication system or a middleware, and enables a wide variety of audio-video communications such as high quality digital video distributions, distance educations, remote lessons and customized communication systems. On the other hand, MidField System had realized a dynamic session configuration method, but unfortunately in many cases communication systems are restricted to use IP multicast. Now, we are trying to design and develop a new adaptive multipoint session configuration method as a future work, which will be able to improve our communication environments.

8. References

- Antonio, L.; George, P. & Graham, K. (2002). Exploiting agent mobility for large-scale network monitoring, *IEEE Network*, No.3, pp.7-15.
- Bak, Y.; Kim, Kapdong., Lee,Y., Sok, S., Kim, C., Kim, H., Kim, M., Kim, Kyongsok. (2006). Design and Implementation of High Performance SMART for HD Video On Demand, *IEEE Int. Conf. on Hybrid Information Technology*.
- Bodecek, K. & Novotny, V. (2007). From standard definition to high definition migration in current digital video broadcasting, *IEEE Int. Multi-Conference on Computing in the Global Information Technology*.
- Campbell, A.T.; De Meer, H., Kounavis, M.E., Miki, K., Vicente, J. & Villela, D.A. (1999). A survey of programmable networks, *ACM Computer Communications Review*.
- Dong, H.N. & Seung, K.P. (2001). Adaptive multimedia stream service with intelligent proxy, *IEEE Int. Conf. on Information Networking*, pp.291-296.
- Ferreira, P.; Veiga, L. & Ribeiro, C. (2003). OBIWAN: design and implementation of a middleware platform, *IEEE Trans. Parallel and Distributed Systems*, Vol.14, No.11, pp.1086-1099.
- Gharai, L.; Perkins, C., Riley, R. & Mankin, A. (2002). Large scale video conferencing: a digital amphitheater, *IEEE Int. Conf. on Distributed Multimedia System*.
- Gharai, L.; Lehman, T., Saurin, A. & Perkins, C. (2006). Experiences with High Definition Interactive Video Conferencing, *IEEE Int. Conf. on Multimedia & Expo*.
- Guedes, L.A.; Oliveira, P.C., Faina, L.F. & Cardozo, E. (1997). QoS agency: an agent-based architecture for supporting quality of service in distributed multimedia systems, *IEEE Int. Conf. on Protocols for Multimedia Systems - Multimedia Networking*, pp.204-212.
- Guo, L.; Chen, S., Ren, S., Chen, X. & Jiang, S. (2004). PROP: a scalable and reliable P2P assisted proxy streaming system, *IEEE Int. Conf. on Distributed Computing Systems*, pp.778-786.
- Hashimoto, K. & Shibata, Y. (2008). Design and Implementation of Adaptive Streaming Modules for Multipoint Video Communication, *IEEE Int. Symposium on Frontiers in Networking with Applications*.
- Jameela, A.J.; Nader, M., Hong J. & David S. (2003). Middleware Infrastructure for Parallel and Distributed Programming Models in Heterogeneous Systems, *IEEE Trans. Parallel and Distributed Systems*, Vol.14, No.11, pp.1100-1111.
- Ohsuga, A.; Nagai, Y., Irie, Y., Hattori, M. & Honiden, S. (1997). PLANGENT: an approach to making mobile agents intelligent, *IEEE Internet Computing*, Vol.1, No.4, pp.50-57.
- Ookuzu, H.; Sato, Y., Hashimoto, K. & Shibata, Y. (2008). A New Teleconference System by GigaEther Omni-directional Video Transmission for Healthcare Applications, *International Computer Symposium*, pp.431-436.
- Sato, Y.; Hashimoto, K. & Shibata, Y. (2009). A Wide Area Surveillance Video System by Combination of Omni-directional and Network Controlled Cameras, *Int. Conf. on Complex Intelligent and Software Intensive System*.
- Schulzrinne, H.; Casner, S., Frederick, R. & Jacobson, V. (2003). RTP: a transport protocol for real-time applications, *RFC3550*.

Video Content Description Using Fuzzy Spatio-Temporal Relations

Archana M. Rajurkar¹, R.C. Joshi²,

Santanu Chaudhary³ and Ramchandra Manthalkar⁴

¹*Dept. of Computer Sc. and Engg. M.G.M.'s College of Engineering, Nanded - 431 605,*

²*Dept. of Electronics and Computer Engg. Indian Institute of Technology Roorkee,
Roorkee - 247 667,*

³*Dept. of Electrical Engg., Indian Institute of Technology, Delhi, New Delhi,*

⁴*Dept. of Electronics and Telecommunication Engg. S.G.G.S. Institute of Engineering,
and Technology Nanded - 431 605,*

India

1. Introduction

There has been an explosive growth in multimedia data such as images, video and audio in the past few years. The rapid proliferation of the Web made a large amount of video data publicly available. This necessitates a system that provides the ability to store and retrieve video in a way that allows flexible and efficient search based on semantic content.

Most of the existing video retrieval techniques search video based on visual features such as color, texture and shape (Zhang et al., 1997; Chang et al., 1998; Hampapur et al., 1997). A video sequence is first segmented into shots, each shot is then represented in terms of a number of key frames and then the visual features are used for retrieval (Zhang et al., 1993; Zhang et al., 1995; Mottaleb et al., 1996). Key frame based methods do not always consider events occurring in the video that involves various objects and choosing the key-frames is a challenging problem. Furthermore, in these approaches the temporal nature of video is neglected. A few systems have addressed the issue of object-based video retrieval (Deng & Manjunath, 1998; Courtney, 1997) and spatial modeling of video data that involve temporal information (Bimbo et al., 1995; Little & Ghafoor, 1993; Dagtas & Ghafoor, 1999; Vazirgiannis et al., 1998).

The semantic modeling of video content is a difficult problem. One simple way to model the video content is by using textual annotation (Oomoto & Tanaka, 1993) or visual features (Zhang et al., 1997; Chang et al., 1998), but the annotation is tedious and time consuming and simple visual features are not always sufficient to represent the rich semantics of digital video. Another way to model the video content is by using spatio-temporal concepts or events that describe the dynamic spatio-temporal behavior and interaction of video objects. As humans think in terms of events and remember different events and objects after watching a video, these high-level concepts are the most important cues in content-based retrieval (Petkovic & Jonker, 2001). For example, in a football game, we usually remember goals, interesting actions *etc.* Therefore there is a great need for development of robust tools

for modeling and extraction of spatio-temporal relations among objects visible in a video sequence.

Some approaches have been suggested in the literature for modeling the spatial and temporal relations of video objects. In most of the approaches, spatial relationships are based on projecting objects on a two or three-dimensional coordinate systems (Dagtas & Ghafoor, 1999), while temporal relations are based on Allen's temporal interval algebra (Allen, 1983). Some models, which use temporal relations for video retrieval, are proposed in (Oomoto & Tanaka, 1993; Little & Ghafoor, 1993). Very few efforts integrate both spatial and temporal aspects (Dagtas, 1998, Vazirgiannis et al., 1998; Nepal & Srinivasan, 2002; Pissiou et al., 2001). A Spatio-Temporal Logic (STL) language is presented in (Bimbo et al., 1995), which is used to describe the contents of an image sequence. A prototype image retrieval system supporting visual querying by example image sequences is discussed in this but; the problem of modeling higher-level concepts of spatio-temporal events is not addressed. A framework for semantic modeling of video data using generalized n-ary operators is presented in (Day et al., 1995). A graph-based model, called Video Semantic Directed graph (VSDG) is proposed in (Dagtas, 1998) for unbiased representation of video data using directed graph model. It also suggests the use of Allen's temporal interval algebra to model relations among objects. A model for complete declarative representation of spatio-temporal composition is presented in (Vazirgiannis et al., 1998). Nepal and Srinivasan (2002) presented a binary representation-based framework for modeling and querying video content using video object relationships at a semantic level that supports conceptual spatio-temporal queries. A topological-directional model for spatio-temporal composition of objects in video sequences is presented in (Rajurkar and Joshi, 2001). It describes the spatio-temporal relationships among objects for each frame and models temporal composition of an object.

Most of the previous spatio-temporal models do not deal with extraction of spatio-temporal relations rather they use handcrafted precise definitions of spatial relations (Egenhofer & Fanzosa, 1991) and temporal relations (Allen, 1983). It is not always possible to precisely characterize spatial and temporal relations between objects in the video sequence because of the inherent dynamic nature of the media. Further, spatial relations such as LEFT, ABOVE and others defy precise definitions, and seem to be best modeled by fuzzy sets (Matsakis & Wendling, 1999). Furthermore, errors may occur in shot segmentation and object detection due to non-robustness of the image/video processing operations leading to erroneous inference of crisp spatio-temporal relations between objects. By using fuzzy definitions of spatio-temporal relations, we can take care of these errors and ambiguities. Consequently, this approach facilitates video modeling as well as query processing based upon spatio-temporal relations extracted automatically from the actual video data without manual intervention.

Motivated by the above observation, we present a new approach for video content description. A fuzzy spatio-temporal model that is based on fuzzy directional and topological relations and fuzzy temporal relations between video objects is proposed. We use the linguistic definitions of spatial relations using *histogram of forces* presented in our earlier work (Rajurkar & Joshi, 2001). Fuzzy definitions and membership functions of temporal relations are presented and the second order fuzzy temporal relations have been proposed. A video representation scheme is presented using the proposed fuzzy spatio-temporal model. We have illustrated query processing with our fuzzy spatio-temporal model.

The rest of the chapter is organized as follows. Section 2 describes the method used for object detection using motion segmentation. The proposed fuzzy spatio-temporal model is presented in section 3. Query processing and retrieval results are given in section 4 and conclusions are presented in section 5.

2. Object detection using motion based segmentation

In our approach, a video sequence is modeled in terms of fuzzy spatio-temporal relations between objects visible in the video sequence. For sake of simplicity it is assumed that the objects of importance in a video sequence are in motion and simple motion based segmentation scheme for identification/detection of the objects is used in this work. A simple approach for motion-based segmentation using difference picture has been described here. More sophisticated approaches can improve the system performance.

In the proposed model, moving objects in every frame are detected in the video sequence using motion segmentation method described in (Courtney, 1997). For each frame F_n in the video sequence, the motion segmentation stage computes segmented image C_n as

$$C_n = T_h \bullet k$$

Where T_h is a binary image resulting from the absolute difference of images I_n and I_0 at threshold h . $T_h \bullet k$ is the morphological close operation on T_h with structuring element k . The image T_h is defined for all pixels (i, j) in T_h as

$$T_h(i, j) = \begin{cases} 1 & \text{if } |I_n(i, j) - I_0(i, j)| \geq h \\ 0 & \text{otherwise} \end{cases}$$

Connected component analysis is then performed on the segmented image C_n resulting in a unique label for each connected region. Each connected component is further recognized and identified manually for constructing the offline video model.

An example of motion segmentation process explained above is shown in Fig. 1. The reference image and the image to be segmented are shown in Fig.1 (a) and (b) respectively. Absolute difference and thresholded image is shown in Fig. 1(c) that detect motion region in the image. Fig. 1(d) shows the morphological close operation that joins small regions together into smoothly shaped objects. The result of connected component analysis that assigns each detected object a unique label is shown in Fig. 1(e). The output of the motion segmentation stage is C_n after discarding components, which are smaller than a given threshold. Fuzzy spatio-temporal relations between these objects are then defined.

Such a motion segmentation technique is best suited for video sequences containing object motion within an otherwise static scene, such as in surveillance and scene monitoring applications (Courtney, 1997).

3. Fuzzy spatio-temporal model

Spatio-temporal relations of the salient objects play a crucial role in characterization of the "content" of video. The spatial characteristics are referred to as spatial relationship between any two objects in a sampled frame, whereas the temporal characteristics are referred as the dynamics of the spatial relationships between every two objects over the frames in a video sequence.

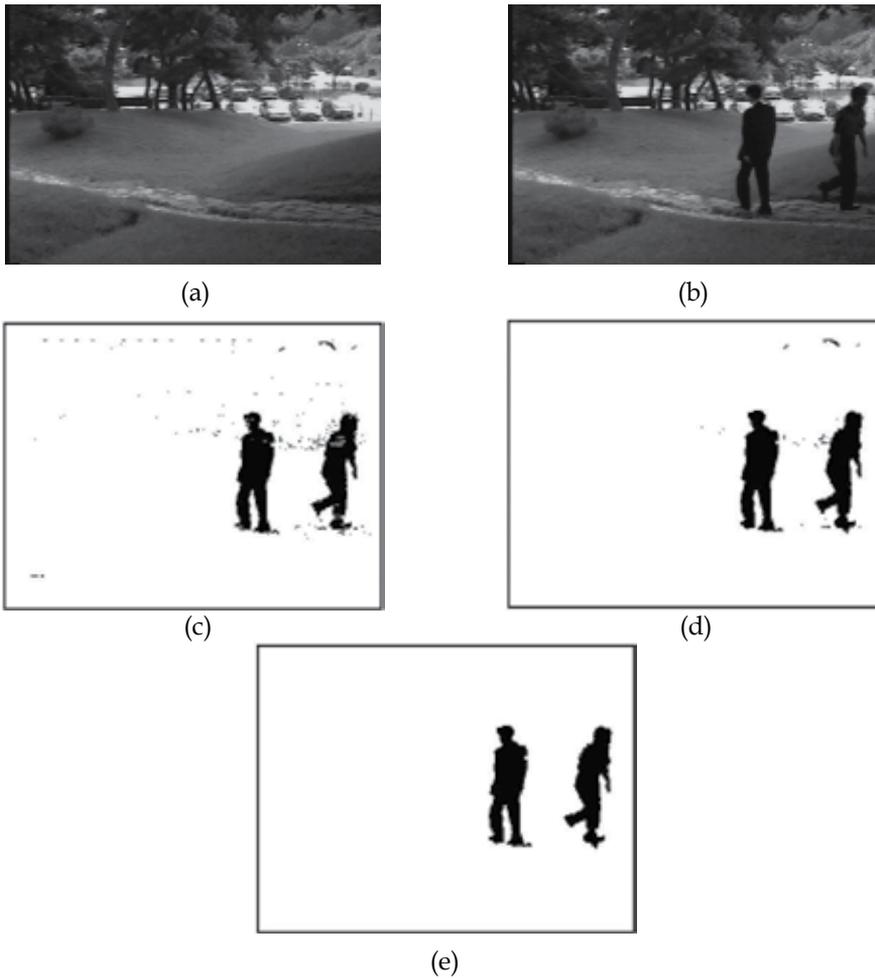


Fig. 1. Motion Segmentation example: (a) Reference Image I_0 (b) Image I_n (c) Thresholded Image T_h (d) Result of Morphological Close Operation (e) Result of Connected Component Analysis

In the proposed model the description of a scene is expressed by the mutual spatial relationships between every two objects and the temporal change in their relationships. Frame intervals specify the period for which a particular spatial relationship holds. For each object O_{it} in a frame t , its fuzzy spatial and temporal relationships, $O_{ST_{ij}t}$, with every other object, O_{jt} , in the same frame are recorded in a vector of size $n(n-1)$, where n is the number of objects in the frame t . The fuzzy spatio-temporal relationship between the two objects is defined as the function:

$$O_{ST_{ij}t}(O_{it}, O_{jt}) = (S_{ijt}, T_{ijt}) \quad (1)$$

where S_{ijt} and T_{ijt} are the spatial and temporal relationships, respectively, between the objects O_{it} and O_{jt} in the frame t .

3.1 Spatial relationships

The relative positions between two objects O_i and O_j can be captured as a fuzzy spatial relation using *histogram of forces* (F-Histogram) (Matsakis & Wendling, 1999). The spatio-temporal models presented earlier in (Dagtas & Ghafoor, 1999; Nepal and Srinivasan, 2002; Pissiou et al., 2001) use precise definitions of spatial relations using either angle measurements or minimum bounding rectangles (MBR). Spatial relations such as *left of*, *above* and others defy precise definitions, and seem to be best modeled by fuzzy sets (Matsakis & Wendling, 1999). Miyajima and Ralescu (1994) developed the idea of representing the relative position between two objects by *histogram of angles*. Matsakis and Wendling (1999) introduced the notation of the F-histogram, which generalizes and supersedes that of the histogram of angles is discussed in the following section.

3.2 F-histograms

The F-histogram represents the relative position of a 2D object A with regard to another object B by a function FAB from \mathbf{R} in to \mathbf{R}^+ . For any direction θ , the value $FAB(\theta)$ is the total weight of the arguments that can be found in order to support the proposition "A is in direction θ of B". Object A is the argument and object B is the referent. The directional spatial relations between objects are defined from a fuzzy subset of \mathbf{R} . Its membership function μ is continuous, with a period 2π that decreasing on $[0, \pi]$, and takes the value 1 at 0 and the value 0 at $\pi/2$ shown in Fig. 2(a). It can be employed to define a family of fuzzy directional relations between points (Rajurkar & Joshi, 2001). We used the typical triangular function graphed in Fig. 2(a). Let α and β be two reals and A and B be two distinct points. If β is an angle measure, then (Fig. 2(b)): $R_\alpha(A, B) = \mu(\beta - \alpha)$. In our earlier work (Rajurkar & Joshi, 2001) we have presented definitions of fuzzy directional spatial relations and topological relations using F-histogram. We make use of these definitions of fuzzy spatial relations as perceived by humans for capturing relative position of a 2D object O_i with regards to another object O_j . Each pair of objects in every frame in the video sequence is represented by relative position histograms and then the degree of truth of a proposition "A is in direction θ of B" is computed. The degree of truth is a real number generated greater than or equal to 0 (proposition completely false) and less than or equal to 1 (proposition completely true).

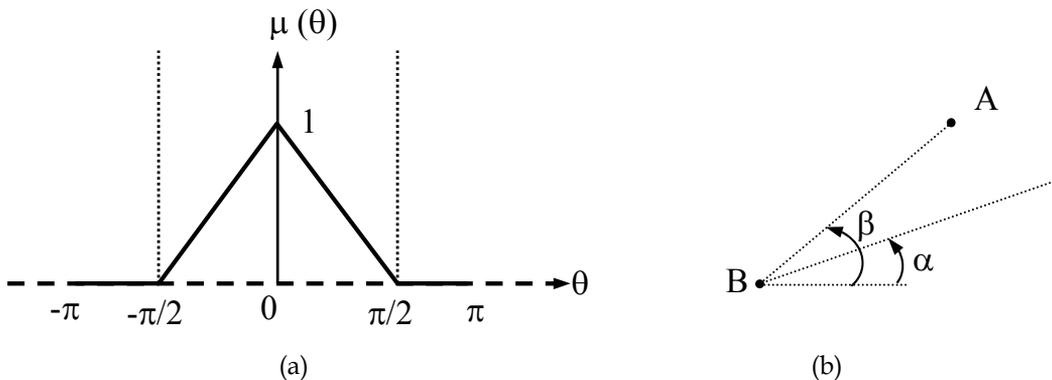


Fig. 2. (a) Example of directional spatial relation between points. (b) The degree of truth of the proposition "A is in direction α of B" is $\mu(\beta - \alpha)$

The spatial relationships, S_{ijt} , between two objects are defined as follows:

$$S_{ijt} = (R_{ijt}, O_{it}, O_{jt}) \quad (2)$$

The R_{ijt} represents the degree of truth of a proposition "A is in direction θ of B" computed as described in (Rajurkar & Joshi, 2001).

3.3 Temporal relationships

Allen (1983) introduced temporal interval algebra for representing and reasoning about temporal relations between events represented as intervals. The elements of the algebra are sets of seven basic relations that can hold between two intervals. The relations are not reflective except equals and a set of inverse relations exists for the other six relations. A list and graphical illustrations of the binary interval relations is provided in Fig. 3. A qualitative description of relative position of objects (MBRs) when their projections on X and Y-axes are considered as interval operands to the relations are provided by interval relations.

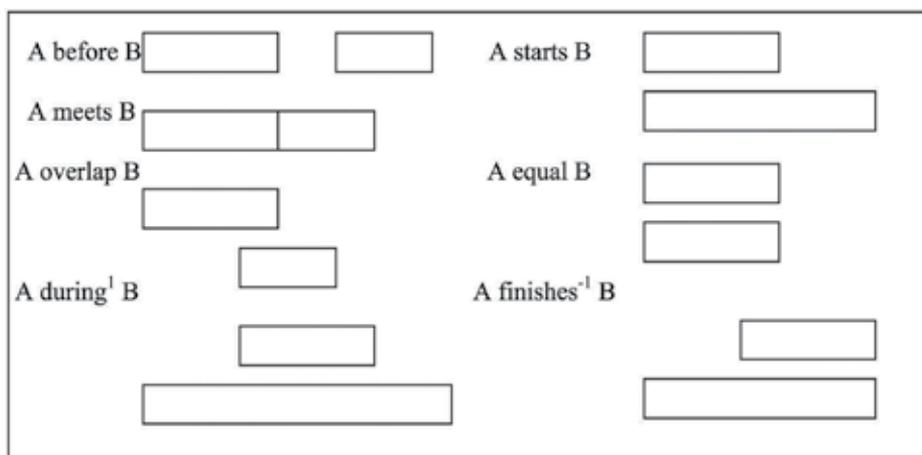


Fig. 3. Binary Interval Relations

Crisp description of temporal relations may not be suited for defining spatio-temporal models of video because of errors in segmentation, event and object detection due to noise and processing errors. We present some examples to illustrate this point. Figure 4 shows an example of error occurred in segmentation. Original image is shown in Fig. 4(a). Thresholded absolute difference image is shown in Fig. 4(b), and the results of motion segmentation after morphological close operation and connected component analysis is shown in Fig. 4(c). It is observed that though the objects *Person 1* and *Person 2* are disjoint after the segmentation they appear to be meet. Further it is also shown that how errors can occur in event detection and segmentation of noisy images. Fig. 5(a) shows an example of noisy image. The thresholded absolute difference image is shown in Fig. 5(b), and results of segmentation are given in Fig 5(c). It can be seen that though the objects *Person 1* and *Person*

2 are disjoint they appear to be meet after segmentation. Fig. 6 shows an example of error occurred in object detection. Fig. 6(a) shows the original image and the results of segmentation are presented in Fig. 6(b). It can be observed that half of the part of *Person2* is missing in the result, which may lead to wrong event and object detection. We present fuzzy definitions of temporal relations in this section that can take care of these errors in event-detection, segmentation and object detection and allows flexible description of video scene (see Table 1). Relations defined in Table 1 are applicable between objects, in particular for the intervals for which two objects are visible. For example, *A before B* with a membership function " μ " means that the *object A* is not visible, in general, in the scene when *B* appears. The temporal relation T_{ij} between the spatial relationships, $S_{ij}t$, of objects $O;t$ and O_jt can be described at two levels. In the first level, the temporal interval Δf for which a fuzzy spatial relationship between the two objects is valid is determined. In the second level, the second order fuzzy temporal relationships between the two spatial relations are described. The advantage of the second order fuzzy temporal relations is that they are more informative and provide global description of a sequence. The proposed second order fuzzy temporal relations using fuzzy spatial relations are shown in the Table 2. Graphical illustration of two second order fuzzy temporal relations are presented in Fig. 7. The temporal relationship, T_{ij} , between the temporal intervals of two spatial relationships, $S_{ij}t$ and $S_{ij}t'$ is defined as follows:

$$T_{ij}^1 = (S_{ij}t_{\Delta f})$$

$$T_{ij}^2 = (S_{ij}t_{\Delta f} \langle FTOP \rangle S_{ij}t'_{\Delta f}) \quad (4)$$

where $\langle FTOP \rangle$ is one of the temporal operators representing the fuzzy temporal relationship between two intervals and Δf is temporal interval for which fuzzy spatial relationship S_{ij} is valid.

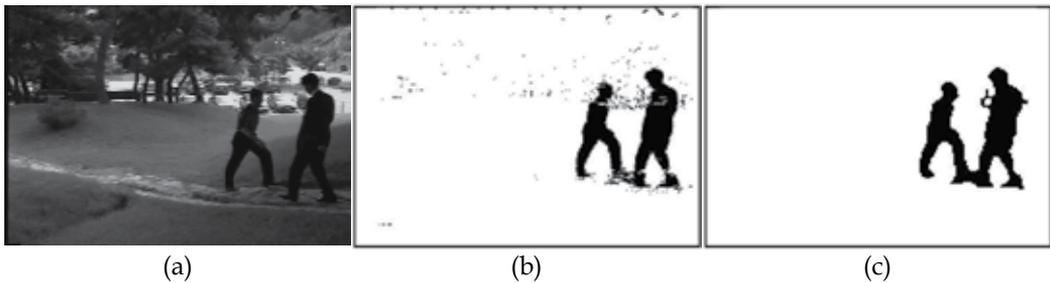


Fig. 4. Error Occurred in Segmentation (a) Original Image (b) Thresholded Absolute Difference Image (c) Results of Segmentation

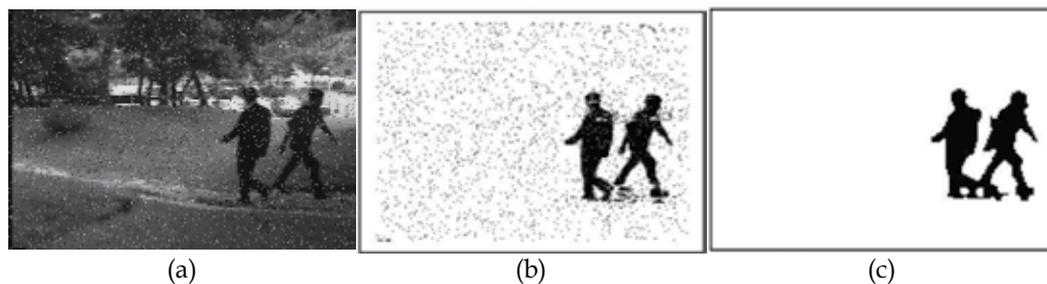


Fig. 5. Error Occurred in Segmentation (a) Noisy Image (b) Thresholded Absolute Difference Image (c) Results of Segmentation

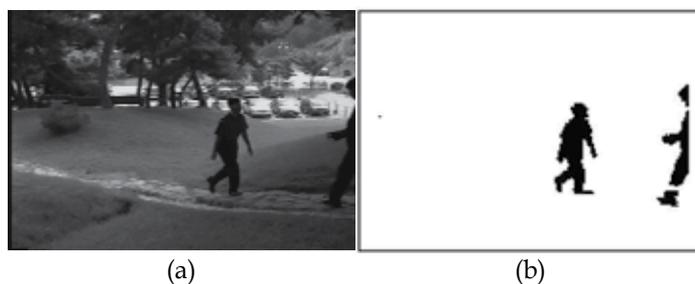


Fig. 6. Error Occurred in Event and Object Detection (a) Original Image (b) Results of Segmentation

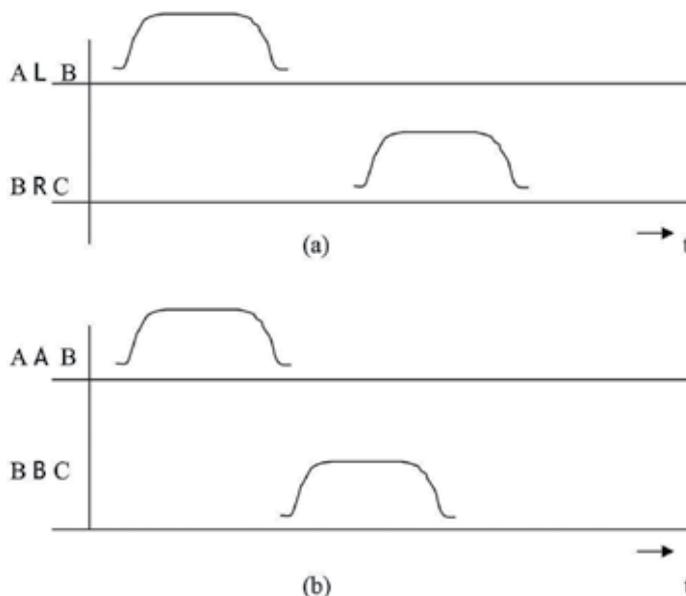


Fig. 7. Graphical Illustration of Second Order Fuzzy Temporal Relations (a) Relation A L B μ_{before} B R C (b) Relation A A B μ_{meets} B B C

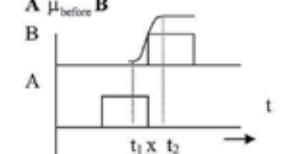
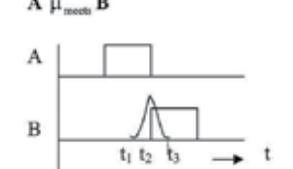
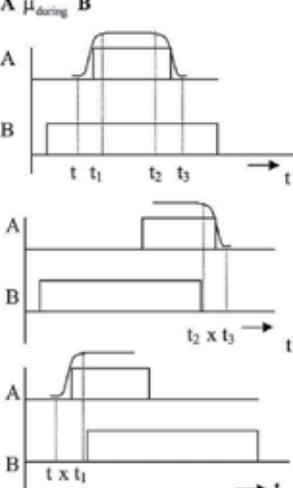
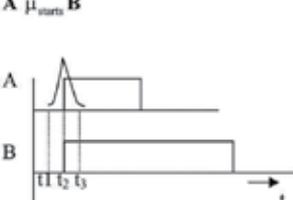
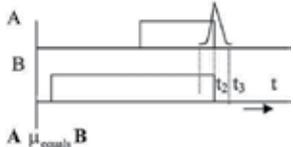
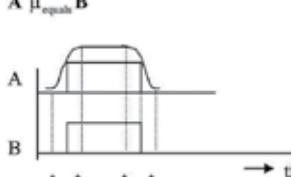
| Fuzzy Temporal Relationship | Membership Function | Logical Description |
|---|--|--|
|  | $\mu_{\text{before}}(x) = 1 \quad x \geq t_2$ $\mu_{\text{before}}(x) = e^{-(x-t_2)^2} \quad x < t_2$ | <p>Object A disappear before the appearance of the object B, but there may be some overlap</p> |
|  | $\mu_{\text{meet}}(x) = 1 \quad x = t_1$ $\mu_{\text{meet}}(x) = e^{-(x-t_2)^2} \quad x < t_2, x > t_2$ | <p>Object A meet object B, but there may be some overlap or they may be disjoint</p> |
|  | $\mu_{\text{during}}(x) = 1 \quad t_1 \leq x \leq t_2$ $\mu_{\text{during}}(x) = e^{-(x-t_1)^2} \quad x < t_1$ $\mu_{\text{during}}(x) = e^{-(x-t_2)^2} \quad x > t_2$ | <p>Object A is appearing and disappearing during object B, but it may appear before object B or it may disappear after object B.</p> |
|  | $\mu_{\text{starts}}(x) = 1 \quad x = t_2$ $\mu_{\text{starts}}(x) = e^{-(x-t_2)^2} \quad x < t_2, x > t_2$ | <p>Object A starts exactly with Object B, but it may start little before or after object B.</p> |
|  | $\mu_{\text{finishes}}(x) = 1 \quad x = t_2$ $\mu_{\text{finishes}}(x) = e^{-(x-t_2)^2} \quad x > t_2, x < t_2$ | <p>Object A finishes exactly with Object B, but it may finish little before or after object B.</p> |
|  | $\mu_{\text{equals}}(x) = 1 \quad t_1 \leq x \leq t_2$ $\mu_{\text{equals}}(x) = e^{-(x-t_1)^2} \quad x < t_1$ $\mu_{\text{equals}}(x) = e^{-(x-t_2)^2} \quad x > t_2$ | <p>Object A starts and finishes exactly with Object B, but it may start or finish little before or after object B.</p> |

Table 1. Fuzzy Temporal Relationships and Corresponding Logical Descriptions

3.4 Representation of video sequence using fuzzy spatio-temporal model

This section presents the representation of a video sequence using the proposed fuzzy spatio-temporal model. The video database V_{db} contains video sequences S_1, S_2, \dots, S_n as follows:

$$V_{db} = \langle S_1 \rangle \langle S_2 \rangle \langle S_3 \rangle \dots \langle S_n \rangle \tag{5}$$

A video sequence S is an ordered set of n frames (t), denoted $S = \{t_0, t_1, t_3, \dots, t_n\}$, where t_n is the n^{th} frame in the sequence. For each frame t in the video sequence S the moving objects are detected and labeled using motion segmentation method outlined in Section 2. Then the attributes of labeled objects are derived.

| Spatial Relation 1 | Temporal Relation | Spatial Relation 2 | Definition of Second Order Fuzzy Temporal Relations |
|--------------------|-------------------|--------------------|---|
| A left of B | Before | B right of C | A L B μ_{before} B R C |
| A above B | Before | B below C | A A B μ_{before} B B C |
| A left of B | Meets | B left of C | A L B μ_{meets} B L C |
| A above B | Meets | B below C | A A B μ_{meets} B B C |
| A left of B | Overlap | B right of C | A L B μ_{overlaps} B R C |
| A above B | Overlap | B below C | A A B μ_{overlaps} B B C |
| A left of B | During | B right of C | A L B μ_{during} B R C |
| A above B | During | B below C | A A B μ_{during} B B C |
| A left of B | Starts | B right of C | A L B μ_{starts} B R C |
| A above B | Starts | B below C | A A B μ_{starts} B B C |
| A left of B | Finishes | B right of C | A L B μ_{finishes} B R C |
| A above C | Finishes | B below C | A A C μ_{finishes} B B C |
| A left of B | Equals | B right of C | A L B μ_{equals} B R C |
| A above B | Equals | B below C | A A B μ_{equals} B B C |

Note: Set of inverse relations exists for all above relations except *equals*. The symbols used for spatial relations are L - left of, R - right of, A - above, B - below.

Table 2. Second Order Fuzzy Temporal Relations

For each object $O_i t$ of a frame t , its fuzzy spatial and temporal relationship $O_{ST_j} t$, with every other object, $O_j t$, is represented using the proposed fuzzy spatio-temporal model as discussed in Section 3.

$$O_{ST_j} t(O_i t, O_j t) = (S_{ij} t, T_{ij} t) \tag{6}$$

where S_{ijt} and T_{ijt} are defined by Eq. (2) and (3) respectively.

For each frame t fuzzy spatio-temporal relationship between all object pairs (e.g. suppose there are four objects in the frame t) is represented as follows:

$$O_{ST}t = (O_{ST_{12}}, O_{ST_{13}}, O_{ST_{14}}, O_{ST_{23}}, O_{ST_{24}}, O_{ST_{34}}) \quad (7)$$

To capture the dynamic change in the fuzzy spatial relationship of two objects $O_i t$ and $O_j t$ over the video scene, interval of l frames in which the corresponding spatial relation is valid is determined. The temporal interval of a spatial relationship is found from the frame t of the initial appearing of a particular spatial relationship, which represent the beginning of the temporal interval in which that spatial relationship is valid. Then the frame t' of the initial appearance of the first different relationship is determined. Thus the duration of the temporal interval in which a particular spatial relation is valid is $t' - t$. Based on these durations, fuzzy spatio-temporal relations are computed for describing the video sequences. The detailed algorithm for representing video sequences in a video database using the proposed fuzzy spatio-temporal model is presented in Fig. 8.

Algorithm Represent video_{stm}

Input: V_{db} (video database)

Output: Representation of video database sequences.

Procedure:

For every video sequences in the video database do the following

1. Motion segmentation

For each frame t in the database video sequence, the motion segmentation stage computes segmented image C_n as $C_n = T_n \bullet k$. Connected component analysis is then performed on the segmented image C_n and objects are identified, labeled and their attributes are derived.

2. Compute fuzzy spatio-temporal relations

For every frame in the video sequence S , for each object pair ($O_i t$, $O_j t$)

Compute fuzzy spatial relations $S_{ijt} = (R_{ijt}, O_i t, O_j t)$

Compute the temporal interval Δt for which the spatial relationship S_{ijt} is valid

$$T_{ij}^1 = (S_{ijt} \Delta t)$$

3. Compute the second order fuzzy temporal relations between every two different spatial relations.

$$T_{ij}^2 = (S_{ij} t_{\Delta t} < FTOP > S_{ij} t'_{\Delta t})$$

Fig. 8. Algorithm Represent_{video}_{stm}

We used MPEG-7 video content *ETRI_od_A.mpg* to illustrate the proposed video sequence representation scheme. Fig. 9 shows few frames in the video sequence *ETRI_od_A.mpg*. Moving objects in the sequence are detected and labeled as *Person1*, *Person2*, *Person3* and *Person4* using the motion segmentation method described earlier in section 2 and their attributes are derived. Initial appearing of two objects *Person1* and *Person 2* is detected in frame 173 and their spatial relation is “*Person1 in white shirt is left of Person2 in black shirt*” shown in Fig.9. The temporal interval in which this spatial relationship is valid starts from

frame 173 and ends at frame 290. The same spatial relationship holds for another temporal intervals from frame 2316 to frame 2343 and from the frame 2792 to frame 2900.

The fuzzy spatio-temporal relationship is given by

$$O_{ST_{12}}t(O_1 173, O_2 173) = (S_{12} 173, T_{12} 173) \quad (8)$$

$$S_{12} 173 = (\text{left of}, \text{Person1}, \text{Person2})$$

$$T_{12}^1 = 354$$

Similarly initial appearing of the different spatial relationships between objects *Person1* and *Person2* is detected in frame 636 and its temporal interval is computed as described above. The procedure is repeated for the initial appearing of the other object pairs in the video sequence and their fuzzy spatio-temporal relationships are computed.



Fig. 9. Few Sample Frames in Video ETRI_od_A.mpg

The second level temporal relationship is determined by computing the second order fuzzy temporal relationships between the two different spatial relationships. In the video sequence *ETRI_od_A.mpg* the first two spatial relations are detected in frame 173 and 636 and their first level temporal intervals are 116 and 187. The second level temporal relationship is given by:

$$T_{12}^2 = (S_{12_{116}} 173 \langle \mu_{\text{before}} \rangle S_{12_{187}} 636) \quad (9)$$

$$S_{12} 173 = (\text{left of, Person1, Person2})$$

$$S_{12} 636 = (\text{right of, Person1, Person2})$$

The second level temporal relationships between all the different spatial relationships found in the video sequence are determined as described above. In this way all the video sequences in the database are represented by the fuzzy spatio-temporal relationships among the objects in the sequence.

4. Querying and retrieval

Formulation of meaningful and clear query is very important for searching and retrieving video. Content-based retrieval of video is rather difficult than that of the image due to the temporal information in the video. The issue of efficient handling queries related to the spatio-temporal relationships among objects is discussed here. We perform the ranking of the database sequences that are similar to the query using the relevance membership function (RMF). The temporal intervals for all spatial relations between two objects over a sequence are computed to decide the maximum number of frames for which a particular spatial relationship between two objects is valid. Depending on the query, the relevance membership function, which is the ratio of total number of frames in the sequence to the maximum number of frames for which the spatial relationship in the query is valid for every sequence in the database, is computed. The detailed algorithm *Search_Video_{stm}* for searching and ranking the video sequences in the database similar to the query V_{qr} is presented in Fig. 10.

In the following, we present some examples of queries using the spatio-temporal properties.

Query1: Find a video sequence in which a *Person P1* approaches *Person P2* from the left.

```
Select      S
From        sequences  $S_n$ 
Where       S contains P1 and P2 and P1 L P2
```

Query 2: Find a video sequence in which a *Car* is moving to the right reaches to a *House*

```
Select      S
From        sequences  $S_n$ 
Where       S contains a Car and a House and Car R House
```

Query 3: Find a video sequence in which *Car A* is in the left of *Car B* before *Car C* is in the right of *Car D* in the race.

```
Select      S
From        sequences  $S_n$ 
Where       S contains Car A, Car B, Car C and Car D and Car A L Car B  $\mu_{\text{before}}$  Car C R Car D
```

Query 4: Find a video sequence in which *Flowers* are right to the *Lake* and a *Jeep* is passing besides the *Lake*.

Select S
 From sequences S_n
 Where S contains *Flowers, Lake* and *Jeep* and *Flowers R Lake* μ_{overlaps} *Jeep*.

We used *MPEG-7* video content to evaluate the effectiveness of our query processing scheme. Video sequences in the database are represented using the proposed fuzzy spatio-temporal model as described in section 3.4.

Algorithm Search_Video_{fstm}

Input: V_{qr}

Output: Ranking of video sequences that are similar to the query V_{qr}

Procedure:

1. For every video sequences in the video database compute the relevance membership function (RMF)

$$RMF = \frac{\text{Number of frames in the database video sequence } S}{\text{Maximum number of frames for which the spatial relation in the query } V_{qr} \text{ is valid}}$$

2. Rank the video sequences in the database in the increasing order of the value of RMF

Fig. 10. Algorithm Search_Video_{fstm}

Now, consider a query “*Find video sequences in which Person1 in white shirt is left of Person2 in black shirt*”. The query results are shown in Table 3. The temporal intervals representing the number of frames for which a spatial relationship *left of* is valid for every sequence and the RMF values are given in the Table 3. Low value of RMF function corresponds to the more number of frames while higher value of RMF corresponds to the less number of frames that satisfy the spatial relationship *left* in the sequence. The sequences in the database are then ranked depending on the value of RMF. The video sequences having low RMF values are retrieved as similar to that of the spatio-temporal relation described in the query. For the query in question the most similar sequence in the video database is *ETRI_od_A.mpg*, which has maximum number of frames *i.e.* 273 for which spatial relationship *Person1 in white shirt is left of Person 2 in black shirt* is valid. The database video sequences are ranked as 1, 3, 2 depending on the RMF value for the query in question.

Consider a query “*Find video sequences in which a blue car is going towards North*”. Table 4 shows the results for this query. The temporal intervals representing the number of frames for which a spatial relationship *blue car is going towards north* is valid for every sequence and the RMF values are given in the Table 4. The video sequences in the database are ranked depending on the RMF values. For the above query the most similar video sequence is *speedwa2.mpg*, which has the maximum number of frames *i.e.* 1132 that satisfy the spatial relationship mentioned in the query. The ranking of the database video sequences for this query is 2, 3, 4, 1.

Consider another query “*Find video sequences in which red car is standing right side of the road*”. The results for this query are shown in Table 5. The most similar video sequence for this query in the database is *speedwa5.mpg* and the ranking obtained for the database video sequences is 5, 4, and 3.

5. Conclusions

We have presented a fuzzy spatio-temporal model for video content description that supports spatio-temporal queries. The proposed model is based on fuzzy directional and topological relations and fuzzy temporal relation. The problems associated with the use of crisp spatio-temporal relations were highlighted. It is shown that errors may occur in segmentation, event detection and object detection due to noisy video data and use of precise spatial relations. In order to minimize these errors fuzzy definitions of temporal relations are proposed. In addition, the second order temporal relations are presented that are more informative and provide global information about the sequence. The proposed model provides a mechanism that represents the fuzzy spatio-temporal relationships among the objects in video sequences in the database and ranks the database sequences based on the query for effective content-based retrieval. We reported the results of our experiment on a sample video from MPEG-7 data set.

| Video Sequence | Spatial Relation <i>Person 1 in white shirt is left of Person 2</i> | Relevance Membership Function (RMF) * |
|----------------------|--|---------------------------------------|
| 1 (ETRI_od_A.mpg) | 273 | $3086/273 = 11.3040$ |
| 2 (ETRI_od_B.mpg) | 0 | $3455/0 = \text{Positive Infinity}$ |
| 3 (ETRI_od_C.mpg) | 219 | $8910/219 = 40.6849$ |

* RMF value 1 indicate maximum relevance and positive infity indicate minimum relevance

Table 3. Temporal Intervals and Relevance Membership Function (RMF) Values for the Query "Find The VideoSequences in Which Person1 in White Shirt is Left of Person 2 in Black Shirt" for the Video Sequences in MPEG-7 Video Content Set

| Video Sequence | Spatial Relation <i>Blue car is going towards north</i> | Relevance Membership Function (RMF) |
|---------------------|--|-------------------------------------|
| 1 (speedwa1.mpg) | 87 | $1420/87 = 16.32$ |
| 2 (speedwa2.mpg) | 1132 | $13185/1132 = 11.64$ |
| 3 (speedwa3.mpg) | 388 | $14019/388=36.13$ |
| 4 (spedwa4.mpg) | 100 | $7496/100=74.96$ |
| 5 speedwa5.mpg) | 0 | $7497/0=\text{Positive Infinity}$ |

Table 4. Temporal Intervals and Relevance Membership Function (RMF) Values for the Query "Find the Video Sequences in which Blue Car is Going Towards North" for the Video Sequences in the MPEG-7 Video Content Set

| Video Sequence | Spatial Relation <i>Red car is standing right side of the road</i> | Relevance Membership Function (RMF) |
|---------------------|---|-------------------------------------|
| 1 (speedwa1.mpg) | 0 | 1420/0 = Positive Infinity |
| 2 (speedwa2.mpg) | 0 | 13185/0 = Positive Infinity |
| 3 (speedwa3.mpg) | 452 | 14019/452=31.01 |
| 4 (spedwa4.mpg) | 1506 | 7496/1506=4.97 |
| 5 speedwa5.mpg) | 1834 | 7497/1834=4.08 |

Table 5. Temporal Intervals and Relevance Membership Function (RMF) Values for the Query "Find the VideoSequences in which Red Car is Standing Right Side of the Road" for the Video Sequences in the MPEG-7Video Content Set

6. References

- Allen J.F., (1983). Maintaining knowledge about temporal intervals. *Communications of ACM* 26 (11), pp. 832-843.
- Bimbo A.D., Vicario E., Zingoni D., (1995). Symbolic description and visual querying of image sequences using spatio-temporal logic. *IEEE Trans. Knowledge and Data Engineering* 7 (4), pp. 609-621.
- Chang S. F., Chen W., Meng H., Sundaram H., Zhong D., (1998). A fully automated content-based video search engine supporting spatiotemporal queries. *IEEE Trans. on Circuits and Systems for Video Technology* 8 (5),pp. 602-615.
- Courtney J.D., 1997. Automatic video indexing via object motion analysis. *Pattern Recognition* 30 (4), pp. 607-625.
- S. Dagtas, A. Ghafoor, (1999). Indexing and retrieval of video based on spatial relation sequences, *Proc. ACM International Multimedia Conf. (part 2) Oriando, FL, USA, , pp. 119-121.*
- Dagtas S., (1998). *Spatio-Temporal Content-Characterization and Retrieval in Multimedia Databases*, Ph.D. Thesis, Purdue University.
- Deng Y., Manjunath B.S., (1998). NeTra-V: Toward an object-Based video representation. *IEEE Trans. Circuits and Systems for Video technology* 8 (5), pp. 616-627.
- Day Y.F., Dagtas S., Lino M., Khokhar A., and Ghafoor A., 1995. Object-oriented conceptual modeling of video data. *Proc. Eleventh Int. Conf. on Data Engineering, Taipei, Taiwan, pp.401-408.*
- Egenhofer M.J., Fanzosa R., (1991). Point-set topological spatial relations. *International Journal on Geographic Information Systems* 5 (2), pp.161-174.
- El-Kwae E., Kabuka M. R., (1999). A robust framework for content-based retrieval by spatial similarity in image databases. *ACM Trans. on Information Systems* 17(2), pp.174-198.

- Gudivada V., Raghvan V., (1995). Design and evaluation of algorithms for image retrieval by spatial similarity. *ACM Trans. on Information Systems* 13 (2), pp.115-144.
- Hampapur A., Gupta A., Horowitz B., Shu C.F., Fuller C., Bach J., Gorkani M., Jain R., (1997). Virage Video Engine. *Proc. SPIE Storage and retrieval for Image and Video Databases V*, San Jose, pp.188-197.
- Little T.D.C., Ghafoor A., (1993). Interval-based conceptual models for time dependent multimedia data. *IEEE Trans. Knowledge and Data Engineering* 5 (4), pp.551-563.
- Li J. Z., Ozsu M. T., Szafron D., (1997). Modeling of video spatial relationships in an object database management system. *Proc. IS & T/ SPIE Int'l Symposium on Electronic Imaging: Multimedia Computing and Networking*, San Jose, USA, pp.81-90.
- Li J. Z., Ozsu M. T., Szafron D., Oria V., (1997). MOQL: A multimedia object query language. *Proc. Third International Workshop on Multimedia Information Systems*, Italy, pp. 19-28.
- Mottaleb M.A., Dimitrova N., Desai R., Martino J., (1996). CONIVAS: COnent-based image and video access system. *Proc. ACM Multimedia Conf.*, Boston, MA USA, pp. 427-428.
- Matsakis P., Wendling L., (1999). A new way to represent the relative position between areal objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21 (7), pp. 634-643.
- Miyajima K., Ralescu A., (1994). Spatial organization in 2D segmented images: representation and recognition of primitive spatial relations. *Fuzzy Sets and Systems* 65, pp. 225-236.
- Nabil M., Ngu A., Shepherd J., (1996). Picture similarity retrieval using the 2D projection interval representation. *IEEE Trans. on Knowledge and Data Engineering* 8 (4), pp. 533-539.
- Nepal S., Srinivasan U., (2002). Spatio-temporal modeling and querying video databases using high level concepts. *Proc. Sixth Working Conference on Visual Database System (VDB 6)*, 29-31 May, Brisbane, Australia .
- Oomoto E., Tanaka K., (1993). OVID: Design and implementation of a video-object database system. *IEEE Trans. on Knowledge and Data Engineering* 5 (4), pp. 629-643.
- Petkovic M., Jonker W., (2001). Content-based video retrieval by integrating spatio-temporal and stochastic recognition of events. *Proc. IEEE workshop on Detection and Recognition of Events in Video*, Vancouver, Canada.
- Pissiou N., Radev I., Makki K., Campbell W.J.,(2001). Spatio-temporal composition of video objects: representation and querying in video database systems. *IEEE Trans. on Knowledge and Data Engineering* 13 (6), pp. 1033-1040.
- Rajurkar A. M., Joshi R.C., (2001). Content-Based Image Retrieval: A fuzzy spatial similarity approach. *Proc. of the International Symposium on Artificial Intelligence, ISAI'2001*, Fort Panhala (Kolhapur), INDIA, Dec pp. 18-20.
- Vazirgiannis M., Theodoridis Y., Sellis T., (1998). Spatio-temporal composition and indexing for large multimedia applications. *Multimedia Systems* 6, pp. 284-298.
- Zhang H.J., Wu J., Zhong D., Smoliar S.W., (1997). An integrated system for content-based video retrieval and browsing, *Pattern Recognition* 30 (4), pp. 643-658.

-
- Zhang H.J., Kankanhalli A., Smoliar S.W., (1993). Automatic partitioning of full-motion video. *Multimedia Systems* 1 (1), pp. 10-28.
- Zhang H.J., Low C.Y., Smoliar S.W., Wu J.H., (1995). Video parsing, retrieval and browsing: an integrated and content-based solution. *Proc. of the ACM Multimedia Conf.*, San Francisco, CA, November 5-9, pp. 15-24.

Trick play on Audiovisual Information for Tape, Disk and Solid-State based Digital Recording Systems

O. Eerenberg and P.H.N. de With
*NXP Semiconductors Research, CycloMedia Technology /
Eindhoven University of Technology
The Netherlands*

1. Introduction

Digital video for consumer applications became available in the early 1990s, enabled by the advances in video compression techniques and associated standards and their efficient implementation in integrated circuits. Convergence of transform coding and motion compensation into a single hybrid coding scheme, resulting in various standards in the early 1990s, such as MPEG-1 (ISO/IEC11172-2, 1993), MPEG-2 (ISO/IEC13818-2, 2000) and DV (CEI/IEC61834, 1993). These standards are capable of compressing video at different quality levels with modest up to high compression ratios and differing complexity. Each of the previous standards has been deployed in a digital storage system, based on different storage media.

The above hybrid coding schemes, distinguish intraframe and interframe coded pictures. The former, also labelled as I-type pictures, can be independently decoded so that it can be used for video navigation. The latter coding form result in P-type and B-type pictures requiring surrounding reference pictures for reconstruction. The availability of these reference pictures cannot be guaranteed during fast-search trick play, which makes these pictures unsuitable for certain navigation functions. Digital consumer storage standards are equipped with locator information facilitating fast-search trick play. The locator information considers the data dependencies and enables the system entry points for proper decoding. These mechanisms form the basis for selectively addressing coded data and the associated data retrieval during trick play.

This chapter discusses trick play for push- and pull-based architectures and elaborates on their implementation for tape, optical and disk-based storage systems. Section 2 introduces traditional and advanced video navigation. Section 3 presents the concepts of low-cost trick play. Section 4 elaborates on trick play for tape-based helical-scan digital video recording. Section 5 discusses trick play in relation with three popular optical-storage systems. Section 6 introduces trick play for push- or pull-based personal video recording deploying a hard-disk drive or solid-state disc. The chapter concludes and presents a future outlook in Section 7.

2. Navigation methods

Video navigation is defined as video playback in non-consecutive or non-chronological order as compared to the original capturing order. Video navigation can be divided into

traditional fast forward or fast rewind playback and advanced search methods, which are modern forms of video navigation. The former is found in analogue and digital video recorders. The latter has become possible for random accessible media such as disc and silicon-based memories. This section covers the basic aspects of traditional video navigation and presents two forms of advanced video navigation. The navigation methods are presented without implementation aspects. Implementation aspects will be discussed in consecutive sections addressing tape, optical disk, silicon-based storage solutions.

2.1 Traditional video navigation

For traditional video navigation a distinction is made between fast search and slow search mode, also known as trick play. Let P_s be the relative playback speed, which is unity for normal play, then fast-search trick play is obtained for $P_s > 1$ and slow-motion trick play for $P_s < 1$. Although this is a firm separation between fast search and slow motion, there is an overlapping area for playback speeds in the vicinity $L_b < P_s < U_b$ of normal play.

Fast-search video navigation is characterized in the sense that the pictures forming the trick-play sequence are derived from the normal-play sequence by applying a temporal equidistant sub-sampling factor, which corresponds to the intended playback speed P_s . In practice, P_s is limited but not restricted to integer values.

Slow-motion search is obtained by repetitive display of each normal-play picture. The amount of repetitions is equal to the reciprocal of P_s . Again, practical values for P_s are limited but not restricted to integer values. Although the basic operation to obtain slow motion has low costs, a distinction is made between slow motion on progressive video and interlaced video. When video originates from an interlaced video source, repetition of an interlaced picture may result in motion judder. Such a situation occurs when the spatial area contains an object that is subject to motion between the capture time of the odd and even field, forming a single frame picture. Repetition of such an interlaced picture causes the object repetitively traveling along the trajectory, which is perceived by the viewer as motion judder (van Gassel et al., 2002).

The above solutions for trick play are based solely on using resampled video information. For playback speeds in the vicinity of unity, there is an alternative implementation for trick play, that builds on also re-using the normal play audio information. For this situation, time-scaled normal-play audio information is used to create an audiovisual trick-play sequence. To maintain audibility of the time-scaled audio information, pitch control is required. The playback speeds P_s that can be used for this type of trick play depend heavily on aspects of the normal play sequence such as speed of the oral information and the algorithm used for processing the audio signal. Algorithms for time- and pitch-scaling can be divided into frequency-domain and time-domain methods. Good results are obtained using Pointer Interval Controlled OverLap and Add (PICOLA) (ISO/IEC 14496-3, 1999), an algorithm that operates in the time-domain on mono-channel audio signals for playback speeds in the range $0.8 \leq P_s \leq 1.5$ for audiovisual content with spoken text. This form of trick play indicates that P_s does not need to be limited to integer-based values.

2.2 Advanced video navigation

Traditional video navigation methods are adequate for navigation in video sequences with a duration of a few minutes. Navigation becomes already a daunting task when a two-hour movie is searched for a particular scene. Increasing the playback speed P_s up to a factor 50

or even higher, does not result in a better navigation performance. There is a twofold reason for this. First, at high playback speeds, pictures forming the trick-play video sequence are less correlated. As a result, it is difficult for the viewer to interpret each individual picture. Lowering the refresh rate by a factor three, which means that each picture is rendered three times and maintaining the playback speed P_s , results in an increment of the normal-play temporal sub-sampling factor, thereby effectively tripling P_s . For example, if P_s is equal to 50, this value will be scaled to 150. A typical scene duration lasts 3–4 seconds. With a P_s of 150, the trick-play sequence may not contain information of all scenes, which makes this approach less suitable for fast search. A more effective method to navigate through hours of video is *hierarchical* navigation, which is based on the usage of mosaic screens to obtain an instant overview of a certain time interval (Eerenberg & de With, 2003). When descending the hierarchy, the mosaic screens contain images that correspond to a smaller time interval, increasing the temporal resolution. Images used for the mosaic screen construction can be either based on a fixed temporal sub-sampling factor, or based on the outcome of a certain filter e.g. a scene-change detection filter. Mosaic screens based on fixed temporal sub-sampling, results in selection of normal-play pictures that in traditional video navigation are used to create a trick-play video sequence. When using these pictures for hierarchical navigation, the navigation efficiency (presence of many similar sub-pictures) of the lowest hierarchical layer depends on the applied temporal sub-sample factor. The usage of scene-change algorithms results in the selection of normal-play pictures that are effective from information point of view. This becomes apparent when only a single picture of a scene is used for construction of a mosaic screen, avoiding the navigation efficiency problem of equidistant temporal sub-sampling.

Hierarchical video navigation forms a good extension on traditional video navigation, but involves user interaction for browsing through the individual mosaic screens. Another advanced video navigation method which minimizes user interaction, is obtained when combining normal-play audiovisual information and fast-search video trick-play (Eerenberg et al., 2008). Time- and pitch scaling of audio information to create an audiovisual navigation sequence is only effective for trick-play speeds about unity, as indicated in Section 2.1. Although the human auditory system is a powerful sensory system, concatenation of normal-play audio samples corresponding to the time interval of a picture selected for fast-search trick play, is not an effective approach. The auditory system requires depending on the audio type, e.g. speech or music, hundreds of milliseconds of consecutive audio to become effective with respect to interpretation of the received audio signal. Based on this observation, an audiovisual trick-play method is proposed, in which normal-play audiovisual fragments are combined with fast-search video, resulting in a double window video with corresponding audio. The strength of this navigation method is that the viewer is not only provided with a course overview of the stored video information via the fast-search video trick-play sequence. Additionally, the viewer is also provided with detailed audiovisual information via the normal-play fragments. The duration of the normal-play fragments can be freely chosen obeying the minimal duration required by the auditory system to recognize well-known audio content or interpret new audio information. For the former situation, an audio fragment duration of about 1 second is suitable for recognition the origin of that particular fragment. For the latter situation, a longer duration is required leading to a practical value of 3 seconds.

3. Low-cost trick play

The basic principles of traditional trick play have been presented in Section 2. This section discusses low-cost trick play for fast search as well as slow motion in the context of the applied storage architecture. Section 3.1 will discuss two types of architectures. In digital consumer recording systems, trick play is a low-cost feature, resulting in the need for simple high-efficient signal processing algorithms fulfilling the cost requirement. Low-cost trick play is in general characterized by signal processing in the compressed domain, avoiding expensive transcoding. This involves selection and manipulation of coded pictures, which is covered in succeeding sections.

3.1 Impact of pull and push-based architectures on trick play

The chosen architecture of the storage system influences the involved trick-play signal processing. A digital storage system can have either a pull- or a push-based architecture. In a pull-based architecture, the video decoder pulls the data from the storage medium, whereas in a push-based architecture the video decoder receives the audiovisual information at a certain rate. For traditional trick play, there is a difference in performance due to the different architectures. Moreover, a push-based architecture allows two approaches. The first approach is based on a solution provided by the MPEG-2 system standard (ISO/IEC13818-2, 2000). This involves trick-play signalling information provided by the Packetized Elementary Stream (PES) header. This information is used to control the output of the video decoder during trick play. The second solution applies signal processing in the compressed domain, avoiding the usage of previous MPEG-2 system signalling information (van Gassel et al., 2002). Although the first solution is described by MPEG-2, its usage is optional and not obligatory (ETSI_TS_101154, 2009). The second solution is a generic approach, which delivers a compliant MPEG-2 video stream. This makes it an attractive solution from the video decoder point of view.

3.2 Influence of video encoding parameters on trick play

Modern compression schemes such as MPEG-2 or H.264, achieve high compression ratios by exploiting spatial and temporal correlation in the video signal. Compression of pictures exploiting only spatial information are intraframe compressed, whereas pictures that exploit temporal correlation are interframe compressed. The latter can be split in two categories: uni-directional (P-type) and bi-directional (B-type) predictive pictures. In a compressed video sequence, the distance between two successive intraframe compressed pictures is expressed by N , which is also known as the Group-Of-Pictures (GOP) length, whereas the distance between P-type predictive pictures is expressed by M . For the situation that $M > 1$, the number of B-type pictures preceding a P-picture is equal to $M - 1$. For the common situation that N is an integer multiple of M , the construction of trick play sequences is simplified because P-pictures are a sub-integer fraction of the GOP length N .

In general, trick play for digital consumer storage equipment is a low-cost feature, which limits the amount of involved signal processing. Low-cost trick-play algorithms deploy pictures that are selected from the compressed normal-play sequence. For fast-search trick play, the minimum fast-forward search speed is equal to M , whereas other fast-forward speeds are obtained for speed-up factors equal to N , or an integer multiple of N . Note that there is basically a gap between speed-up factor M and N if $N \gg M$. This gap is caused by the fact that

P-type pictures can only be decoded if the reference (anchor) picture has been decoded. For typical video compression applications such as Digital Video Broadcast (DVB), or digital recording, the GOP size $N=12$ and P-picture distance $M=3$, resulting in fast-search speed-up factors $P_s=\{3, 12, 24, \dots, 12n\}$, with $n=\{1, 2, 3, \dots\}$. Fast-search trick play in reverse direction is obtained in a similar way, but for a speed-up factor equal to M , buffering is required to store the decompressed pictures to facilitate reordering. This is required to match (potential) motion with the reverse playback direction, as the decoding is always performed in the positive time direction. This forward decoding direction introduces an extra delay, which occurs only once when switching to the reverse search mode with speed-up factor equal to M .

From the concept point of view, trick play in the compressed domain is equal for push or pull-based architecture. However, from a compliancy point of view, they are different. In a pull-based architecture, the decoder retrieves either fragments containing the intended intraframe compressed pictures, or the whole compressed video sequence at a higher rate. For either case, there will most probably be a frame-rate and bit-rate violation. From trick-play point of view, a high quality is perceived by the viewer both for low and high search speeds.

For push-based architectures, fast-search trick play based on concatenation of normal-play intraframe compressed pictures may cause a bit-rate violation. A low-cost method to overcome this bit-rate violation is the usage of repetition pictures, i.e. interframe compressed pictures, of which the decoded result is identical to the reference image (anchor picture), which is the last transmitted intraframe or P-type interframe picture, see Fig. 1. A repetition picture precedes the transmission of an intraframe coded picture. Multiple repetition pictures are involved when the transmission time exceeds two or more display periods, where one display period is equal to the reciprocal of the frame rate. For the situation that the normal-play video is MPEG-2 coded, a repetition picture has an extreme small size of 342 Bytes for a picture size 720×576 and 4:2:0 sampling format. The transmission time for such a compressed picture is negligible compared to the display period, leaving the remaining time for transmission of intraframe compressed data. This concept introduces fast-search trick play with a reduced refresh rate, so that the frame rate is not jeopardized. This concept yields a proper result for progressive as well as interlaced video, although motion judder may be introduced for interlaced video. Motion judder is avoided when the two fields that form the repetition pictures are field-based coded, where

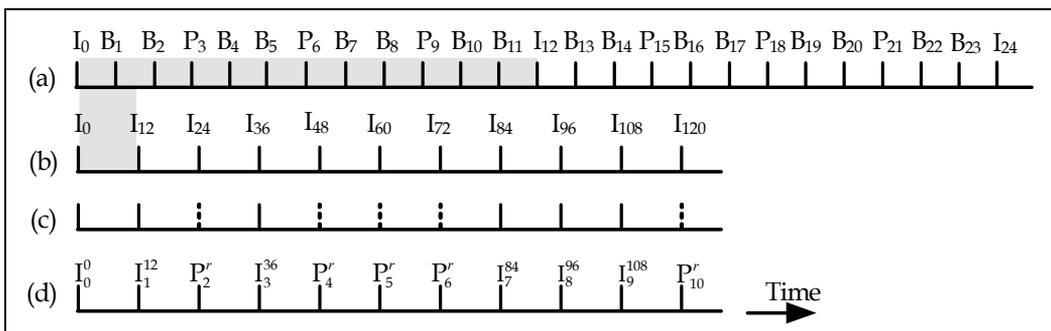


Fig. 1. Low-cost MPEG-compliant fast-search trick play. (a) Normal play compressed pictures with absolute picture index, (b) Intraframe compressed normal-play pictures, (c) Sub-sampling of the normal-play sequence, using a speed-up factor of 12, where the dashed lines indicate the skipped pictures, (d) MPEG-compliant trick play for speed-up factor 12 equipped with repetition pictures to avoid bit-rate violation.

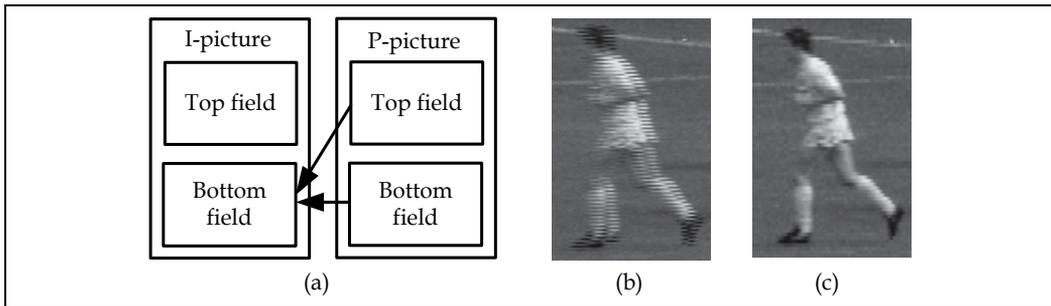


Fig. 2. Effect of “interlace kill”. (a) Predictive picture causing top and bottom field to have equal content. (b) Snapshot of I-picture region constructed of two fields from different time instances. (c) Snapshot of the same region after applying the “interlace kill” feature.

both fields refer to the same reference field, resulting in an “interlace-kill” operation see Fig. 2 (van Gassel et al., 2002). In this operation, one field is removed and replaced by the remaining field.

The reference field depends on the search direction and is bottom field for forward and top field for the reverse search direction. The result of the required decoding operation is that after decoding, both fields are equal, eliminating the possibility of motion portrayal. Using the MPEG-2 coding format, the bit cost for such a predictive-field coded picture is 954 Bytes, with a picture size of 720×288 pixels and 4:2:0 sampling format. The coding of two involved field pictures results in a total of 1,944 Bytes, including the required picture header and corresponding extension header.

3.3 Slow-motion for push-based storage systems

Slow-motion trick play is based on repetitive display of a picture. The number of display periods that a normal-play picture is displayed, is defined by Eqn. (1), where D_p indicates the number of display repetitions and P_s the slow-motion speed factor, giving

$$D_p = \frac{1}{P_s}. \quad (1)$$

Also here, there is a difference between push- and pull-based architectures. In a pull-based architecture, the video decoder outputs the decoded image data D_p times. In a push-based architecture, a slow-motion sequence is derived on the basis of the normal-play compressed video sequence. The slow motion operation requires that newly created display periods, are filled up with the repetition of normal-play pictures. For progressive video, all newly created frame periods are inserted after each original frame. Repetition of normal-play I-type and P-type pictures is achieved by uni-directional coded B-type repetition pictures. Any normal play B-type picture is repeated via repetitive transmission and decoding of the original B-picture. Special attention is required for the repetition of normal-play anchor pictures when they are originating from an interlaced video source. To avoid corruption of the decoder reference pictures, only uni-directional B-type coded “interlace-kill” pictures can be used for repetition of the normal-play anchor pictures, as they do not modify the anchor picture memory.

Anchor pictures maintaining their proper content is a basic requirement for the slow-motion decoding process, as each predictive-coded normal-play picture may use reference data

from two or more fields. The basic problem is that an “interlace-kill” operation can only be applied to the anchor pictures, which are intraframe or P-type predictive normal-play pictures. There is no low-cost algorithm for removing interlacing to process the B-type coded pictures, when they are frame-based coded. For the situation that B-type pictures are field based, only one field out of two B-type field pictures is used. As a result, frame-based B-type coded pictures can only be used once, if the video originates from an interlaced source.

Slow motion results in $N \cdot D_p$ pictures per normal-play GOP. For the situation that the video originates from an interlaced video source, a display error D_e occurs as indicated in Eqn. (2), the display error per normal-play GOP. From Eqn. (2), it becomes clear that for $M=1$, which conforms to MPEG-2 simple profile, no display error occurs, as indicated by

$$D_e = N(D_p (1 - \frac{1}{M}) - (1 - \frac{1}{M})). \quad (2)$$

In order to avoid a speed-error, normal-play anchor pictures which are displayed D_p times, must be displayed an additional A_r times, as specified in Eqn. (3), resulting in

$$A_r = D_p (M - 1) - (M - 1). \quad (3)$$

4. Digital tape-based helical-scan video recording

This section discusses Digital Video (DV) and Digital Video Home System (D-VHS), two tape-based Video Cassette Recording (VCR) standards developed in the 1990s. The DV standard differs from the D-VHS standard in the sense that video coding, on tape storage and trick play form an integral approach. The standardization of D-VHS followed a two-step approach. First, a basic recording engine was established capable of storing an MPEG-2 transport stream. In the second stage of the standardization process, trick play was added to this system. This section is divided in three sections. Section 4.1 discusses the impact of helical-scan recording on trick play. Section 4.2 elaborates on the DV trick-play solution. Finally, Section 4.3 presents trick play for the D-VHS STD mode.

4.1 Helical-scan video recording

Tape has been a popular storage medium for many decades due its good price/storage capacity ratio and the high capacity storage per volume unit. Tape-based video recorders apply helical-scan recording, where the magnetic heads are mounted on a rotary head wheel inside a cylindrical drum, and the tape is helically wrapped around it. Such a recording solution provides a large recording bandwidth, due to the fast rotation speed of the scanner and the relatively slow tape speed. This concept creates high-density recording with modest sized cassettes and sufficient playing time. The heads mounted on the rotary wheel have a different azimuth and write the information in slanted tracks on tape, resulting in a high track density, due to partial overlapped writing. Moreover, each track forms a fixed bit-rate channel, see Fig. 3 (a).

During playback, the heads with the proper azimuth scan the corresponding tracks on tape, a process that is controlled by the tracking servo. During normal play, see Fig 3 (b), all information that was recorded on tape is read from tape. A different situation arises for trick

play. During fast-search trick play, the tape travels at a higher speed along the scanner. As a result, the heads that scan the tape, under control of the tracking servo, follow a different path when compared to normal play, see Fig 4 (a). The tracking servo controls the scanner scan-path via adaptation of the tape speed and allows for two concepts. The first concept is based on speed-lock. Here the heads scan the fast travelling tape, where there is no guarantee which head scans the tape at a particular moment. The second concept is phase-lock. In this concept, the scanner is positioned such that at the beginning of the scan path, the head with a particular azimuth scans the track with the corresponding azimuth. The impact of the different servo approaches has its influence on the available trick-play bit rate. For the situation that the more complex phase-lock servo system is deployed, there is no need for duplicating the trick-play data, which results in a higher available trick-play bit rate because the head starts at the track with the corresponding azimuth, see Fig. 4 (b). For the situation that a speed-lock servo system is used, trick-play data needs to be stored multiple times, such that it reads once during fast search, see Fig. 4 (c).

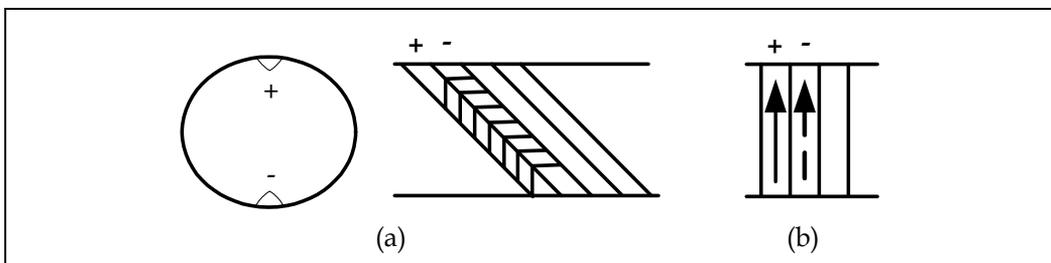


Fig. 3. Helical-scan recording. (a) Two-head scanner and corresponding azimuth tracks, (b) Simplified track diagram showing normal scan path, the non-dashed arrow indicates scan path of '+' head and dashed arrow indicates scan path of '-' head.

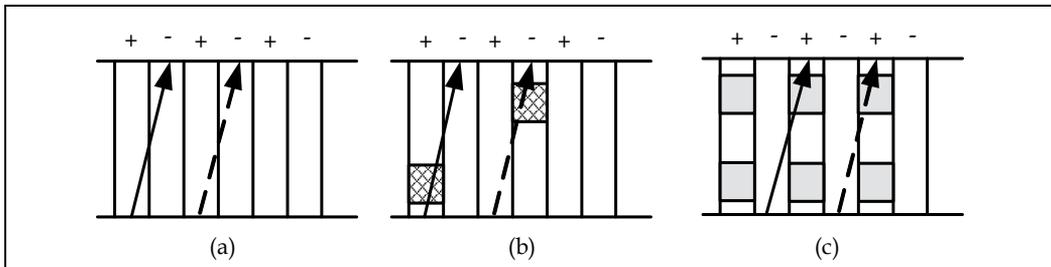


Fig. 4. Helical-scan fast search scan path. (a) Fast scan path for speed-up factor $P_s=2$, (b) Example of trick-play data location for phase-lock-based servo solution, (c) Example of trick-play data location for speed-lock-based servo solution.

The obtained trick-play quality is not only influenced by the applied servo system, but also by the applied video compression scheme. In the DV recording system, tape and compression related aspects have been combined, thereby facilitating trick play on the basis of re-using normal-play compressed video information. D-VHS is a recording system where storage part (bit engine) and compression have been separated, resulting in a trick-play solution that facilitates individual trick play (virtual) channels for various trick-play speed-up factors and their corresponding directions.

4.2 Trick play for digital video

The DV system (CEI/IEC61834, 1998) is a recording standard intended for 25-Hz (PAL) and 30-Hz (NTSC) television standards and was first announced in 1993 (Matsushita et al., 1993). The applied video compression was largely guided by system aspects such as editing on a picture basis, robustness for repetitive (de-)compression, number of tracks to store a compressed picture, very high forward and backward search on tape, overall robustness and high picture quality. In order to understand the DV trick-play mechanism, a brief system overview is presented, which introduces the essential aspects that enable trick play.

4.2.1 System architecture

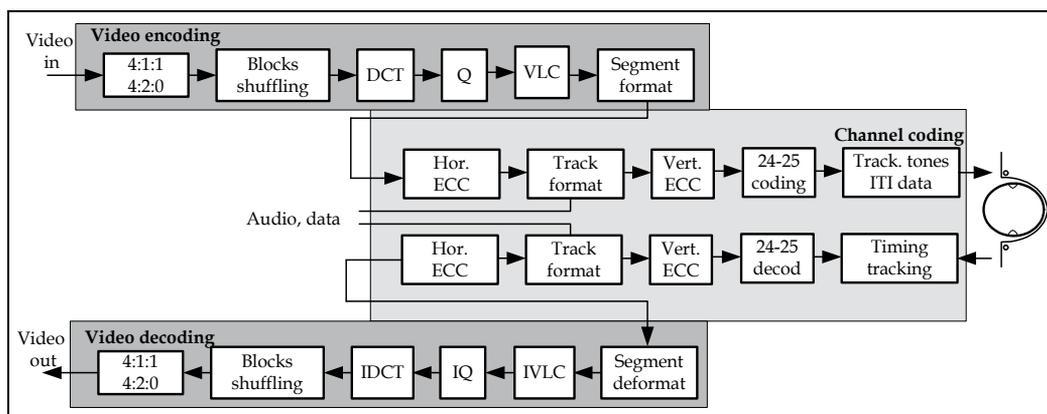


Fig. 5. Block diagram of complete processing of consumer DV recording system.

Figure 5 depicts the block diagram of the DV recording system. At the left-hand side, video enters the system and is compressed using a 4:2:0 sampling format for 25-Hz video or 4:1:1 sampling format in the case of 30-Hz video. Prior to compression, the video data is shuffled increasing the normal-play quality and enabling trick play. The compression is based on intraframe coding, using a Discrete Cosine Transform (DCT), and subsequent quantization and Variable-Length Coding (VLC) of the transformed video data. The compressed video or uncompressed audio data together with signalling information is packetized into Sync Blocks (SB), which are protected with parities from the Reed-Solomon-based horizontal forward error correction. To further improve the system robustness, a vertically-oriented error-correction layer is added, calculating parities over multiple SBs within a track, which are stored in dedicated parity SBs. Finally, the SBs are stored on tape using a highly efficient DC-free 24 -25 channel code with embedded tracking tones required for tracking purposes. For a 25-Hz frame rate, a picture is stored using 12 consecutive tracks or 10 tracks for a 30-Hz frame rate.

4.2.2 DV video compression and sync block mapping

DV video compression is based on macroblocks (MB) constructed from luminance and chrominance DCT blocks, see Fig. 6. Each macroblock contains a full-color area of 256 pixels, where the physical dimensions differ between 32×8 rectangular (30 Hz) and 16×16 square for (25 Hz).

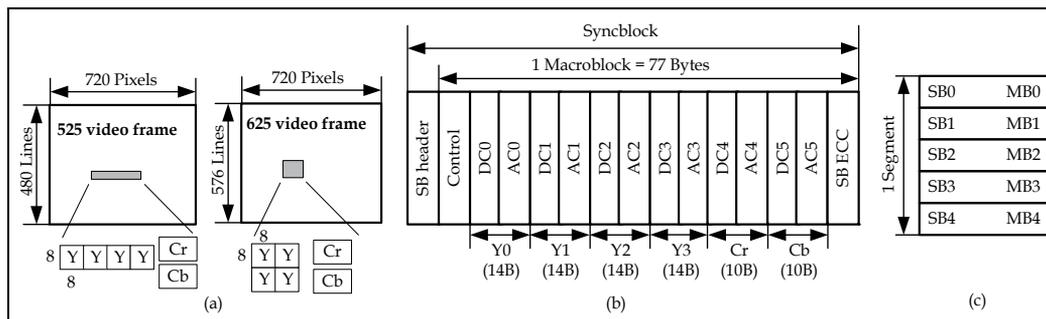


Fig. 6. DV Macroblocks. (a) Construction of macroblocks for 25-Hz and 30-Hz video, (b) SB data format showing fixed predetermined positions of low-frequency DCT coefficients, (c) Five pre-shuffled macroblocks forming a single segment.

The number of DCT blocks differ for a 25-Hz or 30-Hz system, but result in an equal number of 135 macroblocks per track (one MB per SB), which leads to a uniform recording concept that lowers the system costs. Segments are constructed from 5 pre-shuffled MBs, see Fig. 6 (c), to smoothen the coding performance and create a fixed bit cost per segment.

The compression of a segment is realized by applying a feedforward coding scheme, resulting, which high robustness and individual decodability of each segment. Although the segments are of fixed length, this does not mean that the macroblocks constructing the segment are also fixed-length coded. The optimal mapping of macroblocks on sync blocks is achieved when each macroblock is stored in a single sync block. This also enables the highest search speed (up to 100 times), due to individual decodability. By applying a fixed mapping, see Fig 6 (b) of the DCT low-frequency coefficients on a sync block, each sync block is individually decodable at the expense of a somewhat lower PSNR, due to the fact that some high-frequency coefficients may be carried by neighboring sync blocks belonging to the same segment.

The trick-play quality is determined by two factors: the trick play speed-up factor and the data shuffling. The system is designed such that the picture quality gradually deteriorates with increasing search speed. For very high search speeds (50-100 times), the reconstructed picture is based on individually decoded MBs, which are found in the individually retrieved SBs. In this mode, the vertical ECC cannot be deployed and therefore the decoder relies only on a very robust storage format where the low-frequency information of each DCT block is available on fixed positions inside the SB. For lower tape-speeds, larger consecutive portions of a track can be retrieved, so that full segments can be decoded in full quality and individual MB from partial segments. The shuffling is organized such that the quality is further increased in this playback mode. Figure 7 (b) depicts that consecutive segments contain MBs that are neighbors, which form a superblock resulting in a larger coherent spatial area from the same image. Hence, the lower the tape speed, the more neighboring MBs are successfully retrieved and the larger the coherent area that is constructed.

Figure 7 (a) depicts how the individual MBs of one segment are chosen. It can be seen that within one segment, MB data is sampling the full image both horizontally and vertically. This ensures a smoothening of the data statistics, enabling a fixed bit cost per segment. Moreover, the shaded area in the upper-left corner indicates a superblock with that is constructed from the neighboring MBs, deploying a pattern as depicted in Fig. 7 (b).

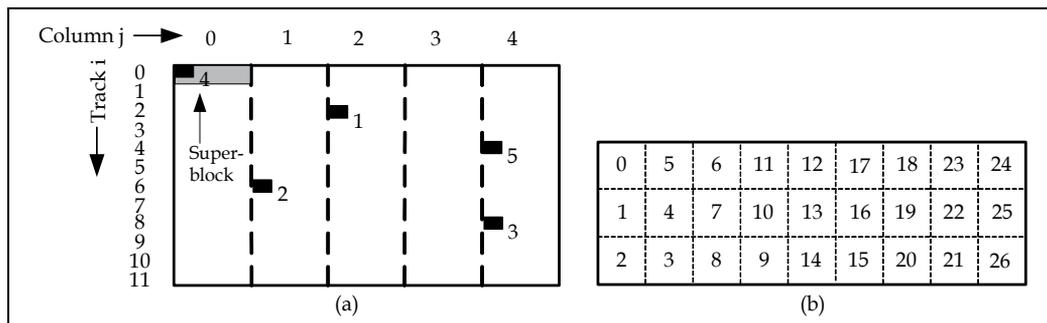


Fig. 7. DV tracks. (a) Selection of MBs for segment construction and assignment of picture areas on tracks, (b) Ordering of macroblocks in repetitive clusters, called superblocks.

4.3 Trick play for D-VHS STD mode format

D-VHS (ISO60774-5, 2004) is a tape-based recording system, following the push-based architecture and developed in two stages. The first stage covers the record and playback of MPEG-2 compressed normal-play information (Fujita et al., 1996), whereas the second stage addresses trick-play aspects. The D-VHS standard describes three recording modes, each intended for a specific data rate, resulting in a format for STanDard (STD) mode, Low-Speed (LS) mode and High-Speed (HS) mode. The first two recording modes share the same trick-play speed-up factors ± 4 , ± 6 , ± 12 and ± 24 , whereas the HS recording mode supports the speed-up factors ± 3 , ± 6 and ± 12 . This section describes high-end and low-cost trick-play signal processing for the STD mode format supporting speed-up factors ± 4 , ± 12 and ± 24 .

4.3.1 Trick play based on information carried by a virtual channel

Trick play for the D-VHS STD mode assumes that the recording system deploys a phase-lock tracking servo (track select). The standard describes areas within the tracks, which are allocated for a particular trick-play speed-up factor for forward as well as reverse playback direction. These trick-play regions form a pattern, which is repetitive modulo 48 tracks, see Fig. 8. For a particular playback direction, i.e. forward or reverse, the trick-play regions corresponding to a particular trick-play speed-up factor, form a virtual information channel. This channel is only present during fast search at a particular speed-up factor and corresponding direction. During record, a track is filled with 336 sync blocks, see Fig. 9 (b), which are data units of 112 Bytes. Each Sync Block (SB) consists of synchronization information, main data area and inner parity based on a Reed-Solomon forward error correction, see Fig. 9 (a). Two main data areas store a time-stamped TS packet, which is constructed of a 4-Byte header containing the time stamp and a single TS packet of 188 Bytes. The physical storage position depends on the SB number and corresponding time stamp value.

Each SB, regardless whether it contains normal-play or trick-play data, is protected by means of inner parity. The corresponding inner parity is stored at the last Byte positions of a SB. To increase the playback robustness, a second Reed-Solomon forward error correction is applied. The corresponding parity data (outer parity) are stored in the last SBs of a track. The usage of outer parity information for trick play is optional. In a push-based architecture, the isochrone nature of the stored audiovisual information is reconstructed at playback with the aid of time stamps. Time stamping is a method that captures the position of an MPEG-2

Transport Stream (TS) packet on the time axis. At playback, the time stamp is used to position the TS packet at the proper time location. This reconstructs the isochrone behavior of the TS sequence, which enables the usage of an external MPEG-2 decoder connected via a digital interface (IEC61883-4, 1994). Time stamping in D-VHS, is based on a more accurate 27-MHz clock ($\pm 20\text{ppm}$ versus $\pm 40\text{ppm}$ in broadcast), which is locked to the incoming TS stream. D-VHS stores this time stamp together with the TS packet in either two consecutive SBs or in two separated SBs, with trick play SBs in between. The trick-play virtual information channel contain an MPEG-2 compliant TS stream. The isochronous playback of such a trick-play sequence requires the presence of time stamps. Moreover, these time stamps are also used for adequate placement of the TS packets in the virtual channel.

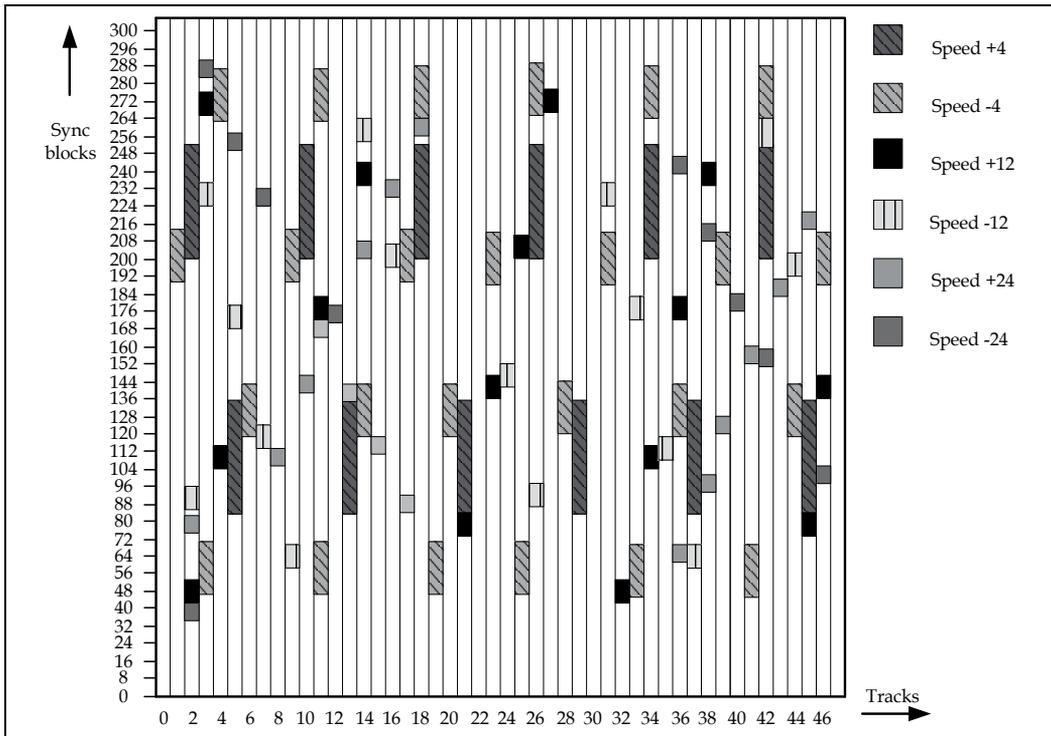


Fig. 8. D-VHS STD-mode trick play regions for speed-up factors: ± 4 , ± 12 and ± 24 .

Due to the tight relation between the time stamp and the physical storage location, the drum rotation phase needs to be synchronized during record and playback. Synchronization is essential because the local time stamp generator is reset modulo three revolutions of the scanner. Trick play in D-VHS is based on retrieving scattered fragments from various tracks. For the situation that the drum rotates at 1,800 rpm (30 Hz), the sync blocks allocated for each trick-play speed-up factor, result in a capacity of 2,301,120 bits/sec, see Table 1, regardless the playback direction. Trick-play data can be divided into two categories: high-end and low-cost. The former is characterized by high video quality and a refresh rate equal to the frame rate, regardless of the involved implementation cost. The latter is determined by minimizing the involved system costs. In any case, as trick play is in general a low-cost feature, the amount of signal processing should be bounded.

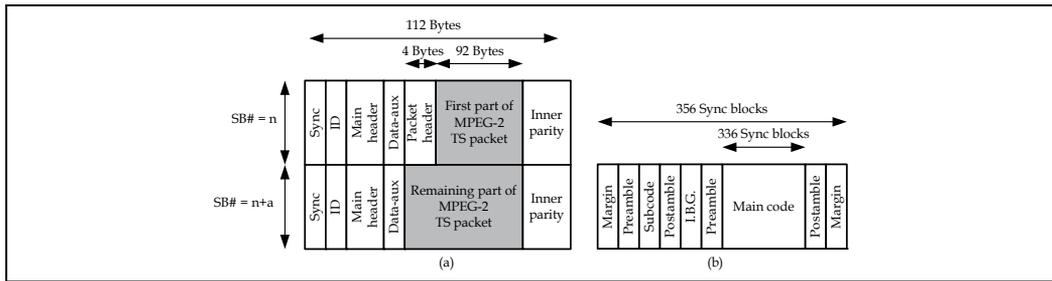


Fig. 9. D-VHS sync block. (a) MPEG-2 TS packet mapped at two SBs, (b) 336 SBs containing normal play, trick play, or padding data constructing one track.

| Scanner revolution (Hz) | Trick play channel bit rate (bits/s) |
|-------------------------|--------------------------------------|
| 30 | 2,301,120 |
| 29.97 | 2,298,821.17 |

Table 1. Channel bit rate for 30-Hz and 29.97-Hz scanner frequency.

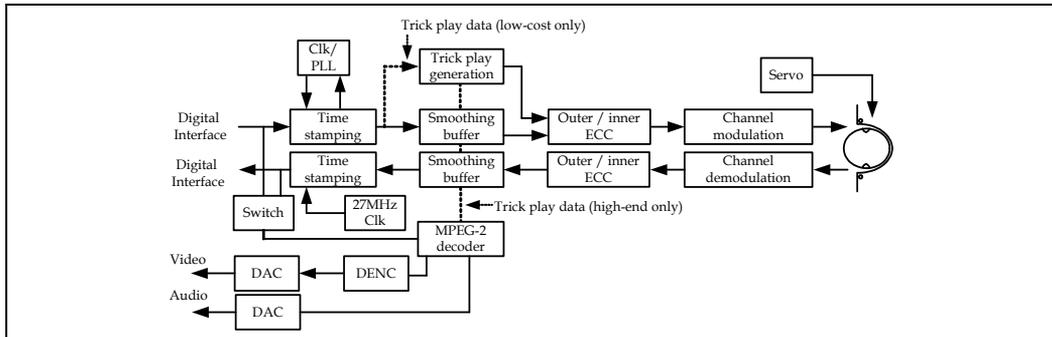


Fig. 10. D-VHS block diagram. The dashed connections indicate the information flow for the two trick-play flavors.

A basic functional block diagram that enables both high-end and low-cost trick-play implementation, is depicted in Fig. 10. The difference between the two concepts is only in the selection of normal-play pictures for the lowest forward trick-play speed, while the remaining processing steps are equal.

High-end trick play is implemented in the following way. During record of the normal play sequence, the normal play sequence is decoded and the decoded pictures are made available to the trick-play functional block, indicated by the dashed line, see Fig. 10. During trick play, the stored trick-play sequence is de-smoothed and reconstructed based on the time stamps. Low-cost trick play requires de-multiplexing of the normal-play sequence and video elementary stream parsing to locate the intraframe-compressed normal-play pictures. Due to the low trick-play bandwidth, see Table 1, high-end trick play is based on Common Intermediat Format (CIF) resolution. Almost 90,000 bits are available on a per picture basis for video with 25-Hz frame rate. Images of CIF-based resolution can be coded with sufficiently high quality using fixed bit-cost compression, resulting in a Peak Signal-to-Noise Ratio (PSNR) between 30 dB and 40 dB, depending on the spatial complexity, see Fig. 11 (b).

For low-cost trick play, the available bandwidth during trick play is not sufficient to enable a full refresh rate and preserve the picture quality when re-using normal-play intraframe-coded pictures. Hence, the concept for low-cost trick play is based on using transcoded normal-play intraframe-compressed pictures. Transcoding resulting in lower bandwidth can be obtained in two ways: removing picture energy (AC coefficients, see Table 3) or reduced refresh rate. Table 2 indicates the normal-play intraframe-compressed bit cost for various encoder settings. The minimum intraframe bit cost that remains when removing all AC energy is depicted in Table 3.

| Normal play bit rate (Mbit/s) | Number of pixels per line | GOP parameters | | Average intraframe bit cost (bits) | Minimum intraframe bit cost (bits) | Maximum intraframe bit cost (bits) |
|-------------------------------|---------------------------|----------------|---|------------------------------------|------------------------------------|------------------------------------|
| | | N | M | | | |
| 9.4 | 720 | 12 | 3 | 770,084 | 430,896 | 1,099,696 |
| 3.4 | 480 | 12 | 3 | 281,126 | 45,984 | 564,568 |
| 5.0 | 528 | 12 | 1 | 417,819 | 68,344 | 640,244 |
| 8.0 | 720 | 12 | 3 | 578,032 | 451,616 | 909,848 |

Table 2. Normal-play MPEG-2 intraframe bit cost for various bit rates, frequently used GOP structures, 576 lines per image, 25-Hz frame rate and 4:2:0 sampling format.

| Normal play bit rate (Mbit/s) | GOP parameters | | Average intraframe bit cost (bits) | Minimum intraframe bit cost (bits) | Maximum intraframe bit cost (bits) |
|-------------------------------|----------------|---|------------------------------------|------------------------------------|------------------------------------|
| | N | M | | | |
| 9.4 | 12 | 3 | 108,326 | 83,600 | 122,488 |
| 3.4 | 12 | 3 | 77,329 | 53,944 | 97,200 |
| 5.0 | 12 | 1 | 55,012 | 48,032 | 60,176 |
| 8.0 | 12 | 3 | 78,915 | 75,336 | 81,840 |

Table 3. Transcoded DC-only bit costs for various bit rates, frequently used GOP structures, 576 lines per image, 25-Hz frame rate and 4:2:0 sampling format.

The normal-play intraframe-compressed pictures, as depicted in Table 2, result in refresh rates varying between 3 Hz and 8 Hz, which is insufficient from trick play perception (fast changing images) point of view. The DC-only intraframe-compressed pictures as depicted in Table 3, result in refresh rates up to 25 Hz, which is sufficient for trick-play perception, but the pictures lack detail due to the strong energy removal. An acceptable solution is obtained when the refresh rate is reduced to 8.3 Hz. In order to create an MPEG-2 compliant trick-play sequence, repetition pictures with or without "interlace kill" are applied, see also Section 3, resulting in a fixed bit cost of roughly 270,000 bits per trick-play GOP, see Fig. 11 (a). Various approaches have been reported to reduce AC energy from the intraframe-compressed normal-play pictures. The approaches vary between selecting the number of AC coefficients based on the amplitude of the AC coefficients (Ting & Hang, 1995) by means of prioritization (Boyce & Lane 1993), or on the basis of the differential DC value between successive DCT blocks (US6621979, 1998). Based on the concepts for low-cost trick play, the final trick-play quality depends on the normal-play quality and associated factors, due to the re-use of pictures. This results for luminance-only information, in a 25-dB PSNR for DC-only trick-play pictures, up to the original normal-play PSNR, when the spatial content of the pictures are of low complexity.

The trick-play information channel is filled with a multiplex of six individual transport streams, which are all derived from a single information stream, in this case the lowest fast-forward sequence. During trick-play, the higher and all supported reverse speeds are derived from fragments belonging to that particular trick-play sequence. Once the fixed bit-cost trick-play GOP for speed-up factor $P_s=+4$, is available (regardless whether high-end or low-cost), the trick-play video sequences for the other forward trick-play speed-up factors ($P_s=+12$ and $P_s=+24$) are derived from this trick-play sequence by the “search speed data selection” block, see Fig. 11 (c), via sub-sampling in the compressed domain of the fixed bit-cost GOPs.

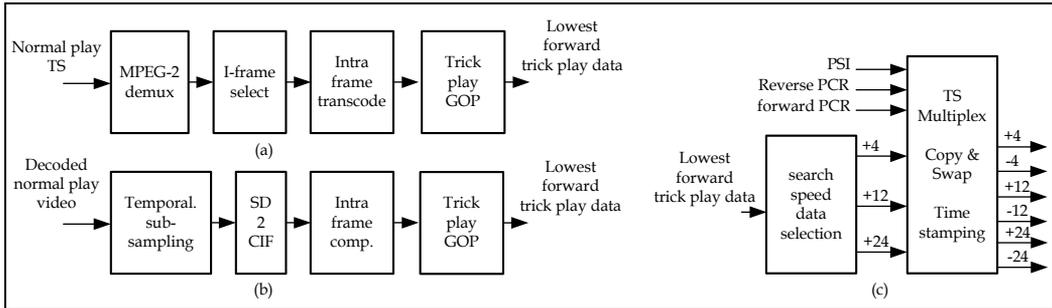


Fig. 11. D-VHS trick-play processing. (a) Low-cost trick-play video elementary stream generation, (b) High-end trick-play video elementary stream generation, (c) MPEG-2 TS multiplexing, time stamping and reverse trick-play generation.

The six trick-play streams are constructed as follows, assuming that the lowest search speed GOP is available, see Fig. 11 (a) and (b) . The functional block “TS multiplex Copy & Swap Time stamping”, see Fig. 11 (c), performs three final operations on the forward trick-play elementary streams. The first operation is multiplexing the elementary streams into MPEG-2 Transport Streams (TS). The second operation is to copy and swap the TSs creating the trick-play sequences for the reverse search direction. Due to the fact that MPEG-2 decoding uses an increasing time axis, the reverse trick-play sequences are generated with a declining time axis. The swap operation on TS data is required to read the GOP in the correct order during trick play. Finally, the TS packets are equipped with a time stamp, such that the TS packets are properly mapped on the corresponding trick-play regions.

5. Trick play for optical storage systems

Optical storage systems differ from the traditional tape-based storage system in the sense that they offer random access to the stored audiovisual information, enabling trick-play generation at playback using the stored normal-play audiovisual information. Although no separate trick-play sequences are stored on the storage medium, additional information is stored facilitating trick play. In this section, trick play is described for two optical standards: SuperVCD and DVD.

5.1 Super Video Compact Disc (SuperVCD or SVCD)

SuperVCD is the successor of Video Compact Disc (VCD) and is also known as SVCD, offering up to 800 MByte storage capacity, using a 780 nm laser. This standard appeared on the market in the early 1990's (IEC62107, 2000). SuperVCD utilizes better audio and video

quality compared to VCD, which used MPEG-1 video compression at Common Intermediat Format (CIF) resolution with a fixed bit rate of 1.1458 Mbit/s and MPEG-1 layer 2 audio at 256 kBit/s. The improved video quality of SuperVCD was later obtained by going to higher spatial resolution 480×576 (H×V) at 25 Hz for PAL TV system and 480×480 (H×V) at 29.97 Hz for the NTSC TV system. To code the interlaced video more efficiently, MPEG-2 video coding is applied in combination with variable bit rate. Multi-channel audio has been added to the system to increase the audio quality. Figure 12 depicts the SuperVCD reference model. From Fig. 12, it can be seen that the amount of data delivered by the CD module differs for MPEG sectors and data sectors, which is caused due to unequal error protection resulting in a higher (14%) storage capacity for an MPEG sector and less protection of the MPEG-compressed data. The layout of an SVCD consists of a lead-in, program area and a lead-out area, as depicted in Fig. 13 (a).

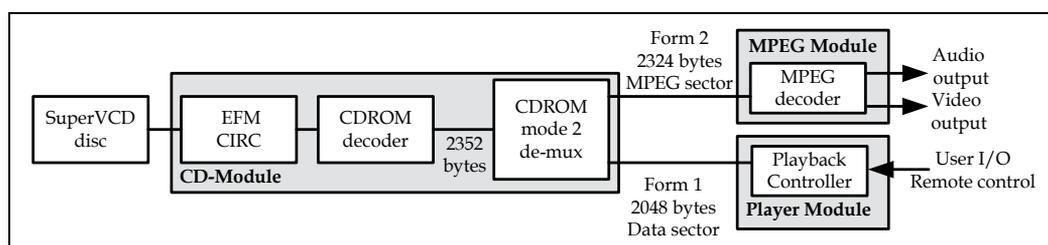


Fig. 12. SuperVCD system reference Model.

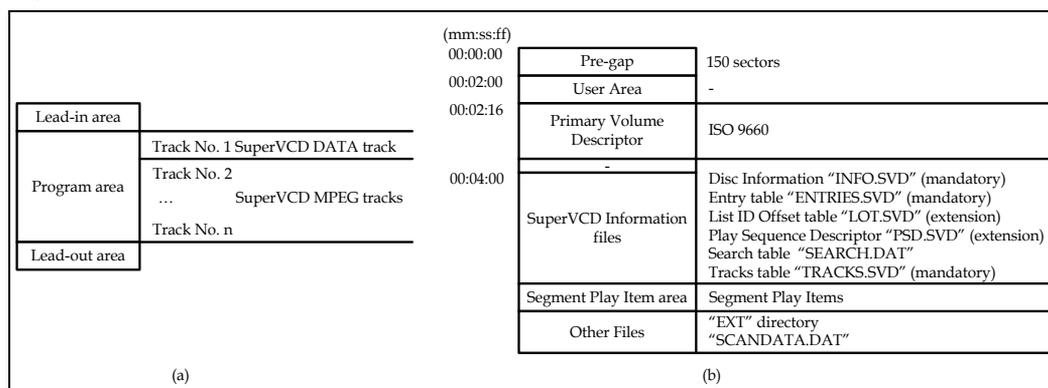


Fig. 13. SuperVCD disc and data track. (a) SuperVCD disc layout, (b) Data track layout example.

The program area is split into a SuperVCD data track and SuperVCD MPEG tracks. The former track stores additional information required by the SuperVCD player. The latter contains the MPEG-2 Program Stream containing the audiovisual information, which can be sub-divided into parts or chapters and are stored on the SuperVCD directory "MPEG2". SuperVCD is based on a pull-based architecture. Traditional trick play for such a system, as discussed in Section 3, can be on the basis of using only normal-play intraframe pictures or intraframe pictures in combination with uni-directional coded pictures. In SuperVCD, only intraframe-compressed pictures are used to realize traditional trick play. Traditional trick play on a SuperVCD stored program is established by means of locator information indicating the start position of an intra-coded picture. Furthermore, due to the random

access of the stored information, a new navigation feature has been added to the system enabling to jump forward or backward in a stored program. To support this new form of navigation, another list with locator information is available.

Traditional trick play on a SuperVCD stored program is established by means of the "SEARCH.DAT" information or via the scan information data, which is multiplexed as user data together with the video elementary stream. The presence of the "SEARCH.DAT" file depends on whether the system profile tag is set to 0x00. Due to the Variable Bit Rate (VBR) coding and buffering applied by MPEG-2, there is no longer a direct relation between playing time and the position on disc. To solve this relation problem for trick play, the "SEARCH.DAT" file contains a list with access point sector addresses, indicating the nearest intraframe compressed picture in the MPEG track on a regular incremental time grid of 0.5 seconds. The total number of entries is minimally 1 and maximally 32,767 using a storage format of (*mm:ss:ff*), where *mm*={0..99} indicates the minutes part of the sector offset value, *ss*={0..59} the seconds part of the sector offset value and *ff*={0..74} represents the fraction part of the sector offset value. This information is useful for features such as time search. It should be noted that, as indicated in Fig. 13 (b), the MPEG-2 coded information and associate data are embedded into the regular CD-Audio sectors, which form the fundamental storage framework on a CD disc. Each sector corresponds to a certain time instance indicates in minutes, seconds and fragments.

If the "SEARCH.DAT" is not present, a second information mechanism, called scan information data, is provided by the SuperVCD standard to enable trick play. Scan information data is information indicating the location of the next and previous intraframe coded picture. This information, which is transmitted using the MPEG-2 user data structure, precedes each intra-coded picture and consists of two field pairs, one pair for a video stream and one pair for a still image stream. The fields are coded as a sector offset value referencing to the start sector of the MPEG track, as indicated by the table of contents. The scan information fields share the same storage format, which is *mm:ss:ff*. The scan information data refers to an access point sector.

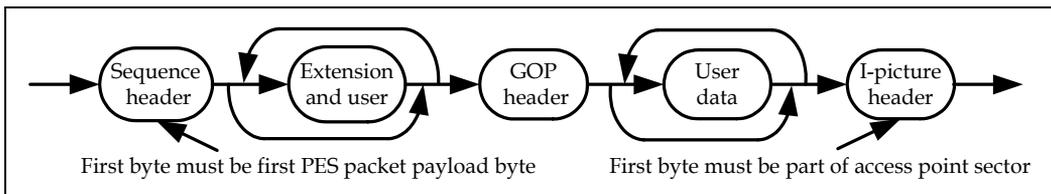


Fig. 14. SuperVCD access-point sector structure.

An access point sector is an MPEG video sector, in which the first byte of a Packetized Elementary Stream (PES) packet belongs to a sequence header, see Fig. 14. MPEG-2 allows the usage of different quantization tables. MPEG-2 video encoders that use this feature will transmit the different quantization matrices in the sequence header. It is for this reason that in order to have proper trick play, these quantization matrices have to arrive at the MPEG-2 decoder, which is guaranteed by the access-point structure as depicted in Fig. 14.

Time-based jumping is another navigation method supported by SuperVCD and involves jumping forward or backward in time to the next chapter. This feature is enabled by means of the "ENTRIES.SVD" file, which stores the entry point addresses of the MPEG-2 tracks on

disc. Due to the limited size of 2,048 Bytes, which corresponds to one sector, the number of entries of the "ENTRIES.SVD" file is limited to a maximum of 500.

5.2 Digital Video Disc (DVD)

The DVD is the optical media successor of the SuperVCD featuring higher recording density of up to 4.7 GByte for the basic format, using a 650 nm laser and deploying a pull-based architecture. The DVD-Video format (Taylor, 2001) specification is a video playback application of the DVD-ROM standard, applying MPEG-2 video compression and allowing MPEG-1 and run-length coding for still images. Various audio formats are supported such as Pulse Code Modulation (PCM), MPEG-1 or MPEG-2 compressed audio, or Dolby Digital multi-channel, which are packetized into a Packetized Elementary Stream (PES) and multiplexed into an MPEG-2 Program Stream. Also included in the MPEG-2 Program Stream multiplex are the real-time private stream, called Presentation Control Information (PCI), and Data Search Information (DSI), resulting in a maximum multiplexed bit rate of 10.08 Mbit/s. A player based on the DVD format, is typically connected to a standard TV display, either via baseband or modulated composite coax, or via analog component YCrCb/RGB video. Moreover, the signal can be of either interlaced or progressive nature. Figure 15 indicates the block diagram of a DVD player. A DVD player is based on a presentation engine and a navigation manager, as shown in Fig. 16. The presentation engine uses the information in the presentation data stream, see Fig. 15, to control what is presented to the viewer. For example, the navigation manager creates menus, provides user interfaces and controls branching, all on the bases of retrieved information from the DVD disc. This allows content providers to disable search-mode functions on particular DVD content, e.g. a trailer indicating that the audiovisual content is only to be used in the domestic environment. Per video title set, a particular movie can be made available as a single MPEG-2 Program Stream, or can be split up in maximally 9 parts, as indicated in Fig. 17 (a). This figure shows an example of a DVD volume layout containing a DVD-Video zone (DVD-Audio zone have been left out for simplicity). A Video Object Block (VOB) file, see Fig. 17 (b), is constructed of one or more cells containing a group of pictures or audio blocks; the cell forms the smallest addressable entity for random access. A cell, also known as scene, can be as short as 1 second, or cover the whole sequence and is uniquely identified with its cell ID and corresponding video object ID. A cell is further sub-divided into video object units, an entity containing zero or more Group-Of-Pictures (GOP). For the situation that a video object unit contains a single GOP, this GOP should start with a sequence header followed by a GOP header, which by default is followed by an intra-coded picture. Video object units are further split into packs and packets, which are compliant with the MPEG-2 Program Stream standard (ISO/IEC13818-1, 2000). Navigation data is a collection of information that determines how the physical data is accessed and controls interactive playback. The information is grouped into four categories: control, search, user interface and navigation commands, resulting in navigation information being split over five levels. The levels "program chain" and "data search" contain information required to perform navigation as discussed previously. For example, the "program chain" information enables the remote control navigation buttons "previous" and "next". The "data search" information is situated in the "navigation pack", see Fig. 17 (b) and thereby an integral part of the Program Stream.

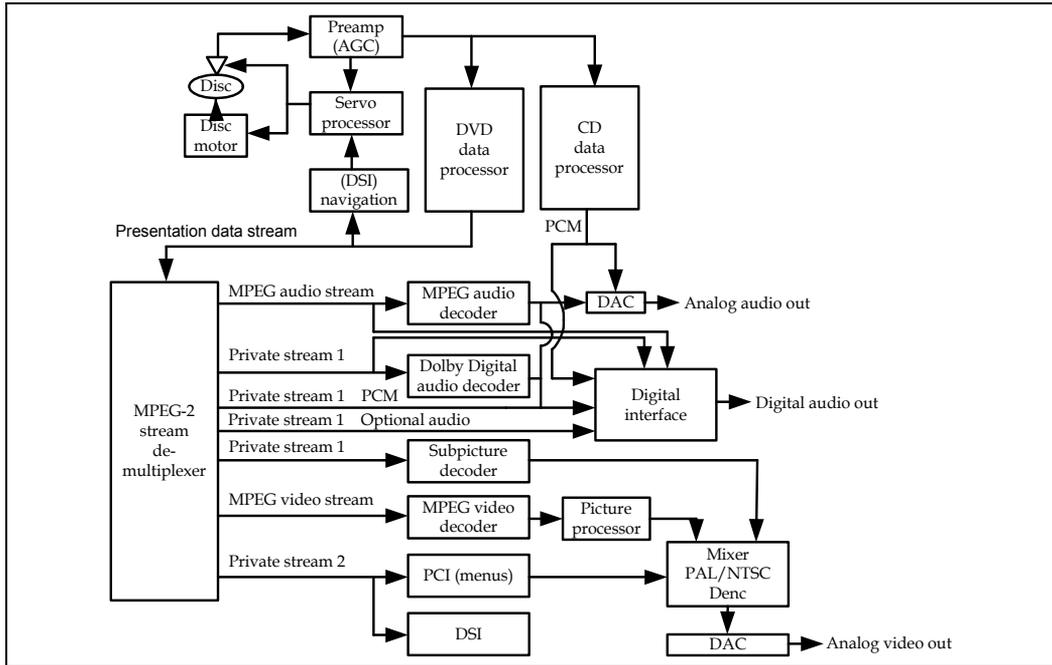


Fig. 15. DVD-Video player block diagram.

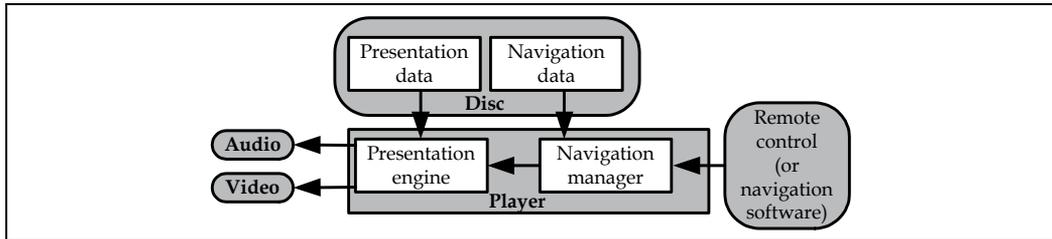


Fig. 16. DVD navigation and presentation model.

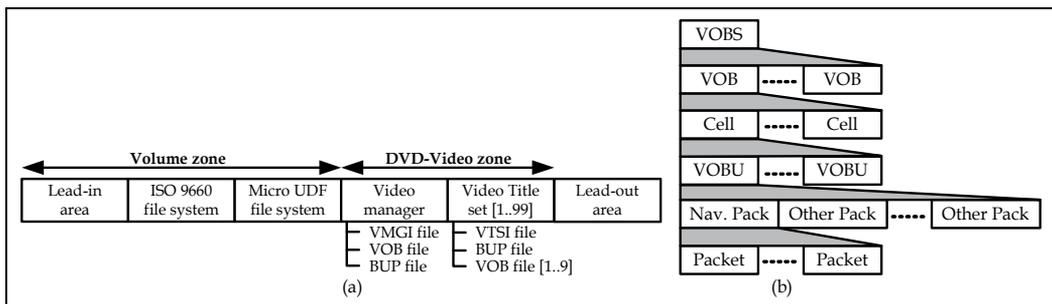


Fig. 17. DVD-Video data construction. (a) DVD volume layout containing a DVD-Video zone, (b) DVD abstraction layers on top of an MPEG-2 Program Stream.

The “navigation pack” holds for each video object unit a pointer for forward / reverse scanning, either referring to I-type or P-type pictures, which is an improvement with respect to SuperVCD that is only supporting I-type picture search.

5.3 Blu-ray Disc (BD)

Blu-ray Disc (BD, 2006) is the optical media successor of DVD, featuring higher recording density of up to 25 GByte for the basic format enabled by a 405 nm laser. BD is an optical storage system suitable for recording as well as playback-only of audiovisual information applying a pull-based architecture, see Fig. 18 (b). Video can be compressed using either: MPEG-2, MPEG-4 AVC, also known as H.264, or SMPTE VC-1 video compression. Multi channel audio is supported up to a 7.1 audio system and encoding is based on either, using Dolby Digital AC-3, DTS, or uncompressed using PCM. Besides the previous audio coding standards, three other standards are optionally supported: Dolby Digital Plus and two lossless coding methods, called Dolby TrueHD and DTS HD. The BD-ROM standard defines two platforms: a High Definition Movie (HDMV) and a Java™ platform, also known as BD-J, and these platforms categorize the BD-ROM features. In the BD-ROM system, exactly one mode, either HDMV mode or BD-J mode, is active at any given point of time during playback. HDMV supports features such as seamless multi-angle and multi-story, Language Credits, Directors Cuts, etc., see Fig. 18 (b).

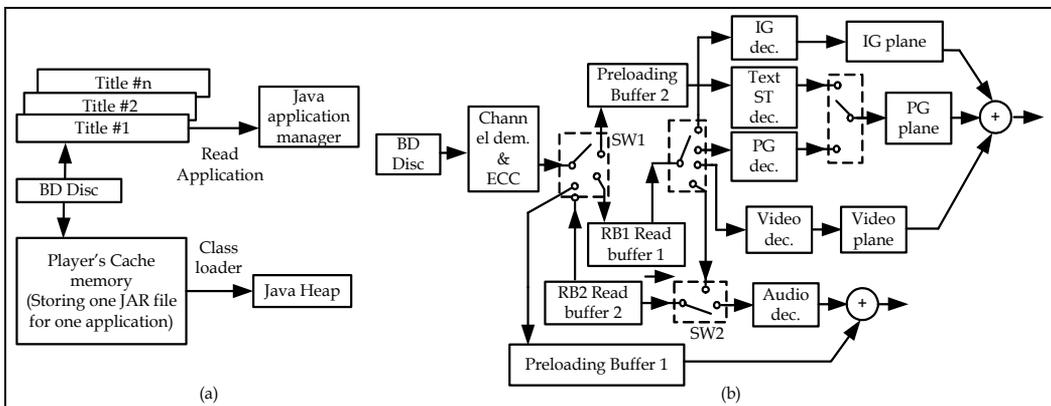


Fig. 18. BD-ROM system standard. (a) Java application tables in BD-ROM, (b) HDMV decoder model.

BD-J enables a fully programmable application environment with network connectivity, thereby enabling the content provider to create highly interactive, updateable BD-ROM titles, see Fig. 18 (a), and is based on the Java-2 Micro-Edition (J2ME) Personal Basis Profile (PBP), a Java profile that was developed for consumer electronics devices. It is assumed that the BD-ROM disc will be the primary source for media files, but it alternatives are the studio's web server and local storage. The unit of playback in BD-J is the Playlist, just as in HDMV. All features of HDMV, except Interactive Graphics (IG), which is replaced by BD-J graphics, can be used by a BD-J application. Supported features include Video, Audio, Presentation Graphics (PG), Text Subtitle component selection, media-time and playback-rate (trick-mode) control. The BD-J video device is a combination of the HDMV video and presentation graphics planes. Both video and presentation graphics will playback in the video device. BD supports traditional trick play via additional locator information, situated in the Clip Information file. Moreover, BD optionally supports Title Scene Search (TSS), a metadata-based advanced navigation method.

An AV stream file, together with its associated database attributes, is considered to be one object. The AV stream file is called a "Clip AV stream file", and the associated database

attribute file is called a "Clip Information file". An object consisting of a "Clip AV stream file" and its corresponding "Clip information file" is called a Clip. A "Clip AV stream file" stores data, which is an MPEG-2 Transport Stream defined in a structure called the BDAV MPEG-2 Transport Stream. In general, a file is a sequence of data bytes. But the contents of the "Clip AV stream file" are developed on a time axis, therefore the access points into a "Clip AV stream file" are specified with time stamps. The "Clip Information file" stores the time stamps of the access point into the corresponding AV stream file. The Presentation Engine reads the clip information, to determine the position where it should begin to read the data from the AV stream file. There is a one-to-one relationship between a "Clip AV stream file" and a "Clip Information file", i.e., for every "Clip AV stream file", there is one corresponding "Clip Information file".

The Playback Control Engine in the BD-ROM Player Model uses PlayList structures, see Fig. 19 (b). A "Movie PlayList" is a collection of playing intervals in the Clips, see Fig. 20 (a). A single playing interval is called a PlayItem and consists of an IN-point and an OUT-point, each of which refers to positions on a time axis of the Clip, see Fig. 20 (a). Hence, a PlayList is a collection of PlayItems. Here, the IN-point denotes a start point of a playing interval, and the OUT-point indicates an end point of the playing interval.

In Section 3, it was discussed that traditional trick play on compressed video data depends on the GOP structure, see also Fig. 1 (a). To enable random access, BD stores an EP_map (Entry Point), which is part of the Clip Information file, for each entry point of an AV stream. Unlike the previous optical storage standards, not only information regarding the position of intraframe-compressed pictures is stored, but optionally, also information regarding the length of each GOP and the coding type of the pictures constructing a particular GOP. This information is stored in the so-called *GOP_structure_map* located in the Supplemental Enhancement Information (SEI), which is stored in the user data container of the firstly decoded Access Unit (AU) of a GOP. The meaning of trick play is defined by the manufacturer of the BD-ROM Player, e.g. the manufacturer may define that $P_s=1.2$ forward play is not trick play, and instead define that $P_s>2$ forward play is considered trick play.

EP_map is a part of the "Clip Information file", and this information is mainly used for finding addressing information of data positions, where the BD-ROM player should start to read the data in the AV stream file. The corresponding access point is given to the Clip, see Fig. 19 (b), in the form of time indications, referring ultimately to a particular data byte in the AV stream. EP_map has a list of Entry Point data (EP-data) that is extracted from the AV stream, where decoding can start. Each EP-data is a pair of a PTS (Presentation Time Stamp) for an access unit and a data address of the access unit in the AV stream file. EP_map is used for two main purposes. First, it enables to find the data address of the access unit in the AV stream file that is pointed to by the PTS in PlayList. Second, it is used for facilitating fast forward and reverse trick play. The EP_map gives the relationships between Presentation Time Stamp (PTS) values and addresses in the AV stream file. The PTS entry point is called PTS_EP_start, where the actual AV stream entry point address is called SPN_EP_start, see Fig. 20 (b). In order to reduce the size of the table and to improve the searching performance, the EP_map for one stream is split in two sub-tables: EP_coarse and EP_fine. The "Clip Information file", which has a maximum file size of 2 MBytes is stored in "Clipinf", a sub-directory of "BDMV". Note that it is allowed that the CLIPINF directory contains no Clip Information file.

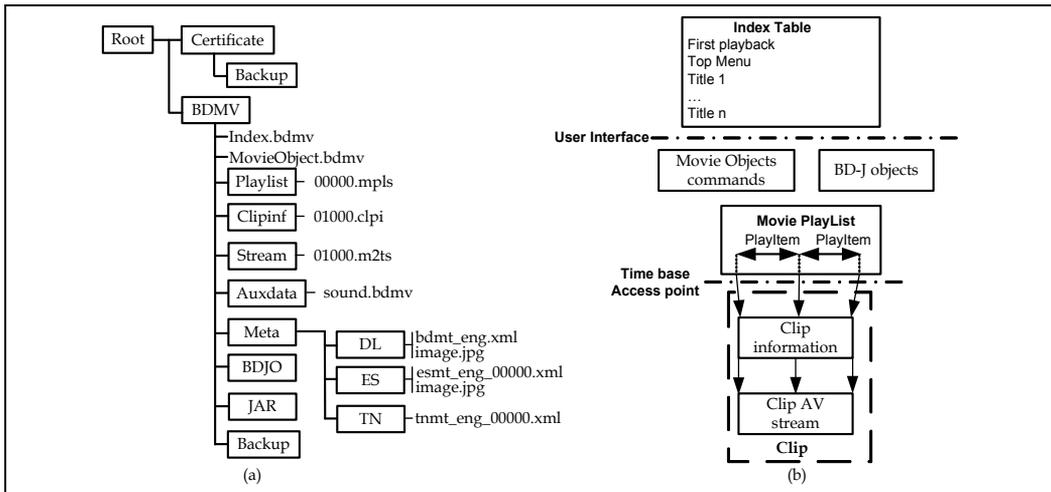


Fig. 19. BD-ROM data. (a) Example BD directory, (b) Simplified structure of BD-ROM.

If Title Scene Search is present, the corresponding information is separately stored from the content in “ES”, a sub-directory of the “META” directory, which also stores other metadata, see Fig. 19 (a). The value part of the metadata filename links the file to the corresponding playlist. Other sub-directories of the “META” directory are “DL” and “TN”. The directory “DL” (Disc Library) contains the disc-library metadata, which is mainly composed of two types of information: disc information and title information. Disc information contains the metadata of a disc itself, and title information includes the metadata of a title. The directory “TN” (Track/chapter Name) stores files containing metadata information regarding the names of tracks and chapters, which are sequentially stored. The track/chapter name metadata file exists per PlayList of a movie title. The corresponding metadata file is linked to the PlayList file via the value part of the file name.

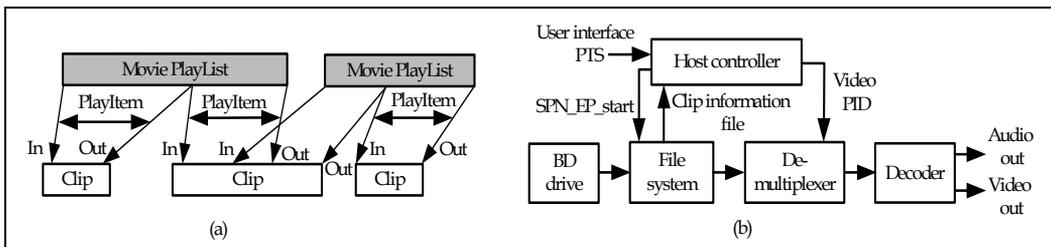


Fig. 20. Playback BD movie. (a) Movie playback using information from the PlayList, (b) Player model for I-picture search using EP_map.

6. Trick play for hard or solid-state disk based storage systems

With the advances in storage capacity and an associated firm price erosion, Hard Disk Drive (HDD) based storage systems have replaced the traditional tape-based storage media in the consumer arena. Recently, the Solid-State Drive (SSD) has made its appearance in the consumer arena, working its way into systems dominated by the HDD. Due to the enormous storage capacity of this type of media, a massive amount of audiovisual material

can be stored. Efficient navigation through this data requires fast random access, enabling high-speed search and new forms of audiovisual navigation. HDD and SSD storage solutions have an advantage regarding the access time, see Fig. 21, when used for audiovisual storage applications. For the situation that the access time of the storage medium is less than the reciprocal of the frame rate, the storage system responsiveness enables traditional trick play at full refresh rate and paves the way for advanced navigation methods.

The disk-based storage system architecture can be either push- or pull-based and is mainly determined by the context in which the storage system is deployed. A good solution is to offer both architectures, and make a choice depending on where the audiovisual decoding is performed, so that the best of both architectures is deployed. When audiovisual decoding is performed locally, the pull-based solutions give the best trick-play quality. When operated in a networked manner, trick-play quality cannot be guaranteed, as this depends on the normal-play encoding settings. It should be kept in mind, that the involved network communication is minimal for push-based architectures, in which only commands are required when changing the mode of operation, i.e. play, stop, fast search forward, etc. Moreover, when supporting both architectures, also advanced navigation features become possible, where the availability of the navigation mode depends on the location of decoding.

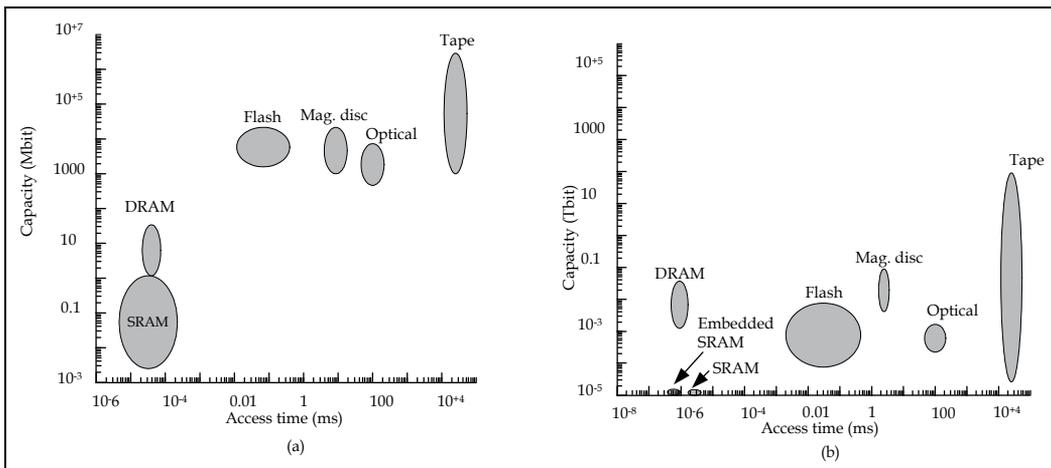


Fig. 21. Storage solutions and their access time. (a) Storage technology mid 1990s, (b) Storage technology 2009.

As a result, some advanced navigation modes may only be partly available when decoding is performed on remotely located decoders. Figure 22 depicts a functional block diagram of an HDD or SSD-based PVR, supporting both pull- and push-based playback. The pull mode requires the decoder to request for information, indicated via the dashed line. Let us elaborate on the two advanced navigation concepts introduced in Section 2. The pull-based architecture is suitable for traditional trick play and supports both advanced navigation concepts, whereas the push-based mode supports traditional trick play, hierarchical navigation and partly audiovisual trick play as discussed in Section 3.

During record, an MPEG-2 Transport Stream enters the system at the left-hand side, see Fig. 22. The CPI block adds time stamps to the incoming TS packets, which are stored together

with the corresponding TS packet on disk. Furthermore, the MPEG-2 TS is de-multiplexed and the video signal is parsed.

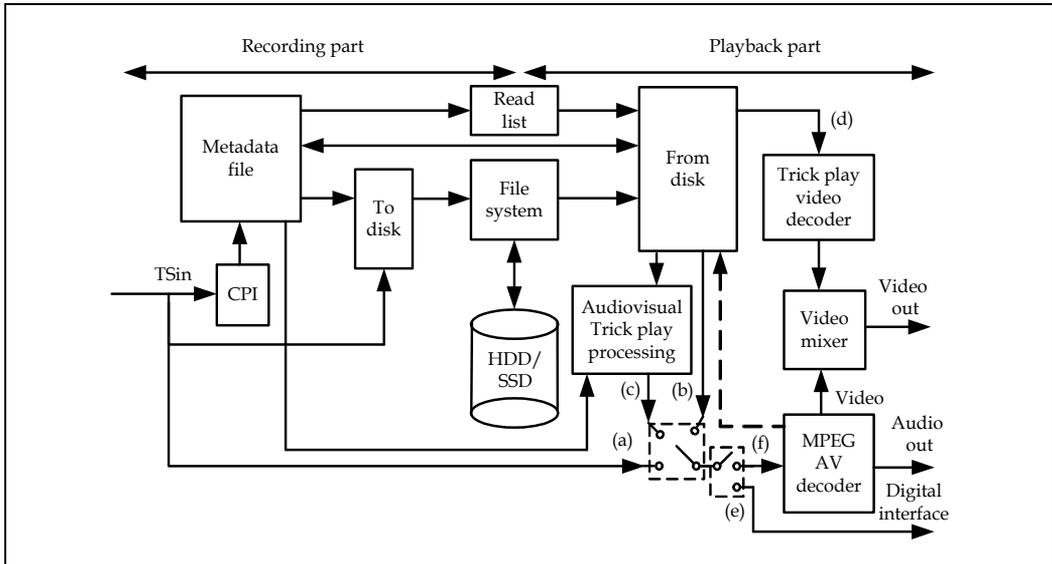


Fig. 22. Functional block diagram of HDD or SSD-based PVR. Operation modes: (a) View real-time, (b) View time-shift, (c) Traditional trick play or trick play based on normal-play AV fragments, (d) Trick play PiP to be used in combination with normal-play AV fragments, (e) AV playback according to push mode, (f) Pull-mode playback for local decoding.

During the parsing process, various characteristics of the MPEG-2 Video Elementary Stream are stored, such as relative positions of I- and P-type pictures, GOP length and transmission time of the intraframe-encoded pictures. This data is stored in a metadata file, which is stored as an extra descriptor file on the disk. Besides the information for traditional trick play, sub-pictures are derived from the normal-play video sequence and stored in the metadata file. The sub-pictures are generated in such a way that they can be used to construct mosaic screens via signal processing in the MPEG-2 domain, resulting in MPEG-2 compliant mosaic screen pictures (Eerenberg & de With, 2003). The information in the metadata file offers navigation features such as traditional trick play, but also the more advanced navigation methods for local (pull-mode) as well as remote users (push-mode).

For traditional trick play, the read-list block (using data from the metadata file) controls the disk retrieval process, resulting in fetching MPEG-2 TS fragments. These fragments are either processed by the audiovisual trick-play processing, which creates an MPEG-2 compliant signal, see Fig. 22 (c) and (e), or send the fragments directly to the local MPEG-2 decoder, see Fig. 22 (c) and (f).

For hierarchical navigation, the audiovisual trick-play block generates the mosaic screens for the pull- and push-based playback. For video navigation reinforced with audible sound, the audiovisual fragments are retrieved from disk and made available to the local decoder. A fast-search trick-play sequence is generated by means of a second video decoder, which could also be the control processor. The CIF-based trick-play sequence is merged with the normal-play video fragments by means of a video mixer.

7. Conclusions

Trick play is a feature that needs to be embedded within the design of a storage system. The system implementation should have low costs and be of suitable quality for quickly searching through the data. However, implementation of trick play differs for the various storage systems. Based on the discussed systems in this chapter, the following techniques have become apparent.

Trick play can be realized by re-use of normal-play compressed video data, or via additional dedicated trick-play information streams. Digital helical-scan recorders apply both mechanisms. For example, DV shuffles normal-play data to facilitate trick-play for a broad range of speed-up factors and D-VHS deploys dedicated virtual channel for each supported speed-up factor, which is filled with special trick-play information. The disadvantage of this solution is the reduction of the normal-play recording bandwidth, e.g. in D-VHS the reduction is about 5 Mbit/s.

Common optical-disc storage systems all deploy the re-usage of normal-play information and selectively address individual decodable pictures in the normal-play video stream. This is achieved via so-called locator information stored within the recorded stream. The optical pick-up unit is positioned, using this locator information, such that the retrieved information contains individually decodable pictures. In this way, the proper placement of the optical-pickup unit becomes a dominant factor in the random access time. As a consequence, the final trick-play quality depends on the normal-play GOP structure. The BD-based optical storage solution also optionally deploy normal-play GOP information, enabling the usage of P-type and B-type pictures for trick play. Moreover, metadata-based retrieval is facilitated by an optional descriptor file. With the current advances in optical drive technology, random access time can be avoided due the readout of the optical disc at a significant higher rate, causing the video source decoder to be the dominant factor for trick-quality.

With the introduction of HDD-based PVRs, trick play can be deployed in full temporal quality due to the short access time, when operated in pull-mode. Both HDD and SSD-based systems deploy locator information, enabling basic and advanced forms of navigation. When operated in a push-mode, the trick-play temporal quality most probably decreases somewhat, but the decoder always receives a compliant stream for decoding, thereby making this approach suitable for networked-based storage systems. Drawback of a push-based architecture is that it limits the advanced navigation modes, as the final signal must comply with the deployed standards, when system cost should be kept low.

Future work in the field of trick play should focus on low-cost algorithms which are capable of finding useful metadata to create the metadata search files. This will facilitate searching techniques, which are based on semantic understanding of the data content, so that high-level searching becomes possible.

8. References

- BD (2006). System Description Blu-ray Disc Read-Only Format, Part 3: Audio Visual Basic Specifications (3-1 Core Specifications) Version 2.01 DRAFT2
- Boyce, J. & Lane, F. (1993). Fast scan technology for digital video tape recorders. *IEEE Trans. Consumer Electron.*, Vol. 39, No. 3, August 1993, pp. 186-191, ISSN 0098-3063

- CEI/IEC61834 (1993). Helical-scan digital video cassette recording system using 6,35 mm magnetic tape for consumer use (525-60, 625-50, 1125-60 and 1250-50 systems)
- Eerenberg, O. & de With, P.H.N. (2003). System requirements and considerations for visual table of contents in PVR, *Proceedings of ICCE 2003*, pp. 24-25, ISBN 0-7803-7721-4, USA, June 2003, Los Angeles, CA
- Eerenberg, O. & Aarts, R.M. & de With, P.H.N. (2008). System Design of Advanced Video Navigation Reinforced with Audible Sound in Personal Video Recording, *Proceedings of ICCE 2008*, pp. 1-2, ISBN 978-1-4244-1458-1, USA, Jan. 2008, Las Vegas, NV
- ETSI_TS_101154 (2009). Technical Specification Digital Video Broadcasting (DVB); Specification for the use of Video and Audio Coding in Broadcasting Applications based on the MPEG 2 Transport Stream, Version 1.9.1, March 2009
- Fujita, M.; Higurashi, S. & Hirano, S. & Ohishi, T. & Zenno, Y. (1996). Newly developed D-VHS digital tape recording system for the multimedia era. *IEEE Trans. Consumer Electron.*, Vol. 42, No. 3, (August 1996), pp. 617-622, ISSN 0098-3063
- van Gassel, J.P. & Kelly, D.P. & Eerenberg, O. & de With, P.H.N. (2002). MPEG-2 compliant trick play over a digital interface, *Proceedings of ICCE 2002*, pp. 170-171, ISBN 0-7803-7300-6, USA, June 2002, Los Angeles, CA
- IEC62107 (2000). Super Video Compact Disc - Disc-interchange system-specification International Standard
- IEC61883-4 (2004). Digital Interface For Consumer Audio/Video Equipment Part 4: MPEG-TS data transmission
- ISO/IEC11172-2 (1993). Information technology - Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s - Part 2: Video
- ISO/IEC 14496-3 (1999). PICOLA Speed change algorithm, Annex 5.D.
- ISO/IEC13818-1 (2000). International Standard Information technology – Generic coding of moving pictures and associated audio information: Systems, Dec. 2000
- ISO/IEC13818-2 (2000). International Standard Information technology – Generic coding of moving pictures and associated audio information: Video, Dec. 2000
- ISO60774-5 (2004). Helical-scan video tape cassette system using 12,65 mm (0,5 in) magnetic tape on type VHS -Part 5: D-VHS
- Matsushita, Philips, Sony & Thomson. (1993). Outline of Basic Specification for Consumer-use Digital VCR
- Taylor, J. (2001). *DVD Demystified*, McGraw-Hill, ISBN 0-07-135026-8
- Ting, H. & Hang, H. (1995). Trick play schemes for advanced television recording on digital VCR. *IEEE Trans. Consumer Electron.*, Vol. 41 No. 4, Nov. 1995, pp. 1159-1168, ISSN 0098-3063
- US6621979 (1998). Trick play signal generation for a digital video recorder using retrieved intra-encoded pictures and generated inter-encoded pictures, Patent

Video Quality Metrics

Mylène C. Q. Farias
*Department of Computer Science
University of Brasília (UnB)
Brazil*

1. Introduction

Digital video communication has evolved into an important field in the past few years. There have been significant advances in compression and transmission techniques, which have made possible to deliver high quality video to the end user. In particular, the advent of new technologies has allowed the creation of many new telecommunication services (e.g., direct broadcast satellite, digital television, high definition TV, video conferencing, Internet video). To quantify the performance of a digital video communication system, it is important to have a measure of video quality changes at each of the communication system stages. Since in the majority of these applications the transformed or processed video is destined for human consumption, humans will ultimately decide if the operation was successful or not. Therefore, human perception should be taken into account when trying to establish the degree to which a video can be compressed, deciding if the video transmission was successful, or deciding whether visual enhancements have provided an actual benefit.

Measuring the quality of a video implies a direct or indirect comparison of the test video with the original video. The most accurate way to determine the quality of a video is by measuring it using psychophysical experiments with human subjects (ITU-R, 1998). Unfortunately, psychophysical experiments are very expensive, time-consuming and hard to incorporate into a design process or an automatic quality of service control. Therefore, the ability to measure video quality accurately and efficiently, without using human observers, is highly desirable in practical applications. Good video quality metrics can be employed to monitor video quality, compare the performance of video processing systems and algorithms, and to optimize the algorithms and parameter settings for a video processing system.

With this in mind, fast algorithms that give a physical measure (objective metric) of the video quality are used to obtain an estimate of the quality of a video when being transmitted, received or displayed. Customarily, quality measurements have been largely limited to a few objective measures, such as the mean absolute error (MAE), the mean square error (MSE), and the peak signal-to-noise ratio (PSNR), supplemented by limited subjective evaluation. Although the use of such metrics is fairly standard in published literature, it suffers from one major weakness. The outputs of these measures do not always correspond well with human judgements of quality.

In the past few years, a big effort in the scientific community has been devoted to the development of better video quality metrics that correlate well with the human perception of quality (Daly, 1993; Lubin, 1993; Watson et al., 2001; Wolf et al., 1991). Although much

has been done in the last ten years, there are still a lot of challenges to be solved since most of the achievements have been in the development of full-reference video quality metrics that evaluate compression artifacts. Much remains to be done, for example, in the area of no-reference and reduced-reference quality metrics. Also, given the growing popularity of video delivery services over IP networks (e.g. internet streaming and IPTV) or wireless channel (e.g. mobile TV), there is a great need for metrics that estimate the quality of the video in these applications.

In this chapter, we introduce several aspects of video quality. We give a brief description of the Human Visual System (HVS), discuss its anatomy and a number of phenomena of visual perception that are of particular relevance to video quality. We also describe the main characteristics of modern digital video systems, focusing on how visible errors (artifacts) are perceived in digital videos. The chapter gives a description of a representative set of video quality metrics. We also discuss recent developments in the area of video quality, including the work of the Video Quality Experts Group (VQEG).

2. The Human Visual System (HVS)

In the past century, the knowledge about the human visual system (HVS) has increased tremendously. Although much more needs to be learned before we can claim to understand it, the current state of the art of visual information-processing mechanisms is sufficient to provide important information that can be used in the design of video quality metrics. In fact, results in the literature show that video quality metrics that use models based on the characteristics of the HVS have better performance, i.e., give predictions that are better correlated with the values given by human observers (VQEG, 2003).

In this section, we introduce basic aspects of the anatomy and psychophysical features of the HVS that are considered relevant to video processing algorithms and, more specifically, to the design of video quality metrics.

2.1 Anatomy of the HVS

The eyes are far more than a simple camera. A more accurate description would be a self-focusing, self-adjusting for light intensity, and self-cleaning camera that provides a real-time output to a very advanced computer. The main components of the eye are the cornea, the pupil, the lens, and the fluids that fill the eye. A transverse section of the human eye is shown in Fig. 1.

The *optics* of the eye is composed by three major elements: the cornea, the pupil, and the lens. The light (visual stimulus) comes in through the optics and it is projected on the retina – the membrane located on the back of the eye. The optics works just like camera lens and their function is to project a clear and focused image on the retina – the retinal image. Given the physical limitation of the optics, the retinal image is only an approximation of the original image (the visual stimulus). As a result, the retinal image main contain some distortions, among which the most noticeable one is blurring. Since the response of optics is roughly linear, shift-invariant, and low-pass, the resulting retinal image can be approximated by convolving the input visual image with a blurring point spread function (PSF) (Marr, 1982).

The *retina* has the main function of translating the incoming light into nerve signals that can be understood by the brain. It has the shape of a plate and it is composed of many layers of neurons, as depicted in Fig. 2. The light projected on the retina has to pass through several layers before it reaches the photoreceptors cells and is absorbed by the pigment layer.

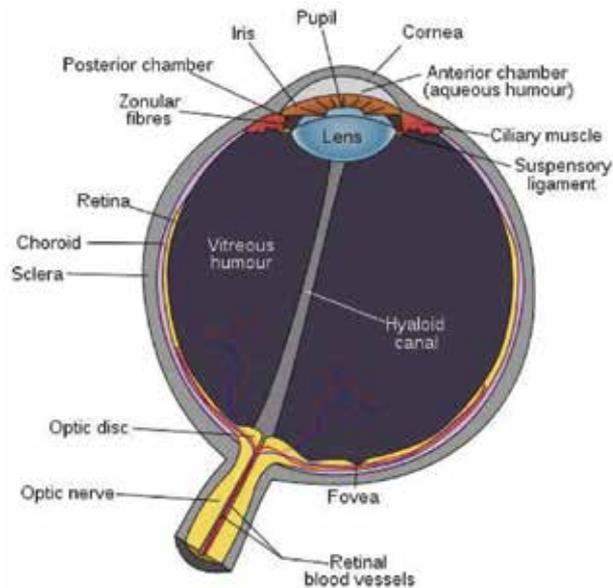


Fig. 1. Transverse section of the human eye (Wikimedia Commons, 2007).

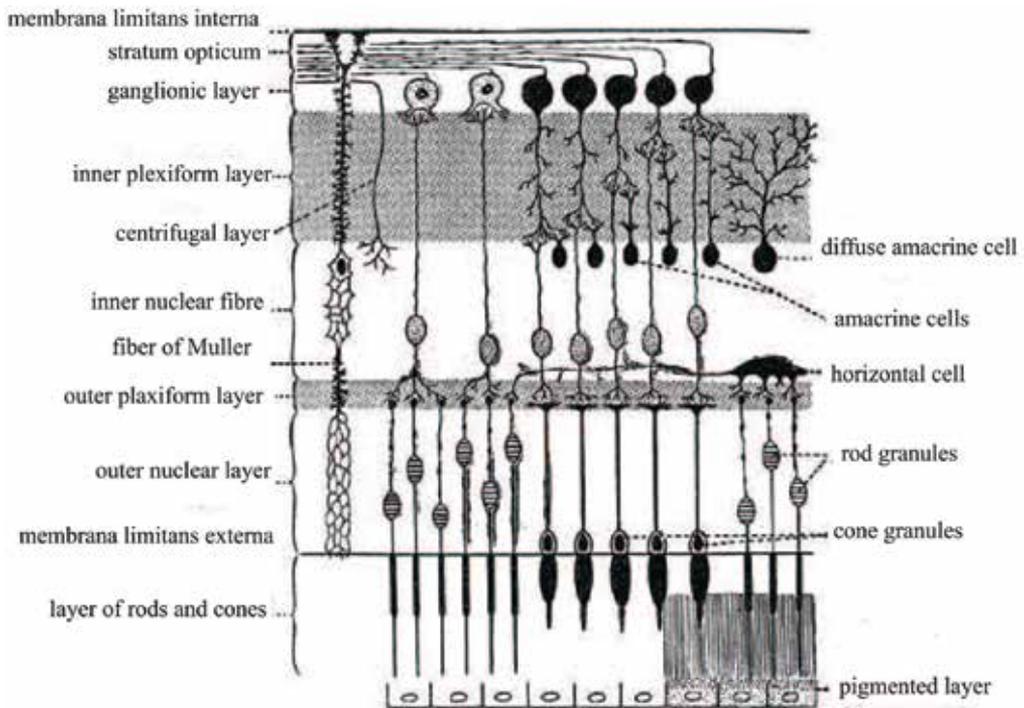


Fig. 2. Plan of retinal neurons. The retina is a stack of several neuronal layers. Light has to pass these layers (from top to bottom) to hit the photoreceptors (layer of rods and cones). The signal propagates through the bipolar and horizontal cells (middle layers) and, then, to the amacrine and ganglion cells. (Adapted from H. Grey (Grey, 1918))

The photoreceptor cells are specialized neurons that convert light energy into signals which can then be understood by the brain. There are two types of photoreceptors cells: *cones* and *rods*. Observe from Fig. 2 that the names are inspired by the shape of the cells. The rods are responsible for vision in low-light conditions. Cones are responsible for vision in normal high-light conditions, color vision, and have the ability to see fine details.

There are three types of cones, which are classified according to the spectral sensitivity of their photochemicals. The three types are known as *L-cones*, *M-cones*, and *S-cones*, which stand for long, medium, and short wavelengths cones, respectively. Each of them has peak sensitivities around 570nm, 540nm, and 440nm, respectively. These differences are what makes color perception possible. The incoming light from the retina is split among the three types of cones, according to its spectral content. This generates three visual streams that roughly correspond to the three primary colors red, green, and blue.

There are roughly 5 million cones and 100 million rods in a human eye. But their distribution varies largely across the surface of the retina. The center of the retina has the highest density of cones and ganglion cells (neurons that carry the electrical signal from the eye to the brain through the optic nerve). This central area is called *fovea* and is only about half a millimeter in diameter. As we move away from it, the density of both cones and ganglion cells falls off rapidly. Therefore, the fovea is responsible for our fine-detail vision and, as a consequence, we cannot perceive the entire visual stimulus at uniform resolution.

The majority of cones in the retina are L- and M-cones, with S-cones accounting for less than 10% of the total number of cones. Rods, on the other hand, dominate outside the fovea. As a consequence, it is much easier to see dim objects when they are located in the peripheral field of vision. Looking at Fig. 1, we can see that there is a hole or *blind spot*, where the optic nerve is. In this region there are no photoreceptors.

The signal collected from the photoreceptors has to pass through several layers of neurons in the retina (retinal neurons) before being carried off to the brain by the optic nerve. As depicted in Fig. 2, different types of neurons can be found in the retina:

- *Horizontal cells* link receptors and bipolar cells by relatively long connections that run parallel to the retinal layers.
- *Bipolar cells* receive input from the receptors, many of them feeding directly into the retinal ganglion cells.
- *Amacrine cells* link bipolar cells and retinal ganglion cells.
- *Ganglion cells* collect information from bipolar and amacrine cells. Their axons form the optic nerve that leaves the eye through the optic disc and carries the output signal from the retina to other processing centers in the brain.

The signal leaves the eye through the *optic nerve*, formed by the axons of the ganglion cells. A scheme showing central connections of the optic nerves to the brain is depicted in Fig. 3. Observe that the optic nerves from the left and right eye meet at the *optic chiasm*, where the fibers are rearranged. About half of these fibers cross to the opposite side of the brain and the other half stay on the same side. In fact, the corresponding halves of the field of view (right and left) are sent to the left and right halves of the brain. Considering that the retinal images are reversed by the optics of the eye, the right side of the brain processes the left half (of the field of view) of both eyes, while the left side processes the right half of both eyes. This is illustrated by the red and blue lines in Fig. 3.

From the optic chiasm, the fibers are taken to several parts of the brain. Around 90% of them finish at the two *lateral geniculate body*. Besides serving as a relay station for signals from the

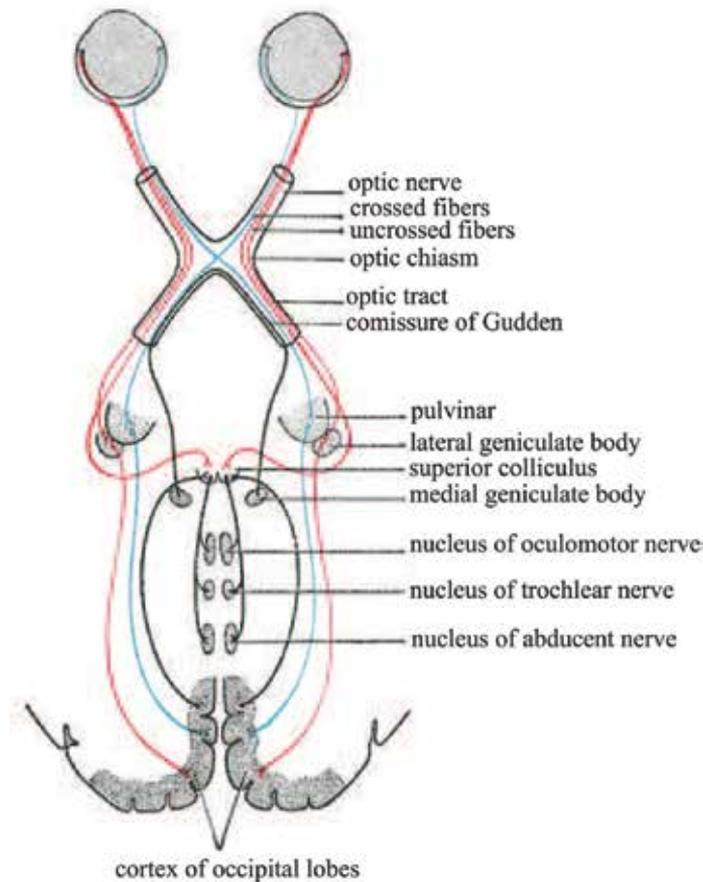


Fig. 3. Scheme showing central connections of the optic nerves and optic tracts. (Adapted from H. Grey (Grey, 1918))

retina to the visual cortex, the lateral geniculate body controls how much information is allowed to pass. From there, the fibers are taken to the visual cortex.

The *virtual cortex* is the region of the brain responsible for processing the visual information. It is located on the back of the cerebral hemispheres. The region that receives the information from the lateral geniculate body is called the *primary visual cortex* (also known as V1). In addition to V1, more than 20 other areas receiving visual input have been discovered, but little is known about their functionalities.

V1 is a region specialized on processing information about static and moving objects and recognizing patterns. There is a big variety of cells in V1 that have selective sensitivity to certain types of information. In other words, one particular cell may respond strongly to patterns of a certain orientation or to motion in a certain direction. Others are tuned to particular frequencies, color, velocities, etc. An interesting characteristic of these neurons is the fact that their outputs saturates as the input contrast increases.

The selectivity of the neurons in V1 is the heart of the multichannel organization characteristic of the human vision system. In fact, the neurons in V1 can be modeled as an octave-band Gabor filter bank, where the spatial frequency spectrum (in polar

representation) is sampled at octave intervals in the radial frequency dimension and at uniform intervals in the orientation dimension (Marr, 1982). This model is used by several algorithms in image processing and video quality assessment.

2.2 Perceptual features

A number of visual perception phenomena are a consequence of the characteristics of the optics of the human eye. The phenomena described in this section are of particular interest to the area of image processing and, more specifically to video quality.

2.2.1 Foveal and peripheral vision

The densities of the photoreceptors and ganglion cells in the retina are not uniform, increasing towards the center of the retina (fovea) and decreasing on the contrary direction. As a consequence, the resolution of objects in the visual field is also not uniform. The point where the observer fixates is projected on the fovea and, consequently, resolved with the highest resolution. The objects in the peripheral area are resolved with progressively lower resolution (peripheral vision).

2.2.2 Light adaptation

In the real world, the amount of light intensity varies tremendously, from dim (night) to high intensity (sun day). The HVS adapts to this large range by controlling the amount of light that enters the eye. This is done by increasing/decreasing the diameter of the pupils and, at the same time, adjusting the gain of post-receptor neurons in the retina. As a result, instead of coding absolute light intensities, the retina encodes the contrast of the visual stimulus.

The phenomenon that keeps the contrast sensitivity over a wide range of light intensity is known as Weber's law:

$$\Delta I / I = K$$

where I is the background luminance, ΔI is the just noticeable incremental luminance over the background, and K is a constant called the Weber fraction.

2.2.3 Contrast Sensitivity Functions (CSF)

CSF models the sensitivity of the HVS as a function of the spatial frequency of the visual stimuli. A typical CSF is shown in Fig. 4(a). Spatial contrast sensitivity peaks at 3 cycles per degree (cpd), and declines more rapidly at higher than at lower spatial frequencies. Frequencies higher than 40 cpd (8 cpd scotopic) are undetectable even at maximum contrast. For illustration purposes, consider the image in Fig. 4(b) that corresponds to the intensities of a sinusoidal luminance grating. In this image, the spatial frequency (number of luminance cycles the grating repeats in one degree of visual angle) increases from left to right, while contrast (difference between the maximum and minimum luminance) increases from top to bottom. The shape of the visible lower part of the image gives an indication of our relative sensitivity to different spatial frequencies. If the perception of contrast were determined solely by the image contrast, then the alternating bright and dark bars should appear to have equal height across any horizontal line across the image. However, the bars are observed to be significantly higher at the middle of the image, following the shape of the CSF (see Fig. 4(a)).

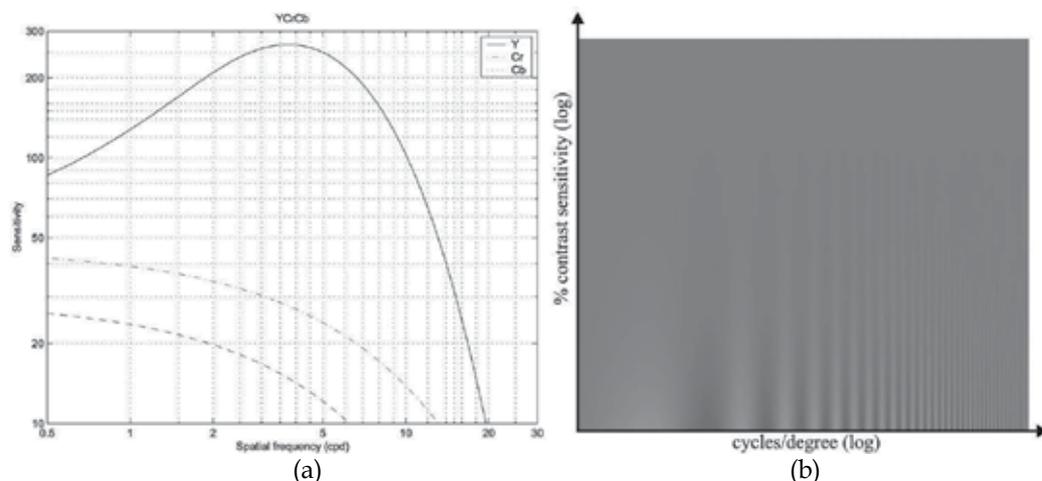


Fig. 4. (a) Contrast sensitivity functions for the three channels YCbCr (after Moore, 2002 (Moore, 2002)). (b) Pelli-Robson Chart, where spatial frequency increases from left to right, while contrast increases from top to bottom.

2.2.4 Masking and facilitation

Masking and facilitation are important aspects of the HVS in modeling the interactions between different image components present at the same spatial location. Specifically, these two effects refer to the fact that the presence of one image component (*the mask*) will decrease/ increase the visibility of another image component (*test signal*). The mask generally reduces the visibility of the test signal in comparison with the case where the mask is absent. However, the mask may sometimes facilitate detection as well. Usually, the masking effect is the strongest when the mask and the test signal have similar frequency content and orientations. Most quality metrics incorporate a model for masking and/or facilitation.

2.2.5 Pooling

Pooling refers to the task of arriving at a single measurement of quality from the outputs of the visual streams. It is not quite understood how the HVS performs pooling. But, it is clear that a perceptible distortion may be more annoying in some areas of the scene (such as human faces) than in others. Most quality metrics use the *Minkowski metric* to pool the error signals from the streams with different frequency and orientation selective and arrive at a fidelity measurement (de Ridder, 1992; 2001). The Minkowski metric is also used to combine information across spatial and temporal coordinates.

3. Digital video systems

In this section, we give a brief overview of the available video compression and transmission techniques and their impact on the quality of a digital video.

3.1 Video compression

Video compression (or video coding) is the process of converting a video signal into a format that takes up less storage space or transmission bandwidth. Given the video

transmission and storage requirements (up to 270 Mbits/s for Standard Definition and 1.5 Gbit/s for High Definition), video compression is an essential technology for applications such as digital television (terrestrial, cable or satellite transmission), optical storage/reproduction, mobile TV, videoconferencing and Internet video streaming (Poynton, 2003).

There are two types of compression: *lossy* and *lossless* compression (Bosi & Goldberg, 2002). Lossless compression algorithms have the characteristic of assuring perfect reconstruction of the original data. Unfortunately, this type of compression only allows around 2:1 compression ratios, which is not sufficient for video applications. Lossy compression is the type of compression most commonly used for video because it provides much bigger compression ratios. There is, of course, a trade-off: the higher the compression ratio, the lower the quality of the compressed video.

Compression is achieved by removing the redundant information from the video. There are four main types of redundancies that are typically explored by compression algorithms:

- *Perceptual redundancy*: Information of the video that cannot be easily perceived by the human observer and, therefore, can be discarded without significantly altering the quality of the video.
- *Temporal redundancy*: Pixels in successive video frames have great similarity. So, even though motion tend to change the position of blocks of pixels, it does not change their values and therefore their correlation.
- *Spatial redundancy*: There is a significant correlation among pixels around the same neighborhood in a frame.
- *Statistical redundancy*: This type of redundancy is related to the statistical relationship within the video data (bits and bytes).

Each stage of a video compression algorithm is responsible for mainly reducing one type of redundancy. Fig. 5 depicts the functional components in a typical video compression algorithm. Different algorithms differ in what tools are used in each stage. But, most of them share the same principles: motion compensation and block-based transform with subsequent quantization. Currently, there are several standards for video compression, which standardize the decoding process. The encoding process is not fixed, what leaves room for innovation.

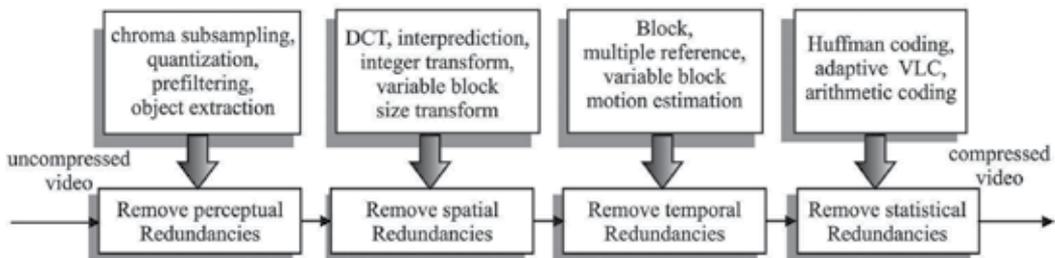


Fig. 5. Functional components in a typical video compression algorithm.

The most popular compression standards were produced by the Motion Picture Experts Group (MPEG) (ITU, 1998) and the Video Coding Experts (VCEG). The MPEG is a working group of the International Organization for Standardization (ISO) and of the International Electrotechnical Commission (IEC), formally known as ISO/IEC - JTC1/SC29/WG11. Among the standards developed by MPEG are MPEG-1, MPEG-2, and MPEG-4. The MPEG-2

is a very popular standard used not only for broadcasting, but also in DVDs (Haskell et al., 1997; ITU, 1998). The main advantage of MPEG-2 is its low cost, given its popularity and the large scale of production. MPEG-2 is also undoubtedly a very mature technology.

The VCEG is a working group of the Telecommunication Standardization Sector of the International Telecommunication Union (ITU-T). Among the standards developed by VCEG are the H.261 and H.263. A joint collaboration between MPEG and VCEG resulted in the development of the H.264, also known as MPEG-4 Part 10 or AVC (Advanced Video Coding) (Richardson, 2003; ITU, 2003). The H.264 represents a major advance in the technology of video compression, providing a considerable reduction of bitrate when compared to previous standards (Lambert et al., Jan. 2006). For the same quality level, H.264 provides a bitrate of about half the bitrate provided by MPEG-2.

3.2 Digital video transmission

Compressed video streams are mainly intended for transmission over communication networks. But, there are different types of video communication and streaming applications. Each one has particular operating conditions and properties. The channels used for video communication may be static or dynamic, packet-switched or circuit-switched. Also, the channels may support a constant or variable bit rate transmission, and may support some form of Quality of Service (QoS) or may only provide best effort support. Finally, the transmission may be point-to-point, multicast, and broadcast.

In most cases, after the video has been digitally compressed, the resulting bitstream is segmented into fixed or variable packets and multiplexed with other data types, such as audio. The next stage is the channel encoder, which will add error protection to the data. The characteristics of the specific video communication application will, of course, have a great impact on the quality of the video displayed at the receiver.

3.3 Common artifacts in digital video systems

An impairment is a property of the video that is perceived as undesirable, whether it is in the original or not. Impairments can be introduced during capture, transmission, storage, and/or display, as well as by any image processing algorithm (e.g. compression) that may be applied along the way (Yuen & Wu, 1998). They can be very complex in their physical descriptions and also in their perceptual descriptions. Most of them have more than one perceptual feature, but it is possible to have impairments that are relatively pure. To differentiate impairments from their perceptual features, we will use the term *artifact* to refer to the perceptual features of impairments and *artifact signal* to refer to the physical signal that produces the artifact.

The most common artifacts present in digital video are:

- *Blockiness* or *blocking* – A type of artifact characterized by a block pattern visible in the picture. It is due to the independent quantization of individual blocks (usually of 8x8 pixels in size) in block-based DCT coding schemes, leading to discontinuities at the boundaries of adjacent blocks. The blocking effect is often the most visible artifact in a compressed video, given its periodicity and the extent of the pattern. More modern codecs, like the H.264, use a deblocking filter to reduce the annoyance caused by this artifact.

- *Blur or blurring* – It is characterized for a loss of spatial detail and a reduction of edge sharpness. In the in the compression stage, blurring is introduced by the suppression of the high-frequency coefficients by coarse quantization.
- *Color bleeding* – It is characterized by the smearing of colors between areas of strongly differing chrominance. It results from the suppression of high-frequency coefficients of the chroma components. Due to chroma subsampling, color bleeding extends over an entire macroblock.
- *DCT basis image effect* – It is characterized by the prominence of a single DCT coefficient in a block. At coarse quantization levels, this results in an emphasis of the dominant basis image and reduction of all other basis images.
- *Staircase effect* – These artifacts occurs as a consequence of the fact that DCT basis are best suited for the representation of horizontal and vertical lines. The representation of lines with other orientations require higher-frequency DCT coefficients for accurate reconstruction. Therefore, when higher frequencies are lost, slanted lines appear.
- *Ringling* – Associated with the Gibbs phenomenon. It is more evident along high contrast edges in otherwise smooth areas. It is a direct result of quantization leading to high-frequency irregularities in the reconstruction. Ringing occurs with both luminance and chroma components.
- *Mosquito noise* – Temporal artifact that is seen mainly in smoothly textured regions as luminance/chrominance fluctuations around high contrast edges or moving objects. It is a consequence of the coding differences for the same area of a scene in consecutive frames of a sequence.
- *Flickering* – It occurs when a scene has a high texture content. Texture blocks are compressed with varying quantization factors over time, which results in a visible flickering effect.
- *Packet loss* – It occurs when parts of the video are lost in the digital transmission. As a consequence, parts (blocks) of video are missing for several frames.
- *Jitter* – It is the result of skipping regularly video frames to reduce the amount of video information that the system is required to encode or transmit. This creates motion perceived as a series of distinct snapshots, rather than smooth and continuous motion.

The performance of a particular digital video system can be improved if the type of artifact that is affecting the quality of the video is known (Klein, 1993). This type of information can also be used to enhance the video by reducing or eliminating the identified artifacts (Caviedes & Jung, 2001). In summary, this knowledge makes it possible to implement a complete system for detecting, estimating and correcting artifacts in video sequences. Unfortunately, there is not yet a good understanding of how visible/annoying these artifacts are, how the content influences their visibility/annoyance, and how they combine to produce the overall annoyance. A comprehensive *subjective* study of the most common types of artifacts is still needed.

An effort in this direction has been done by Farias *et al* (Farias, Moore, Foley & Mitra, 2002; Farias *et al.*, 2003a;b; Farias, Foley & Mitra, 2004; Farias, Moore, Foley & Mitra, 2004). Their approach makes use of synthetic artifacts that look like “real” artifacts, yet are simpler, purer, and easier to describe. This approach makes it possible to control the type, proportion, and strength of the artifacts being tested and allows to evaluate the performance of different combination models of the artifact metrics. The results gathered from the psychophysical experiments performed by Farias *et al* show that the synthetic artifacts,

besides being visually similar to the real impairments, have similar visibility and annoyance properties. Their results also show that there is an interaction between among different types of artifacts. For example, the presence of noisy artifact signals seem to decrease the perceived strength of the other artifacts, while the presence of blurry artifact signals seem to increase it. The authors also modeled annoyance by combining the artifact perceptual strengths (MSV) using both a Minkowski metric and a linear model (de Ridder, 1992).

4. Subjective video quality assessment

Subjective experiments (also called psychophysical experiments) represent the most accurate way of measuring the quality of a video. In subjective experiments, a number of subjects (observers or participants) are asked to watch a set of test sequences and give judgements about their quality or the annoyance of the impairments. The average of the values collected for each test sequence are known as Mean Observer Score (MOS).

In general, subjective experiments are expensive and time-consuming. The design, execution, and data analysis consume a great amount of the experimenter's time. Running an experiment requires the availability of subjects, equipment, and physical space. As a result, the number of experiments that can be conducted is limited and, therefore, an appropriate methodology should be used to get the most out of the resources.

The International Telecommunication Union (ITU) has recommendations for subjective testing procedures. The two most important documents are the ITU-R Rec. BT.500-11 (ITU-R, 1998), targeted at television applications, and the ITU-T Rec. P.910 (ITU-T, 1999), targeted at multimedia applications. These documents give information regarding the standard viewing conditions, the criteria for selections of observers and test material, assessment procedures, and data analysis methods. Before choosing which method to use, the experimenter should take into account the application in mind and the accuracy objectives.

According to ITU, there are two classes of subjective assessments:

- *Quality assessments* - The judgements given by subjects are in a quality scale, i.e., how good or bad is the quality of the displayed video. These assessments establish the performance of systems under optimum conditions;
- *Impairment assessments* - The judgements given by subjects are in an impairment scale, i.e., how visible or imperceptible are the impairments in the displayed video. These assessments establish the ability of systems to retain quality under non-optimum conditions that relate to transmission.

According to the type of scale, quality or impairment judgements can be classified as *continuous* or *discrete*. Judgements can also be categorical or non-categorical, adjectival or numerical. Depending on the form of presentation of the stimulus (sequences), the assessment method can be classified as *double* or *single* stimulus. In the single stimulus approach the test sequence is presented by itself, while in the double stimulus method a pair of sequences (test sequence and the corresponding reference) are presented together.

The most popular assessment procedures of ITU-R Rec. BT.500-11 are:

- *Double Stimulus Continuous Quality Scale (DSCQS)* - This method is specially useful when the test conditions exhibit the full range of quality. The observer is shown multiple pairs of sequences consisting of a test sequence and the corresponding reference. The sequences have a short duration of around 10s and are presented twice, alternated by each other. The observers are not told which is the reference and which is the test sequence. In each trial, their positions are changed randomly. The observer is

asked to assess the overall quality of both sequences by inserting a mark on a vertical scale. Fig. 6 shows a section of a typical score sheet. The continuous scales are divided into five equal lengths, which correspond to the normal ITU five-point quality continuous scale.

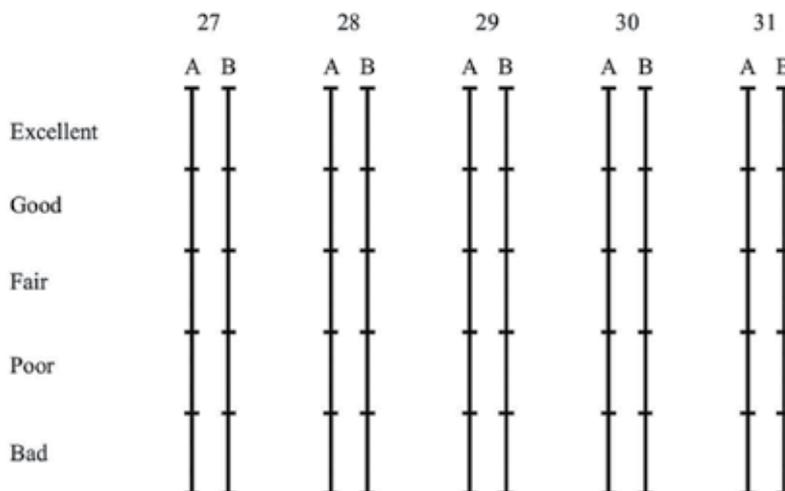


Fig. 6. Continuous quality scale used in DSCQS.

- *Double Stimulus Impairment Scale (DSIS)* – For this method, the reference is always shown before the test sequence and the pair is not repeated. Observers are asked to judge the amount of impairment in the test sequence using a five-level scale. The categories in the scale are ‘imperceptible’, ‘perceptible, but not annoying’, ‘slightly annoying’, ‘annoying’, and ‘very annoying’. This method is adequate for evaluating visible artifacts.
- *Single Stimulus Continuous Quality Evaluation (SSCQE)* – In this method, observers are asked to watch a video (program) of around 20-30 minutes. The content is processed using the conditions under test and the reference is not presented. The observer uses a slider to continuously rate the quality, as it changes during the presentation. The scale (ruler) goes from ‘bad’ to ‘excellent’.

The most popular assessment procedures of ITU-T Rec. P.910 are:

- *Absolute Category Rating (ACR)* – Also known as Single Stimulus Method (SSM), this method is characterized by the fact that the test sequences are presented one at a time, without the reference. This makes it a very efficient method, compared to DSIS or DSCQS, which have durations of around 2 to 4 times longer. After each presentation, observers are asked to judge the overall quality of the test sequence using a five-level scale. The categories in this scale are ‘bad’, ‘poor’, ‘fair’, ‘good’, and ‘excellent’. A nine-level scale may be used if a higher discriminative power is desired. Also, if additional ratings of each test sequence are needed, repetitions of the same test conditions at different points in time of the test can be used.
- *Degradation Category Rating (DCR)* – This method is identical to the DSIS described earlier.
- *Pair Comparison (PC)* – In this method, all possible pair combinations of all test sequences are shown to viewers, i.e., if there are n test conditions, a total of $n \cdot (n - 1)$

pairs are presented for each reference. The observers have to choose which sequence of the pair he/she thinks has the best quality. This method allows a very fine distinction between conditions, but also requires a longer period of time when compared to other methods.

Although each assessment method has its own requirements, the following recommendations are valid in most cases:

- The choice of test sequences must take into account the goal of the experiment. The spatial and temporal content of the scenes, for example, are critical parameters. These parameters determine the type and severeness of the impairments present in the test sequences.
- It is important that the set of test scenes spans the full range of quality commonly encountered for the specific conditions under test.
- When a comparison among results from different laboratories is the intention, it is mandatory to use a set of common source sequences to eliminate further sources of variation.
- The test sequences should be presented in a pseudo-random order and, preferably, the experimenter should avoid that sequences generated from the same reference be shown in a subsequent order.
- The viewing conditions, which include the distance from the subject's eye to the monitor and the ambient light, should be set according to the standards.
- The size and the type of monitor or display used in the experiment must be appropriate for the application under test. Calibration of the monitor may be necessary.
- It is best to use the whole screen for displaying the test sequences. In case this is not possible, the sequences must be displayed on a window of the screen, with a 50% grey ($Y=U=V=128$) background surrounding it.
- Before the experiment starts, the subjects should be tested for visual acuity. After that, written and oral instructions should be given to them, describing the intended application of the system, the type of assessment, the opinion scale, and the presentation methodology.
- At least 15 subjects should be used in the experiment. Preferably, the subjects should not be considered 'experts', i.e., have considerable knowledge in the area of image and video processing.
- Before the actual experiment, indicative results can be obtained by performing a pilot test using only a couple (4-6) of subjects (experts or non-experts).
- A training section with at least five conditions should be included at the beginning of the experimental session. These conditions should be representative of the ones used in the experiment, but should not be taken into account in the statistical analysis of the gathered data. It should be made clear to the observer that the worst quality seen in the training set does not necessarily correspond to the worst or lowest grade on the scale.
- Include at least two replications (i.e. repetitions of identical conditions) in the experiment. This will help to calculate individual reliability per subject and, if necessary, to discard unreliable results from some subjects.
- Statistical analysis of the gathered data can be performed using standard methods (Snedecor & Cochran, 1989; Hays, 1981; Maxwell & Delaney, 2003; ITU-R, 1998). For each combination of the test variables, the mean value and the standard deviation of the collected assessment grades should be calculated. Subject reliability should also be estimated.

5. Objective video quality metrics

Video quality metrics can be employed to:

- *monitor* video quality;
- *compare* the performance of video processing systems and algorithms; and
- *optimize* the algorithms and parameter settings for a video processing system.

The choice of which type of metric should consider the application and its the requirements and limitations.

In general, video quality metrics can be divided in three different categories according to the availability of the original (reference) video signal:

- *Full Reference (FR)* metric - Original and distorted (or test) videos are available.
- *Reduced Reference (RR)* metric - Besides the distorted video, a description of the original and some parameters are available.
- *No-reference (NR)* metric - Only the distorted video is available.

Figs. 7, 8, and 9 depict the block diagrams corresponding to the full reference, reduced reference, and no-reference video quality metrics, respectively. Observe that on the FR approach the entire reference is available at the measurement point. On the RR approach only part of the reference is available through an auxiliary channel. In this case, the information available at the measurement point generally consists of a set of features extracted from the reference. For the NR approach no information concerning the reference is available at the measuring point.

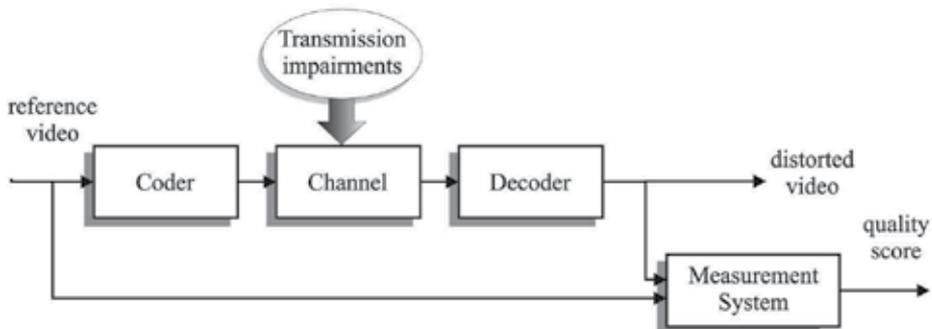


Fig. 7. Block diagram of a full reference video quality assessment system.

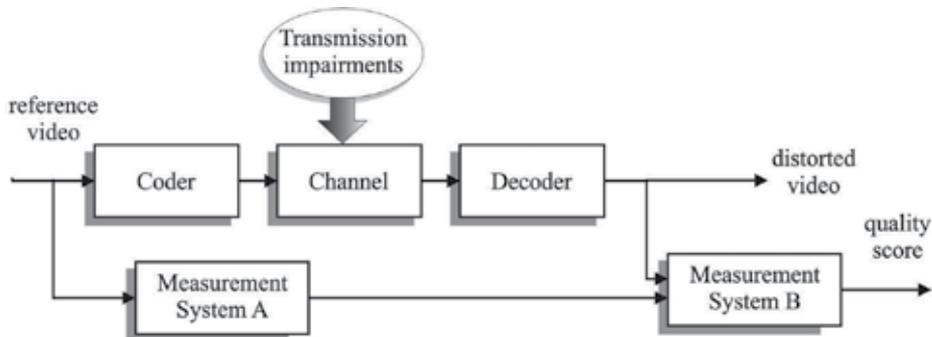


Fig. 8. Block diagram of a reduced reference video quality assessment system.

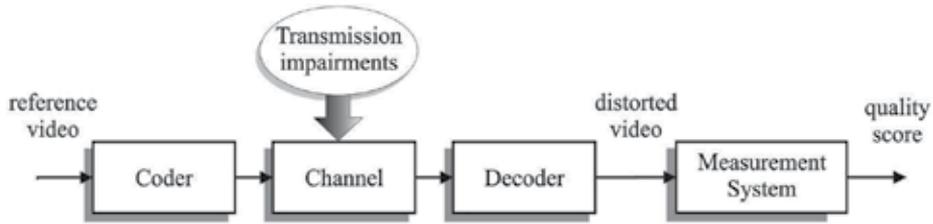


Fig. 9. Block diagram of a no-reference video quality assessment system.

These three classes of metrics are targeted at different applications. FR metrics are more suitable for offline quality measurements, for which a detailed and accurate measurement of the video quality is of higher priority than having immediate results. NR and RR metrics are targeted at real-time applications, where the computational complexity limitations and the lack of access to the reference are the main issues. Comparisons among the performances of several video quality metrics were done by Yubing Wang (Wang, 2006), Eskicioglu and Fisher (Eskicioglu & Fisher, 1995), Sheikh *et al* (Sheikh *et al.*, 2006), and Avcibas *et al* (Avcibas *et al.*, 2002).

The quality metrics can be classified according to the approach they take for estimating the amount of impairment in a video. There are basically two main approaches. The first one is the *error sensitivity* approach that tries to analyze visible differences between the test and reference videos. This approach is mostly used for full reference metrics, since this is the only type of metric where a pixel-by-pixel difference between the original and test videos can be generated.

The second approach is the *feature extraction* approach that looks for higher-level features that do not belong to the original video to obtain an estimate of the quality of the video. No-reference and reduced reference metrics frequently use the feature extraction approach making use of some a priori knowledge of the features of the original video.

Finally, quality metrics can also be classified according to what type of information they consider when processing the video. Metrics that take into account the how the HVS works are typically called *picture metrics* or *perceptual metrics*. More simple metrics that only measure the fidelity of the signal without considering its content are called *data metrics*.

In this section, a brief description of a representative set of FR, RR, and NR metrics is presented. Also, a description of data metrics and metrics based on data hiding is presented.

5.1 Data FR fidelity metrics

Data fidelity metrics measure the physical differences between two signals without considering its content. Two of the most popular data fidelity metrics are the mean squared error (MSE) and the peak signal-to-noise ratio (PSNR), which are defined as:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (X_i - Y_i)^2, \quad (1)$$

and

$$\text{PSNR} = 10 \cdot \log_{10} \frac{255^2}{\text{MSE}}, \quad (2)$$

where N is the total number of pixels in the video, 255 is the maximum intensity value of the images, and X_i and Y_i are the i -th pixels in the original and distorted video, respectively.

Strictly speaking, the MSE measures image differences, i.e. how different two images are. PSNR, on the other hand, measures image fidelity, i.e. how close two images are. In both cases, one of the pictures is always the reference (uncorrupted original) and the other is the test or distorted sequence.

The MSE and PSNR are very popular in the image processing community because of their physical significance and of their simplicity, but over the years they have been widely criticized for not correlating well with the perceived quality measurement (Teo & Heeger, 1994; Eskicioglu & Fisher, 1995; Eckert & Bradley, 1998; Girod, 1993; Winkler, 1999). More specifically, it has been shown that simple metrics like PSNR and MSE can only predict subjective rating with a reasonable accuracy, as long as the comparisons are made for the same content, the same technique or the same type of artifact (Eskicioglu & Fisher, 1995).

One of the major reasons why these simple metrics do not perform as desired is because they do not incorporate any HVS features in their computation. In fact, it has been discovered that in the primary visual cortex of mammals, an image is not represented in the pixel domain, but in a rather different manner. The measurements produced by metrics like MSE or PSNR are simply based on a pixel to pixel comparison of the data, without considering what is the content. These simple metrics do not consider, for example, what are the relationships among pixels in an image (or frames). They also do not consider how the spatial and frequency content of the impairments are perceived by human observers.

5.2 Full reference video quality metrics

In general, full reference (FR) metrics have the best performance among the three types of metrics. This is mainly due to the availability of the reference video. Also, since FR are intended for off-line applications, they can be more computationally complex and incorporate several aspects of the HVS. The major drawback of the full reference approach is the fact that a large amount of reference information has to be provided at the final comparison point. Also, a very precise spatial and temporal alignment of reference and impaired videos is needed to guarantee the accuracy of the metric.

A large number of FR metrics are *error sensitivity* metrics, which attempt to analyze and quantify the error signal in a way that simulates the human quality judgement. Some examples include the works by Daly (Daly, 1993), Lubin (Lubin, 1995), Teo and Heeger (Teo & Heeger, 1994), Watson (Watson, 1990; 1998; Watson et al., 2001), Van den Branden Lambrecht and Kunt (van den Branden Lambrecht & Kunt, 1998), and Winkler (Winkler, 1999). The group of *full reference* metrics that uses a *feature extraction* approach is much smaller and includes the works of Algazi and Hiwasa (Algazi & Hiwasa, 1993), Pessoa *et al.* (Pessoa et al., 1998), and Wolf and Pinson (Wolf & Pinson, 1999). In this section, we present a brief description of a representative set of full reference video quality metrics.

5.2.1 Visible Differences Predictor (VDP)

The full reference model proposed by Daly (Daly, 1993; 1992) is known as visible differences predictor (VDP). The general approach of the model consists of finding what limits the visual sensitivity and taking this into account when analysing the differences between distorted and reference videos. The main sensitivity limitations (or variations) considered by the model are *light level*, *spatial frequency*, and *signal content*. Each of these sensitivity variations corresponds to one of the stages of the model, as described below:

- Amplitude non-linearity – It is well known that sensitivity and perception of lightness are non-linear functions of luminance. The amplitude non-linearity stage of the VDP

describes the sensitivity variations as a function of the gray scale. It is based on a model of the early retina network.

- Contrast Sensitivity Function (CSF) - The CSF describes the variations in the visual sensitivity as a function of spatial frequency. The CSF stage changes the input as a function of light adaptation, noise, color, accommodation, eccentricity, and image size.
- Multiple detection mechanism - It is modeled with four subcomponents:
 - Spatial cortex transform - It models the frequency selectivity of the visual system and creates the framework for multiple detection mechanisms. This is modeled by a hierarchy of filters modified from Watson's cortex transform (Watson, 1987) that separates the image into spatial levels followed by six orientation levels.
 - Masking function - Models the magnitude of the masking effect.
 - Psychometric function - Describes the details of the threshold.
 - Probability summation - Combines the responses of all detection mechanisms into an unified perceptual response.

A simplified block-diagram of the VDP is depicted in Fig. 10. The output of Daly's metric is a probability-of-detection map, which indicates the areas where the reference and test images differ in a perceptual sense.

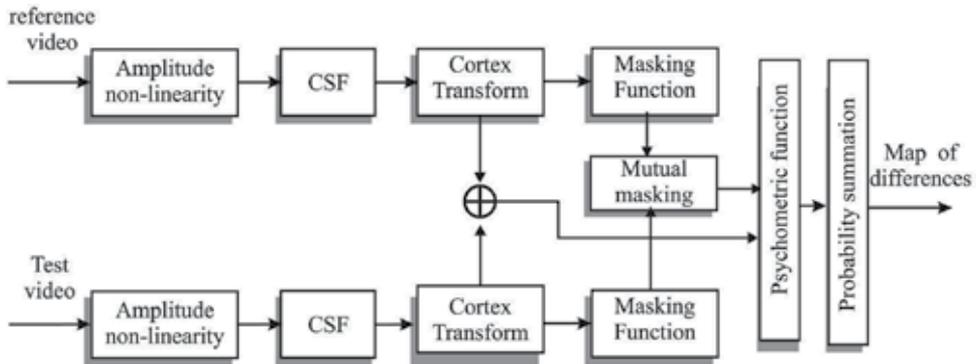


Fig. 10. Block diagram of the visible differences predictor (VDP) (Daly, 1993; 1992).

5.2.2 Sarnoff JND model

The Sarnoff JND model is based on multi-scale spatial vision model proposed by Lubin (Lubin, 1993; 1995). The model takes into account color and temporal variation. Like the metric by Daly, it is designed to predict the probability of detection of artifacts in an image. But, it uses the concept of *just noticeable differences* (JNDs) that are visibility thresholds for changes in images.

The JND unit of measure is defined such that 1 JND corresponds to a 75% chance that an observer viewing the two images detects the difference. JND values above 1 are calculated incrementally. For example, if image A is 1 JND higher than Image B, and image C is 1 JND higher than image A, then image C is 2 JNDs higher than image B. In terms of probability of detection, a 2 JND difference corresponds to 93.75% chance of discrimination, while a 3 JND difference corresponds to 98.44%.

The block diagram of the Sarnoff JND model is depicted in Fig. 11. First, the picture is transformed to the CIE $L^*u^*v^*$ uniform color space (Poynton, 2003). Next, each sequence is filtered and down-sampled using a Gaussian pyramid operation (Burt & Adelson, 1983).

Then, the normalization stage sets the overall gain with a time-dependent average luminance, modelling the HVS insensitivity to overall light level.

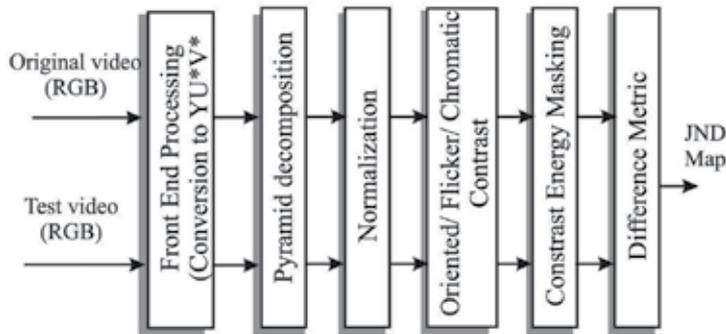


Fig. 11. Block diagram of the Sarnoff JND (Lubin, 1993; 1995).

After normalization, contrast measures are obtained. At each pyramid level, the contrast arrays are calculated by dividing the local difference of the pixel values by the local sum. The result is, then, scaled to be 1 when the image contrast is at the human detection threshold. This gives the definition of 1 JND, which is used on subsequent stages. This scaled contrast arrays are then passed to the contrast energy masking stage in order to desensitize to image “busyness”. Then, test and reference are compared to produce the JND map.

5.2.3 Structural Similarity and Image Quality (SSIM)

The Structural SIMilarity and Image Quality (SSIM) (Wang et al., 2004) is based on the idea that natural images are highly “structured”. In other words, image signals have strong relationships amongst themselves, which carry information about the structures of the objects in the scene.

To estimate the similarity between a test image and the corresponding reference, the SSIM algorithm measures the luminance $l(x,y)$, contrast $c(x,y)$, and structure $s(x,y)$ of the test image y and the corresponding reference image x , using the following expressions:

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \quad (3)$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \quad (4)$$

and

$$s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}, \quad (5)$$

where C_1 , C_2 , and C_3 are small constants given by $C_1 = (K_1 \cdot L)^2$, $C_2 = (K_2 \cdot L)^2$, and $C_3 = C_2/2$. L is the dynamic range of the pixel values (for 8 bits/pixel gray scale images, $L = 255$), $K_1 \ll 1$, and $K_2 \ll 1$.

The general formula of the SSIM metric is given by

$$SSIM(x,y) = [l(x,y)]^\alpha \cdot [c(x,y)]^\beta \cdot [s(x,y)]^\gamma, \quad (6)$$

where α , β , and γ are parameters that define the relative importance of the luminance, contrast, and structure components. If $\alpha = \beta = \gamma = 1$, the above equation is reduced to

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}. \quad (7)$$

The SSIM has a range of values varying between '0' and '1', with '1' being the best value possible. A block diagram of the SSIM algorithm is depicted in Fig. 12. A study of the performance of SSIM has shown that this simple metric presents good results (Sheikh et al., 2006).

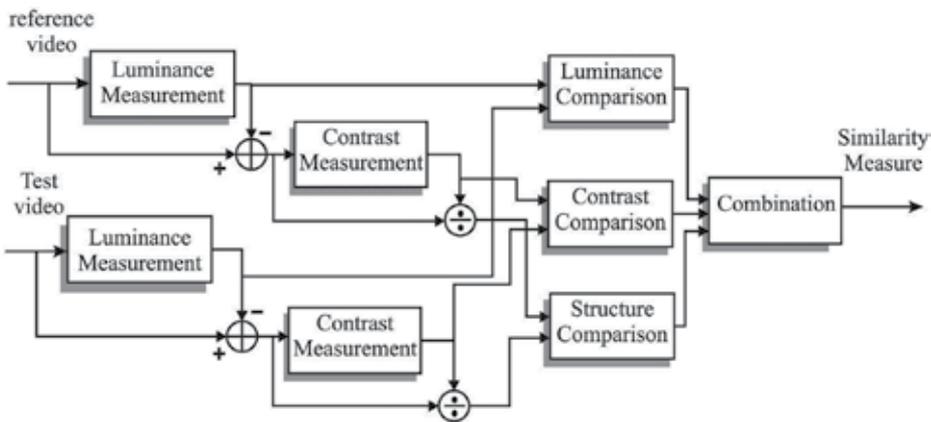


Fig. 12. Block diagram of the SSIM algorithm (Wang et al., 2004).

5.2.4 NTIA Video Quality Metric (VQM)

The video quality metric (VQM) is a metric proposed by Wolf and Pinson from the National Telecommunications and Information Administration (NTIA) (Wolf & Pinson, 1999; Pinson & Wolf, 2004). This metric has recently been adopted by ANSI as a standard for objective video quality. In VQEG Phase II (VQEG, 2003), VQM presented a very good correlation with subjective scores. VQM presented one of the best performances among the competitors.

The algorithm used by VQM includes measurements for the perceptual effects of several video impairments, such as blurring, jerky/unnatural motion, global noise, block distortion, and color distortion. These measurements are combined into a single metric that gives a prediction of the overall quality. The VQM algorithm can be divided into the following stages:

- Calibration – This first stage has the goal of calibrating the video in preparation for the feature extraction stage. With this propose, it estimates and corrects the spatial and temporal shifts, as well as the contrast and brightness offsets, of the processed video sequence with respect to the original video sequence.
- Extraction of quality features – In this stage, the set of quality features that characterizes perceptual changes in the spatial, temporal, and chrominance domains are extracted from spatial-temporal sub-regions of the video sequence. For this, a perceptual filter is

applied to the video to enhance a particular type of property, such as edge information. Features are extracted from spatio-temporal (ST) subregions using a mathematical function and, then, a visibility threshold is applied to these features.

- Estimation of quality parameters - In this stage, a set of quality parameters that describe the perceptual changes is calculated by comparing features extracted from the processed video with those extracted from the reference video.
- Quality estimation - The final step consists of calculating an overall quality metric using a linear combination of parameters calculated in previous stages.

5.3 Reduced reference video quality metrics

Reduced reference (RR) video quality metrics require only partial information about the reference video. To help evaluate the quality of the video, certain features or physical measures are extracted from the reference and transmitted to the receiver as a *side information*. One of the interesting characteristics of RR metrics is the possibility of choosing the amount of side information. In practice, the exact amount of information will be dictated by the characteristics of the side channel that is used to transmit the auxiliary data or, similarly, by the available storage to cache them. Bit rates of the reduced-reference channel can go from zero (for no-reference metrics) to 15 kbps, 80 kbps, or 256 kbps (VQEG, 2009). Metrics in this class may be less accurate than the *full reference* metrics, but they are also less complex, and make real-time implementations more affordable. Nevertheless, synchronization between the original and impaired data is still necessary. Works in this area include the work by Webster *et al.* (Webster et al., 1993), Brètillon *et al.* (Bretillon et al., 2000), Gunawan and Ghanbari (Gunawan & Ghanbari, 2005), and the work by Carnec *et al.* (Carnec et al., 2003). In this section, we describe the RR metrics by Webster *et al.* and Gunawan and Ghanbari.

5.3.1 Objective video quality assessment system based on human perception

One of the earliest *reduced reference* metrics was proposed by Webster *et al.* (Webster et al., 1993). Their metric is a *feature extraction* metric that estimates the amount of impairment in a video by extracting localized spatial and temporal activity features using especially designed filters. The block-diagram of this metric is depicted in Fig. 13.

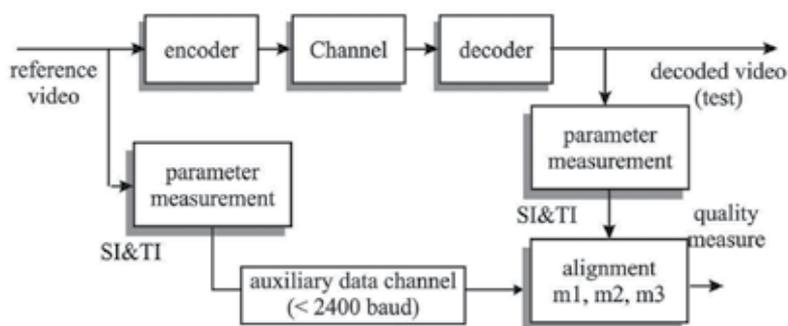


Fig. 13. Block diagram of Webster's algorithm (Webster et al., 1993).

The spatial information (SI) feature corresponds to the the standard deviation of edgenhanced frames, assuming that degradation will modify the edge statistics in the frames. The temporal information (TI) feature corresponds to the standard deviation of

difference frames, i.e., the amount of perceived motion in the video scene. Three comparison metrics are derived from the SI and TI features of the reference and the distorted videos. The metrics for the reference video are transmitted over the RR channel. The size of the RR data depends upon the size of the window over which SI & TI features are calculated.

5.3.2 Local Harmonic Strength (LHS) metric

The work by Gunawan and Ghanbari (Gunawan & Ghanbari, 2005) proposes a RR video quality metric that is based on a *local harmonic strength* (LHS) feature. The harmonic strength can be interpreted as a spatial activity measure, estimated in terms of vertical/horizontal edges of the picture. In summary, the quality measure is based on harmonic gain and loss estimates obtained from a discriminative analysis of the LHS feature computed on gradient images. A simplified block diagram of the LHS RR video quality metric is depicted in Fig. 14.

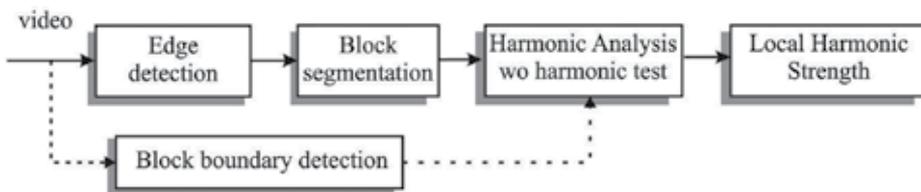


Fig. 14. Block diagram of the LHS algorithm stage of the metric (Gunawan & Ghanbari, 2008; 2005).

The first step of the algorithm is a simple edge-detection stage (3×3 Sobel operator) that generates a gradient image. This resulting image is, then, processed by a non-overlap block segmentation algorithm, with a block size large enough to account for any vertical or horizontal activity within the blocks (typically 32×32 pixels). An optional alignment with the DCT blocks may also be included to increase the precision of the algorithm.

The LHS is calculated as the accumulated strength of the harmonic frequencies of the blocks on the pictures. Since non-overlapped blocks on a picture are also identified by their spatial location, the collected features from all blocks are organized as a matrix. This matrix has a size which is 32×32 smaller than the full resolution picture. LHS matrix features from test and reference pictures are computed separately, and then compared to each other.

The harmonical analysis is performed on the segmented blocks of the gradient image. For this step, a 2-D Fast Fourier Transform (FFT) is applied to each block. The resulting image reveals the appearance of frequency components at certain interval along the two principle axes (horizontal and vertical axes). These frequencies are known as harmonics. The harmonic analysis, then, isolates and accumulates the harmonic components of the resulting FFT spectrum, estimating the value of the LHS feature.

The discriminative analysis is performed after all features from the reference and test images are calculated. Applied to the blocks, the analysis will differentiate between an increase (gain) or decrease (loss) in their strengths, giving an insight on how degradations are distributed over the frame/image. The two quality features produced, namely *harmonic gain* and *harmonic loss*, correspond to blockiness and blurriness on the test image, respectively. To produce a single overall quality measure, spatial collapsing functions (e.g. arithmetic average) are used.

In a more recent algorithm, the author improved the performance of the algorithm by compressing the side information (Gunawan & Ghanbari, 2008), what results in a reduction of the amount of data that needs to be transmitted or stored.

5.4 No-reference video quality metrics

Requiring the reference video or even a small portion of it becomes a serious impediment in many real-time transmission applications. In this case, it becomes essential to develop ways of blindly estimating the quality of a video using a *no-reference* video quality metric. It turns out that, although human observers can usually assess the quality of a video without using the reference, designing a *no-reference* metric is a very difficult task. Considering the difficulties faced by the *full reference* video quality metrics (Eskicioglu & Fisher, 1995; Martens & Meesters, 1997; Rohaly, 2000; VQEG, 2003), this is no surprise.

Except for the metric by Gastaldo *et al.* that uses a neural network (Gastaldo *et al.*, 2001), most of the proposed metrics are *feature extraction* metrics that estimate features of the video. Due to the difficulties encountered in designing NR reference metrics, several metrics rely on one or two features to estimate quality. In most cases, the features used in the algorithms are *artifact signals*, with the most popular being blockiness, blurriness, and ringing. For example, the metrics by Wu *et al.* and Wang *et al.* estimate quality based solely on a blockiness measurement (Wang *et al.*, 2000; Wu & Yuen, 1997; Keimel *et al.*, 2009). The metrics by Farias and Mitra (Farias & Mitra, 2005) and by Caviedes and Jung (Caviedes & Jung, 2001) use four and five artifacts, respectively. In this section, we describe the metrics by Farias and Mitra (Farias & Mitra, 2005) and Oprea *et al.* (Oprea *et al.*, 2009).

5.4.1 No-reference video quality metric based on artifact measurements

As a representation of the metrics based on artifact measurements, we will describe the algorithm proposed by Farias and Mitra (Farias & Mitra, 2005). This approach is based on the assumption that the perceived quality of a video can be affected by a variety of artifacts and that the strengths of these artifacts contribute to the overall annoyance (Ahumada & Null, 1993).

The multidimensional approach requires a good knowledge of the types of artifacts present in digital videos and an extensive study of the most relevant artifacts. The authors performed a series of psychophysical experiments to understand how artifacts depend on the physical properties of the video and how they combine to produce the overall annoyance.

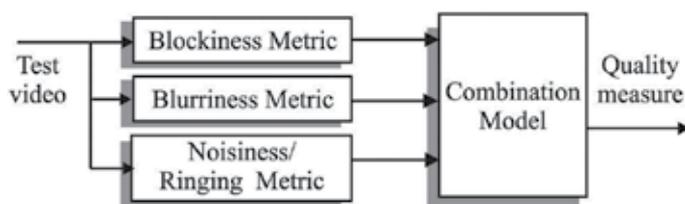


Fig. 15. Block diagram of the *no-reference* metric proposed by Farias and Mitra (Farias & Mitra, 2005).

The block diagram of the metric is as depicted in 15. The algorithm is composed by a set of three artifact metrics (artifact physical strength measurements) for estimating *blockiness*, *blurriness*, *ringing/noisiness*. The metrics are simple enough to be used in real-time applications, as briefly described below.

- The blockiness metric is a modification of the metric by Vlachos (Vlachos, 2000). It estimates the blockiness signal strength by comparing the cross-correlation of pixels inside (intra) and outside (inter) the borders of the coding blocking structure of a frame.

- The blurriness metric is based on the idea that blur makes the edges larger or less sharp (Marziliano et al., 2004; Lu, 2001; Ong et al., 2003). The algorithm measures blurriness by estimating the width of the edges in the frame.
- The noisiness/ringing metric is based on the work by Lee (Lee & Hoppel, 1989) that uses the well known fact that the noise variance of an image can be estimated by the local variance of a flat area. To reduce the content effect a cascade of 1-D filters was used as a pre-processing stage (Olsen, 1993).

To evaluate the performance of each artefact metric, their ability to detect and estimate the artifact signal strength is tested using test sequences containing only the artifact being measured, artifacts other than the artifact being measured, and a combination of all artifacts. The outputs of the individual metrics are also compared to artifact perceptual strengths gathered from psychophysical experiments. A model for overall annoyance is obtained based on a combination of the artifact metrics using a Minkowski metric.

5.4.2 Perceptual video quality assessment based on salient region detection

A recent work by Oprea *et al.* (Oprea et al., 2009) proposes a video quality metric that weighs the distortion measurements on the perceptual importance of the region where it is located.

The first step of this algorithm is to find which are the perceptually important areas of the video frame. For this, the model estimates key features that attract attention: color contrast, object size, orientation, and eccentricity. The measurement of these features will determine which are the important (or salient) areas, producing a saliency map. It is worth pointing out that extracting saliency from video sequences is a complex task because both the spatial extent and dynamic evolution of regions should be considered.

For the detected *salient* areas, a distortion measure is computed using a specialized no-reference metric. The metric considered by the algorithm is a blurriness metric. The blurriness algorithm is based on previous algorithms that estimate the blur by measuring the width of the edges in a frame (Marziliano et al., 2004).

An experiment performed by the authors revealed that the metric has a correlation of about 85% with subjective scores. The algorithm has, nevertheless, limitations concerning the fact that the saliency maps are calculated for the frames individually.

5.5 Metrics using data hiding

An alternative way of implementing a video quality metrics is to use embedding techniques (Sugimoto et al., 1998; Farias, Mitra, Carli & Neri, 2002; Holliman & Young, 2002). The idea consists of embedding in the original video the necessary information to estimate its quality at the time of display. One example of metrics that uses this approach is the work by Farias *et al.* (Farias, Mitra, Carli & Neri, 2002). Fig. 16 depicts the block diagram of this video quality assessment system.

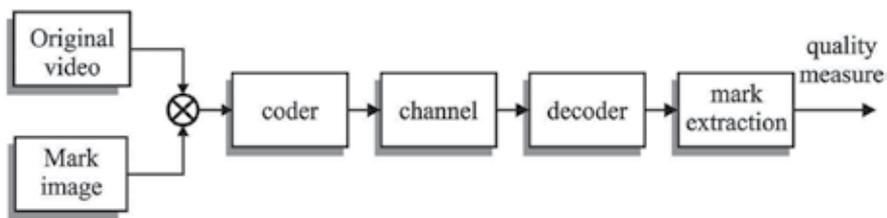


Fig. 16. Block diagram of the video quality assessment system based on data hiding.

At the transmitter, the mark is embedded in each frame of the video using a spread-spectrum technique (Cox et al., 1997). The embedding procedure can be summarized as follows. A pseudo random algorithm is first used to generate pseudo-noise (PN) images $\mathbf{p} = p(i, j, k)$, with values -1 or 1 and zero mean. The final mark to be embedded, \mathbf{w} , is obtained by multiplying the binary image, \mathbf{m} , by the PN image \mathbf{p} . Only one binary image is used for all frames, but the PN images vary from frame to frame.

Then, the logarithm of the luminance of the video frame, \mathbf{y} , is taken and the DCT transform, \mathbf{LY} , is computed. The mark, \mathbf{w} , is multiplied by a scaling factor, α , before being added to the luminance DCT coefficients. After the embedding, the DCT coefficients are given by the following expression:

$$LY'(i, j, k) = \begin{cases} LY(i, j, k) + \alpha \cdot w(i, j, k), & 120 \leq i \leq 240, 120 \leq j \leq 240; \\ LY(i, j, k), & \text{elsewhere.} \end{cases} \quad (8)$$

where i and j are the frequency coordinates and k is the temporal coordinate (index of the frame). For the purpose of assessing the quality of a video, the mark is inserted in the mid-frequencies.

The scaling factor, α , is used to vary the strength of the mark. An increase in the value of α increases the robustness of the mark, but also decreases the quality of the video. After the mark is inserted, the exponential of the video is taken and then the inverse DCT (IDCT). The video is then coded (compressed) and sent over the communication channel.

The process of extracting the mark from the received video is summarized as follows. First, the logarithm of the luminance of the received video, \mathbf{y}'' , is first taken and its DCT calculated. Then, we multiply the mid-frequency DCT coefficients where the mark was inserted by the corresponding pseudo-noise image. Considering that $p(i, j, k) \cdot p(i, j, k) = 1$ because $p(i, j, k)$ is either -1 or +1, we obtain:

$$LY''(i, j, k) \cdot p(i, j, k) = LY(i, j, k) \cdot p(i, j, k) + \alpha \cdot m(i, j), \quad (9)$$

The result of Eq.(9) is then averaged for a chosen number of frames N_f to eliminate the noise (PN signal) introduced by the spread spectrum embedding algorithm. The extracted binary mark is obtained by taking the sign of this average, as given by the following expression:

$$m_r(i, j) = \text{sgn} \left(\frac{1}{N_f} \sum_{k=1}^{N_f} LY(i, j, k) \cdot p(i, j, k) + \alpha \cdot m(i, j, k) \right), \quad (10)$$

Since the PN matrix has zero-mean, the sum $\sum_{k=1}^{N_f} LY(i, j, k) \cdot p(i, j, k)$ approaches zero for a large value of N_f . In general, for $N_f \geq 10$ the mark is recovered perfectly, i.e., $\mathbf{m}_r = \mathbf{m}$. When errors are added by compression or transmission, $Y'' = Y' + \eta$ and the extracted mark \mathbf{m}_r is an approximation of \mathbf{m} . A measure of the degradation of the mark is given by the Total Square Error (TSE) between the extracted mark \mathbf{m}_r and the original binary image:

$$E_{tse} = \sum_i \sum_j [m(i, j) - m_r(i, j)]^2. \quad (11)$$

The less the amount of errors caused by processing, compression or transmission, the smaller E_{tse} is. On the other hand, the more degraded the video, the higher E_{tse} is. Therefore, the measure given by E_{tse} can be used as an estimate of the degradation of the host video.

6. The work of the video quality experts group

The Video Quality Experts Group (VQEG) was formed in October 1997 in Turin, Italy, to address video quality issues. Since then, it has been conducting formal evaluations of video quality metrics on common test material.

The first task of VQEG was to perform a validation of full reference video quality metrics targeted at TV applications (FR-TV). VQEG outlined, designed and executed a test program to compare subjective video quality evaluations to the predictions of a number of proposed objective metrics (VQEG, 1999). The result of the test was inconclusive (VQEG, 2000). From this first phase, a database of test sequences and their corresponding subjective rating was made available publicly.

In 2003, the second phase of the FR-TV test was completed (VQEG, 2003). The results of these tests have become part of two ITU recommendations (ITU-T, 2004b; a). Contrary to what happened in the first phase, this time the best metrics reached a correlation of around 94% with the subjective scores. The PSNR had a performance of around 70%.

After concluding the FR-TV tests, VQEG conducted a round of tests to evaluate video quality metrics targeted at multimedia applications. The videos considered for this phase had lower bitrates and smaller frames sizes. Besides that, a larger number of codecs and transmission conditions were considered. But, at this first phase the audio signal was not tested. On the 19th of September 2008, the Final Report of VQEGs Multimedia Phase I was released (VQEG, 2008). The correlation results for the submitted FR, RR, and NR metrics were of about 80%, 78%, and 56%, respectively. PSNR had a correlation of around 65%. VQEG has already started working on the second phase of the multimedia tests. In the second phase, both audio and video signals will be tested (simultaneously).

VQEG is currently finishing the tests for evaluation of reduced- and no-reference video quality metrics for television applications (“RR/NR-TV”). The draft report is currently available and the final report should be made available by the time of this publication. VQEG is also working on the evaluation of metrics to be used with High Definition TV (HDTV) content.

One more recent development of VQEG is the test for “hybrid metrics”. These metrics estimate quality by looking at not only the decoded video, but also the encoded bitstream. These metrics are targeted at applications like broadcasting, lab applications, live monitoring in network (bitstream) or at end-user (hybrid no reference).

In June 2009, a new activity of VQEG has been launched. It’s named Joint Effort Group (JEG) and it consists of an alternative collaborative action. The idea is to work jointly on both mandatory actions to validate metrics (subjective dataset completion and metrics design).

7. Conclusions and new perspectives

In this chapter, we introduced several aspects of video quality. We described the anatomy of the Human Visual System (HVS), discussing a number of phenomena of visual perception that are of particular relevance to video quality. We briefly introduced the main characteristics of modern digital video systems, focusing on the errors (artifacts) commonly present in digital video applications.

We described both objective and subjective methods for assessing video quality. For the subjective methods, we discussed the most common techniques standardized by ITU. We also discussed the main ideas used in the design of objective metric algorithms, listing a

representative set of video quality metrics (FR, RR, and NR). Finally, we discussed the work of the Video Quality Experts Group (VQEG).

It is worth pointing out that, although much has been done in the last ten years in the area of video quality metrics, there are still a lot of challenges to be solved. Most of the achievements have been in the development of full-reference video quality metrics that evaluate compression artifacts. Much remains to be done, for example, in the area of no-reference and reduced-reference quality metrics.

Also, there has been a growing interest in metrics that estimate the quality of video digitally transmitted over wired or wireless channels/networks. This is due to the popularity of video delivery services over IP networks (e.g. internet streaming and IPTV) or wireless channel (e.g. mobile TV). Another area that has attracted attention is the area of multimedia quality. So far, very few metrics have addressed the issue of simultaneously measuring the quality of all medias involved (e.g. video, audio, text). There is also been an interest in 3D video and HDTV applications.

A new trend in video quality design is the development of *hybrid metrics*, which are metrics that use a combination of packet information, bitstream or decoded video as input (Verscheure et al., 1999; Kanumuri, Cosman, Reibman & Vaishampayan, 2006; Kanumuri, Subramanian, Cosman & Reibman, 2006). The idea here is to also consider parameters extracted from the transport stream and the bitstream (without decoding) in the computation of quality estimation. The main advantage of this approach is the lower bandwidth and processing requirements, when compared to metrics that only consider the fully decoded video.

8. References

- Ahumada, A. J., J. & Null, C. (1993). Image quality: a multidimensional problem, *Digital Images and Human Vision* pp. 141–148.
- Algazi, V. & Hiwasa, N. (1993). Perceptual criteria and design alternatives for low bit rate video coding, *Proc. 27th Asilomar Conf. on Signals, Systems and Computers*, Vol. 2, Pacific Grove, California, USA, pp. 831–835.
- Avcibas, I., Avcba, I., Sankur, B. & Sayood, K. (2002). Statistical evaluation of image quality measures, *Journal of Electronic Imaging* 11: 206–223.
- Bosi, M. & Goldberg, R. E. (2002). *Introduction to Digital Audio Coding and Standards*, Springer International Series in Engineering and Computer Science.
- Bretillon, P., Montard, N., Baina, J. & Goudezeune, G. (2000). Quality meter and digital television applications, *Proc. SPIE Conference on Visual Communications and Image Processing*, Vol. 4067, Perth, WA, Australia, pp. 780–90.
- Burt, P. J. & Adelson, E. H. (1983). The laplacian pyramid as a compact image code, *IEEE Transactions on Communications* COM-31, 4: 532–540.
- Carnec, M., Le Callet, P. & Barba, D. (2003). New perceptual quality assessment method with reduced reference for compressed images, *Proc. SPIE Conference on Visual Communications and Image Processing*, Vol. 5150, Lugano, Switzerland, pp. 1582–93.
- Caviedes, J. & Jung, J. (2001). No-reference metric for a video quality control loop, *Proc. Int. Conf. on Information Systems, Analysis and Synthesis*, Vol. 13.
- Cox, I., Kilian, J., Leighton, F. & Shamoon, T. (1997). Secure spread spectrum watermarking for multimedia, *IEEE Trans. on Image Processing* 6(12).

- Daly, S. (1992). The visible difference predictor: An algorithm for the assessment of image fidelity, *Proc. SPIE Conference on Human Vision and Electronic Imaging XII*, p. 2.
- Daly, S. (1993). The visible differences predictor: an algorithm for the assessment of image fidelity, in A. B. Watson (ed.), *Digital Images and Human Vision*, MIT Press, Cambridge, Massachusetts, pp. 179–206.
- de Ridder, H. (1992). Minkowski-metrics as a combination rule for digital-image-coding impairments, *Proc. SPIE Conference on Human Vision, Visual Processing and Digital Display III*, Vol. 1666, San Jose, CA, USA, pp. 16–26.
- de Ridder, H. (2001). Cognitive issues in image quality measurement, *Electronic Imaging* 10(1): 47–55.
- Eckert, M. & Bradley, A. (1998). Perceptual quality metrics applied to still image compression, *Signal Processing* 70: 177–200.
- Eskicioglu, E. & Fisher, P. (1995). Image quality measures and their performance, *IEEE Trans. Image Processing* 43(12): 2959–2965.
- Farias, M., Foley, J. & Mitra, S. (2003a). Perceptual contributions of blocky, blurry and noisy artifacts to overall annoyance, *Proc. IEEE International Conference on Multimedia & Expo*, Vol. 1, Baltimore, MD, USA, pp. 529–532.
- Farias, M., Foley, J. & Mitra, S. (2003b). Some properties of synthetic blocky and blurry artifacts, *Proc. SPIE Conference on Human Vision and Electronic Imaging*, Vol. 5007, Santa Clara, CA, USA, pp. 128–136.
- Farias, M., Foley, J. & Mitra, S. (2004). Detectability and annoyance of synthetic blurring and ringing in video sequences, *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, IEEE, Montreal, Canada.
- Farias, M. & Mitra, S. (2005). No-reference video quality metric based on artifact measurements, *Proc. IEEE Intl. Conf. on Image Processing*, pp. III: 141–144.
- Farias, M., Mitra, S., Carli, M. & Neri, A. (2002). A comparison between an objective quality measure and the mean annoyance values of watermarked videos, *Proc. IEEE Intl. Conf. on Image Processing*, Vol. 3, Rochester, NY, pp. 469–472.
- Farias, M., Moore, M., Foley, J. & Mitra, S. (2002). Detectability and annoyance of synthetic blocky and blurry artifacts, *Proc. SID International Symposium*, Vol. XXXIII, Number II, Boston, MA, USA, pp. 708–712.
- Farias, M., Moore, M., Foley, J. & Mitra, S. (2004). Perceptual contributions of blocky, blurry, and fuzzy impairments to overall annoyance, *Proc. SPIE Conference on Human Vision and Electronic Imaging*, San Jose, CA, USA.
- Gastaldo, P., Rovetta, S. & Zunino, R. (2001). Objective assessment of MPEG video quality: a neural-network approach, *Proc. International Joint Conference on Neural Networks*, Vol. 2, pp. 1432–1437.
- Girod, B. (1993). What's wrong with mean-squared error?, in A. B. Watson (ed.), *Digital Images and Human Vision*, MIT Press, Cambridge, Massachusetts, pp. 207–220.
- Grey, H. (1918). Anatomy of the human body. All images are online and are public domain. URL: <http://www.bartley.com/107/>
- Gunawan, I. & Ghanbari, M. (2005). Image quality assessment based on harmonics gain/loss information, *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, Vol. 1, pp. I-429–32.
- Gunawan, I. & Ghanbari, M. (2008). Efficient reduced-reference video quality meter, *Broadcasting, IEEE Transactions on* 54(3): 669–679.

- Haskell, B. G., Puri, A. & Netravali, A. N. (1997). *Digital video: An Introduction to MPEG-2, Digital multimedia standards*, Chapman & Hall: International Thomson Pub., New York, NY, USA.
- Hays, W. (1981). *Statistics for the social sciences*, 3 edn, LLH Technology Publishing, Madison Avenue, New York, N.Y.
- Holliman, M. & Young, M. (2002). Watermarking for automatic quality monitoring, *Proc. SPIE Conference on Security and Watermarking of Multimedia Contents*, Vol. 4675, San Jose, CA, USA.
- ITU (1998). *ISO/IEC 13818: Generic coding of moving pictures and associated audio (MPEG-2)*.
- ITU (2003). *Recommendation ITU-T H.264: Advanced Video Coding for Generic Audiovisual Services*.
- Kanumuri, S., Cosman, P., Reibman, A. & Vaishampayan, V. (2006). Modeling packet-loss visibility in mpeg-2 video, *Multimedia, IEEE Transactions on* 8(2): 341–355.
- Kanumuri, S., Subramanian, S., Cosman, P. & Reibman, A. (2006). Predicting h.264 packet loss visibility using a generalized linear model, *Image Processing, 2006 IEEE International Conference on*, pp. 2245–2248.
- Keimel, C., Oelbaum, T. & Diepold, K. (2009). No-reference video quality evaluation for highdefinition video, *Acoustics, Speech, and Signal Processing, IEEE International Conference on* pp. 1145–1148.
- Klein, S. (1993). Image quality and image compression: A psychophysicist's viewpoint, in A. B. Watson (ed.), *Digital Images and Human Vision*, MIT Press, Cambridge, Massachusetts, pp. 73–88.
- Lambert, P., de Neve, W., de Neve, P., Moerman, I., Demeester, P. & de Walle, R. V. (Jan. 2006). Rate-distortion performance of h.264/avc compared to state-of-the-art video codecs, *Circuits and Systems for Video Technology, IEEE Transactions on* 16(1): 134–140.
- Lee, J. & Hoppel, K. (1989). Noise modeling and estimation of remotely-sensed images, *Proc. International Geoscience and Remote Sensing*, Vol. 2, Vancouver, Canada, pp. 1005–1008.
- Lu, J. (2001). Image analysis for video artifact estimation and measurement, *Proc. SPIE Conference on Machine Vision Applications in Industrial Inspection IX*, Vol. 4301, San Jose, CA, USA, pp. 166–174.
- Lubin, J. (1993). The use of psychophysical data and models in the analysis of display system performance, in A. B. Watson (ed.), *Digital Images and Human Vision*, MIT Press, Cambridge, Massachusetts, pp. 163–178.
- Lubin, J. (1995). A visual discrimination model for imaging system design and evaluation, in E. Peli (ed.), *Vision models for target detection and recognition*, World Scientific Publishing, Singapore.
- Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, Freeman.
- Martens, J. & Meesters, L. (1997). Image dissimilarity, *Signal Processing* 70: 1164–1175.
- Marziliano, P., Dufaux, F., Winkler, S. & Ebrahimi, T. (2004). Perceptual blur and ringing metrics: Application to JPEG2000, *Signal Processing: Image Communication* 19(2): 163–172.
- Maxwell, S. E. & Delaney, H. D. (2003). *Designing experiments and analyzing data: A model comparison perspective*, Lawrence Erlbaum Associates, Mahwah, NJ.

- ITU-R (1998). *Recommendation BT.500-8: Methodology for subjective assessment of the quality of television pictures.*
- ITU-T (1999). *Recommendation P.910: Subjective Video Quality Assessment Methods for Multimedia Applications.*
- ITU-T (2004a). *Objective perceptual video quality measurement techniques for standard definition digital broadcast television in the presence of a full reference.*
- ITU-T (2004b). *Recommendation J.144: Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference.*
- Moore, M. S. (2002). *Psychophysical Measurement and Prediction of Digital Video Quality*, PhD thesis, University of California Santa Barbara.
- Olsen, S. I. (1993). Estimation of noise in images: an evaluation, *CVGIP-Graphical Models & Image Processing* 55(4): 319–23.
- Ong, E.-P., Lin, W., Lu, Z., Yao, S., Yang, X. & Jinag, L. (2003). No-reference JPEG2000, *Proc. IEEE International Conference on Multimedia and Expo*, Vol. 1, Baltimore, USA, pp. 545–548.
- Oprea, C., Pirnogi, I., Paleologu, C. & Udrea, M. (2009). Perceptual video quality assessment based on salient region detection, *Telecommunications, 2009. AICT '09. Fifth Advanced International Conference on*, pp. 232–236.
- Pessoa, A. C. F., Falcao, A. X., Silva, A., Nishihara, R. M. & Lotufo, R. A. (1998). Video quality assessment using objective parameters based on image segmentation, *Proc. SBT/IEEE International Telecommunications Symposium*, Vol. 2, Sao Paulo, Brazil, pp. 498–503.
- Pinson, M. & Wolf, S. (2004). A new standardized method for objectively measuring video quality, *Broadcasting, IEEE Transactions on* 50(3): 312–322.
- Poynton, C. (2003). *Digital Video and HDTV - Algorithms and Interfaces*, 5th edn, Morgan Kaufmann.
- Richardson, I. E. (2003). *H.264 and MPEG-4 Video Compression*, John Wiley & Sons, New York, NY, USA.
- Rohaly, A. M. al., e. (2000). Final report from the video quality experts group on the validation of objective models of video quality assessment, *Technical report*, Video Quality Experts Group.
- Sheikh, H., Sabir, M. & Bovik, A. (2006). A statistical evaluation of recent full reference image quality assessment algorithms, *Image Processing, IEEE Transactions on* 15(11): 3440–3451.
- Snedecor, G. W. & Cochran, W. G. (1989). *Statistical methods*, 8th edn, Iowa State University Press,
- Ames. Sugimoto, O., Kawada, R., Wada, M. & Matsumoto, S. (1998). Objective measurement scheme for perceived picture quality degradation caused by MPEG encoding without any reference pictures, *Proc. SPIE Conference on Human Vision and Electronic Imaging*, Vol. 4310, San Jose, CA, USA, pp. 932–939.
- Teo, P. C. & Heeger, D. J. (1994). Perceptual image distortion, *Proc. IEEE International Conference on Image Processing*, Vol. 2, Austin, TX, USA, pp. 982–986.
- van den Branden Lambrecht, C. J. & Kunt, M. (1998). Characterization of human visual sensitivity for video imaging applications, *Signal Processing* 67(3): 255–69.
- Verscheure, O., Frossard, P. & Hamdi, M. (1999). User-oriented qos analysis in mpeg-2 video delivery, *Real-Time Imaging* 5(5): 305–314.

- Vlachos, T. (2000). Detection of blocking artifacts in compressed video, *Electronics Letters* 36(13): 1106–1108.
- VQEG (1999). *VQEG subjective test plan (Phase 1)*. <ftp://ftp.crc.ca/crc/vqeg/phase1-docs>.
- VQEG (2000). Final report from the video quality experts group on the validation of objective models of video quality assessment, *Technical report*, <http://www.vqeg.org>.
- VQEG (2003). Final report from the video quality experts group on the validation of objective models of video quality assessment - Phase II, *Technical report*, <http://ftp.crc.ca/test/pub/crc/vqeg/>.
- VQEG (2008). Final report of vqegs multimedia phase i validation test, *Technical report*, <http://www.vqeg.org>.
- VQEG (2009). Rnr-tv group - test plan draft version 2, *Technical report*, <http://www.vqeg.org>.
- Wang, Y. (2006). Survey of objective video quality measurements, *Technical Report T1A1.5/96- 110*, Worcester Polytechnic Institute.
- Wang, Z., Bovik, A. C., Sheikh, H. R., Member, S., Simoncelli, E. P. & Member, S. (2004). Image quality assessment: From error visibility to structural similarity, *IEEE Transactions on Image Processing* 13: 600–612.
- Wang, Z., Bovik, A. & Evan, B. (2000). Blind measurement of blocking artifacts in images, *Proc. IEEE International Conference on Image Processing*, Vol. 3, pp. 981–984.
- Watson, A. (1987). The cortex transform: rapid computation of simulated neural images, *Computer Vision, Graphics, and Image Processing* 39(3): 311–327.
- Watson, A. B. (1990). Perceptual-components architecture for digital video, *Journal of the Optical Society of America, A-Optics & Image Science* 7(10): 1943–54.
- Watson, A. B. (1998). Towards a visual quality metric for digital video, *Proc. European Signal Processing Conference*, Vol. 2, Island of Rhodes, Greece.
- Watson, A. B., James, H. & McGowan, J. F. (2001). Digital video quality metric based on human vision, *Journal of Electronic Imaging* 10(1): 20–9.
- Webster, A. A., Jones, C. T., Pinson, M. H., Voran, S. D. & Wolf, S. (1993). An objective video quality assessment system based on human perception, *Proc. SPIE Conference on Human Vision, Visual Processing, and Digital Display IV*, Vol. 1913, San Jose, CA, USA, pp. 15–26.
- Winkler, S. (1999). A perceptual distortion metric for digital color video, *Proc. SPIE Conference on Human Vision and Electronic Imaging*, Vol. 3644, San Jose, CA, USA, pp. 175–184.
- Wolf, S. & Pinson, M. H. (1999). Spatial-temporal distortion metric for in-service quality monitoring of any digital video system, *Proc. SPIE Conference on Multimedia Systems and Applications II*, Vol. 3845, Boston, MA, USA, pp. 266–77.
- Wolf, S., Pinson, M. H., Voran, S. D. & Webster, A. A. (1991). Objective quality assessment of digitally transmitted video, *Proc. IEEE Pacific Rim Conference on Communications, Computers and Signal Processing*, Victoria, BC, Canada, pp. 477–82 vol.
- Wu, H. & Yuen, M. (1997). A generalized block-edge impairment metric for video coding, *IEEE Signal Processing Letters* 4(11): 317–320.
- Yuen, M. & Wu, H. R. (1998). A survey of hybrid MC/DPCM/DCT video coding distortions, *Signal Processing* 70(3): 247–78.

Video Analysis and Indexing

Hui Ding, Wei Pan and Yong Guan
Capital Normal University
China

1. Introduction

In recent years, hardware technologies and standards activities have matured to the point that it is becoming feasible to transmit, store, process, and view video signals that are stored in digital formats, and to share video signals between different platforms and application areas. Manual annotation of video contents is a tedious, time consuming, subjective, inaccurate, incomplete, and - perhaps more importantly - costly process. Over the past decade, a growing number of researchers have been attempting to fulfil the need for creative algorithms and systems that allow (semi-) automatic ways to describe, organize, and manage video data with greater understanding of its semantic contents.

In this chapter, we study methods for automatic segmentation in digital videos. We also demonstrate their suitability for indexing and retrieval. The first section is related to granularity and addresses the question: what to index, and reviews existing techniques in video analysis and indexing. The modalities and their analysis are presented in section 2. Pre-processing, feature extraction and representation are discussed in Section 3. We discuss the actual process of developing models for semantic concepts in Section 4. Finally, directions for future research and conclusions are presented in Section 5.

1.1 What is a video database?

Multimedia data, such as text, audio, images and video, is rapidly evolving as the main form for the creation, exchange, and storage of information in the modern era. Numerous types of videos are created in the world. Conservative estimates state that there are more than 6 million hours of video already stored and this number grows at a rate of about 10 percent a year (Hjelsvold, 1995). Projections estimate that by the end of 2010, 50 percent of the total digital data stored worldwide will be video and rich media (Brown, 2001). The amount of video information stored in archives worldwide is huge. Because the amount of video we all must manage is growing at an exponential rate, it makes the creation of the video databases all the more essential. And consequently, the design and implementation of video database systems has become a major topic of interest.

What is a Video Database? The video database is a storehouse of information on the various aspects related to the videos. The video database is one of the most accessed types of databases. One way to think about a digital video system is in the context of a database. Such a database contains a large collection of information elements of a certain granularity. These elements are described according to a range of attributes and can be accessed according to the intent of the user.

The combination of the growing number of applications for video-intensive products and solutions - from personal video recorders to multimedia collaborative systems - with the many technical challenges behind the design of contemporary video database systems. Progress in visual information analysis has been fostered by many research fields (Fig. 1), particularly: (text-based) information retrieval, image processing and computer vision, pattern recognition, multimedia database organization, multidimensional indexing, data mining, machine learning, and visualization, psychological modeling of user behavior, man-machine interaction, among many others.

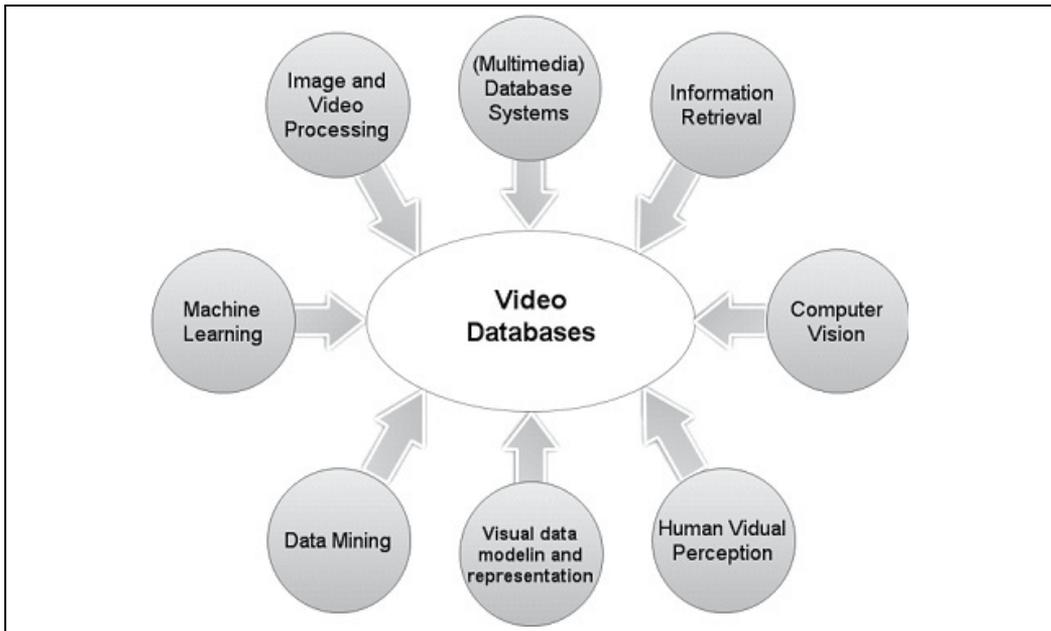


Fig. 1. Visual information retrieval blends together many research disciplines

1.2 A short overview of related work

However, raw video data by itself has limited usefulness, since it takes far too long to search for the desired piece of information within a videotape repository or a digital video archive. Attempts to improve the efficiency of the search process by adding extra data (henceforth called metadata) to the video contents do little more than transferring the burden of performing inefficient, tedious, and time-consuming tasks to the cataloguing stage. The challenging goal is to devise better ways to automatically store, catalog, and retrieve video information with greater understanding of its contents. Researchers from various disciplines have acknowledged such challenge and provided a vast number of algorithms, systems, and papers on this topic during recent years.

As in conventional information retrieval, the purpose of a Visual Information Retrieval (VIR) system is to retrieve all the images (or image sequences, video) that are relevant to a user query while retrieving as few non-relevant images as possible. The emphasis is on the retrieval of information as opposed to the retrieval of data. Similarly to its text-based counterpart a visual information retrieval system must be able to interpret the contents of the documents (images) in a collection and rank them according to a degree of relevance to the user query.

The interpretation process involves extracting (semantic) information from the documents (images) and using this information to match the user needs (Baeza-Yates & Ribeiro-Neto, 1999). VIR systems can be classified in three main generations, according to the attributes used to search and retrieve a desired image or video file:

- First-generation VIR systems: Use query by text, allowing queries such as “all pictures of red Ferraris” or “all images of Van Gogh’s paintings”. They rely strongly on metadata, which can be represented either by alphanumeric strings, keywords, or full scripts.
- Second-generation (CB)VIR systems: Support query by content, where the notion of content, for still images, includes, in increasing level of complexity: perceptual properties (e.g., color, shape, texture), semantic primitives (abstractions such as objects, roles, and scenes), and subjective attributes such as impressions, emotions and meaning associated to the perceptual properties. Many second-generation systems use content-based techniques as a complementary component, rather than a replacement, of text-based tools.
- Third-generation (SB)VIR systems (Zhang, 2006): The semantic gap has dictated that solutions to image and video indexing could only be applied in narrow domains using specific concept detectors. When recent advances in data-driven image and video analysis on the one hand, and top-down ontology engineering and reasoning on the other hand, are combined we are reaching the point where the semantic gap can be bridged for many concepts in broad domains such as film, news, sports and personal albums. The research in these disciplines has traditionally been within different communities.

2. Techniques of video analysis

In order to organize the video database, a segmentation algorithm is applied to the video data to obtain video intervals. Parsing, which segments the video stream into generic clips? These clips are the elemental index units in the database. Ideally, the system decomposes individual images into semantic primitives. On the basis of these primitives, a video clip can be indexed with a semantic description using existing knowledge-representation techniques.

2.2 Architecture for video data

Generally, data represents the facts or observations on any phenomenon that is worth formulating and recording. Even though there can be many data structures for the same application, there are some fundamental features that the structure should reflect. Unfortunately, current video characterization techniques rely on image representations based on low-level visual primitives (such as color, texture, and motion), while practical and computationally efficient, and fails to capture most of the structure that is relevant for the perceptual decoding of the video.

Video metadata are data used for the description of video data, including the attributes and the structure of videos, video content and relationships that exist within a video, among videos and between videos and real world objects. A key aspect for the definition of a video metadata model is the imposed video structure. Video data are often represented either as a set of still images that contain salient objects or as clips that have specific spatial (e.g. color, position etc.) or temporal (e.g. motion) features or are related to semantic objects (Li & Özsu,

1997) (Dağtas et al., 2000) (Al-Khatib et al., 1999). More sophisticated approaches are either a hierarchical representation of video objects (Analyti & Christodoulakis, 1995) (Yeo & Yeung, 1997) (Kyriakaki, 2000) based on their structure, or an event-based approach that represents a video object as a set of (non-contiguous, even overlapping) video segments called strata or temporal cohesions (Hacid et al, 2000) that correspond to individual events.

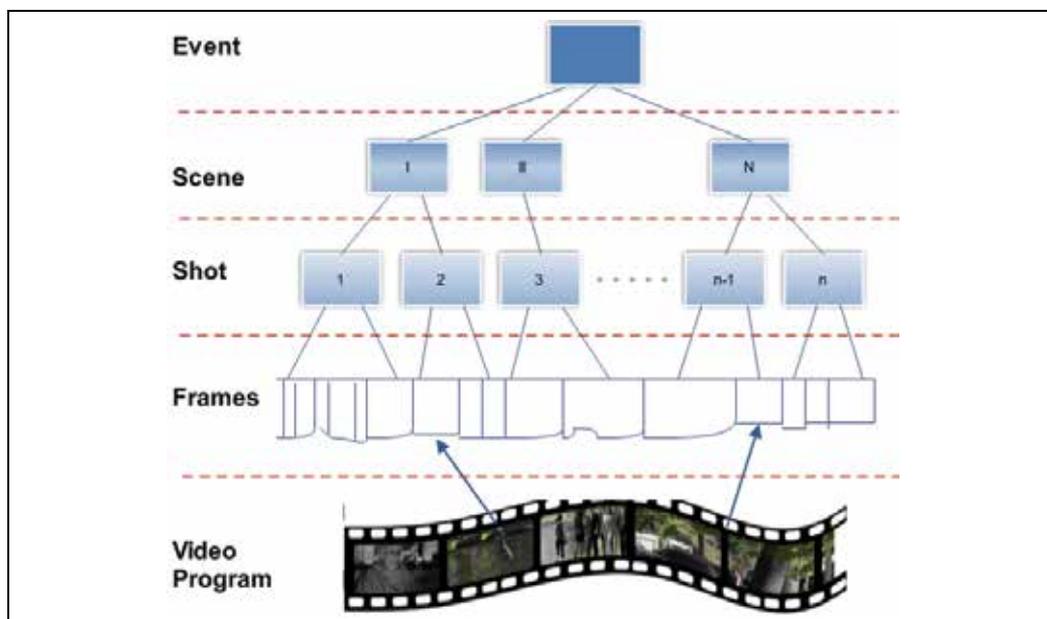


Fig. 2. General structure of a video data

The imposed structure of a video is shown in Fig. 2: A video is represented as an instance of the (Video) Program class and is comprised of a set of Stories. Each story is a logical section of the video object and is further divided in a set of Scenes. A scene represents an event and may be either composite or simple: a simple scene contains a simple event and is comprised of video Shots, while a composite one contains a composite event and is comprised of other scenes. Shots are sets of “similar” consequent frames that are usually recognized using automatic segmentation techniques.

2.3 Temporal segmentation of digital video

Following Fig. 2, temporal video segmentation is the essential first step towards automatic analysis of digital video sequences. Its goal is to divide the video stream into a set of meaningful and manageable segments (shots). Shots are considered to be the primitives for higher level content analysis. A video shot is defined as a series of interrelated consecutive frames taken contiguously by a single camera and representing a continuous action in time and space.

According to whether the transition between shots is abrupt or not, the shot boundaries are categorized into two types: cuts and gradual transitions. Abrupt transitions (cuts) are simpler. In the case of shot cuts, the content change is usually large and easier to detect than the content change during a gradual transition (Lienhart, 1999). Many metrics and the classification algorithms have been proposed in the literature during the past decade, e.g.,

starting with the initial work of (Zhang et al., 1993) (Hampapur et al., 1994) to some recent work of (Smeaton, 2007), (Teng & Tan, 2008). In the following paragraphs of this section, we shall briefly review the metrics used and the classification strategy adopted.

2.3.1 Pixels difference metrics

Metrics classified as pixels difference metrics (PDM) are based on the intensity variations of pixels in equal position in consecutive frames. Temporal segmentation techniques using interframe differences based on color are conceptually similar, but they are not very popular as they have a greater computational burden and almost the same accuracy of their intensity based counterpart.

A basic PDM is the sum of the absolute differences of intensity of the pixels of two consecutive frames (Ford et al., 1997). In particular, indicating with $Y(x, y, j)$ and $Y(x, y, k)$ the intensity of the pixels at position (x, y) and frames j and k , the metric can be expressed in the following way:

$$\Delta f = \sum_x \sum_y |Y(x, y, j) - Y(x, y, k)| \quad (1)$$

where the summation is taken all over the frame.

The same authors propose other metrics based on first and second order statistical moments of the distributions of the levels of intensity of the pixels. Indicating with μ_k and σ_k respectively the values of mean and standard deviation of the intensity of the pixels of the frame k , it is possible to define the following interframe metric between the frames j and k :

$$\lambda = \frac{\left[\frac{\sigma_j + \sigma_k}{2} + \left(\frac{\mu_j - \mu_k}{2} \right)^2 \right]^2}{\sigma_j \sigma_k} \quad (2)$$

This metric has been also used in (Dugad et al., 1998), (Sethi & Patel, 1995), and is usually named likelihood ratio, assuming a uniform second order statistic.

2.3.2 Histogram difference metrics

Metrics classified as histograms difference metrics (HDM) are based on the evaluation of the histograms of one or more channels of the adopted color space. As it is well known, the histogram of a digital image is a measure that supplies information on the general appearance of the image. With reference to an image represented by three color components, each quantized with 8 bit/pixel, a three-dimensional histogram or three one-dimensional histograms can be defined. Although the histogram does not contain any information on the spatial distribution of intensity, the use of interframe metrics based on image histograms is very popular because it represents a good compromise between the computational complexity and the ability to represent the image content.

In recent years several histogram-based techniques have been proposed (Gargi et al., 2000), (Dailianas et al., 1995); some of them are based only on the luminance channel, others on the conversion of 3-D or 2-D histograms to linear histograms (Ardizzone & Cascia, 1997). In what follows some of the most popular HDM metrics are reviewed. In all the equations M and N are respectively the width and height (in pixels) of the image, j and k are the frame

indices, L is the number of intensity levels and $H[j, i]$ the value of the histogram for the i -th intensity level at frame j . A commonly used metric is the bin-to-bin difference, defined as the sum of the absolute differences between histogram values computed for the two frames:

$$f_{db2b}(j, k) = \frac{1}{2MN} \sum_{i=0}^{L-1} |H(j, i) - H(k, i)| \quad (3)$$

The metric can easily be extended to the case of color images, computing the difference separately for every color component and weighting the results. For example, for a RGB representation we have:

$$f_{db2b}(j, k) = \frac{r}{s} f_{db2b}(j, k)^{(red)} + \frac{g}{s} f_{db2b}(j, k)^{(green)} + \frac{b}{s} f_{db2b}(j, k)^{(blue)} \quad (4)$$

where r , g and b are the average values of the three channels and s is:

$$s = \frac{r + g + b}{3} \quad (5)$$

Another metric is called intersection difference and is defined in the following way:

$$f_{dint}(j, k) = 1 - \frac{1}{MN} \sum_{i=0}^{L-1} \min[H(j, i), H(k, i)] \quad (6)$$

In other approaches [28], the chi-square test has been used, which is generally accepted as a test useful to detect if two binned distributions are generated from the some source:

$$f_{dchi2}(j, k) = \sum_{i=0}^{L-1} \frac{[H(j, i) - H(k, i)]^2}{[H(j, i) + H(k, i)]^2} \quad (7)$$

Also the correlation between histograms is used:

$$f_{dcorr}(j, k) = 1 - \frac{\text{cov}(j, k)}{\sigma_j \sigma_k} \quad (8)$$

where $\text{cov}(j, k)$ is the covariance between frame histograms:

$$\text{cov}(j, k) = \frac{1}{L} \sum_{i=0}^{L-1} [H(j, i) - \mu_j][H(k, i) - \mu_k] \quad (9)$$

and μ_j and σ_j represent the mean and the standard deviation, respectively, of the histogram of the frame j :

$$\mu_j = \frac{1}{L} \sum_{i=0}^{L-1} H(j, i); \quad \sigma_j = \sqrt{\frac{1}{L} \sum_{i=0}^{L-1} [H(j, i) - \mu_j]^2} \quad (10)$$

All the metrics discussed so far are global, i.e., based on the histogram computed over the entire frame. Both PDM and HDM techniques are based on the computation of a similarity

measure of two subsequent frames and on comparison of this measure with a threshold. The choice of the threshold is critical, as too low threshold values may lead to false detections and too high threshold values may cause the opposite effect of missed transitions. To limit the problem of threshold selection, several techniques have been proposed.

It has been pointed out that the aim of temporal segmentation is the decomposition of a video in camera-shots. Therefore, the temporal segmentation must primarily allow to exactly locating transitions between consecutive shots. Secondly, the classification of the type of transition is of interest. Basically, temporal segmentation algorithms are based on the evaluation of the quantitative differences between successive frames and on some kind of threshold. In general an effective segmentation technique must combine an inter-frame metric computationally simple and able to detect video content changes with robust decision criteria.

2.4 Techniques operating on compressed video

The Moving Picture Experts Group (MPEG) is a working group of ISO/IEC in charge of the development of international standards for compression, decompression, processing, and coded representation of moving pictures, audio and their combination. So far MPEG has produced:

- MPEG-1, the standard for storage and retrieval of moving pictures and audio on storage media (approved Nov. 1992)
- MPEG-2, the standard for digital television (approved Nov. 1994)
- MPEG-4, the standard for multimedia applications
- MPEG-7 the content representation standard for multimedia information search, filtering, management and processing (to be approved July 2001).
- MPEG-21, the multimedia framework.

MPEG uses two basic compression techniques: 16×16 macroblock-based motion compensation to reduce temporal redundancy and 8×8 Discrete Cosine Transform (DCT) block-based compression to capture spatial redundancy. An MPEG stream consists of three types of pictures, I, P and B, which are combined in a repetitive pattern called group of picture (GOP).

- I (Intra) frames provide random access points into the compressed data and are coded using only information present in the picture itself. DCT coefficients of each block are quantized and coded using Run Length Encoding (RLE) and entropy coding. The first DCT coefficient is called DC term and is proportional to the average intensity of the respective block.
- P (Predicted) frames are coded with forward motion compensation using the nearest previous reference (I or P) pictures.
- B (Bi-directional) pictures are also motion compensated, this time with respect to both past and future reference frames

A well known approach to temporal segmentation in the MPEG compressed domain, useful for detecting both abrupt and gradual transitions, has been proposed in (Yeo & Liu, 1995) using the DC sequences. Since this technique uses I, P and B frames, a partial decompression of the video is necessary. The DC terms of I frames are directly available in the MPEG stream, while those of B and P frames must be estimated using the motion vectors and the DCT coefficients of previous I frames. This reconstruction process is computationally very expensive.

Both PDM and HDM metrics are suited as similarity measures, but pixel differences-based metrics give satisfactory results as DC images are already smoothed versions of the corresponding full images. Gradual transitions are detected through an accurate temporal analysis of the metric.

3. Content-based analysis of digital video

Video content analysis is a strongly multidisciplinary research area. The ever increasing amount of multimedia data creates a need for new sophisticated methods to retrieve the information one is looking for. Analysis of the ground truths provided for development data revealed that the important sections to be included in summarized video were of four types: shots containing camera motion, shots of people entering or leaving a scene, shots showing certain objects, and shots of distinct events. Since high-level features can be indicators of the relative importance of a particular video segment (Todd, 2005), appropriate features were extracted to capture these four types. The video analysis (Ojala et al., 2001) task consists in recovering the object shape, object texture, and object motion from the given video, shown in Fig. 3.

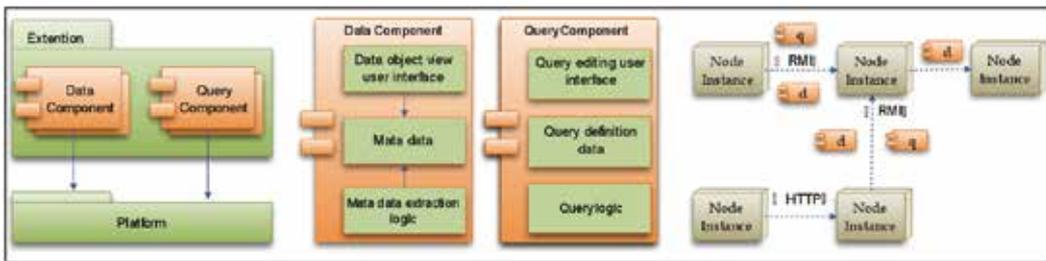


Fig. 3. Principal components of the CMRS architecture described using UML notation. (a) CMRS architecture comprises of a media independent platform and media specific extensions, which contain one or more Data and Query components. (b) Data and Query components encapsulate the data, user interface and operations related to physical media items and queries, respectively. (c) Runtime view of the platform.

3.1 Feature extraction and selection

Current efforts in the automatic video content analysis are directed primarily towards the decomposition of video streams into meaningful subsequences. By reducing the dimensionality of the input set correlated information is eliminated at the cost of a loss of accuracy (Addison, 2003). Dimensionality reduction can be achieved either by eliminating data closely related with other data in the set, or combining data to make a smaller set of features. The feature extraction techniques used in this study are principal components analysis and auto-associative neural networks. Feature selection is achieved by the use of genetic algorithms, sensitivity analysis.

3.1.1 Linear principal components analysis

Using the example of projecting data from two dimensions to one, a linear projection requires the optimum choice of projection to be a minimisation of the sum-of-squares error. This is obtained first by subtracting the mean \bar{x} of the data set. The covariance matrix is

calculated and its eigenvectors and eigenvalues are found. The eigenvectors corresponding to the M largest eigenvalues are retained, and the input vectors x^n are subsequently projected onto the eigenvectors to give components of the transformed vectors z^n in the M -dimensional space. Retaining a subset $M < d$ of the basis vectors u_i so that only M coefficients z_i are used allows for replacement of the remaining coefficients by constants b_i . This allows each x vector to be approximated by an expression of the form:

$$\tilde{x} = \sum_{i=1}^M z_i u_i + \sum_{i=M+1}^d b_i u_i \quad (11)$$

Where u_i represents a linear combination of d orthonormal vectors.

3.1.2 Auto Associative Networks (AAN)

An auto associative network (AAN) consists of a multi-layer perception with d inputs, d outputs, and M hidden units with $M < d$. (Fausett, 1994). The targets used to train the network are the input vectors themselves, which means the network is attempting to map each input vector onto itself. Because the number of units in the middle layer is reduced, a perfect reconstruction of the input vectors may not always be possible. The network is then trained using a sum of squares error of the following form.

$$E = \frac{1}{2} \sum_{n=1}^N \sum_{k=1}^d \{y_k(x^n) - x_k^n\}^2 \quad (12)$$

where N is the number of patterns in the sample and x_k^n represents the target value for the output unit k when the input vector is x^n . The error minimisation here performs a form of unsupervised training, even though we are using a supervised architecture as no independent target data is provided. Such networks perform a non-linear principal components analysis, which has the advantage of not using linear transformations.

3.1.3 Genetic algorithms

We use the Holland (Holland, 1975) algorithm. The algorithm can be expressed as follows.

1. Set generation counter $I \leftarrow 0$
2. Create initial population, $\text{Pop}(i)$, by random generation of N -individuals
3. Apply objective function to the individual, record the value found (determines data fitness)
4. Increment next generation, $I \leftarrow I + 1$
5. Select N individuals randomly from the previous population $\text{Pop}(I-1)$ based on their fitness
6. Select R parents from new population to form new population to form new children by applying the genetic operators
7. Evaluate fitness of newly formed children by applying the objective function
8. If $I < \text{maximum number of generations to be considered}$, go to step (4)
9. Write out best solution found

3.1.4 Sensitivity analysis

Sensitivity analysis (SA) is the study of how the variation (uncertainty) in the output of a mathematical model can be apportioned, qualitatively or quantitatively, to different sources

of variation in the input of a model (Bishop, 1997). We conducted sensitivity analysis by treating each input variable in turn as if it were “unavailable” (Hunter et al., 2000). A neural network is trained using all of the input attributes, and the values of training and test set errors are produced. Afterwards the network is “pruned” of input variables whose training and verification errors are below the threshold. In this way variables can be assessed according to the deterioration effect they have upon network performance if removed.

3.2 Content comparison techniques

“Content-based” means that the search will analyze the actual contents of the frame image. Unfortunately, automatic indexing and feature extraction from digital video is even harder than still-image analysis. Several techniques are proposed to automatically segment digital video into scenes, shots and subshots based on color histogram, motion, texture and shape features (Volker, 1999). The term ‘content’ in this context might refer to colors, shapes, textures, or any other information that can be derived from the frame image itself.

- *Color* represents the distribution of colors within the entire image. This distribution includes the amounts of each color.
- *Texture* represents the low-level patterns and textures within the image, such as graininess or smoothness. Unlike shape, texture is very sensitive to features that appear with great frequency in the image.
- *Shape* represents the shapes that appear in the image, as determined by color-based segmentation techniques. A shape is characterized by a region of uniform color.

3.2.1 Color

Color reflects the distribution of colors within the entire frame image. A color space is a mathematical representation of a set of colors. The three most popular color models are RGB (used in computer graphics); YIQ, YUV or YCbCr (used in video systems); and CMYK (used in color printing). However, none of these color spaces are directly related to the intuitive notions of hue, saturation, and brightness. This resulted in the temporary pursuit of other models, such as HIS and HSV, to simplify programming, processing, and end-user manipulation.

- RGB Color Space

The red, green, and blue (RGB) color space is widely used throughout computer graphics. Red, green, and blue are three primary additive colors (individual components are added together to form a desired color) and are represented by a three-dimensional, Cartesian coordinate system (Figure 3.1). The indicated diagonal of the cube, with equal amounts of each primary component, represents various gray levels. Table 1 contains the RGB values for 100% amplitude, 100% saturated color bars, a common video test signal.

| | Nominal Range | White | Yellow | Cyan | Green | Magenta | Red | Blue | Black |
|---|---------------|-------|--------|------|-------|---------|-----|------|-------|
| R | 0 to 255 | 255 | 255 | 0 | 0 | 255 | 255 | 0 | 0 |
| G | 0 to 255 | 255 | 255 | 255 | 255 | 0 | 0 | 0 | 0 |
| B | 0 to 255 | 255 | 0 | 255 | 0 | 255 | 0 | 255 | 0 |

Table 1. 100% RGB Color Bars

The RGB color space is the most prevalent choice for computer graphics because color displays use red, green, and blue to create the desired color. However, RGB is not very efficient when dealing with “real-world” images. All three RGB components need to be of equal band-width to generate any color within the RGB color cube. The result of this is a frame buffer that has the same pixel depth and display resolution for each RGB component. Also, processing an image in the RGB color space is usually not the most efficient method. For these and other reasons, many video standards use luma and two color difference signals. The most common are the YUV, YIQ, and YCbCr color spaces. Although all are related, there are some differences.

- YCbCr Color Space

The YCbCr color space was developed as part of ITU-R BT.601 during the development of a world-wide digital component video standard. YCbCr is a scaled and offset version of the YUV color space. Y is defined to have a nominal 8-bit range of 16–235; Cb and Cr are defined to have a nominal range of 16–240. There are several YCbCr sampling formats, such as 4:4:4, 4:2:2, 4:1:1, and 4:2:0.

RGB - YCbCr Equations: SDTV

The basic equations to convert between 8-bit digital R'G'B' data with a 16–235 nominal range and YCbCr are:

$$\begin{aligned}
 Y_{601} &= 0.2999R' + 0.587G' + 0.114B' \\
 Cb &= -0.172R' - 0.399G' + 0.511B' + 128 \\
 Cr &= 0.511R' - 0.428G' - 0.083B' + 128
 \end{aligned}
 \tag{13}$$

$$\begin{aligned}
 R' &= Y_{601} + 1.371(Cr - 128) \\
 G' &= Y_{601} - 0.698(Cr - 128) - 0.336(Cb - 128) \\
 B' &= Y_{601} + 1.732(Cb - 128)
 \end{aligned}$$

When performing YCbCr to R'G'B' conversion, the resulting R'G'B' values have a nominal range of 16–235, with possible occasional excursions into the 0–15 and 236–255 values. This is due to Y and CbCr occasionally going outside the 16–235 and 16–240 ranges, respectively, due to video processing and noise. Note that 8-bit YCbCr and R'G'B' data should be saturated at the 0 and 255 levels to avoid underflow and overflow wrap-around problems. Table 2 lists the YCbCr values for 75% amplitude, 100% saturated color bars, a common video test signal.

| | Nominal Range | White | Yellow | Cyan | Green | Magenta | Red | Blue | Black |
|------|---------------|-------|--------|------|-------|---------|-----|------|-------|
| SDTV | | | | | | | | | |
| Y | 16 to 235 | 180 | 162 | 131 | 112 | 84 | 65 | 35 | 16 |
| Cb | 16 to 240 | 128 | 44 | 156 | 72 | 184 | 100 | 212 | 128 |
| Cr | 16 to 240 | 128 | 142 | 44 | 58 | 198 | 212 | 114 | 128 |
| HDTV | | | | | | | | | |
| Y | 16 to 235 | 180 | 168 | 145 | 133 | 63 | 51 | 28 | 16 |
| Cb | 16 to 240 | 128 | 44 | 147 | 63 | 193 | 109 | 212 | 128 |
| Cr | 16 to 240 | 128 | 136 | 44 | 52 | 204 | 212 | 120 | 128 |

Table 2. 75% YCbCr Color Bars.

RGB - YCbCr Equations: HDTV

The basic equations to convert between 8-bit digital R'G'B' data with a 16–235 nominal range and YCbCr are:

$$\begin{aligned}
 Y_{709} &= 0.213R' + 0.751G' + 0.072B' \\
 Cb &= -0.117R' - 0.394G' + 0.511B' + 128 \\
 Cr &= 0.511R' - 0.464G' - 0.047B' + 128
 \end{aligned}
 \tag{14}$$

$$\begin{aligned}
 R' &= Y_{709} + 1.540(Cr - 128) \\
 G' &= Y_{709} - 0.459(Cr - 128) - 0.183(Cb - 128) \\
 B' &= Y_{709} + 1.816(Cb - 128)
 \end{aligned}$$

When performing YCbCr to R'G'B' conversion, the resulting R'G'B' values have a nominal range of 16–235, with possible occasional excursions into the 0–15 and 236–255 values. This is due to Y and CbCr occasionally going outside the 16–235 and 16–240 ranges, respectively, due to video processing and noise. Note that 8-bit YCbCr and R'G'B' data should be saturated at the 0 and 255 levels to avoid underflow and overflow wrap-around problems. Table 2 lists the YCbCr values for 75% amplitude, 100% saturated color bars, a common video test signal.

- HSI, HLS, and HSV Color Spaces

The HSI (hue, saturation, intensity) and HSV (hue, saturation, value) color spaces were developed to be more “intuitive” in manipulating color and were designed to approximate the way humans perceive and interpret color. They were developed when colors had to be specified manually, and are rarely used now that users can select colors visually or specify Pantone colors. These color spaces are discussed for “historic” interest. HLS (hue, lightness, saturation) is similar to HSI; the term lightness is used rather than intensity. The difference between HSI and HSV is the computation of the brightness component (I or V), which determines the distribution and dynamic range of both the brightness (I or V) and saturation(S). The HSI color space is best for traditional image processing functions such as convolution, equalization, histograms, and so on, which operate by manipulation of the brightness values since I is equally dependent on R, G, and B. The HSV color space is preferred for manipulation of hue and saturation (to shift colors or adjust the amount of color) since it yields a greater dynamic range of saturation.

3.2.2 Texture

Texture reflects the texture of the entire image. Texture is most useful for full images of textures, such as catalogs of wood grains, marble, sand, or stones. A variety of techniques have been developed for measuring texture similarity. Most techniques rely on comparing values of what are known as second-order statistics calculated from query and stored images. These methods calculate measures of image texture such as the degree of contrast, coarseness, directionality and regularity; or periodicity, directionality and randomness (Liu & Picard, 1996). Alternative methods of texture analysis for image retrieval include the use of Gabor filters and fractals (Kaplan, 1998). Gabor filter (or Gabor wavelet) is widely adopted to extract texture features from the images for image retrieval and has been shown to be very efficient. Manjunath and Ma (Manjunath & Ma, 1996) have shown that image

retrieval using Gabor features outperforms that using pyramid-structured wavelet transform (PWT) features, tree-structured wavelet transform (TWT) features and multiresolution simultaneous autoregressive model (MR-SAR) features.

Haralick (Haralick, 1979) and Van Gool (Gool et al., 1985) divide the techniques for texture description into two main categories: statistical and structural. Most natural textures can not be described by any structural placement rule, therefore the statistical methods are usually the methods of choice. One possible approach to reveal many of the statistical texture properties is by modelling the texture as an autoregressive (AR) stochastic process, using least squares parameter estimation. Letting s and r be coordinates in the 2-D coordinate system, a general causal or non-causal auto-regressive model may be written:

$$y(s) = \sum_{r \in N} \theta_r y(s-r) + e(s) \quad (15)$$

Where $y(s)$ is the image, θ_r are the model parameters, $e(s)$ is the prediction error process, and N is a neighbour set. The usefulness of this modelling is demonstrated with experiments showing that it is possible to create synthetic textures with visual properties similar to natural textures.

3.2.3 Shape

Shape represents the shapes that appear in the image. Shapes are determined by identifying regions of uniform color. In the absence of color information or in the presence of images with similar colors, it becomes imperative to use additional image attributes for an efficient retrieval. Shape is useful to capture objects such as horizon lines in landscapes, rectangular shapes in buildings, and organic shapes such as trees. Shape is very useful for querying on simple shapes (like circles, polygons, or diagonal lines) especially when the query image is drawn by hand. Incorporating rotation invariance in shape matching generally increases the computational requirements.

4. Semantic-based annotation for digital video

An annotation represents any symbolic description of a video, or an excerpt of a video. Semantic concepts do not occur in isolation. There is always a context to the co-occurrence of semantic concepts in a video scene. To locate and represent the semantic meanings in video data is the key to enable intelligent query. This section presents the methodology for knowledge elicitation and, in particular, introduces a preliminary semantic representation scheme that bridges the information gap not only between different media but also between different levels of contents in the media.

4.1 Framework of object-based video indexing

Object segmentation and tracking is a key component for new generation of digital video representation, transmission and manipulations. The schema provides a general framework for video object extraction, indexing, and classification. By video objects, here we refer to objects of interest including salient low-level image regions (uniform color/texture regions), moving foreground objects, and group of primitive objects satisfying spatio-temporal constraints (e.g., different regions of a car or a person). Automatic extraction of video objects at different levels can be used to generate a library of video data units, from which various

functionalities can be developed. For example, video objects can be searched according to their visual features, including spatio-temporal attributes. High-level semantic concepts can be associated with groups of low-level objects through the use of domain knowledge or user interaction.

As mentioned above, in general, it is hard to track a meaningful object (e.g., a person) due to its dynamic complexity and ambiguity over space and time. Objects usually do not correspond to simple partitions based on single features like color or motion. Furthermore, definition of high-level objects tends to be domain dependent. On the other hand, objects can usually be divided into several spatial homogeneous regions according to image features. These features are relatively stable for each region over time. For example, color is a good candidate for low-level region tracking. It does not change significantly under varying image conditions, such as change in orientation, shift of view, partial occlusion or change of shape. Some texture features like coarseness and contrast also have nice invariance properties. Thus, homogenous color or texture regions are suitable candidates for primitive region segmentation. Further grouping of objects and semantic abstraction can be developed based on these basic feature regions and their spatio-temporal relationship. Based on these observations, we proposed the following model for video object tracking and indexing (Fig. 4).

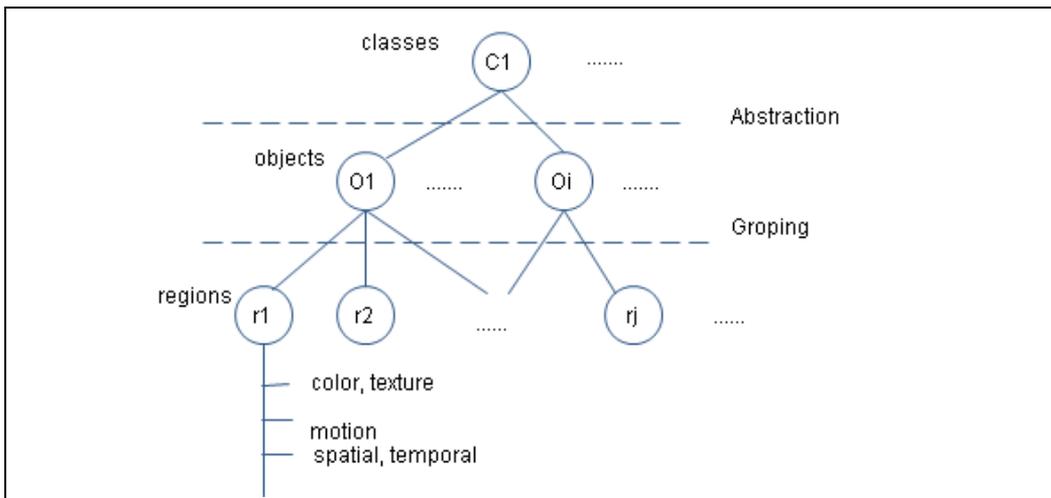


Fig. 4. Hierarchical representation of video objects

At the bottom level are primitive regions segmented according to color, texture, or motion measures. As these regions are tracked over time, temporal attributes such as trajectory, motion pattern, and life span can be obtained. The top level includes links to conceptual abstraction of video objects. For example, a group of video objects may be classified to moving human figure by identifying color regions (skin tone), spatial relationships (geometrical symmetry in the human models), and motion pattern of component regions. We propose the above hierarchical video object schema for content-based video indexing. One challenging issue here is to maximize the extent of useful information obtained from automatic image analysis tasks. A library of low-level regions and mid-level video objects can be constructed to be used in high-level semantic concept mapping. This general schema can be adapted to different specific domains efficiently and achieve higher performance.

4.2 Video indexing using face detection and face recognition methods

In this section, we will construct such a signature by using semantic information, namely information about the appearance of faces of distinct individuals. We will not concern ourselves with the extraction of face-related information, since ample work has been performed on the subject. Instead we will try to solve the problems of consistency and robustness with regards to face-based indexing, to represent face information with minimal redundancy, and also to find a fast (logarithmic-time) search method. All works on face-related information for video indexing until now have focused on the extraction of the face-related information and not on its organization and efficient indexing. In effect, they are works on face recognition with a view to application on indexing. The face detection stage is presented in Fig. 5.



Fig. 5. Detection algorithm scheme

The skin pixel detection is performed with a simple colormap, using the YCbCr color space. The second block corresponds to the segmentation algorithm which is performed in two stages, where the chrominance and luminance information is used consecutively. For each stage an algorithm which combines pixel and region based color segmentation techniques is used. After the segmentation, a set of connected homogenous skin-like regions is obtained. Then, potential face candidates (FC) are obtained by an iterative merging procedure using an adjacency criterion. Once the set of FC is built, it is necessary to remove the ones that do not match to any face. To that end some constrains regarding shape, size and overlapping are used.

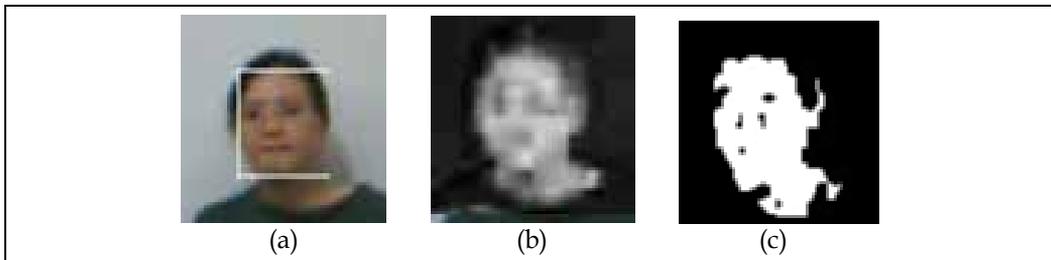


Fig. 6. Results of face detection: (a) the capture frame image; (b) result of similarity; (c) the binary result

Some results are shown in Fig. 6. Notice that the algorithm is able to produce good face candidates but also provides some candidates which do not correspond to any face. Most of these erroneous face candidates will be not recognized in the face recognition stage and will be discarded, but at the expenses of increasing unnecessarily the computational cost. Thus, if there are many erroneous face candidates, the overall system becomes inefficient.

Face areas are characterized to have a homogeneous chrominance component. Taking into account this fact, a new selection criterion has been designed in order to remove all the FC composed of regions whose average color differs substantially, as is the case when hair and face are included in the same candidate.

4.3 Text extraction for video indexing

The increasing availability of online digital video has rekindled interest in the problems of how to index multimedia information sources automatically and how to browse and manipulate them efficiently (David, 1998) (Snoek & Worring, 2005) (Zhu et al., 2006). The need for efficient content-based video indexing and retrieval has increased due to the rapid growth of video data available to consumers. For this purpose, text in video, especially the superimposed text, is the most frequently used since it provides high level semantic information about video content and it has distinctive visual characteristic.

Extraction of text information involves detection, localization, tracking, extraction, enhancement, and recognition of the text from a given video. Although a large number of techniques have been done on solving this problem, very little work has ever considered the temporal aspects of video. Text may span tens or even hundreds of frames, providing a tremendous amount of redundant information. In this section we address some of the efficient and reliable aspects of the text matching algorithms, by using fast correlation, rectangular sub-images techniques in a multi-resolution scheme, which enables real-time text tracking and locating applications on a standard personal computer (Ding et al., 2008).

4.3.1 Matching strategy

Similarity is the guiding principle for solving the matching problem. Among the different similarity measures proposed in the literature, Normalized Cross-Correlation (NCC) is widely used to their robustness in template matching. It has also been shown that NCC tends to give better results (Wu et al., 1995). Suppose T is a template to be tracking in an image I . Template matching by normalized correlations computes the following value at each point (x, y) of the image I :

$$c(x, y) = \frac{\text{cov}_{x,y}(T, I)}{\sqrt{Q_{x,y}(T)Q_{x,y}(I)}} \quad (16)$$

in which,

$$\begin{aligned} \text{cov}_{x,y}(T, I) &= \sum_{u,v} T(u, v)I(x + u, y + v) \\ Q_{x,y}(T) &= \sum_{u,v} T^2(u, v) \\ Q_{x,y}(I) &= \sum_{u,v} I^2(x + u, y + v) \end{aligned} \quad (17)$$

where the summations are over all template coordinates. A large value of $c(x, y)$ indicates a likely match at the coordinate (x, y) . It can be shown that a match that maximizes c is identical to the template T up to scaling.

4.3.2 Computational optimisation

In this section we outline the optimization techniques adopted to avoid redundant calculations. Fig. 7 shows the coordinate relations between template and candidate images. The arrow denotes the moving direction of template. Expansion (16):

$$\text{cov}_{x,y}(T, I) = \sum_{u=1}^W \sum_{v=1}^H T(u, v)I(x + u, y + v) \quad (18)$$

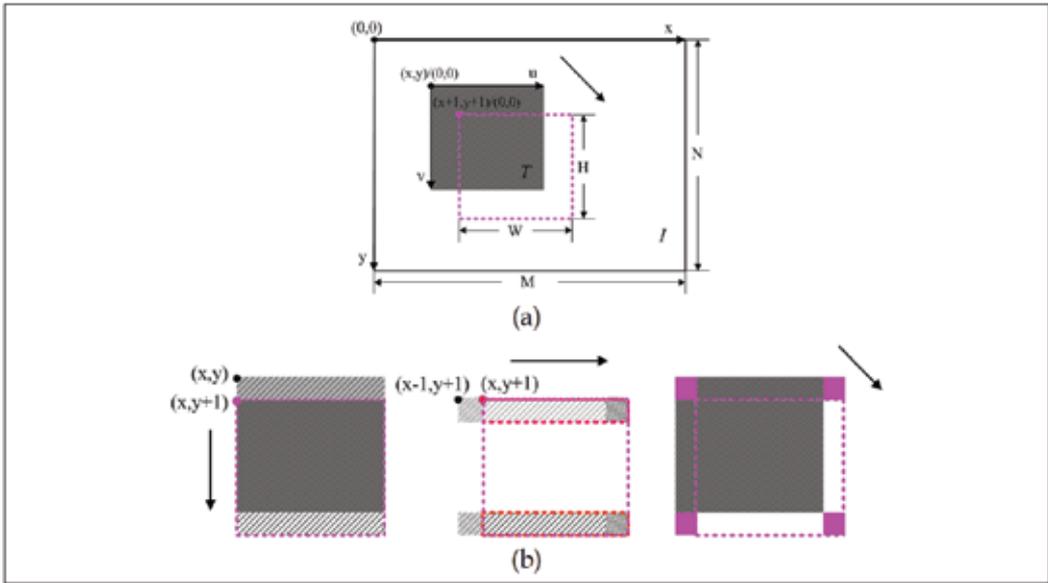


Fig. 7. (a)The coordinate relations between template and candidate images; (b) the steps of the template movement

Observing Figure 2, and following Eq. (18), suppose that $cov_{x,y}(T,I)$ is the NCC score between a template T and a candidate image I at point (x,y) . It is easy to notice that $cov_{x,y+1}(T,I)$ can be attained from $cov_{x,y}(T,I)$:

$$cov_{x,y+1}(T,I) = cov_{x,y}(T,I) + diff_{x,y+1}(T,I) \tag{19}$$

With $diff_{x,y+1}(T,I)$ representing the difference between the $c(x,y)$ associated with the lowermost and uppermost rows of the matching window (bias region shown in Figure 2(a)):

$$diff_{x,y+1}(T,I) = \sum_{u=1}^W [T(u,H)I(x+u,y+H+1) - T(u,1)I(x+u,y+1)] \tag{20}$$

And

$$diff_{x,y+1}(T,I) = diff_{x-1,y+1}(T,I) + [T(W,H)I(x+W,y+H+1) - T(1,H)I(x,y+H+1)] + [T(W,1)I(x+W,y+1) - T(1,1)I(x,y+1)] \tag{21}$$

This allows for keeping complexity small and independent of the size of the matching window, which is the size of the template, since only four elementary operations are needed to obtain the NCC score at each new point.

With the optimized algorithm, which is associated with the 4 purple points of Figure 2. The computational scheme of Eq. (16) and (21) makes use of a vertical recursion and a horizontal recursion to obtain the updating term.

When using the conventional NCC method, there is a computation of $(M \times N \times W \times H)$ times sum-of-squares operations for each matching point, however, this computation will be degraded to $(M \times N + W \times H)$ times operations with the optimized algorithm assuming the diff is unknown, otherwise only 4 times operations needed, similarly calculating Q.

With the optimized algorithm, which is associated with the 4 purple points of Figure 2. The computational scheme of Eq. (16) and (21) makes use of a vertical recursion and a horizontal recursion to obtain the updating term.

4.3.3 Multi-Resolution matching and result

The NCC-based text matching is region-based and its computational cost is considerable when tracking a large text line. The coarse-to-fine technique uses a hierarchical representation of image called Gaussian Pyramid (Bergen et al., 1992).

We collected a large amount of video sequences for experiments on text tracking from a wide variety of video sources, including movie credits, TV programs and news etc. A description of running times and match score of the algorithm on different video frames is listed in Table 3.

| Algorithm | Candidate Image Size | Template Image Size | Match Percentage | Runtime (second) |
|-----------------|----------------------|---------------------|------------------|------------------|
| Fast NNC | 640×480 | 428×34 | 98.783% | 0.0239 |
| cvMatchTemplate | 640×480 | 428×34 | 98.784% | 0.1670 |
| Fast NNC | 352×288 | 173×31 | 98.088% | 0.0283 |
| cvMatchTemplate | 352×288 | 288×384 | 98.088% | 0.1698 |
| Fast NNC | 512×384 | 370×30 | 92.906% | 0.0258 |
| cvMatchTemplate | 512×384 | 370×30 | 92.907% | 0.1710 |

Table 3. Running times and match score



Fig. 8. Results of news detection: (a) template (b)1057th frame, 97.348%*(c)1158th frame, 97.859%* (d)1280th frame, 95.586%*

The time spent in the stage for obtaining coefficients for the normalized cross-correlation is obviously reduced comparing to the `cvMatchTemplate` in the same search window size, but the match score is almost invariant. Various types of videos have been measured by using our proposed method. Figure 4 shows result of news in 25 fps. Where the green box in Fig. 8 (a) is the text region referred to as template, it is the text detection result of frame 996. Where * denote the match percentage.

The measure results of American TV programs (the Apprentice) in 25fps. The shot, as well as background, is change frequently in this video. Where the green boxes in Figure 8(a) are the text detection result of frame 5564.

Generally, when the match score is lower than 90% we think the result can not be confident and new detection needed. Unfortunately, quantitative evaluation of tracking accuracy is not easy because of the lack of truth data. Our experiments show that the matcher can work well when the text is in simple (rigid, linear) motion, even in complex backgrounds.

5. Future research and conclusions

We have presented a novel method for performing fast retrieval of video segments based on the output of face detectors and recognizers, with possible uses to indexing of video databases. Finally, we have developed a fast text matching method using fast correlation in the coarse-to-fine framework. By using the normalized cross-correlation (NCC) similarity measure rather than the simple SSD or SAD, the reliability of the algorithm is increased. The fast cross-correlation has been realized by using the recursive technique. These systems are similar to ours in that they use features from the visual data to segment and index video content. We also focus on content-based retrieval that consumers would want in their homes such as automatic personalization of content retrieval based on user profiles.

During the last thirty years we have witnessed a tremendous explosion in research and applications in the visual communications field. The field is now mature as is proven by the large number of applications that make use of this technology. There is no doubt that the beginning of the new century revolves around the "information society." Technologically speaking, the information society will be driven by audio and visual applications that allow instant access to multimedia information.

6. Acknowledgments

This work is supported by the research and application of intelligent equipment based on untouched techniques for children under 8 years old of BMSTC & Beijing Municipal Education Commission (No. 2007B06 & No. KM200810028017).

7. References

- Hjelsvold, R. (1995). *VideoSTAR - A database for video information sharing*, Ph.D. Thesis. Norwegian Institute of Technology
- Brown, K. (2001). *A rich diet: Data-rich multimedia has a lot in store for archiving and storage companies*, Broadband Week
- Baeza-Yates R. & Ribeiro-Neto B. (1999). *Modern Information Retrieval*, Addison-Wesley / ACM Press, ISBN-13: 978-0201398298, New York

- Zhang Y.J. (2006). *Semantic-Based Visual Information Retrieval*, IRM Press, ISBN-13: 978-1599043715, USA
- Li J. Z. & Özsu M. T. (1997). STARS: A Spatial Attributes Retrieval System for Images and Videos, *Proceedings of the 4th International Conference on Multimedia Modeling (MMM'97)*, pp 69-84, Singapore
- Dağtas, Al-Khatib W.; Ghafoor A. & Kashyap R. L. (2000). Models for Motion-Based Indexing and Retrieval, *IEEE Transactions on Image Processing*, Vol. 9, No. 1, January 2000
- Al-Khatib Q.; Day F.; Ghafoor A. & Berra B. (1999). Semantic Modelling and Knowledge Representation in 7 Multimedia Databases, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 11, No. 1, January/February 1999
- Analyti A. & Christodoulakis S. (1995). Multimedia Object Modeling and Content-Based Querying, *Proceedings of Advanced Course - Multimedia databases in Perspective*, Netherlands 1995.
- Yeo B-L. & Yeung M. (1997). Retrieving and Visualizing Video, *Communications of the ACM*, Vol. 40, No. 12, December 1997
- Kyriakaki G. (2000) *MPEG Information Management Services for Audiovisual Applications*, Master Thesis, Technical University of Crete, March 2000
- Hacid M-S.; Declair C. & Kouloumdjian J. (2000). A Database Approach for Modeling and Querying Video Data, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 12, No. 5, September/October 2000
- Lienhart, R. (1999). Comparison of automatic shot boundary detection algorithms. In: *SPIE Conf. on Storage and Retrieval for Image and Video Databases VII*, vol. 3656, pp. 290-301
- Zhang H. J.; Kankanhalli A. & Smoliar S. W. (1993). Automatic Partitioning of Full Motion Video, *Multimedia Systems*, vol.1, pp:10 - 28
- Hampapur A.; Jain R. & T. Weymouth. (1994). Digital Video Segmentation. In *Proc. ACM Multimedia*, pp: 357 - 364
- Teng S.H. & Tan W.W. (2008). Video Temporal Segmentation Using Support Vector Machine, *Contact Information Retrieval Technology*, pp: 442-447
- Smeaton A.F. (2007). Techniques used and open challenges to the analysis, indexing and retrieval of digital video. *Information Systems*, vol.32(4), pp: 545-559
- Ford R. M.; Robson C.; Tample D.; & Gerlach M. (1997). Metrics for Scene Change Detection in Digital Video Sequences, In *IEEE International Conference on Multimedia Computing and Systems (ICMCS '97)*, pp: 610-611, Ottawa, Canada
- Dugad R.; Ratakonda K. & Ahuja N. (1998). Robust Video Shot Change Detection, *IEEE Second Workshop on Multimedia Signal Processing*, pp: 376-381, Redondo Beach, California
- Sethi K. & Patel N. V. (1995). A Statistical Approach to Scene Change Detection, in *IS&T SPIE Proceedings: Storage and Retrieval for Image and Video Databases III*, Vol. 2420, pp. 329-339, San Jose
- Gargi U.; Kasturi R. & Strayer S. H. (2000). Performance Characterization of Video-Shot-Change Detection Methods, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 10, No. 1, pp. 1-13.
- A. Dailianas, R. B. Allen, and P. England, Comparison of Automatic Video Segmentation Algorithms, in *Proceedings of SPIE Photonics West, 1995.SPIE*, Vol. 2615, pp. 2-16, Philadelphia, 1995.

- Ardizzone E. & La Cascia M. (1997). Automatic Video Database Indexing and Retrieval, *Multimedia Tools and Applications*, Vol. 4, pp. 29–56
- Yeo L. & Liu B. (1995). Rapid Scene Analysis on Compressed Video, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 5, No. 6, pp. 533–544
- Todd R. R. (2005). *Digital Image Sequence Processing, Compression, and Analysis*, CRC Press, ISBN: 0849315263, 9780849315268, USA
- Ojala T.; Kauniskangas H.; Keränen H.; Matinmikko E.; Aittola M.; Hagelberg K.; Rautiainen M. & Häkkinen M. (2001). CMRS: Architecture for content-based multimedia retrieval. *Proc. Infotech Oulu International Workshop on Information Retrieval*, Oulu, Finland, pp: 179-190.
- Addison J. F. D.; MacIntyre J. (editors). (2003). Intelligent techniques: A review, *Springer Verlag (UK) Publishing Company*, 1st Edition, Chapter 9
- Fausett, L. (1994). *Fundamentals of Neural Networks*, Englewood Cliffs, NJ: Prentice Hall.
- Holland J. (1975). *Adaptation in Natural and Artificial systems*. MIT Press.
- Bishop C. M. (1997). *Neural networks for pattern recognition*, Oxford University Press, pp: 6-9
- Hunter A.; Kennedy L.; Henry J. & Ferguson R.I. (2000). Application of Neural Networks and Sensitivity Analysis to improved prediction of Trauma Survival, *Computer Methods and Algorithms in Biomedicine* 62, pp: 11-19
- Volker R. (1999). Content-based retrieval from digital video, *Image and Vision Computing*, Volume 17, Issue 7, May 1999, pp: 531-540
- Liu F. & Picard R. W. (1996). Periodicity, directionality and randomness: World features for image modelling and retrieval, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(7), pp: 722-733
- Manjunath B. S. & Ma W. Y. (1996). Texture features for browsing and retrieval of large image data, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (Special Issue on Digital Libraries), Vol. 18 (8), August 1996, pp: 837-842.
- Kaplan L. M. et al. (1998). Fast texture database retrieval using extended fractal features, *In Storage and Retrieval for Image and Video Databases VI* (Sethi, I K and Jain, R C, eds), Proc SPIE 3312, pp: 162-173
- Haralick R. M. (1979). Statistical and structural approaches to texture. *Proc. IEEE*. Vol 67, pp: 786-804
- Gool L. V. & Dewaele P. and Oosterlinck A. (1985). Texture analysis anno 1983, *Computerr Vision, Graphics and Image Processing*, vol 29, pp: 336-357.
- Ding H.; Ding X. Q.; Wang S. J. (2008). Texture Fusion Based Wavelet Transform Applied to Video Text Enhancement, *Journal of Information and Computational Science*, pp: 2083-2090
- Ding H.; Ding X. Q.; Wang S. J. (2008). Fast Text Matching in Digital Videos, *International Symposium on Computational Intelligence and Design 2008*, China, pp: 273-276, 2008.
- David D. (1998). The indexing and retrieval of document images: a survey, *International Journal of Computer Vision and Image Understanding*, 70(3), pp: 287-298.
- Snoek C.G.M. & Worring M. (2005). Multimodal Video Indexing: A Review of the State-of-the-art, *Multimedia Tools and Applications*, 25(1), pp: 5-35
- Zhu Q.; Yeh M.C. & K.T. Cheng. (2006). Multimodal fusion using learned text concepts for image categorization, *Proc. 14th ACM Conf. on Multimedia*, Santa Barbara, CA, 2006, pp: 211–220

- Wu Q.X.; McNeill S.J. & Pairman D. (1995). Fast algorithms for correlation-relaxation technique to determine cloud motion fields?, *Proc. Digital Image Computing: Techniques and Applications*, Brisbane, Australia, 1995, pp: 330-335
- Schweitzer H.; Bell J.W. & F. Wu. (2002). Very fast template matching, *Proc. 7th European Conf. on Computer Vision*, Copenhagen, Denmark, pp: 358-372.
- Gonzalez R.C.; Woods R.E. (1992). *Digital Image Processing*, Massachusetts: Addison-Wesley,
- Bergen J.R.; Anandan P.; Hanna K.J. & R. Himgorani. (1992). Hierarchical Model-based Motion Estimation, *Proc. 2nd European Conf. on Computer Vision*, Santa Margherita Ligure, Italy, pp: 237-252.

Video Editing Based on Situation Awareness from Voice Information and Face Emotion

Tetsuya Takiguchi, Jun Adachi and Yasuo Ariki
*Kobe University
Japan*

1. Introduction

Video camera systems are becoming popular in home environments, and they are often used in our daily lives to record family growth, small home parties, and so on. In home environments, the video contents, however, are greatly subjected to restrictions due to the fact that there is no production staff, such as a cameraman, editor, switcher, and so on, as with broadcasting or television stations.

When we watch a broadcast or television video, the camera work helps us to not lose interest in or to understand its contents easily owing to the panning and zooming of the camera work. This means that the camera work is strongly associated with the events on video, and the most appropriate camera work is chosen according to the events. Through the camera work in combination with event recognition, more interesting and intelligible video content can be produced (Ariki et al., 2006).

Audio has a key index in the digital videos that can provide useful information for video retrieval. In (Sundaram et al, 2000), audio features are used for video scene segmentation, in (Aizawa, 2005) (Amin et al, 2004), they are used for video retrieval, and in (Asano et al, 2006), multiple microphones are used for detection and separation of audio in meeting recordings. In (Rui et al, 2004), they describe an automation system to capture and broadcast lectures to online audience, where a two-channel microphone is used for locating talking audience members in a lecture room. Also, there are many approaches possible for the content production system, such as generating highlights, summaries, and so on (Ozeke et al, 2005) (Hua et al, 2004) (Adams et al, 2005) (Wu, 2004) for home video content.

Also, there are some studies that focused on a facial direction and facial expression for a viewer's behavior analysis. (Yamamoto, et al, 2006) proposed a system for automatically estimating the time intervals during which TV viewers have a positive interest in what they are watching based on temporal patterns in facial changes using the Hidden Markov Model. In this chapter, we are studying about home video editing based on audio and face emotion. In home environments, since it may be difficult for one person to record video continuously (especially for small home parties: just two persons), it will require the video content to be automatically recorded without a cameraman. However, it may result in a large volume of video content. Therefore, this will require digital camera work which uses virtual panning and zooming by clipping frames from hi-resolution images and controlling the frame size and position (Ariki et al, 2006).

In this chapter, our system can automatically capture only conversations using a voice/non-voice detection algorithm based on AdaBoost. In addition, this system can clip and zoom in on a talking person only by using the sound source direction estimated by CSP, where a two-channel (stereo) microphone is used. Additionally, we extract facial feature points by EBGM (Elastic Bunch Graph Matching) (Wiskott et al, 1997) to estimate atmosphere class by SVM (Support Vector Machine).

One of the advantages of the digital shooting is that the camera work, such as panning and zooming, is adjusted to user preferences. This means that the user can watch his/her own video produced by his/her own virtual editor, cameraman, and switcher based on the user's personal preferences. The main point of this chapter is that home video events can be recognized using techniques based on audio and face emotion and then used as the key indices to retrieve the events and also to summarize the whole home video.

The organization of this chapter is as follows. In Section 2, the overview of the video editing system based on audio and face emotion is presented. Section 3 describes voice detection with AdaBoost in order to capture conversation scenes only. Section 4 describes the estimation of the talker's direction with CSP in order to zoom in on the talking person by clipping frames from the conversation scene videos. Section 5 describes facial emotion recognition. Section 6 describes the digital camera work.

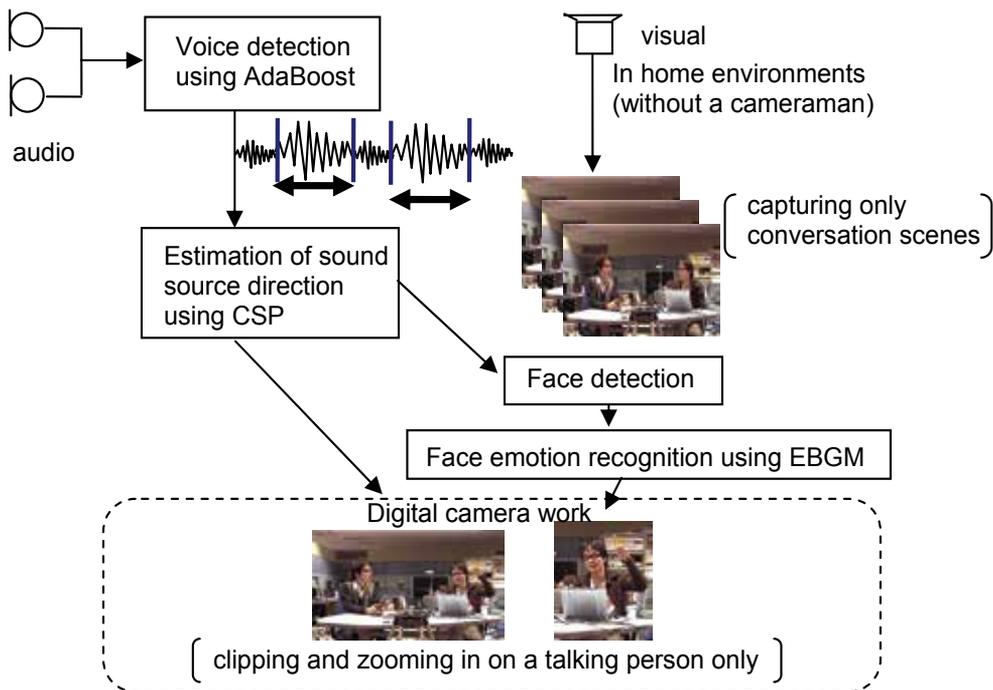


Fig. 1. Video editing system using digital camera work based on audio and face emotion

2. Overview of the system

Figure 1 shows the overview of the video editing system using digital camera work based on audio and face emotion. The first step is voice detection with AdaBoost, where the

system identifies whether the audio signal is a voice or not in order to capture conversation scenes only. When the captured video is a conversation scene, the system performs the second step. The second step is estimation of the sound source direction using the CSP (Crosspower-Spectrum Phase) method, where a two-channel microphone is used. Using the sound source direction, the system can clip and zoom in on a talking person only. The third step is face emotion recognition. Using the emotion result, the system can zoom out on persons who have positive expressions (happiness, laughter, etc).

3. Voice detection with AdaBoost

In automatic production of home videos, a speech detection algorithm plays an especially important role in capture of conversation scenes only. In this section, a speech/non-speech detection algorithm using AdaBoost, which can achieve extremely high detection rates, is described.

"Boosting" is a technique in which a set of weak classifiers is combined to form one highperformance prediction rule, and AdaBoost (Freund et al, 1999) serves as an adaptive boosting algorithm in which the rule for combining the weak classifiers adapts to the problem and is able to yield extremely efficient classifiers.

Figure 2 shows the overview of the voice detection system based on AdaBoost. The audio waveform is split into a small segment by a window function. Each segment is converted to the linear spectral domain by applying the discrete Fourier transform (DFT). Then the logarithm is applied to the linear power spectrum, and the feature vector is obtained. The AdaBoost algorithm uses a set of training data,

$$\{(X(1), Y(1)), \dots, (X(N), Y(N))\} \quad (1)$$

where $X(n)$ is the n -th feature vector of the observed signal and Y is a set of possible labels. For the speech detection, we consider just two possible labels, $Y = \{-1, 1\}$, where the label, 1, means voice, and the label, -1, means noise. Next, the initial weight for the n -th training data is set to

$$w_1(n) = \begin{cases} \frac{1}{2m}, & Y(n) = 1 \quad (\text{voice}) \\ \frac{1}{2l}, & Y(n) = -1 \quad (\text{noise}) \end{cases} \quad (2)$$

where m is the total voice frame number, and l is the total noise frame number.

As shown in Figure 2, the weak learner generates a hypothesis $h_t : X \rightarrow \{-1, 1\}$ that has a small error. In this chapter, single-level decision trees (also known as decision stumps) are used as the base classifiers. After training the weak learner on t -th iteration, the error of h_t is calculated by

$$e_t = \sum_{n: h_t(X(n)) \neq Y(n)} w_t(n). \quad (3)$$

Next, AdaBoost sets a parameter α_t . Intuitively, α_t measures the importance that is assigned to h_t . Then the weight w_t is updated.

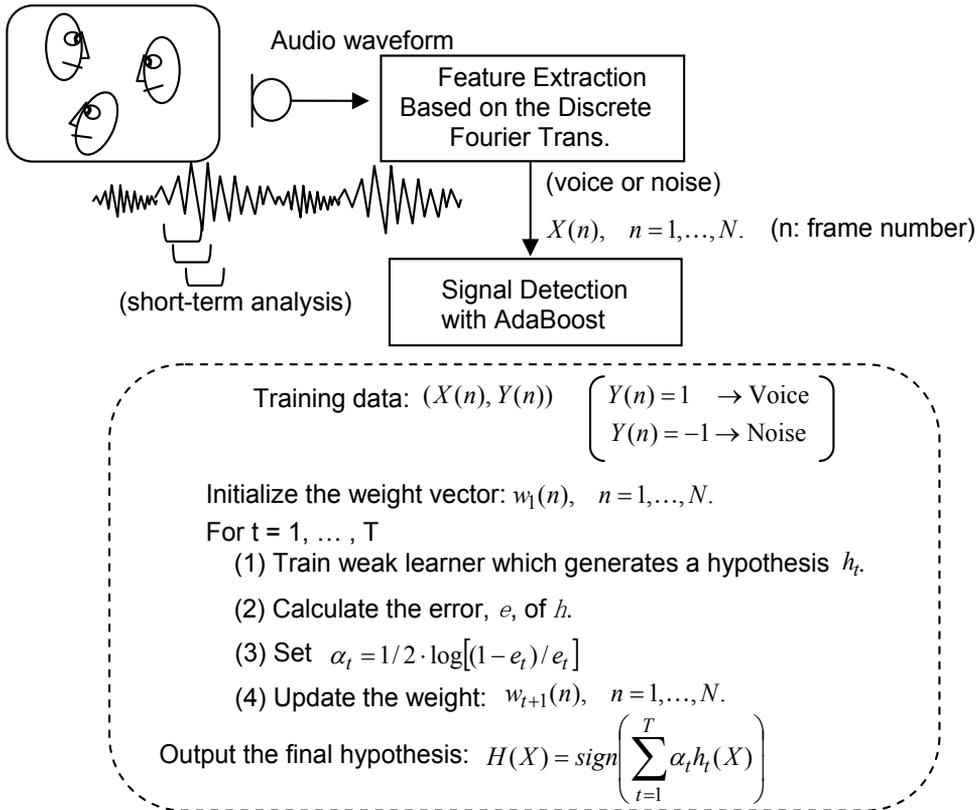


Fig. 2. Voice detection with AdaBoost

$$w_{t+1}(n) = \frac{w_t(n) \exp\{-\alpha_t \cdot Y(n) \cdot h_t(X(n))\}}{\sum_{n=1} w_t(n) \exp\{-\alpha_t \cdot Y(n) \cdot h_t(X(n))\}} \quad (4)$$

The equation (4) leads to the increase of the weight for the data misclassified by h_t . Therefore, the weight tends to concentrate on "hard" data. After T -th iteration, the final hypothesis, $H(X)$, combines the outputs of the T weak hypotheses using a weighted majority vote.

In home video environments, speech signals may be severely corrupted by noise because the person speaks far from the microphone. In such situations, the speech signal captured by the microphone will have a low SNR (signal-to-noise ratio) which leads to "hard" data. As the AdaBoost trains the weight, focusing on "hard" data, we can expect that it will achieve extremely high detection rates in low SNR situations. For example, in (Takiguchi et al, 2006), the proposed method has been evaluated on car environments, and the experimental results show an improved voice detection rate, compared to that of conventional detectors based on the GMM (Gaussian Mixture Model) in a car moving at highway speed (an SNR of 2 dB)

4. Estimation of sound source direction with CSP

The video editing system is requested to detect a person who is talking from among a group of persons. This section describes the estimation of the person's direction (horizontal

localization) from the voice. As the home video system may require a small computation resource due to its limitations in computing capability, the CSP (Crosspower-Spectrum Phase)-based technique (Omologo et al, 1996) has been implemented on the video-editing system for a real-time location system.

The crosspower-spectrum is computed through the short-term Fourier transform applied to windowed segments of the signal $x_i[t]$ received by the i -th microphone at time t

$$CS(n; \omega) = X_i(n; \omega)X_j^*(n; \omega) \quad (5)$$

where $*$ denotes the complex conjugate, n is the frame number, and ω is the spectral frequency. Then the normalized crosspower-spectrum is computed by

$$\phi(n; \omega) = \frac{X_i(n; \omega)X_j^*(n; \omega)}{|X_i(n; \omega)| |X_j^*(n; \omega)|} \quad (6)$$

that preserves only information about phase differences between x_i and x_j . Finally, the inverse Fourier transform is computed to obtain the time lag (delay) corresponding to the source direction.

$$C(n; l) = \text{InverseDFT}(\phi(n; \omega)) \quad (7)$$

Given the above representation, the source direction can be derived. If the sound source is non-moving, $C(n; l)$ should consist of a dominant straight line at the theoretical delay. In this chapter, the source direction has been estimated averaging angles corresponding to these delays. Therefore, a lag is given as follows:

$$\hat{l} = \arg \max_l \left\{ \sum_{n=1}^N C(n; l) \right\} \quad (8)$$

where N is the total frame in a voice interval which is estimated by AdaBoost. Figure 3 shows the overview of the sound source direction by CSP.

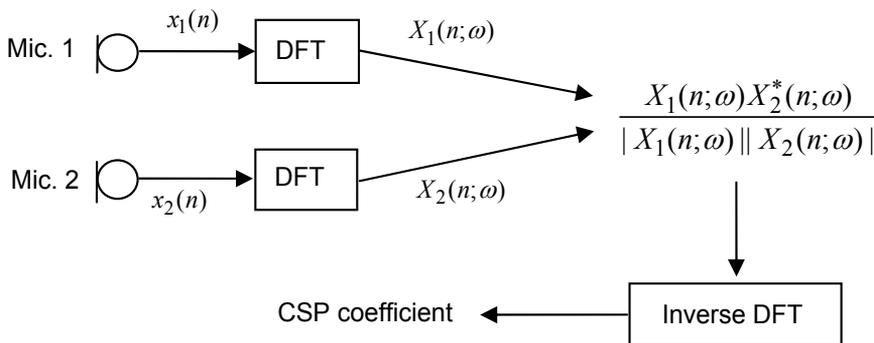


Fig. 3. Estimation of sound source direction using CSP

Figure 4 shows the CSP coefficients. The left is the result for a speaker direction of 60 degrees, and the right is that for two speakers' talking. As shown in Figure 4, the peak of the CSP coefficient (in the left figure) is about 60 degrees, where the speaker is located at 60 degrees.

When only one speaker is talking in a voice interval, the shape peak is obtained. However, plural speakers are talking in a voice interval, a sharp peak is not obtained as shown in the bottom figure. Therefore, we set a threshold, and the peak above the threshold is selected as the sound source direction. In the experiments, the threshold was set to 0.08. When the peak is below the threshold, a wide shot is taken.

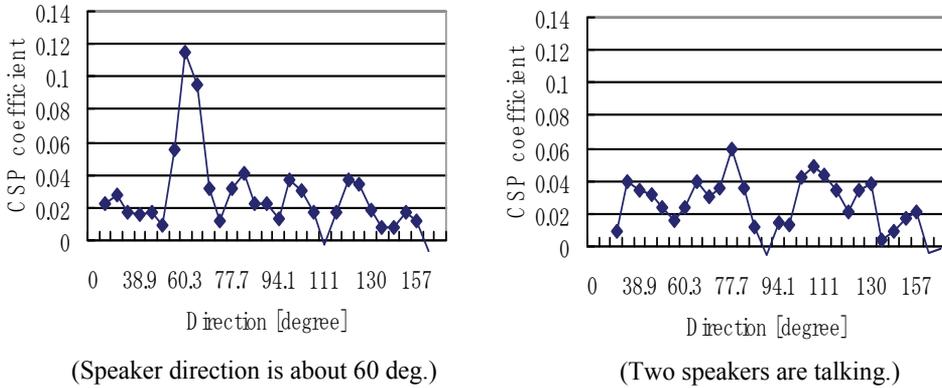


Fig. 4. CSP coefficients

5. Facial feature point extraction using EBG

To classify facial expressions correctly, facial feature points must be extracted precisely. From this viewpoint, Elastic Bunch Graph Matching (EBGM) (Wiskott et al, 1997) is employed in the system. EBG was proposed by Laurenz Wiskott and proved to be useful in facial feature point extraction and face recognition.

5.1 Gabor wavelets

Since Gabor wavelets are fundamental to EBG, it is described here. Gabor wavelets can extract global and local features by changing spatial frequency, and can extract features related to a wavelet's orientation.

Eq. (9) shows a Gabor Kernel used in Gabor wavelets. This function contains a Gaussian function for smoothing as well as a wave vector \vec{k}_j that indicates simple wave frequencies and orientations.

$$\varphi_j(\vec{x}) = \frac{\vec{k}_j^2}{\sigma^2} \exp\left(-\frac{\vec{k}_j^2 \vec{x}^2}{2\sigma^2}\right) \left[\exp(i\vec{k}_j \vec{x}) - \exp\left(-\frac{\sigma^2}{2}\right) \right] \tag{9}$$

$$\vec{k}_j = \begin{pmatrix} k_{jx} \\ k_{jy} \end{pmatrix} = \begin{pmatrix} k_v \cos \varphi_\mu \\ k_v \sin \varphi_\mu \end{pmatrix} \tag{10}$$

Here, $k_v = 2^{-\frac{\nu+2}{2}\pi}$, $\varphi_\mu = \mu \frac{\pi}{8}$. We employ a discrete set of 5 different frequencies, index $\nu = 0, \dots, 4$, and 8 orientations, index $\mu = 0, \dots, 7$.

5.2 Jet

A jet is a set of convolution coefficients obtained by applying Gabor kernels with different frequencies and orientations to a point in an image. To estimate the positions of facial feature points in an input image, jets in an input image are compared with jets in a facial model.

A jet is composed of 40 complex coefficients (5 frequencies * 8 orientations) and expressed as follows:

$$J_j = a_j \exp(i\phi_j) \quad j = 0, \dots, 39 \quad (11)$$

where $\bar{x} = (x, y)$, $a_j(\bar{x})$ and ϕ_j are the facial feature point coordinate, magnitude of complex coefficient, and phase of complex coefficient, which rotates the wavelet at its center, respectively.

5.3 Jet similarity

For the comparison of facial feature points between the facial model and the input image, the similarity is computed between jet set $\{J\}$ and $\{J'\}$. Locations of two jets are represented as \bar{x} and \bar{x}' . The difference between vector \bar{x} and vector \bar{x}' is given in Eq. (12).

$$\vec{d} = \bar{x} - \bar{x}' = \begin{pmatrix} dx \\ dy \end{pmatrix} \quad (12)$$

Here, let's consider the similarity of two jets in terms of the magnitude and phase of the jets as follows:

$$S_D(J, J') = \frac{\sum_j a_j a'_j \cos(\phi_j - \phi'_j)}{\sqrt{\sum_j a_j^2 \sum_j a'^2_j}} \quad (13)$$

where the phase difference $(\phi_j - \phi'_j)$ is qualitatively expressed as follows:

$$\phi_j - \phi'_j = \bar{k}_j \bar{x} - \bar{k}_j \bar{x}' = \bar{k}_j (\bar{x} - \bar{x}') = \bar{k}_j \vec{d} \quad (14)$$

To find the best similarity between $\{J\}$ and $\{J'\}$ using Eq. (13) and Eq. (14), phase difference is modified as $\phi_j - (\phi'_j + \bar{k}_j \vec{d})$ and Eq. (13) is rewritten as

$$S_D(J, J') = \frac{\sum_j a_j a'_j \cos(\phi_j - (\phi'_j + \bar{k}_j \vec{d}))}{\sqrt{\sum_j a_j^2 \sum_j a'^2_j}} \quad (15)$$

In order to find the optimal jet J' that is most similar to jet J , the best \vec{d} is estimated that will maximize similarity based not only upon phase but magnitude as well.

5.4 Displacement estimation

In Eq. (15), the best \vec{d} is estimated in this way. First, the similarity at zero displacement ($dx = dy = 0$) is estimated. Then the similarity of its North, East, South, and West neighbors is

estimated. The neighboring location with the highest similarity is chosen as the new center of the search. This process is iterated until none of the neighbors offers an improvement over the current location. The iteration is limited to 50 times at one facial feature point.

5.5 Facial feature points and face graph

In this chapter, facial feature points are defined as the 34 points shown in Fig. 5. A set of jets extracted at all facial feature points is called a face graph. Fig. 5 shows an example of a face graph.

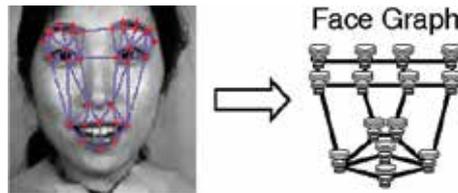


Fig. 5. Jet extracted from facial feature points

5.6 Bunch graph

A set of jets extracted from many people at one facial feature point is called a bunch. A graph constructed using bunches at all the facial feature points is called a bunch graph. In searching out the location of facial feature points, the similarity described in Eq. (15) is computed between the jets in the bunch graph and a jet at a point in an input image. The jet with the highest similarity, achieved by moving \vec{d} as described in Section 5.4, is chosen as the target facial feature point in the input image. In this way, using a bunch graph, the locations of the facial feature points can be searched allowing various variations. For example, a chin bunch may include jets from non-bearded chins as well as bearded chins, to cover the local changes. Therefore, it is necessary to train data using the facial data of various people in order to construct the bunch graph. The training data required for construction of bunch graph was manually collected.

5.7 Elastic bunch graph matching

Fig. 6 shows an elastic bunch graph matching flow. First, after a facial image is input into the system, a bunch graph is pasted to the image, and then a local search for the input face commences using the method described in Section 5.4. A set of jets extracted from many people at one facial feature point is called a bunch. Finally, the face graph is extracted after all the locations of the feature points are matched.

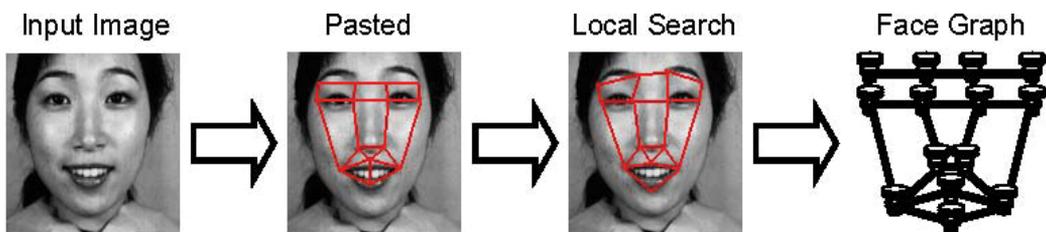


Fig. 6. Elastic Bunch Graph Matching procedure

6. Facial expression recognition using SVM

In this study, three facial expression classes are defined; “Neutral,” “Positive” and “Rejective”. Table 1 shows the class types and the meanings.

| Classes | Meanings |
|-----------|--|
| Neutral | Expressionless |
| Positive | Happiness, Laughter, Pleasure, etc. |
| Rejective | Watching other direction, Occluding part of face, Tilting the face, etc. |

Table 1. Facial expression classes

The users of our system register their neutral images as well as the personal facial expression classifier in advance. The differences between the viewer's facial feature points extracted by EBGM and the viewer's neutral facial feature points are computed as a feature vector for SVM. In our experiments, Radial Basis Function (RBF) is employed as a kernel function of SVM.

7. Camera work module

In the camera work module, the only one digital panning or zooming is controlled in a voice interval. The digital panning is performed on the HD image by moving the coordinates of the clipping window, and the digital zooming is performed by changing the size of the clipping window.

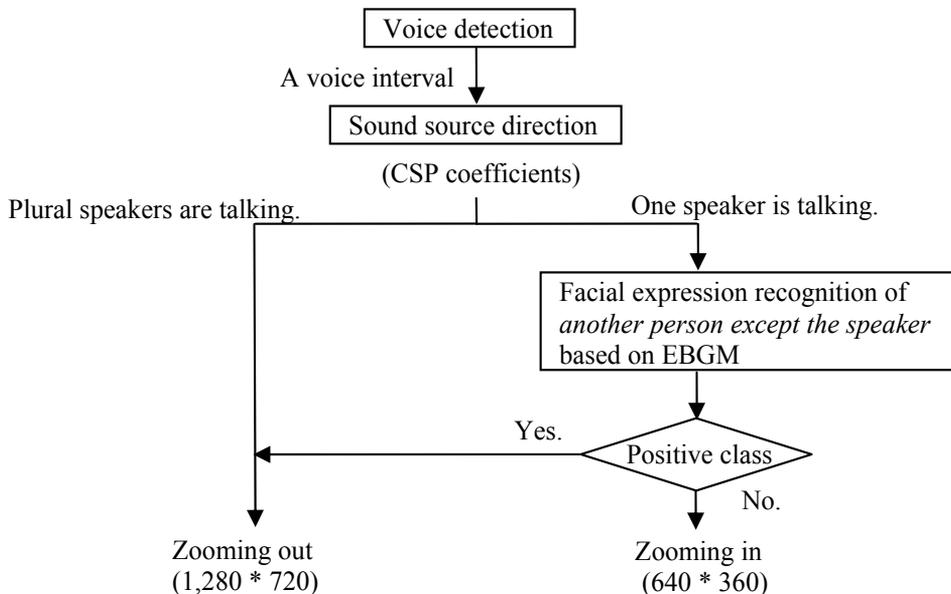


Fig. 7. Processing flow of digital zooming in and out

7.1 Zooming

Figure 7 shows the processing flow of the digital camera work (zooming in and out). After capturing a voice interval by AdaBoost, the sound source direction is estimated by CSP in order to zoom in on the talking person by clipping frames from videos.

As described in Section 4, we can estimate that one speaker is talking or plural speakers are talking in a voice interval. In the camera work, when plural speakers are talking, a wide shot (1280*720) is taken. On the other hand, when one speaker is talking in a voice interval, the system estimates the atmosphere of another person. When the atmosphere of another person is "positive" class, a wide shot (1280*720) is taken. When the atmosphere of another person except the speaker is not "positive" class, the digital camera work zooms in the speaker. In our experiments, the size of the clipping window (zooming in) is fixed to 640*360.

7.2 Clipping position (Panning)

The centroid of the clipping window is selected according to the face region estimated by using the OpenCV library. If the centroid of the clipping window is changing frequently in a voice interval, the video becomes not intelligible so that the centroid of the clipping window is fixed in a voice interval.

The face regions are detected within the 200 pixels of the sound source direction in a voice interval as shown in Figure 8. Then the average centroid is calculated in order to decide that of the clipping window.

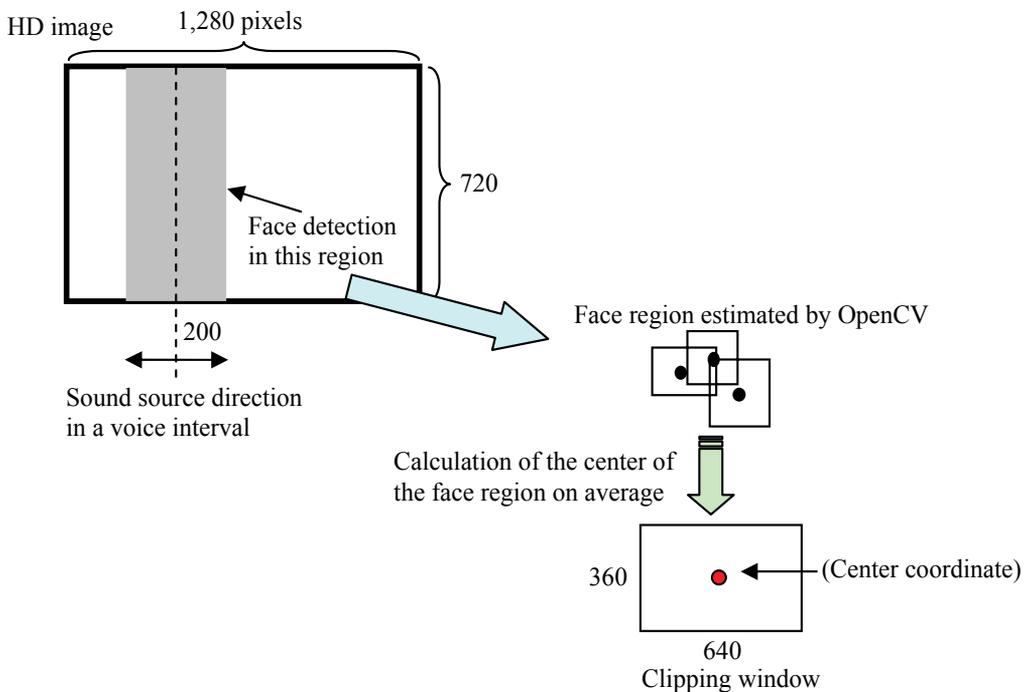


Fig. 8. Clipping window for zooming in

8. Experiments

8.1 Voice detection and sound source direction

Preliminary experiments were performed to test the voice detection algorithm and the CSP method in a room. Figure 9 shows the room used for the experiments, where a two-person conversation is recorded. The total recording time is about 303 seconds.

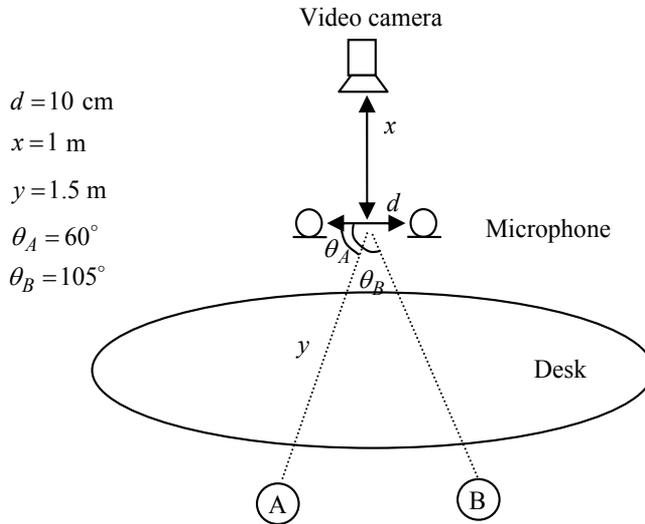


Fig. 9. Room used for the experiments. A two-person conversation is recorded.

In the experiments, we used a Victor GR-HD1 Hi-vision camera (1280*720). The focal length is 5.2 mm. The image format size is 2.735 mm (height), 4.864 mm (width) and 5.580 mm (diagonal). From these parameters, we can calculate the position of a pixel number corresponding to the sound source direction in order to clip frames from high-resolution images. (In the proposed method, we can calculate the horizontal localization only.)

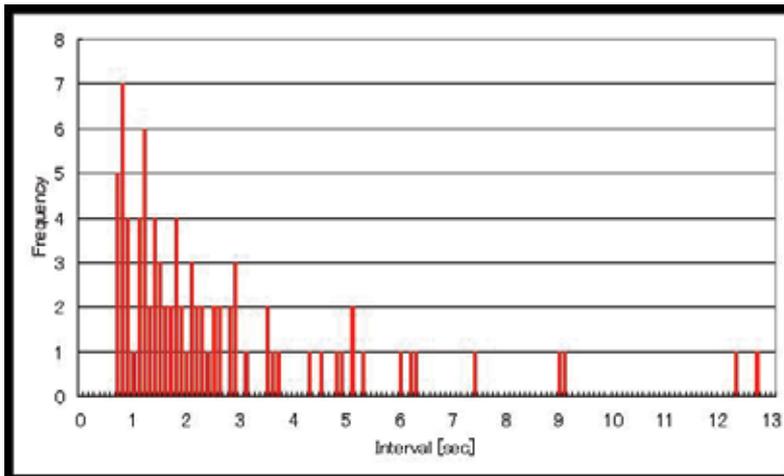


Fig. 10. Interval of conversation scene that was estimated by AdaBoost

Figure 10 shows the interval of the conversation scene that was estimated by AdaBoost. The max interval is 12.77 sec., and the minimum is 0.71 sec. The total number of conversation scenes detected by AdaBoost is 84 (223.9 sec), and the detection accuracy is 98.4%.

After capturing conversation scenes only, the sound source direction is estimated by CSP in order to zoom in on the talking person by clipping frames from videos. The clipping accuracy is 72.6% in this experiment. Some conversation scenes cause a decrease in the accuracy of clipping. This is because two speakers are talking in one voice (conversation) interval estimated by AdaBoost, and it is difficult to set the threshold of the CSP coefficient. Figure 11 shows an example of time sequence for zooming in and out.

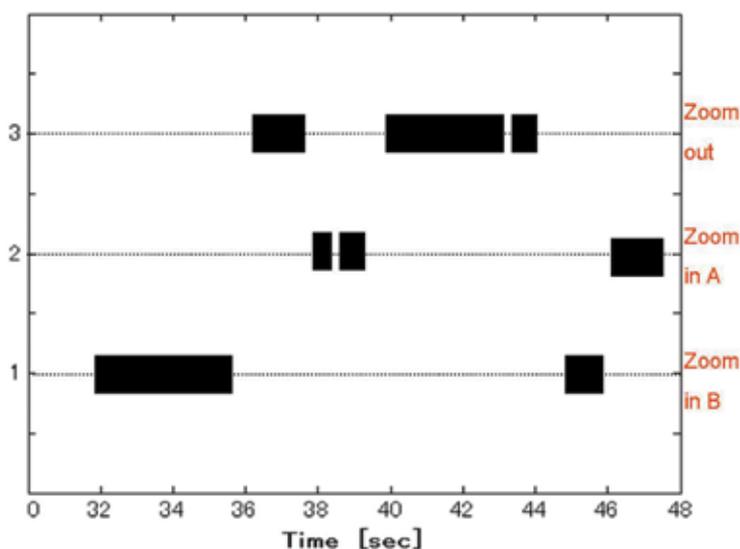


Fig. 11. Example of time sequence for zooming in and out

8.2 Facial expression recognition

| | Neutral | Positive | Rejective | Total |
|-----------|---------|----------|-----------|-------|
| Subject A | 3,823 | 2,517 | 153 | 6,493 |
| Subject B | 3,588 | 2,637 | 268 | 6,493 |

Table 2. Tagged results (frames)

Next, we tagged the video with three labels, “Neutral,” “Positive,” and “Rejective.” Tagged results for all conversational frames in the experimental videos are shown in Table 2.

Facial regions were extracted using AdaBoost based on Haar-like features (Viola et al, 2001) in all frames of conversation scenes except Reject frames. Extracted frames were checked manually to confirm whether they were false regions or not. The experimental results are shown in Table 3 and Table 4. The extraction rate of the facial region for subject B was not good, compared with that for subject A. The reason for the worse false extraction rate for subject B is attributed to his face in profile, where he often looks toward subject A.

| | Neutral | Positive |
|------------------|---------|----------|
| False extraction | 13 | 16 |
| Total frames | 3,823 | 2,517 |
| Error rate [%] | 0.34 | 0.64 |

Table 3. Experimental results of facial region extraction for subject A

| | Neutral | Positive |
|------------------|---------|----------|
| False extraction | 1,410 | 1,270 |
| Total frames | 3,588 | 3,719 |
| Error rate [%] | 39.3 | 48.2 |

Table 4. Experimental results of facial region extraction for subject B

For every frame in the conversation scene, facial expression was recognized by Support Vector Machines. The 100 frames for each subject were used for training data, and the rest for testing data. The experiment results were shown in Table 5 and Table 6.

| | Neutral | Positive | Rejective | Sum. | Recall [%] |
|---------------|---------|----------|-----------|-------|------------|
| Neutral | 3,028 | 431 | 364 | 3,823 | 79.2 |
| Positive | 230 | 2,023 | 264 | 2,517 | 80.4 |
| Rejective | 32 | 10 | 121 | 153 | 79.1 |
| Sum. | 3,280 | 2,464 | 749 | 6,493 | - |
| Precision [%] | 92.3 | 82.1 | 16.2 | - | - |

Table 5. Confusion matrix of facial expression recognition for subject A

| | Neutral | Positive | Rejective | Sum. | Recall [%] |
|---------------|---------|----------|-----------|-------|------------|
| Neutral | 1,543 | 214 | 1,831 | 3,588 | 43.0 |
| Positive | 194 | 1,040 | 1,403 | 2,637 | 39.4 |
| Rejective | 34 | 24 | 210 | 264 | 78.4 |
| Sum. | 1,771 | 1,278 | 3,444 | 6,493 | - |
| Precision [%] | 87.1 | 81.4 | 6.1 | - | - |

Table 6. Confusion matrix of facial expression recognition for subject B

The averaged recall rates for subject A and B were 79.57% and 53.6%, respectively, and the averaged precision rates for subject A and B were 63.53% and 58.2%, respectively. The result of the facial expression for subject B was lower than that for subject A because of the worse

false extraction rate of the facial region. Moreover, when the subjects had an intermediate facial expression, the system often made a mistake because one expression class was only assumed in a frame.

Figure 12 and Figure 13 show an example of the digital shooting (zooming in) and an example of zoom out, respectively. In this experiment, the clipping size is fixed to 640*360. In the future, we need to automatically select the size of the clipping window according to each situation, and subjective evaluation of video production will be described.

9. Conclusion

In this chapter, we investigated about home video editing based on audio with a two-channel (stereo) microphone and facial expression, where the video content is automatically recorded without a cameraman. In order to capture a talking person only, a novel voice/non-voice detection algorithm using AdaBoost, which can achieve extremely high detection rates in noisy environments, is used. In addition, the sound source direction is estimated by the CSP (Crosspower-Spectrum Phase) method in order to zoom in on the talking person by clipping frames from videos, where a two-channel (stereo) microphone is used to obtain information about time differences between the microphones. Also, we extract facial feature points by EBGM (Elastic Bunch Graph Matching) to estimate atmosphere class by SVM (Support Vector Machine). When the atmosphere of the other person except the speaker is not "positive" class, the digital camera work zooms in only the speaker. When the atmosphere of the other person is "positive" class, a wide shot is taken. Our proposed system can not only produce the video content but also retrieve the scene in the video content by utilizing the detected voice interval or information of a talking person as indices. To make the system more advanced, we will develop the sound source estimation and emotion recognition in future, and we will evaluate the proposed method on more test data.

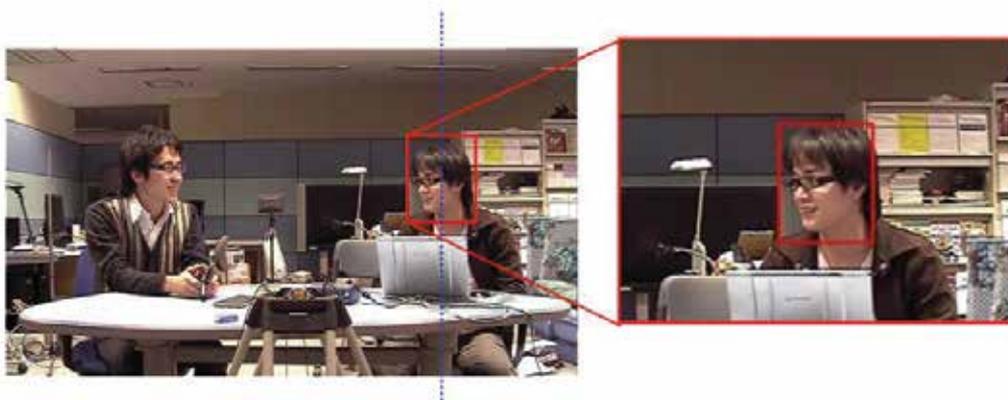


Fig. 12. Example of camera work (zooming in)



Fig. 13. Example of camera work (zoom out)

10. References

- Y. Ariki, S. Kubota, & M. Kumano (2006). Automatic production system of soccer sports video by digital camera work based on situation recognition, *Eight IEEE International Symposium on Multimedia (ISM)*, pp. 851-858, 2006.
- H. Sundaram & S.-F. Chang (2000). Video scene segmentation using audio and video features, *Proc. ICME*, pp. 1145-1148, 2000.
- K. Aizawa. Digitizing personal experiences: Capture and retrieval of life log, *Proc. Multimedia Modelling Conf.*, pp. 10-15, 2005.
- T. Amin, M. Zeytinoglu, L. Guan, & Q. Zhang. Interactive video retrieval using embedded audio content, *Proc. ICASSP*, pp. 449-452, 2004..
- F. Asano & J. Ogata. Detection and separation of speech events in meeting recordings, *Proc. Interspeech*, pp. 2586-2589, 2006.
- Y. Freund & R. E. Schapire. A short introduction to boosting, *Journal of Japanese Society for Artificial Intelligence*, 14(5), pp. 771-780, 1999.
- Y. Rui, A. Gupta, J. Grudin, & L. He. Automating lecture capture and broadcast: technology and videography, *ACM Multimedia Systems Journal*, pp. 3-15, 2004.
- M. Ozeke, Y. Nakamura, & Y. Ohta. Automated camerawork for capturing desktop presentations, *IEEProc.-Vis. Image Signal Process.*, 152(4), pp. 437-447, 2005.
- X.-S. Hua, L. Lu, & H.-J. Zhang. Optimization-based automated home video editing system, *IEEE Transactions on circuits and systems for video technology*, 14(5), pp.572-583, 2004.
- B. Adams & S. Venkatesh. Dynamic shot suggestion filtering for home video based on user performance, *ACM Int. Conf. on Multimedia*, pp. 363-366, 2005.
- P. Wu. A semi-automatic approach to detect highlights for home video annotation, *Proc. ICASSP*, pp. 957-960, 2004.
- M. Yamamoto, N. Nitta, N. Babaguchi, Estimating Intervals of Interest During TV Viewing for Automatic Personal Preference Acquisition. *Proceedings of The 7th IEEE Pacific-Rim Conference on Multimedia*, pp. 615-623, 2006.

-
- L. Wiskott, J.-M. Fellous, N. Kruger, & C. von der Malsburg, Face Recognition by Elastic Bunch Graph Matching, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(7), pp. 775-779, 1997.
- T. Takiguchi, H. Matsuda, & Y. Ariki. Speech detection using real AdaBoost in car environments, *Fourth Joint Meeting ASA and ASJ*, page 1pSC20, 2006.
- M. Omologo & P. Svaizer. Acoustic source location in noisy and reverberant environment using CSP analysis, *Proc. ICASSP*, pp. 921-924, 1996.
- P. Viola & M. Jones, Rapid object detection using a boosted cascade of simple features, *Proc. IEEE conf. on Computer Vision and Pattern Recognition*, pp. 1-9, 2001.

Combined Source and Channel Strategies for Optimized Video Communications

François-Xavier Coudoux^{1,2,3}, Patrick Corlay^{1,2,3},
Marie Zwingelstein-Colin^{1,2,3}, Mohamed Gharbi^{1,2,3},
Charlène Mouton-Goudemand^{1,2,3}, and Marc-Georges Gazelet^{1,2,3}

¹University Lille Nord de France, F-59000 Lille,

²UVHC, IEMN-DOAE, F-59313 Valenciennes,

³CNRS, UMR 8520, F-59650 Villeneuve d'Ascq,
France

1. Introduction

Digital video is becoming more and more popular with the wide deployment of multimedia applications and networks. In the actual context of Universal Media Access (UMA), one of the main challenges is to flexibly deliver video content with the best perceived image quality for end-users having different available resources, access technologies and terminal capabilities. In this chapter, we look in detail at the basic source and channel coding techniques for digital video communication systems, and show how they can be combined efficiently in order to fulfill the quality of service (QoS) constraints of video communication applications. The chapter includes several illustrative examples and references on the related topics.

The chapter begins with an overview of digital video compression basics, including MPEG-2 and H.264/AVC. We discuss the most common coding artifacts due to digital compression, and show how the compressed bitstream is made more sensitive to channel errors (Section 2). Since both compression and channel distortions affect the final perceived video quality, image quality metrics are needed in order to estimate the resulting visual quality. Both subjective and objective metrics are presented briefly and discussed in section 3.

The second half of the chapter concerns channel coding and error control for video communication (Section 4). The most common existing techniques are presented, with a focus on forward error correction (FEC) and the hierarchical modulation, used in the DVB standard, for example. We show that such channel coding schemes can be used to increase the error resilience of compressed video data. For example, scalable video coding can be combined with hierarchical modulation in order to allow the most important information in the compressed video bitstream to be transmitted with better protection against channel distortions. Section 5 explains this concept of scalability, and gives the great benefits of this coding tool for robust video transmission. Since all the various bits of the transmitted video data don't have the same level of importance, unequal error protection (UEP), with its different priority levels can be applied successfully.

For a given application, it is necessary to determine the best combination of lossy compression and channel encoding schemes in order to offer the optimal quality to the end-user. In section 6, we illustrate such a combined source and channel approach by presenting a quality-oriented study for a broadband video distribution system using digital subscriber lines (DSL). Our system allows the coverage area for a given DSL infrastructure to be extended, thus increasing the number of potential end-users. This system applies an adaptation mechanism that determines the “bottleneck” bit-rate at which the reconstructed video has the best quality. First, we detail the complete system architecture, and then we explain how to jointly determine the optimal source and channel coding parameters. Finally, we report our experimental results in order to demonstrate the effectiveness of our system in terms of extended coverage and optimal quality for a given eligibility level. In section 7, we conclude the chapter with a discussion of the current issues and the perspectives for future research on video communications.

2. Digital video compression basics

Digital video signals generally include some redundant information. For still images, the redundancy is spatial and is due to the important correlation between neighboring pixels. In order to reduce this redundancy, lossless data compression can be used, thus allowing a reconstructed image equal to original image to be obtained. The compression ratio obtained with this method is generally very low, close to 2 or 3. Lossy compression insures that a higher compression ratio will be obtained. The deterioration of the reconstructed image is a function of the compression rate, constituting a rate-quality trade-off. Like transmitting still images, transmitting a video stream also requires compressing the video data. This compression is made possible by the video stream's redundancy, both spatial (intra-image) and temporal (inter-image).

To eliminate the redundancy—or, in other words, the correlation—between two images, each image in a video stream is predicted in terms of the previous and/or following images, with only the prediction error being encoded. The first image in the stream (or the group of pictures described below) is always fully encoded without reference to the other images; this is the "Intra" mode. This encoding without reference facilitates the synchronisation of the receiver (i.e., the decoder). The images that follow can be predicted by motion compensation, and the prediction error can be then encoded before being transmitted with motion vectors.

During the encoding process, a video stream is split into a group of pictures (GOP) of a fixed size. The GOP contains:

- Intra-Picture (I): This is the first frame in each GOP. It is the reference, representing a still image, independent of other pictures.
- Inter-Picture (P or B): These frames contain the motion-compensated differences. A P-frame is the prediction based on a previous image, while a B-frame is the result of the encoding from two images, one previous and one following. An error in a predicted picture will propagate up to the final image in the GOP. Figure 1 shows the typical structure of a MPEG sequence.

In digital video broadcasting, the widely used video coding standards MPEG-2 (Mitchell et al., 1996) and, more recently, H264/MPEG-4 (Wiegang et al., 2003) are both based on a hybrid encoding method using transformation and motion compensation (Tekalp, 1996), as illustrated in Fig. 2.

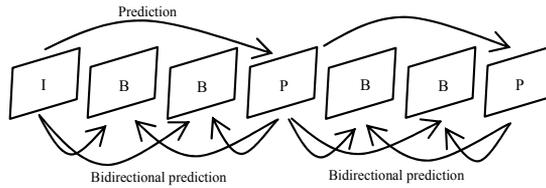


Fig. 1. Structure of MPEG sequence

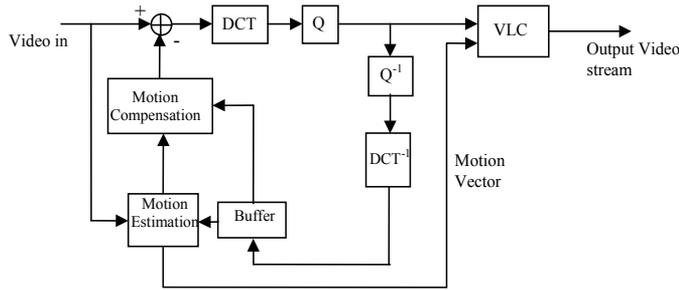


Fig. 2. Hybrid encoding method using transformation and motion compensation

Since the advent of the well-known JPEG algorithms (Pennebaker & Mitchell, 1993), the encoding process has consisted of several steps:

- *Conversion:* The image's color representation is converted into the Y (luminance) and Cr,Cb (chroma) components. The chroma resolution is generally reduced by a factor of 2 both horizontally and vertically.
- *Transformation:* The image is split into blocks, and each block (Intra-Picture or prediction residue) undergoes a discrete cosine transform (DCT). This transform exhibits excellent energy compaction for highly correlated images (Rabbani & Jones, 1991). DCT is independent of the signal to be encoded, and many fast DCT computation algorithms exist. A DCT applied on blocks of 8x8 pixels generates 64 coefficients. The first coefficient represents the constant value (DC), and the others represent waveforms at gradually increasing frequencies. In the most recent H.264 compression standard, the image can be decomposed into blocks of different sizes to adapt to local image statistics, and thus increase encoding efficiency.
- *Quantization:* The amplitudes of the frequency components generated in the previous step are quantized by removing small values. Quantization, which plays a major part in lossy compression, reduces the amount of data needed to represent an image. Since human eye is more sensible to errors in low frequency compared with high frequency (Glenn, 1993), each DCT frequency coefficient is quantized with an adequate step.
- *Scanning:* The quantized DCT coefficients are subjected to zigzag scanning, which arranges the DCT coefficients in order of increasing frequency.
- *Compression:* The resulting data is further compressed using entropic encoding, which is a form of lossless data compression. The entropy encoders compress data by replacing each value, defined by a fixed length, with a corresponding variable length codeword. Since the length of each codeword is an inverse function of its appearance probability, the most common value is represented by the shortest codeword. In JPEG and MPEG2, the entropic encoding is always Huffman encoding.

Typically, in the above encoding process, every 12th frame is an I-frame, and the GOP contains the image string, IBBPBBPBBPBB. The (I) intra-coded frame is split into block of $8 * 8$ pixels, and each block is coded independently of the others, using the above encoding process. An Inter-frame (P or B) is divided into $16*16$ pixel macro blocks. For each macro block, the motion is estimated in terms of the reference frame, and then the estimation error is compressed.

The H.264/AVC standard allows a compression ratio equal to twice MPEG-2 ratio to be obtained. The H.264/AVC standard is similar to the MPEG-2 standard, but with some rather important differences (Richardson, 2003):

- The intra-coded blocks are predicted in terms of the pixels located above and to the left (causal neighborhood) of those that have previously been encoded and reconstructed. The error prediction is then encoded.
- A deblocking filter is applied to blocks in a decoded video to improve the subjective visual quality.
- The motion is compensated by accounting for different block sizes ($16*16$, $16*8$, $8*16$, $8*8$, $8*4$, $4*8$ and $4*4$). Using previously-encoded frames as references is much more flexible than past standards, and the precision of the motion compensation is equal to quarter pixel.
- The entropy encoding is enhanced by providing Context-adaptive Binary Arithmetic Coding or Context-adaptive Variable-length Coding for residual data and Exponential-Golomb Coding for many of the syntax elements.

But... what about image quality?

We mentioned in the previous sub-section that digital video compression algorithms use lossy quantization in order to achieve high compression ratios. This quantization results in various kinds of coding artifacts, which may greatly affect the visual quality of the reconstructed video signal, especially for low bit-rate coding. The compression artifacts and their visual significance have been widely studied in the literature. Yuen (1998) provides a comprehensive classification and analysis of most coding artifacts in digital video, compressed using MC/DPCM/block-based DCT hybrid coding methods. In particular, this author shows how the visual impact of coding artifacts is strongly related to the spatial and temporal characteristics, both local and global, of the video sequence, as well as the properties of the human visual system (HVS).

Among the various coding artifacts, *blocking effect* (also called *blockiness*) is the most well-known distortion introduced by video compression algorithms. The blocking effect manifests itself as a discontinuity located at boundaries between adjacent blocks. This spurious phenomenon is due to the fact that common block-based compression algorithms encode adjacent blocks as independent units without taking into account the correlation that exists between them. Hence, at the decoding stage, the quantization error differs from one block to another, resulting in inter-block discontinuities. Figure 3 illustrates this particular coding distortion, which is clearly visible on the face and the building in the background.

This blocking effect is very annoying and mainly affects the visual perception of end-users. The perceptual relevance of this distortion is strongly related to HVS sensitivity: the regular geometric spacing of blocks, the specific horizontal or vertical alignment of block edges, and the spatial frequency of repeated blocks are highly apparent to the human eye. Because of its visual prominence, several methods have been proposed in order to reduce the visibility of the blocking effect in compressed images or sequences. (See the post-processing algorithms

proposed by Ramamurthi (1986), Lee (1998) or Coudoux (2001), for example.) Recently, the H.264/AVC codec has introduced a deblocking filter as a standardized tool to reduce the visibility of this coding artifact.



Fig. 3. Illustrative example of the blocking effect (*Foreman* sequence, MPEG-2, QCIF@128Kbps)

Unfortunately, digital image and video impairments are not restricted to coding artifacts, since errors may also occur when transmitting compressed video bitstreams over a noisy channel. Once the video sequence has been compressed by the encoder, the resulting bitstream is typically packetized in the network adaptation layer using transport protocols, such as ATM or TCP/IP, and then the packets are sent over the transmission network.

Different types of impairments can occur in transmissions over noisy channels: packets may be corrupted, or they may be affected by extensive delays that are incompatible with video applications. In all cases, erroneous packets are considered to be lost and are not available for decoding. Depending on the packet size, this loss may corrupt a limited part of a decoded picture, the entire picture or, in the worst case, a complete group of pictures. In the latter case, error concealment techniques can be used at the decoding stage to limit the visual impact of channel errors. In addition, error control mechanisms may fail at the transport level. In this case, errors occur when decoding at the application level. The wide variety of configurations leads to very different visual distortions depending on the kind of corrupted data. Figure 4 shows examples of localized false macroblocks, erroneous or misaligned slices, and frozen parts of pictures. For example, an erroneous VLC may be decoded as a false value and will subsequently result in a localized distortion in the display.



Fig. 4. Examples of visual impairments due to transmission errors (*Foreman* sequence, MPEG-2, QCIF@128Kbps, BER = 10^{-4})

Visual impairments due to transmission errors generally have a much more severe effect on end-user quality, compared to compression artifacts. In particular, the use of compression techniques based on spatial (e.g., differential encoding of DC coefficient) and temporal (e.g., MC) prediction makes the compressed bitstream very sensitive to channel errors, due to

spatial or temporal loss propagation of the corrupted data up to the next synchronization point (e.g., end of slice/intra-frame).

3. Image quality metrics

Digital video quality obviously constitutes one of the key points for any video service to insure a satisfying level of quality for the end-user. For this reason, it is crucial for researchers, broadcasters and network providers to be able to reliably assess and control the perceived video quality. There are two main approaches to image and video quality assessment: the first one relies on subjective evaluation metrics, and the second one is based on objective quality metrics. These two approaches are described in the following subsections.

3.1 Subjective metrics

The best way to measure image quality as perceived by a human observer is to use the subjective viewing experience of human observers themselves. The International Telecommunication Union (ITU) has developed and standardized several subjective methods that provide reliable test conditions and measurements. The ITU recommendations, ITU-R Rec. BT.500-11 and ITU-T Rec. P.910, specify:

- the test conditions (e.g., viewing distance, observer selection process, test material),
- the evaluation procedure (e.g., single vs. double stimuli, type of rating scale), and
- the methods for exploiting the data collected (e.g., statistical tools used for accurate analysis of viewers scores).

For example, ITU-R Rec. BT500-11 specifies the Double Stimulus Impairment Scale (DSIS) method shown in Figure 5. With this method, the reference and the test sequence are shown only once, and the observers have to evaluate the corresponding impairment using a five-level impairment scale. A mean opinion score (MOS) is obtained by averaging the evaluations/scores of all observers. More details on subjective video quality assessment are available in the reference books by Wu (2005) and Winkler (2005).

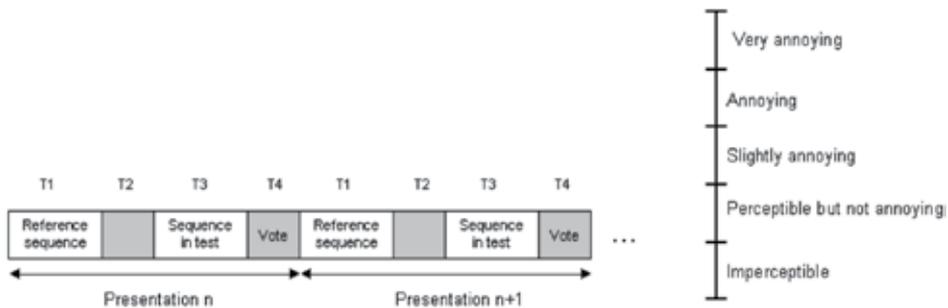


Fig. 5. DSIS method: a) presentation sequence; b) five-level impairment scale

Without any doubt, subjective visual quality assessment is the most correlated with human perception. However, this approach has several drawbacks: it is complex, time-consuming and expensive. Thus, alternative solutions relying on objective metrics are often preferred to such subjective methods. In practice, objective metrics can be used at different places in the broadcasting chain to assess and monitor video quality. They can also be used to optimize the different components of a video communication system, as illustrated later in Section 6.

Nonetheless, in order to be meaningful, objective metrics should be well correlated with the results obtained from the subjective methods.

3.2 Objective metrics

Objective metrics for assessing image and video quality can be divided into three distinct groups:

- signal-based methods,
- structural methods, and
- HVS-based methods.

The most common signal-based method is the Peak Signal-to-Noise Ratio (PSNR), expressed in decibels (dB), and defined as:

$$PSNR(dB) = 10 \cdot \log_{10} (d^2 / MSE) \quad (1)$$

where $d = 255$ for 8-bit encoded picture, and MSE is the mean squared error, defined as:

$$MSE = \frac{1}{M \cdot N \cdot T} \sum_{i=1}^M \sum_{j=1}^N \sum_{k=1}^T (f_0(i, j, k) - f_d(i, j, k))^2 \quad (2)$$

where f_0 and f_d are the pictures from the original and distorted sequence, respectively. Parameters M and N represent the size (in pixels) of each picture, the video sequence being made up of T pictures. Unfortunately, this objective metric has been shown to be poorly correlated with human visual assessment (Girod, 1993). PSNR-like metrics do not translate the visibility of the distortions, which depend on picture content, the location of the distortions and the HVS masking properties. For example, the two images presented below have the same PSNR numerical value, although the one located on the left side clearly exhibits worst visual quality (especially due to strong blocking artifacts).



Fig. 6. Image compressed by the JPEG algorithm using high-frequency emphasis (left) and Standard Y (right) quantization tables. Both images have the same PSNR: 30.7 dB.

Recently, the Structural SIMilarity (SSIM) method was proposed by Wang et al. (Wang, 2004a). Instead of measuring a simple pixel-by-pixel difference like MSE does, SSIM relies on the fact that human perception depends greatly on the presence of structures inside the video scene: any change in these structures will affect the way that a human observer will perceive the scene. Practically, SSIM is locally computed as a combination of three similarity indicators, one for luminance (based on local means), one for contrast (based on local variance), and one for structural information (based on local covariance). The SSIM values are averaged over the entire image in order to give a single SSIM index by picture. In the case of video sequences, a global note is computed by appropriate weighted averaging of all

the pictures in the sequence, as described by Wang et al. (Wang, 2004b). These authors claimed that the SSIM index gives very satisfying correlation results with respect to subjective assessment methods.

Finally, the most preferred methods model Human Visual System (HVS) properties and then integrate them into the objective metric. There has been plenty of research done on HVS-based objective quality metrics since the beginning of 1970s (e.g., Mannos, 1974; Faugeras, 1979; Lukas, 1982). Such methods can be divided into 2 groups:

- *single-channel methods* -- in single-channel methods, the HVS is modelled as a single spatial filter and is characterized by its Contrast Sensitivity Function (CSF). The final metric is generally obtained by weighting the error signal based on HVS sensitivity.
- *multi-channel methods* -- multi-channel methods are more complex because they assume that the visual signal is processed by separate spatial frequency channels. In the past, several multi-channel vision models have been developed by researchers, including Daly (1993) and Watson (2001), for example. These models generally integrate a more complete representation of significant visual phenomena, such as spatiotemporal masking or orientation selectivity. The global score is typically obtained by error pooling using Minkowski summation. Though more complex, these quality metrics provide better prediction accuracy.

Created in 1997, the Video Quality Expert Group (VQEG) has for the last ten years been evaluating the capacity of various video quality metrics to predict subjective quality ratings, as measured by MOS. Many vision-based video quality metrics have been shown to achieve very good correlation scores with MOS, outperforming the PSNR method. In particular, the so-called Video Quality Metric (VQM) has been recently included in two international Recommendations (Pinson, 2004). Nevertheless, for the moment, no method has ever been judged optimal and subjective rating remains the only reference method for accurately evaluating video quality.

4. Channel coding and error control for video communications

Once a digital video has been appropriately coded and compressed (as described in the previous sections), the concern becomes how to reliably transmit this video over a transmission medium—the channel—that by its nature deteriorates the signal quality. This section presents and analyzes the techniques used to improve digital video transmission quality.

For a given compression algorithm, compressed video quality is clearly a monotone increasing function of the bit-rate: the higher the bit-rate, the higher the video quality. In digital communications, transmission quality is measured as the probability of one bit being flipped during the transmission process – a received “1” corresponding to a transmitted “0” and vice-versa. This probability is referred to as the Bit Error Rate (BER). The BER is a monotone decreasing function of the Signal-to-Noise Ratio (SNR) of the transmission, and a monotone increasing function of the bit-rate. Thus, the higher the video bit-rate, the higher the video quality before transmission but also the higher the BER of the transmission (which results in poor quality of the received video). Thus, a good compromise has to be found for the bit-rate so that the overall video quality deterioration (i.e., the deterioration inherent to the compression algorithm and degradation inherent to the transmission) is acceptable.

In the context of digital video transmission—not to be confused with digital communications in general—the main question is "What are the properties that are specific

to video in terms of its transmission?". One answer is that all the bits in a digital video bitstream do not have the same importance in terms of video quality, which means that the consequences of transmission failure with respect to the video quality can change dramatically depending on which bit(s) in the bitstream fail.

A few examples

- In Pulse Code Modulation (PCM) source coding, each video sample is quantified and binary coded. Clearly, the loss of the least significant bits (LSB) will only slightly degrade its quality, whereas the loss of the most significant bits (MSB) will lead to totally erroneous reconstructed video signal. Thus, the most significant bits are of greater importance than the least significant bits in terms of video transmission.
- In the Discrete Cosine Transform (DCT) coding used for MPEG formats, high-frequency coefficients generally correspond to fine granularity details in the images, whereas low-frequency coefficients correspond to the structure of the images (Richardson, 2003). Thus, in terms of video transmission, low-frequency coefficients can be considered to be of greater importance than high-frequency coefficients.
- In the packetization and framing process inherent to every digital system, headers are added to the video information in order to insure its proper progress in the network, as well as its decoding at the destination. These headers are of crucial importance, since their loss would involve a transmission or decoding failure of a complete packet of video data (corresponding to from a part of an image to several consecutive images). Thus, in terms of video transmission, headers are of greater importance than the other bits transmitted.

Aware of that some bits are more important than others, recent video compression standards (Richardson, 2003) generally make it possible to partition these bits into several different bitstreams, generally 2 or 3. This means that, from the transmission system design perspective, which is the one covered in this section, the idea is to globally optimize the quality of a video transmission by:

- Partitioning the bits transmitted into several bitstreams of different importance, if this has not already been done at the source coding level (possible with the MPEG-2 or H.264 formats – See Section 5) , then
- Providing a way to give different transmission BERs to these bitstreams, where the values of the BERs shall be adapted to the relative importance of the bitstreams. In the following, we use the term *Unequal Error Protection (UEP)* to speak of the transmission techniques that provide different BERs to different bitstreams in a single video sequence.

UEP transmission techniques can have several different effects. For example, they can either improve the video quality of a given user or, for a fixed video quality, extend the range of users eligible for a video service. For example, in a Asymmetric Digital Subscriber Line (ADSL) environment, the UEP techniques would allow users with a poor-quality subscriber line to benefit video services, which would have been impossible in a traditional transmission environment in which all bits are transmitted with the same BER.

Another desirable feature of UEP digital video transmission systems are their ability to provide a simulcast of different video formats—for example, the Standard Definition (SD) and High Definition (HD) formats. This allows a graceful degradation of the video quality as the channel quality degrades, which is frequently the case in mobile communications. It is also an elegant means of adapting the video quality to the receiver's screen definition level.

Although UEP techniques for the different video bitstreams can be considered at the different layers of the Open System Interconnection (OSI) model, we focus on the techniques that do it at the physical layer. This section is organized as follows. In section 4.1, we review the basic modulation principles, which are used to map digital information onto an analog waveform constituted of a sequence of successive symbols. We also examine how the BER is related to the transmission link's signal-to-noise ratio, as well as to the number of information bits conveyed by each symbol. In section 4.2, we focus on UEP modulation techniques. Two independent types of modulation are presented: hierarchical modulation (e.g., Hierarchical Quadrature Amplitude Modulation (HQAM)) and multi-carrier modulation (e.g., Orthogonal frequency-division multiplexing (OFDM)). We also evoke the possibility of using the two types in combination (e.g., HQAM-OFDM). Since the BER performance achieved by any modulation scheme is generally insufficient in terms of the application requirements, forward error correction (FEC) techniques are usually necessary to insure an acceptable quality of service. In section 4.3, we introduce the FEC principles, as well as the UEP techniques that operate at the FEC layer.

4.1 Modulation basics

This subsection does not aim to cover the subject of modulation exhaustively like (Proakis, 1995) did, but rather to provide the essential information needed to understand the techniques developed in the subsequent subsections.

Modulation is the operation that makes the link between digital information – a sequence of bits at a given bit-rate $1/T_b$, where T_b is the duration of one bit – and an analog waveform appropriate for transmission over a channel. Generally, bits are not transmitted one at a time, but are rather grouped in small quantities (typically 2 to 10 bits), called symbols, which are transmitted sequentially. For a modulation scheme that maps b bits into one symbol, the symbol duration will be $T=b \times T_b$, resulting in a symbol rate of $1/T$ bauds.

As an example, let us consider the rectangular Quadrature Amplitude Modulation with eight modulation levels (8-QAM). A temporal representation of the modulated signal appears in Figure 7a. In this figure, the symbols are formed of $\log_2(8) = 3$ bits. These bits are used to code the amplitudes of two orthogonal waves of the same frequency, or to code the amplitude A and the phase π of a single wave, since: $x \cos(2\pi f_0 t) + y \sin(2\pi f_0 t) = A \cos(2\pi f_0 t + \Theta)$, where: $A = \sqrt{x^2 + y^2}$, and: $\Theta = \arctan(y/x)$.

A signal-space representation of the signal, called the *constellation*, is a useful visualization tool that facilitates the analytical evaluation of the modulation scheme's performance (Figure 7b). A constellation represents the set of possible values for the amplitude and phase of the sinusoid in the complex plan - 8 points for the 8-QAM constellation. In Figure 7b, the point of the constellation corresponding to each symbol of the transmitted waveform is highlighted in red.

This notion of *symbol* leads to defining a second transmission quality parameter, which is of great practical importance: Symbol Error Rate (SER), or the probability that a symbol will be received erroneously. The relationship between the SER and the BER depends on the way the bits are mapped on each symbol. Generally, this mapping is optimized in order that an erroneous symbol results in no more than one erroneous bit (Gray mapping). Such general cases yield the relation $BER = SER/b$, where b is the number of bits per symbol. Figure 8 illustrates a Karnaugh-style Gray map for the 8-QAM constellation. If this map was used to

construct the modulated signal of Figure 7a, the corresponding transmitted information binary sequence would be ...011110000...

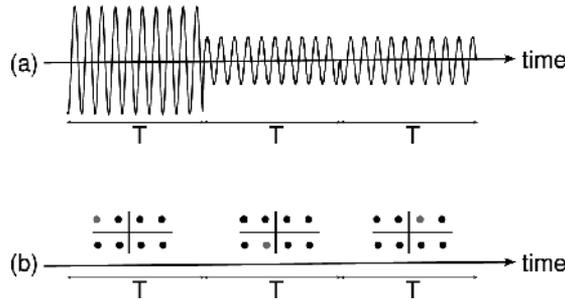


Fig. 7. Different representations of a 8-QAM modulated signal: a) temporal representation; b) signal-space representation (i.e., a constellation)

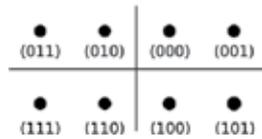


Fig. 8. Karnaugh-style Gray map of a 8-QAM constellation.

For general QAM transmissions over the additive white Gaussian noise channel, the SER can be shown to be related to the distance d between nearest points in the constellation to the noise standard deviation ratio σ , or to the SNR channel and to the number of bits per symbol. This relationship is expressed by the equation (Proakis, 1995):

$SER = 4\beta Q(\alpha)(1 - Q(\alpha))$, where: $\alpha = \sqrt{3SNR/(2^b - 1)}$ and: $\beta = 1 - 2^{-b/2}$. In this equation, the β term represents the average number or nearest neighbors of a constellation point. Figure 9 presents the SER performance of several classical QAM modulations.

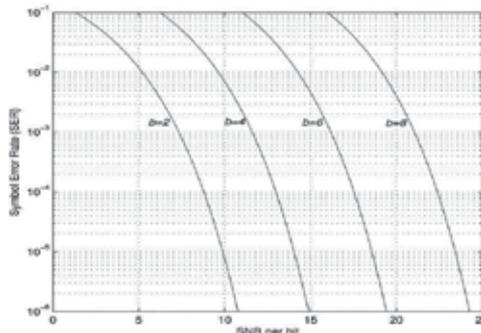


Fig. 9. SER performance of several classic QAM modulations

4.2 Achieving unequal error protection at the modulation level: examples of HQAM, OFDM and HQAM-OFDM combinations

Several possible methods for achieving unequal error protection during modulation are described in this subsection. One possibility is to use hierarchical modulation methods, such as HQAM. The idea of hierarchical modulation comes from the fact that the SER is directly

related to the distance between nearest points in the constellation. Figure 10 illustrates the principle, using the example of a two-level 4/8-HQAM and its associated Karnaugh-style Gray map of two highly important bits and two less important bits. As the figure shows, the points are grouped into four clouds, where the distance between two nearest points in a cloud is d_2 and the distance between clouds is $d_1 > d_2$. The map indicates that the least significant bits serve to differentiate between the four points within each cloud and that the two most significant bits serve to differentiate between the clouds. Hence, the BER for the least significant bits is directly related to the distance d_2 . For the two most significant bits, the constellation can also be seen as a 4-QAM constellation with distance between points equal to $d_1 + d_2/2$, where $d_2/2$ can be seen as a noise component since it contains no information about these most significant bits.

The ratio $\lambda = d_2/d_1$ controls the bits' relative priority. When $\lambda = 0$, the result is a uniform 4-QAM (the two least significant bits are merely discarded); when $\lambda = 1$, the result is a uniform 8-QAM (equal priority for all bits). When $0 < \lambda < 1$, the most significant bits have a greater priority.

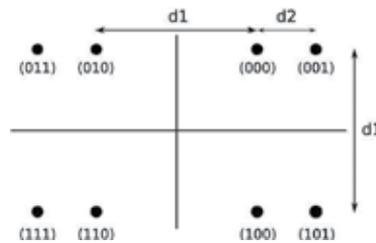


Fig. 10. 8-HQAM constellation and Karnaugh-style Gray map for two most significant bits and one least significant bit

BER analytical expressions have been formulated by Vitthaladevuni (2001) for the class of 4/M-HQAM constellations, in which the number of clouds is 4 (i.e., $\log_2(4) = 2$ most significant bits) and the number of points within each cloud is M (i.e., $\log_2(M)$ least significant bits). Clearly, the BER expressions are related to the ratio d_2/d_1 , which, in practice, must be chosen appropriately according to the requirements of the video bitstreams. For the 4/16-HQAM constellations, these BERs can be approximated as:

$$BER_{most_significant} = \frac{1}{4} \left(\operatorname{erfc} \left(\frac{d_1}{\sqrt{N_0}} \right) + \operatorname{erfc} \left(\frac{d_1 + 2d_2}{\sqrt{N_0}} \right) \right) \tag{3}$$

$$BER_{less_significant} = \frac{1}{4} \left(\operatorname{erfc} \left(\frac{d_2}{\sqrt{N_0}} \right) + \operatorname{erfc} \left(\frac{2d_1 + d_2}{\sqrt{N_0}} \right) - \operatorname{erfc} \left(\frac{2d_1 + 3d_2}{\sqrt{N_0}} \right) \right)$$

Figure 11 illustrates these BERs for a 4/16-HQAM constellation as a function of the SNR, for a ratio of $\lambda = d_2/d_1 = 0.8$.

A second possibility for achieving unequal error protection during modulation is based on multi-carrier modulation methods, such as DMT or OFDM. These methods are used extensively in current digital transmission systems (e.g., DVB, ADSL, Wimax, LTE). In DMT and OFDM, a Quadrature Amplitude Modulation (QAM) is carried out on a sub-channel, using IFFT/FFT processing (Starr, 1999). For a multi-carrier symbol rate $1/T$, the spacing between the sub-channels is also $1/T$, and the overall bandwidth is N/T , where N is half the

size of the IFFT/FFT used for processing. One of the key issues in designing efficient DMT systems is the bit-loading algorithm that optimizes the bit and power allocation over the QAM sub-channels, based on their power gains and noise levels.

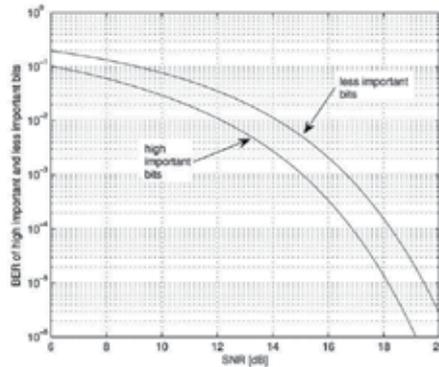


Fig. 11. Approximate BERs of a 4/16-HQAM constellation as a function of the SNR ($\lambda=d_2/d_1=0.8$)

Due to their multi-carrier structure, DMT and OFDM are inherently adapted to provide Frequency Division Multiplexing (FDM) of the different bitstreams. Thus, a natural approach for providing UEP is to multiplex the different streams using FDM, adapting the bit-loading algorithm so that it will provide different BERs for the different bitstreams. Generally, these algorithms assign the “best” sub-carriers to the bits of higher importance, and the other ones to the less important bits, as is suggested by the theoretical water-filling analysis proposed by Starr (1999). Such approaches have been studied by Zheng (2000) and Goudemand (2006), who point out different quality/complexity trade-offs. Figure 12 shows the bit-allocation produced by Zheng's algorithm (2000) for unequal error protection in an ADSL environment, using the European Telecommunications Standards Institute's loop 2 channel model (ETSI, 1996) and two video bitstreams from a data-partitioning MPEG-2 video coder at 6 Mbps (3.6 Mbps for the highly important bitstream and 2.4 Mbps for the less important one). The BERs are 10^{-7} and 10^{-4} respectively.

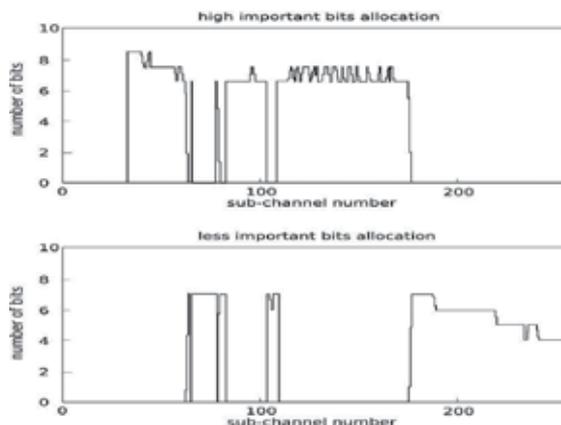


Fig. 12. Bit-allocation result for FDM unequal error protection based on DMT modulation (Goudemand, 2006)

Another possibility for achieving unequal error protection on multi-carrier transmission systems using DMT and OFDM is to combine the two methods described above. This method, called *hierarchical multi-carrier modulation*, involves modulating each sub-carrier using HQAM so that, unlike the pure FDM approach, each sub-channel carries both important and less important bits with embedded unequal error protection. The bit and power allocation over hierarchical multi-carrier modulation minimizes the total transmitted power while maintaining a constant bit-rate and BER for each bitstream. In practice, the bit and power allocation is calculated in two steps, providing two-level unequal error protection. The first step allocates higher priority bits and their corresponding power, and the second one allocates the remaining less important bits (Goudemand1, 2006).

As usual, the bit-loading algorithm must be computationally efficient and must reflect the water-filling principle, in that more bits must be allocated to the sub-channels with the lowest noise levels. Figure 13 shows a typical bit allocation in a hierarchical multi-carrier modulation system.

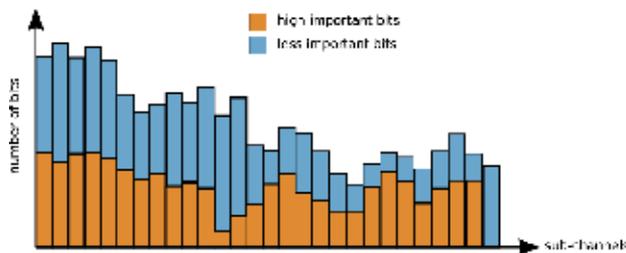


Fig. 13. A typical bit allocation for a hierarchical multi-carrier system

4.3 Achieving unequal error protection at the FEC level

Subsection 4.1 presented our analysis of SER and BER performance in general QAM modulation. This performance is generally insufficient for the needs of video communication. It is thus interesting to compare the bit-rate of a QAM modulation (b bits/symbol) to the maximum bit-rate that could theoretically be transmitted reliably (i.e., at a BER as close to zero as desired) at the same SNR. This maximum bit-rate is known as the channel capacity: $c = \log_2(1 + \text{SNR})$ bits/symbol (Shannon, 1948). Figure 14 facilitates this comparison. In the figure, the SER is plotted as a function of the ratio bit-rate/channel capacity $= b/c$ for different channel capacity values.

Clearly, satisfactory values of the BER for video transmission (below 10^{-4}) can be achieved at a rate that is very far from the channel capacity. In order to improve transmission efficiency (i.e., in order to achieve sufficiently low BER at rates close to the channel capacity), error correcting codes have to be used. These error correcting codes operate by adding controlled redundancy to the information bits so that only certain bit patterns can be transmitted. The error correcting decoder, which is aware of these patterns, will then be able to detect and even correct some erroneous bits.

Among the class of error correcting codes, linear block codes, also known as Forward Error Correction (FEC) codes, are of practical importance (Costello, 1998). One of them, the Reed-Solomon codes, are used extensively in recent popular communication systems, such as DVB or ADSL. Each Reed-Solomon block code is formed of N bytes, of which K bytes are information bytes and $(N - K)$ bytes are redundancy bytes. Figure 15 illustrates the error correcting capabilities of such codes.

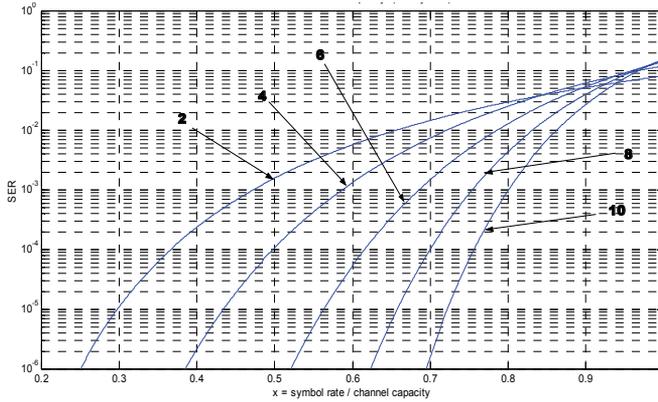


Fig. 14. QAM SER without FEC, as a function of the ratio bit-rate/channel capacity. Parameter: channel capacity (bit/symbol).

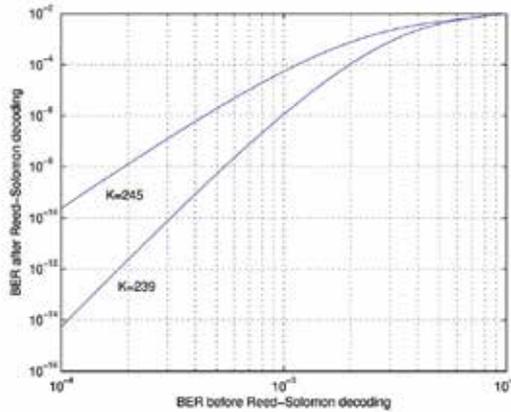


Fig. 15. Error correcting capabilities of some Reed-Solomon codes ($N=255, K$)

Including FEC in video transmission systems offers new possibilities for achieving UEP. Instead of, or in addition to, providing UEP at the modulation level, UEP can easily be provided at the FEC level by adding different FEC redundancies (and thus different error correction capabilities) to the bits, according to their perceptual importance.

Figure 16 provides a synoptic version of such a UEP approach in the context of a DMT + Reed-Solomon FEC coding, comparing it to a classic DMT + Reed-Solomon FEC coding without UEP.

In this figure, R_H and R_L , respectively, represent the bit-rates of the bitstreams of higher and lesser importance, and (N_H, K_H) and (N_L, K_L) represent the Reed-Solomon parameters of these two bitstreams. For the classic transmission scheme, $R_0 = R_H + R_L$ is the bit-rate, and (N_0, K_0) are the Reed-Solomon parameters. For the two UEP approaches, the DMT modulators are the same, with the same bit and power loading, which makes the BER before Reed-Solomon decoding the same in both approaches.

In order to evaluate the possibilities of this approach to UEP, the cumulative redundancy of the two Reed-Solomon codes, (N_H, K_H) and (N_L, K_L) , shall equal the redundancy of (N_0, K_0) (i.e., $R_0 \cdot N_0 / K_0 = R_H \cdot N_H / K_H + R_L \cdot N_L / K_L$). Figure 17 illustrates the relative variations of the BER of the highly important bits (BER_H) versus the relative variations of the BER of the less

important bits (BER_L), in terms of the BER of the classic method without UEP (BER_0) for an ADSL environment (ANSI CSA5 test loop) in which $(N_0, K_0) = (255, 239)$ and for different proportions of less important bits (10% to 50%).

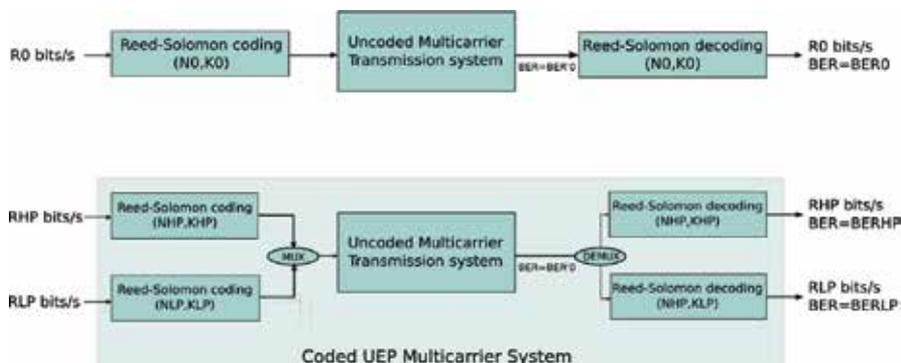


Fig. 16. Synoptic version of a DMT system using FEC with UEP and a classic DMT system using FEC without UEP.

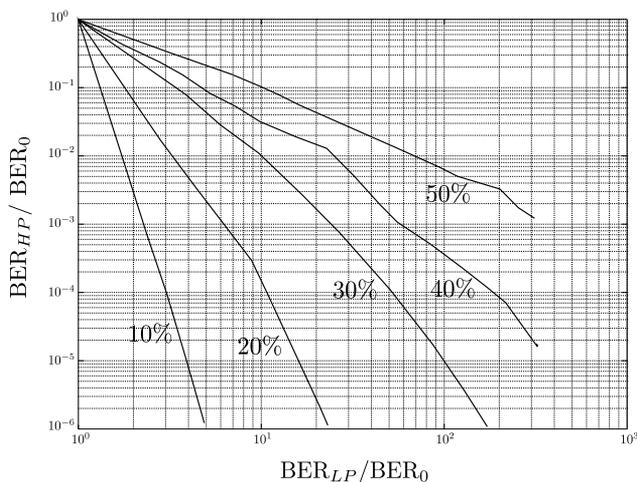


Fig. 17. Evaluation of the possibilities offered by providing unequal error protection at the FEC level.

More sophisticated UEP methods based on error control codes have been explored in the literature. In particular, the UEP methods using multilevel codes and multistage decoding accomplish Unequal Error Protection quite satisfactorily (Wachsmann, 1999; Chui, 2008).

5. Scalability tools

In the previous section, we underlined the fact that UEP with different priority levels can be applied successfully in modern video transmission systems to provide flexibility as well as the QoS level required by end-users. This is true because not all the various bits of the transmitted video data streams have the same importance with respect to reconstructed video quality. Recently, digital video compression standards have introduced the concept of *scalability* to allow video content to be encoded with different levels of resolution and

different levels of quality. In this section, we provide a brief overview of the existing scalability modes, as well as the corresponding coding tools.

Typically, video coding standards propose four main scalability modes: SNR scalability, spatial scalability, temporal scalability and data partitioning. Each of the modes is presented briefly below, with more detail given for the first and the last modes.

SNR scalability allows the delivery of a video bitstream compressed into several separate layers with the same spatio-temporal resolution but different quality levels. The base layer, which contains a reduced quality version of the encoded video signal, is typically transmitted with a high protection level—for example, using one of the techniques described in Section 4—to guarantee that video will be decoded even with high error rates. Then, each additional layer enhances the quality of the reconstructed video, but is transmitted with a reduced protection level. Hence, SNR scalability provides graceful degradation of decoded video according to the transmission quality.

A block diagram of a SNR scalable video encoder is shown in Figure 18. First, the base layer is encoded using coarse quantization. Then, this base layer is decoded, and the residual error between the reconstructed base layer and the original video signal, which constitutes the enhancement layer, is computed and then re-encoded during a second encoding stage, using finer quantization.

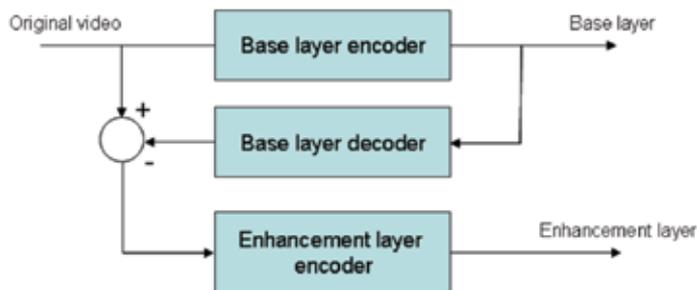


Fig. 18. Block diagram of a SNR scalable video encoder.

Other scalability tools rely on the same basic scheme. With *spatial scalability*, however, the video signal is encoded into separate streams corresponding to different spatial resolutions. The base layer consists of an encoded version of the video signal at lowest spatial resolution. Then, each enhancement layer contains additional data related to higher resolutions. *Temporal scalability* allows the simultaneous delivery of video signals with different frame rates.

The last mode, *data partitioning* (DP) (Mathew & Arnold, 1999), is quite similar to SNR scalability. The video signal is encoded into two separate bitstreams, with the additional layer allowing increased video quality if received at the decoding stage. However, data partitioning has the great advantage of being easier to implement. In fact, the DP mode is applied directly to the encoded bitstream and does not require any significant modification of the codec. In MPEG-2, the compressed video bitstream is split into two separate streams: the first one, corresponding to the base layer, contains the most important data for each of the 8x8 encoded blocks, such as low frequencies DCT coefficients or motion vectors. The remaining high frequency DCT coefficients, corresponding to details of each block, constitute the enhancement layer. The priority breakpoint (PBP) parameter marks the place where the bitstream is split into two parts. In the next section, we demonstrate that data partitioning is a great advantage for increasing error resilience, if unequal error protection is applied before transmission.

Since 2003, the scalability concept has been put forward through the design of the scalable video coding (SVC) amendment of the H.264/AVC standard. The SVC technology is based on the coding tools available in H.264/AVC, but includes also specific hierarchical and inter-layer predictive coding techniques. For further details on SVC, see (Schwarz et al., 2007) and (Huang et al., 2007). SVC provides a way to encode a video sequence into a single compressed bit-stream composed of a H.264/AVC compatible base layer and multiple additional layers that enhance spatial resolution, frame rate and quality. It is therefore possible to extract on-the-fly (inside the network or at the terminal) a limited number of layers in order to decode video with given resolution and quality level. Another obvious advantage of SVC is that UEP can be applied efficiently to the different parts of the bit stream having different importance in terms of reconstructed video quality. Hence, using SVC allows supporting a broad range of devices with different capabilities (display resolution, battery power) and access networks. Several papers on SVC applications have been proposed recently in the literature (Schierl et al., 2006) (Kouadio et al., 2008) (Thang et al., 2008). For example, Hellge et al. propose to combine of the spatial and SNR scalability features of SVC with the hierarchical modulation of DVB-H (Hellge et al., 2009). The authors show the benefits of jointly using SVC with hierarchical modulation in terms of error robustness or increased number of services.

6. Application to digital video delivery over digital subscriber lines

In the following Section, we present a complete quality-oriented transmission system for video distribution over DSL, optimized by applying techniques previously described.

The diffusion of audiovisual content, or Video on Demand (VoD), was one of the major objectives in the development of ADSL technology. This technology was initially a huge success for high-speed internet access and gaming. With increased bit-rates, in recent years, there has been a return to one of the original objectives, which was audio and video diffusion through ADSL.

ADSL involves transmission over the physical link portion of the telephone network, called subscriber loop, which connects DSL Multiplexer (DSLAM) located at the central office (CO) and Customer Premises Equipment (CPE) on the Subscriber side. The subscriber loop consists of a twisted-pair copper line. ADSL coexists with voiceband service by using the high frequency band above the one allocated to Plain Old Telephony Services (POTS). This relatively inexpensive technology benefits from the existence of a reliable widespread copper-wire infrastructure. Typically, several customer lines from the same CO ends at the same DSLAM, and the DSLAM outputs are connected to a high-speed Internet backbone line. The modulation used by International Telecommunications Union (ITU) in various standardized versions of ADSL is the DMT (see paragraph 4).

ADSL experiments and knowledge have resulted in several developments (e.g., ADSL1, ADSL2, ADSL2+). Initially, ADSL had a theoretical maximum bit-rate of about 6 Mb/s, which was rapidly multiplied by a factor of almost five in ADSL2+. For a fixed error rate and a fixed transmission power, throughput is strongly dependent on the line characteristics:

- attenuation of the copper line, which depends on line length; and
- noise, including Near End CrossTalk and Far End CrossTalk interference from lines in a same bundle, and impulsive noise.

Unfortunately, high bit-rates are available only to users located near the DSLAM. For users more than 2.5 kilometers from the CO, line attenuation restricts the bit-rate so that it rarely exceeds 4 Mbps. As a result, the subscribers cannot all receive the same bit-rate with the same error rate. If a video is sent with the same bit-rate for all the subscribers, then each subscriber receives the video at a different error rate, which may cause unacceptable visual degradations for a lot of users if the error rate is high. This usually happens for users with limited link capacity below the required bit-rate. In the same way, if the same error rate is set for all subscribers, then the bit-rate must be adapted for each subscriber. Generally, the latter solution is preferable to insure reliability of the received data.

However, in the context of video transmission, it is also necessary to take the real time constraint into account in the video broadcast. All subscribers must be able to receive video simultaneously with an acceptable quality level. To allow this, transcoding is usually done in platform somewhere between the server and the customer. Generally, using transcoding makes it possible, for example, to multicast the video over heterogeneous links. Transcoding is an all-embracing term, which can involve adapting the standard, the format, the bit-rate or the frequency. In this Section, we deal only with transrating, which consists of adapting the video bit-rate to each subscriber loop by removing part of the transmitted video information. Of course, this information removal procedure must be carried out in such a manner that the video received is the least degraded version possible.

In the literature, there are many transrating algorithms that are more, or less, complex. Some are based, for example, on the suppression of high-frequency spatial information (i.e., removal of high-frequency DCT coefficients) (Werner, 1999; Assuncao & Ghanbari, 1997a; Sun et al., 1996; Celandroni et al., 2000). Others are based on the suppression of certain images in order to reduce the display frequency, as in temporal scalability (Horn et al., 1999; Lagendijk et al., 2000). Still others are based on a stronger quantization of predefined image areas (Tudor & Verner, 1997; Assuncao & Ghanbari, 1997a; Assuncao & Ghanbari, 1997b; Sun et al., 1996; Celandroni et al., 2000). These algorithms can be divided into two groups: *closed-loop transrating*, which require a complete decoding and re-encoding of the compressed video source, and *open-loop transrating*, which generally do not resort to a complete decoding, making them much less complex (Assuncao & Ghanbari, 1997a).

In this section, we present a Joint Source and Channel Coding (JSCC) approach for broadband MPEG-2 video distribution over DSL. Originally published by Coudoux et al. (2008), this approach combines a layered hierarchical video transrating scheme with an unequal error protection (UEP) technique and multi-carrier modulation for DSL video distribution in order to optimize the end-to-end video quality. We first highlight the existence of a "bottleneck bandwidth" of the video source after transrating; at this bandwidth, the quality of the video received by the subscriber is optimal. Following that, we present the JSCC system, and then explain the transrating optimization procedure. Finally, we report our simulation results.

6.1 Motivations

Let us consider an example of a MPEG-2 MP@ML single layer video source encoded at 6 Mbps. For the lines not reaching 6 Mbps with a sufficiently low error rate under practical power constraint, decreasing the bit-rate will decrease the transmission error rate. However, this improvement is counterbalanced by the fact that decreasing the bit-rate of the source video (and, as a result, removing some useful video information), increases the deterioration

of the quality of the video received. Thus, lowering the bitrate clearly has two contradictory effects on the quality of the video available to the end-user.

Consequently, it can be assumed that, for each subscriber loop that cannot achieve 6 Mbps with appropriate QoS level, there is an ideal bit-rate, called the “bottleneck” bit-rate, at which the overall deterioration is minimal. This minimal deterioration is the result of a compromise between the distortion generated by the bit-rate reduction and the distortion generated by the transmission. Figure 19 illustrates this basic idea.

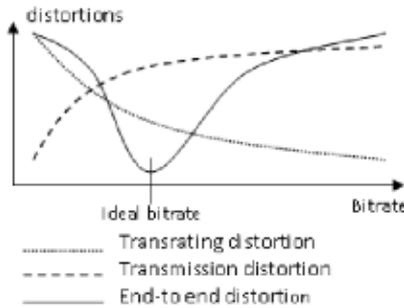


Fig. 19. End-to-end distortion in terms of the overall bit-rate

To insure a minimum level of deterioration, and thus a received video quality threshold, we used a hierarchical video source coding system, coupled with UEP at channel coding level as described in the preceding sections.

6.2 System architecture

Our JSCC system is based on the Data Partitioning (DP) mode used in the MPEG-2 standard, which splits the MPEG-2 bitstream into two hierarchical bitstreams (see Section 5). For simplicity, DP was chosen in order to conform to real-time constraint: indeed, it makes possible to split on-the-fly the incoming video bitstream into two separate ones. The most important bitstream is the base layer containing the most important data (e.g., headers, motion vectors and low-frequency DCT coefficients). The second bitstream is the enhancement layer (EL), containing all the other data, typically high-frequency DCT coefficients. Thus, these two bitstreams can be protected differently against channel errors, using UEP as explained in section 4.

The proposed JSCC system can be decomposed in a four-step adaptation process, as shown in Figure 20:

1. *Estimating the channel* -- The channel-to-noise ratios of the considered line are estimated for each carrier. These ratios are used by the bit- and power-loading algorithm (see Section 4).
2. *Defining the transrating parameters* -- The single layer input video bitstream is transcoded into two video bitstreams. Starting with the estimated channel parameters, the bit-rates of both the base layer bitstream and the enhanced layer bitstream must be determined.
3. *Transrating using data partitioning* -- The MPEG-2 bitstream coming from the very high bit-rate network (from a video server) and received at the CO is transcoded based on the transrating parameters defined in step 2.
4. *Coding and transmission* -- UEP is applied to the two bitstreams. Reed Solomon (RS) codes are used as in the ADSL standard in this step. As explained in section 4, two different RS codes can be applied to the two bitstreams before transmission. The two

bitstreams are then transmitted at fixed BER (with $BER_1 < BER_2$). The RS code is applied to the base layer, so that the decoded base bitstream can be considered to be error-free at the receiver.

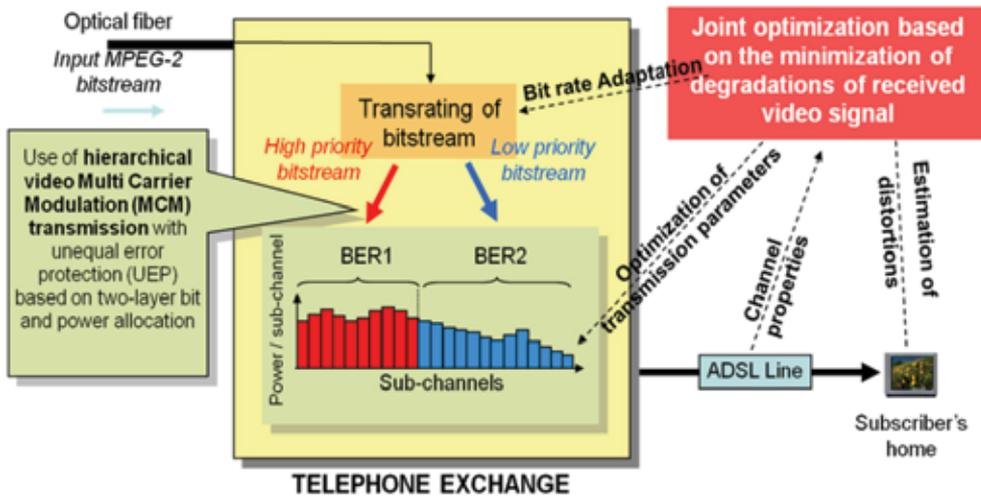


Fig. 20. Overview of the complete JSCC architecture

The first two steps are performed by the ADSL modem during initialization, assuming quasi-static transmission conditions; the last two are specific to our transrating system.

The Data partitioning step uses the Priority Break Point (denoted PBP1 in this chapter) defined in the MPEG-2 standard. PBP1 defines the separation point of the two bitstreams, and its value is related to the number of non-zero DCT coefficients preserved in the base layer (i.e., only the first low-frequency DCT coefficients). PBP1 depends on the number of VLC codes preserved in the base layer after the zigzag serialization MPEG procedure. The JSCC system maintains PBP1 constant in order to insure a minimal video quality level at the receiver side.

As the goal is to reduce the total video bit-rate, we introduce a second Priority Break Point, denoted PBP2 ($PBP_1 \leq PBP_2 \leq 127$), which defines the point at which the remaining VLC codes are discarded. The extreme values of PBP2 correspond to the situation in which the base layer only is transmitted ($PBP_2 = PBP_1$) and to the situation in which all the bit-rates are preserved before transrating ($PBP_2 = 127$). The other intermediate values make it possible to adjust the video bit-rate of the enhanced layer. This adjustment can propagate a quantization error (i.e., drift) and thus lead to a reduction in video quality. However, the quality metric used for determining transrating parameters takes into account of this error (C. Goudemand et al., 2007).

The simulations performed by Goudemand et al. (2006) showed that the optimal value of the PBP2 determines the best compromise between the perceived distortion and the bit-rate reduction obtained. Simulations on different MPEG-2 video sequences transmitted at 6 Mbps were performed with different PBP2 values. From one sequence to another, the bit-rate curves obtained after transrating versus the PBP2 value are very similar. For this reason, the average of these curves was used to determine the video bit-rate after transrating based on PBP2 (Figure 21). As illustrated, the bit-rate increases rapidly from 2.6 to 5.95 Mbps when the PBP2 goes from 64 to 85. Then, from $PBP_2 = 85$ to 127, the bit-rate increases slowly. This

evolution of the bit-rate with PBP2 can be easily explained by the fact that the DCT coefficient's energy decreases as the spatial frequency increases.

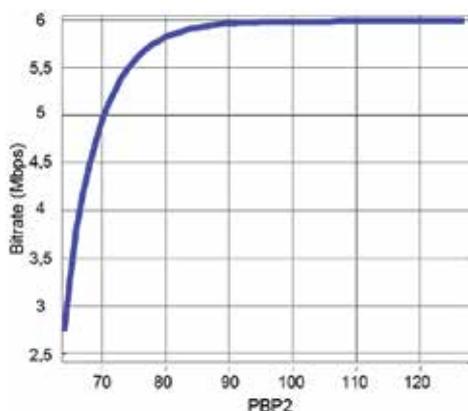


Fig. 21. Evolution of total bit rate as a function of the PBP2 parameter.

We also verified that the bit-rate overhead introduced by data partitioning does not exceed 1.5% of the aggregate video rate and therefore can be neglected. Let us consider our example of a coded sequence at 6 Mbps, transrated with the PBP1 and PBP2 equal to 67 and 85, respectively. According to Figure 21, the bit-rate at PBP1=PBP2=67 is about 4.3 Mbps, which is the base layer bit-rate. With PBP2=85, the bit-rate is 5.95 Mbps, which corresponds to the total bit-rate after transrating (i.e., the total bit-rate of the base and enhanced layers). The bit-rate corresponding to the enhancement layer is therefore equal to 1.65 Mbps.

In order to optimize the overall video quality, it is necessary to have a quality metric which permits to evaluate practically the visual impact of eliminating high-frequency DCT coefficients by transrating. First, let us introduce some features of the quality metric used in our system for the case of still images (i.e., the (I) intra-coded pictures of an MPEG-2 bitstream). Remember that, for the intra-pictures, the DCT coefficients are uniformly quantized. A quantization table defines the quantization step values that were obtained from subjective measurements.

The Normalized Weighted Mean Square Error (NWMSE) quality metric, proposed by Goudemand (2007), is based on the quantization property of the DCT coefficients involved in the MPEG-2 coder. As each quantization step value is related to the importance of the visual impact of the corresponding DCT coefficient, the weights used for this metric are related to these quantization steps. Such WMSE metric has been first introduced by Vandendorpe (1991) in the context of sub-band coding.

In the present case, the reference for calculating the quality metric is the input video sequence compressed at 6 Mbps. First considering the intra-coded pictures, WMSE is typically calculated from the DCT coefficients perceptually weighted according to the contrast sensitivity function (CSF) of the human eye. The weights used in the WMSE computation are inversely proportional to the square of the quantization step values. Therefore, the WMSE of the DCT coefficients can be seen for each block as the MSE of the corresponding quantized values. The perceptual distortion depends only on the error magnitude of the quantized DCT coefficients, independent of the spatial frequencies. The quantized coefficients thus have the same visual importance in the DCT domain. Any additional transrating or transmission distortion can be taken into account in the WMSE of the quantized coefficients.

The previous approach has been extended to (P) and (B) inter-coded pictures, such that WMSE can be computed from quantized DCT coefficients whatever the image type. Normalization is also introduced for each picture type, which depends on the different macroblock types as well as the rate control. However, the suppression of DCT coefficients due to transrating or transmission errors typically results in *drift error* among successive pictures. This phenomenon is accounted for by introducing weighting factors noted W_I, W_P, W_B defined as the average number of frames affected by an error in I, P, B frame, respectively. Finally, the average normalized perceptually weighted MSE of a sequence $NWMSE_{sequence}$ is the weighted sum of the NWMSE averaged for I, P, B frames defined as:

$$NWMSE_{sequence} = W_I E\{NWMSE_{I_frame}\} + W_P E\{NWMSE_{P_frame}\} + W_B E\{NWMSE_{B_frame}\} \quad (4)$$

6.3 System parameter optimization

The optimization process consists in determining the bit-rate of the two bitstreams, and therefore the values of the two Priority Break Points: PBP1 and PBP2, such that the received video quality (measured thanks to the above metric) is the best. The end-to-end perceptual quality of the video is considered to depend on the total bit-rate. In our JSCC architecture, the two-layer bitstreams received undergoes two independent types of degradations: transrating degradation and transmission degradation. Thus, the total degradation of the system is the sum of these two types, which can be evaluated using the NWMSE quality metric. Figure 22 shows two curves. For a given and fixed value of PBP1, the left side of the figure illustrates the evolution of the total transmission power P with respect to the PBP2.

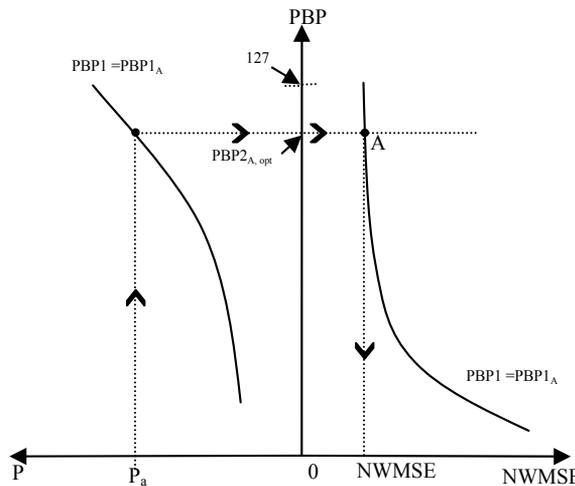


Fig. 22. Proposed approach for the determination of the end-to-end NWMSE

Let us consider the case of a transmission at power P , where $P=P_a$. This is a case for example in which all the total available power P_a is used for the transmission, making it possible to have the maximum transmission bit-rate. For any value of the $PBP1=PBP1_A$ given, the optimal value of the $PBP2=PBP2_{A,opt}$ is immediately obtained, as shown on the left curve in Figure 22. Thus, the operating point of the system given by $(PBP1_A, PBP2_{A,opt})$ can be determined. In addition, the NWMSE is a decreasing function of PBP2, as illustrated on the

curve shown on the right side part of Figure 22. Thus, the $NWMSE = NWMSE_A$ can be therefore determined from the value of $PBP2_{opt}$.

If another similar but different value of $PBP1$ is considered, other curves must be used, such as those in Figure 22. Since the value of $PBP1$ is unknown, the direct determination of $PBP2_{opt}$ is compromised. If all the values of $PBP1$ from 64 to 127 are considered, $PBP2_{opt}$ and the $NWMSE$ can be determined for each one of these values. A series of $NWMSE$ values will be obtained. The operating point of the optimal system is thus the couple $(PBP1_{opt}, PBP2_{opt})$, which will minimize the global end-to-end $NWMSE$.

In short, by varying the $PBP1$ from 64 to 127, determining the optimal value of the $PBP2$ and the $NWMSE$ for each value of the $PBP1$, and selecting the couple $(PBP1_{opt}, PBP2_{opt})$ that produces the minimal $NWMSE$, the locus C of all the operating points for the $PBP1$ values between 64 and 127 can be obtained, as shown in Figure 23.

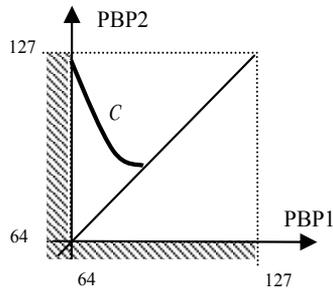


Fig. 23. Evolution of the $PBP2$ parameter as a function of $PBP1$, under transmit power constraint (Coudoux et al., 2008).

6.4 Simulation results

Experiments were conducted in two steps: 1) the determination of the optimal transrating parameters and 2) the simulation of the transmission of transrated video sequences over ADSL lines.

The optimal parameters were determined as follows: for each subscriber loop, the transrating couples $(PBP1; PBP2)$ were estimated, which allowed the base and enhancement layers to be transmitted. These parameter couples were obtained using the bit and power allocations at a fixed bit-rate for the two bitstreams, based on the values of $PBP1$ and $PBP2$ at a maximum total power of 110 mW as stated in the ADSL standard. The optimal couple was then the one corresponding to the minimum total end-to-end $NWMSE$.

The simulations were carried out at 6 Mbps initial bit-rate using MPEG-2 (MP@ML) video sequences. ADSL Reed Solomon RS (255, 239) were applied to the base layer, which generated a bit-rate overcost of about 6.7%. The BER of the base layer transmission was fixed at 10^{-6} before RS decoding ($BER_{BL} = 10^{-6}$), which corresponds to about $6.4 \cdot 10^{-33}$ after RS decoding. For this reason, the BL transmission was assumed to be error-free. The BER of the enhanced layer transmission was fixed at 10^{-4} ($BER_{HL} = 10^{-4}$) so that visual degradations would be minimized whatever the values of $PBP1$ and $PBP2$. The bit and power allocation algorithm used frequency division multiplexing of the two layers, as explained in Section 4. Once the optimal $PBP1$ and $PBP2$ parameters for the transmission were determined, the transmission of the video processed with these values was simulated.

For purposes of comparison, the ADSL MPEG-2 transmission was used in our experiments as the reference transmission. Figure 24 represents the evolution of the end-to end $NWMSE$

as a function of the subscriber line length. The European Telecommunications Standards Institute (ETSI) test loop 1 was used (ETSI, 1996). Remember the combined source/channel transrating system described in this section was not designed to improve the quality of the received image on “a good” line, but rather to increase the range of ADSL video transmission. Given this reminder, a single layer ADSL transmission at a BER of 10^{-6} over the 1800 m-long ETSI loop 1 required 2.5 dB more than the authorized ADSL maximum transmission power of 110 mW. In order to reduce the total power to 110 mW without changing all other transmission parameters, it was necessary to reduce the line length by approximately 110 m. The traditional ADSL transmission on the same line was thus limited to 1700 m around the telephone exchange. The solution proposed by the authors led to a slightly degraded transmission in an area with an 1800 m radius, which can possibly be extended if progressive degradations, increasing with the line length, are accepted.

The two-layer JSCC used in these experiments was compared to a single-layer transrating system based on the same principle, but using only one priority break point PBP2=PBP1. With the same average system quality, the two-layer JSCC system led to an increase in the zone covered compared to the single-layer transmission system.

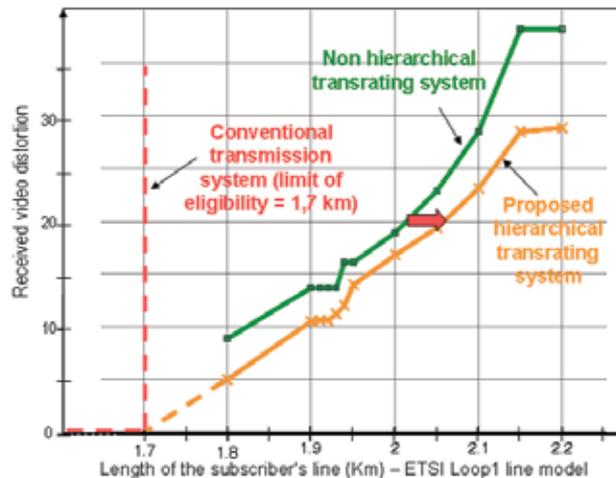


Fig. 24. Evolution of end-to-end perceptual distortions as a function of subscriber’s line length (from Coudoux et al., 2008).

7. Conclusion

We demonstrated in this chapter how source and channel coding can be efficiently combined in order to fulfil QoS requirements of modern video communication systems. The underlying philosophy relies on the fact that not all the bits of a compressed video bitstream have the same visual importance. Thus, both digital video compression and transmission parameters should be jointly designed in an end-to-end approach, aiming at delivering the best video quality to the final end-user. We also demonstrated that digital video signals can be encoded into several layers of varying perceptual relevance using scalability tools. Subsequently, if unequal error protection is applied to these layers, the most visually important layers will be the best protected. Such video transmission methods typically increase the flexibility and reliability of video communications in heterogeneous multimedia

environments, and produce a better quality of experience for the end-user. The effectiveness of a combined source/channel coding strategy was illustrated for the particular case of DSL video distribution. We showed that the eligibility level of video services can be optimally extended, by transrating the compressed video bitstream available on a distant server for the optimal bit-rate at which the end-user's visual quality is the best. The extension of these results to the case of the scalable extension of the H.264/AVC standard as well as wireless residential networks is currently under consideration in the TOSCANE project (see: <http://www.lien.uhp-nancy.fr/lien/index.php?p=toscane>).

Combined source and channel coding strategies constitute a promising research topic in the field of video communications. This is particularly true when considering the emergence of new services, such as mobile video streaming on small display devices or 3D-TV, as well as new transmission architectures, such as wireless vision sensor networks, for example. Future work will try to develop more efficient solutions to respond to the new challenges of video communications in terms of error resilience, accessibility, interactivity and universality.

8. References

- Chui, J.; Calderbank, A.R. (2008). *Multilevel diversity-embedded space-time codes for video broadcasting over WiMAX*, Proceedings of the IEEE International Conference on Information Theory, pp. 1068-1072, Nice, France, June 2008.
- Coudoux, F.-X.; Gazalet, M.G.; Mouton-Goudemand, C.; Corlay & P.; Gharbi, M. (2008). Extended Coverage for DSL Video Distribution Using a Quality-Oriented JSCC Architecture, *IEEE Transactions on Broadcasting*, Volume 54, Issue 3, Sept. 2008, pp. 525 - 531, ISSN: 0018-9316.
- Coudoux F.X., Gazalet M., Derviaux C., Corlay P. (2001). Picture quality measurement based on block visibility in discrete cosine transform video sequences, *Journal of Electronic Imaging*, Vol.10(2), pp.498-510.
- Daly S. (1993). The visible differences predictor: an algorithm for the assessment of image fidelity. In Watson A.B. (ed.), *Digital Images and Human Vision*, pp. 179-206, MIT Press.
- ETSI Technical Report ETR 328 (1996). *Transmission and Multiplexing (TM; Asymmetric Digital Subscriber Line (ADSL); Requirements and Performance*, Reference DTR/TM-06001, November 1996.
- Faugeras O.D. (1979). Digital color image processing within the framework of a human visual model, *IEEE Trans. On Acoustics, Speech and Signal Processing*, vol. 22, no. 4, pp. 380-393.
- Girod B. (1993). What's wrong with mean-squared error? , in *Visual Factors of Electronic Image Communications*. Cambridge, MA: MIT Press.
- Glenn W.E. (1993), *Digital image compression based on visual perception and scene properties, video processing*, SMPTE Journal, pp. 395-397, May 1993
- Goudemand, C., Coudoux, F.X., Gazalet, M. (2006). A study on Power and Bit Assignment of Embedded Multi-Carrier Modulation Schemes for Hierarchical Image Transmission over Digital Subscriber Line, *IEICE Transactions on Communications*, Vol. E89, No.7 (July 2006), page numbers (2071-2073).
- Goudemand, C.; Gazalet, M.; Coudoux, F.X.; Gharbi, M. (2006). Reduced complexity power minimization algorithm for DMT transmission - Application to layered multimedia

- services over DSL, *Proceedings of the 13th International Conference of Electronics, Circuits and Systems, ICECS 2006*, pp. 728-731, Nice, France, December 2006.
- Goudemand C., Coudoux F.X., Gazalet M.G. (2006). Optimal bit rate adaptation for layered video transmission over spectrally shaped channels using multicarrier modulation, *IEEE ICIP-2006*, pp.13-16, Atlanta, USA.
- Goudemand C., Gazalet M.G., Coudoux F.X., Corlay P., Gharbi M. (2007). A Low Complexity Image Quality Metric for Real-Time Open-Loop Transcoding Architectures, *IEEE ICC' 07*, Glasgow.
- Huang, H-C., Peng, W-H., Chiang, T., Hang, H-M., "Advances in the Scalable Amendment of H.264/AVC", *IEEE Commun. Mag.*, pp. 68-76, Jan. 2007.
- Hellge, C., Mirta, S., Schierl, T., Wiegand, T., „Mobile TV with SVC and Hierarchical Modulation for DVB-H Broadcast Services“, *IEEE BMSB*, Bilbao, 13-15 May 2009.
- Kouadio, A., Clare, M., Noblet, L., Bottreau, V., "SVC - a highly scalable version of H.264/AVC", EBU Technical Review, 2008.
- Legendijk R.L., Frimout E.D., Biemond J. (2000). Low-Complexity Rate-Distortion Optimal Transcoding of MPEG I-frames, *Signal Processing: Image Communication*, no. 15, p531 - 544.
- Lee Y.L., Kim H.C., Park H.W. (1998). Blocking effect reduction of JPEG images by signal adaptive filtering, *IEEE Trans. Image Processing*, vol. 7, no. 2, pp. 229-234.
- Lin, S., Costello, D. (1983), *Error Control Coding: Fundamentals and Applications*, Prentice-Hall Series in Computer Applications in Electrical Engineering, ISBN 0-13-283796-X, N.J. 07632.
- Lukas F.X., Budrikis Z.L. (1982). Picture quality prediction based on a visual model, *IEEE Transactions on Communications*, vol. 30, no. 7, pp. 1679-1692.
- Mannos J.L., Sakrison D.J. (1974). The effects of a visual fidelity criterion on the encoding of images, *IEEE Trans. On Information Theory*, vol. 20, no. 4, pp. 525-535.
- Mathew R., Arnold J.F. (1999), Efficient Layered Video coding using Data Partitioning, *Signal Processing: Image Communication*, n°14, 1999.
- Mitchell J., Pennebacker W., Fogg C., Le Gall D. (1995). *MPEG Video Standard Compression*, Van Nostrand Reinhold, New York.
- Pennebaker W.B., Mitchell J.L. (1993). *JPEG Still Image Data Compression Standard*, Van Nostrand Reinhold, New York.
- Pinson M., Wolf S. (2004). A new standardized method for objectively measuring video quality. *IEEE Transactions on Broadcasting*, vol. 50, no. 3, pp 312-322.
- Proakis, G.P. (1995). *Digital Communications*, third edition, McGraw-Hill Electrical Engineering Series, ISBN 0-07-051726-6, International Edition.
- Rabbani M., Jones P.W. (1991). *Digital Image Compression Techniques*, SPIE Press, Bellingham
- Ramamurthi B., Gersho A. (1986). Nonlinear space variant postprocessing of block coded images, *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1258-1267.
- Richardson, I. (2003). *H.264 and MPEG-4 Video Compression: Video Coding for Next Generation Multimedia*, John Wiley & Sons, ISBN 0-470-84837-5.
- Schierl, T., Gänger, K., Wiegand, T., Stockhammer, T., "SVC-based multisource streaming for robust video transmission in mobile ad hoc networks", *IEEE Wireless Commun. Mag.*, Special Issue on Multimedia in Wireless/Mobile Ad Hoc Networks, Vol. 13, N° 5, pp. 96-103, Oct. 2006.

- Schwarz, H., Marpe, D., Wiegand, T., "Overview of the Scalable Video Coding extension of the H.264/AVC standard", *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 17, N° 9, pp. 1103-1120, Sept. 2007.
- Shannon, C.E. (1948). A mathematical Theory of Communications, *Bell System Technical Journal*, Vol. 27 (October 1948), page numbers (379-423,623-656).
- Starr, T., Cioffi, J.M., Silverman P.J. (1999). *Understanding Digital Subscriber Line Technology*, Prentice Hall PTR, ISBN 0-13-780545-4, N.J. 07458.
- Sun H., Kwok W., Zdepski J. (1996). Architectures for MPEG Compressed Bitstream Scaling, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 6, No. 2.
- Talluri R., Moccagatta I., Nag Y., Gene C. (1999). «Error Concealment by Data Partitioning », *Signal Processing : Image Communication*, No. 14.
- Tekalp, M. (1996). *Digital video processing*, Prentice Hall: Signal processing, Chapman&Hall, International Thomson Publishing.
- Thang, T-C., Kang, J-W., Yoo, J-J., Ro, Y-M., "Optimal Multilayer Adaptation of SVC Video over Heterogeneous Environments", *Advances in Multimedia*, Volume 2008, Article ID 739192, 8 pages, doi:10.1155/2008/739192
- Vandendorpe L. (1991). Optimized quantization for image subband coding, *Signal Processing: Image Commun.*, 4, pp. 65-79.
- Vitthaladevuni, P.K.; Alouini, M.S. (2001). BER Computation of 4/M-QAM Hierarchical Constellations, *IEEE Transactions on Broadcasting*, Vol.47, No.3 (September 2001), page numbers (228-239).
- Wachsmann, U.; Fischer, F.H; Huber, J.B; (1999). Multilevel Codes: Theoretical Concepts and Practical Design Rules, *IEEE Transactions of Information Theory*, Vol.45, No.5 (July 1999), page numbers (1361-1391).
- Wang Z., Bovik A.C., Sheikh H.R., Simoncelli E.P. (2004). Image quality assessment; from error visibility to structural similarity, *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 300-612.
- Wang Z., Lu L., Bovik A. (2004). Video quality assessment based on structural distortion measurement. *Signal Processing: Image Communication*, special issue on objective video quality metrics, vol. 19, pp. 121-132.
- Watson A.B., Hu J., McGowan, J.F. (2001). Digital video quality metric based on human vision. *Journal of Electronic Imaging*, vol. 10, no. 1, pp. 20-29.
- Werner O. (1999). Requantization for Transcoding of MPEG-2 Intra frames, *IEEE Trans. Image Processing*, Vol. 8, No. 2.
- Wiegand T., Sullivan G. J., Bjntegaard G., Luthra A. (2003), Overview of the H.264/AVC video coding standard. *IEEE Trans. Circuits Syst. Video Techn.* 13(7), pp 560-576 (2003)
- Winkler S., *Digital Video Quality: Vision Models and Metrics*, Wiley Press, 2005.
- Wu H.R. (Ed.), Rao K.R. (Ed.), *Digital Video Image Quality and Perceptual Coding*, CRC Press, Nov. 2005.
- Yuen M., "Coding Artifacts and Visual Distortions", in *Digital Video Image Quality and Perceptual Coding*, H.R. Wu (Ed.), K.R. Rao (Ed.), CRC Press, 2005.
- Zheng, H. ; Liu, K.J.R. (2000). Power Minimization for Delivering Integrated Multimedia Services over Digital Subscriber Line, *IEEE Journal of Selected Areas on Communications*, Vol.18, No.6, (June 2000), page numbers 841-849.

Amplitude Phase Shift Keying Constellation Design and its Applications to Satellite Digital Video Broadcasting

Konstantinos P. Liolis¹, Riccardo De Gaudenzi², Nader Alagha³,
Alfonso Martinez⁴, and Albert Guillén i Fàbregas⁵

¹*Space Hellas S.A., R&D and Applications Division, 312 Messogion Ave., 153 41, Athens, Greece*

^{2,3}*European Space Agency (ESA/ESTEC), Keplerlaan 1, P.O. Box 299, 2200 AG, Noordwijk, The Netherlands*

⁴*Centrum Wiskunde & Informatica (CWI), Science Park 123, 1098 XG, Amsterdam, The Netherlands*

⁵*University of Cambridge, Department of Engineering, Trumpington Street, CB2 1PZ, Cambridge, UK*

1. Introduction

Satellite communications are providing a key role in the worldwide digital information networks. Among other applications, satellite communications provide the platform for Direct-to-Home (DTH) digital TV broadcasting as well as interactive and subscription TV services, mobile services to ships, aircraft and land-based users, and data distribution within business networks. Satellite networks are essential part of the Internet backbone enabling both broadband and narrowband Internet access services from remote and rural areas where the satellite access provide a unique way to complement the terrestrial telecommunication infrastructure. Moreover, satellite networks will play a crucial role in the framework of the Future Internet [Future Internet, 2009].

Recent trends in satellite communications show an increasing demand to replace or complement conventional modulation schemes, such as Quaternary Phase Shift Keying (QPSK), with higher-order M -ary modulation schemes. Some contributing factors that enable such trends are as follows [Alberty *et al.*, 2007; Benedetto *et al.*, 2005; Rinaldo & De Gaudenzi, 2004a; Rinaldo & De Gaudenzi, 2004b]:

- The usage of higher frequency bands (e.g., from X, Ku, up to Ka and Q).
- The exploitation of capacity boosting techniques, such as Adaptive Coding and Modulation (ACM) which are enabling the commercial exploitation of high frequency bands.
- The higher frequency reuse achievable with multibeam satellite antennas.
- The increasing satellite Effective Isotropic Radiated Power (EIRP) at RF as well as satellite antenna Gain over noise Temperature (G/T) becoming available thanks to payload and antenna technology improvements.

- The availability of wideband satellite transponders as well as the deep submicron Application Specific Integrated Circuit (ASIC) technology allowing the support of high data rate modems (i.e. single carrier transponder operation with hundreds of Mbaud).

Nevertheless, these technical enhancements require the exploitation of highly efficient coding techniques associated with power- and spectrally- efficient modulation schemes designed to operate over the satellite channel environment. Higher-order M -ary modulation schemes can provide greater spectral efficiency and thus the high data rate required for either digital multimedia applications or other applications such as point-to-point high data-rate backbone connectivity and future Earth observation missions requiring downlink data rates exceeding 1 Gbps. In this regard, Amplitude Phase Shift Keying (APSK) represents an attractive modulation scheme for digital transmission over nonlinear satellite channels due to its power and spectral efficiency combined with its inherent robustness against nonlinear distortion. The concept of circular APSK modulation and its suitability for nonlinear channels was already proposed in 1970's [Thomas *et al.*, 1974] but, at that time, it was concluded that APSK performs worse than PSK schemes for single carrier operation over nonlinear channel. However, based on the recent pioneering work of some of the authors reported in [De Gaudenzi *et al.*, 2006a; De Gaudenzi *et al.*, 2006b], this conclusion was reverted and thus APSK has become nowadays a state-of-the-art modulation scheme for advanced satellite communications. As an illustration, APSK has been recently adopted in the following commercial standards related to satellite digital video broadcasting:

- DVB-S2 (Digital Video Broadcasting via Satellite – 2nd generation) [DVB-S2, 2005]: It is the ETSI standard for the forward link of satellite digital video broadcasting systems mainly operating at Ku (12/14 GHz) and Ka (20/30 GHz) frequency bands and targeting mainly fixed users but also mobile collective platforms, such as airplanes, ships and trains.
- DVB-SH (Digital Video Broadcasting via Satellite to Handheld devices) [DVB-SH, 2007]: It is the ETSI standard for the forward link of hybrid satellite/terrestrial digital video broadcasting systems operating at L (1/2 GHz) and S (2/4 GHz) frequency bands targeting mainly mobile users, such as handheld devices and vehicles.
- IPoS (Internet Protocol over Satellite) [IPoS, 2006]: It is the ETSI standard for broadband interactive satellite networks that has adopted DVB-S2 with ACM for its forward link. This standard has been proposed and adopted by the biggest interactive satellite networks manufacturer [Hughes, 2009].
- GMR-1 3G (Geostationary Mobile Radio – 3rd Generation) [GMR-1 3G, 2008]: It is a new ETSI standard for Mobile Satellite Systems (MSS) whose Release 3 has evolved to a 3G satellite-based packet service equivalent to 3GPP terrestrial mobile standard for the support of voice, data and video applications.
- ABS-S (Advanced Broadcasting System via Satellite) [Yuhai Shi *et al.*, 2008]: It is the Chinese standard for the forward link of satellite digital video broadcasting systems operating at Ku frequency band and targeting mainly fixed users and DTH applications.

The abovementioned standards have adopted the 16- and 32-APSK modulation schemes, whose constellation design has been described in [De Gaudenzi *et al.*, 2006a; De Gaudenzi *et al.*, 2006b]. These modulation schemes are mainly considered for professional applications where the satellite terminal sizing (i.e., antenna size and High Power Amplified (HPA)) allows for adequate Signal-to-Noise Ratio (SNR) at the receiver. However, APSK modulations can also be used for interactive consumer applications in case of multibeam

satellites or digital video broadcasting when higher spectral efficiency is needed. More recently, there has also been an increasing commercial interest to go beyond the standardized 16- and 32-APSK modes especially in professional applications such as Digital Satellite News Gathering (DSNG) where even higher requirements in terms of available SNR apply. Thus, design optimization for 64-APSK modulation has been addressed in [Liolis & Alagha, 2008; Benedetto *et al.*, 2005]. The reported results particularly in [Benedetto *et al.*, 2005] were obtained as part of the ESA project “Modem for High Order Modulation Schemes” whose technical baseline has been recently proposed as potential standard for space telemetry related applications [CCSDS, 2007]. The APSK constellation design approaches considered in [De Gaudenzi *et al.*, 2006a; De Gaudenzi *et al.*, 2006b; Liolis & Alagha, 2008] follow two main optimization criteria: (i) maximization of the minimum Euclidean distance, and (ii) maximization of the channel mutual information. Unlike the former optimization criterion which refers to the high SNR asymptotic case, the latter provides an optimum M -APSK constellation for each SNR operating point. The design optimization analysis and results obtained for 16-, 32- and 64-APSK modulation schemes based on the mutual information maximization criterion are presented in detail in this chapter assuming error correction coding and taking into account both cases of equiprobable and non-equiprobable constellations.

Although these higher-order M -ary APSK modulation schemes have been specifically designed for operating over nonlinear satellite channels, they still show signal envelope fluctuations and are particularly sensitive to the characteristics of the satellite transponders which introduce channel nonlinearities. Examples of performance analysis of APSK signal transmission over nonlinear satellite channel and error rate bounds are reported in [Sung *et al.*, 2009]. As has been shown in [Casini *et al.*, 2004; De Gaudenzi *et al.*, 2006b] and in references therein, the power efficiency of APSK modulation schemes can be improved by applying pre-distortion on the transmit data, avoiding large input and output back-off values on the satellite transponder (or in the ground terminal HPA). Pre-distortion means intentionally modifying the location of the data symbols on the complex plane with respect to their nominal position. Such a technique only calls for a modification of the transmitted constellation points and this is particularly straightforward and effective for circular constellations such as APSK. Different pre-distortion schemes have been investigated in the literature, based on either “instantaneous” evaluation of the constellation centroids distortion at the receiver (adaptive *static* pre-distortion) [De Gaudenzi *et al.*, 2006b] or consideration of a certain amount of “memory” in the combined phenomenon of nonlinear distortion plus matched filtering at the receiver (adaptive *dynamic* pre-distortion) [Karam & Sari, 1991]. Despite the hardware complexity impact, commercial satellite modems have already adopted relevant dynamic pre-distortion techniques for standardized 16- and 32-APSK modes [Newtec, 2009], which are presented in this chapter.

The rest of this chapter is organized as follows. Section 2 describes the system model and the main effects of the nonlinear amplifier. Section 3 provides a formal description of APSK signal sets, describes the APSK constellation design optimization criterion of maximum mutual information and discusses some of the properties of the optimized constellations for the equiprobable and non-equiprobable cases. In Section 4, practical static and dynamic nonlinearity distortion pre-compensation techniques of commercial interest are described. Useful numerical results illustrating the system design tradeoffs for APSK constellations over typical satellite channels along with some practical design guidelines for system engineers

with regard to the application of APSK constellations for satellite digital video broadcasting are also provided along Sections 3-4. Finally, concluding remarks are drawn in Section 5.

2. System model

We consider a communication system composed of a digital modulator, a Square-Root Raised Cosine (SRRC) band-limiting filter and a nonlinear HPA with typical Travelling Wave Tube Amplifier (TWTA) characteristic for a satellite operating in Ku or Ka band. This model is representative of a satellite bent-pipe transponder with uplink noise negligible compared to the downlink. Due to the tight signal band-limiting, the impact of the satellite input and output analog filters is assumed negligible. In any case, satellite filters' linear distortions can be easily compensated through simple demodulator digital equalizers. The baseband equivalent of the transmitted signal at time t , $s_T(t)$, is given by

$$s_T(t) = \sqrt{P} \sum_{k=0}^{L-1} x(k) p_T(t - kT_s) \quad (1)$$

where P is the signal power, $x(k)$ is the k -th transmitted symbol drawn from a complex-valued APSK signal constellation \mathcal{X} with cardinality $|\mathcal{X}| = M$, $p_T(t)$ is the SRRC transmission filter impulse response and T_s is the transmitted symbol duration (in seconds). Without loss of generality, we consider transmission of frames with L symbols. The coded modulation spectral efficiency R is the number of information bits normalised by the modulator baud rate. Equivalently, $R = r \log_2 M$, where r is the coding rate.

The signal $s_T(t)$ passes through the HPA which is operated close to the saturation point. In this region, the HPA shows nonlinear characteristics that induce phase and amplitude distortions to the transmitted signal. The HPA is modelled by a memoryless nonlinearity, with an output signal $s_A(t)$ at time t given by

$$s_A(t) = F(|s_T(t)|) \exp \left[j \left(\phi(s_T(t)) + \Phi(|s_T(t)|) \right) \right] \quad (2)$$

where we have implicitly defined $F(A)$ and $\Phi(A)$ as the Amplitude-to-Amplitude Modulation (AM/AM) and Amplitude-to-Phase Modulation (AM/PM) characteristics of the HPA amplifier for a signal with instantaneous amplitude A . As an illustration, the AM/AM and AM/PM characteristics of a typical Ka band TWTA is depicted in Fig. 1. The signal amplitude is the instantaneous complex envelope so that the baseband equivalent signal is denoted as $s_T(t) = |s_T(t)| \exp[j\varphi(s_T(t))]$.

2.1 Linear AWGN channel model

In this case, we assume an (ideal) signal modulating a train of rectangular pulses, which do not create Inter-Symbol Interference (ISI) when passed through the HPA operated in the nonlinear region [De Gaudenzi *et al.*, 2006a]. Under these conditions, the channel reduces to an Additive White Gaussian Noise (AWGN), where the modulation symbols are distorted following (2). Let x_A denote the distorted symbol corresponding to $x = |x| \exp[j\varphi(x)] \in \mathcal{X}$, i.e.,

$$x_A = F(|x|) \exp \left[j \left(\phi(x) + \Phi(|x|) \right) \right] \quad (3)$$

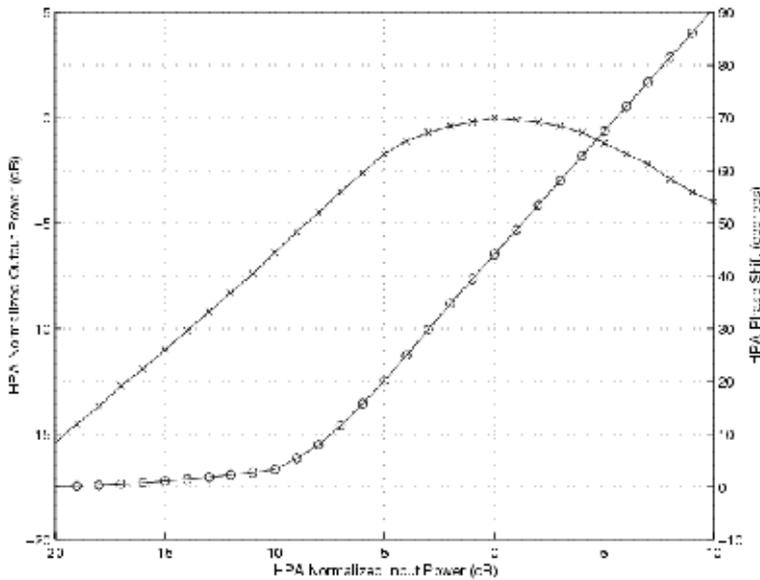


Fig. 1. AM/AM (crosses '+') and AM/PM (circles 'o') characteristics of the reference Ka band TWTA.

After matched filtering and sampling at time kT_s , the discrete-time received signal at time k is then given by

$$y(k) = \sqrt{E_s} x_A(k) + n(k), \quad k = 0, 1, \dots, L-1 \quad (4)$$

where E_s is the symbol energy given by $E_s = PT_s$, $x_A(k)$ is the symbol at the k -th time instant as defined in (3) and $n(k) \sim N_C(0, N_0)$ is the corresponding noise sample.

2.2 Non-Linear channel model

In this case, we introduce the parameter $E_b/N_0|_{\text{sat}}$ defined as the ratio between the transmitted energy per bit E_b when the amplifier is driven at saturation by a continuous wave carrier and the noise power spectral density N_0 at the demodulator input. Note that the subscript "sat" refers to the HPA saturation. The SNR at the demodulator input $E_b/N_0|_{\text{inp}}$ is reduced by the Output Back-Off (OBO, in dB) with respect to the value in a system operating with a single constant-envelope signal at amplifier saturation (in dB):

$$\frac{E_b}{N_0} \Big|_{\text{sat}} (IBO) = \frac{E_b}{N_0} \Big|_{\text{inp}} (IBO) + OBO (IBO) \quad (5)$$

where IBO is the satellite transponder's Input Back-Off.

Additionally, due to constellation warping and satellite channel induced ISI, the demodulator performance is degraded by an amount D (in dB) with respect to an ideal linear AWGN channel [De Gaudenzi *et al.*, 2006b]. This quantity D depends on the HPA distortion and hence on the IBO/OBO satellite characteristics (see e.g., Fig. 1). With this degradation, the effective SNR at the demodulator input $E_b/N_0|_{\text{eff}}$ is given by (in dB)

$$\frac{E_b}{N_0} \Big|_{\text{sat}} (IBO) = \frac{E_b}{N_0} \Big|_{\text{eff}} + OBO(IBO) + D(IBO) \quad (6)$$

Eq. (6) allows us to calculate an optimum HPA operating point which minimizes $E_b/N_0|_{\text{sat}}$. This point represents the best trade-off between the increasing power loss (OBO) related to the higher IBO and the reduction of the distortion (D) due to the improved linearity experienced by a larger IBO.

3. APSK constellation design optimization

3.1 Constellation description

M -APSK constellations are composed of n_R concentric rings, each with uniformly spaced PSK points. The signal constellation points x are complex numbers, drawn from a set \mathfrak{X}

$$\mathfrak{X} = \begin{cases} r_1 \exp \left[j \left((2\pi/n_1)i + \theta_1 \right) \right] & i = 0, \dots, n_1 - 1 \quad (\text{Ring } \ell = 1) \\ r_2 \exp \left[j \left((2\pi/n_2)i + \theta_2 \right) \right] & i = 0, \dots, n_2 - 1 \quad (\text{Ring } \ell = 2) \\ \vdots & \vdots \\ r_{n_R} \exp \left[j \left((2\pi/n_R)i + \theta_{n_R} \right) \right] & i = 0, \dots, n_{n_R} - 1 \quad (\text{Ring } \ell = n_R) \end{cases} \quad (7)$$

where n_ℓ, r_ℓ and θ_ℓ ($\ell = 1, \dots, n_R$) are defined as the number of points, the radius and the relative phase shift for the ℓ -th ring, respectively. Such modulation schemes are termed hereinafter as $n_1 + n_2 + \dots + n_{n_R}$ -APSK. Fig. 2 depicts the 4+12-APSK, 4+12+16-APSK and 4+12+16+32-APSK constellations. As mentioned above, for next generation standardized digital video broadcasting satellite systems, the constellation sizes of interest are $|\mathfrak{X}| = 16$ and $|\mathfrak{X}| = 32$ with $n_R = 2$ and $n_R = 3$ rings, respectively [DVB-S2, 2005; DVB-SH, 2007; GMR-1 3G, 2008; IPoS, 2006; Yuhai Shi *et al.*, 2008]. In addition, there has been an increasing commercial interest to go well beyond the standardized 16- and 32-APSK modes especially in professional DSN applications where the constellation size of interest is $|\mathfrak{X}| = 64$ with $n_R = 4$ rings. In general, we consider that \mathfrak{X} is normalized in energy, i.e., $E[|x|^2] = 1$ which further implies that the radii r_ℓ are normalized such that $\sum_{\ell=1}^{n_R} n_\ell r_\ell^2 = 1$. Note also that the radii r_ℓ are ordered such that $r_1 < r_2 < \dots < r_{n_R}$.

Furthermore, in order to reduce the dimensionality of the optimization problem, instead of optimizing the phase shifts θ_ℓ and the ring radii r_ℓ in absolute terms, the objective here is their optimization in relative terms. That is, we are looking for the optimum values of the phase shift of the ℓ -th ring with respect to the inner ring, $\phi_\ell = \theta_\ell - \theta_1$ ($\ell = 1, \dots, n_R$), and of the relative radius of the ℓ -th ring with respect to the inner ring, $\rho_\ell = r_\ell / r_1$ ($\ell = 1, \dots, n_R$), which satisfy a design optimization criterion. In particular, $\phi_1 = 0$ and $\rho_1 = 1$. Thus, we are interested in finding an M -APSK constellation defined by the parameters $\mathbf{p} = (\rho_1, \rho_2, \dots, \rho_{n_R})$ and $\mathbf{\varphi} = (\varphi_1, \varphi_2, \dots, \varphi_{n_R})$ such that a given cost function $f(\mathfrak{X})$ reaches a maximum. As discussed next, the cost function employed here is the mutual information of the AWGN channel [De Gaudenzi *et al.*, 2006a; De Gaudenzi *et al.*, 2006b; Liolis & Alagha, 2008] which, unlike the more classical optimization criterion of the minimum Euclidean distance referring to the high SNR asymptotic case, provides an optimum M -APSK constellation \mathfrak{X} for each coding

rate and modulation format Quasi-Error Free (QEF) SNR operating point or, equivalently, for each spectral efficiency R . We particularly assume a linear AWGN channel whereas robustness against nonlinear distortion is achieved in a subsequent step through exploitation of constellation pre-compensation (see Section 4).

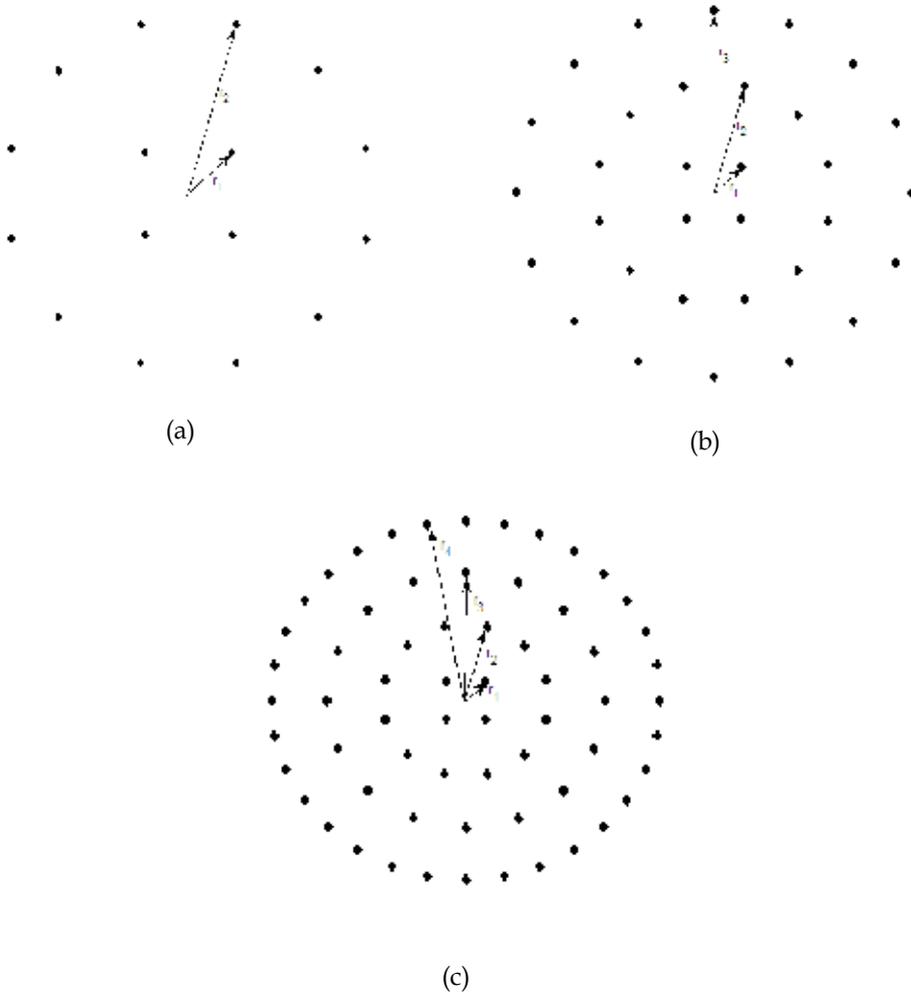


Fig. 2. M -APSK Constellations: (a) 16-APSK (4+12-APSK), (b) 32-APSK (4+12+16-APSK), (c) 64-APSK (4+12+16+32-APSK). The radii r_ℓ ($\ell = 1, \dots, n_R$) of the multiple rings are also shown.

Next, two separate cases of APSK constellations are specifically examined with respect to the probability distribution of the constellation points $x \in \mathcal{X}$:

- Equiprobable constellation points, that is, all constellation points $x \in \mathcal{X}$ are equiprobable with probability $1/M$;
- Non-equiprobable constellation points, that is, the constellation points on each ℓ -th ring are assumed to be equiprobable but the *a priori* probability P_ℓ associated per ring is assumed different for each ring such that $\sum_{\ell=1}^{n_R} n_\ell P_\ell = 1$.

3.2 Equiprobable constellation optimization

In this case, all constellation points $x \in \mathcal{X}$ are equiprobable with probability $1/M$, as considered in all standardized APSK modes so far [DVB-S2, 2005; DVB-SH, 2007; GMR-1 3G, 2008; IPoS, 2006; Yuhai Shi *et al.*, 2008]. The mutual information for a given APSK signal set \mathcal{X} provides the maximum transmission rate (in bits/channel use) at which error-free transmission is possible with such signal set and is given by [Ungerboeck, 1982]

$$f_{eq}(\mathcal{X}) = I_{eq}(X; Y) = \log_2 M - \frac{1}{M} \sum_{k=0}^{M-1} E_w \left\{ \log_2 \sum_{i=0}^{M-1} \exp \left[-\frac{E_s}{N_0} \left(|x^k + w - x^i|^2 - |w|^2 \right) \right] \right\} \quad (8)$$

Thus, the optimization problem to be solved is formulated as

$$C_{eq}^* = \max_{\rho, \Phi} f_{eq}(\mathcal{X}) \quad (9)$$

In (8), $E\{\cdot\}$ denotes the expectation operator. We have particularly used expectation over the normally distributed noise variable w which is complex with variance N_0 , i.e., $w \sim N_c(0, N_0)$. In general, obtaining a closed-form expression for C_{eq}^* in (9) is a daunting task and so the Gauss-Hermite quadrature rules are employed for its numerical computation [Abramowitz & Stegun, 1964]. Numerical calculations of (9) for 16-, 32- and 64-APSK constellations are provided next in Sections 3.2.1-3.2.3.

However, for the asymptotic case $E_s/N_0 \rightarrow \infty$, it is possible to obtain a closed-form expression as follows. First, note that the expectation in (8) can be rewritten as

$$\lambda(\mathcal{X}) \triangleq \frac{1}{M} \sum_{k=0}^{M-1} E_w \left\{ \log_2 \sum_{i=0}^{M-1} \exp \left[-\frac{E_s}{N_0} \left(|x^k - x^i|^2 + 2 \operatorname{Re}((x^k - x^i)w) \right) \right] \right\} \quad (10)$$

Using the *Dominated Convergence Theorem* and the analysis presented in [De Gaudenzi *et al.*, 2006a], the influence of the noise term w vanishes asymptotically, since the limit can be pushed inside the expectation. Furthermore, the only remaining terms in the summation over $x^i \in \mathcal{X}$ are $x^i = x^k$ and those closest in Euclidean distance $\delta_{\min}^2 = \min_{x^i \in \mathcal{X}} |x^i - x^k|^2$ which are in total $n_{\min}(x^k)$. Therefore, taking the above into account as well as the approximation $\log_2(1+x) \approx x \log_2 e$ for $|x| \ll 1$, $\lambda(\mathcal{X})$ in (10) comes up in the high SNR asymptotic case

$$\begin{aligned} \lambda(\mathcal{X}) &= \frac{1}{M} \sum_{k=0}^{M-1} \left\{ \log_2 \left[1 + n_{\min}(x^k) \exp \left(-\frac{E_s}{N_0} \delta_{\min}^2 \right) \right] \right\} \approx \frac{1}{M} \sum_{k=0}^{M-1} \left\{ n_{\min}(x^k) \exp \left(-\frac{E_s}{N_0} \delta_{\min}^2 \right) \log_2 e \right\} \\ &= \varepsilon \exp \left(-\frac{E_s}{N_0} \delta_{\min}^2 \right) \end{aligned} \quad (11)$$

where ε is a constant which does not depend neither on the constellation minimum distance δ_{\min} nor on SNR. Thus, after substituting (11) in (8), (9) yields

$$C_{eq}^* \Big|_{SNR \rightarrow \infty} = \max_{\rho, \Phi} \left\{ \log_2 M - \varepsilon \exp \left(-\frac{E_s}{N_0} \delta_{\min}^2 \right) \right\} = \max_{\rho, \Phi} \delta_{\min}^2 \quad (12)$$

Eq.(12) clearly indicates that the maximization of the mutual information of the AWGN channel corresponds to the maximization of the minimum Euclidean distance in the asymptotic case of high SNR.

3.2.1 Numerical results for 16-APSK

Fig. 3 shows the numerical evaluation of (8) for a given range of values of $\rho_2 = r_2/r_1 = r_2$ and $\phi_2 = \theta_2 - \theta_1 = \theta_2$ for the 4+12-APSK constellation at $E_s/N_0=12$ dB. Surprisingly, there is no sensible dependence on ϕ . Therefore, the two-dimensional optimization can be achieved by simply finding the ρ_2 that maximizes AWGN channel capacity. This result is found to hold true also for the other constellations and hence, in the following, capacity optimization results do not account for relative phase shifts ϕ . Fig. 4 shows the union bound on the Symbol Error Probability (SER) for several 16-APSK modulations and for the optimum value of ρ_2 at $R=3$ bps/channel use (calculated as described above). Continuous lines indicate $\phi=0$ while dotted lines refer to the maximum value of the relative phase shift, i.e. $\phi=\pi/n_2$, showing no dependence on ϕ at high SNR. This absence of dependence on ϕ is justified by the fact that the optimum constellation separates the rings by a distance larger than the number of points in the ring itself, so that the relative phase ϕ has no significant impact in the distance spectrum of the constellation.

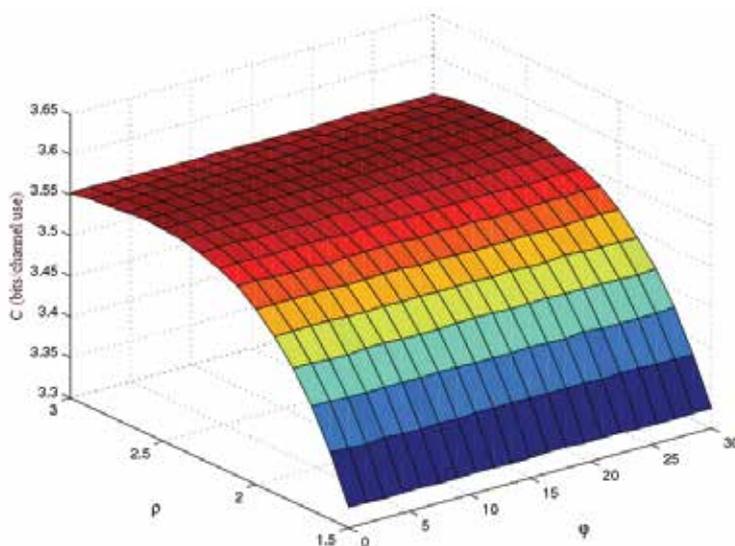


Fig. 3. Capacity surface for the 16-APSK (4+12) constellation at $E_s/N_0=12$ dB.

For 16-APSK it is also interesting to investigate the capacity dependency on n_1 and n_2 . Fig. 5 depicts the modulation constrained capacity curves for several configurations of optimized 16-APSK constellations and compared with classical 16-QAM and 16-PSK signal sets. As can be observed, capacity curves are very close to each other, showing a slight advantage of 6+10-APSK over the rest. However, as discussed in [De Gaudenzi *et al.*, 2006b] and also in Section 4 below, 6+10-APSK and 1+5+10-APSK show other disadvantages compared to 4+12-APSK for phase recovery and nonlinear channel behaviour. All APSK optimized constellations are doing equal or better than 16QAM and significantly better than 16PSK.

Note also that there is a small gain, of about 0.2 dB, in using the optimized constellation for every spectral efficiency R rather than the one calculated with the minimum Euclidean distance (referring to high SNR asymptotic case).

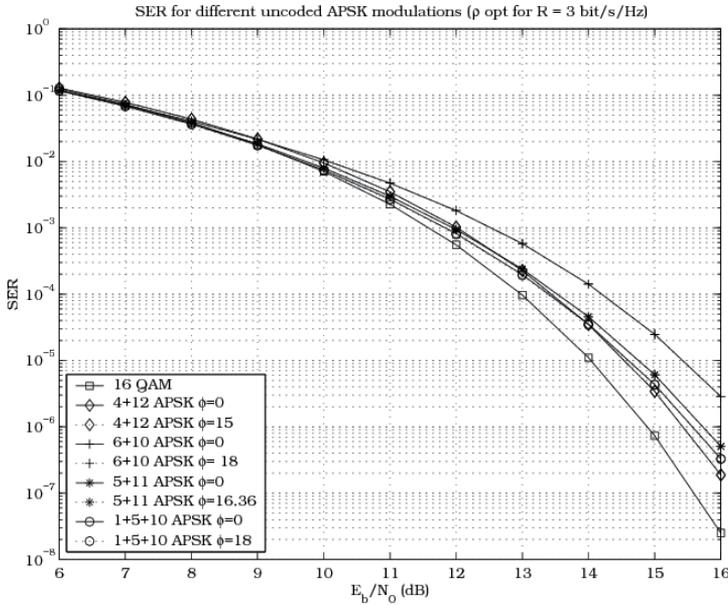


Fig. 4. Union bound on the uncoded SER for several 16-APSK modulations.

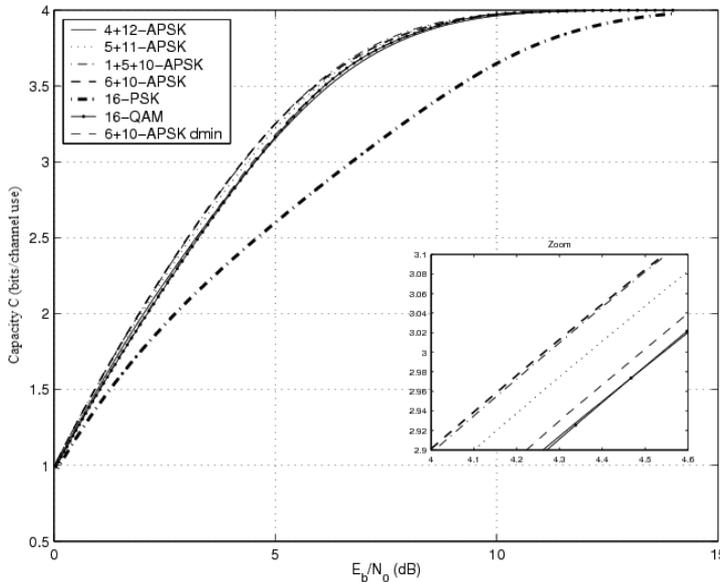


Fig. 5. Capacity of optimized 16-APSK constellations vs. 16-QAM and 16-PSK.

Table 1 below provides the optimized 16-APSK parameters for various coding rates r giving an optimum constellation for each given spectral efficiency R .

| Modulation Order | Coding Rate r | Spectral Efficiency R (bps/Hz) | ρ_2^{opt} |
|------------------|-----------------|-------------------------------------|----------------|
| 4+12-APSK | 2/3 | 2.67 | 3.15 |
| 4+12-APSK | 3/4 | 3.00 | 2.85 |
| 4+12-APSK | 4/5 | 3.20 | 2.75 |
| 4+12-APSK | 5/6 | 3.33 | 2.70 |
| 4+12-APSK | 8/9 | 3.56 | 2.60 |
| 4+12-APSK | 9/10 | 3.60 | 2.57 |

Table 1. Optimized parameters for equiprobable 16-APSK constellation.

3.2.2 Numerical results for 32-APSK

Similarly, Fig. 6 presents the modulation constrained capacity of optimized 4+12+16-APSK (with the corresponding optimal values of ρ_2, ρ_3) compared to 32-QAM and 32-PSK. Again we observe slight capacity gain of 32-APSK over 32-QAM constellations for moderate to low SNR. The 32-APSK performance advantage versus 32-PSK is much more evident. Other 32-APSK constellations with different distribution of points in the three rings did not provide significantly better results.

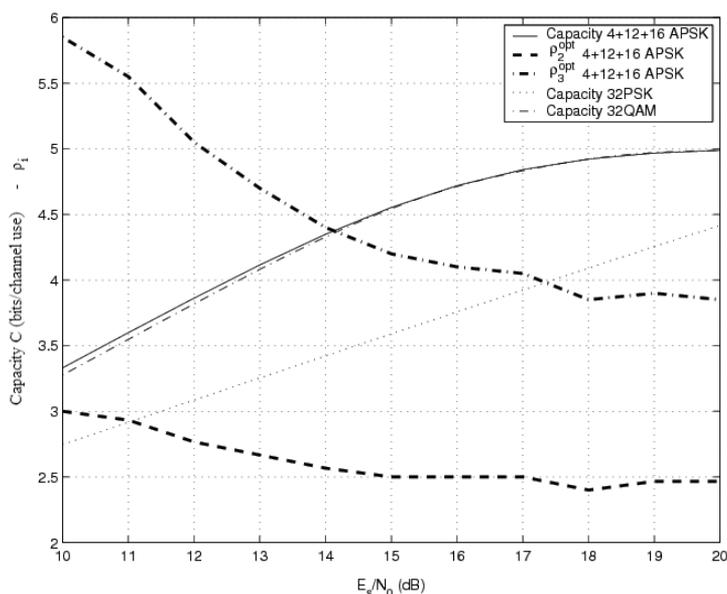


Fig. 6. Capacity and ρ^{opt} for the optimized 32-APSK constellations vs. 32-QAM and 32-PSK.

Table 2 below provides the optimized 32-APSK parameters for various coding rates r giving an optimum constellation for each given spectral efficiency R .

Note that the optimization results for 16- and 32-APSK reported in Tables 1 and 2 above have been adopted by the commercial standards related to satellite digital video broadcasting [DVB-S2, 2005; DVB-SH, 2007; GMR-1 3G, 2008; IPoS, 2006; Yuhai Shi *et al.*, 2008].

| Modulation Order | Coding Rate r | Spectral Efficiency R (bps/Hz) | ρ_2^{opt} | ρ_3^{opt} |
|------------------|-----------------|-------------------------------------|----------------|----------------|
| 4+12+16-APSK | 3/4 | 3.75 | 2.84 | 5.27 |
| 4+12+16-APSK | 4/5 | 4.00 | 2.72 | 4.87 |
| 4+12+16-APSK | 5/6 | 4.17 | 2.64 | 4.64 |
| 4+12+16-APSK | 8/9 | 4.44 | 2.54 | 4.33 |
| 4+12+16-APSK | 9/10 | 4.50 | 2.53 | 4.30 |

Table 2. Optimized parameters for equiprobable 32-APSK constellation.

3.2.3 Numerical results for 64-APSK

Fig. 7 presents the modulation constrained capacity along with the Shannon capacity bound versus the operating SNR for the 4+12+20+28-APSK constellation. In this case, the SNR range corresponds to coding rates r from 0.7 to 0.9. As can be observed, the penalty with respect to the Shannon limit ranges between 0.5 and 1.5 dB within the SNR range of interest. In the same figure, depicted is also the capacity penalty due to the suboptimal 4+12+20+28-APSK signal constellation setting if the minimum Euclidean distance maximization results ($\rho_2=2.73$, $\rho_3=4.52$ and $\rho_4=6.31$ [Benedetto *et al.*, 2005]) are instead taken into account. These results refer to the high SNR asymptotic case and so they are not optimized over the whole SNR range for each given spectral efficiency R . The latter statement is also illustrated in Fig. 7 where the optimal results in terms of mutual information maximization approach converge to the ones optimized based on minimum Euclidean distance maximization approach at high SNR values. It can also be observed that there is a small gain (of about 0.25 dB at SNR corresponding to coding rate $r=0.7$ which gets even less as the SNR - or r or R - increases) in using the optimized constellation for every R , rather than the calculated one with the minimum Euclidean distance criterion. Moreover, it can be seen that the relatively large penalty with respect to the Shannon limit within the SNR range of interest still remains in this case, as well.

Based on the findings in the case of 4+12+20+28-APSK above, there is no significant capacity gain in using the optimized constellation for each spectral efficiency R rather than the calculated one with the minimum Euclidean distance (or high SNR) criterion. Therefore, for the 4+12+16+32-APSK case, we have simply tested the previous optimization approach at high SNR and considered the (slightly) suboptimal results obtained over the whole SNR range for each coding rate r or spectral efficiency R . Following this approach, the resulting capacity achieved along with the Shannon capacity bound is plotted versus the operating SNR in Fig. 8. The SNR range depicted corresponds to coding rates r from 0.7 to 0.9. As can be seen, the penalty with respect to the Shannon limit ranges between 0.75 and 1.5 dB within the SNR range of interest.

From Fig. 7 and Fig. 8, it can be observed that there is no significant difference between the capacities of 4+12+20+28-APSK and 4+12+16+32-APSK constellations. Therefore, the preference of one constellation over the other has to be based on other performance criteria such as the signal Peak-to-Average Power Ratio (PAPR), synchronization issues, impact on nonlinearity, etc [De Gaudenzi *et al.*, 2006b].

Table 3 provides the optimized parameters taken into account in Fig. 7 and Fig. 8 for the cases of 4+12+20+28-APSK and 4+12+16+32-APSK constellations, respectively, for each

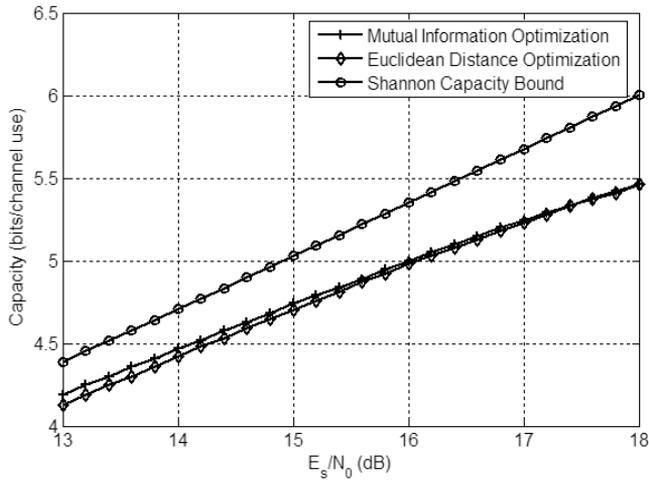


Fig. 7. Capacity of 64-APSK (4+12+20+28) constellation obtained through mutual information maximization and minimum Euclidean distance maximization approaches. Comparison with Shannon capacity bound.

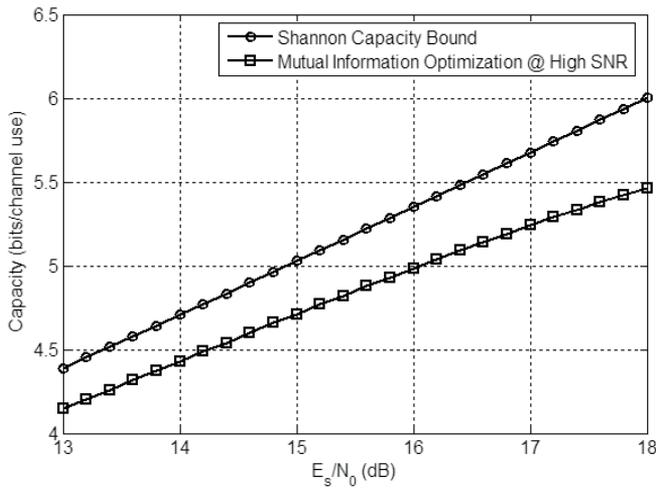


Fig. 8. Capacity of 64-APSK (4+12+16+32) constellation based on suboptimal parameters obtained for high SNR (minimum Euclidean distance maximization) and comparison with Shannon capacity bound.

given spectral efficiency R . The coding rates reported refer to the ones specified in the CCSDS Orange Book [CCSDS, 2007] which correspond to the ACM formats 25, 26, 27. As can be seen, only the optimized values of the relative radius of the ℓ -th ring with respect to the inner ring, ρ_ℓ^{opt} ($\ell = 2, 3, 4$) are of rather importance and, therefore, reported. This is because there is no noticeable dependence on the relative phase shift φ as is the case also for 16- and 32-APSK. Particularly, in the case of 4+12+16+32-APSK, the (slightly) suboptimal parameters obtained for constellation design at high SNR are provided.

| Modulation Order | Coding Rate r | Spectral Efficiency R (bps/Hz) | ρ_2^{opt} | ρ_3^{opt} | ρ_4^{opt} |
|------------------|-----------------|----------------------------------|----------------|----------------|----------------|
| 4+12+20+28-APSK | 0.798 | 4.79 | 2.62 | 4.58 | 7.00 |
| 4+12+20+28-APSK | 0.841 | 5.05 | 2.58 | 4.40 | 6.56 |
| 4+12+20+28-APSK | 0.896 | 5.38 | 2.50 | 4.14 | 6.00 |
| 4+12+16+32-APSK | - | - | 2.60 | 4.20 | 6.00 |

Table 3. Optimized parameters for equiprobable 64-APSK constellation.

3.3 Non-Equiprobable constellation optimization

In the case of non-equiprobable constellation points, which has not been considered so far in any of the APSK related standards [DVB-S2, 2005; DVB-SH, 2007; GMR-1 3G, 2008; IPoS, 2006; Yuhai Shi *et al.*, 2008], shaping of M -APSK constellations is examined here in order to achieve the so called “shaping gain” [Calderbank & Ozarow, 1990; Forney & Wei, 1989]. To this end, assuming equiprobable constellation points on each ℓ -th ring which allows different *a priori* probabilities on different rings, a new APSK constellation design optimization problem is formulated and numerically solved in order to calculate the *a priori* probabilities of constellation points on each ℓ -th ring and also to calculate the corresponding shaping gain. The possible achieved shaping gain allows the reduction of the relatively large penalty with respect to the Shannon limit experienced with equiprobable constellation points (as shown in Section 3.2.3). The main idea behind constellation shaping is that signals with large norm are used less frequently than signals with small norm, thus improving the overall gain by adding shaping gain to the original coding gain. Theoretically, when constellation points are selected according to a continuous Gaussian distribution at every dimension, the maximum achievable shaping gain in the limit for infinite transmission rates is 1.53 dB [Calderbank & Ozarow, 1990; Forney & Wei, 1989]. Practically, a smaller shaping gain can be achieved in finite constellations as is the case in M -APSK constellations. From the above, it follows that the shaping gain achieved in M -APSK constellations increases with the cardinality of signal set M and so it is expected to be greater in the case of 64-APSK than in the cases of 16- and 32-APSK.

Considering that the constellation points on each ring are equiprobable but the *a priori* symbol probability P_ℓ associated per each ℓ -th ring is different so that $\sum_{\ell=1}^{n_r} n_\ell P_\ell = 1$, and fixing the values of the relative radius and phase shift of each ℓ -th ring with respect to the inner ring to the optimized ones found in the equiprobable case (see Section 3.2 and Tables 1-3), a new constellation design optimization problem is formulated, which allows the calculation of the shaping gain. The cost function in this case is given by [Ungerboeck, 1982]

$$f_{non-eq}(\mathfrak{X}) = I_{non-eq}(X; Y) = - \sum_{k=0}^{M-1} Q(k) E_w \left\{ \log_2 \left[\sum_{i=0}^{M-1} Q(i) \exp \left[- \frac{E_s}{N_0} \left(|x^k + w - x^i|^2 - |w|^2 \right) \right] \right] \right\} \quad (13)$$

where $Q(k)$ denotes the *a priori* probability associated with each point $x^k \in \mathfrak{X}$. Equivalently,

$$Q(k) = \begin{cases} P_1 & k = 0, 1, \dots, n_1 - 1 & (\text{Ring } \ell = 1) \\ P_2 & k = n_1, n_1 + 1, \dots, n_1 + n_2 - 1 & (\text{Ring } \ell = 2) \\ P_3 & k = n_1 + n_2, n_1 + n_2 + 1, \dots, n_1 + n_2 + n_3 - 1 & (\text{Ring } \ell = 3) \\ P_4 & k = n_1 + n_2 + n_3, n_1 + n_2 + n_3 + 1, \dots, n_1 + n_2 + n_3 + n_4 - 1 & (\text{Ring } \ell = 4) \\ \vdots & \vdots & \vdots \\ P_{n_R} & k = M - n_R, M - n_R + 1, \dots, M - 1 & (\text{Ring } \ell = n_R) \end{cases} \quad (14)$$

Thus, optimization over probability distribution is pursued in this case and the respective optimization problem to be solved is formulated as

$$C_{non-eq}^* = \max_{P_1, P_2, \dots, P_{n_R}} f_{non-eq}(\mathfrak{X}) \quad (15)$$

where (as opposed to the equiprobable case addressed in Section 3.2) the parameters \mathbf{p} , $\mathbf{\varphi}$ are now assumed fixed and the parameters to be optimized are the *a priori* probabilities P_ℓ ($\ell = 1, \dots, n_R$) which satisfy the condition $\sum_{\ell=1}^{n_R} n_\ell P_\ell = 1$. Note that, since the probabilities P_ℓ are now varying, the signal set \mathfrak{X} needs to be normalized in energy accordingly so that unit power is maintained [Calderbank & Ozarow, 1990]. C_{non-eq}^* in (15) is numerically computed using the Gauss-Hermite quadrature rules [Abramowitz & Stegun, 1964]. Numerical calculations of (15) for 16-, 32- and 64-APSK constellations are provided next in Sections 3.3.1-3.3.3.

It is worth noting that in this new optimization problem, the number of the (free) optimization parameters is equal to the number of the *a priori* probabilities P_ℓ ($\ell = 1, \dots, n_R$) whereas the rest parameters of normalized ring radii $\rho_\ell = r_\ell / r_1$ are constrained (and fixed) to their optimum values obtained in Section 3.2. That is, the formulated optimization problem does not consider the optimum solution where all key optimization parameters (P_ℓ and ρ_ℓ) are considered free (in that case, the complexity of the optimization problem would significantly increase). However, it provides a good approximation to the near optimum solution for the *a priori* probabilities P_ℓ taking into account prior work on the optimization of the normalized ring radii ρ_ℓ . It also provides a rapid and efficient way to calculate the shaping gain achieved in each APSK mode as an upper bound on the potential spectral efficiency improvement. Note though that issues such as a possible increase in the transmit signal PAPR may limit the feasibility of achieving such gain.

3.3.1 Numerical results for 16-APSK

Following the optimization approach described above for the non-equiprobable 4+12-APSK constellation (its equiprobable 4+12-APSK counterpart refers to the standardized 16-APSK mode [DVB-S2, 2005; DVB-SH, 2007; GMR-1 3G, 2008; IPoS, 2006; Yuhai Shi *et al.*, 2008]), we obtain the near-optimal *a priori* probabilities P_1 , P_2 for SNR operating points corresponding to coding rates r in the range of 2/3 to 9/10. The optimized constellation settings \mathbf{p} , $\mathbf{\varphi}$ previously found in the equiprobable case and reported in Table 1 have been considered fixed. These optimization results are reported in Table 4 and illustrated in Fig. 9, as well. For the sake of comparison, the Shannon capacity bound and the capacity results for the equiprobable case have been plotted, as well. As can be seen, the constellation shaping

decreases the penalty with respect to the Shannon capacity bound. Namely, for spectral efficiency $R=3$ bits/channel use, the penalty with respect to the Shannon limit decreases from 0.9 dB (equiprobable case) to 0.6 dB (non-equiprobable case), whereas for $R=3.5$ bits/channel use, it decreases respectively from 1.5 dB to 1.2 dB. Therefore, the shaping gain for $R=3$ and 3.5 bits/channel use is about 0.3 dB in both cases.

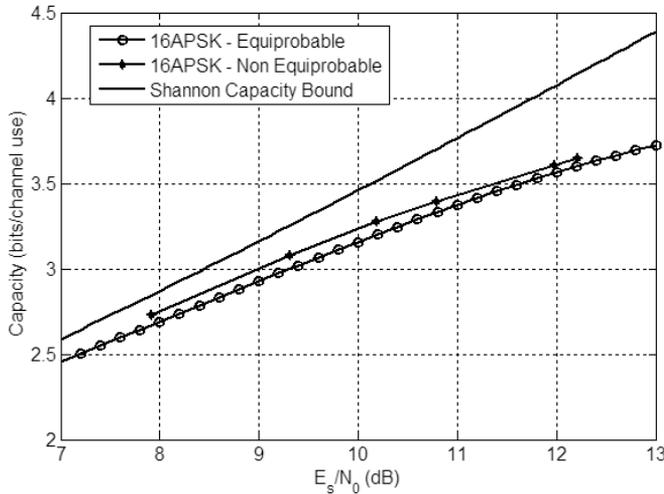


Fig. 9. Capacity and shaping gain of non-equiprobable 16-APSK (4+12) constellation.

| Modulation Order | Coding Rate r | Spectral Efficiency R (bps/Hz) | P_1^{opt} | P_2^{opt} |
|------------------|-----------------|-------------------------------------|-------------|-------------|
| 4+12-APSK | 2/3 | 2.73 | 0.116 | 0.045 |
| 4+12-APSK | 3/4 | 3.08 | 0.115 | 0.045 |
| 4+12-APSK | 4/5 | 3.27 | 0.109 | 0.047 |
| 4+12-APSK | 5/6 | 3.40 | 0.105 | 0.048 |
| 4+12-APSK | 8/9 | 3.61 | 0.095 | 0.052 |
| 4+12-APSK | 9/10 | 3.64 | 0.093 | 0.052 |

Table 4. Optimized parameters for non-equiprobable 16-APSK constellation.

3.3.2 Numerical results for 32-APSK

Similarly, for the non-equiprobable 4+12+16-APSK constellation (its equiprobable 4+12+16-APSK counterpart refers to the standardized 32-APSK mode [DVB-S2, 2005; DVB-SH, 2007; GMR-1 3G, 2008; IPoS, 2006; Yuhai Shi *et al.*, 2008]), we obtain the near-optimal *a priori* probabilities P_1, P_2, P_3 for SNR operating points corresponding to coding rates r in the range of 3/4 to 9/10. The given optimized constellation settings $\mathbf{p}, \mathbf{\varphi}$ are fixed to those previously found in the equiprobable case and reported in Table 2. These optimization results are reported in Table 5 and illustrated in Fig. 10, as well. Similar results are achieved as in the case of non-equiprobable 4+12-APSK. As an illustration, the shaping gain for spectral efficiency $R=4$ and 4.5 bits/channel use is about 0.3 dB in both cases.

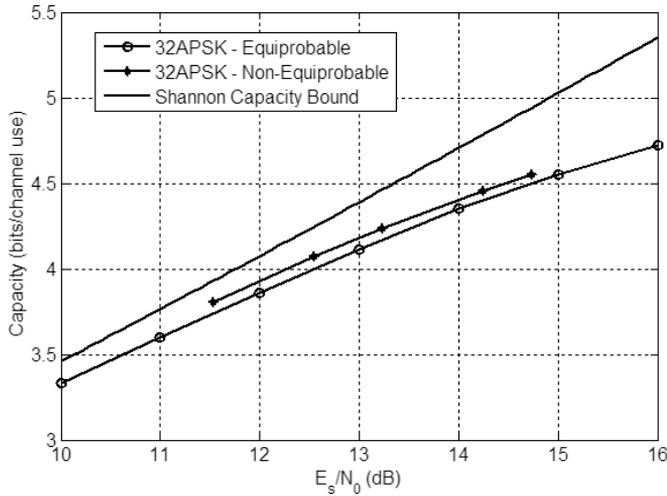


Fig. 10. Capacity and shaping gain of non-equiprobable 32-APSK (4+12+16) constellation.

| Modulation Order | Coding Rate r | Spectral Efficiency R (bps/Hz) | P_1^{opt} | P_2^{opt} | P_3^{opt} |
|------------------|-----------------|----------------------------------|-------------|-------------|-------------|
| 4+12+16-APSK | 3/4 | 3.81 | 0.055 | 0.042 | 0.017 |
| 4+12+16-APSK | 4/5 | 4.07 | 0.056 | 0.040 | 0.019 |
| 4+12+16-APSK | 5/6 | 4.24 | 0.055 | 0.038 | 0.020 |
| 4+12+16-APSK | 8/9 | 4.46 | 0.052 | 0.037 | 0.022 |
| 4+12+16-APSK | 9/10 | 4.55 | 0.049 | 0.036 | 0.023 |

Table 5. Optimized constellation parameters for non-equiprobable 32-APSK.

3.3.3 Numerical results for 64-APSK

Similarly, for the non-equiprobable 4+12+20+28-APSK signal constellation set, we obtain the near optimal *a priori* probabilities P_1, P_2, P_3, P_4 for SNR operating points corresponding to coding rates r in the range of 0.82 to 0.91 for the given optimized constellation settings ρ, φ previously found in the equiprobable case and reported in Table 3. These new optimization results are reported in Table 6 and illustrated in Fig. 11, as well. As can be seen, the constellation shaping decreases the penalty with respect to the Shannon capacity bound. Namely, for spectral efficiency $R=5$ bits/channel use, the penalty with respect to the Shannon limit decreases from 1 dB (equiprobable case) to 0.5 dB (non-equiprobable case), whereas for $R=5.5$ bits/channel use, it decreases respectively from 1.5 dB to 1.1 dB. Therefore, the shaping gain for $R=5$ and 5.5 bits/channel use is 0.5 dB and 0.4 dB, respectively.

Moreover, in the case of non-equiprobable 4+12+16+32-APSK signal constellation set, the near optimal *a priori* probabilities P_1, P_2, P_3, P_4 are obtained following a similar optimization approach whose results are presented in Table 7 and Fig. 12, as well. As an illustration in this case, the shaping gain achieved is 0.5 dB for spectral efficiency $R=5$ bits/channel use, whereas for $R=5.5$ bits/channel use, it is 0.3 dB.

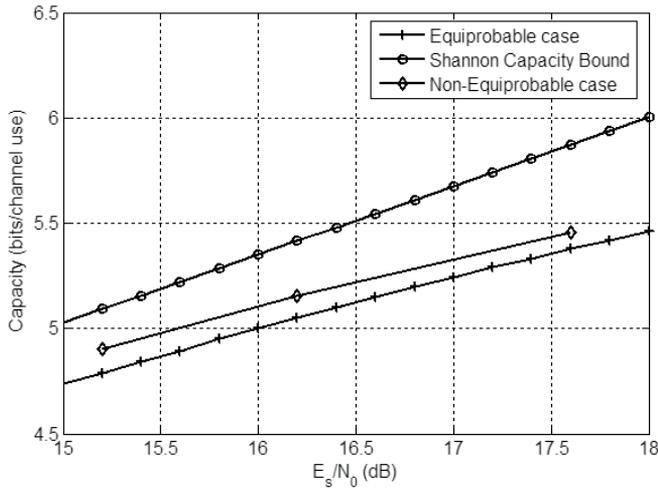


Fig. 11. Capacity and shaping gain of non-equiprobable 64-APSK (4+12+20+28) constellation.

| Modulation Order | Coding Rate r | Spectral Efficiency R (bps/Hz) | P_1^{opt} | P_2^{opt} | P_3^{opt} | P_4^{opt} |
|------------------|-----------------|----------------------------------|-------------|-------------|-------------|-------------|
| 4+12+20+28-APSK | 0.817 | 4.90 | 0.040 | 0.025 | 0.020 | 0.0050 |
| 4+12+20+28-APSK | 0.859 | 5.15 | 0.045 | 0.025 | 0.015 | 0.0079 |
| 4+12+20+28-APSK | 0.910 | 5.46 | 0.040 | 0.025 | 0.015 | 0.0086 |

Table 6. Optimized constellation parameters for non-equiprobable 64-APSK (4+12+20+28).

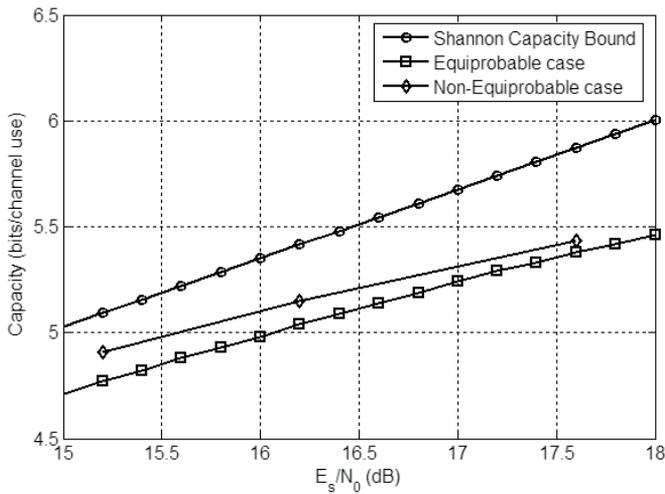


Fig. 12. Capacity and shaping gain of non-equiprobable 64-APSK (4+12+16+32) constellation.

Note that the present chapter does not go into further details on how to actually achieve the calculated shaping gain in the examined non-equiprobable 16-, 32- and 64-APSK modes; however, the interested readership might refer elsewhere in this regard, e.g., in [Calderbank & Ozarow, 1990; Forney & Wei, 1989].

| Modulation Order | Coding Rate r | Spectral Efficiency R (bps/Hz) | P_1^{opt} | P_2^{opt} | P_3^{opt} | P_4^{opt} |
|------------------|-----------------|----------------------------------|-------------|-------------|-------------|-------------|
| 4+12+16+32-APSK | 0.818 | 4.91 | 0.051 | 0.028 | 0.019 | 0.0049 |
| 4+12+16+32-APSK | 0.858 | 5.15 | 0.047 | 0.025 | 0.019 | 0.0065 |
| 4+12+16+32-APSK | 0.905 | 5.43 | 0.040 | 0.022 | 0.018 | 0.0090 |

Table 7. Optimized constellation parameters for non-equiprobable 64-APSK (4+12+16+32).

4. Satellite channel distortion pre-compensation

Generally speaking, the HPA nonlinearity has two major effects. First, the varying constellation is distorted as the constellation points are mapped by the HPA nonlinear characteristic to a different point (amplitude, phase). Furthermore, the relative positions of the constellation points change. As briefly outlined above and discussed in detail next, this impairment can be efficiently reduced by ad-hoc pre-compensation at the transmitter. Despite its hardware complexity impact, commercial satellite modems have already adopted advanced dynamic pre-compensation techniques for standardized 16- and 32-APSK modes [Newtec, 2009]. Also, experimental laboratory measurement results for static pre-distortion techniques for standardized 16- and 32-APSK modes are reported in [Bischl *et al.*, 2009].

Second, ISI appears at the receiver as the HPA, although memoryless, is driven by a signal with controlled ISI due to the presence of the modulator SRRC filter. This leads to an overall nonlinear channel with memory. As a consequence, the demodulator SRRC is not matched anymore to the incoming signal. This issue is to be tackled mainly with a pre-equalization at the modulator [Karam & Sari, 1990] or equalization at the demodulator or a combination of the two techniques.

Although results presented in Section 3.2.1 indicated a slight superiority of 6+10-APSK, for nonlinear transmission over an amplifier, 4+12-APSK is preferable to 6+10-APSK because the presence of more points in the outer ring allows us to maximize the HPA DC power conversion efficiency. It is better to reduce the number of inner points, as they are transmitted at a lower power, which corresponds to a lower DC efficiency (the HPA power conversion efficiency is monotonic with the input power drive up to its saturation point). Fig. 13 shows the Probability Distribution Function (PDF) of the transmitted signal envelope for 16-QAM, 4+12-APSK, 6+10-APSK, 5+11-APSK, and 16-PSK; the shaping filter is an SRRC with roll-off factor $\alpha=0.35$. As can be observed, the 4+12-APSK envelope is more concentrated around the outer ring amplitude than 16-QAM and 6+10-PSK, being remarkably close to the 16-PSK case. This shows that the selected constellation represents a good trade-off between 16-QAM and 16-PSK, with error performance close to 16-QAM, and

resilience to nonlinearity close to 16-PSK. Therefore, 4+12-APSK is preferable to the rest of 16-ary modulations considered. Similar advantages have been observed for 32-APSK compared to 32-QAM [De Gaudenzi *et al.*, 2006a].

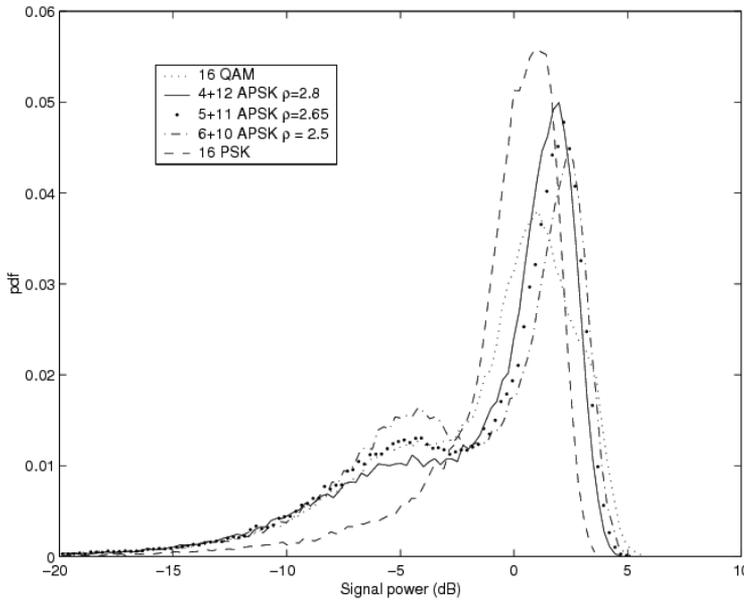


Fig. 13. Simulated histogram of the transmitted signal envelope power for 16-ary APSK, QAM and PSK constellations.

4.1 Static pre-compensation

The simplest approach for counteracting the HPA nonlinear characteristic is to modify the complex-valued constellation points at the modulator side. Thanks to the multiple-ring nature of the APSK constellation, pre-compensation is easily done by a simple modification of the parameters ρ , ϕ . The known AM/AM and AM/PM HPA characteristics (see, e.g., Fig.1) are exploited in order to obtain a good replica of the desired signal constellation geometry after the HPA, as if it had not suffered any distortion. This can be simply obtained by artificially increasing the relative radii ρ_ℓ and modifying the relative phases ϕ_ℓ ($\ell = 1, \dots, n_R$) at the modulator side.

The calculation of the pre-distorted constellation parameters can be made with the technique described in [De Gaudenzi & Luise, 1995] for the computation of the distorted constellation center of mass (*centroids*) seen at the demodulator matched filter output. With knowledge of the satellite link characteristics, the static pre-compensation parameters can be calculated off-line by taking the following steps:

1. Generation of S blocks of W symbols over which the Symbol Matched Filter (SMF) centroids are computed (transmission in the absence of white noise);
2. Computation of the error signal at the end of each block;
3. Pre-distorted constellation point update.

The latter task can be readily achieved through an iterative Least Mean Square (LMS) type of algorithm illustrated by the following set of equations:

$$\left\{ \begin{array}{l}
 \left| x_{pre}^{(n)}(s+1) \right| = \left| x_{pre}^{(n)}(s) - \gamma_r \cdot e_c^{(n)}(s) \right| \\
 \arg \left(x_{pre}^{(n)}(s+1) \right) = \arg \left(x_{pre}^{(n)}(s) \right) - \gamma_\phi \cdot \psi(s) \\
 e_c^{(n)}(s) = r_c^{(n)}(s) \cdot \exp \left(j\theta_c^{(n)}(s) \right) - \left| x^{(n)} \right| \\
 r_c^{(n)}(s) \cdot \exp \left(j\theta_c^{(n)}(s) \right) = \frac{1}{W} \cdot \sum_{k \in I^n, sW+1 \leq k \leq (s+1)W} y(k) \\
 \psi(s) = \begin{cases} \arg \left(e_c^{(n)}(s) \right) - 2\pi & , \arg \left(e_c^{(n)}(s) \right) > \pi \\ \arg \left(e_c^{(n)}(s) \right) + 2\pi & , \arg \left(e_c^{(n)}(s) \right) < -\pi \\ \arg \left(e_c^{(n)}(s) \right) & , \left| \arg \left(e_c^{(n)}(s) \right) \right| \leq \pi \end{cases}
 \end{array} \right. \quad (16)$$

where the index n refers to the constellation point, $l^{(n)}$ indicates the conditioning to the constellation point n , s refers to the iteration step of the algorithm, $y(k)$ represents the k -th SMF output complex sample, $x^{(n)}$ represents the APSK complex constellation reference point, $r_c^{(n)}(s)$ and $\theta_c^{(n)}(s)$ are the modulus and the phase of the SMF output complex n -th centroid computed at step s , $x_{pre}^{(n)}(s)$ is the pre-distorted n -th constellation point computed at step s , γ_r and γ_ϕ the adaptation steps for the pre-distorted constellation point modulus and the phase, respectively.

As an illustration, based on Fig. 14(a), the optimal 4+12-APSK pre-distortion parameters are $\rho'_2 = 3.5$ and $\Delta\phi = 25$ deg for an IBO=3 dB whereas $\rho'_2 = 3.7$ and $\Delta\phi = 27$ deg for a smaller IBO=2 dB. As expected, the pre-distorted constellation is expanded, e.g., $\rho'_2 > \rho_2$. For the new constellation points x' , they are $x' \in \mathcal{X}$ i.e., they still follow (7) but with new radii r'_i such that $F(r'_i) = r_i$ ($i=1,2$).

Concerning the phase, it is possible to pre-correct for the effect of the HPA on the phase HPA between inner and outer rings through a simple change in the relative phase shift by $\phi'_2 = \phi_2 + \Delta\phi$, with $\Delta\phi = \phi(r'_2) - \phi(r'_1)$. These operations can be readily implemented in the digital modulator by simply modifying the reference constellation parameters ρ' , ϕ' with *no hardware complexity impact* or out-of-band emission increase at the linear modulator output. However, it should be remarked that static pre-compensation may imply an increase of the transmitted signal PAPR, which in turn may have a potential impact of the out-of-band emissions after the HPA. The compensation effort is shifted into the modulator side, allowing the use of an optimal demodulator/decoder for AWGN channels even when the HPA is close to saturation. Based on (1), the signal at the modulator output is then

$$s_T^{pre} = \sqrt{P} \sum_{k=0}^{L-1} x'(k) p_T(t - kT_s) \quad (17)$$

where now $x'(k) \in \mathcal{X}'$ being the pre-distorted symbols with ρ'_ℓ, ϕ'_ℓ ($\ell=1, \dots, n_R$).

To show the effect of this practical static pre-distortion technique, the scatter diagram at the output of the SRRC (with roll-off factor $a=0.35$) receiver matched filter is shown in Fig. 14(b) for 16-APSK with IBO=2 dB. For clarity, the scatter diagram at the SMF has been obtained in

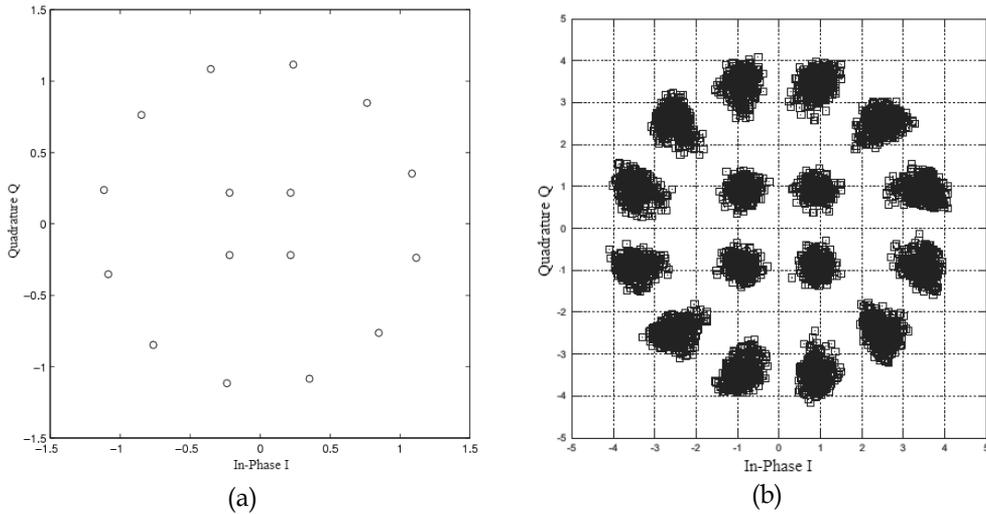


Fig. 14. Static pre-compensation for 4+12-APSK constellation: (a) Modulator output roll-off factor $a=0.35$ with static pre-compensation ($\rho_2=3.7$ and $\Delta\varphi=27$ deg); (b) Demodulator SRRC filter output noiseless scatter diagram in the nonlinear channel for IBO=2 dB, roll-off factor $a=0.35$ with static pre-compensation ($\rho_2=3.7$ and $\Delta\varphi=27$ deg).

the absence of AWGN. Despite the strong channel nonlinearity, the center of mass corresponding to the scattered diagram closely follows the optimum 4+12-APSK constellation, for which the optimum parameters are $\rho_2=2.7$ and $\varphi=0$.

Measurements showed that the HPA characteristic sensitivity to temperature or aging results in a limited change of gain but not in a modification of the AM/AM, AM/PM characteristics shape. The limited gain variations are compensated by the satellite transponder Automatic Level Control (ALC) device, thus off-line pre-compensation has a long term value. If required, the compensated parameters can be adapted to track larger slow variations in HPA characteristic due to aging.

Further experimental laboratory measurement results for such static pre-distortion techniques for the standardized 16- and 32-APSK modes are reported in [Bischl *et al.*, 2009].

4.2 Dynamic pre-compensation

Static pre-distortion is able to compensate for the constellation warping effects but not for the clustering phenomenon. To this end, an extension of the static pre-distortion algorithm described above has been adopted and further improved in [Casini *et al.*, 2004]. The dynamic pre-distortion algorithm takes into account the memory of the channel, conditioning the pre-distorted modulator constellation not only to the current symbol transmitted but also to the $(L-1)/2$ preceding and $(L-1)/2$ following symbols (L being the number of symbols in total). This calls for an increased look-up table size of ML points. By exploiting the APSK constellation symmetry, the amount of memory size can be reduced to $3ML/16$.

The advantage provided by the dynamic pre-compensation on the reduction of the clustering effect can be assessed by observing the scatter diagram at the SMF sampler output shown in Fig. 15(b) and 15(c). Looking at Fig. 15(a), it appears that the clustering effect reduction is obtained at the expenses of an increased outer constellation points' amplitude. This corresponds to a peak to average signal envelope ratio of 10 dB compared to 6.77 dB in

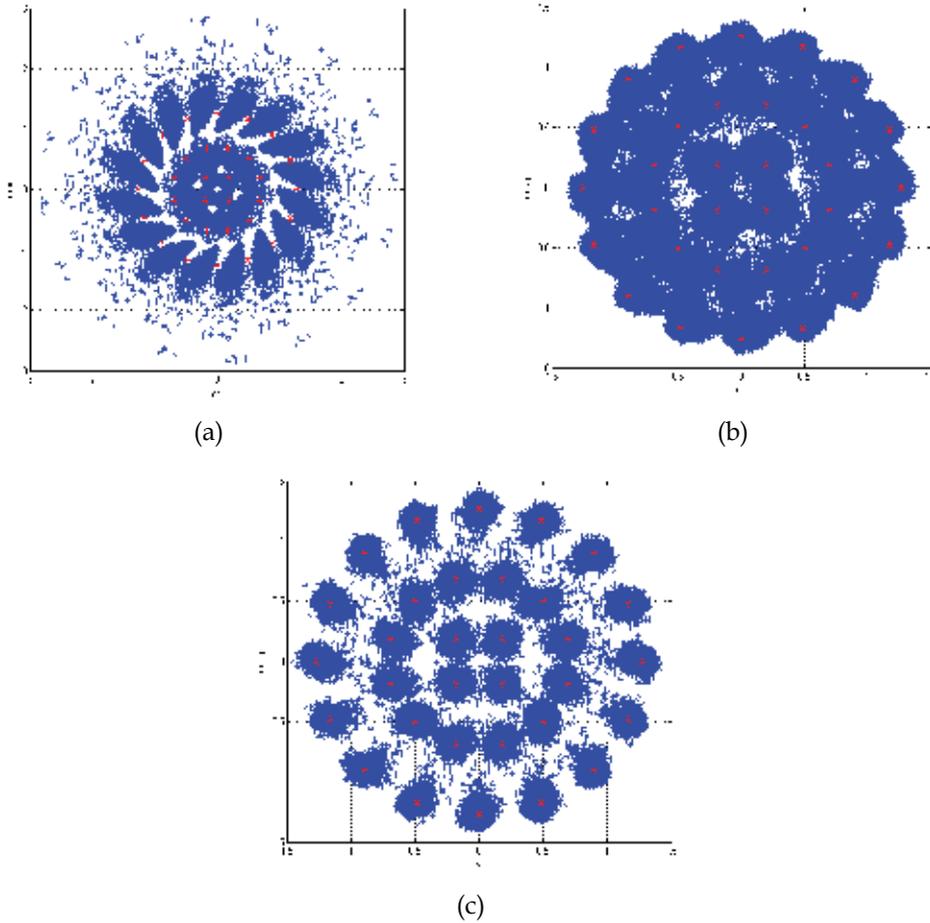


Fig. 15. Dynamic pre-distortion on 32APSK, with IBO=3.6 dB, $W=5000$ symbols, $s=85$ blocks, $L=3$: a) blue crosses pre-distorted constellation, red circles nominal constellation, b) blue crosses constellation centroids at the demodulator SMF, red circles nominal constellation without pre-compensation, c) blue crosses constellation centroids at the demodulator SMF, red circles nominal constellation with dynamic pre-compensation.

the case of static pre-compensation. The dynamic pre-distortion induced PAPR increase has two main drawbacks: a) the augmentation of the HPA OBO which negatively affects the overall system efficiency, b) the possible impact on the HPA TWTA safe operation due to the higher PAPR making the instantaneous signal power occasionally well beyond the saturation point. Based on these considerations, an improved dynamic pre-distortion approach has been devised. The quantity to minimize is in fact not the Root Mean Squared (RMS) of the centroids conditioned to a certain data pattern but rather the total link degradation D_{TOT} (in dB) given by [Casini *et al.*, 2004]

$$D_{TOT}(s) = \frac{E_s}{N_0} \Big|_{req}^{NL}(s) - \frac{E_s}{N_0} \Big|_{req}^{AWGN}(s) + OBO(s) \quad (18)$$

where $(E_s/N_0)_{req}^{NL}$ and $(E_s/N_0)_{req}^{AWGN}$ are the average symbol energy over noise density required to achieve the target Frame Error Rate (FER) in the nonlinear and linear AWGN channel, respectively. By using the Gaussian approximation for the ISI at the SMF sampler output one can write:

$$\left[\frac{E_s}{N_0} \right]_{req}^{NL}(s) = \left[\frac{E_s}{N_0} \right]_{req}^{AWGN}(s) \left[1 + \frac{\sigma_{ISI}^2(s)}{N_0} \right] \quad (19)$$

where $\sigma_{ISI}^2(s)$ represents the ISI power at the output of the SMF averaged over the constellation points at step s , that is, in mathematical terms

$$\sigma_{ISI}^2(s) = \frac{1}{M} \sum_{n=1}^M \sum_{k \in I^{(n)}, sW+1 \leq k \leq (s+1)W} \{z(k) - c^{(n)}\}^2 \quad (20)$$

By replacing (19) into (18) we get

$$D_{TOT}(s) = \left[1 + \frac{\sigma_{ISI}^2(s)}{N_0} \right] OBO(s) \quad (21)$$

Eq. (21) is valid in the case of absence of intra-system co-channel and adjacent channel interference. These two quantities, in fact, depend on the signal energy so that an increase of the OBO does not affect the signal-to-interference ratio. It is easy to show that, if I_0^{SAT} denotes the total power of the intra-system interference samples at the SMF output with $OBO=0$ dB, (21) can be generalized as

$$D_{TOT}(s) = \left[1 + \frac{\sigma_{ISI}^2(s) + \frac{I_0^{SAT}}{OBO(s)}}{N_0} \right] OBO(s) \quad (22)$$

Since I_0^{SAT} is system-dependent, next we assume a situation where intra-system interference is negligible, so that its contribution does not affect the pre-distortion optimization results. The dynamic pre-compensation is now performed as described before computing every block of W symbols also the σ_{ISI}^2 and OBO and computing (21) at each step. The dynamic pre-compensation is now stopped when the minimum of D_{TOT} is achieved. This approach ensures the best trade-off between ISI minimization and the OBO penalty due to the increased PAPR caused by dynamic pre-compensation. To speed-up process convergence and to avoid bias in the σ_{ISI}^2 estimate, the static pre-compensation is first applied then the dynamic pre-compensation is started. This approach allows beginning the dynamic pre-compensation with the SMF centroids very close to the nominal constellation points. An example of the D_{TOT} -based optimized dynamic pre-compensation is illustrated in Fig. 16. In this figure the D_{TOT} and OBO evolution versus the iteration number s is plotted for the case of 8PSK. It appears that after the minimum of D_{TOT} occurring around $s=30$ blocks, after this value the total loss is growing following the OBO growth due to the transmit constellation outer points expansion.

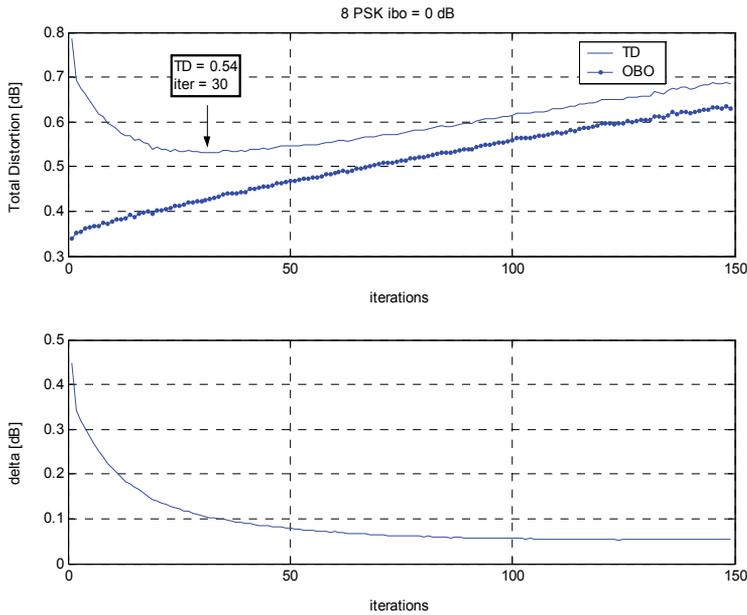


Fig. 16. D_{TOT} and OBO evolution vs. time with dynamic pre-distortion on 8PSK, with IBO = 0 dB, $W=50000$ symbols.

Performance improvements for dynamic pre-distortion are quite remarkable for high order constellations. As reported in [Casini *et al.*, 2004], in case of 16-APSK, the advantage for exploiting dynamic pre-compensation amounts to about 0.8 dB compared to static pre-compensation while the dynamic optimization brings about 0.2 dB improvement. In case of 32-APSK, the advantage for exploiting dynamic pre-compensation amounts to about 1.8 dB compared to static pre-compensation and about 3.5 dB compared to non pre-distorted constellations.

It is also worth noting that commercial satellite modems have already adopted similar advanced dynamic pre-compensation techniques [Newtec, 2009].

5. Conclusions

This chapter has presented analysis and numerical results on the design of APSK signal constellation sets which are well suited for advanced satellite digital video broadcasting systems. The APSK constellation modes examined are the 16-, 32- and 64-ary whereas the design optimization criterion employed has been the maximization of the AWGN channel mutual information. The presented analysis has taken into account both equiprobable and non-equiprobable APSK constellations where a respective optimization problem has been formulated and numerically solved. In addition, practical static and dynamic pre-distortion techniques for pre-compensation of the satellite channel nonlinearities have been addressed. In particular, the design optimization technique based on the mutual information maximization has been shown to extend the more traditional one based on the minimum Euclidean distance maximization, yielding a small but significant improvement, especially for the 16- and 32-APSK modes. Following this optimization approach, the optimal

constellation settings, more specifically, the optimum normalized radii of the APSK constellation rings, have been determined as a function of the operating SNR. Optimization results have indicated no significant dependency on the relative phase shifts between the multiple rings of APSK constellation in all cases of 16-, 32- and 64-ary modes at the target SNR operating points. The presented results for 16- and 32-APSK equiprobable constellations have already been adopted by commercial standards related to satellite digital video broadcasting.

In the case of non-equiprobable 16-, 32- and 64-APSK signal sets, constellation shaping has been examined, the near optimal *a priori* probabilities associated with each ring have been evaluated and the respective shaping gain has been calculated reaching values of up to 0.5 dB for the target SNR operating regions.

In addition, the impact of typical satellite channel nonlinearities as well as techniques to counteract their impact on the demodulator performance have been analyzed. It has been shown that coded APSK with simple digital pre-distortion techniques based on look-up tables can achieve very good performance with satellite HPA driven at saturation (16-APSK) or with limited back-off (32-APSK). Special emphasis has been put on practical dynamic pre-distortion techniques which achieve remarkable performance improvements for 16- and 32-APSK constellations and which have recently been adopted in commercial satellite modems for the relevant nonlinear channel pre-compensation.

6. References

- [Abramowitz & Stegun, 1964] Abramowitz M., and Stegun I.A. (eds.), *Handbook of Mathematical Functions*, Appl. Math. Ser. No. 55, National Bureau of Standards, Washington, D.C., 1964.
- [Alberty *et al.*, 2007] Alberty E., Defever S., Moreau C., De Gaudenzi R., Ginesi A., Rinaldo R., Gallinaro G., and Vernucci A, "Adaptive Coding and Modulation for the DVB-S2 Standard Interactive Applications: Capacity Assessment and Key System Issues", *IEEE Wireless Communications*, vol. 14, no. 4, pp. 61-69, August 2007.
- [Benedetto *et al.*, 2005] Benedetto S., Garello R., Montorsi G., Berrou C., Douillard C., Giancristofaro D., Ginesi A., Giugno L., and Luise M., "MHOMS: High-Speed ACM Modem for Satellite Applications", *IEEE Wireless Communications Magazine*, vol. 12, no. 2, pp. 66-77, April 2005.
- [Bischi *et al.*, 2009] Bischi H., Brandt H., De Cola T., De Gaudenzi R., Eberlein E., Girault N., Alberty E., Lipp S., Rinaldo R., Rislow B., Arthur Skard J., Tusch J., and Ulbricht J., "Adaptive coding and modulation for satellite broadband networks: From theory to practice", to appear in *International Journal on Satellite Communications and Networking*, DOI: 10.1002/sat.932, 2009.
- [Calderbank & Ozarow, 1990] Calderbank A.R., and Ozarow L.H., "Nonequiprobable signalling on the Gaussian channel", *IEEE Transactions on Information Theory*, vol.36, no.4, pp.726-740, July 1990.
- [Casini *et al.*, 2004] Casini E., De Gaudenzi R., and Ginesi A., "DVB-S2 modem algorithms design and performance over typical satellite channels", *International Journal on Satellite Communications and Networking*, vol. 22, no. 3, pp. 281-318, May/June 2004.
- [CCSDS, 2007] Flexible Serially Concatenated Convolutional Turbo Codes with Near-Shannon Bound Performance for Telemetry Applications, *CCSDS Orange Book*, Issue 1, September 2007, CCSDS 131.2-O-1.

- [De Gaudenzi & Luise, 1995] De Gaudenzi R., and Luise M., "Design and analysis of an all-digital demodulator for trellis coded 16-QAM transmission over a nonlinear satellite channel," *IEEE Transactions on Communications*, vol. 43, no. 2/3/4-Part I, February/March/April, 1995.
- [De Gaudenzi *et al.*, 2006a] De Gaudenzi R., Guillén i Fàbregas A. & Martinez A., "Turbo-coded APSK modulations design for satellite broadband communications", *International Journal of Satellite Communications and Networking*, vol. 24, pp. 261-281, 2006.
- [De Gaudenzi *et al.*, 2006b] De Gaudenzi R., Guillén i Fàbregas A. & Martinez A., "Performance analysis of turbo-coded APSK modulations over nonlinear satellite channels", *IEEE Transactions on Wireless Communications*, vol.5, no.9, pp. 2396-2407, September 2006.
- [DVB-S2, 2005] ETSI EN 302 307 V1.1.1 (2005-03), Digital Video Broadcasting (DVB); "Second generation framing structure, channel coding and modulation systems for broadcasting interactive services, news Gathering and other broadband satellite applications".
- [DVB-SH, 2007] ETSI EN 302 583 V1.1.1 (2007-07), Digital Video Broadcasting (DVB); "Framing structure, channel coding and modulation for Satellite services to Handheld devices (SH) below 3 GHz".
- [Forney & Wei, 1989] Forney G.D., and Wei L.F., "Multidimensional constellations - Part I: Introduction, figures of merit, and generalized cross constellations", *IEEE Journal on Selected Areas in Communications*, vol.7, no.6, pp.877-892, August 1989.
- [Future Internet, 2009] Future Internet: The Cross-ETP (European Technology Platform) Future Internet Strategic Research Agenda, Version 1.0, July 2009 [Available on-line at EU Future Internet portal: www.future-internet.eu].
- [GMR-1 3G, 2008] ETSI TS 101 376-5-4 V2.3.1 (2008-08), GMR-1 3G, "GEO-Mobile Radio Interface Specifications (Release 2); General Packet Radio Service; Part 5: Radio interface physical layer specifications; GMPRS-1".
- [Hughes, 2009] Hughes Network Systems (HNS) IP over Satellite Technology [Available on-line at: www.hughes.com].
- [IPoS, 2006] ETSI TS 102 354 V1.2.1 (2006-11), Satellite Earth Stations and Systems (SES); Broadband Satellite Multimedia (BSM); Transparent Satellite Star - B (TSS-B); "IP over Satellite (IPoS) Air Interface Specification".
- [Karam & Sari, 1990] Karam G., and Sari H., "Data predistortion techniques using intersymbol interpolation," *IEEE Transactions on Communications*, vol. 38, no. 10, pp. 1716-1723, October 1990.
- [Karam & Sari, 1991] Karam G., and Sari H., "A Data Predistortion Technique with Memory for QAM Radio Systems," *IEEE Transactions on Communications*, vol. 39, no. 2, pp 336-44, February 1991.
- [Liolis & Alagha, 2008] Liolis K.P., and Alagha N.S., "On 64-APSK Constellation Design Optimization", in Proc. 10th International Workshop on Signal Processing for Space Communications (SPSC'08), Rhodes, Greece, October 2008.
- [Newtec, 2009] Newtec EL470 IP Satellite Modem [Available on-line at: www.newtec.eu].
- [Rinaldo & De Gaudenzi, 2004a] Rinaldo R., and De Gaudenzi R., "Capacity analysis and system optimization for the forward link of multi-beam satellite broadband

- systems exploiting adaptive coding and modulation", *International Journal on Satellite Communications and Networking*, vol. 22, no. 3, pp. 401-423, May/June 2004.
- [Rinaldo & De Gaudenzi, 2004b] Rinaldo R., and De Gaudenzi R., "Capacity analysis and system optimization for the reverse link of multi-beam satellite broadband systems exploiting adaptive coding and modulation", *International Journal on Satellite Communications and Networking*, vol. 22, no. 4, pp. 425-448, July/August 2004.
- [Sung *et al.*, 2009] Sung W., Kang S., Kim P., Chang D.-I., and Shin D.-J., "Performance analysis of APSK modulation for DVB-S2 transmission over nonlinear channels", *International Journal on Satellite Communications and Networking*, vol. 27, 2009.
- [Thomas *et al.*, 1974] Thomas C.M., Weidner M.Y., and Durrani S.H., "Digital amplitude phase keying with M-ary alphabets," *IEEE Transactions on Communications*, vol. 22, no. 2, pp. 168-180, February 1974.
- [Ungerboeck, 1982] Ungerboeck G., "Channel coding with multilevel phase signals", *IEEE Transactions on Information Theory*, vol.28, no.1, pp. 55-67, 1982.
- [Yuhai Shi *et al.*, 2008] Yuhai Shi, Ming Yang, Ju Ma, and Rui Lv, "A New Generation Satellite Broadcasting System in China: Advanced Broadcasting System-Satellite", in Proc. 4th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM'08), Dalian, China, October 2008.

Non-Photo Realistic Rendering for Digital Video Intaglio

Kil-Sang Yoo, Jae-Joon Cho and Ok-Hue Cho
*Department of Advanced Image,
Graduate School of Advanced Imaging Science, Multimedia and Film,
Chung-Ang University
Republic of Korea*

1. Introduction

The recent surge of interest in non-photorealistic rendering is a testimony to our fascination with distinctive artistic style. Rendering algorithms have been introduced to mimic various classical art forms, ranging from pen-and-ink illustrations and line art drawings to expressive paintings. Classic printmaking offers another suite of styles that can be effectively applied to images. We shall be interested in transforming digital images into renderings that resemble the work of traditional engravers.

Engraving is a classic graphic technique originating in the printing industry of the fifteenth century. It is actually one of several intaglio printing techniques, whereby the image is inscribed into a metal plate with the use of sharp instruments. The resulting recessed lines are filled with ink and the image is transferred to paper by means of a press. Another common technique is etching, whereby acid is used to achieve the inscription. Although engraving has a technical definition that is closely coupled to the manner in which an image is inscribed, loose colloquial usage has come to make it synonymous with intaglio printing.

Each method places different demands on the printmaker, giving rise to distinctive styles. We shall be interested in imbuing digital images with the stylistic effects of intaglio printing, which include the familiar styles of engravings and etchings. This process is familiar to anyone who has looked at modern currency. Figure. 1 illustrates several examples of physical intaglio that conform to the characteristic styles we seek to emulate.

In our study we realized real-time digital intaglio image system. We realized digital Intaglio NPR system that can get result immediately from any kinds of image source. For example, picture images or real-time video streams from video camera. Until present, there have been many kinds of NPR systems by graphic programmers and media artists. But our algorithm will be very new approaching using traditional etching technique. We analyzed traditional etching to two directions of line. Each vertical lines and horizontal lines have all different width to express etching style image.

Rest of this paper is organized as follows: previous works are described in Section 3; Section 4 presents proposed algorithm and experimental results are shown; and conclusions are shown in Section 5.



Fig. 1. Exemplae of physical intaglio

2. Previous works

In order to place our discussion in proper context, we now review the four major printmaking processes and their histories. We shall confine our attention to the stylistic advances introduced by intaglio printing.

Printmaking dates back to the fifteenth century where it was first used to reproduce illustrations. It played a supportive role in the growing book-printing industry at that time. Great demand for technical achievement in this area has resulted in four major printmaking processes: relief printing, intaglio printing, lithography, and serigraphy. These processes are outlined below.

- **Relief printing:** A wooden block or metal plate is carved such that the nonprinting background areas are cut away below the surface, leaving the image areas of the print in relief. After ink is applied to the raised surface with a roller, paper is pressed against the block and the image is transferred.
- **Intaglio printing:** The image areas are depressed below the surface of the plate. Lines are incised into a metal plate with the use of sharp tools or acids. The plate is covered with ink, and then wiped clean leaving ink in the recessed lines. A press forces the paper into these ink-filled lines and the image is transferred.
- **Lithography:** Images are drawn in greasy crayon on a flat slab of limestone, or a metal plate. The stone is treated chemically so that ink, when rolled onto the stone, adheres only where the drawing was done. A high pressure press is used to transfer the image onto paper.
- **Serigraphy:** Image areas are drawn on a fabric mesh, usually silk or nylon, and the nonimage areas are nonporous. A squeegee pulled across the screen forces ink through the image areas and onto the printing paper directly below. Also known as silkscreen or screen printing. Adaptation of the basic stencil-making technique.

Relief printing uses a raised surface to represent the image. Carved reliefs capable of making an impress were the precursors to relief printing. They predate actual printing by three thousand years. There were wooden stamps in Egypt, brick seals in Babylonia, and clay seals in Rome. The first use of carved reliefs to print images onto paper originated in China, where hand cut wood blocks were used for printing as far back as the Tang dynasty (618 - 906 A.D.). Woodcut and typographic printing are now the most common form of relief printing.

The intaglio process is distinguished from relief printing in that the image areas are depressed below the surface of the plate. This distinction is reflected directly in the word intaglio, which comes from the Italian and means to engrave or cut into. Intaglio techniques include engraving, etching, drypoint, mezzotint, and aquatint. These techniques differ only in the manner in which they incise lines into the plate. Engraving, drypoint, and mezzotint use sharp tools, while etching and aquatint use acid solutions. Several intaglio techniques are outlined below.

Engraving: Crisp lines are incised into a metal plate with the use of a burin. Heavier and wider lines are produced by pushing the burin deeper into the metal. Tonalities are achieved by engraving parallel lines close together (hatching), by making parallel lines that intersect at various angles (cross-hatching), or by many closely spaced fine dots (stippling).

Etching: The metal plate is first covered with a layer of wax or acid-resistant ground into which the artist scratches a design with a stylus or needle, revealing the bare metal below. The plate is then dipped in acid, which etches the exposed metal, leaving the impression permanently on the plate. This corrosive property is reflected in the root of the word etching, which is derived from the German *aetzung*, which means to eat away or corrode. Etched lines are usually sharply defined, having uniform thickness.

Drypoint: Similar to etching, but the hard point of a needle is scratched directly onto the metal plate to produce soft, thick lines. No acids are used.

Mezzotint: A drypoint process in which a metal plate is textured with many fine dots so as to hold a great deal of ink and print a solid black field. The texture is produced with a rocker, essentially a large curved blade with many fine teeth. After the blade is rocked back and forth, the stippled texture is scraped away where lighter tones are needed. Areas that are vigorously scraped hold less ink and are printed whiter. Unlike other intaglio techniques, the image is developed from dark to light.

Aquatint: An etching technique that creates areas of tone through the use of powdered resin that is sprinkled on the metal plate before applying the etching acid. The result is a finely textured tonal area whose darkness is determined by how long the plate is bitten by the acid.

The earliest engravings can be traced back some 17,000 years in the Lascaux Caverns in the Dordogne region of France. Widely regarded as the most outstanding of known prehistoric art, the paintings and more than 1000 engravings found on the walls and ceilings of these caves indicate that engraving was as widely practiced and respected among primitive man as the technique of painting. Sharp flint implements and stone scrapers were used to carve images into cave walls. Later, engraved bronze vessels and mirrors made by Etruscan artisans became highly regarded and prized objects in ancient Greece and Rome.

Although the underlying principles of intaglio were known to goldsmiths in the Middle Ages, it was not until the fifteenth century, when paper became more generally available, that intaglio printing emerged as a specific art medium. Engraving and etching were two

intaglio techniques that flourished under the early masters. Durer was a master engraver and pioneered the use of variable thickness lines that elegantly swell and taper. A classic example of his work is shown in Fig. 2.



Fig. 2. A classic durer engraving



Fig. 3. Etching of two peasants by Leibl (George Wolberg)

Etching as a graphic arts technique was perfected in the early seventeenth century. Unlike engraving where lines may swell and taper, etched lines have uniform width. Dark regions are created by modulating the spatial frequency of the lines. This effect is demonstrated in Wilhelm Leibl's etchings of two peasants in Figure 3. (Wasserman, Y., 1995)

3. Proposed algorithm

We propose a new technique which is user driven, providing a versatile tool to the digital artist. Using our technique we are able to simulate etching by modulating. Our digital intaglio etching system using images or video frames has five step processes.

Step 1) Acquisition video streams from video camera or digital image.

First we acquire video streams from video camera.

Step 2) The affine transformation in color space.

In step2 [R G B] are changed to [YUV] and we extract the luminance component. Step2 is to extract the luminance component from the color image size. If the color image format is in RGB, then we need to convert it to YUV color space to get the luminance. The new value $Y=0.299*R+0.587*G+0.114*B$ is called the luminance. One of the main advantages of this format is that gray-scale information is separated form color data, so the same signal can be used for both color and black and white sets. It is the value used by the monochrome monitors to represent an RGB color. Physiologically, it represents the intensity of an RGB color perceived by the eye.

Step 3) Decomposed luminance through DWT in one level.

The Haar wavelet transform (Albert Bogges Francis J. Narcowich., 2001) is identical to a hierarchical sub band system, where the sub bands are logarithmically spaced in frequency. An image frame is first decomposed into two parts of vertical etchings and horizontal etchings by critically sub-sampling horizontal and vertical channels by following equation:
Basic one Step function

$$\psi_n(t) = \begin{cases} 1 & 0 < t < 1/2 \\ -1 & 1/2 \leq t < 1 \\ 0 & \text{Otherwise} \end{cases}$$

A video frame and its DWT decomposition are shown in Fig. 4.

- Scaled and translated: scaling the transformed image to double of its original size

$$\psi_{jk}(t) = \psi_H(2^j t - k)$$

Step 4) Modulating line thickness to produce varied tones

The goal of this work is to emulate the distinctive style of intaglio printing for artistic effect. We define digital etching to be the effect of modulating line density to produce varied tones. According to the following equation, it is very simple, fast, easy, understandable and very effective and expressive. The flow of engraved lines is determined by following:

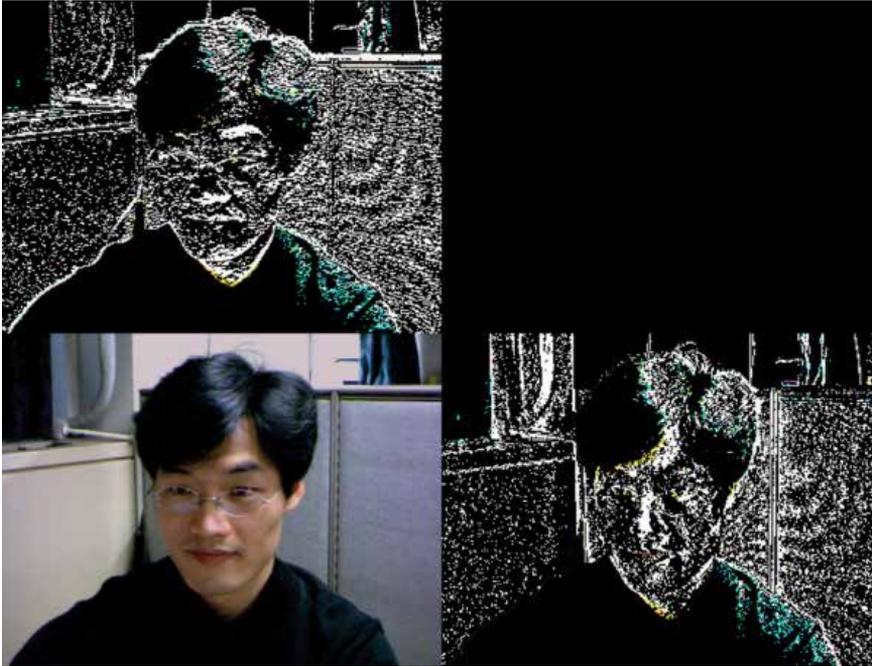


Fig. 4. Transformed coefficients of DWT magnitude

```

for i=1 : 3 : display_width
  for j=1:11: display_height
    for k=0:1:9
      if (j,i) >= 128 then ImageX(j+k,i)= ImageX(j+k,i) + 100 ;
      else Image X(j+k,i)= 0
        end
      end
    end
  end
end
end

```

Depending on the image pixel values along those lines (and between them), the output is converted into black and white pixels. An example is shown in Fig. 6 and 7.

Step 5) Displaying Video Digital Intaglio on real-time

Other experimental results of digital intaglio are given below. Our process is not only tedious and demanding, but also do not require great artistic skill and time. We performed some numerical experiments from web camera (15 frames per second).



Fig. 5. Original source image

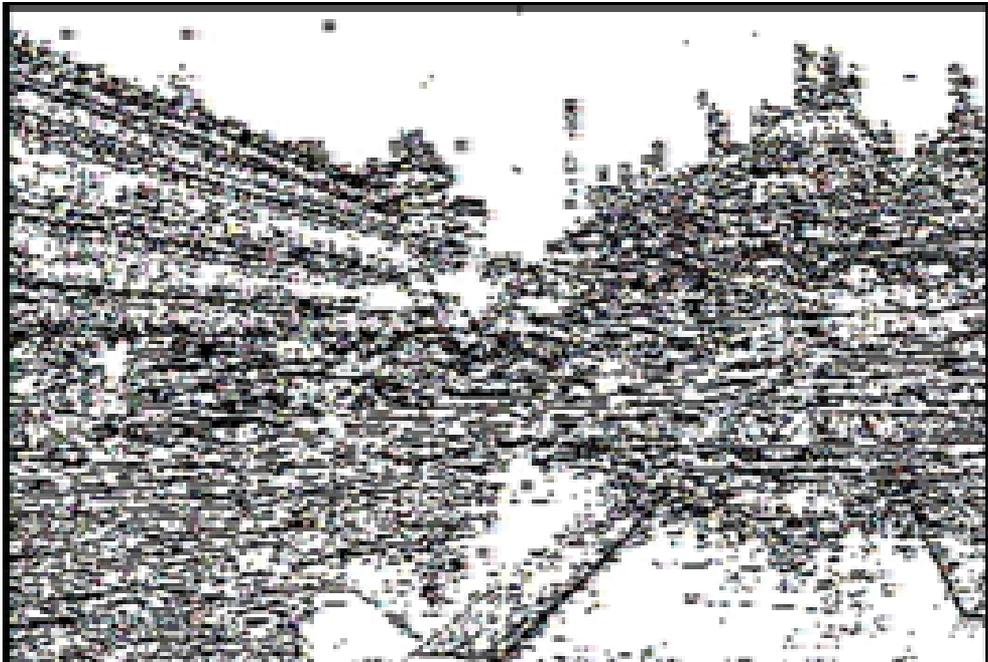


Fig. 6. Result image in horizontal

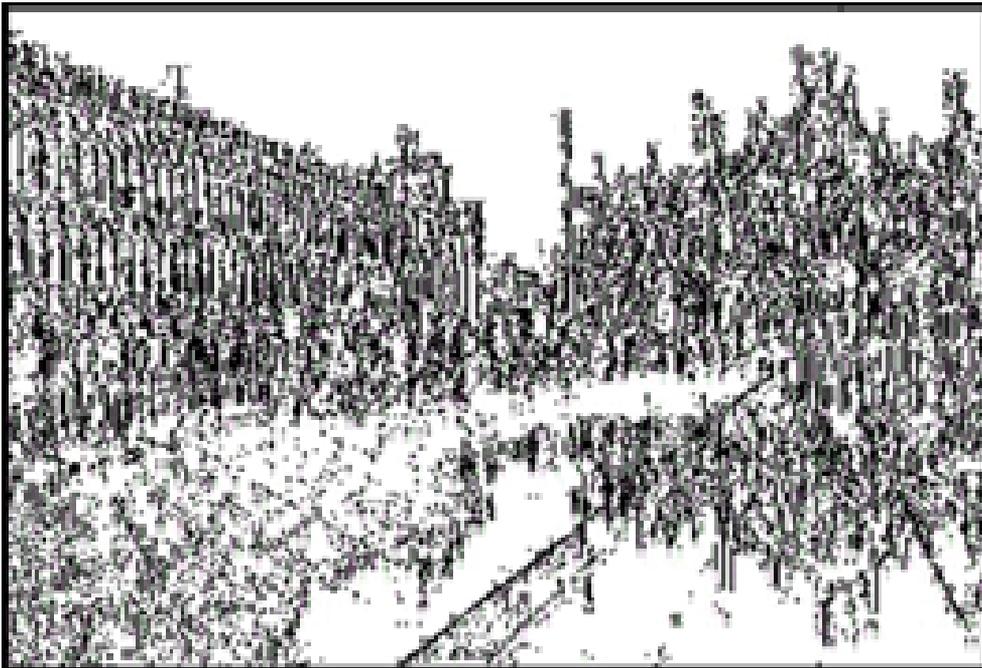


Fig. 7. Result image in vertical

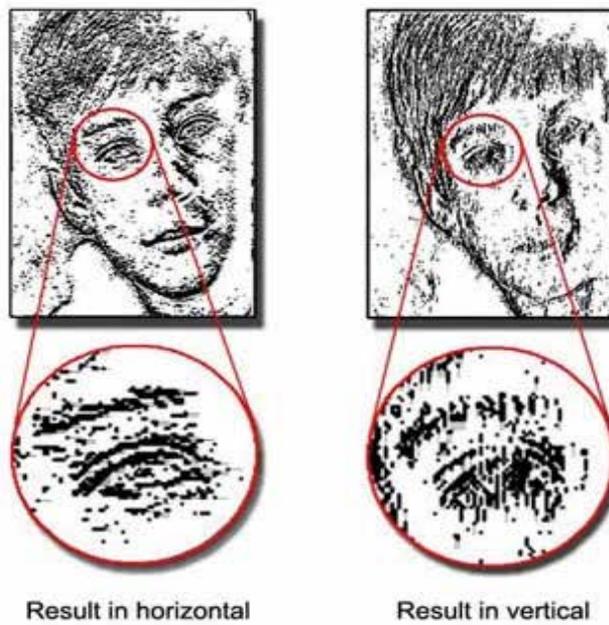


Fig. 8. Result images in both horizontal and vertical

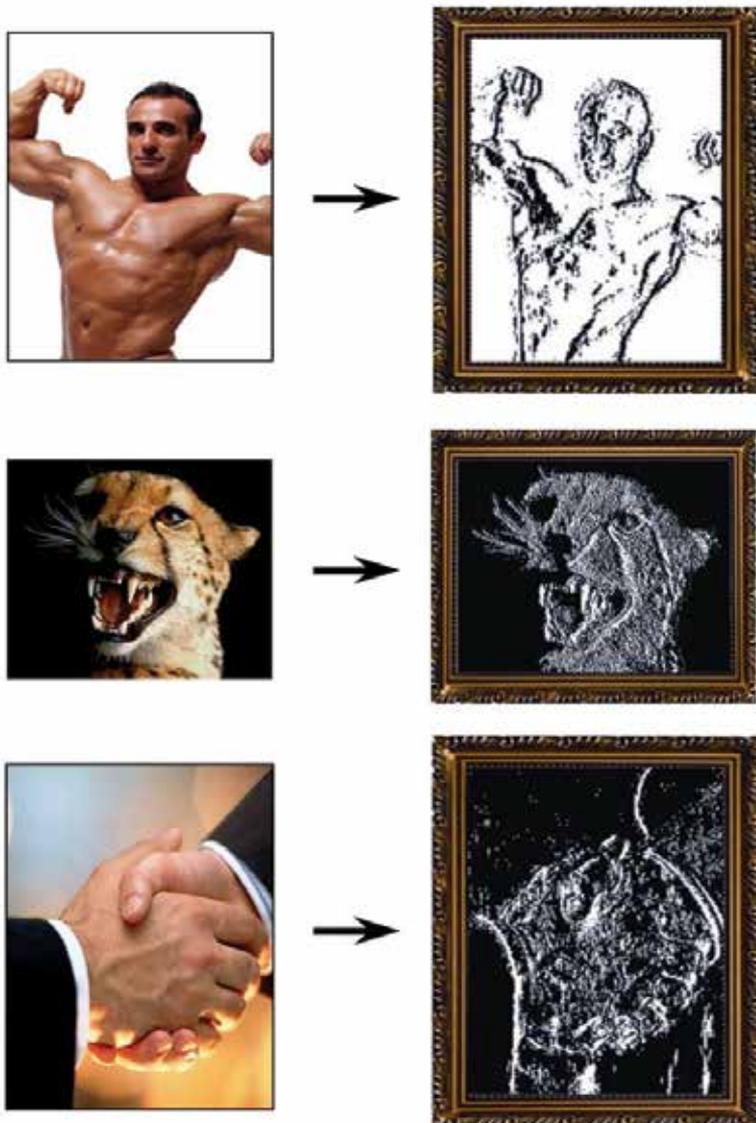


Fig. 9. Result images

5. Conclusion

In this paper, we proposed in transforming digital video streams with the etching effects of intaglio printing, which include the familiar styles of engravings that conform to the characteristic styles we seek to emulate. It displays instantly two types of etchings video art: vertical etchings and horizontal etchings.

Experimental results show that proposed algorithm close to physical etching. The resulting recessed lines are filled with black and the video frame is transferred to monitor by means of

the algorithm on real time. The newest features are preview video window while real time, saved drawing time, displayed two types of etchings video art which are vertical etchings and horizontal etchings at once.

We plan to further develop the software for use in the visual effects industry. The potential markets include advertising, television, and film production. There are currently no commercially available systems that robustly offer such tools. A successful implementation of the proposed work has great potential for introducing a high impact visual effect to the marketplace. Such tools can be offered as Adobe Photoshop plug-ins or as a stand-alone product. We also have to plan to augment the graphical user interface in response to additional testing among artists.

6. References

- Albert Bogges Francis J. Narcowich. (2001). *A First Course in Wavelets with Fourier Analysis*, Prentice Hall, ISBN 0-13-022809-5, Upper Saddle River.
- DigiDuffer. (1993). a digital engraving system. *The Visual Computer: International Journal of Computer Graphics archive*, Vol. 10 , Issue 5 (1993) , pp. 277 - 292, ISSN:0178-2789
- Goeorge Wolberg, Digital Engraving Techniques For Artistic Rendering, City College of New York.
- Line Art Rendering via a Coverage of Isoparametric Curves, *IEEE Transactions on Visualization and Computer Graphics archive*, Vol. 1 , Issue 3 (September 1995) table of contents, pp. 231 - 239, ISSN:1077-2626
- Wasserman, Y. (1995). Integrated Single-Wafer RP Solutions for 0.25-micron Technologies. *IEEE Trans-CPMT-A*, Vol. 17, No. 3, pp. 346-351.

Building Principles and Application of Multifunctional Video Information System

Mahmudov Ergash B. and Fedosov Andrew A.
Tashkent University of Information Technologies
Uzbekistan

1. Introduction

The beginning of 3 millennium marks evolutionary development of telecommunication and communication branch, including digital TV and broadcasting. It was promoted by introduction of new digital integrated microcircuits, computer engineering, use of new transformation methods, processing, compression and archiving, and also digital methods of great volumes signals modulation.

For today all over the world create a global information infrastructure, a special place in which the multipurpose interactive video information system occupies. And it is not casual - after all over 80 % of the information the person receives through organ of vision.

Basing on the theoretical and experimental research results for last 25 years, structural principles of perspective television system, actually video informational system (VIS) for multipurpose application or multifunctional video informational system (MFVIS) are stated in this chapter. The possibility and expediency of interactive and integrated broadcasting TV system designing necessary for society provision with information on its base is also considered.

2.

Digital TV-broadcasting strengthens an interactivity role. The concept of interactivity accepted by the international organizations, provides complex use of landline, satellite and cable means, harmonious connecting with other telecommunication services, with mobilization of funds for maintenance of reverse channels. The model of digital TV-broadcasting system is presented on Fig. 1.1.

In this system the subscriber equipment forms a local domestic complex basically with cable and other wire connections. Such connections limit opportunities for the equipment location change, its use in an interactive mode. In this connection the complex is supplemented with receive/transmit radio systems of a subscriber providing bidirectional "wireless" access of subscriber devices to direct and return interactive channel. For increase of interest to transition to a digital TV broadcasting a task for transformation of the functions of STB subscriber terminals (multi-purpose interactive devices capable to service broadcasting and many info communication services) limited originally has been applied. Thus, due to multifunctional achieved by software, a cost of such devices practically does not grow. The basic hardware-software means of a receive/transmit path of video

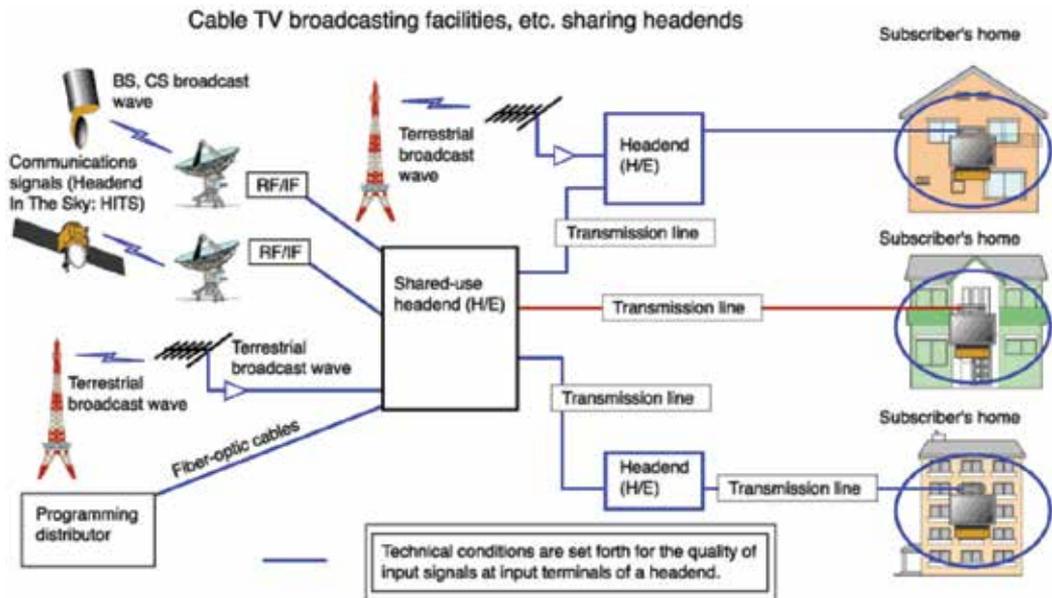


Fig. 1. 1. Global model of digital broadcasting system

information systems offered by author of MFVIS with a universal purpose is a complex of coding and decoding device with adaptively controlled parameters allowing signals compression and decompression of multimedia and packing and unpacking of packets. At the first stage of introduction it is expedient to use analog and digital methods in a combination to technical opportunities and equipment with new digital means of transmission, reception and transmitting environment.

In this case reception on domestic TVs can directly be carried out through a terrestrial network of television broadcasting by means of application of universal multipurpose prefixes for reception both analog and digital TV programs on usual TV made after 1990. Taking into account that a TV image represents huge data file so a speed of a stream about 270 Mbit/s that considerably surpasses opportunities of modern transmission channels is required for their transmission on a TV communication channel with quality of the broadcast standard. Therefore, various methods of coding image and a sound with the purpose of compression of initial volumes of video-audio information due to elimination of their redundancy apply for coordination of parameters of TV program signals and transmission channels.

The main advantages which digital technologies give us are an excellent image and sound, and also increase in number of transmitted channels with expansion of service. For example on a satellite transponder, intended for the normal analog channel, about 8-10 channels of digital television together with a binaural sound are possible to transfer. In result the choice of channels has already visibly grown, and from the satellite now channels can broadcast that could not earlier, at that rent of "place" in 8 times became cheaper. Also, new conveniences which were inaccessible earlier have appeared.

However, needs for transmission of video information still essentially exceed the opportunities, represented for it by TV communication networks, both on availability of communication, and on volume of information which can be transferred through this network.

In this connection now, the problems of development of information space necessary for achievement of economic and political independence are urgent for our republic.

These circumstances urgently demand an increase in effectiveness of operating TV networks, introduction of new broadband digital means and video information systems of communication and also integration of existing local TV networks with the purpose of a single information system creation in the long term in the republic, ensuring an output in to international and other regional communication networks.

The conducted analysis of increases in use efficiency of a throughput capacity of video informational systems (VIS) will allow to reduce essentially a TV signal redundancy that creates a potentiality for increasing use efficiency of a throughput capacity of VIS channel due to transmission on this channel a rather great volume of additional information.

It is known, that resulting transmission speed of a digital stream of TV program signals at component encoding of components of color TV signals with use of normal PCM can be presented as the expression:

$$V_{P0} = V_{\alpha} + 2V_{\beta} + V_{\gamma} \quad (1)$$

Where:

$V_{\alpha} = n_{\alpha} f_{\alpha}$ is transmission speed of brightness signal - U_{γ} ;

$V_{\beta} = n_{\beta} f_{\beta}$ - transmission speed of color-difference signal U_{b-y} ;

$V_{\gamma} = n_{\gamma} f_{\gamma}$ - transmission speed of audio signal;

$n_{\alpha}, n_{\beta}, n_{\gamma}, f_{\alpha}, f_{\beta}, f_{\gamma}$ - length of codes and a sampling frequency of signals U_{γ} and U_{b-y} (U_{b-y}) accordingly. In non-adaptive digital VIS length of codes, a sampling frequency or resulting speed do not change, i.e. $V_{P0} = \text{const}$.

At organizing of multiprogramming transmission of TV signals on VIS with compression and decompression of its components, resulting speed is defined by the expression:

$$V_{p2}^1 = \begin{cases} \Pi(2k-1)m^{-1}(f_{\alpha} + 2f_{\beta} + f_{\gamma}), n_{puk} = 1, n; \\ \Pi(2k \cdot m^{-1}) \cdot (f_{\alpha} + 2f_{\beta} + f_{\gamma}), n_{puk} = 2, 3; \end{cases} \quad (2)$$

Where:

Π - number of TV programs;

f_{α} - Sampling frequency of brightness signal

f_{β} - Sampling frequency of color-difference signals

f_{γ} - Sampling frequency of audio signal

One of directions of TV systems realization allowing to improve a level of information service is transmission of additional data, convergence, texts, including on TV viewers requests. VIS model contains a logic and physical structure which has both a direct and return communication channel.

On the basis of logic and physical structures, and also structures of an information data packet it is possible to make a generalized model of VIS, intended for providing various additional services except basic video information, and also giving an opportunity for computer processing and archiving images with compression on subscribers' requests. (Fig.1.2.)

Now the level of communication equipment of users has considerably increased, and the majority of users are expressing now desires to communicate with an external world. In this

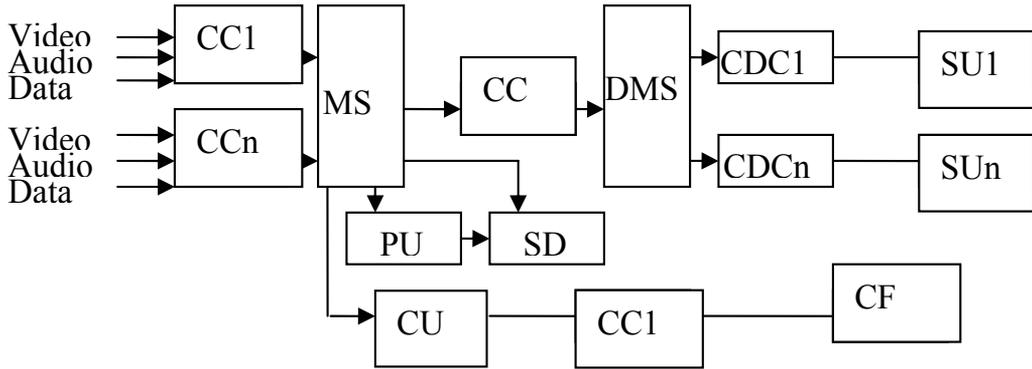


Fig. 1. 2. The generalized circuit model with compressions of a TV programs signal and additional information.

- Where CC1, CCn- complex coder
- MS- multiplexer
- PU-processing unit
- CU- control unit
- CF-call facility
- CDC1, CDCn -complex decoder
- SU- subscriber's unit

connection many foreign firms have started developing technologies simplifying such an exchange with providing to users an opportunity for reception of a single integrated access to information by means of processing universal system [1].

The basic directions of the research are connected with the search of optimum methods of orthogonal (mathematical) transformation on the base of Wavelet and Fourier analysis.

Efficiency of MFVIS in many respects is defined by a method of signals transformation with the purpose of maintenance of high compression ratio at a preservation of the set requirement to the quality of a restored image.

Most frequently the Fourier analysis which represents an original signal as a set of harmonic signals is used for these purposes. It allows to transfer an initial electric signal A (x, y) from time area into spectral area F (w_x, w_y).

In Fig.1.3 the opportunity for the adaptive coder designing based on a combination of methods of adaptive linear prediction (ALP), correlation prediction (CP) is submitted, and is discrete cosine transformation (DCT)

Optimization of VIS parameters with an adaptive linear prediction was made by computer simulation with application of the mathematical device of a difference equation. The final difference equation connecting input signal U_n (k,l) with a target restored signal U_n B(k,l) looks like:

$$U_n \epsilon(k, l) - \sum_{k=1}^n \sum_{l=1}^m a_{k,l} U_b(k-1, l-1) \cdot K_{oc} + \sum_{k=1}^n \sum_{l=1}^m a_{k,l} U_b(k-1, l-1) \frac{K_{oc}}{m_0} = \frac{1}{m_0} U_n(k, l) \quad (3)$$

a_{k,l} weight ratio of predictors:

N, M- number of elements participating in a prediction:

Where d is a size of DCT fragment, U_{k,l}- values of luminance, k,l - points of a fragment, U_{k,l} - sum of luminance values of adjoining points.

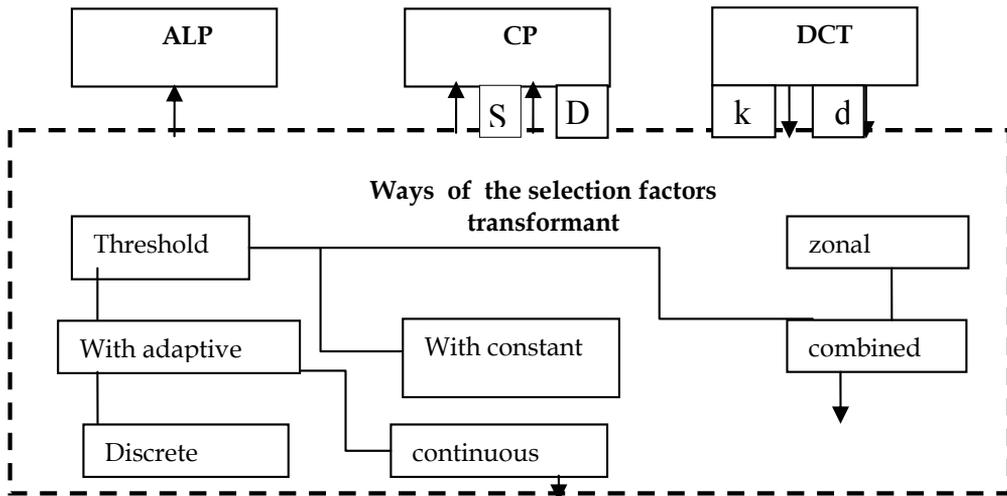


Fig. 1. 3. Structure of construction the compression signals device on the basis of images complex coding method.

The resulting compression factor at simultaneous use of all three coding methods can be written down for adaptive (K_{pca}) and non-adaptive encoders (K_{pc}).

$$K_{pc} = V_1 / V_2 = K_1 \cdot K_2 \cdot K_3, \tag{4}$$

$$K_{pca} = K_1 \cdot K_{2a} \cdot K_3, \tag{5}$$

Where:

$$V_1 = f_{\text{cyan}} \cdot n_{\text{yan}} + 2f_{\text{cyan}} \cdot n_{\text{yn}}, \tag{6}$$

$f_{\text{cyan}}, f_{\text{cyan}}; n_{\text{yan}}, n_{\text{yn}}$ - nominal sampling frequencies and codes length of luminance signals and color-difference signals.

$$K_1 = \text{Log}_2 L_n / \text{Log}_2 L_k, \tag{7}$$

Where:

$L_n=2^{n_n}$ -quantizing level of an original signal ($n_n=8$).

$L_k=2^{n_k}$ -quantizing level of the transformed signal ($n_k=5$)

$$K_2 = \frac{N^2 \cdot \log_2 L_m}{\sum_{U=0}^{N-1} \sum_{V=0}^{N-1} n_{U,V}}; \tag{8}$$

L_m - maximal value of luminance.

$$n_{U,V} = \frac{\sqrt{F^2(u,v)}}{F_n} \text{ -number of elements participating in a prediction}$$

F_n - threshold of restriction.

$$K_3 = 2^l; \quad (9)$$

l -number of recurrences.

$$K_{2a} = \frac{N^2 d \cdot}{\sum_{U=0}^{N-1} \sum_{V=0}^{N-1} \frac{\sqrt{F^2(u,v)}}{dy - \arctg((Q_c - dx)/e)}}; \quad (10)$$

Where:

$$F(u,v) = \sum_{k=1}^N \sum_{l=1}^N L_{\text{BX}}(k,l) \cdot W(j,g, u, v); \quad (11)$$

$W(j,g, u, v)$ - nucleus of direct transformation

j,g - number of a line and column in the image's block

u,v - number of a line and column in the spectral factors block

$L_{\text{BX}}(k,l)$ - luminance samples of an original signal

dx, dy - center symmetry coordinates of \arctg function

e,d - factors influencing a compression degree

$Q_c = \frac{1}{(d-2)(dS-2)} \sum_{i=1}^{d=2} \sum_{j=1}^{d=2} (8U_{i,j} - U_{i,j}^*),$ - complexity of a fragment

$$K_{23} = \frac{N^2}{\frac{N^2}{S \cdot K_{2a}} + \frac{N}{S} (S-1) 16h_0}; \quad (12)$$

S - lines number in transformation block.

h_0 - the normalizing factor, which taking into account a volume of the additional information

The results of a computer simulation have shown that distortions of a restored signal reduces due to block artifacts at VIS variant combining ALP and DCT or CP and DCT methods, alongside with increase of compression ratio by 10-15 %.

3.

The possibility of broadcasting television signal real-time transmission over cellular networks has been studied. The results of experimental investigation of discrete cosine and wavelet based video and audio compression techniques efficiencies have been compared. It was shown that wavelet based compression algorithms allow achieving the necessary compression ratios while conserving sufficient video and audio quality for bandwidth limited cellular networks transmission.

It is known that efficiency of TV program preparing for further broadcasting is directly related with a time required to transfer the filmed material to television center editorial office. However, sometimes lack of a wide bandwidth communication link does not allow performing the data transfer at once. That's why on the news mostly only reporter's voice is broadcasted with a static background instead of a full video. This problem can be solved by using cellular network communication links for effective TV-signal transmission.

Besides cell-phone main functionality, modern cellular communication facilities allow data transfer that permits to transfer multimedia information and access the Internet. Nokia-N92™ cell-phone announced by Nokia® is a first apparatus of the N-family of Nokia with embedded function of DVB-H television programs receiving and recording. TV program view is performed with 2.8 inches (71mm) screen which has resolution of 240x320 pixels of 16 millions colors. This is not enough for quality television image which requires 720x576 pixels resolution.

Cellular communication systems of generations 2.5 and 3 provide data transmission with up to 2 Mbps bandwidth that allows digital TV signal transferring. However, the TV digital data transmission rate should be limited by the value of 2 Mbps while conserving sufficient video and audio quality. Video transmission rate can be significantly reduced by using an appropriate compression scheme. For this purpose the MPEG video coding standard [2] defines three types of frames to be transmitted: intra frames (I-frames), predicted frames (P-frames) and bi-directional interpolated frames (B-frames). I-frames hold an exact input image structure. These are reference frames for decoding starting. P-frames hold the difference between current frame and the previous one (I- or P- frame). Finally, B-frames store only the most significant changes between previous and next frames.

In practice I-frames compression ratio is not very high (up to 30% of initial data size) as far as I-frames are compressed just in order to eliminate the image spatial redundancy. P-frames usually take one third of I-frames data size, while B-frames take only one eighth. So, to achieve maximal video data compression rates I-frames should not be used frequently. For decoder stability it is desirable to transmit every 10th frame as I-frame, that is also required when a subject image changes. The frames order of MPEG-2 data stream is represented in the Fig. 1.4.

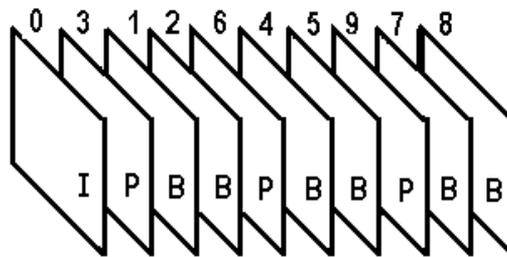


Fig. 1. 4. MPEG-2 frames order

To eliminate image spatial redundancy the most common approach is to use discrete cosine transformation (DCT) [3] proposed by V. Chen in 1981. DCT is used in MPEG-1 and MPEG-2 standards. In general DCT is just a two-dimensional discrete Fourier transformation, but uses only cosines as basic functions. The advantage of DCT usage is its' fast series convergence for most of images. This provides high compression ratios for minimal image reconstruction error.

Image encoding method using DCT is based on frame decomposition by square blocks of 64 pixels (8 by 8) each referred as signal matrices. The idea of this method is to transform the initial image signal matrix to frequency values matrix of the same dimension. The frequency values matrix is not related directly with pixels position of the initial image, but presents a convenient mathematical form for DCT coefficients. In this way frequency matrix can be considered as two-dimensional spectral representation of image fragment. In this matrix the

upper-left corner stands for low-frequency DCT values while the bottom-right is high-frequency ones.

For most of images the significant part of energy is concentrated in low-frequency region. This is very important DCT feature in the sense of data compression. High-frequency coefficients usually take zero or near-zero values and can be discarded, so DCT coefficients with significant values above defined threshold is transmitted. The values lying under the defined threshold are concerned as zero values. Subjectively loss-less video compression is achieved if only zero values of frequency matrices are discarded while data transmission. In such case the reconstructed image will not be distinguished from the original one by a viewer (Fig. 1.11), however the compression ration will not exceed value of 10-20 depending of input image details.

Increase of DCT values threshold leads to increase of image compression ratio as well as loss of reconstructed image quality. So, the rational choose of the threshold value allows balancing between the compression ratio and image quality. Fig. 1.12 illustrates the result of image compression with the ratio value of 45; small details of the images are lost while overall reconstructed image quality is quite acceptable (areas 1 and 2 of the Fig. 1.12).

To achieve a higher compression ratio DCT values quantization can be used by dividing each element of the DCT frequency matrix by quantization values taken from quantization matrix. Each color component (Y, U and V) has its own quantization matrix $q[u,v]$:

$$Y_q[u, v] = \left[\frac{Y[u, v]}{q[u, v]} \right]$$

Other specific features of this compression scheme are also related with quantization. High compression ratios can lead to losses even in low-frequency region, in this case the result image breaks apart to 8 by 8 pixels blocks. As it's shown in Fig. 1.13 quality of image compressed with ratio of 75 become unacceptable (details in region 1 and 2 can't be distinguished).

Thus, considering experimental results of different images compression it is ascertained that TV image compression ratio can not exceed the value of 20-30 while conserving sufficient quality. Theoretically pure video data rate for TV digital signal of resolution of 720x576 can reach the value of 2.5-2.3 Mbps. Indeed, the real video stream including audio and supplementary data transmission rate is much higher, that does not allow using DCT based compression methods for TV signal transmission over narrowband cellular network links.

Besides the DCT based compression techniques, there are alternative video and audio compression algorithms based on wavelet transformation introduced by Grossman and Morlet [3] in the middle eighties.

The main drawback of Fourier transformation (and particularly that of DCT) is an attempt to approximate real non-periodic signals with a help of periodical basic functions. On the contrary wavelet-transformation is based on signal representation by space-limited basic functions of finite energy (wavelets) but in general sense resembles Fourier transformation [4]. However there are a lot of differences between Fourier and wavelet transformations. First of all, wavelets are limited (non-periodic) functions, secondly to obtain a complete transformation basic set maternal wavelet stretching is used, finally basic functions liner composition is replaced by a composition of wavelets temporally shifted and scaled. In general wavelet transformation can be concerned as an act of a certain basic filter, whose impulse response is defined by maternal wavelet. So, a complete wavelet decomposition of

an input signal can be represented as a set of filtering processed followed by operation of decimation (Fig. 1.5).

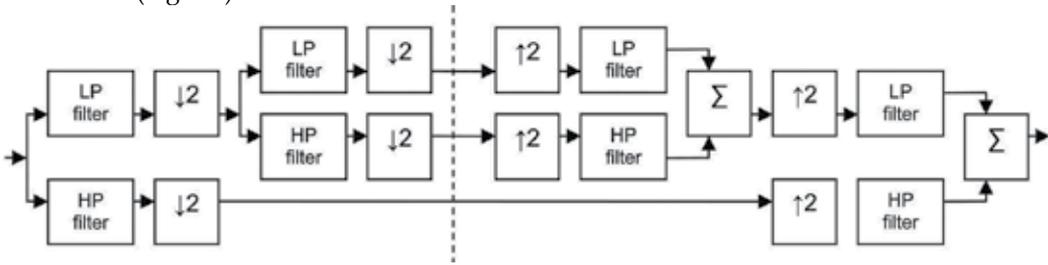


Fig. 1. 5. Generalized block diagram of wavelet decomposition.

Input signal is passed through low-pass (LP) and high-pass (HP) filters, which divide signal spectral band into two parts. That’s why low-frequency and high-frequency input signal spectral bands become two times narrower after the filters passed.

To perform complete wavelet decomposition the input data length should be a power of two: 2^n , where n is a positive integer value. This is explained by the transformation method which divides input signal on two equal parts (low-frequency and high-frequency) at every decomposition step. Another element introduced in wavelet filters is decimation. The decimation operation passes on output every second value taken from input, as far as every second value holds redundant information.

Since an image is a two-dimensional signal, each step of wavelet decomposition is implemented in two stages: at first image rows are treated, then image columns (or vice versa).

Let’s take an image row with a length of N . This row can be represented as an array $S_{[N]}$. Having this the simplest Haar wavelet transformation can be accomplished by meaning of two neighbor pixels colors (low-pass filter) and numerical differentiation (high-pass filter). This gives two array on output $A_{[N/2]}$ and $D_{[N/2]}$ defined as followed:

$$A_k = \frac{S_{2k} + S_{2k+1}}{2}, D_k = \frac{S_{2k} - S_{2k+1}}{2}, \text{ where } k \in [0, N/2).$$

$A_{[N/2]}$ values are referred as signal approximation, $D_{[N/2]}$ values are referred as signal detailing. Obviously, having A and D arrays one can easily reconstruct the initial signal as shown in figure 3, where the numerical values of A and D are represented by pixels brightness.

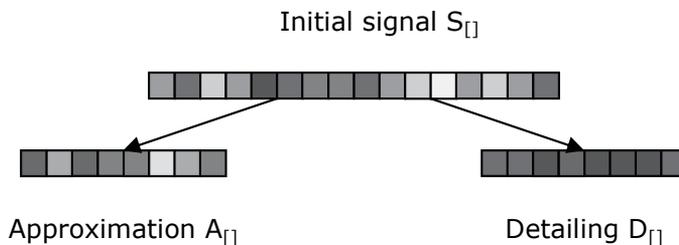


Fig. 1. 6. Single step of wavelet decomposition

Complete image decomposition can be implemented in several steps (Fig. 1.6). First of all image rows decomposition is implemented (Fig. 1.8). After this step passed, the image approximation appears on the left side, while the image detailing - on the right one. A similar treatment is for the image columns (Fig. 1.9). After these two steps, the image approximation appears in the upper-left corner, upper-right corner stands for differential image in the line of X, bottom-left corner is for differential image in the line of Y, and finally, bottom-right corner consists of differential image in the line of X and Y. One can notice that image detailing coefficients take zero or near-zero values (zero values are represented by grey color with a value of 128).



Fig. 1. 7. Initial image

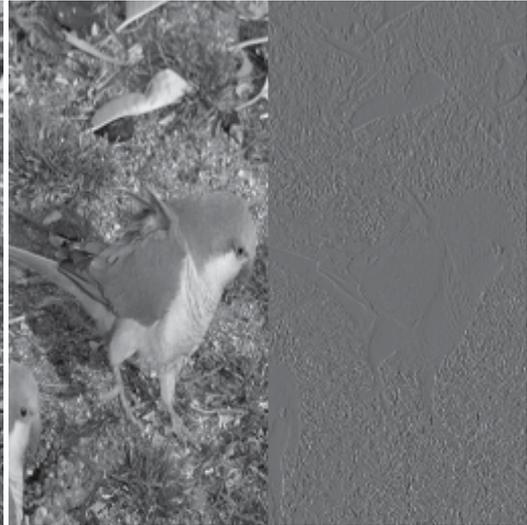


Fig. 1. 8. Rows wavelet decomposition



Fig. 1. 9. Columns wavelet decomposition

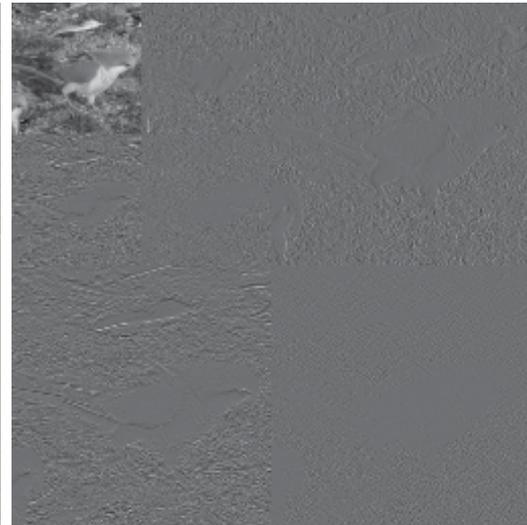


Fig. 1. 10. Two steps decomposition

The further decomposition is performed on the image approximation, which is the initial image two-times reduced copy. The result of the second decomposition step is represented in Fig. 1.10. The decomposition process can be carried out up to the single pixel in the upper-left corner of the image. Obviously, the brightness of this pixel will represent the mean brightness of the whole initial image.



Fig. 1. 11. Initial image used for compression



Fig. 1. 12. Image compressed with MPEG-2 (compression ratio is 45)

Difference between wavelet based image compression and MPEG-2 compression is illustrated in Fig. 1.13 and 1.15. As one can notice, wavelet based compression provides better image quality for the same compression ratio (areas 1 and 2). So, the advantage of wavelets over DCT methods is wavelets' localization within the range of transformation.



Fig. 1. 13. Image compressed with MPEG-2 (compression ratio is 75)



Fig. 1. 14. Wavelet based image compression (compression ratio is 45)

This allows obtaining higher compression ratios for quality image reconstruction. However when the compression ratio is too high, wavelet based compression algorithms can produce ripple artifacts, especially for inhomogeneous regions of the image.

Along with image compression wavelets can be efficiently used for audio data compression. In this case coder deals with one-dimensional input data set (or several data sets for multi-

channel audio). Wavelets also introduce some advantages for audio coding and don't yield to MPEG in coding efficiency [5].



Fig. 1. 15. Wavelet based image compression (compression ratio is 75)

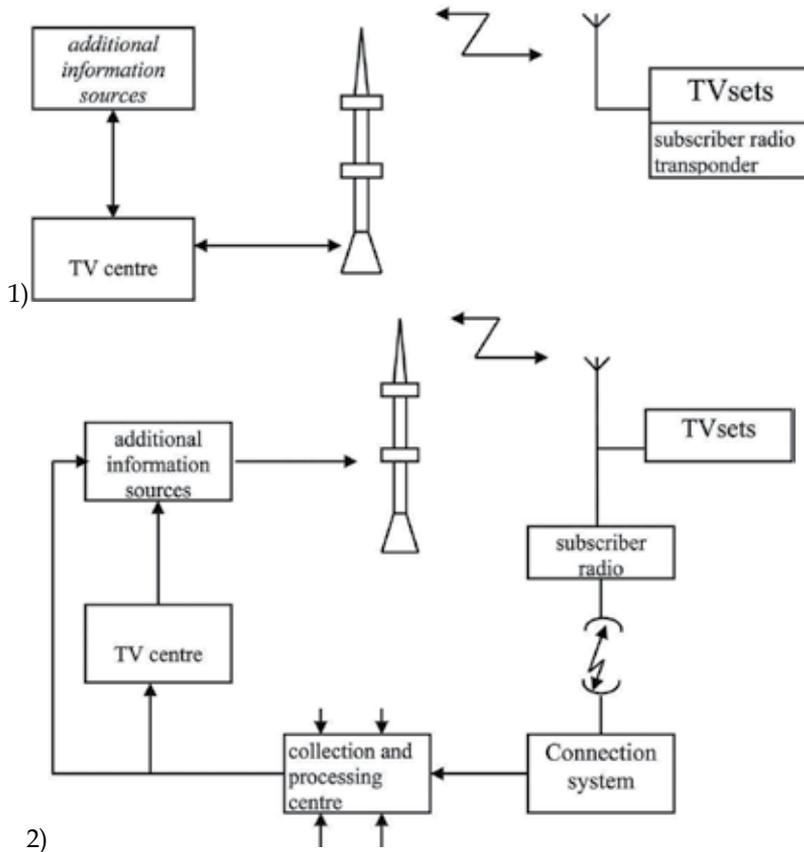


Fig. 1. 16. The organization of interactive digital TV broadcasting system

According to literature data, modern wavelet based coding methods are used effectively in MPEG-4, JPEG-2000 and DejaVu compressors. Analog Devices company produces ADV601

commercial chip which provides real-time video compression using wavelet transformation with up to 350 compression ratio for VHS quality video [6].

Thus, wavelet based compression progress are expected to be sufficient enough to keep video data rate at the level of 1.5-2 Mbps that will give a possibility to use cellular network links to transmit high quality TV directly from the place of event to any spot on the globe served by cellular operator.

4.

Multiprogramme stream formed by the broadcasting module and multimedia data formed in the module of multimedia after multiplexing according to global broadcasting TV model are transferred on the monochannel. Which exits are connected to multipurpose user's installations or branch off in other directions using existing environments of transfers for delivery of the corresponding information to remote users.

It is important to note combination of existing digital television and broadcasting systems with a view of coverage difficult access region and territories with small number of inhabitants. Thus construction of the reverse channel is necessary. The complex system of tele-radio broadcasting is based on standards DVB-T and DRM.

In TV systems and broadcasting with integration telecommunication services takes place, that separate groups or individual users wish to receive the information protected from other users, such VIS systems with the conditional (limited) access are called.

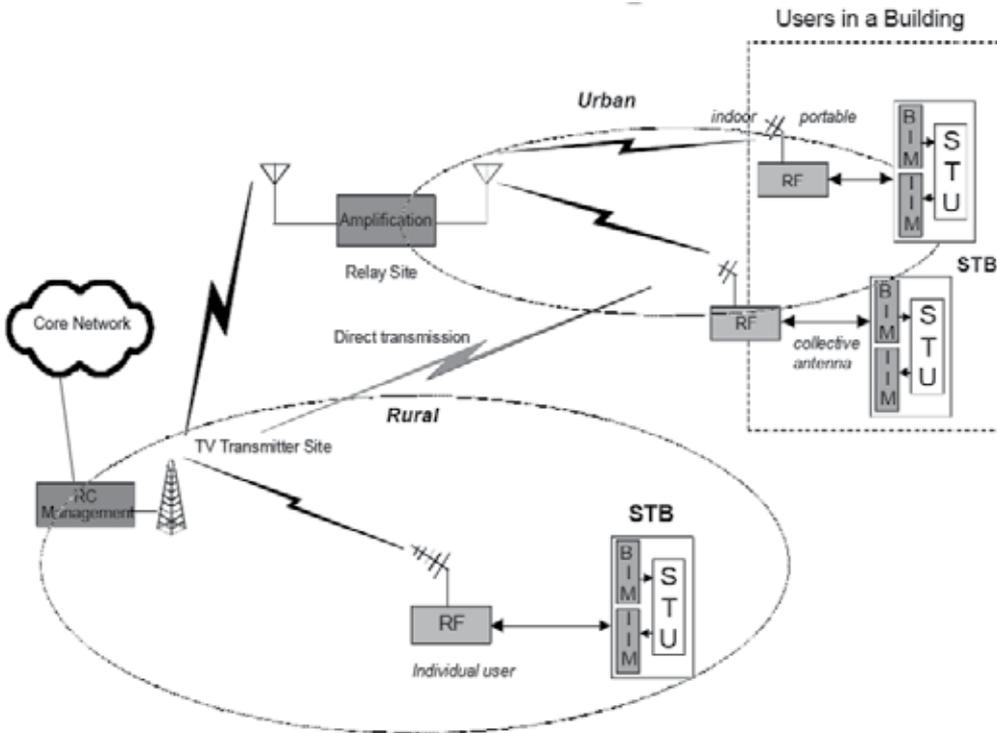


Fig. 1. 17. The block diagram of Interactive terrestrial digital TV system

In such systems various cryptography algorithms are used: symmetric and asymmetric. Some examples of popular and well-respected symmetric algorithms include Twofish, Serpent, AES (Rijndael), Blowfish, CAST5, RC4, TDES, and IDEA.

The most popular asymmetric algorithms is RSA. Generalisation of Cocks' scheme was independently invented in 1977 by Rivest, Shamir and Adleman, all then at MIT. The latter authors published their work in 1978, and the algorithm appropriately came to be known as RSA.

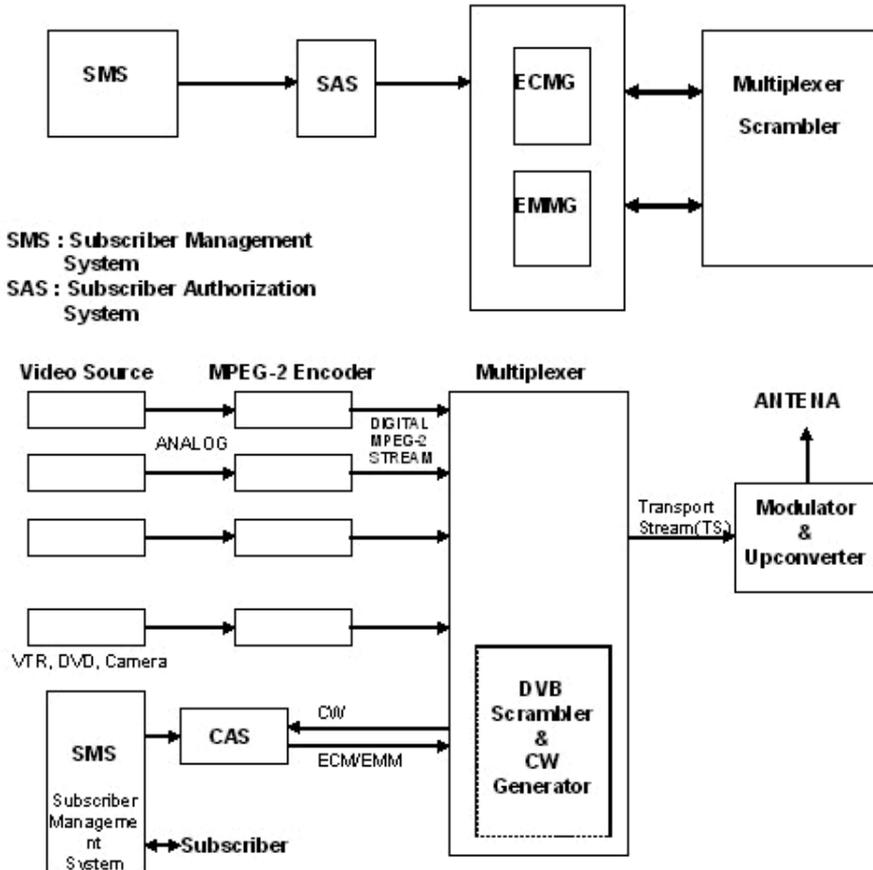


Fig. 1. 18. The block diagramme of conventional access system

EIS (Event Information Scheduler)(option)

EMMG (EMM Generator)

ECMG (ECM Generator)

5.

Nowadays for observation of artificial celestial bodies (ACB) broad application as received by photo-television methods, which allows to essentially increase penetrating ability of a telescope, to intensify a visible brightness of stars and object, to increase their contrast concerning a hum noise of the sky [7,8].

On the basis of use of photo-television methods the following principles are fixed:

- Photography of a sidereal field with located observable object (LOO) from the monitor of the video-control devise (VCD);
- Visual observation of a sidereal field and object on the monitor of VCD with simultaneous processing on PC, counting of brightness and coordinates immediately on scales of the control panel.

Advantage of the second principle is the high efficiency of obtaining the image with an evaluation of quality and possibilities of reaching the necessary accuracy of installation of brightness and coordinates of the images at the expense of their direct counting (without operations of photography, chemical processing of photos and writing them in the PC for the consequent digital processing).

The creation of a television telescope, i.e. system of optic-television equipment and telescope has opened more capabilities of automation of astronomical observations with use of the PC. It allows not only fast and precisely to determine brightness and coordinates of satellites in relation to rather reference stars, but also expands functionality of TV complexes. In the present time there is a plenty such TV complexes intended for observation ACB.[7-10]

Most important for modern TVIMS is the capability of adaptation of their parameters depending on change of an astroclimate or other factors both fulfillment of diverse functions and problems. In developed and researched TVMIS the multifunctionality is provided at their designing and technological fulfillment. In particular, are used in transmitting cameras CCD matrix actually being multifunctional devices, working on various algorithms depending on pilot signals, modern PC for processing of videosignal and control, and also multifunctional RAM of volume in some television frames (1-4 Mb). This RAM can execute buffering of the videoinformation, in case of low-frame-rate mode- slow record of chosen fragments of a series of frames and slow rewriting in the RAM of the PC.

The increase of MITVS speed is possible by actuation of the external RAM in address space of the PC. Besides, the multifunctional RAM provides visualization both received, and video information, processed in PC. In case of removal of a chamber part on rather large distance, the transmission of the video information and control commands is expedient for executing on FOLC. [9]

Similar MITVS executed on the basis of a television telescope and the PC can decide many problems faster and more effective than person. At observation of AO the received signals of the image are not so great and difference of signals from each other is possible to distinguish with the help of simple tags, i.e. the images of interesting us AO differ in the main size and appropriate amplitude of a signal. [10] Therefore, most perspective is the using for this purpose the television information measuring system (TVIMS), in which are combined hardware and software ensuring automation of process of observation of the objects [11].

The analysis of known, selection and implementation of the substantiation of a structure of construction IMTV has allowed developing structures MFDMITVS distinguished from existing;

- Capability of rigid binding to the beginning of selected frame and maintenance synchronization at the closed and open-ended version of its activity [11,12]
- High effective using of chips of dynamic memory of the RAM [11];
- Maintenance of a reasonable combination manual (through the operator) and program (Through PC) modes and parameters management [9];
- Maintenance of effective coding of the digital video information with the purpose to decrease the time for storage and processing on PC [11-13];

Capability of interface with PC of a type Pentium-133 and control of exchange by the videodata;

Capability of maintenance by a feed from one independent (autonomous) stabilized source; On Fig.1.19 the block diagram of researched MFDMITVS is shown which contains the following devices and units:

- Television camera on Charge Coupled Device matrix -1;
- Unit of processing of videosegment -2;
- Amplifier with an adjustable transmission factor -3;
- Unit of suppression television synchrony signals -4;
- Unit of binding to reference voltage -5;
- Source of reference voltage -6;
- Analog-to-digital converter -7;
- Unit of packing of figures of videosegment -8;
- Storage device for a record and storage of figures odd half-frame of videosegment -9;
- Storage device for a record and storage of figures even half-frame of videosegment -10;
- Unit of formation of current date and time codes -11;
- Unit of input and adjusting of current date and time -12;
- Shaper of the sign information -13;
- Unit of allocation lower case and personnel synchrony signals -14;
- Lock-in generator of television synchrony signals -15;
- Unit of unpacking of digital data -16;
- Basic digital-to-analog converter -17;
- Personal computer -18;
- Unit of transceivers -19;
- Operation decoder -20;
- Register of output data -21;
- Buffer register of input data -22;
- Register of management and condition -23;
- Unit of clocking and gating pulses -24;
- Counter of addresses -25;
- Multiplexer of the address -26;
- Unit of formation of pilot signals -27;
- Decoder of groups of state elements in 3Y 1 AND 3Y2-28;
- Unit of reading of figures and videosegment -29;
- Monitor -30;
- Videorecorder -31;
- Generator of clock television signals -32;
- Monitoring digital-to-analog converter -33;

MFDMITVS functions as follows. Videosegment from the television camera acts on 2, in which the preliminary analogue processing is made with the purpose of suppression of pulse interference and increase of visibility of the transmitted images. [14]

From an output 2 full TV signal through 3 and 4 acts in 5. The unit 3 is intended for obtaining an optimum level of videosegment ensuring an effective digital conversion. The amplified videosegment 3 simultaneously acts on an input 7 through the unit of binding the recovery of a constant making of videosegment and in the unit of allocation of personnel and

line signal clock pulses 14 is made. In 14 the allocation of a signal 64 harmonics of line signal clock pulses. Marked lower case and personnel clock pulses from an output 14 implements act in 15 and provide necessary phasing phased synchronix, and the signal of 64 harmonics acts in 24. In 24 with the help of generator with impulse phase selftuning of frequency (IFSTF) [12] the consequent multiplying of frequency 64 harmonics of a line signal clock pulses up to 24 MHz implements. The generator IFSTF is used by activity 15 in a driven mode. By activity 15 in the independent mode, which can be used at review of the images stored in memory 18, the signals of reference frequency are formed with the help of generator stabilized by a quartz resonator. The selected personnel pulse from an output of 14 acts in 15 and provides necessary phasing of formed signals of synchronix: the generator of synchronix is executed by the classical scheme. From an output of 15 the signal of synchronix acts 17 or 33, where the full videosal is formed.

The synchronization of activity of MFDMITVS units is provided with the unit of clock and gating pulses located in a module of the control unit. Unit 24 forms clock frequency of synchronization of videosal 12 MHz, and also group diverse on a phase clock pulses by frequency 3 MHz.

MFDMITVS can function in three operational modes: Automated mode; with control from the PC; Usual TV mode or low-frame-rateTV mode.

From an output of 7 the digital videosal through the unit of packing of figures of videosal 8 acts in units 9 and 10. The unit of packing provides a record in rather slowly functioning elements of memory of units of the RAM (9 or 10) figures of videosal acting with frequency 12 MHz. In the unit 9 the digital data of the first half-frame, and in the unit 10 digital data of the second half-frame are recorded. The sign generator is intended to form digital-letter information and provides data display in a window formed on a field of a plot frame. . The digital-letter data contain the information about current date and time, number of frame. The initial entry about date and time data is made with the help of unit of input and adjusting of current value 11. The unit of unpacking provides issue of digital data on 7 in rate of adopted sampling rate of videosal at rather slow reading of the data from the RAM (9 or 10). The principle of activity of the unit of unpacking, sign generator and unit of input does not require the special explanations. The address multiplexer of the 26 is intended for byte input of digital datas of videosal written in the RAM (9 or 10), in 18 with the purpose of digital processing. The decoder of groups of state elements 28 is intended for selection of group of elements of the RAM (9 or 10), by which it is necessary to address at a record or reading of videosal in 18.

The reading of digital data from units of the RAM (9 or 10) in 18 with the purpose of conversion and return filing of the treated digital information in the RAM (9 or 10) for a visual evaluation of the image on 30 implements through the unit of interface (interface unit), in which structure enter the unit of transceivers 19, operation decoder 20, register of output data 21, buffer register of input data, both register of control and condition 23.

The interface of digital conversion equipment with 18 implements the special interface unit (MOUSTACHE) 29 and for change of rate of input and format of the observable image the standard videoblaster of domestic production of a type SE-100 is applied.

In developed transmitting TV camera on CCD matrix the capability of automatic adjustment of sensitivity by means of the electron obturator is stipulated. The mode of the obturator is selected with the help of switch with constant exposure 1.120 and 40 mlsec. (usual mode)

During the research of the mockup devices is installed, that the operating change of parameters of decomposition TV measuring system on CCD matrixes enables to execute adjustment from frame to frame - time of accumulation and time of frame, and within the limits of frame - number of lines, number of elements and transition from a mode with personnel accumulation to a mode of full flare or to a mode of temporary delay of accumulation. Thus upgraded TV complex allows to provide measurement of brightness of various astronomical objects and their coordinates of rather reference stars with a brightness from 2,0 up to 16,0 sidereal sizes ensuring accommodation of close stars with intersidereal distance 1,0-1,5 angular seconds at identical brightness of their spectra.

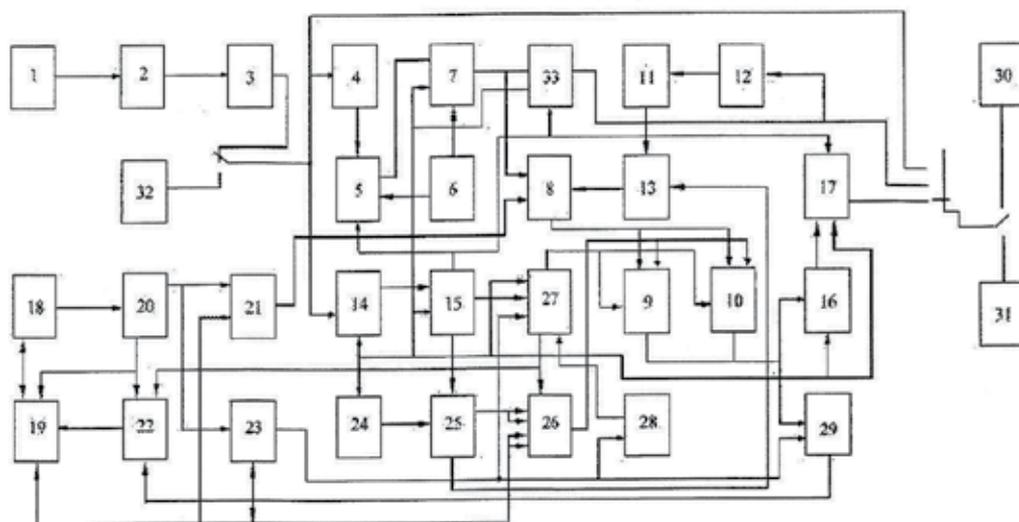


Fig. 1. 19. Block-scheme of MFDMTS

The modified multipurpose digital information measuring system for definition of dynamic objects parameters differs from the first pre-production model, inclusion the sign generator, necessary for registration a number of the frame and time of video information extraction, the device of digital processing and video data compression on the basis of an adaptive linear prediction and complex orthogonal transformation. Besides in the modified variant accuracy of object parameters measurement is reached by use of the new fundamental stars catalogue and the developed technique of their calculation realized in the software. Possibilities of carrying out and day supervision over spacecraft with use of special optical filters are considered. With a view of the international exchange of video data taken from objects are transferred on communication channels using various environments of transfers with a signals compression. The highest degree of video compression ratio at preservation of demanded quality is reached by an adaptive method and optimum algorithms realised in the coding and decoding device.

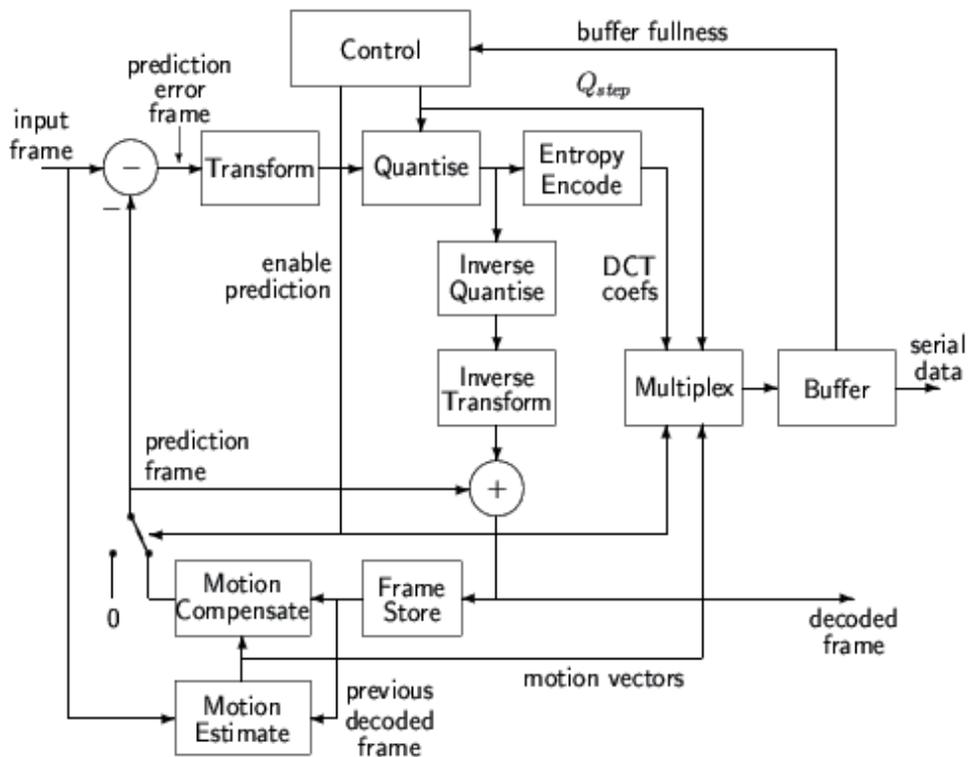


Fig. 1.19. The functional block diagramme of the video coder

6.

Now, for compression of video, there are some various standards to which standards of family MPEG (MPEG-2, MPEG-4), MJPEG, based on discrete-cosine transformation of block structure of images with motion compensation, and JPEG-2000, using wavelet transformations, concern. Thus each standard has the merits and demerits. So, format MPEG-2, using 3 types of the frame, owing to the mechanism of motion compensation of the adjacent frames, allows to get the better general compression of a digital stream, than separately taken keyframes, is provided with comprehensible quality of images at speeds of a digital stream more than 4-5 Mbit/s [15]. However, at smaller speeds of a stream, smoothness change of brightness on borders of blocks is broken, that leads to occurrence of distortions in the form of block effect, reducing legibility and quality of the restored image, as shown in Fig.1.20. In standards MPEG for reduction of digital stream speed the basic compression of video information is carried out due to elimination of interframe redundancy. Usually, it is reached by means of motion compensation methods, thus the information on changes of images in video stream is transferred only. There are very many various motion compensation methods, but in basic their principle of work consists, that original frame is broken on macroblocks for which the maximal conformity in the following frame (Fig.1.21) is searched [16]. If conformity is reached, macroblock is not transferred, since it already exists. If such conformity is not found, the opportunity of such conformity in some area of prospective displacement of

macroblock (to the left, to the right, upwards, downwards) is checked. At detection such macroblock in other position, value of its new coordinates (motion vector) is transferred the decoder. In case of not found conformity it will be transferred the block completely. It allows to receive videostream compression rates much more rather than separately compressed keyframes. However, distortions in the form of block structure of images are inherent in all standards of family MPEG.



Fig. 1. 20. The original image and display of block structure at compression ratio $Cr=100$ with JPEG.

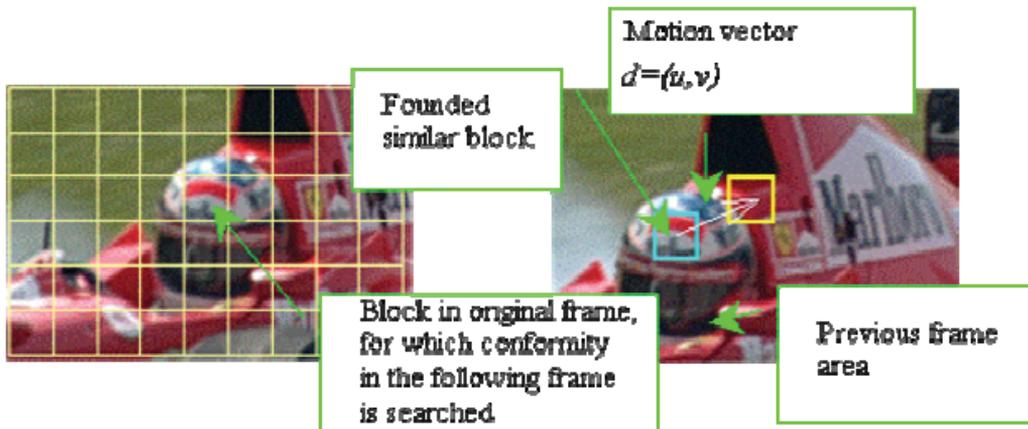


Fig. 1. 21. Principle of motion compensation on the basis of macroblocks searching.

Now rough rates alternative methods of images compression on the basis of wavelet-transformations, at which the image isn't splitted into blocks, and it is processed entirely, develop. It eliminates occurrence of distortions in the form of block effect, therefore images at greater compression rates don't break up to blocks, and simply lose clearness due to border blurriness, but as a whole quality turns out considerably above, than in JPEG (Fig.1.22), that allows to increase compression rate in 1,5-2 times without essential deterioration of the image. However, for wavelet-codecs the effective mechanism of motion compensation as it is made for standards MPEG for the present is not developed, therefore such codecs usually work in standard MJPEG-2000, where everyone image-frame

of videostream is processed and compressed separately, and output videostream consists of a set of static images (keyframes), as shown in Fig.1.23., in which intraframe redundancy is eliminated only [17].



Fig. 1. 22. Comparative quality of images on the basis of JPEG and JPEG-2000 at Cr=100

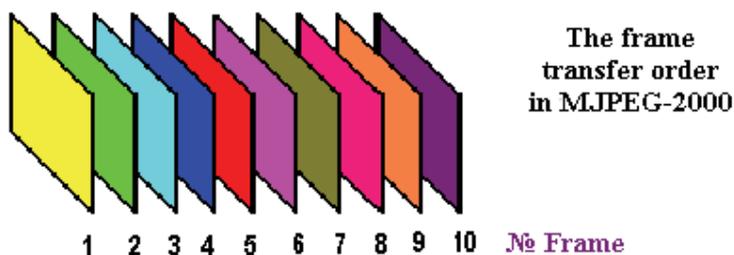


Fig. 1. 23. The frame transfer order in MJPEG-2000.

On the one hand, at identical quality of images it allows to receive compression rate in 1.5-2 times more than the single frame (keyframe) in MPEG, but with reference to video stream, wavelet-codecs considerably concede to them on total compression rate due to absence of motion compensation, which provides the basic compression in MPEG-codec. Basically, wavelet transformations can be applied not to the whole image, but to its fragments or blocks, to which it is possible to apply a method of motion compensation, but such splitting into blocks will lead to occurrence of block distortions, and with more expressed borders, due to mismatch number of samples with a length of wavelet-filter on borders of blocks. Therefore it has been carried out research on estimations of videostream compression efficiency by the wavelet-codec at use of an interframe difference of the next frames. For this purpose 2 algorithms were investigated. The first is based on formation of the interframe difference of the adjacent frames, and the second – using logic operation "exclusive OR", as shown in Fig.1.24. For an estimation of interframe difference efficiency in wavelet-codec of format JPEG-2000 experimental researches of videosequences compression from 10 frames of 3 various plots, in which the first frame was compressed as basic (keyframe) with elimination of intraframe statistical redundancy, and the others 9 with an interframe difference, were spent. Results of images processing are shown in comparative Tables 1-3 and presented in the form of graphic in Fig.1.25.

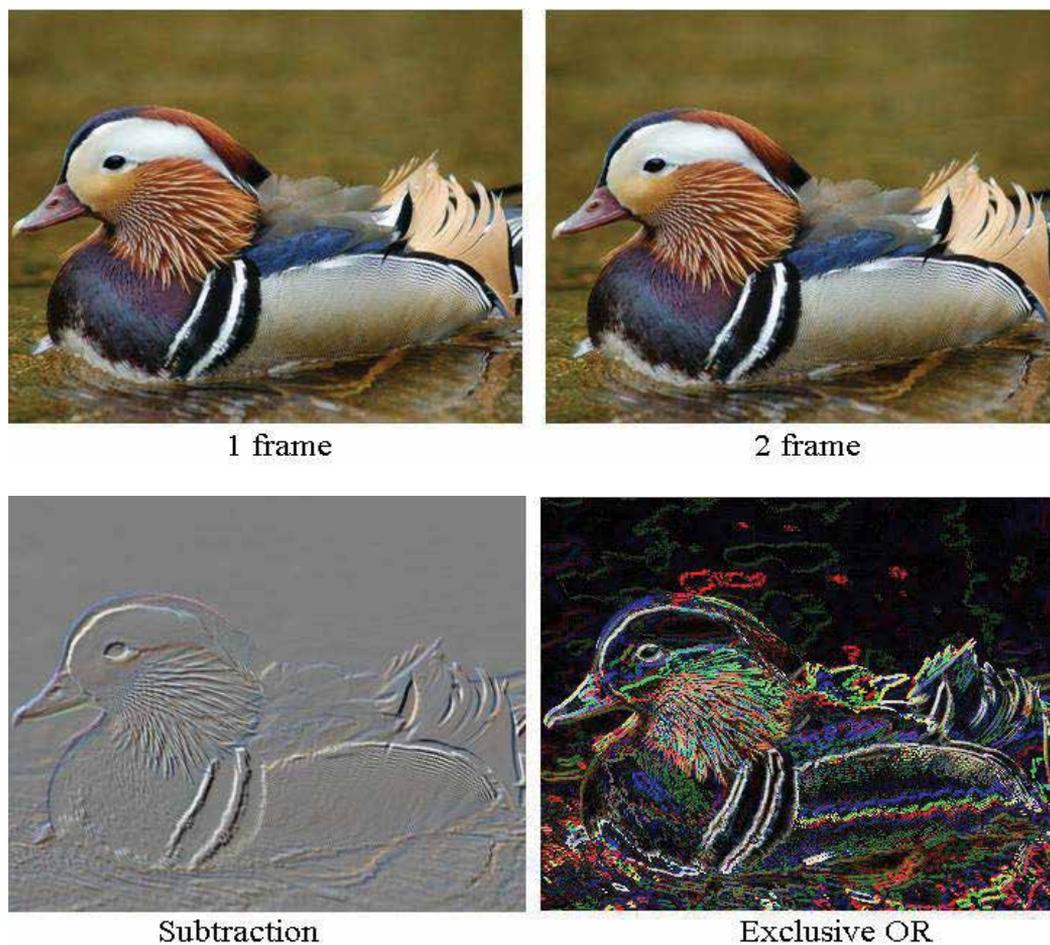


Fig. 24. Formation of interframe difference image on a method of subtraction and logic operation "exclusive OR".

TABLE 1

Results of compression by codec MJPEG-2000 of 1 frame with dimensions 510x265 and size of 406 KB

| № frame | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Σ |
|----------------------|------|------|------|------|------|------|------|------|------|------|----------|
| Frame size, KB | | | | | | | | | | | |
| Lossless compression | 33.1 | 33.6 | 34.5 | 34.3 | 34.7 | 35.3 | 34.8 | 34.6 | 34.6 | 31.1 | 340.6 |
| Compression -20 | 20.4 | 20.4 | 20.4 | 20.4 | 20.2 | 20.2 | 20.4 | 20.2 | 20.4 | 20.4 | 203.4 |
| Compression -50 | 8.1 | 8.1 | 8.1 | 8.1 | 8.1 | 8.1 | 8.1 | 8.1 | 8.1 | 8.1 | 81 |
| Compression -100 | 4.2 | 4.2 | 4.2 | 4.2 | 4.1 | 4.2 | 4.2 | 4.2 | 4.1 | 4.1 | 41.7 |

TABLE 2

Results of compression of 1 frame with an interframe difference and method of subtraction

| № frame Frame size, KB | 0 | 0-1 | 1-2 | 2-3 | 3-4 | 4-5 | 5-6 | 6-7 | 7-8 | 8-9 | Σ |
|---------------------------|------|------|------|-----|-----|-----|------|------|------|------|-------|
| Lossless compression | 33.1 | 11 | 10 | 9.6 | 9.4 | 9.4 | 10 | 10.4 | 10.5 | 10.5 | 123.9 |
| Compression -20 | 20.4 | 11.1 | 10.1 | 9.6 | 9.5 | 9.5 | 10.1 | 10.5 | 10.6 | 10.5 | 111.9 |
| Compression -50 | 8.1 | 8.0 | 8.0 | 8.1 | 8.0 | 8.0 | 8.1 | 8.1 | 8.0 | 8.0 | 80.4 |
| Compression -100 | 4.2 | 4.2 | 4.1 | 4.2 | 4.2 | 4.2 | 4.2 | 4.2 | 4.2 | 4.2 | 41.9 |

TABLE 3

Results of compression of 1 frame with an interframe difference and "exclusive OR" method

| № frame Frame size, KB | 0 | 0⊕1 | 1⊕2 | 2⊕3 | 3⊕4 | 4⊕5 | 5⊕6 | 6⊕7 | 7⊕8 | 8⊕9 | Σ |
|---------------------------|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Lossless compression | 33.1 | 222.8 | 219.3 | 218.2 | 221.2 | 222.9 | 224.4 | 226.6 | 223.6 | 217.9 | 2030 |
| Compression -20 | 20.4 | 29.3 | 29.4 | 29.4 | 29.3 | 29.3 | 29.4 | 29.3 | 29.4 | 29.4 | 284.6 |
| Compression -50 | 8.1 | 8.1 | 8.0 | 8.0 | 8.1 | 8.1 | 8.1 | 8.1 | 8.1 | 8.1 | 80.8 |
| Compression -100 | 4.2 | 4.2 | 4.2 | 4.2 | 4.2 | 4.2 | 4.1 | 4.2 | 4.1 | 4.2 | 41.8 |

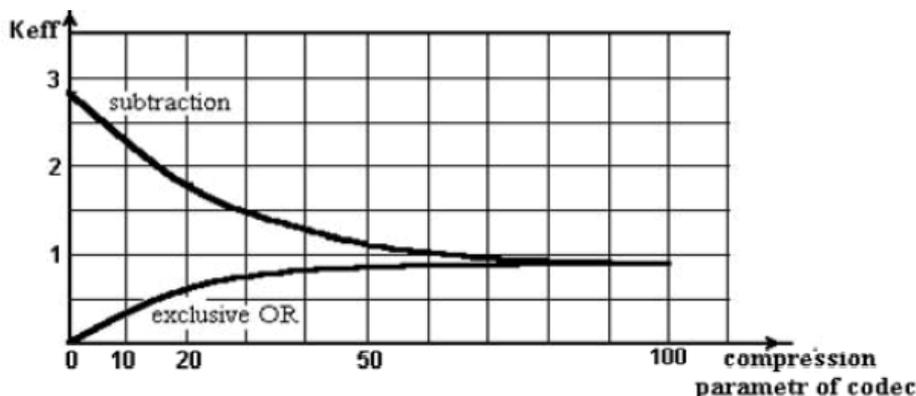


Fig. 1. 25. Dependence of videostream compression efficiency on a method of interframe difference calculation at various wavelet-codec parameters of compression.

Apparently from results of researches, application of an interframe difference gives a positive effect only at small compression rates (10-20 times), and then efficiency falls up to 1, as in an interframe difference high-frequency components which are compressed badly prevail. Especially it is strongly expressed at formation of an interframe difference on a method "exclusive or", as in this case there is a modification of correlation between samples. Thus, as a result of the lead researches, it is established, that use of an interframe difference at greater compression rates doesn't improve videostream compression efficiency in the wavelet-codecs, therefore works on elimination of interframe redundancy of images for wavelet transformations will be continued.

7. References

- [1] Махмудов Э.Б. Разработка, исследование и внедрение методов и устройств повышающих эффективность МЦВИС представляющей различные виды услуг. В сборнике е дународной НТК, Чимкент
- [2] K. Glassman "Videocompression," "625" (1997), No.7
- [3] A. Grossman, J. Morlet, "Decomposition of functions into wavelets of constant shape, and related transforms", in Mathematics_Physics, Lectures on Recent Results Vol. 1 (Ed. L. Streit) (Singapore: World Scientific, 1985)
- [4] В.Воробьёв и В.Грибунин "Теория и практика вейвлет преобразований" 1999
- [5] M. Raad, A. Mertins, and I. Burnett, "Audio compression using the MLT and SPIHT," Proceedings of DSPCS' 02, pp. 128--132, 2002.
- [6] А. Зубинский "Другое лицо Интернета?" Компьютерный обзор (1999) No. 47
- [7] Gorelin S.L., Kan B.M., Kivirin V.I. Television measuring systems. Communication 1980 - 168 pages.
- [8] Television astronomy. Communication, Moscow. 1986.
- [9] Lebedev N.V., Skurlov K.V., Sharov S.L. Television system with controlled parameters of expansion on CCD. Engineering of communication facilities. A series engineering of a television. № 2 Moscow 1990.
- [10] Lebson M.N., Khesin A.Y., Yansen B.A. Automation of recognition of television images. Energia. Moscow 1975.-160 page.
- [11] Ptachek M. Digital television. The theory and engineering. Moscow. Radio and communication 1990.-528 page.
- [12] Systems of a single synchronization with elements of digitization. 2 issue under edition V.V.Shakhgildyan. Moscow. Radio and communication 1989.-320 page.
- [13] Grigoryev B.S. The driven programmed television synchro-generator. Engineering of communication facilities. A series engineering of a television. № 2 Moscow 1990. 98-108.
- [14] 1) The patent of the Republic of Uzbekistan № 4042. The device of a correction of TV sign (Makhmudov E.B., Mirakhmedov V.S., Tikunov S.S.) B.I.O. № IHDP from 9500661.1 from 30.12.1996. 2)The patent of the Republic of Uzbekistan № 4259 . The device of a Tvsignals compression (Makhmudov E.B., Holmatov O.A.) B.I.O. № IHDP from 9500628.1 from 13.05.1998
- [15] Gavrilov I.A., Ibraimov R.R., Benilov A.I., Ibraeva S.M., Ignatieva O.S., Chernyshov A.A. - "Study of TV-signals over cellular networks transmission possibility", The Second International Conference In Central Asia on Internet The Next Generation of Mobile, Wireless and Optical Communications Networks (ICI2006), 2006, Tashkent.
- [16] Кубасов, Д. Ватолин «Обзор методов компенсации движения» <http://cgm.graphicon.ru/content/view/76/65/>
- [17] MichaelD.Adams «JPEG-2000StillImageCompression Standard» Dept. ofElectricala nd Computer Engineering. University of British Columbia Vancouver, BC, Canada V6T 1Z4

Digital Video Image Quality

Tomáš Kratochvíl and Martin Slanina

*Department of Radio Electronics, Brno University of Technology
Czech Republic*

1. Introduction

The video image quality in digital television systems is subject to quite different effects and influences than that in analog television systems. There are mainly two sources which can disturb the digital video image quality and which can cause visible degradation of video image quality. These are the source coding and the related compression and transmission link from the modulator to the receiver. The perturbation, noise, transmission channel influence or transmission distortions can cause an increase of channel bit error rate and due to the error protection, e.g. FEC (Forward Error Correction) in DVB (Digital Video Broadcasting) (Fisher, 2008), included in the signal, most of the bit errors can be repaired up. It leads to QEF (Quasi Error Free) transmission conditions, and the errors are not noticeable in the video image. If the transmission channel is too noisy, the transmission totally breaks down. This situation is well known as the “fall of the cliff”, or simply “cliff off”. The linear or nonlinear distortion has no direct effect on the video image, but in an extreme case it can also lead to a breakdown. No matter if the picture quality is good, bad or indifferent, it needs to be evaluated differently and detected by different means in DTV (Digital Television) and DVB systems than in ATV (Analog Television). An example of the video image quality in ATV and DTV system is shown in Fig. 1.

There are several dimensions of digital video image quality evaluation, generally splitted into the subjective and objective methods. The subjective evaluation is a result of human observers providing their opinion on the video image quality. The objective evaluation is performed with the aid of instrumentation, calibrated scales and mathematical algorithms. Direct measurements are performed with the video images (picture quality measurement) and indirect measurements are made processing specially designed test signals in the same manner as the pictures (signal quality measurement) (Tektronix, 1997). The test video image sequences are used for both direct measurements, subjective and objective, but in a compressed digital video image system, they can not be used for the compression encoder/decoder part of the system because a comparison of the codec influence on the common test scenes and natural scenes is not possible. To specify, evaluate and compare digital video systems with video image artifacts caused by compression or transmission, the quality of the digital video and image presented to the observer has to be determined. Video image quality is inherently subjective and is affected by many subjective factors. It could be difficult to obtain accurate measures and results. Measuring video image quality using objective criteria results is an accurate and repeatable evaluation, but there is still no general objective evaluation. It should naturally cover the subjective experience of a human observer and performance of a video display and viewing conditions (Richardson, 2002).

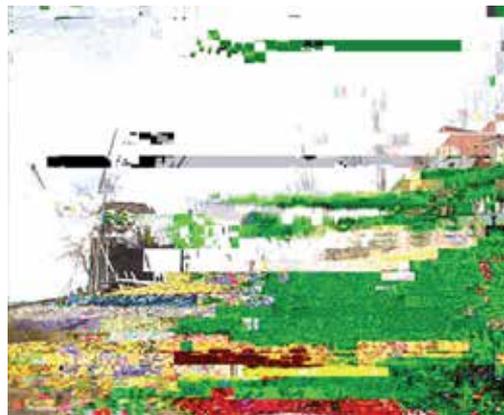
a) $\text{PSNR}_Y = 29.85$ dBd) $\text{PSNR}_Y = 29.74$ dBb) $\text{PSNR}_Y = 20.65$ dBe) $\text{PSNR}_Y = 19.97$ dBc) $\text{PSNR}_Y = 13.58$ dBf) $\text{PSNR}_Y = 12.85$ dB

Fig. 1. Examples of typical distortion artifacts in the video transmission over noisy channels. Uncompressed video sequence a) low level of noise, b) average level of noise, c) high level of noise. MPEG-2 algorithm compressed video sequence d) low bit-error rate and QEF transmission, e) average bit error rate and blockiness, f) high bit-error rate and cliff off effect.

2. Subjective test procedures

The test procedures for subjective test are defined especially in ITU-R recommendation BT.500-11 (ITU, 2001). The most popular is evaluation by the DSCQS (Double Stimulus Continuous Quality Scale) method. An assessor evaluates a pair of digital video image short sequences, called A and B, one after another. Then he is asked to give a score to A and B sequences on a continuous scale. The scale is divided into five intervals of the subjective quality scores reaching from excellent through good, fair, poor to bad quality. The impairment scale related to the mentioned five intervals is in Tab. 1.

| Score | Quality | Impairment |
|-------|-----------|--------------------------|
| 5 | Excellent | Imperceptible |
| 4 | Good | Perceptible but annoying |
| 3 | Fair | Slightly annoying |
| 2 | Poor | Annoying |
| 1 | Bad | Very annoying |

Table 1. Score and related subjective quality evaluation criteria

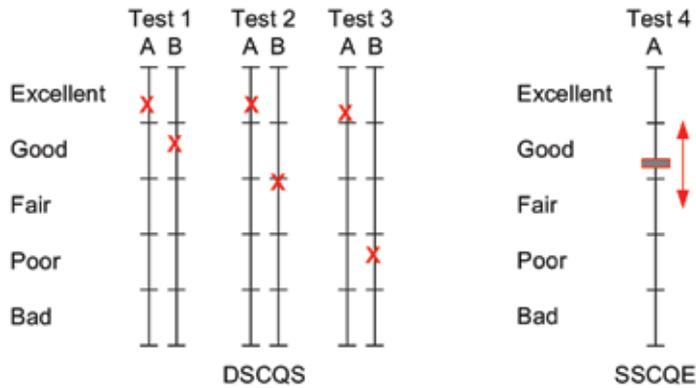


Fig. 2. Subjective methods and continuous quality rating forms for DSCQS and SSCQE

In a typical test session, the assessor is presented a series of sequence pairs and is asked to evaluate a grade of each individual pair, as shown in Fig. 2. In each pair of digital video image sequences, one is unimpaired (so-called reference sequence) and the other is the same sequence modified by a compression algorithm or process under test, e.g. video image compression or transmission. A typical example is a video coding system as shown in Fig. 3. In this case, the original sequence is compared to the same video image sequence which was subject to encoding and decoding. The order of the evaluated sequences is randomized during the evaluation, so the assessor does not know which sequence (original or impaired), he is currently evaluating. This prevents the assessor from predicting and prejudging the results. At the end of the test, the scores are normalized and the result is a score that indicates the relative video image quality of the impaired and reference sequences. The resulting score is denoted to as MOS (Mean Opinion Score).

The DSCQS test can be used as a realistic measure of subjective digital video image quality. In its application it must be considered that it suffers from several practical problems. The evaluation can vary significantly and depends on selection of the assessors and also on the characteristics of the video image sequence under test. This variation can be reduced by

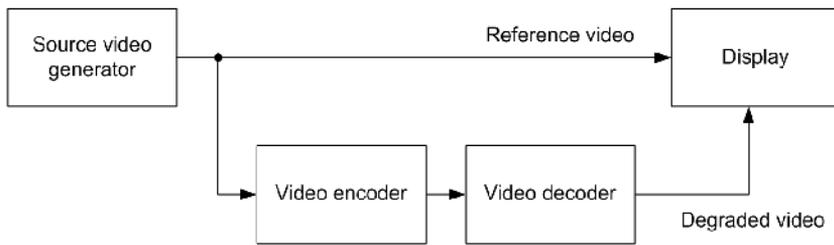


Fig. 3. General arrangement for the DSCQS method evaluation

repeating the test with several sequences and several assessors. An expert assessor, who is familiar with the digital video image artifacts caused by compression, can give a biased score. That is why the non-expert assessors are preferred. Additionally, non-expert assessors can quickly learn to recognize compression artifacts in the digital video image sequences. Subjective tests are also influenced by the viewing conditions. A test carried out in a comfortable environment will be evaluated with a higher score than the same test carried out in a less comfortable setting. It has been proved that the “recency effect” (Richardson, 2002) means that the assessor’s opinion is significantly biased by the last few seconds of a video image sequence. The quality of the last section will strongly influence the score of a whole longer sequence. That is why the subjective test is really suitable only for short video image sequences.

The second popular subjective test defined by ITU-R BT.500 recommendation is the SSCQE (Single Stimulus Continuous Quality Evaluation) method (ITU, 2001). In this subjective test the assessor evaluates video image quality without the reference sequence. Since the SSCQE deliberately dispenses the reference sequence, this method can be used more widely in practice. In this method a group of test persons assesses the processed video image sequence only and evaluates the score again from excellent to bad, which also provides a video image quality profile over time (see Fig. 2).

The advantages of subjective testing are in obtaining valid results for conventional and compressed television systems and evaluation of scalar MOS that works over a wide range of still and motion picture applications. The disadvantages of subjective testing are in a wide variety of possible methods which must be considered for the test. Many observers must be selected and it is very time consuming in case the procedure respects all the requirements.

3. Objective tests

Objective test methods are based on automated, computational approach. Depending on the original video image sequence, the objective test results are not always correlated with the impression of quality in a subjective observation. The degree of correlation to subjective results can be considered a benchmark of subjective tests.

The first choice when selecting a metric for full-reference quality evaluation is usually the peak signal-to-noise ratio (PSNR). For video sequences, it can be easily computed as (Winkler, 2005)

$$\text{PSNR} = 10 \cdot \log_{10} \frac{m^2}{\text{MSE}}, [\text{dB}] \quad (1)$$

where m is the number of values by which a pixel can be represented (e.g. $m = 255$ for 8-bit luma samples) and MSE is the mean squared error, computed as

$$\text{MSE} = \frac{1}{T \cdot M \cdot N} \sum_{k=1}^T \sum_{i=1}^M \sum_{j=1}^N [f(k, i, j) - \tilde{f}(k, i, j)]^p \quad (2)$$

The constants M, N, T are the horizontal and vertical dimensions in pixels and the number of frames (fields), respectively, f and \tilde{f} are the sample values (luma or chroma) of the degraded and the reference video sequence, respectively. The peak signal-to-noise ratio is very simple and easy to implement, but its disadvantage is in poor correlation with subjective tests.

A great effort has been devoted to developing objective video quality metrics in recent years. An ideal objective quality metric should closely simulate the results of subjective tests. Many approaches have been proposed to achieve this, with different success. To select the metrics suitable for real applications, two phases of tests were performed by the Video Quality Experts Group (VQEG) in 2000 and 2003, respectively (VQEG, 2000; VQEG, 2003).

In the first testing phase ten proposed quality evaluation algorithms were considered, and the correlation of their results with subjective scores obtained for video sequences with different characteristics (different scene contents) and subject to different quality degradations (compressed with H.263, MPEG-2 encoders using different settings, Betacam with drop-out) was examined. All the tested video sequences were in standard definition, considering both 625- and 525-line television systems. The first phase of testing was completed only with a limited success - the performance of the proposed quality evaluation algorithms was very close to the performance of PSNR. As a result, none of the tested algorithms was proposed by the VQEG to be included in an ITU Recommendation.

Another testing (Phase II) was realized by the VQEG a couple of years later, considering a set of six proposed quality evaluation algorithms. The testing procedures were very close to those performed in Phase I. However, out of the six proposed algorithms, four were selected to be included in an ITU Recommendation, published in 2004 as ITU-R Recommendation BT.1683. In the following subsections, the principles of the four standardized algorithms will be briefly described (ITU, 2004).

3.1 The BTFR algorithm

This metric was designed by the British Telecom, United Kingdom, and is denoted to as the BTFR algorithm (British Telecom full-reference automatic video quality assessment tool).

The algorithm computes several measures comparing the degraded video and the reference video, to finally combine the measures together to get a quality prediction. A simple diagram of the algorithm operation is shown in Fig. 4. The preprocessing block in the diagram consists of several steps. It includes format conversion, cropping, offset and matching operations. These are performed on the luma (Y) as well as chroma (U, V) components of both the degraded and the reference video sequences. Matching operations are also included in the preprocessing block. They consist in finding the best match for blocks within each degraded field from a buffer of neighboring reference fields. A matched video sequence is then used instead of the reference sequence in some of the following analyses:

- *PSNR analysis* - a PSNR calculation is performed using the degraded and matched reference sequences.

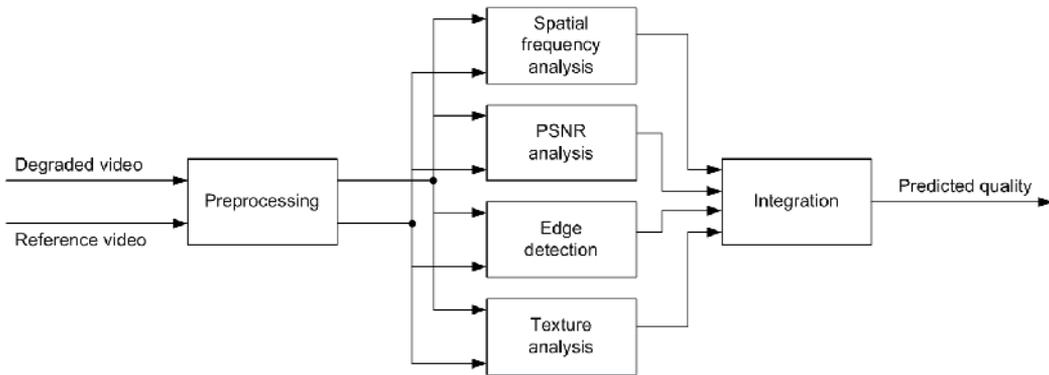


Fig. 4. Operation of the BTFR algorithm

- *Spatial frequency analysis* – a pyramid transformation of the degraded and matched reference sequences is performed. The differences between pyramid arrays are then calculated using SNR.
- *Edge analysis* – an edge detector is used to create edge maps of the matched reference and degraded video sequences. The total numbers of edge-marked pixels are then calculated for both edge maps.
- *Texture analysis* – the texture properties are measured by recording the number of turning-points in the luma signal along horizontal picture lines.

Finally, all the parameters gained from the analyses are put together in the Integration block to form the final quality measure. The integration is nothing but computing a linear combination of the parameters, with specified weights and offset.

3.2 The EPSNR algorithm

The second metric described in Rec. BT.1683 was designed in cooperation of the Yonsei University, Radio Research Laboratory and SK Telecom, Republic of Korea. It is based on the fact that human observers are very sensitive to degradations around edges – when the edges are blurred, the subjective scores are likely to be worse. Additionally, many compression algorithms tend to produce artifacts around the edges. The metric computes a value called Edge PSNR (EPSNR) and uses it as a quality measure after post-processing. A block diagram of the metric is shown in Fig. 5.

Using an arbitrary edge detection algorithm, edge areas are located in the first step using the reference video sequence. For each field (or frame when processing progressive video), an edge mask image is created – the algorithm operates on a field-by-field basis. Then differences between the reference and the degraded video fields are computed, based on simple mean squared error evaluation, limited to the edge areas. Finally, PSNR of the edge areas (EPSNR) is computed from the mean squared error.

In the final phase, post-processing is applied to the EPSNR value of the actual field, taking into account the following:

- For high PSNR values, the EPSNR overestimates perceptual quality. The solution is in piecewise linear scaling (reduction) of EPSNR values over 35.
- If the degraded video is severely blurred (the number of edges detected in the degraded video is significantly lower than the number of edges in the reference video), the EPSNR is reduced.
- Scaling is performed at the end to reach the range of outputs between 0 and 1.

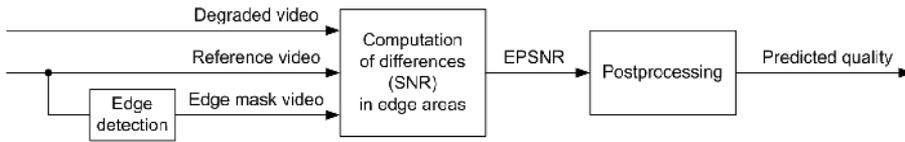


Fig. 5. Metric based on Edge PSNR

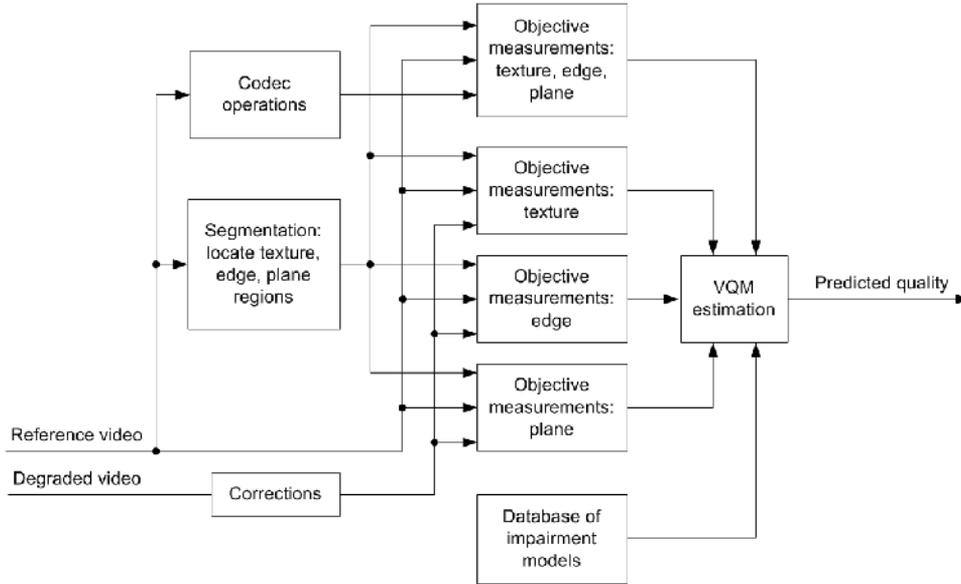


Fig. 6. Simplified block diagram of the IES metric

3.3 The IES algorithm

This metric is designed by the Center of Telecommunications Research and Development, Brazil (CPqD). It is denoted to as the IES system (image evaluation based on segmentation). Its simplified block diagram is shown in Fig. 6.

Based on analysis of the reference video sequence, each field (frame) is segmented into texture, edge, and plane regions. Each of these regions is then processed separately: objective quality measures are computed for the reference and the degraded video fields after correction (offset, gain). As each of the objective measures is performed for the luma as well as both chroma components, it results in nine parameters.

An interesting approach is in including a block called Codec operations. This is nothing but encoding and decoding the reference video sequence using two different codecs in parallel: the MPEG-2 4:2:0 and the MPEG-1 CIF algorithms (with fixed encoder settings). The resulting video sequences together with the reference video are then subject to the same objective analysis as the reference - corrected degraded video sequence pair from the input. Thus, for two codecs, three region types, and three components, we get the total of eighteen additional objective parameters for the quality estimation.

The IES system uses a database of impairment models for scenes different from the reference video scenes in order to estimate the subjective quality of the degraded video sequence. The database consists of information about sequences with different degrees of motion and

detail and different context. Together with spatial and temporal attributes extracted from the reference video sequence, it forms another input for the quality estimation algorithm. Finally, subjective impairment levels are estimated from the objective parameters and the resulting predicted quality measure is achieved through a linear combination of the estimated impairment levels.

3.4 The VQM algorithm

The fourth and the last metric described in Rec. 1683 was designed by the National Telecommunication and Information Administration / Institute for Telecommunication Science, US. The acronym VQM used by the authors stands for Video Quality Metric. The metric is quite complex, involves preprocessing (matching) operations as well as a thorough evaluation of video sequence properties. An important feature is that this metric natively operates not only on separate fields (frames), but breaks the video sequences into S-T (spatial-temporal) sub-regions, including a block of pixels in several consequent fields.

The metric can also be called a reduced-reference metric, as it extracts a certain information (quality feature) from the reference and the degraded video sequences, and then forms a quality measure based on the extracted information only – just several values. The quality features include information on spatial gradients of the video scenes, chrominance information, contrast information, temporal information, etc. Their proper combination gives the metric output value. Typically, one output value is calculated for one video clip of 5 – 15 seconds in length.

The Fig. 7 shows frame of a video sequence compressed with the H.264/AVC algorithm using different settings. An original uncompressed sequence is shown in Fig. 7a). The sequence in Fig. 7b) is compressed with high bit rate, fine quantization, and the resulting PSNR is almost 35 dB. The output value of the VQM metric is almost zero, which means there will be probably no noticeable difference from the original. Indeed, both pictures look identical. Now look at the pictures in Fig. 7c) and Fig. 7d). Even though the PSNR computed for the whole sequence (100 frames) has almost the same values, the VQM differs considerably. By taking a close look at the pictures, especially on the bush in front of the house, much more blur is visible on the picture in Fig. 7d). This proves that the different degradation was not captured by the peak signal-to-noise ratio, but the VQM metric exhibits quite different values showing that the quality in the bottom right picture in Fig. 7d) is worse. The PSNR is computed only in the luminance channel, which has the highest impact on the perceived quality. The range of the VQM values is from zero to one, and the best quality with no degradation is represented by a zero.

4. Objective tests with no reference

The no-reference video quality assessment metrics cannot rely on any information about the original material. What information is then available at the receiver side and can be used for measurement? Usually, no-reference metrics use some a-priori information about the processing system. For example, a usual DVB-T broadcasting system using MPEG-2 source coding is known to have the block artifacts as the most annoying impairment (Fischer, 2008). Tracking these artifacts down in the video image may provide enough information to judge the overall quality. In the following text, metrics for still image quality evaluation will be considered as well as those for video sequences only. In fact, most of the still image metrics can be used for video sequences when applied on each video frame.

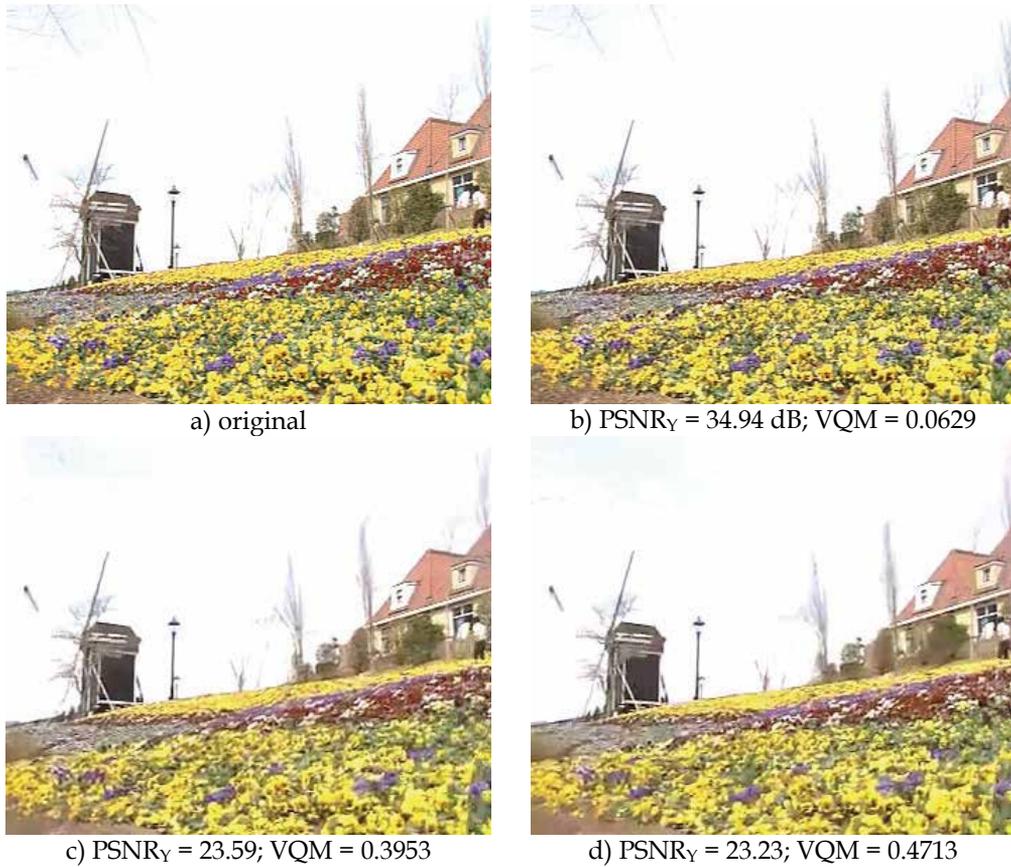


Fig. 7. Frame of a compressed video sequence using the H.264 algorithm. PSNR and according VQM video image quality evaluation results are shown.

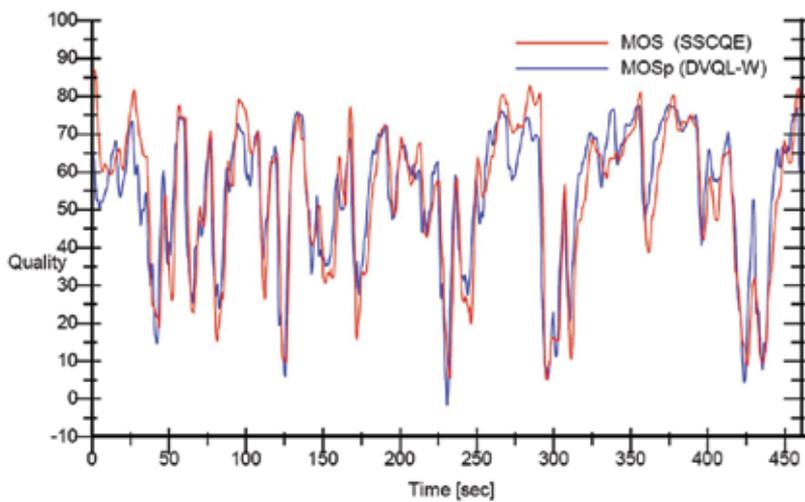


Fig. 8. Comparison of subjective and objective picture quality (Lauterjung, 1998)

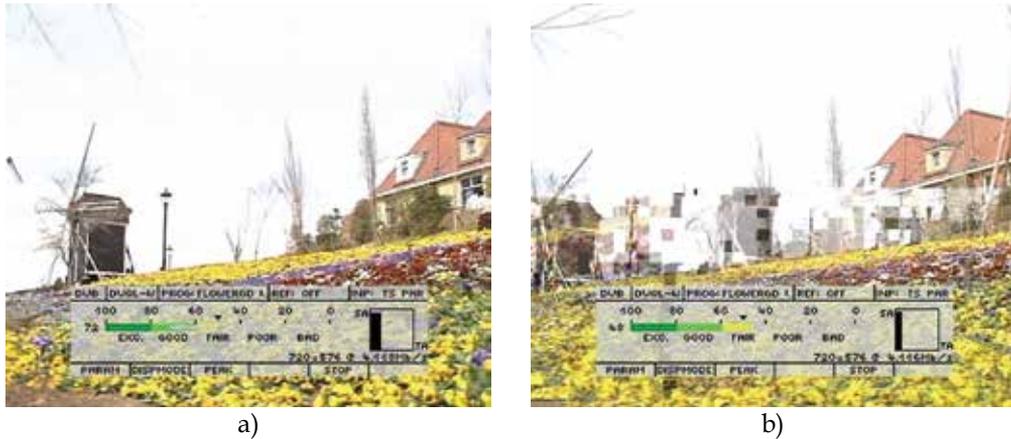


Fig. 9. Example of a real-time objective picture quality analysis using DVQ analyzer. The metric DVQL-W is used to demonstrate the achieved video image quality in one frame of the a) reference video, b) degraded video with blockiness and cliff-off effect present.

4.1 No-reference analysis using pixel values

The block artifact detection approach is used in DVQ analyzer - video quality measurement equipment supplied by Rohde & Schwarz and is briefly described in (Fischer, 2008). The principle of the method is in assumption that block artifacts create a regular grid with constant distances. Neighboring pixel differences are computed for the whole image and averaged in such manner that only 16 values remain (since MPEG-2 is supposed to create 16 x 16 blocks). If the average pixel value difference is significantly larger on block boundaries, a statement can be made that block artifacts are present in the image.

To bring the objective test results closer to the subjectively perceived quality, other quantities in the moving picture are also taken into consideration. These are spatial and temporal activities (Fischer, 2008). The spatial activity is a measure of the existence of fine structures in the video image and temporal activity is a measure of change and movement in successive frames. Both activities can render the blocking structure invisible or mask them. Such artifacts in the video image are then simply not seen by the human eye.

If masking is incorporated, the DVQL-W (Digital Video Quality Level - Weighted) metric applied in DVQ analyzer delivers a prediction of the MOS. With the masking included, the algorithm shows an excellent correlation with subjective assessment results as it is shown in the Fig. 8. The results of the subjective evaluation were obtained by the SSCQE method. The compiled test sequence consisted of 11 well-known test sequences such as "Flowergarden", etc. The data rates for the sequences varied between 1 MBit/s and 9 MBit/s. From the subjective assessment about 1000 measurement values were obtained. Their scaling factor was re-based and a fixed delay of 1 second was introduced. With this optimization, an overall correlation of more than 94 % was achieved (Lauterjung, 1998).

An example of a real-time measurement using DVQ analyzer is shown in Fig. 9 and numerical results are in the Fig 10. The DVQL-W metric evaluates blocking structure in the video image of a selected DTV program in an MPEG-2 TS. It is obvious from Fig. 9 that the quality decreases with the blockiness in the video. The temporal and spatial activity and evaluation in the luminance and chrominance video channels are considered (Fisher, 2008).

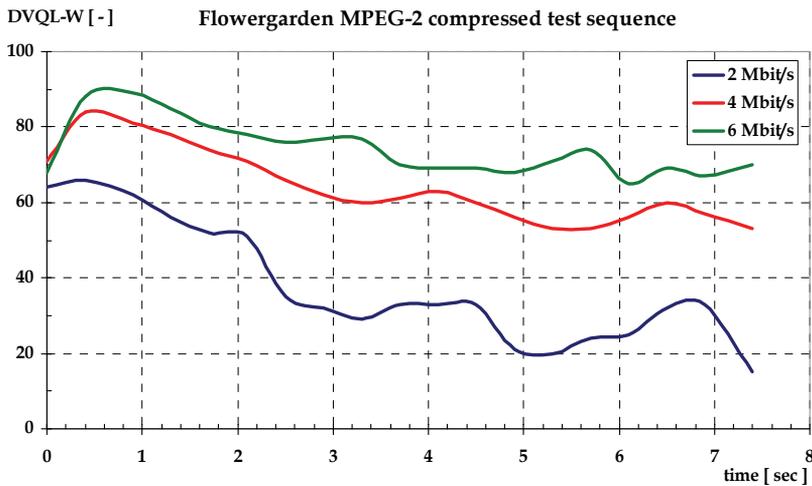


Fig. 10. Video image quality analysis of the „Flowergarden“ test sequence. The DVQL-W metric and according quality varies with the MPEG-2 compressed test sequence bitrate.

In (Pan et al., 2004), a no-reference algorithm was presented capable of detecting block artifacts in a block-by-block manner and, as an extension, detects flatness of the image. A no-reference block and blur detection approach was introduced in (Horita et al., 2004), designed to measure quality of JPEG and JPEG2000 images. Another no-reference algorithm for block artifact detection was described in (Wang et al., 2000). Another common distortion, blur, can also be used for quality evaluation. Of course, depending on the characteristics of the processing system (whether or not the system is likely to introduce blur). An interesting metric was presented in (Ong et al., 2003). A very similar approach is also used in (Marziliano et al., 2002). In principle, these metrics analyze how steep the changes in pixel values within the line are. The main difference is that (Ong et al., 2003) analyzes not only the horizontal direction, but measures blur in four directions instead.

An interesting no-reference approach was used in (Tong et al., 2005), using a learning algorithm to assess the overall quality of an image. The metric uses pixel values of a decoded picture, which was subject to JPEG or JPEG2000 compression.

4.2 No-reference analysis using transform coefficients and encoded stream values

In (Sheikh et al., 2005), a metric was presented for JPEG2000 compressed static images. The JPEG2000 standard uses wavelet transform. The authors analyze the wavelet coefficients to gain a quality measure. An observation was made that in natural images, these coefficients have some characteristic properties. If the wavelet coefficients do not behave in a desired manner, quality degradation can be expected. However, this metric is only applicable for wavelet transform compressed images, and thus not applicable for any of the wide-spread present-day video compression standards. Anyway, coefficient analysis for video sequences is also possible.

In (Gastaldo et al., 2002), such analysis was performed for MPEG-2 compressed video sequences. First of all, a statistical distribution analysis was performed to say which of the features available in the MPEG-2 transport stream may be used for evaluation. Over twenty features were then used to feed an artificial neural network for learning and consequently

for quality evaluation. For this metric, a correlation as high as 0.85 was achieved by the authors. A different approach was published in (Eden, 2007), where the author computes PSNR values of a H.264/AVC using transform coefficient and quantization parameter values, which means computation can be done on the encoded bit stream only.

4.3 No-reference metric developed at the Brno University of Technology

A no-reference quality metric was recently developed at the Brno University of Technology and published in (Slanina & Říčný, 2008). The metric operates on a compressed bit stream conforming to the H.264 / AVC standard. The idea is based on the fact that the encoder can adaptively select the sizes of blocks to be coded, and the coarseness of quantization of residual transform coefficients. A very simple artificial neural network is then used to process the input parameters, represented as ratios of the block sizes used by the encoder, the quantization parameter, and information on the quality of the reference frames for the inter predicted (using motion compensation) frames. The metric is not supposed to output values simulating subjective tests – the artificial neural network is trained to simulate PSNR values for a given compressed video sequence without reference.

The attainable correlation of the metric with real PSNR values is above 0.95. This value is somewhat lower than the correlation achieved in (Eden, 2007). Anyway, the algorithm is designed in such manner that it can be easily changed to predict different values than just the PSNR. The authors are currently working on predicting output values of the standardized full-reference algorithms. So far, it turns out that to achieve satisfactory results, the number of parameters extracted from the bit stream needs to be increased (the bit stream carries other low cost information, such as the bit rate, GOP format, etc.).

5. Conclusion

Measuring video image quality is difficult and very often not precise. There are many factors that can affect the results and their interpretation. The advantages of subjective testing are in obtaining valid results for conventional and compressed television systems and the possibility of evaluating scalar MOS over a wide range of still and motion video image applications. Their disadvantages are in a wide variety of possible methods and tests to be considered, the high number of observers required and in high time demands. There are many objective testing approaches. It can be stated that the algorithms with video image feature analysis correlate better with subjective results than just simple pixel-based methods. A combination of different measurements and features gives the best results and correlation between subjective and objective scores but it is hardly technology independent.

6. Acknowledgement

This work was supported by the Research program of Brno University of Technology no. MSM0021630513, “Electronic Communication Systems and New Generation Technology (ELKOM)” and the research project of the Czech Science Foundation no. 102/08/P295, “Analysis and Simulation of the Transmission Distortions of the Digital Television DVB-T/H”. The research leading to these results has received funding from the European Community's Seventh Framework Programme under grant agreement no. 230126.

7. References

- Fisher, W. (2008). *Digital Video and Audio Broadcasting Technology. A Practical Engineering Guide*. 2nd edition. Springer, ISBN978-3-540-76357-4, Berlin
- Eden, A. (2007). A no-reference estimation of the coding PSNR for H.264-coded sequences, *IEEE Transactions on Consumer Electronics*, Vol. 53, No. 2, May 2007, pp. 667-674, ISSN 0098-3063
- Gastaldo, P.; Zunino, R. & Rovetta, S. (2002). Objective assessment of MPEG-2 video quality, *Journal of Electronic Imaging*, Vol. 11, No. 3, July 2002, pp. 365-374, ISSN 1017-9909
- Horita, Y.; Arata, S. & Murai, T. (2004). No-reference image quality assessment for JPEG/JPEG2000 coding, *Proceedings of XII. European Signal Processing Conference EUSIPCO-2004*, Vol. 2, pp. 1301-1304, ISBN 3-200-00148-8, Vienna, September 2004, Vienna University of Technology
- ITU (2001). *Methodology for the subjective assessment of the quality of television pictures*. Recommendation ITU-R BT.500-11. 2001.
- ITU (2004). *Objective Perceptual Video Quality Measurement Techniques for Standard Definition Digital Broadcast Television in the Presence of a Full Reference*. Recommendation ITU-R BT.1683. 2004.
- Lauterjung, J. (1998). *Picture Quality Measurement*. Rohde & Schwarz GmbH & Co KG. Reprint of paper done at IBC Sept. 1998, Amsterdam, 1998. [online] http://www2.rohde-schwarz.com/file_3609/PQM.pdf
- Marziliano, P.; Dufaux, F.; Winkler, S. & Ebrahimi, T. (2002). A no-reference perceptual blur metric, *Proceedings of the International Conference on Image Processing*, Vol. 3, pp. 57-60, ISBN 0-7803-7622-6, Rochester, NY, September 2002
- Ong, E. P. et al. (2003). A No-reference quality metric for measuring image blur, *Proceedings of the Seventh International Symposium on Signal Processing and Applications*, Vol. 1, pp. 469-472, ISBN 0-7803-7946-2, Paris, July 2003, IEEE
- Pan, F et al. (2004). A Locally-Adaptive Algorithm for Measuring Blocking Artifacts in Images and Videos, *Proceedings of the 2004 International Symposium on Circuits and Systems ISCAS '04*, Vol. 3, pp. 925-928, ISBN 0-7803-8251-X, Vancouver, May 2004, IEEE Circuits and Systems Society, Piscataway
- Richardson, I.E.G. (2002). *Video Codec Design. Developing Image and Video Compression Systems*, Wiley, ISBN 0-471-48553-5, Chirchester
- Sheikh, H. R.; Bovik, A. C. & Cormack, L. K. (2005). No-reference quality assessment using natural scene statistics: JPEG2000, *IEEE Transactions on Image Processing*, Vol. 14, No. 11, November 2005, pp. 1918-1927, ISSN 1057-7149
- Slanina, M. & Říčný, V. (2008). Estimating PSNR in high definition H.264/AVC video sequences using artificial neural network, *Radioengineering*, Vol. 17, No. 3, September 2008, ISSN 1210-2512 [online] http://www.radioeng.cz/fulltexts/2008/08_03_103_108.pdf
- Tektronix, Inc. (1997). *A guide to Picture Quality Measurements for Modern TV Systems*. Tektronix Inc., 1997. [online] http://www.tek.com/Measurement/App_Notes/PicQuality/25W_11419_0.pdf
- Tong, H. et al. (2005). Learning no-reference quality metric by examples, *Proceedings of the 11th International Multimedia Modelling Conference MMM '05*, pp. 247-254, ISSN 1550-5502, Melbourne, 2005, IEEE Computer Society

- VQEG (2000). *Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment*. ITU, 2000. [online]
http://www.its.bldrdoc.gov/vqeg/projects/frtv_phaseI/COM-80E_final_report.pdf
- VQEG (2003). *Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment, Phase II*. ITU, 2003. [online].
http://www.its.bldrdoc.gov/vqeg/projects/frtv_phaseII/downloads/VQEGII_Final_Report.pdf
- Wang, Z.; Bovik, A. C. & Evans, B. L. (2000). Blind measurement of blocking artifacts in images, *Proceedings of 2000 International Conference on Image Processing*, Vol. 3, pp. 981-984, ISBN 0-7803-6297-7, Vancouver, September 2000
- Winkler, S. (2005). *Digital Video Quality: Vision Models and Metrics*. Wiley, ISBN 0-470-02404-6, Chichester

Edited by Floriano De Rango

This book tries to address different aspects and issues related to video and multimedia distribution over the heterogeneous environment considering broadband satellite networks and general wireless systems where wireless communications and conditions can pose serious problems to the efficient and reliable delivery of content. Specific chapters of the book relate to different research topics covering the architectural aspects of the most famous DVB standard (DVB-T, DVB-S/S2, DVB-H etc.), the protocol aspects and the transmission techniques making use of MIMO, hierarchical modulation and lossy compression. In addition, research issues related to the application layer and to the content semantic, organization and research on the web have also been addressed in order to give a complete view of the problems. The network technologies used in the book are mainly broadband wireless and satellite networks. The book can be read by intermediate students, researchers, engineers or people with some knowledge or specialization in network topics.

Photo by mirceagrau / iStock

IntechOpen

