IntechOpen

Face Recognition

Edited by Milos Oravec

Face Recognition

Edited by

Miloš Oravec

Face Recognition
http://dx.doi.org/10.5772/207
Edited by Milos Oravec

Notice

Statements and opinions expressed in the chapters are these 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

Face Recognition
Edited by Milos Oravec

p. cm.

ISBN 978-953-307-060-5

eBook (PDF) ISBN 978-953-51-4550-9

Contents

Preface

Face recognition has been studied for many years in the context of biometrics. The human face belongs to the most common biometrics, since humans recognize faces throughout their whole lives; at the same time face recognition is not intrusive. Face recognition systems show many advantages, among others easy implementation, easy cooperation with other biometric systems, availability of face databases.

Nowadays, automatic methods of face recognition in ideal conditions (for two-dimensional face images) are generally considered to be solved. This is confirmed by many recognition results and reports from tests running on standard large face databases. Nevertheless, the design of a face recognition system is still a complex task which requires thorough choice and proposal of preprocessing, feature extraction and classification methods. Many tasks are still to be solved, e.g. face recognition in an unconstrained and uncontrolled environment (varying pose, illumination and expression, a cluttered background, occlusion), recognition of non-frontal facial images, the role of the face in multimodal biometric systems, real-time operation, one sample problem, 3D recognition, face recognition in video; that is why many researchers study face biometric extensively.

This book aims to bring together selected recent advances, applications and original results in the area of biometric face recognition. They can be useful for researchers, engineers, graduate and postgraduate students, experts in this area and hopefully also for people interested generally in computer science, security, machine learning and artificial intelligence.

Various methods, approaches and algorithms for recognition of human faces are used by authors of the chapters of this book, e.g. PCA, LDA, artificial neural networks, wavelets, curvelets, kernel methods, Gabor filters, active appearance models, 2D and 3D representations, optical correlation, hidden Markov models and others. Also a broad range of problems is covered: feature extraction and dimensionality reduction (chapters 1-4), 2D face recognition from the point of view of full system proposal (chapters 5-10), illumination and pose problems (chapters 11-13), eye movement (chapter 14), 3D face recognition (chapters 15-19) and hardware issues (chapters 19-20).

Chapter 1 reviews the most relevant feature extraction techniques (both holistic and local feature) used in 2D face recognition and also introduces a new feature extraction technique. Chapter 2 presents the n-dimensional extension of PCA, which solves numerical difficulties and provides near optimal linear classification property. Chapter 3 is devoted to curvelets; authors concentrate on fast digital curvelet transform. In chapter 4, a dimensionality reduction method based on random projection is proposed and compressive classification algorithms that are robust to random projection dimensionality reduction are reviewed.

In chapter 5, the author presents a modular system for face recognition including a method that can suppress unwanted features and make useful decisions on similarity irrespective of the complex nature of the underlying data. Chapter 6 presents discussion of appearance-based methods vs. local description methods and the proposal of a novel face recognition system based on the use of interest point detectors and local descriptors. Chapter 7 focuses on wavelet-based face recognition schemes and presents their performance using a number of benchmark databases of face images and videos. Chapter 8 presents a complex view on the proposal of a biometric face recognition system including methodology, settings of parameters and the influence of input image quality on face recognition accuracy. In chapter 9, authors propose a face recognition system built as a cascade connection of an artificial neural network and pseudo 2D hidden Markov models. In chapter 10, an experimental evaluation of the performance of VG-RAM weightless neural networks for face recognition using well-known face databases is presented.

Chapter 11 addresses the problem of illumination in face recognition including mathematical illumination modeling, influence of illumination on recognition results and the current state-of-art of illumination processing and its future trends. Chapter 12 brings the proposal of a novel face representation based on phase responses of the Gabor filter bank which is characterized by its robustness to illumination changes. Chapter 13 presents illumination and pose-invariant face alignment based on an active appearance model.

Chapter 14 reviews current literature about eye movements in face recognition and provides answers to several questions relevant to this topic.

Chapter 15 gives an overview of surface representations for 3D face recognition; also surface representations promising in terms of future research that have not yet been reported in current face recognition literature are discussed. Chapter 16 presents framework for 3D face and expression recognition taking into account the fact that the deformation of the face surface is always related to different expressions. Chapter 17 addresses security leakages and privacy protection issues in biometric systems and presents latest results of template protection techniques in 3D face recognition systems. Chapter 18 presents a 3D face recognition system based on pseudo 2D hidden Markov models using an expression-invariant representation of faces. Chapter 19 covers some of the latest developments in optical correlation techniques for face recognition using the concept of spectral fusion; also a new concept of correlation filter called segmented composite filter is employed that is suitable for 3D face recognition.

Chapter 20 presents an implementation of the Neocognitron neural network using a high-performance computing architecture based on a graphics processing unit.

The editor owes special thanks to authors of all included chapters for their valuable work.

April 2010

Miloš Oravec
Slovak University of Technology
Faculty of Electrical Engineering and Information Technology
Department of Applied Informatics and Information Technology
Ilkovičova 3, 812 19 Bratislava, Slovak Republic
e-mail: milos.oravec@stuba.sk

1

Feature Extraction and Representation for Face Recognition

[1]M. Saquib Sarfraz, [2]Olaf Hellwich and [3]Zahid Riaz
[1]*Computer Vision Research Group, Department of Electrical Engineering*
COMSATS Institute of Information Technology, Lahore
Pakistan
[2]*Computer Vision and Remote Sensing, Berlin University of Technology*
Sekr. FR 3-1, Franklin str. 28/29, 10587, Berlin
Germany
[3]*Institute of Informatik, Technical University Munich*
Germany

1. Introduction

Over the past two decades several attempts have been made to address the problem of face recognition and a voluminous literature has been produced. Current face recognition systems are able to perform very well in controlled environments e.g. frontal face recognition, where face images are acquired under frontal pose with strict constraints as defined in related face recognition standards. However, in unconstrained situations where a face may be captured in outdoor environments, under arbitrary illumination and large pose variations these systems fail to work. With the current focus of research to deal with these problems, much attention has been devoted in the facial feature extraction stage. Facial feature extraction is the most important step in face recognition. Several studies have been made to answer the questions like what features to use, how to describe them and several feature extraction techniques have been proposed. While many comprehensive literature reviews exist for face recognition a complete reference for different feature extraction techniques and their advantages/disadvantages with regards to a typical face recognition task in unconstrained scenarios is much needed.

In this chapter we present a comprehensive review of the most relevant feature extraction techniques used in 2D face recognition and introduce a new feature extraction technique termed as Face-GLOH-signature to be used in face recognition for the first time (Sarfraz and Hellwich, 2008), which has a number of advantages over the commonly used feature descriptions in the context of unconstrained face recognition.

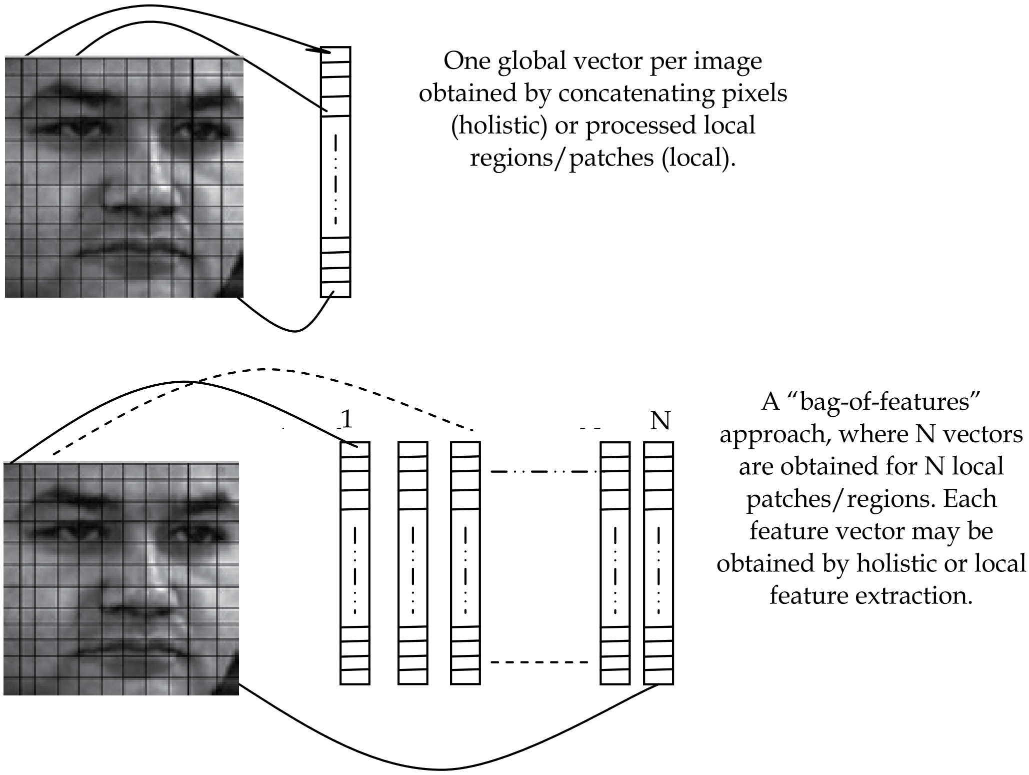
The goal of feature extraction is to find a specific representation of the data that can highlight relevant information. This representation can be found by maximizing a criterion or can be a pre-defined representation. Usually, a face image is represented by a high dimensional vector containing pixel values (holistic representation) or a set of vectors where each vector summarizes the underlying content of a local region by using a high level transformation (local representation). In this chapter we made distinction in the holistic and local feature extraction and differentiate them qualitatively as opposed to quantitatively. It is argued that a global feature representation based on local feature analysis should be preferred over a bag-of-feature approach. The problems in current feature extraction techniques and their reliance on a strict alignment is discussed. Finally we introduce to use face-GLOH signatures that are invariant with respect to scale, translation and rotation and therefore do not require properly aligned images. The resulting dimensionality of the vector is also low as compared to other commonly used local features such as Gabor, Local Binary Pattern Histogram 'LBP' etc. and therefore learning based methods can also benefit from it.

A performance comparison of face-GLOH-Signature with different feature extraction techniques in a typical face recognition task is presented using FERET database. To highlight the usefulness of the proposed features in unconstrained scenarios, we study and compare the performance both under a typical template matching scheme and learning based methods (using different classifiers) with respect to the factors like, large number of subjects, large pose variations and misalignments due to detection errors. The results demonstrate the effectiveness and weakness of proposed and existing feature extraction techniques.

2. Holistic Vs Local Features-What Features to Use?

Holistic representation is the most typical to be used in face recognition. It is based on lexicographic ordering of raw pixel values to yield one vector per image. An image can now be seen as a point in a high dimensional feature space. The dimensionality corresponds directly to the size of the image in terms of pixels. Therefore, an image of size 100x100 pixels can be seen as a point in a 10,000 dimensional feature space. This large dimensionality of the problem prohibits the use of any learning to be carried out in such a high dimensional feature space. This is called the curse of dimensionality in the pattern recognition literature (Duda et al, 2001). A common way of dealing with it is to employ a dimensionality reduction technique such as Principal Component Analysis 'PCA' to pose the problem into a low-dimensional feature space such that the major modes of variation of the data are still preserved.

Local feature extraction refers to describing only a local region/part of the image by using some transformation rule or specific measurements such that the final result describes the underlying image content in a manner that should yield a unique solution whenever the same content is encountered. In doing so, however it is also required to have some degree of invariance with respect to commonly encountered variations such as translation, scale and rotations. A number of authors (Pentland et al, 1994; Cardinaux et al, 2006; Zou et al, 2007) do not differentiate the holistic and local approaches according to the very nature they are obtained, but rather use the terms in lieu of global (having one feature vector per image) and a bag-of-feature (having several feature vectors per image) respectively. Here we want to put the both terms into their right context, and hence a holistic representation can be obtained for several local regions of the image and similarly a local representation can still be obtained by concatenating several locally processed regions of the image into one global vector, see figure 1 for an illustration. An example of the first usage is local-PCA or modular- PCA (Gottumukkal and Asari, 2004; Tan and Chen, 2005), where an image is divided into several parts or regions, and each region is then described by a vector



One global vector per image obtained by concatenating pixels (holistic) or processed local regions/patches (local).

A "bag-of-features" approach, where N vectors are obtained for N local patches/regions. Each feature vector may be obtained by holistic or local feature extraction.

Fig. 1. Global and bag-of-feature representation for a facial image

comprising underlying raw-pixel values, PCA is then employed to reduce the dimensionality. Note that it is called local since it uses several local patches of the same image but it is still holistic in nature. An example of the second is what usually found in the literature, e.g. Gabor filtering, Discrete Cosine Transform 'DCT', Local Binary Pattern 'LBP' etc where each pixel or local region of the image is described by a vector and concatenated into a global description (Zou et al, 2007), note that they still give rise to one vector per image but they are called local in the literature because they summarize the local content of the image at a location in a way that is invariant with respect to some intrinsic image properties e.g. scale, translation and/or rotation.

Keeping in view the above discussion it is common in face recognition to either follow a global feature extraction or a bag-of-features approach. The choice, of what is optimal, depends on the final application in mind and hence is not trivial. However, there are a number of advantages and disadvantages with both the approaches. For instance, a global description is generally preferred for face recognition since it preserves the configural (i.e., the interrelations between facial parts) information of the face, which is very important for preserving the identity of the individual as have been evidenced both from psychological (Marta et al, 2006), neurobiological (Schwaninger et al, 2006; Hayward et al, 2008) and computer vision (Belhumeur et al, 1997; Chen et al, 2001) communities. On the other hand, a bag-of-features approach has been taken by a number of authors (Brunelli and Poggio, 1993; Martınez, 2002; Kanade and Yamada, 2003) and shown improved recognition results in the presence of occlusion etc., nonetheless, in doing so, these approaches are bound to preserve the configural information of the facial parts either implicitly or explicitly by comparing only the corresponding parts in two images and hence puts a hard demand on the requirement of proper and precise alignment of facial images.

Note that while occlusion may be the one strong reason to consider a bag-of-features approach, the tendency of preserving the spatial arrangement of different facial parts (configural information) is largely compromised. As evidenced from the many studies from interdisciplinary fields that this spatial arrangement is in fact quite crucial in order to preserve the identity of an individual, we therefore, advocate the use of a global representation for a face image in this dissertation, as has also been used by many others.

One may, however, note that a global representation does not necessarily mean a holistic representation, as described before. In fact, for the automatic unconstrained face recognition, where there may be much variation in terms of scale, lighting, misalignments etc, the choice of using local feature extraction becomes imperative since holistic representation cannot generalize in these scenarios and is known to be highly affected by these in-class variations.

3. Holistic Feature Extraction

Holistic feature extraction is the most widely used feature description technique in appearance based face recognition methods. Despite its poor generalization abilities in unconstrained scenarios, it is being used for the main reason that any local extraction technique is a form of information reduction in that it typically finds a transformation that describes a large data by few numbers. Since from a strict general object recognition stand point, face is one class of objects, and thus discriminating within this class puts very high demands in finding subtle details of an image that discriminates among different faces. Therefore each pixel of an image is considered valuable information and holistic processing develops. However, a holistic-based global representation as been used classically (Turk and Pentland, 1991) cannot perform well and therefore more recently many researchers used a bag-of-features approach, where each block or image patch is described by holistic representation and the deformation of each patch is modeled for each face class (Kanade and Yamada, 2003; Lucey and Chen, 2006; Ashraf et al, 2008).

3.1 Eigenface- A global representation

Given a face image matrix *F* of size *Y* x *X*, a vector representation is constructed by concatenating all the columns of *F* to form a column vector $\vec{f}$ of dimensionality *YX*. Given a set of training vectors $\{\vec{f}_i\}_{i=1}^{Np}$ for all persons, a new set of mean subtracted vectors is formed using:

$$g_i = \vec{f}_i - \vec{f}_\mu, \qquad i = 1, 2, ..., N_p \tag{1}$$

The mean subtracted training set is represented as a matrix $G = [\vec{g}_1, \vec{g}_2, ..., \vec{g}_{Np}]$. The covariance matrix is then calculated using, $\sum = GG^T$. Due to the size of $\sum$, calculation of the eigenvectors of $\sum$ can be computationally infeasible. However, if the number of training vectors (N_p) is less than their dimensionality (*YX*), there will be only N_p-1 meaningful eigenvectors. (Turk and Pentland, 91) exploit this fact to determine the eigenvectors using an alternative method summarized as follows. Let us denote the eigenvectors of matrix G^TG as $\vec{v}_j$ with corresponding eigenvalues Λ_j:

$$G^T G\vec{v}_j = \Lambda_j \vec{v}_j \tag{2}$$

Pre-multiplying both sides by G gives us: $GG^T G\vec{v}_j = \Lambda_j G\vec{v}_j$, Letting $\vec{e}_j = G\vec{v}$ and substituting for $\sum$ from equation 1:

$$\sum \vec{e}_j = \Lambda_j \vec{e}_j \tag{3}$$

Hence the eigenvectors of $\sum$ can be found by pre-multiplying the eigenvectors of G^TG by G. To achieve dimensionality reduction, let us construct matrix $E = [\vec{e}_1, \vec{e}_1, ..., \vec{e}_D]$, containing D eigenvectors of $\sum$ with largest corresponding eigenvalues. Here, D<N_p, a feature vector $\vec{x}$ of dimensionality D *is* then derived from a face vector $\vec{f}$ using:

$$\vec{x} = E^T(\vec{f} - \vec{f}_\mu) \tag{4}$$

Therefore, a face vector $\vec{f}$ is decomposed into D eigenvectors, known as eigenfaces. Similarly, employing the above mentioned Eigen analysis to each local patch of the image results into a bag-of-features approach. Pentland *et al.* extended the eigenface technique to a layered representation by combining eigenfaces and other eigenmodules, such as eigeneyes, eigennoses, and eigenmouths(Pentland et al, 1994). Recognition is then performed by finding a projection of the test image patch to each of the learned local Eigen subspaces for every individual.

4. Local Feature Extraction

(Gottumukkal and Asari, 2004) argued that some of the local facial features did not vary with pose, direction of lighting and facial expression and, therefore, suggested dividing the face region into smaller sub images. The goal of local feature extraction thus becomes to represent these local regions effectively and comprehensively. Here we review the most commonly used local feature extraction techniques in face recognition namely the Gabor wavelet transform based features , discrete cosine transform DCT-based features and more recently proposed Local binary pattern LBP features.

4.1 2D Gabor wavelets

The 2D Gabor elementary function was first introduced by Granlund (Granlund, 1978). Gabor wavelets demonstrate two desirable characteristic: spatial locality and orientation selectivity. The structure and functions of Gabor kernels are similar to the two-dimensional receptive fields of the mammalian cortical simple cells (Hubel and Wiesel, 1978). (Olshausen and Field, 1996; Rao and Ballard, 1995; Schiele and Crowley, 2000) indicates that the Gabor wavelet representation of face images should be robust to variations due to illumination and facial expression changes. Two-dimensional Gabor wavelets were first introduced into biometric research by Daugman (Daugman, 1993) for human iris recognition. Lades et al. (Lades et al, 1993) first apply Gabor wavelets for face recognition using the Dynamic Link Architecture framework.

A Gabor wavelet kernel can be thought of a product of a complex sinusoid plane wave with a Gaussian envelop. A Gabor wavelet generally used in face recognition is defined as (Liu, 2004):

$$\psi_{u,v}(z) = \frac{\|k_{u,v}\|^2}{\sigma^2} e^{-\frac{\|k_{u,v}\|^2 \|z\|^2}{2\sigma^2}} [e^{ik_{u,v}z} - e^{-\frac{\sigma^2}{2}}] \tag{5}$$

where z = (x, y) is the point with the horizontal coordinate x and the vertical coordinate y in the image plane. The parameters u and v define the orientation and frequency of the Gabor kernel, $\|.\|$ denotes the norm operator, and σ is related to the standard derivation of the Gaussian window in the kernel and determines the ratio of the Gaussian window width to the wavelength. The wave vector $k_{\mu,v}$ is defined as $k_{u,v} = k_v e^{i\phi_u}$.

Following the parameters suggested in (Lades et al, 1993) and used widely in prior works (Liu, 2004) (Liu and Wechsler, 2002) $k_v = \frac{k_{\max}}{f^v}$ and $\phi_u = \frac{\pi u}{8}$. k$_{\max}$ is the maximum frequency, and f^v is the spatial frequency between kernels in the frequency domain. $v \in \{0,...,4\}$ and $u \in \{0,...,7\}$ in order to have a Gabor kernel tuned to 5 scales and 8 orientations. Gabor wavelets are chosen relative to $\sigma = 2\pi$, $k_{\max} = \frac{\pi}{2}$ and $f = \sqrt{2}$. The parameters ensures that frequencies are spaced in octave steps from 0 to π, typically each Gabor wavelet has a frequency bandwidth of one octave that is sufficient to have less overlap and cover the whole spectrum.

The Gabor wavelet representation of an image is the convolution of the image with a family of Gabor kernels as defined by equation (6). The convolution of image *I* and a Gabor kernel $\psi_{u,v}(z)$ is defined as follows:

$$G_{u,v}(z) = I(z) * \psi_{u,v}(z) \tag{6}$$

where $z = (x, y)$ denotes the image position, the symbol $'*'$ denotes the convolution operator, and $G_{u,v}(z)$ is the convolution result corresponding to the Gabor kernel at scale v and orientation u . The Gabor wavelet coefficient is a complex with a real and imaginary part, which can be rewritten as $G_{u,v}(z) = A_{u,v}(z).\mathrm{e}^{i\theta_{u,v}(z)}$, where $A_{u,v}$ is the magnitude response and $\theta_{u,v}$ is the phase of Gabor kernel at each image position. It is known that the magnitude varies slowly with the spatial position, while the phases rotate in some rate with positions, as can be seen from the example in figure 2. Due to this rotation, the phases taken from image points only a few pixels apart have very different values, although representing almost the same local feature (Wiskott et al, 1997). This can cause severe problems for face

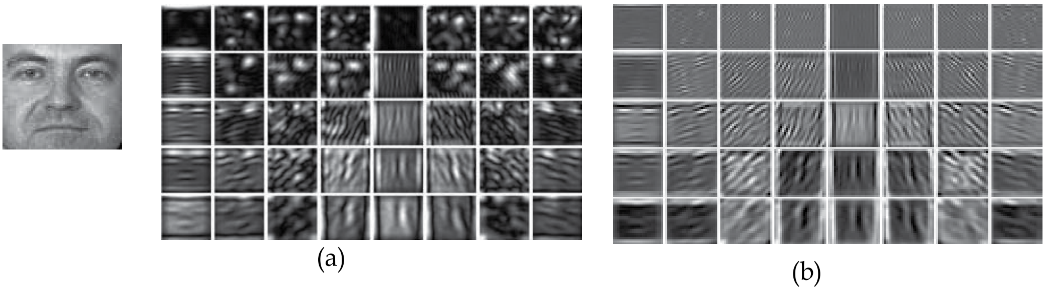


Fig. 2. Visualization of (a) Gabor magnitude (b) Gabor phase response, for a face image with 40 Gabor wavelets (5 scales and 8 orientations).

matching, and it is just the reason that all most all of the previous works make use of only the magnitude part for face recognition. Note that, convolving an image with a bank of Gabor kernel tuned to 5 scales and 8 orientations results in 40 magnitude and phase response maps of the same size as image. Therefore, considering only the magnitude response for the purpose of feature description, each pixel can be now described by a 40 dimensional feature vector (by concatenating all the response values at each scale and orientation) describing the response of Gabor filtering at that location.

Note that Gabor feature extraction results in a highly localized and over complete response at each image location. In order to describe a whole face image by Gabor feature description the earlier methods take into account the response only at certain image locations, e.g. by placing a coarse rectangular grid over the image and taking the response only at the nodes of the grid (Lades et al, 1993) or just considering the points at important facial landmarks as in (Wiskott et al, 1997). The recognition is then performed by directly comparing the corresponding points in two images. This is done for the main reason of putting an upper limit on the dimensionality of the problem. However, in doing so they implicitly assume a perfect alignment between all the facial images, and moreover the selected points that needs to be compared have to be detected with pixel accuracy.

One way of relaxing the constraint of detecting landmarks with pixel accuracy is to describe the image by a global feature vector either by concatenating all the pixel responses into one long vector or employ a feature selection mechanism to only include significant points (Wu and Yoshida, 2002) (Liu et al, 2004). One global vector per image results in a very high and prohibitive dimensional problem, since e.g. a 100x100 image would result in a 40x100x100=400000 dimensional feature vector. Some authors used Kernel PCA to reduce this dimensionality termed as Gabor-KPCA (Liu, 2004), and others (Wu and Yoshida, 2002; Liu et al, 2004; Wang et al, 2002) employ a feature selection mechanism for selecting only the important points by using some automated methods such as Adaboost etc. Nonetheless, a global description in this case still results in a very high dimensional feature vector, e.g. in (Wang et al, 2002) authors selected only 32 points in an image of size 64x64, which results in 32x40=1280 dimensional vector, due to this high dimensionality the recognition is usually performed by computing directly a distance measure or similarity metric between two images. The other way can be of taking a bag-of-feature approach where each selected point is considered an independent feature, but in this case the configural information of the face is effectively lost and as such it cannot be applied directly in situations where a large pose variations and other appearance variations are expected.

The Gabor based feature description of faces although have shown superior results in terms of recognition, however we note that this is only the case when frontal or near frontal facial images are considered. Due to the problems associated with the large dimensionality, and thus the requirement of feature selection, it cannot be applied directly in scenarios where large pose variations are present.

4.2 2D Discrete Cosine Transform

Another popular feature extraction technique has been to decompose the image on block by block basis and describe each block by 2D Discrete Cosine Transform 'DCT' coefficients. An image block $f(p,q)$, where $p,q = \{0,1..,\mathrm{N}-1\}$ (typically N=8), is decomposed terms of orthogonal 2D DCT basis functions. The result is a NxN matrix *C*(*v*,*u*) containing 2D DCT coefficients:

$$C(v,u) = \alpha(v)\alpha(u)\sum_{y=0}^{N-1}\sum_{x=0}^{N-1} f(p,q)\beta(p,q,v,u) \tag{7}$$

where $v,u = 0,1,2,...,N-1$, $\alpha(v) = \sqrt{\frac{1}{N}}$ for v=0, and $\alpha(v) = \sqrt{\frac{2}{N}}$ for v=1,2,…,N-1 and

$$\beta(p,q,v,u) = \cos\left[\frac{(2p+1)v\pi}{2N}\right]\cos\left[\frac{(2q+1)u\pi}{2N}\right] \tag{8}$$

The coefficients are ordered according to a zig-zag pattern, reflecting the amount of information stored (Gonzales and Woods, 1993). For a block located at image position (*x*,*y*), the baseline 2D DCT feature vector is composed of:

$$x = [c_o^{(x,y)} \quad c_1^{(x,y)} \ldots \quad c_{M-1}^{(x,y)}]^T \tag{9}$$

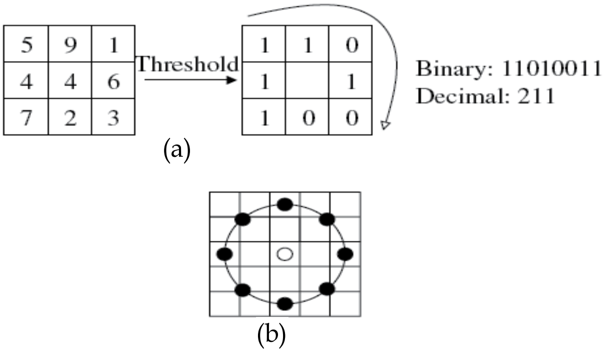
Where $c_n^{(x,y)}$ denotes the *n-th* 2D DCT coefficient and *M* is the number of retained coefficients3. To ensure adequate representation of the image, each block overlaps its horizontally and vertically neighbouring blocks by 50% (Eickeler et al, 2000). M is typically set to 15 therefore each block yields a 15 dimensional feature vector. Thus for an image which has Y rows and X columns, there are $N_D = (2\frac{Y}{N}-1)\times(2\frac{X}{N}-1)$ blocks.

DCT based features have mainly been used in Hidden Markov Models HMM based methods in frontal scenarios. More recently (Cardinaux et al, 2006) proposed an extension of conventional DCT based features by replacing the first 3 coefficients with their corresponding horizontal and vertical deltas termed as DCTmod2, resulting into an 18-dimensional feature vector for each block. The authors claimed that this way the feature vectors are less affected by illumination change. They then use a bag-of-feature approach to derive person specific face models by using Gaussian mixture models.

4.2 Local Binary Pattern Histogram LBPH and its variants

Local binary pattern (LBP) was originally designed for texture classification (Ojala et al, 2002), and was introduced in face recognition in (Ahonen et al, 2004). As mentioned in

(Ahonen et al, 2004) the operator labels the pixels of an image by thresholding some neighbourhood of each pixel with the centre value and considering the result as a binary number. Then the histogram of the labels can be used as a texture descriptor. See figure 3 for an illustration of the basic $LBP_{P,R}^{U2}$ operator. The face area is divided into several small windows. Several LBP operators are compared and $LBP_{8,2}^{U2}$ the operator in 18x21 pixel windows is recommended because it is a good trade-off between recognition performance and feature vector length. The subscript represents using the operator in a (P, R) neighbourhood. Superscript U2 represent using only uniform patterns and labelling all remaining patterns with a single label, see (Ahonen et al, 2004) for details. The χ^2 statistic and the weighted χ^2 statistic were adopted to compare local binary pattern histograms.



(a)

(b)

Fig. 3. (a) the basic LBP operator. (b) The circular (8,2) neighbourhood. The pixel values are bilinearly interpolated whenever the sampling point is not in the centre of a pixel (Ahonen et al, 2004)

Recently (Zhang et al, 2005) proposed local Gabor binary pattern histogram sequence (LGBPHS) by combining Gabor filters and the local binary operator. (Baochang et al, 2007) further used LBP to encode Gabor filter phase response into an image histogram termed as Histogram of Gabor Phase Patterns (HGPP).

5. Face-GLOH-Signatures –introduced feature representation

The mostly used local feature extraction and representation schemes presented in previous section have mainly been employed in a frontal face recognition task. Their ability to perform equally well when a significant pose variation is present among images of the same person cannot be guaranteed, especially when no alignment is assumed among facial images. This is because when these feature representations are used as a global description the necessity of having a precise alignment becomes unavoidable. While representations like 2D-DCT or LBP are much more susceptible to noise, e.g. due to illumination change as noted in (Zou et al, 2007) or pose variations, Gabor based features are considered to be more invariant with respect to these variations. However, as discussed earlier the global Gabor representation results in a prohibitively high dimensional problem and as such cannot be directly used in statistical based methods to model these in-class variations due to pose for instance. Moreover the effect of misalignments on Gabor features has been studied

(Shiguang et al, 2004), where strong performance degradation is observed for different face recognition systems.

As to the question, what description to use, there are some guidelines one can benefit from. For example, as discussed in section 3.1 the configural relationship of the face has to be preserved. Therefore a global representation as opposed to a bag-of-features approach should be preferred. Further in order to account for the in-class variations the local regions of the image should be processed in a scale, rotation and translation invariant manner. Another important consideration should be with respect to the size of the local region used. Some recent studies (Martınez, 2002; Ullman et al, 2002; Zhang et al, 2005) show that large areas should be preferred in order to preserve the identity in face identification scenarios.

Keeping in view the preceding discussion we use features proposed in (Mikolajczyk and Schmid, 2005), used in other object recognition tasks, and introduce to employ these for the task of face recognition for the first time (Sarfraz, 2008; Sarfraz and Hellwich, 2008) Our approach is to extract whole appearance of the face in a manner which is robust against misalignments. For this the feature description is specifically adapted for the purpose of face recognition. It models the local parts of the face and combines them into a global description

We use a representation based on gradient location-orientation histogram (GLOH) (Mikolajczyk and Schmid, 2005), which is more sophisticated and is specifically designed to reduce in-class variance by providing some degree of invariance to the aforementioned transformations.

GLOH features are an extension to the descriptors used in the scale invariant feature transform (SIFT) (Lowe, 2004), and have been reported to outperform other types of descriptors in object recognition tasks (Mikolajczyk and Schmid, 2005). Like SIFT the GLOH descriptor is a 3D histogram of gradient location and orientation, where location is quantized into a log-polar location grid and the gradient angle is quantized into eight orientations. Each orientation plane represents the gradient magnitude corresponding to a given orientation. To obtain illumination invariance, the descriptor is normalized by the square root of the sum of squared components.

Originally (Mikolajczyk and Schmid, 2005) used the log-polar location grid with three bins in radial direction (the radius set to 6, 11, and 15) and 8 in angular direction, which results in 17 location bins. The gradient orientations are quantized in 16 bins. This gives a 272 bin histogram. The size of this descriptor is reduced with PCA. While here the extraction procedure has been specifically adapted to the task of face recognition and is described in the remainder of this section.

The extraction process begins with the computation of scale adaptive spatial gradients for a given image *I(x,y)*. These gradients are given by:

$$\nabla_{xy} \equiv \sum_{t} w(x,y,t)\sqrt{t}\nabla^{t}_{xy} L(x,y;t) \tag{10}$$

where *L(x,y; t)* denotes the linear Gaussian scale space of *I(x,y)* (Lindeberg, 1998) and *w(x,y,t)* is a weighting, as given in equation 11.

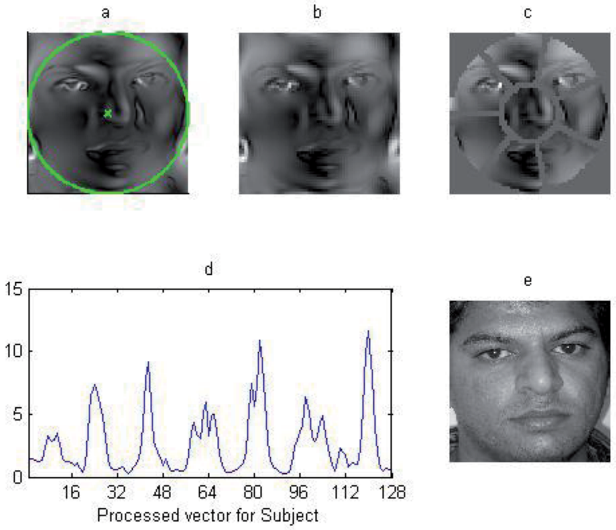


Fig. 5. Face-GLOH-Signature extraction (a-b) Gradient magnitudes (c) polar-grid partitions (d) 128-dimentional feature vector (e) Example image of a subject.

$$w(x,y,t)=\frac{\left|\sqrt{t}\nabla^{t}_{xy}L(x,y;t)\right|^{4}}{\sum\limits_{t}\left|\sqrt{t}\nabla^{t}_{xy}L(x,y;t)\right|^{4}} \tag{11}$$

The gradient magnitudes obtained for an example face image (Figure 5 e) are shown in Figure 5 b. The gradient image is then partitioned on a grid in polar coordinates, as illustrated in Figure 5 c. As opposed to the original descriptor the partitions include a central region and seven radial sectors. The radius of the central region is chosen to make the areas of all partitions equal. Each partition is then processed to yield a histogram of gradient magnitude over gradient orientations. The histogram for each partition has 16 bins corresponding to orientations between 0 and 2п, and all histograms are concatenated to give the final 128 dimensional feature vector, that we term as face-GLOH-signature, see Figure 5 d. No PCA is performed in order to reduce the dimensionality.

The dimensionality of the feature vector depends on the number of partitions used. A higher number of partitions results in a longer vector and vice versa. The choice has to be made with respect to some experimental evidence and the effect on the recognition performance. We have assessed the recognition performance on a validation set by using ORL face database. By varying the partitions sizes from 3 (1 central region and 2 sectors), 5, 8, 12 and 17, we found that increasing number of partitions results in degrading performance especially with respect to misalignments while using coarse partitions also affects recognition performance with more pose variations. Based on the results, 8 partitions seem to be the optimal choice and a good trade off between achieving better recognition performance and minimizing the effect of misalignment. The efficacy of the descriptor is demonstrated in the presence of pose variations and misalignments, in the next section. It should be noted that, in practice, the quality of the descriptor improves when care is taken to minimize aliasing artefacts. The recommended measures include the use of smooth partition boundaries as well as a soft assignment of gradient vectors to orientation histogram bins.

6. Performance Analysis

In order to assess the performance of introduced face-GLOH-signature with that of various feature representations, we perform experiments in two settings. In the first setting, the problem is posed as a typical multi-view recognition scenario, where we assume that few number of example images of each subject are available for training. Note that, global feature representations based on Gabor, LBP and DCT cannot be directly evaluated in this setting because of the associated very high dimensional feature space. These representations are, therefore, evaluated in a typical template matching fashion in the second experimental setting, where we assess the performance of each representation across a number of pose mismatches by using a simple similarity metric. Experiments are performed on two of the well-known face databases i.e. FERET (Philips et al, 2000) and ORL face database (http://www.cam-orl.co.uk).

6.1 Multi-view Face recognition

In order to perform multi-view face recognition (recognizing faces under different poses) it is generally assumed to have examples of each person in different poses available for training. The problem is solved form a typical machine learning point of view where each person defines one class. A classifier is then trained that seek to separate each class by a decision boundary. Multi-view face recognition can be seen as a direct extension of frontal face recognition in which the algorithms require gallery images of every subject at every pose (Beymer, 1996). In this context, to handle the problem of one training example, recent research direction has been to use specialized synthesis techniques to generate a given face at all other views and then perform conventional multi-view recognition (Lee and Kim, 2006; Gross et al, 2004). Here we focus on studying the effects on classification performance when a proper alignment is not assumed and there exist large pose differences. With these goals in mind, the generalization ability of different conventional classifiers is evaluated with respect to the small sample size problem. Small sample size problem stems from the fact that face recognition typically involves thousands of persons in the database to be recognized. Since multi-view recognition treats each person as a separate class and tends to solve the problem as a multi-class problem, it typically has thousands of classes. From a machine learning point of view any classifier trying to learn thousands of classes requires a good amount of training data available for each class in order to generalize well. Practically, we have only a small number of examples per subject available for training and therefore more and more emphasis is given on choosing a classifier that has good generalization ability in such sparse domain.

The other major issue that affects the classification is the representation of the data. The most commonly used feature representations in face recognition have been introduced in previous sections. Among these the Eigenface by using PCA is the most common to be used in multi-view face recognition. The reason for that is the associated high dimensionality of other feature descriptions such as Gabor, LBPH etc. that prohibits the use of any learning to

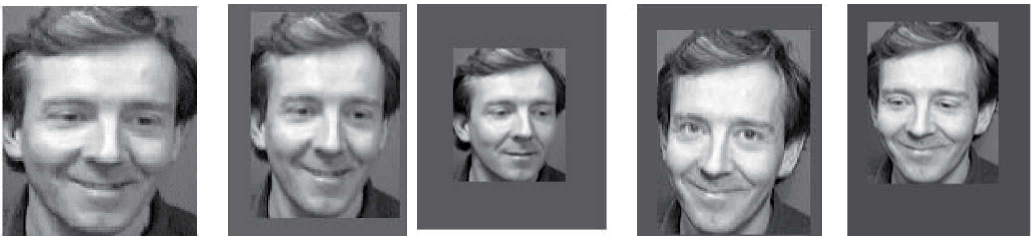


Fig. 6. An example of a subject from O-ORL and its scale and shifted examples from SS-ORL

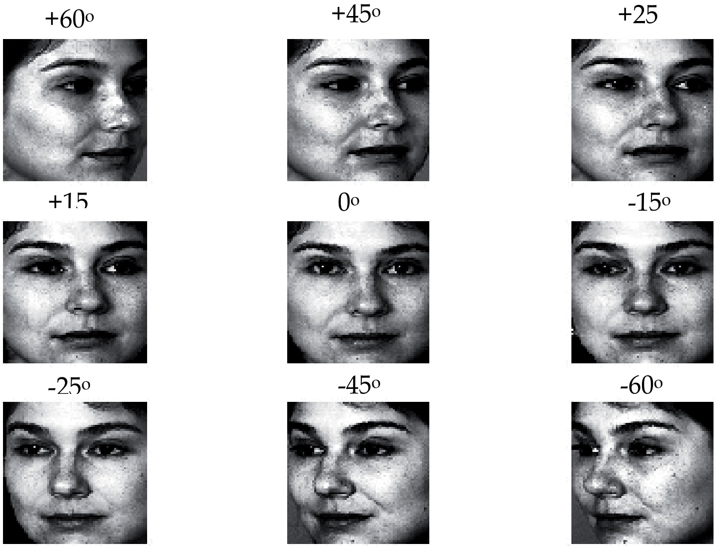


Fig. 7. Cropped faces of a FERET subject depicting all the 9 pose variations.

be done. This is the well known curse of dimensionality issue in pattern recognition (Duda et al, 2001) literature and this is just the reason that methods using such over complete representations normally resort to performing a simple similarity search by computing distances of a probe image to each of the gallery image in a typical template matching manner. While by using PCA on image pixels an upper bound on the dimensionality can be achieved.

In line with the above discussion, we therefore demonstrate the effectiveness of the proposed face-GLOH signatures with that of using conventional PCA based features in multi-view face recognition scenarios with respect to the following factors.

When facial images are not artificially aligned

When there are large pose differences

Large number of subjects

Number of examples available in each class (subject) for training.

In order to show the effectiveness of face-GLOH signature feature representation against misalignments, we use ORL face database. ORL face database has 400 images of 40 subjects (10 images per subject) depicting moderate variations among images of same person due to expression and some limited pose. Each image in ORL has the dimension of 192x112 pixels.

All the images are depicted in approximately the same scale and thus have a strong correspondence among facial regions across images of the same subject. We therefore generate a scaled and shifted ORL dataset by introducing an arbitrary scale change between 0.7 and 1.2 of the original scale as well as an arbitrary shift of 3 pixels in random direction in each example image of each subject. This has been done to ensure having no artificial alignment between corresponding facial parts. This new misaligned dataset is denoted scaled-shifted SS-ORL (see Figure 6). The experiments are performed on both the original ORL denoted O-ORL and SS-ORL using PCA based features and face-GLOH signatures. ORL face database is mainly used to study the effects on classification performance due to misalignments since variations due to pose are rather restricted (not more than 20^o). To study the effects of large pose variations and a large number of subjects, we therefore repeat our experiments on FERET database pose subset. The FERET pose subset contains 200 subjects, where each subject has nine images corresponding to different pose angles (varying from 0^o frontal to left/right profile $\pm 60^o$) with an average pose difference of 15^o. All the images are cropped from the database by using standard normalization methods i.e. by manually locating eyes position and warping the image onto a plane where these points are in a fixed location. The FERET images are therefore aligned with respect to these points. This is done in order to only study the effects on classifier performance due to large pose deviations. All the images are then resized to 92x112 pixels in order to have the same size as that of ORL faces. An example of the processed images of a FERET subject depicting all the 9 pose variations is shown in Figure 7.

We evaluate eight different conventional classifiers. These include nearest mean classifier 'NMC', linear discriminant classifier 'LDC', quadratic 'QDC', fisher discriminant, parzen classifier, k-nearest neighbour 'KNN', Decision tree and support vector machine 'SVM', see (Webb, 2002) for a review of these classifiers.

6.1.1 Experiments on ORL database

We extract one global feature vector per face image by using lexicographic ordering of all the pixel grey values. Thus, for each 92 x 112 ORL image, one obtains a 10384 dimensional feature vector per face. We then reduce this dimensionality by using unsupervised PCA. Where the covariance matrix is trained using 450 images of 50 subjects from FERET set. The number of projection Eigen-vectors are found by analysing the relative cumulative ordered eigenvalues (sum of normalized variance) of the covariance matrix. We choose first 50 largest Eigen vectors that explain around 80% of the variance as shown in figure 4-3. By projecting the images on these, we therefore obtain a 50-dimentional feature vector for each image. We call this representation the PCA-set.

The second representation of all the images is found by using face-GLOH-signature extraction, as detailed in section 5.

In all of our experiments we assume equal priors for training, SVM experiments on O-ORL use a polynomial kernel of degree 2, to reduce the computational effort, since using RBF kernel with optimized parameters C and kernel width σ did not improve performance. For SS-ORL a RBF kernel is used with parameter C=500 and σ = 10, these values were determined using 5-fold cross validation and varying sigma between 0.1 and 50 and C between 1 and 1000. All the experiments are carried out for classifiers on each of two representations for both O-ORL and SS-ORL.

We use a 10-fold cross validation procedure to produces 10 sets of the same size as original dataset each with a different 10% of objects being used for testing. All classifiers are evaluated on each set and the classification errors are averaged. The results from this experiment on both O- ORL and SS-ORL for both feature representations are reported in table 1.

Classifiers	O-ORL Representation sets		SS-ORL Representation sets	
	PCA	face-GLO H	PCA	face-GLOH
NMC	0.137	0.152	0.375	0.305
LDC	0.065	0.020	0.257	0.125
Fisher	0.267	0.045	0.587	0.115
Parzen	0.037	0.030	0.292	0.162
3-NN	0.097	0.062	0.357	0.255
Decision Tree	0.577	0.787	0.915	0.822
QDC	0.64	0.925	0.760	0.986
SVM	0.047	0.037	0.242	0.105

Table 1. Classification errors in 10-fold cross validation tests on ORL

Table 1 shows how classification performance degrades, when the faces are not aligned i.e. arbitrarily scaled and shifted, on PCA based feature representation. The robustness of the face-GLOH-signature representation against misalignments can be seen by comparing the results on O-ORL and SS-ORL, where it still gives comparable performance in terms of classification accuracy. Best results are achieved by using LDC or SVM in both cases.

6.1.2 Experiments on FERET database

As stated earlier, FERET database pose subset is used to assess the performance with regards to large pose variations and large number of subjects. 50 out of 200 FERET subjects are used for training the covariance matrix for PCA. The remaining 1350 images of 150 subjects are used to evaluate classifier performance with respect to large pose differences. In order to assess the small sample size problem (i.e. number of raining examples available per subject), experiments on FERET are performed with respect to varying training/test sizes by using 2, 4, 6, and 8 examples per subject and testing on the remaining. Similarly, tests at each size are repeated 5 times, with different training/test partitioning, and the errors are averaged. Figure 8 shows the averaged classification errors for all the classifiers on FERET set for both the feature representations with respect to varying training and test sizes. As shown in figure 8, increasing number of subjects and pose differences has an adverse affect on the performance of all the classifiers on PCA-representation set while face-GLOH-Signature representation provides relatively better performance.

6.2 Template matching Setting

As stated earlier, due to the associated high dimensionality of the extracted features of GABOR, LBP, DCT etc, we assess the performance of these feature descriptions with that of face-GLOH signature across a number of pose mismatches in a typical template matching

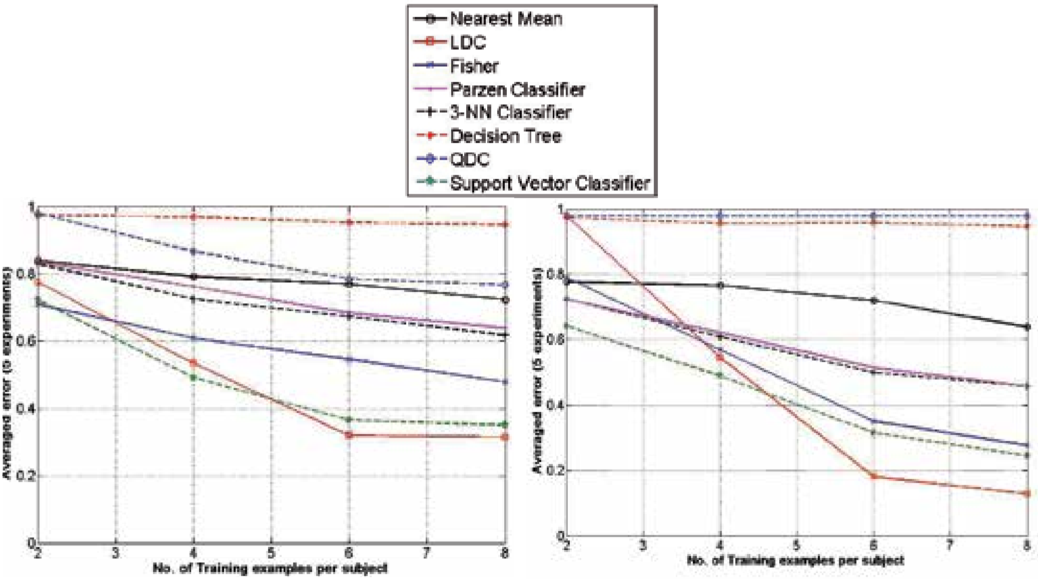


Fig. 8. Classifiers evaluation On FERET by varying training/test sizes (a) Using PCA-set (b) Using face-GLOH-signature set

setting. Frontal images of 200 FERET subjects are used as gallery while images for the remaining eight poses of each subject are used as test probes. Each probe is matched with each of the gallery images by using the cosine similarity metric. Probe is assigned the identity of the gallery subject for which it has the maximum similarity.

6.2.1 Test Results

We obtain each of the three feature descriptions as described in section 4. Gabor features are obtained by considering real part of the bank of Gabor filter kernel response tuned to 8 orientations and 5 scales, at each pixel location. This resulted in 40x92x112=412160 dimensional feature vector for each image. Due to memory constraints we used PCA to reduce the dimensionality to 16000-dimensional vector. For the LBPH feature representation, we use $LBP_{8,2}^{U2}$ operator in 18x21 window as described in (Ahonen et al, 2004) which resulted in a 2124 dimensional feature vector. The recognition scores in each pose are averaged. Table 2 depicts the performance comparison of different feature representations with that of Face-GLOH-Signature across a number of pose mismatches.

Feature Description	Average Recognition across each FERET Probe Pose			
	$\pm 15^{o}$	$\pm 25^{o}$	$\pm 45^{o}$	$\pm 60^{o}$
Eigenface	70.1%	56.2%	31.1%	13.4%
Gabor	91.4%	81.2%	68.5%	32.1%
LBPH	87.3 %	71.4%	56%	18.3%
Face-GLOH-Signature	100%	94.5%	81.1%	53.8%

Table 2. Comparison of face identification performance across pose of different feature representations

7. Conclusion

A comprehensive account of almost all the feature extraction methods used in current face recognition systems is presented. Specifically we have made distinction in the holistic and local feature extraction and differentiate them qualitatively as opposed to quantitatively. It is argued that a global feature representation should be preferred over a bag-of-feature approach. The problems in current feature extraction techniques and their reliance on a strict alignment is discussed. Finally we have introduced to use face-GLOH signatures that are invariant with respect to scale, translation and rotation and therefore do not require properly aligned images. The resulting dimensionality of the vector is also low as compared to other commonly used local features such as Gabor, LBP etc. and therefore learning based methods can also benefit from it.

In a typical multi-view face recognition task, where it is assumed to have several examples of a subject available for training, we have shown in an extensive experimental setting the advantages and weaknesses of commonly used feature descriptions. Our results show that under more realistic assumptions, most of the classifiers failed on conventional features. While using the introduced face-GLOH-signature representation is relatively less affected by large in-class variations. This has been demonstrated by providing a fair performance comparison of several classifiers under more practical conditions such as misalignments, large number of subjects and large pose variations. An important conclusion is to be drawn from the results on FERET is that conventional multi-view face recognition cannot cope well with regards to large pose variations. Even using a large number of training examples in different poses for a subject do not suffice for a satisfactory recognition. In order to solve the problem where only one training example per subject is available, many recent methods propose to use image synthesis to generate a given subject at all other views and then perform a conventional multi-view recognition (Beymer and Poggio, 1995; Gross et al, 2004). Besides the fact that such synthesis techniques cause severe artefacts and thus cannot preserve the identity of an individual, a conventional classification cannot yield good recognition results, as has been shown in an extensive experimental setting. More sophisticated methods are therefore needed in order to address pose invariant face recognition. Large pose differences cause significant appearance variations that in general are larger than the appearance variation due to identity. One possible way of addressing this is to learn these variations across each pose, more specifically by fixing the pose and establishing a correspondence on how a person's appearance changes under this pose one could reduce the in-class appearance variation significantly. In our very recent work (Sarfraz and Hellwich, 2009), we demonstrate the usefulness of face-GLOH signature in this direction.

8. References

Ahonen, T., Hadid, A. & Pietikainen, M. (2004). Face recognition with local binary patterns, *Proceedings of European Conference on Computer Vision ECCV*, pp. 469–481.

Ashraf A.B., Lucey S., and Chen T. (2008). Learning patch correspondences for improved viewpoint invariant face recognition. *Proceedings of IEEE Computer Vision and Pattern Recognition CVPR*, June.

Baochang Z., Shiguang S., Xilin C., and Wen G. (2007). Histogram of Gabor Phase Patterns (HGPP):A Novel Object Representation Approach for Face Recognition, *IEEE Trans. on Image Processing*, vol. 16, No. 1, pp. 57-68.

Beymer D. (1996). Pose-invariant face recognition using real and virtual Views. *M.I.T., A.I. Technical Report* No.1574, March.

Beymer, D. Poggio, T. (1995). Face recognition from one model view. *Proceedings of International conference on computer vision.*

Belhumeur, P. N., Hespanha, J. P. & Kriegman, D. J. (1997). Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp. 711-720.

Brunelli R. and Poggio T. (1993). Face recognition: Features versus templates. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 10, pp. 1042–1052.

Chen, L.F., Liao, H.Y., Lin, J.C. & Han, C.C. (2001). Why recognition in a statistics-based face recognition system should be based on the pure face portion: a probabilistic decision- based proof, *Pattern Recognition*, Vol. 34, No. 7, pp. 1393-1403.

Cardinaux, F., Sanderson, C. & Bengio, D S. (2006). User authentication via adapted statistical models of face images. *IEEE Trans. Signal Processing*, 54(1):361–373.

Daugman, J. (1993). High confidence visual recognition of persons by a test of statistical independence. IEEE *Transactions on Pattern Analysis and Machine Intelligence,* 15 1148–1161.

Duda, R.O., Hart, P.E. & Stork, D.G. (2001). *Pattern Classification*, 2nd edition, Wiley Interscience.

Eickeler, S., Müller, S. & Rigoll, G. (2000). Recognition of JPEG Compressed Face Images Based on Statistical Methods, *Image and Vision Computing*, Vol. 18, No. 4, pp. 279-287.

Gonzales,R. C. & Woods, R. E. (1993) *Digital Image Processing*, Addison-Wesley, Reading, Massachusetts.

Granlund, G. H. (1978). Search of a General Picture Processing Operator, *Computer Graphics and Image Processing*, 8, 155-173.

Gottumukkal, [R. & Asari, V. K. (2004). An improved face recognition technique based on modular PCA approach, *Pattern Recognition Letter*, vol. 25, no. 4, pp. 429–436.

Hubel, D., Wiesel, T. (1978). Functional architecture of macaque monkey visual cortex, *Proceedings of Royal Society on Biology,* 198 (1978) 1–59.

Gross R., Matthews I. and Baker S. (2004). Appearance-based face recognition and light-fields, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 449–465.

Hayward, W. G., Rhodes, G. & Schwaninger, A. (2008). An own-race advantage for components as well as configurations in face recognition, *Cognition* 106(2), 1017-1027.

Kanade, T., & Yamada, A. (2003). Multi-subregion based probabilistic approach towards pose-Invariant face recognition. *Proceedings of IEEE international symposium on computational intelligence in robotics automation,* Vol. 2, pp. 954–959.

Lindeberg T. (1998) Feature detection with automatic scale selection. *Int. Journal of computer vision*, vol. 30 no. 2, pp 79-116.

Lowe D. (2004). Distinctive image features from scale-invariant keypoints. *Int. Journal of computer vision*, 2(60):91-110.

Liu, C. (2004). Gabor-based kernel PCA with fractional power polynomial models for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 26, pp. 572–581.

Liu, C., Wechsler, H. (2002). Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *IEEE Transactions on Image Processing*, 11, pp. 467–476.

Lucey, S. & Chen, T. (2006). Learning Patch Dependencies for Improved Pose Mismatched Face Verification, *Proceedings of IEEE Int'l Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 17-22.

Lades, M., Vorbruggen, J., Budmann, J., Lange, J., Malsburg, C., Wurtz, R. (1993). Distortion invariant object recognition on the dynamic link architecture. *IEEE Transactions on Computers,* 42, 300–311.

Lee H.S., Kim D. (2006). Generating frontal view face image for pose invariant face recognition. *Pattern Recognition letters,* vol. 27, No. 7, pp. 747-754.

Liu, D. H., Lam, K. M., & Shen, L. S. (2004). Optimal sampling of Gabor features for face recognition. *Pattern Recognition Letters*, 25, 267-276.

Marta, P., Cassia, M. & Chiara,T. (2006). The development of configural face processing: the face inversion effect in preschool-aged children, Annual *meeting of the XVth Biennial International Conference on Infant Studies*, Jun 19, Westin Miyako, Kyoto, Japan.

Mikolajczyk and Schmid C. (2002). Performance evaluation of local descriptors, *IEEE Transaction on Pattern Analysis and Machine Intelligence PAMI*, 27(10), 31-47.

Martınez A. M. (2002) Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. *IEEE Transaction on Pattern Analysis and Machine Intelligence PAMI*, vol. 24, no. 6, pp. 748–763.

Ojala, T., Pietikainen, M. & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987.

Olshausen, B., Field, D. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381, 607–609.

Pentland, A., Moghaddam, B. & Starner, T. (1994). View-Based and modular eigenspaces for face recognition," *Proceedings IEEE Conference of Compute Vision and Pattern Recognition*, pp. 84–91.

Phillips, P.J., Moon, H., Rizvi, S.A. & Rauss, P.J. (2000). The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104.

Rao, R., Ballard, D. (1995). An active vision architecture based on iconic representations, *Artificial Intelligence,* 78,461–505.

Sarfraz, M.S. and Hellwich, O. (2008). Statistical Appearance Models for Automatic Pose Invariant Face Recognition, *Proceedings of 8th IEEE Int. conference on Face and Gesture Recognition 'FG',* IEEE computer Society , September 2008, Holland.

Sarfraz, Muhammad Saquib (2008). Towards Automatic Face Recognition in Unconstrained Scenarios", *PhD Dissertation*, urn:nbn:de:kobv:83-opus-20689.

Sarfraz, M.S., Hellwich, O. (2009)" Probabilistic Learning for Fully Automatic Face Recognition across Pose", *Image and Vision Computing*, Elsevier, doi: 10.1016/j.imavis.2009.07.008.

Schwaninger, A., Wallraven, C., Cunningham, D. W. & Chiller-Glaus, S. (2006). Processing of identity and emotion in faces: a psychophysical, physiological and computational perspective. *Progress in Brain Research* 156, 321-343.

Schiele, B., Crowley, J. (2000). Recognition without correspondence using multidimensional receptive field histograms, *International Journal on Computer Vision*, 36 31–52.

Shiguang, S., Wen, G., Chang,Y., Cao,B., Yang, P. (2004). Review the Strength of Gabor features for Face Recognition from the Angle of its Robustness to Misalignment, *Proceedings of International conference on Pattern Recognition ICPR*.

Turk, M. & Pentland, A. (1991). Eigenfaces for Recognition, *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, pp. 71-86.

Tan, K. & Chen, S. (2005). Adaptively weighted sub-pattern PCA for face recognition, *Neurocomputing*, 64, pp. 505–511.

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*, (7), 682–687.

Wang, Y. J., Chua, C. S., & Ho, Y. K. (2002). Facial feature detection and face recognition from 2D and 3D images. *Pattern Recognition Letters*, 23, 1191-1202.

Webb, A.R. (2002). *Statistical pattern recognition*, 2nd edition, John Willey & Sons.

Wiskott, L., Fellous, J., Krüger, N., Malsburg, C. (1997) Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19, 775–779.

Wu, H. Y., Yoshida Y., & Shioyama, T. (2002). Optimal Gabor filters for high speed face identification. *Proceedings of International Conference on pattern Recognition*, pp. 107-110.

Zhang, Lingyun and Garrison W. Cottrell (2005). Holistic processing develops because it is good. *Proceedings of the 27th Annual Cognitive Science Conference*, Italy.

Jie Zou, Qiang Ji, and George Nagy (2007). A Comparative Study of Local Matching Approach for Face Recognition, *IEEE Transaction on image processing*, VOL. 16, NO.10.

Zhang, W., Shan, S., Gao, W., Chen, X. & Zhang, H. (2005). Local Gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition," *Proceedings International Conference of Computer Vision ICCV*, pp. 786–791.

An Extension of Principal Component Analysis

Hongchuan Yu and Jian J. Zhang
National Centre for Computer Animation, Bournemouth University
U.K.

1. Introduction

Principal component analysis (PCA), which is also known as Karhunen-Loeve (KL) transform, is a classical statistic technique that has been applied to many fields, such as knowledge representation, pattern recognition and image compression. The objective of PCA is to reduce the dimensionality of dataset and identify new meaningful underlying variables. The key idea is to project the objects to an orthogonal subspace for their compact representations. It usually involves a mathematical procedure that transforms a number of correlated variables into a smaller number of uncorrelated variables, which are called principal components. The first principal component accounts for as much of the variability in the dataset as possible, and each succeeding component accounts for as much of the remaining variability as possible. In pattern recognition, PCA technique was first applied to the representation of human face images by Sirovich and Kirby in [1,2]. This then led to the well-known Eigenfaces method for face recognition proposed by Turk and Penland in [3]. Since then, there has been an extensive literature that addresses both the theoretical aspect of the Eigenfaces method and its application aspect [4-6]. In image compression, PCA technique has also been widely applied to the remote hyperspectral imagery for classification and compression [7,8]. Nevertheless, it can be noted that in the classical 1D-PCA scheme the 2D data sample (e.g. image) must be initially converted to a 1D vector form. The resulting sample vector will lead to a high dimensional vector space. It is consequently difficult to evaluate the covariance matrix accurately when the sample vector is very long and the number of training samples is small. Furthermore, it can also be noted that the projection of a sample on each principal orthogonal vector is a scale. Obviously, this will cause the sample data to be over-compressed. In order to solve this kind of dimensionality problem, Yang et al. [9,10] proposed the 2D-PCA approach. The basic idea is to directly use a set of matrices to construct the corresponding covariance matrix instead of a set of vectors. Compared with the covariance matrix of 1D-PCA, one can note that the size of the covariance matrix using 2D-PCA is much smaller. This improves the computational efficiency. Furthermore, it can be noted that the projection of a sample on each principal orthogonal vector is a vector. Thus, the problem of over-compression is alleviated in the 2D-PCA scheme. In addition, Wang et al. [11] proposed that the 2D-PCA was equivalent to a special case of the block-based PCA, and emphasized that this kind of block-based methods had been used for face recognition in a number of systems.

For the multidimensional array cases, the higher order SVD (HO-SVD) has been applied to face recognition in [12,13]. They both employed a higher order tensor form associated with people, view, illumination, and expression dimensions and applied the HO-SVD to it for face recognition. We formulated them into the N-Dimensional PCA scheme in [14]. However, the presented ND-PCA scheme still adopted the classical single directional decomposition. Besides, due to the size of tensor, HO-SVD implementation usually leads to a huge matrix along some dimension of tensor, which is always beyond the capacity of an ordinary PC. In [12,13], they all employed small sized intensity images or feature vectors and a limited number of viewpoints, facial expressions and illumination changes in their "tensorface", so as to avoid this numerical challenge in HO-SVD computation.

Motivated by the above-mentioned works, in this chapter, we will reformulate our ND-PCA scheme presented in [14] by introducing the multidirectional decomposition technique for a near optimal solution of the low rank approximation, and overcome the above-mentioned numerical problems. However, we also noted the latest progress – Generalized PCA (GPCA), proposed in [15]. Unlike the classical PCA techniques (i.e. SVD-based PCA approaches), it utilizes the polynomial factorization techniques to subspace clustering instead of the usual Singular Value Decomposition approach. The deficiency is that the polynomial factorization usually yields an overabundance of monomials, which are used to span a high-dimensional subspace in GPAC scheme. Thus, the dimensionality problem is still a challenge in the implementation of GPCA. We will focus on the classical PCA techniques in this chapter.

The remainder of this chapter is organized as follows: In Section 2, the classical 1D-PCA and 2D-PCA are briefly revisited. The ND-PCA scheme is then formulated by using the multidirectional decomposition technique in Section 3, and the error estimation is also given. To evaluate the ND-PCA, it is performed on the FRGC 3D scan facial database [16] for multi-model face recognition in Section 4. Finally, some conclusions are given in Section 5.

2. 1D- AND 2D-PCA, AN OVERVIEW

1D-PCA

Let a sample $X \in R^n$. This sample is usually expressed in a vector form in the case of 1D-PCA. Traditionally, principal component analysis is performed on a square symmetric matrix of the cross product sums, such as the Covariance and Correlation matrices (i.e. cross products from a standardized dataset), i.e.

$$\begin{cases} Cov = E\left\{(X - \bar{X})(X - \bar{X})^T\right\} \\ Cor = (X - X_0)(Y - Y_0)^T \end{cases} \tag{1}$$

where, $\bar{X}$ is the mean of the training set, while X_0, Y_0 are standard forms. Indeed, the analysis of the Correlation and Covariance are different, since covariance is performed within the dataset, while correlation is used between different datasets. A correlation object has to be used if the variances of the individual samples differ much, or if the units of measurement of the individual samples differ. However, correlation can be considered as a special case of covariance. Thus, we will only pay attention to the covariance in the rest of this chapter.

After the construction of the covariance matrix, Eigen Value Analysis is applied to *Cov* of Eq.(1), i.e. $Cov = U\Sigma U^T$. Herein, the first k eigenvectors in the orthogonal matrix *U* corresponding to the first *k* largest eigenvalues span an orthogonal subspace, where the major energy of the sample is concentrated. A new sample of the same object is projected in this subspace for its compact form (or PCA representation) as follows,

$$\alpha = U_k^T (X - \bar{X}) , \tag{2}$$

where, U_k is a matrix consisting of the first *k* eigenvectors of *U*, the projection α is a k-dimensional vector, which calls the *k* principal components of the sample *X*. The estimate of a novel representation of *X* can be described as,

$$\tilde{X} = U_k \alpha + \bar{X} . \tag{3}$$

It is clearly seen that the size of the covariance matrix of Eq.(1) is very large when the sample vectors are very long. Due to the large size of the covariance matrix and the relatively small number of training samples, it is difficult to estimate the covariance matrix of Eq.(1) accurately. Furthermore, a sample is projected on a principal vector as follows,

$$\alpha_i = u_i^T (X - \bar{X}), \quad \alpha_i \in \alpha, u_i \subset U_k, i = 1...k .$$

It can be noted that the projection α_i is a scale. Thus, this usually causes over-compression, i.e. we will have to use many principal components to approximate the original sample *X* for a desired quality. We call these above-mentioned numerical problems as "curse of dimensionality".

2D-PCA

In order to avoid the above mentioned problem, Yang et al. in [10] firstly presented a 2D-PCA scheme for 2D array cases in order to improve the performance of the PCA-style classifiers, that is, SVD is applied to the covariance matrix of, $G = \sum_i (X_i - \bar{X})^T (X_i - \bar{X})$, to get $G = V \Lambda V^T$, where $X_i \in R^{n \times m}$ denotes a sample, $\bar{X}$ denotes the mean of a set of samples, and *V* is the matrix of the eigenvectors and *Λ* is the matrix of the eigenvalues. The low-rank approximation of sample *X* is described as,

$$\begin{cases} \tilde{X} = Y V_k^T + \bar{X} \\ Y = (X - \bar{X}) V_k \end{cases} , \tag{4}$$

where V_k contains the first *k* principal eigenvectors of *G*. It has been noted that 2D-PCA only considers between column (or row) correlations [11].

In order to improve the accuracy of the low rank approximation, Ding et al. in [17] presented a 2D-SVD scheme for 2D cases. The key idea is to employ the 2-directional decomposition to the 2D-SVD scheme, that is, two covariance matrices of,

$$\begin{cases} F = \sum_i (X_i - \bar{X})(X_i - \bar{X})^T = U \Lambda_F U^T \\ G = \sum_i (X_i - \bar{X})^T (X_i - \bar{X}) = V \Lambda_G V^T \end{cases} ,$$

are considered together. Let U_k contain the first *k* principal eigenvectors of *F* and V_s contain the first *s* principal eigenvectors of *G*. The low-rank approximation of *X* can be expressed as,

$$\begin{cases} \tilde{X} = U_k M V_s^T + \bar{X} \\ M = U_k^T (X - \bar{X}) V_s \end{cases} . \tag{5}$$

Compared to the scheme Eq.(5), the scheme Eq.(4) of 2D-PCA only employs the classical single directional decomposition. It is proved that the scheme Eq.(5) of 2D-SVD can obtain a near-optimal solution compared to 2D-PCA in [17]. While, in the dyadic SVD algorithm [18], the sample set is viewed as a 3 order tensor and the HO-SVD technique is applied to each dimension of this tensor except the dimension of sample number, so as to generate the principal eigenvector matrices U_k and V_s as in the 2D-SVD.

3. N-DIMENSIONAL PCA

For clarity, we first introduce Higher Order SVD [19] briefly, and then formulate the N-dimensional PCA scheme.

3.1 Higher Order SVD

A higher order tensor is usually defined as $A \in R^{I_1 \times \ldots \times I_N}$, where N is the order of A, and $1 \leq i_n \leq I_n$, $1 \leq n \leq N$. In accordance with the terminology of tensors, the column vectors of a 2-order tensor (matrix) are referred to as 1-mode vectors and row vectors as 2-mode vectors. The n-mode vectors of an N-order tensor A are defined as the I_n-dimensional vectors obtained from A by varying the index i_n and keeping the other indices fixed. In addition, a tensor can be expressed in a matrix form, which is called matrix unfolding (refer to [19] for details).

Furthermore, the n-mode product, $\times_n$, of a tensor $A \in R^{I_1 \times \ldots \times I_n \ldots \times I_N}$ by a matrix $U \in R^{J_n \times I_n}$ along the n-th dimension is defined as,

$$(A \times_n U)_{i_1,\ldots,i_{n-1},j_n,i_{n+1},\ldots,i_N} = \sum_{i_n} a_{i_1,\ldots,i_n,\ldots,i_N} u_{j_n,i_n} .$$

In practice, n-mode multiplication is implemented first by matrix unfolding the tensor A along the given n-mode to generate its n-mode matrix form $A_{(n)}$, and then performing the matrix multiplication as follows,

$$B_{(n)} = UA_{(n)} .$$

After that, the resulting matrix $B_{(n)}$ is folded back to the tensor form, i.e. $A \times_n U = \text{fold}_n\left(U\text{unfold}_n(A)\right)$. In terms of n-mode multiplication, Higher Order SVD of a tensor A can be expressed as,

$$A = S \times_1 U^{(1)} \times_2 \ldots \times_N U^{(N)} , \tag{6}$$

where, $U^{(n)}$ is a unitary matrix of size $I_n \times I_n$, which contains n-mode singular vectors. Instead of being pseudo-diagonal (nonzero elements only occur when the indices $i_1 = \ldots = i_N$), the tensor S (called the core tensor) is all-orthogonal, that is, two subtensors $S_{i_n=a}$ and $S_{i_n=b}$ are orthogonal for all possible values of n, a and b subject to $a \neq b$. In addition, the Frobenius-norms $s_i^{(n)} = \left\| S_{i_n=i} \right\|_F$ are n-mode singular values of A and are in decreasing order, $s_1^{(n)} \geq ... \geq s_{I_n}^{(n)} \geq 0$, which correspond to n-mode singular vectors $u_i^{(n)} \subset U^{(n)}, i = 1,..., I_n$ respectively. The numerical procedure of HO-SVD can be simply described as,

$$\text{unfold}_n(A) = U^{(n)} \Sigma^{(n)} V^{(n)T}, n = 1,..., N,$$

where, $\Sigma^{(n)} = diag\left(s_1^{(n)},..., s_{I_n}^{(n)}\right)$ and $V^{(n)}$ is another orthogonal matrix of SVD.

3.2 Formulating N-dimensional PCA

For the multidimensional array case, we first employ a difference tensor instead of the covariance tensor as follows,

$$D = \left((X_1 - \bar{X}),...,(X_M - \bar{X})\right), \tag{7}$$

where $X_i \in R^{I_1 \times ... I_i \times ... \times I_N}$ and $D \in R^{I_1 \times ... MI_i \times ... \times I_N}$, i.e. *N*-order tensors $(X_n - \bar{X}), n = 1,..., M$ are stacked along the *i*th dimension in the tensor *D*. Then, applying HO-SVD of Eq.(6) to *D* will generate n-mode singular vectors contained in $U^{(n)}, n = 1,..., N$. According to the n-mode singular values, one can determine the desired principal orthogonal vectors for each mode of the tensor *D* respectively. Introducing the multidirectional decomposition to Eq.(7) will yield the desired N-dimensional PCA scheme as follows,

$$\begin{cases} \tilde{X} = Y \times_1 U_{k_1}^{(1)} \times_2 ... \times_N U_{k_N}^{(N)} + \bar{X} \\ Y = (X - \bar{X}) \times_1 U_{k_1}^{(1)T} \times_2 ... \times_N U_{k_N}^{(N)T} \end{cases}, \tag{8}$$

where $U_{k_i}^{(i)}$ denotes the matrix of i-mode k_i principal vectors, *i* = 1,…*N*. The main challenge is that unfolding the tensor *D* in HO-SVD usually generates an overly large matrix.

First, we consider the case of unfolding *D* along the *i*th dimension, which generates a matrix of size $MI_i \times (I_{i+1} \cdot ... \cdot I_N \cdot I_1 \cdot ... \cdot I_{i-1})$. We prefer a unitary matrix $U^{(i)}$ of size $I_i \times I_i$ to one of the sizes $MI_i \times MI_i$. This can be achieved by reshaping the unfolded matrix as follows.

Let A_j be a $I_i \times (I_{i+1} \cdot ... \cdot I_N \cdot I_1 \cdot ... \cdot I_{i-1})$ matrix and *j* = 1,…*M*. The unfolded matrix is expressed as $A = \begin{pmatrix} A_1 \\ ... \\ A_M \end{pmatrix}$. Reshaping *A* into a $I_i \times M(I_{i+1} \cdot ... \cdot I_N \cdot I_1 \cdot ... \cdot I_{i-1})$ matrix $\tilde{A} = \left(A_1,..., A_M\right)$, we can obtain an unitary matrix $U^{(i)}$ of size $I_i \times I_i$ by SVD.

Then, consider the generic case. Since the sizes of dimensions $I_1,..., I_N$ may be very large, this still leads to an overly large matrix along some dimension of sample *X*. Without loss of generality, we assume that the sizes of dimensions of sample *X* are independent of each other.

Now, this numerical problem can be rephrased as follows, for a large sized matrix, how to carry out SVD decomposition. It is straightforward to apply matrix partitioning approach to the large matrix. As a start point, we first provide the following lemma.

Lemma:

For any matrix $M \in R^{n \times m}$, if each column M_i of *M*, $M = (M_1,...,M_m)$, maintain its own singular value σ_i , i.e. $M_i M_i^T = U_i diag(\sigma_i^2,0,...,0)U_i^T$, while the singular values of *M* are $s_1,...,s_{\min(m,n)}$, i.e. $M = Vdiag(s_1,...,s_{\min(m,n)})U^T$, then $\sum_{i=1}^{\min(m,n)} \sigma_i^2 = \sum_{i=1}^{\min(m,n)} s_i^2$.

Proof:

Let *n* > *m*. Because,

$$MM^T = \sum_{i=1}^{m} M_i M_i^T = \sum_{i=1}^{m} u_i \sigma_i^2 u_i^T = \left(u_1,...,u_m\right) diag(\sigma_1^2,...,\sigma_m^2) \left(u_1,...,u_m\right)^T ,$$

where u_i is the first column of each U_i , while the SVD of MM^T ,

$$MM^T = Vdiag(s_1^2,...,s_m^2,0,...,0)V^T = \sum_{i=1}^{m} v_i s_i^2 v_i^T ,$$

where v_i is the *i*th column of *V*. We have,

$$tr(MM^T) = \sum_{i}^{m} \sigma_i^2 = \sum_{i}^{m} s_i^2 , \qquad \textbf{End of proof.}$$

This lemma implies that each column of *M* corresponds to its own singular value. Moreover, let *M*

let M_i be a submatrix instead of column vector, $M_i \in R^{n \times r}$. We have,

$$M_i M_i^T = U_i diag(s_{1i}^2,...s_{ri}^2,...,0)U_i^T .$$

It can be noted that there are more than one non-zero singular values $s_{1i} \ge ... \ge s_{ri} \ge 0$. If we let $rank(M_i M_i^T) = 1$, the approximation of $M_i M_i^T$ can be written as $M_i M_i^T \approx U_i diag(s_{1i}^2,0,...,0)U_i^T$. In terms of the lemma, we can also approximate it as $M_i M_i^T \approx M_{1i} M_{1i}^T = u_{1i} \sigma_{1i}^2 u_{1i}^T$, where M_{1i} is a column of M_i corresponding to the biggest singular value σ_{1i} of column vector. On this basis, M_{1i} is regarded as the principal column vector of the submatrix M_i.

We can rearrange the matrix $M \in R^{n \times m}$ by sorting these singular values $\{\sigma_i\}$ and partition it into *t* block submatrices, $M = (M_1,...,M_t)$, where $M_i \in R^{n \times m_i}, i = 1,...,t, m = \sum_{i}^{t} m_i$. Indeed, the principal eigenvectors are derived only from some particular submatrices rather than the others as the following analysis. (For computational convenience, we assume *m* ≥ *n* below.)

In the context of PCA, the matrix of the first *k* principal eigenvectors is preferred to a whole orthogonal matrix. Thus, we partition *M* into 2 block submatrices $\tilde{M} = (M_1, M_2)$ in terms of the sorted singular values $\{\sigma_i\}$, so that M_1 contains the columns corresponding to the first *k* biggest singular values while M_2 contains the others. Note that $\tilde{M}$ is different from the original *M* because of a column permutation (denoted as *Permute*). Applying SVD to each

M_i respectively yields,

$$\tilde{M} = (U_1, U_2)\begin{pmatrix} \Lambda_1 & \\ & \Lambda_2 \end{pmatrix}\begin{pmatrix} V_1^T & \\ & V_2^T \end{pmatrix}. \tag{9}$$

Thus, matrix $\tilde{M}$ can be approximated as follows,

$$\tilde{M} \approx \tilde{M}' = (U_1, U_2)\begin{pmatrix} \Lambda_1 & \\ & 0 \end{pmatrix}\begin{pmatrix} V_1^T & \\ & V_2^T \end{pmatrix}. \tag{10}$$

In order to obtain the approximation of *M*, the inverse permutation of *Permute* needs to be carried out on the row-wise orthogonal matrix of $\begin{pmatrix} V_1^T & \\ & V_2^T \end{pmatrix}$ given in Eq.(10). The resulting matrix is the approximation of the original matrix *M*. The desired principal eigenvectors are therefore included in the matrix of U_1.

Now, we can re-write our ND-PCA scheme as,

$$\begin{cases} \tilde{X} = Y \times_1 U_{k_1}^{(1)} \ldots \times_i U_{k_i}^{(i)} \ldots \times_N U_{k_N}^{(N)} + \bar{X} \\ Y = (X - \bar{X}) \times_1 U_{k_1}^{(1)T} \ldots \times_N U_{k_N}^{(N)T} \\ U_{k_i}^{(i)} \text{ is from Eq.(10)} \end{cases}. \tag{11}$$

For comparison, the similarity metric can adopt the Frobenius-norms between the reconstructions of two samples *X* and X' as follows,

$$\varepsilon = \left\| \tilde{X} - \tilde{X}' \right\|_F = \left\| Y - Y' \right\|_F. \tag{12}$$

Furthermore, we can provide the following proposition,

Proposition:

$\tilde{X}$ of Eq.(11) is a near optimal approximation to sample *X* in a least-square sense.

Proof.

According to the property 10 of HO-SVD in [19], we assume that the n-mode rank of $(X - \bar{X})$ be equal to $R_n (1 \le n \le N)$ and $(\tilde{X} - \bar{X})$ be defined by discarding the smallest n-mode singular values $\sigma_{I_n'+1}^{(n)}, \ldots, \sigma_{R_n}^{(n)}$ for given I_n'. Then, the approximation $\tilde{X}$ is a near optimal approximation of sample *X*. The error is bounded by Frobenius-norm as follows,

$$\left\| X - \tilde{X} \right\|_F^2 \le \sum_{i_1 = I_1'+1}^{R_1} \sigma_{i_1}^{(1)2} + \ldots + \sum_{i_N = I_N'+1}^{R_N} \sigma_{i_N}^{(N)2}. \tag{13}$$

This means that the tensor $(\tilde{X} - \bar{X})$ is in general not the best possible approximation under the given n-mode rank constraints. But under the error upper-bound of Eq.(13), $\tilde{X}$ is a near optimal approximation of sample *X*.

Unfolding $(X - \bar{X})$ along *i*th dimension yields a large matrix which can be partitioned into two submatrices as shown in Eq.(9), i.e.

$$\tilde{M} = (M_1, M_2) = (U_1, U_2)\begin{pmatrix} \Lambda_1 & \\ & \Lambda_2 \end{pmatrix}\begin{pmatrix} V_1^T & \\ & V_2^T \end{pmatrix}.$$

Let $\tilde{M}' = \left(U_1, U_2\right)\begin{pmatrix}\Lambda_1 & \\ & 0\end{pmatrix}\begin{pmatrix}V_1^T & \\ & V_2^T\end{pmatrix}$ as shown in Eq.(10). Consider the difference of $\tilde{M}$ and $\tilde{M}' \in R^{n\times m}$ as follows,

$$\tilde{M} - \tilde{M}' = \left(U_1, U_2\right)\begin{pmatrix}0 & \\ & \Lambda_2\end{pmatrix}\begin{pmatrix}V_1^T & \\ & V_2^T\end{pmatrix},$$

where $U_i \in R^{n\times n}, V_i \in R^{m_i \times m_i}, \Lambda_i \in R^{n\times m_i}, i = 1,2$. It can be noted that the 2-norm of $\begin{pmatrix}V_1^T & \\ & V_2^T\end{pmatrix}$ is 1, and that of $\begin{pmatrix}0 & \\ & \Lambda_2\end{pmatrix}$ is $\max\{\sigma : \sigma \in \Lambda_2\}$. Since

$$\left(U_1, U_2\right) = U_1\left(I_{n\times n}, I_{n\times n}\right)\begin{pmatrix}I_{n\times n} & \\ & U_1^T U_2\end{pmatrix},$$

we can note that the 2-norm of both the orthogonal matrix U_1 and $\begin{pmatrix}I_{n\times n} & \\ & U_1^T U_2\end{pmatrix}$ are 1, and that of $\left(I_{n\times n}, I_{n\times n}\right)$ is $\sqrt{2}$ because of identity matrix $I_{n\times n}$. Therefore, we have,

$$\left\|\tilde{M} - \tilde{M}'\right\|_2^2 \leq 2\max^2\{\sigma : \sigma \in \Lambda_2\}, \tag{14}$$

in a 2-norm sense.

Substituting Eq.(14) into Eq.(13) yields the error upper-bound of $\tilde{X}$ as follows,

$$\left\|X - \tilde{X}\right\|_F^2 \leq 2\left(\max^2\left\{\sigma^{(1)} : \sigma^{(1)} \in \Lambda_2^{(1)}\right\} + \ldots + \max^2\left\{\sigma^{(N)} : \sigma^{(N)} \in \Lambda_2^{(N)}\right\}\right). \tag{15}$$

This implies that the approximation $\tilde{X}$ of Eq.(11) is a near optimal approximation of sample X under this error upper bound. **End of proof**.

Remark: So far, we formulated the ND-PCA scheme, which can deal with overly large matrix. The basic idea is to partition the large matrix and discard non-principal submatrices. In general, the dimensionality of eigen-subspace is determined by the ratio of sum of singular values in the subspace to the one of the whole space for solving the dimensionality reduction problems [20]. But, for an overly large matrix, we cannot get all the singular values of the whole matrix here, because of discarding the non-principal submatrices. An alternative is to iteratively determine the dimensionality of eigen-subspace by using reconstruction error threshold.

4. EXPERIMENTS AND ANALYSIS

The proposed ND-PCA approach was performed on a 3D range database of human faces used for the Face Recognition Grand Challenge [16]. In order to establish an analogy with a 3D volume dataset or multidimensional solid array, each 3D range dataset was first mapped to a 3D array and the intensities of the corresponding pixels in the still face image were regarded as the voxel values of the 3D array. For the sake of memory size, the reconstructed volume dataset was then re-sampled to the size of 180×180×90. Figure 1 shows an example of the still face image, corresponding range data and the reconstructed 3D model.

Experiment 1. This experiment is to test the rank of the singular values. In our gallery, eight samples of each person are available for training. Their mean-offset tensors are aligned together along the second index (*x* axis) to construct a difference tensor $D \in R^{180\times1440\times90}$. We applied HO-SVD of Eq.(6) to *D* to get the 1-mode and 3-mode singular values of *D*, which are depicted in Fig.2. One can note that the numbers of 1-mode and 3-mode singular values are different, and they are equal to the dimensionalities of indices 1 and 3 of *D* respectively (i.e. 180 for 1-mode and 90 for 3-mode). This is a particular property of higher order tensors, namely the N-order tensor *A* can have *N* different n-mode ranks but all of them are less than the rank of *A*, $rank_n(A) \le rank(A)$. Furthermore, the corresponding n-mode singular vectors constitutes orthonormal basis which can span independent n-mode orthogonal subspaces respectively. Therefore, we can project a sample to an arbitrary n-mode orthogonal subspace accordingly. In addition, one can also note that the magnitude of the singular values declines very quickly. This indicates that the energy of a sample is only concentrated on a small number of singular vectors as expected.

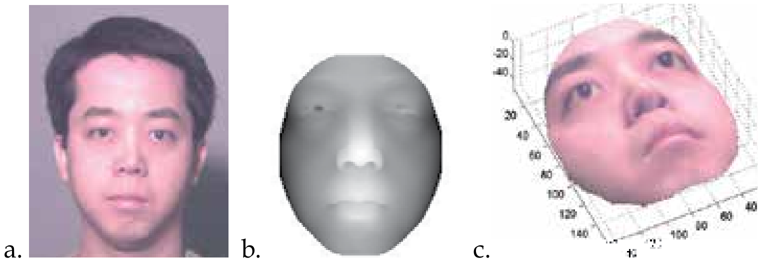


Fig. 1. The original 2D still face image (a), range data (b) and reconstructed 3D model (c) of a face sample.

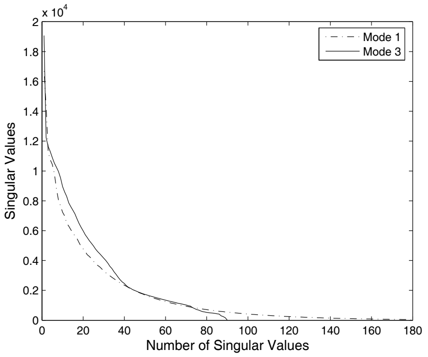


Fig. 2. The singular values in decreasing order.

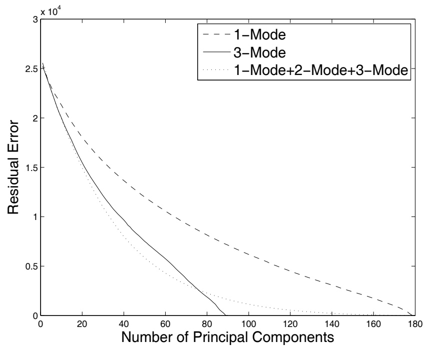


Fig. 3. Comparison of the reconstruction through 1-mode, 3-mode and 1-mode+2-mode+3-mode principal subspace respectively. ND-PCA with multidirectional decomposition converges quicker than ND-PCA with single directional decomposition.

Experiment 2. This experiment is to test the quality of the reconstructed sample. Within our 3D volume dataset, we have 1-mode, 2-mode and 3-mode singular vectors, which could span three independent orthogonal subspaces respectively. The sample could be approximated by using the projections from one orthogonal subspace, two ones or three ones. Our objective is to test which combination leads to the best reconstruction quality. We designed a series of tests for this purpose. The reconstructed sample using the scheme of Eq.(11) was performed on 1-mode, 3-mode and 1-mode+2-mode+3-mode principal subspaces respectively with a varying number of principal components *k*. (Note that 1-mode or 3-mode based ND-PCA adopted the single directional decomposition, while 1-mode+2-mode+3-mode based ND-PCA adopted the multidirectional decomposition.) The residual errors of reconstruction are plotted in Fig.3. Since the sizes of dimensions of $U^{(1)}$ and $U^{(3)}$ are different, the ranges of the corresponding number of principal components *k* are also different. However, *k* must be less than the size of dimension of the corresponding orthogonal matrix $U^{(1)}$ or $U^{(3)}$. As a result of the differing dimensionalities, the residual error of reconstruction in 3-mode principal subspace converges to zero faster than in 1-mode or 1-mode+2-mode+3-mode principal subspaces. Indeed, if the curve of 3-mode (solid curve) is quantified to the same length of row coordinate as the curve of 1-mode (dashed line) in Fig.3, there is no substantial difference compared to the 1-mode test. This indicates that the reconstructed results are not affected by the difference between the different n-mode principal subspaces. Furthermore, in the test of 1-mode+2-mode+3-mode principal subspaces, the number of principal components *k* was set to 180 for both $U^{(1)}$ and $U^{(2)}$ while it was set to 90 for $U^{(3)}$. Comparing the curve of 1-mode+2-mode+3-mode (dot line) with that of 1-mode (dashed line) and 3-mode (solid line), one can note that the approximation of 1-mode+2-mode+3-mode principal subspace converges to the final optimal solution more rapidly.

—·—·—·

Remark: In [9,10], the over-compressed problem was addressed repeatedly. [10] gave a comparison of the reconstruction results between the 1D-PCA case and the 2D-PCA case, which is reproduced in Fig.4 for the sake of completeness. It can be noted that the small number of principal components of the 2D-PCA can perform well compared with the large number of principal components of the 1D-PCA. Moreover, consider the cases of single directional decomposition, i.e. 2D-PCA and 1-mode based ND-PCA scheme, and multidirectional decomposition, i.e. 2D-SVD and 1-mode+2-mode+3-mode based ND-PCA. We respectively compared the reconstructed results of the single directional decomposition and the multidirectional decomposition with a varying number of principal components k (i.e. the reconstruction of the volume dataset by using the ND-PCA of Eq.(11) while the reconstruction of the corresponding 2D image respectively by using 2D-PCA of Eq.(4) and 2D-SVD of Eq.(5)). The training set is the same as in the first experiment. The residual errors of reconstruction are normalized to the range of [0,1], and are plotted in Fig.5. One can note that the multidirectional decomposition performs better than the single directional decomposition in the case of a small number of principal components (i.e. comparing Fig.5a with Fig.5b). But then comparing the 2D-PCA with ND-PCA scheme shown in Fig.5a (or 2D-SVD with ND-PCA scheme shown in Fig.5b), one can also note that 2D-PCA (or 2D-SVD) performs a little better than ND-PCA scheme when only a small number of principal components are used. In our opinion, there is no visible difference in the reconstruction quality between 2D-PCA (or 2D-SVD) and ND-PCA scheme with a small number of singular values. This is because the reconstructed 3D volume dataset is a sparse 3D array (i.e. all voxel values are set to zero except the voxels on the face surface), it is therefore more sensitive to computational errors compared to a 2D still image. If the 3D volume datasets were solid, e.g. CT or MRI volume datasets, this difference between the two curves of Fig.5a or Fig.5b would not noticeably appear.

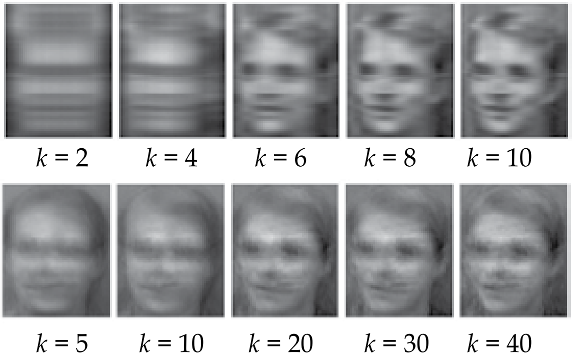
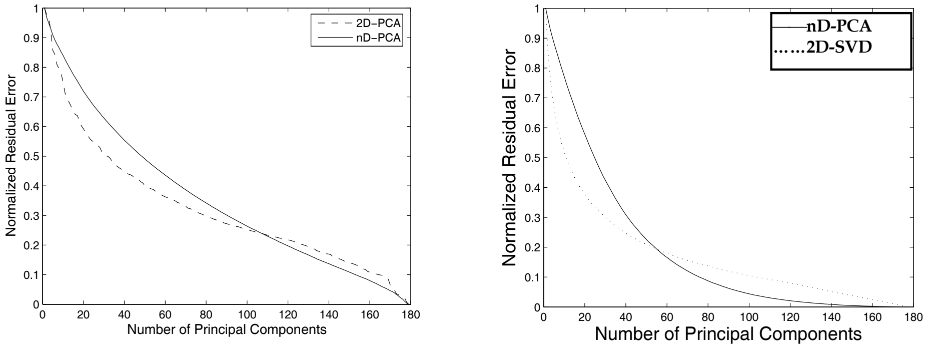


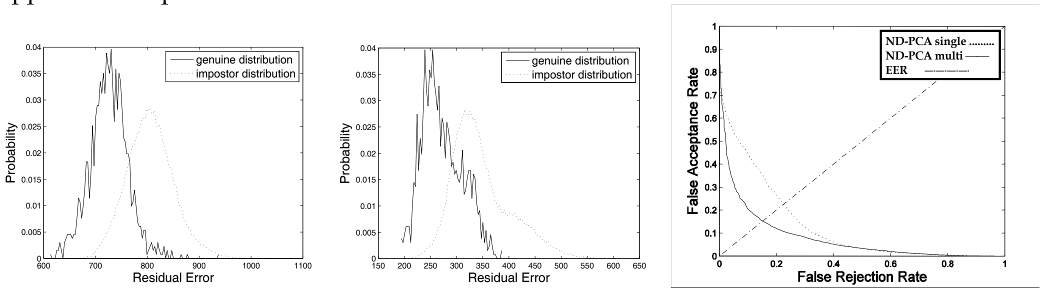
Fig. 4. Comparison of the reconstructed images using 2D-PCA (upper) and 1D-PCA (lower) from [10].



a. single direction decomposition. b. multiple direction decomposition

Fig. 5. Comparison of the reconstruction by using single directional decomposition (a), i.e. 2D-PCA and 1-mode based ND-PCA scheme, and multidirectional composition (b), i.e. 2D-SVD and ND-PCA, in terms of the normalized residual errors.

Experiment 3. In this experiment, we compared the 1-mode based ND-PCA scheme with the 1-mode+2-mode+3-mode based ND-PCA scheme on the performance of the face verification using the Receiver Operating Characteristic (ROC) curves [21]. Our objective is to reveal the recognition performance between these two ND-PCA schemes respectively by using the single directional decomposition and the multidirectional decomposition. The whole test set includes 270 samples (i.e. range datasets and corresponding still images), in which there are 6 to 8 samples for one person. All these samples are from the FRGC database and are re-sampled. Two ND-PCA schemes were carried out directly on the reconstructed volume datasets. Their corresponding ROC curves are shown respectively in Fig.6. It can be noted that the overlapping area of the genuine and impostor distributions (i.e. false probability) in Fig.(6a) is smaller than that in Fig.(6b). Furthermore, their corresponding ROC curves relating to the False Acceptance Rate (FAR) and the False Rejection Rate (FRR) are depicted by changing the threshold as shown in Fig.(6c). At some threshold, the false probability of recognition corresponds to some rectangular area under the ROC curve. The smaller the area under the ROC curve, the higher is the rising of the accuracy of the recognition. For quantitative comparison, we could employ the Equal Error Rate (EER), which is defined as the error rate at the point on ROC curve where the FAR is equal to the FRR. The EER is often used for comparisons because it is simpler to obtain and compare a single value characterizing the system performance. In Fig.(6c), the EER of Fig.(6a) is 0.152 while the EER of Fig.(6b) is 0.224. Obviously, the ND-PCA scheme with multidirectional decomposition can improve the accuracy of face recognition. Of course, since the EERs only give comparable information between the different systems that are useful for a single application requirement, the full ROC curve is still necessary for other potentially different application requirements.



a. b. c.

Fig. 6. Comparison of the recognition performance. a) are the genuine and impostor distribution curves of ND-PCA with multidirectional decomposition; b) are the genuine and impostor distribution curves of ND-PCA with single directional decomposition; c) are the ROC curves relating to the False acceptance rate and False rejection rate.

5. CONCLUSION

In this chapter, we formulated the ND-PCA approach, that is, to extend the PCA technique to the multidimensional array cases through the use of tensors and Higher Order Singular Value Decomposition technique. The novelties of this chapter include, 1) introducing the multidirectional decomposition into ND-PCA scheme and overcoming the numerical difficulty of overly large matrix SVD decomposition; 2) providing the proof of the ND-PCA scheme as a near optimal linear classification approach. We performed the ND-PCA scheme on 3D volume datasets to test the singular value distribution, and the error estimation. The results indicated that the proposed ND-PCA scheme performed as well as we desired. Moreover, we also performed the ND-PCA scheme on the face verification for the comparison of single directional decomposition and multidirectional decomposition. The experimental results indicated that the ND-PCA scheme with multidirectional decomposition could effectively improve the accuracy of face recognition.

6. References

1. Sirovich, L. and Kirby, M. (1987). Low-Dimensional Procedure for Characterization of Human Faces. *J. Optical Soc. Am.*, Vol. 4, pp. 519-524.
2. Kirby, M. and Sirovich, L. (1990). Application of the KL Procedure for the Characterization of Human Faces. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 1, pp. 103-108.
3. Turk, M. and Pentland, A. (1991). Eigenfaces for Recognition. *J. Cognitive Neuroscience*, Vol. 3, No. 1, pp. 71-86.
4. Sung, K. and Poggio, T. (1998). Example-Based Learning for View-Based Human Face Detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 1, pp. 39-51.
5. Moghaddam, B. and Pentland, A. (1997). Probabilistic Visual Learning for Object Representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp. 696-710.
6. Zhao, L. and Yang, Y. (1999). Theoretical Analysis of Illumination in PCA-Based Vision Systems. *Pattern Recognition*, Vol. 32, No. 4, pp. 547-564.
7. Harsanyi, J.C. and Chang, C. (1994). Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach. *IEEE Trans. Geoscience Remote Sensing*, Vol. 32, No. 4, pp. 779-785.
8. Sunghyun, L.; Sohn, K.H. and Lee, C. (2001). Principal component analysis for compression of hyperspectral images. *Proc. of IEEE Int. Geoscience and Remote Sensing Symposium*, Vol. 1, pp. 97-99.
9. Yang, J. and Yang, J.Y. (2002). From Image Vector to Matrix: A Straightforward Image Projection Technique—IMPCA vs. PCA. *Pattern Recognition*, Vol. 35, No. 9, pp. 1997-1999.
10. Yang, J.; Zhang, D.; Frangi, A.F. and Yang, J.Y. (2004). Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 1, pp. 131-137.
11. Wang, L.; Wang, X. and Zhang, X. et al. (2005). The equivalence of the two-dimensional PCA to lineal-based PCA. *Pattern Recognition Letters*, Vol. 26, pp. 57-60.
12. Vasilescu, M. and Terzopoulos, D. (2003). Multilinear subspace analysis of image ensembles. *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2003)*, Vol. 2, June 2003.
13. Wang, H. and Ahuja, N. (2003). Facial Expression Decomposition. *Proc. of IEEE 9th Int'l Conf. on Computer Vision (ICCV'03)*, Vol. 2, Oct. 2003.
14. Yu, H. and Bennamoun, M. (2006). 1D-PCA 2D-PCA to nD-PCA. *Proc. of IEEE 18th Int'l Conf. on Pattern Recognition*, HongKong, pp. 181-184, Aug. 2006.
15. Vidal, R.; Ma, Y. and Sastry, S. (2005). Generalized Principal Component Analysis (GPCA). *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 12.
16. Phillips, P.J.; Flynn, P.J. and Scruggs, T. et al. (2005). Overview of the Face Recognition Grand Challenge. *Proc. of IEEE Conf. on CVPR2005*, Vol. 1.
17. Ding, C. and Ye, J. (2005). Two-dimensional Singular Value Decomposition (2DSVD) for 2D Maps and Images. *Proc. of SIAM Int'l Conf. Data Mining (SDM'05)*, pp:32-43, April 2005.

18. Inoue, K. and Urahama, K. (2006). Equivalence of Non-Iterative Algorithms for Simultaneous Low Rank Approximations of Matrices. *Proc. of IEEE Int'l Conf. on Computer Vision and Pattern Recognition (CVPR'06)*, Vol.1, pp. 154-159.

19. Lathauwer, L.D.; Moor, B.D. and Vandewalle, J. (2000). A Multilinear Singular Value Decomposition. *SIAM J. on Matrix Analysis and Applications*, Vol. 21, No. 4, pp. 1253-1278.

20. Moghaddam, B. and Pentland, A. (1997). Probabilistic Visual Learning for Object Representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp. 696-710.

21. Jain, A.K.; Ross, A. and Prabhakar, S. (2004). An Introduction to Biometric Recognition. *IEEE Trans. on Circuits and Systems For Video Technology*, Vol. 14, No. 1, pp. 4-20.

3

Curvelet Based Feature Extraction

Tanaya Guha[1] and Q. M. Jonathan Wu[2]
[1]*ECE, University of British Columbia, BC, Canada*
[2]*ECE, University of Windsor, ON, Canada*

1. Introduction

Designing a completely automatic and efficient face recognition system is a grand challenge for biometrics, computer vision and pattern recognition researchers. Generally, such a recognition system is able to perform three subtasks: face detection, feature extraction and classification. We'll put our focus on feature extraction, the crucial step prior to classification. The key issue here is to construct a representative feature set that can enhance system-performance both in terms of accuracy and speed.

At the core of machine recognition of human faces is the extraction of proper features. Direct use of pixel values as features is not possible due to huge dimensionality of the faces. Traditionally, Principal Component Analysis (PCA) is employed to obtain a lower dimensional representation of the data in the standard eigenface based methods [Turk and Pentland 1991]. Though this approach is useful, it suffers from high computational load and fails to well-reflect the correlation of facial features. The modern trend is to perform multiresolution analysis of images. This way, several problems like, deformation of images due to in-plane rotation, illumination variation and expression changes can be handled with less difficulty.

Multiresolution ideas have been widely used in the field of face recognition. The most popular multiresolution analysis tool is the Wavelet Transform. In wavelet analysis an image is usually decomposed at different scales and orientations using a wavelet basis vector. Thereafter, the component corresponding to maximum variance is subjected to 'further operation'. Often this 'further operation' includes some dimension reduction before feeding the coefficients to classifiers like Support Vector Machine (SVM), Neural Network (NN) and Nearest Neighbor. This way, a compact representation of the facial images can be achieved and the effect of variable facial appearances on the classification systems can also be reduced. The wide-spread popularity of wavelets has stirred researchers' interest in multiresolution and harmonic analysis. Following the success of wavelets, a series of multiresolution, multidimensional tools, namely contourlet, curvelet, ridgelet have been developed in the past few years. In this chapter, we'll concentrate on Digital Curvelet Transform. First, the theory of curvelet transform will be discussed in brief. Then we'll talk about the potential of curvelets as a feature descriptor, looking particularly into the problem of image-based face recognition. Some experimental results from recent scientific works will be provided for ready reference.

2. Curvelet Transform

Before getting started with curvelet transform, the reader is suggested to go through the theory of multiresolution analysis, especially wavelet transform. Once the basic idea of wavelets and multiresolution analysis is understood, curvelets will be easier to comprehend.

2.1 Theory and Implementation

Motivated by the need of image analysis, Candes and Donoho developed curvelet transform in 2000 [Candes and Donoho 2000]. Curvelet transform has a highly redundant dictionary which can provide sparse representation of signals that have edges along regular curve. Initial construction of curvelet was redesigned later and was re-introduced as Fast Digital Curvelet Transform (FDCT) [Candes et al. 2006]. This second generation curvelet transform is meant to be simpler to understand and use. It is also faster and less redundant compared to its first generation version. Curvelet transform is defined in both continuous and digital domain and for higher dimensions. Since image-based feature extraction requires only 2D FDCT, we'll restrict our discussion to the same.

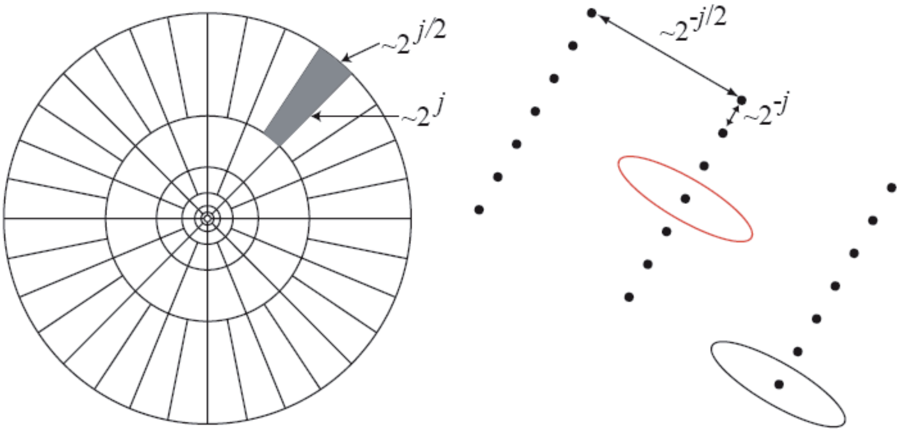


Fig. 1. Curvelets in Fourier frequency (left) and spatial domain (right) [Candes et al. 2006].

In order to implement curvelet transform, first 2D Fast Fourier Transform (FFT) of the image is taken. Then the 2D Fourier frequency plane is divided into wedges (like the shaded region in fig. 1). The parabolic shape of wedges is the result of partitioning the Fourier plane into radial (concentric circles) and angular divisions. The concentric circles are responsible for the decomposition of an image into multiple scales (used for bandpassing the image at different scale) and the angular divisions partition the bandpassed image into different angles or orientations. Thus if we want to deal with a particular wedge we'll need to define its scale j and angle ℓ. Now let's have a look at the spatial domain (fig. 1 right). Each of the wedges here corresponds to a particular curvelet (shown as ellipses) at a given scale and angle. This indicates that the inverse FFT of a particular wedge if taken, will determine the curvelet coefficients for that scale and angle. This is the main idea behind the implementation of curvelet transform. Figure 1 (right) represents curvelets in spatial Cartesian grid associated with a given scale and angle.

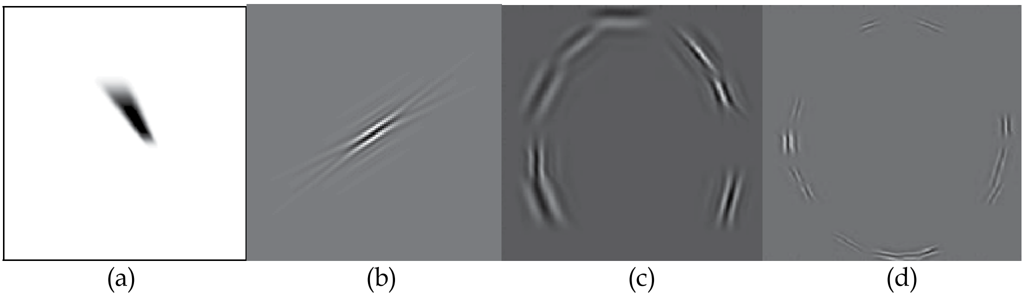


Fig. 2. (a) a real wedge in frequency domain, (b) corresponding curvelet in spatial domain [Candes et al. 2006], (c) curvelets aligned along a curve at a particular scale, (d) curvelets at a finer scale [Starck et al. 2002].

There are two different digital implementations of FDCT: Curvelets via USFFT (Unequally Spaced Fast Fourier Transform) and Curvelets via Wrapping. Both the variants are linear and take as input a Cartesian array to provide an output of discrete coefficients. Two implementations only differ in the choice of spatial grid to translate curvelets at each scale and angle. FDCT wrapping is the fastest curvelet transform currently available [Candes et al. 2006].

Though curvelets are shown to form the shape of an ellipse in fig. 1, looking at fig. 2 (b-d), we can understand that actually it looks more like elongated needles. This follows from the parabolic scaling law (length $\approx$width^2) that curvelets obey. The values of curvelet coefficients are determined by how they are aligned in the real image. The more accurately a curvelet is aligned with a given curve in an image, the higher is its coefficient value. A very clear explanation is provided in figure 3. The curvelet named '**c**' in the figure is almost perfectly aligned with the curved edge and therefore has a high coefficient value. Curvelets '**a**' and '**b**' will have coefficients close to zero as they are quite far from alignment. It is well-known that a signal localized in frequency domain is spread out in the spatial domain or vice-versa. A notable point regarding curvelets is that, they are better localized in both frequency and spatial domain compared to other transforms. This is because the wedge boundary is smoothly tapered to avoid abrupt discontinuity.

2.2 Comparison with wavelets

Fourier series requires a large number of terms to reconstruct a discontinuity within good accuracy. This is the well-known Gibbs phenomenon. Wavelets have the ability to solve this problem of Fourier series, as they are localized and multiscale. However, though wavelets do work well in one-dimension, they fail to represent higher dimensional singularities (especially curved singularities, wavelets can handle point singularities quite well) effectively due to limited orientation selectivity and isotropic scaling. Standard orthogonal wavelet transform has wavelets with primarily vertical, horizontal and diagonal orientations independent of scale.

Curvelet transform has drawn much attention lately because it can efficiently handle several important problems, where traditional multiscale transforms like wavelet fait to act. Firstly, Curvelets can provide a sparse representation of the objects that exhibit '*curve punctuated smoothness*' [Candes, 2003], i.e. objects those are smooth except along a general curve with bounded curvature. Curvelets can model such curved discontinuities so well that the representation becomes as sparse as if the object were not singular. From figure 4, we can have an idea about the sparsity and efficiency of curvelet representation of curved singularities compared to wavelets. At any scale j, curvelets provide a sparse representation $O(2^{j/2})$ of the images compared to wavelets' $O(2^{j})$. If an image function f is approximated by largest m coefficients as $\hat{f}_m$, then the approximation errors are given by:

Fourier transform

$$\left\| f - \hat{f}_m^F \right\|^2 \propto m^{-1/2},\ m \to +\infty$$

Wavelet transform

$$\left\| f - \hat{f}_m^W \right\|^2 \propto m^{-1},\ m \to +\infty$$

Curvelet transform

$$\left\| f - \hat{f}_m^C \right\|^2 \propto m^{-2} \log(m^3),\ m \to +\infty$$

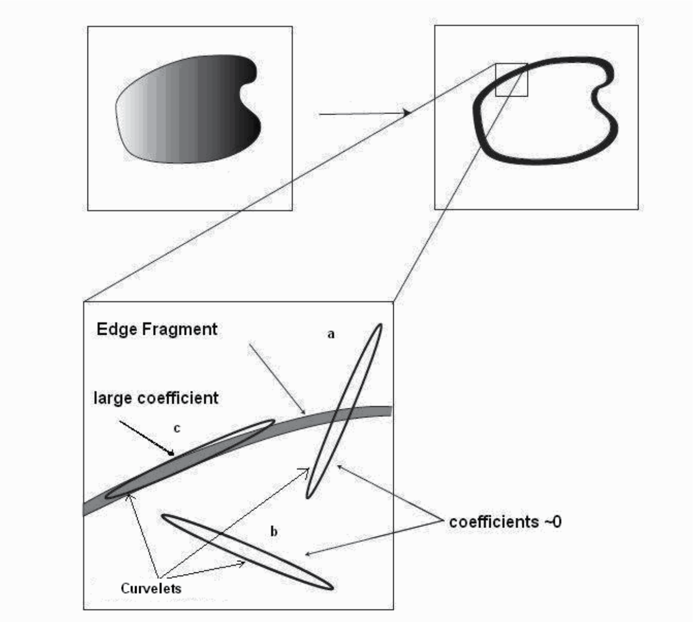


Fig. 3. Alignment of curvelets along curved edges [R]

The main idea here is that the edge discontinuity is better approximated by curvelets than wavelets. Curvelets can provide solutions for the limitations (curved singularity representation, limited orientation and absence of anisotropic element) the wavelet transform suffers from. It can be considered as a higher dimensional generalization of wavelets which have the unique mathematical property to represent curved singularities effectively in a non-adaptive manner.

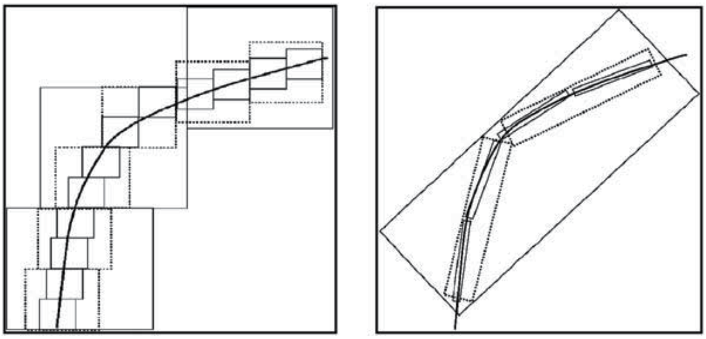


Fig. 4. Representation of curved sigularities using wavelets (left) and curvelets (right) [Starck, 2003].

2.3 Applications

Curvelet transform is gaining popularity in different research areas, like signal processing, image analysis, seismic imaging since the development of FDCT in 2006. It has been successfully applied in image denoising [Starck et al. 2002], image compression, image fusion [Choi et al., 2004], contrast enhancement [Starck et al., 2003], image deconvolution [Starck et al., 2003], high quality image restoration [Starck et al., 2003], astronomical image representation [Starck et al., 2002] etc. Examples of two applications, contrast enhancement and denoising are presented in figures 5 and 6. Readers are suggested to go through the referred works for further information on various applications of the curvelet transform. Recently, curvelets have also been employed to address several pattern recognition problems, such as face recognition [Mandal et al., 2007; Zhang et al., 2007] (discussed in detail in section 3), optical character recognition [Majumdar, 2007], finger-vein pattern recognition [Zhang et al., 2006] and palmprint recognition [Dong et al. 2005].

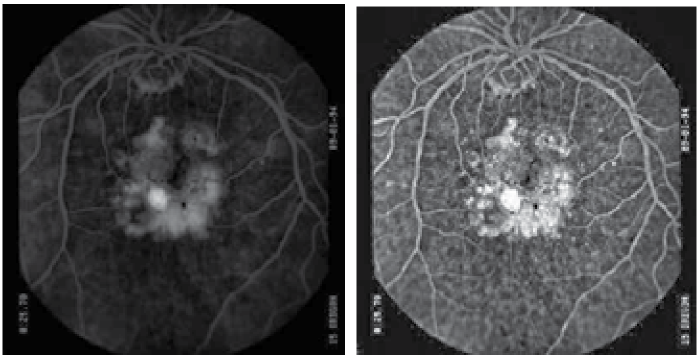


Fig. 5. Contrast enhancement by curvelets [Starck et al., 2003].

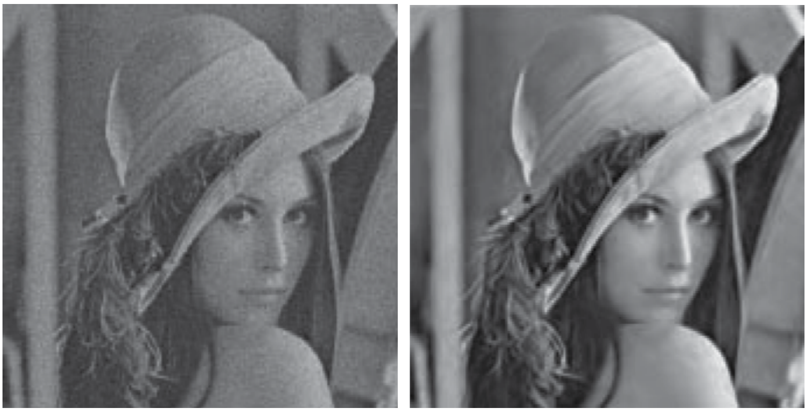


Fig. 6. Image denoising by curvelet [Starck et al. 2002].

3. Curvelet Based Feature Extraction for Faces

In the previous section, we have presented a theoritical overview of curvelet transform and explained why it can be expected to work better than the traditional wavelet transform. Facial images are generally 8 bit i.e. they have 256 graylevels. In such images two very close regions that have differing pixel values will give rise to edges; and these edges are typically curved for faces. As curvelets are good at approximating curved singularities, they are fit for extracting crucial edge-based features from facial images more efficiently than that compared to wavelet transform. We will now describe different face recognition methodologies that employ curvelet transform for feature extraction.

Typically, a face recognition system is divided into two stages: a training stage and a classification stage. In the training stage, a set of known faces (labeled data) are used to create a representative feature-set or template. In the classification stage, a unknown facial image is matched against the previously seen faces by comparing the features. Curvelet based feature extraction takes the raw or the preprocessed facial images as input. The images are then decomposed into curvelet subbands in different scales and orientations. Figure 7 shows the decomposition of a face image of size 112×92 (taken from ORL database) by curvelets at scale 2 (coarse and fine) and angle 8. This produces one approximate (75×61) and eight detailed coefficients (four of those are of size 66×123 and rest are of size 149×54). These curvelet decomposed images are called 'Curveletfaces'. The approximate curveletface contains the low-frequency components and the rest captures the high-frequency details along different orientations. It is sufficient to decompose faces using curvelet transform at scale 3 and angle 8 or 16. Increasing scales and/or orientations does not necessarily lead to significant improvement in recognition accuracy. If required, images can be reduced in size before subjecting them to feature extraction.

3.1 Curvelets and SVM

The first works on curvelet-based face recognition are [Zhang et al., 2007; Mandal et al. 2007]. A simple application of curvelet transform in facial feature extraction can be found in [Zhang et al., 2007]. The authors have used SVM classifier directly on the curvelet decomposed faces. The curvelet based results have been compared with that of wavelets.

Mandal et al. have performed 'bit quantization' before extracting curvelet features. The original 8 bit images are quantized to their 4 bit and 2 bit versions, as shown in figure 8. This is based on the belief that on bit quantizing an image, only bolder curves will remain in the lower bit representations, and curvelet transform will be able to make the most out of this curved edge information. During training, all the original 8 bit gallery images and their two bit-quantized versions are decomposed into curvelet subbands. Selected curvelet coefficients are then separately fed to three different Support Vector Machine (SVM) classifiers. Final decision is achieved by fusing results of all SVMs. The selection of the curvelet coefficients is done on the basis of their variance. The recognition results for these two methods are shown below.

	Curvelet + SVM	Wavelet + SVM
Average Recognition Accuracy	90.44 %	82.57%

Table 1. Face recognition results for ORL database [Zhang et al., 2007]

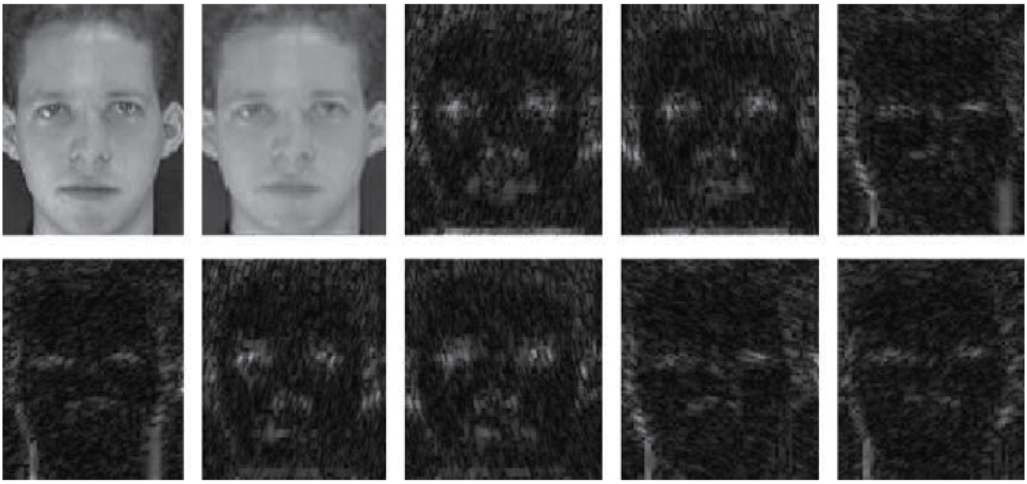


Fig. 7. Curvelet decomposition of a facial image - 1st image in the first row is the original image, 2nd image in the first row is the approximate coefficients and others are detailed coefficients at eight angles (all the images are resized to same dimension for the purpose of illustration only) [Mandal et al., 2009].

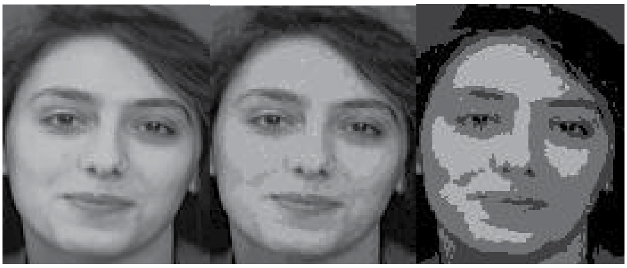


Fig. 8. Bit quantization: left most is the original 8 bit image (from ORL database), next two are 4 bit and 2 bit representations respectively [Mandal et al., 2007].

No. of Bits in Image	Accuracy of each Classifier	Accuracy after majority Voting	Rejection Rate	Incorrect Classification rate
8	96.9	98.8	1.2	0
4	95.6			
2	93.7			

Table 2. Recognition result for bit-quantized, curvelet decomposed images for ORL database [Mandal et al., 2007].

3.1 Curvelets and dimensionality reduction

However, even an image of size 64×64 when decomposed using curvelet transform at scale 3 (coarse, fine, finest) and angle 8 will produce the coarse subband of size 21×21 and 24 detailed coefficients of slightly larger size. Working with such large number of features is extremely expensive. Hence it is important to find a representative feature set. Only important curvelet subbands are selected depending on the amount of total variance they account for. Then dimensionality reduction methods like PCA, LDA and a combined PCA-LDA framework have been applied on those selected subbands to get an even lower dimensional representation [Mandal et al., 2009]. This not only reduces computational load, but also increases recognition accuracy.

The theory of PCA/LDA and will not be discussed here. Readers are requested to consult any standard book and the classical papers of Cootes et al. and Belhumeur et al. to understand the application of PCA and LDA in face recognition. PCA has been successfully applied on wavelet domain for face recognition by Feng et al. PCA has been employed on curvelet decomposed gallery images to form a representational basis. In the classification phase, the query images are subjected to similar treatment and transformed to the same representational basis. However, researchers argue that PCA, though is able to provide an efficient lower dimensional representation of the data, suffers from higher dimensional load and poor discriminative power. This issue can be resolved by the application of LDA that can maximize the within-class dissimilarity, simultaneously increasing the between-class similarity. This efficient dimensionality reduction tool is also applied on curvelet coefficients to achieve even higher accuracy and lower computational load. Often, the size of the training set is less than the dimensionality of the images. In such cases LDA fails to work, since the within-class scatter matrix become singular. Computational difficulty also arises while working with high-dimensional image vectors. In such high-dimensional and singular cases PCA is performed prior to LDA. Curvelet subimages are projected onto PCA-space and then LDA is performed on this PCA-transformed space. Curvelet features thus extracted are also robust against noise. These curvelet-based methods are compared to several existing techniques in terms of recognition accuracy in table 3. Though LDA is expected to work better than PCA that is not reflected in figures 9 and 10. This is because ORL is a small database and PCA can outperform LDA in such cases. In a recent work [Mohammed et al., 2009] Kernal PCA has been used for dimensionality reduction of curvelet features and even higher accuracy is achieved.

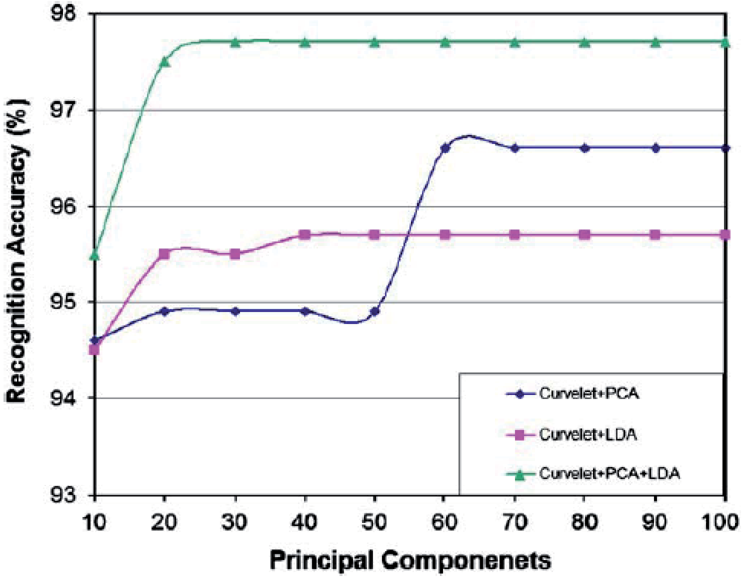


Fig. 9. Curvelet –based recognition accuracy for ORL database [Mandal et al., 2009]

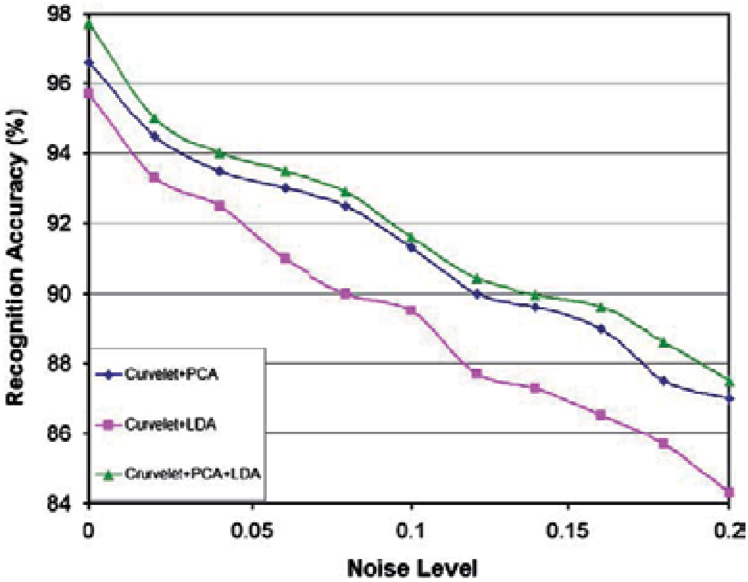


Fig. 10. Performance of curvelet-based methods against noise [Mandal et al, 2009]

Method	Recognition Accuracy (%)
Standard eigenface [Turk et al., 1991]	92.2
Waveletface [Feng et al.]	92.5
Curveletface	94.5
Waveletface + PCA [Feng et al., 2000]	94.5
Waveletface + LDA [Chien and Wu, 2002]	94.7
Waveletface + weighted modular PCA [Zhao et al., 2008]	95.0
Waveletface + LDA + NFL [Chien and Wu, 2002]	95.2
Curveletface + LDA	95.6
Waveletface + kAM [Zhang et al. 2004]	96.6
Curveletface + PCA	96.6
Curveletface + PCA + LDA	97.7

Table 3. Comparative study [Mandal et al., 2009]

4. Conclusion

In this chapter, newly developed curvelet transform has been presented as a new tool for feature extraction from facial images. Various algorithms are discussed along with relevant experimental results as reported in some recent works on face recognition. Looking at the results presented in tables 1, 2 and 3, we can infer that curvelet is not only a successful feature descriptor, but is superior to many existing wavelet-based techniques. Results for only one standard database (ORL) are listed here; nevertheless, work has been done on other standard databases like, FERET, YALE, Essex Grimace, Georgia-Tech and Japanese facial expression datasets. From the results presented in all these datasets prove the superiority of curvelets over wavelets for the application of face recognition. Curvelet features thus extracted from faces are also found to be robust against noise, significant amount of illumination variation, facial details variation and extreme expression changes.

The works on face recognition using curvelet transform that exist in literature are not yet complete and do not fully understand the capability of curvelet transform for face recognition; hence, there is much scope of improvement in terms of both recognition accuracy and curvelet-based methodology.

5. References

Candes, E. J. (2003). What is curvelet, *Notices of American Mathematical Society* vol. 50, pp. 1402–1403, 2003.

Candes, E. J.; Demanet, L.; Donoho, D. L. & Ying, L. (2007). Fast discrete curvelet transform, *SIAM Multiscale Modeling and Simulations*, 2007.

Candes, E. J. & Donoho, D. L. (2000). *Curvelets – A surprisingly effective non- adaptive representation for objects with Edges*, Vanderbilt University Press, Nashville, TN, 2000.

Chien, J. T. & Wu, C. C. (2002). Discriminant waveletfaces and nearest feature classifiers for face recognition, *IEEE Transactions on PAMI,* vol. 24, pp. 1644–1649, 2002.

Choi, M.; Kim, R. Y. & Kim, M. G. (2004). The curvelet transform for image fusion, *Proc ISPRS Congress*, Istanbul, 2004.

Dong, K.; Feng, G. & Hu, D. (2005). Digital curvelet transform for palmprint recognition, *Lecture notes in Computer Science, Springer,* vol. 3338, pp. 639-645, 2005.

Feng, G. C.; Yuen, P. C. & Dai, D. Q. (2000). Human face recognition using PCA on wavelet subband, *Journal of Electronic Imaging,* vol. 9(2), pp. 226–233, 2000.

Majumdar, A. (2007). Bangla basic character recognition using digital curvelet transform, *Journal of Pattern Recognition Research*, vol. 2, pp. 17-26, 2007.

Mandal, T.; Majumdar, A. & Wu, Q. M. J. (2007). Face recognition by curvelet based feature extraction. *Proc. International Conference on Image Analysis and Recognition,* vol. 4633, 2007, pp. 806–817 , Montreal, August, 2007.

Mandal, T.; Yuan, Y.; Wu Q. M. J. (2009). Curvelet based face recognition via dimensional reduction. *Signal Processing*, vol. 89, issue 12, pp. 2345-2353, December, 2009.

Mohammed, A. A.; Minhas, R.; Wu, Q. M. J. & Sid-Ahmed, M. A. (2009). A novel technique for human face recognition using non-linear curvelet feature subspace, *Proc. International Conference on Image Analysis and Recognition,* vol. 5627, July, 2009.

Starck, J. L. (2003). Image Processing by the Curvelet Transform, PPT.

Starck, J. L.; Candes, E. J. & Donoho, D. L. (2000). The curvelet transform for image denosing, *IEEE Transactions on Image Processing,* vol. 11, pp. 670–684, 2000.

Starck, J. L.; Donoho, D. L. & Candes, E. J. (2003). Very high quality image restoration by combining wavelets and curvelets, *Proceedings of SPIE*, vol. 4478, 2003.

Starck, J. L.; Donoho, D. L. & Candes, E. J. (2002). Astronomical image representation by curvelet transform, *Astronomy & Astrophysics,* vol. 398, pp. 785–800, 2002.

Starck, J. L.; Murtagh, F.; Candes, E. J. & Donoho, D. L. (2003). Gray and color image contrast enhancement by the curvelet transform, *IEEE Transactions on Image Processing,* vol. 12, pp. 706–717, 2003.

Starck, J. L.; Nguyen, M. K. & Murtagh, F. (2003). Deconvolution based on the curvelet transform, *Proc. International Conference Image Processing*, 2003.

Turk, M. & Pentland, A. Face recognition using eigenfaces (1991). *Proc. Computer Vision and Pattern Recognition*, pp. 586–591, 1991.

Zhang, J.; Ma, S. & Han, X. (2006). Multiscale feature extraction of finger-vein patterns based on curvelets and local interconnection structure neural network, *Proc. ICPR,* vol. 4, pp. 145-148, 2006.

Zhang, J.; Zhang, Z.; Huang, W.; Lu, Y. & Wang, Y. (2007). Face recognition based on curvefaces, *Proc. Natural Computation*, 2007.

Zhang, B. L.; Zhang, H. & Ge, S. S. (2004). Face recognition by applying wavelet subband representation and kernel associative memory, *IEEE Transactions on Neural Networks*, vol. 15 (1), pp. 166–177, 2004.

Zhao, M.; Li, P. & Liu, Z. (2008). Face recognition based on wavelet transform weighted modular PCA, *Proc. Congress in Image and Signal Processing*, 2008.

4

COMPRESSIVE CLASSIFICATION FOR FACE RECOGNITION

Angshul Majumdar and Rabab K. Ward

1. INTRODUCTION

Face images (with column/row concatenation) form very high dimensional vectors, e.g. a standard webcam takes images of size 320x240 pixels, which leads to a vector of length 76,800. The computational complexity of most classifiers is dependent on the dimensionality of the input features, therefore if all the pixel values of the face image are used as features for classification the time required to finish the task will be excessively large. This prohibits direct usage of pixel values as features for face recognition.

To overcome this problem, different dimensionality reduction techniques has been proposed over the last two decades – starting from Principal Component Analysis and Fisher Linear Discriminant. Such dimensionality reduction techniques have a basic problem – they are data-dependent adaptive techniques, i.e. the projection function from the higher to lower dimension cannot be computed unless all the training samples are available. Thus the system cannot be updated efficiently when new data needs to be added.

Data dependency is the major computational bottleneck of such adaptive dimensionality reduction methods. Consider a situation where a bank intends to authenticate a person at the ATM, based on face recognition. So, when a new client is added to its customer base, a training image of the person is acquired. When that person goes to an ATM, another image is acquired by a camera at the ATM and the new image is compared against the old one for identification. Suppose that at a certain time the bank has 200 customers, and is employing a data-dependent dimensionality reduction method. At that point of time it has computed the projection function from higher to lower dimension for the current set of images. Assume that at a later time, the bank has 10 more clients, then with the data-dependent dimensionality reduction technique, the projection function for all the 210 samples must be recomputed from scratch; in general there is no way the previous projection function can be updated with results of the 10 new samples only. This is a major computational bottleneck for the practical application of current face recognition research.

For an organization such as a bank, where new customers are added regularly, it means that the projection function from higher to lower dimension will have to be updated regularly. The cost of computing the projection function is intensive and is dependent on the number of samples. As the number of samples keeps on increasing, the computational cost keeps on increasing as well (as every time new customers are added to the training dataset, the projection function has to be recalculated from scratch). This becomes a major issue for any practical face recognition system.

One way to work around this problem is to skip the dimensionality reduction step. But as mentioned earlier this increases the classification time. With the ATM scenario there is another problem as well. This is from the perspective of communication cost. There are two possible scenarios in terms of transmission of information – 1) the ATM sends the image to some central station where dimensionality reduction and classification are carried out or 2) the dimensionality reduction is carried out at the ATM so that the dimensionality reduced feature vector is sent instead. The latter reduces the volume of data to be sent over the internet but requires that the dimensionality reduction function is available at the ATM. With the first scenario, the communication cost arises from sending the whole image over the communication channel. In the second scenario, the dimensionality reduction function is available at the ATM. As this function is data-dependent it needs to be updated every time new samples are added. Periodically updating the function increases the communication cost as well.

In this work we propose a dimensionality reduction method that is independent of the data. Practically this implies that the dimensionality reduction function is computed once and for all and is available at all the ATMs. There is no need to update it, and the ATM can send the dimensionality reduced features of the image. Thus both the computational cost of calculating the projection function and the communication cost of updating it are reduced simultaneously.

Our dimensionality reduction is based on Random Projection (RP). Dimensionality reduction by random projection is not a well researched topic. Of the known classifiers only the K Nearest Neighbor (KNN) is robust to such dimensionality reduction [1]. By robust, it is meant that the classification accuracy does not vary much when the RP dimensionality reduced samples are used in classification instead of the original samples (without dimensionality reduction). Although the KNN is robust, its recognition accuracy is not high. This shortcoming has motivated researchers in recent times to look for more sophisticated classification algorithms that will be robust to RP dimensionality reduction [2, 3].

In this chapter we will review the different compressive classification algorithms that are robust to RP dimensionality reduction. However, it should be remembered that these classifiers can also be used with standard dimensionality reduction techniques like Principal Component Analysis.

In signal processing literature random projection of data are called 'Compressive Samples'. Therefore the classifiers which can classify such RP dimensionality reduced data are called 'Compressive Classifiers'. In this chapter we will theoretically prove the robustness of compressive classifiers to RP dimensionality reduction. The theoretical proofs will be validated by thorough experimentation. Rest of the chapter will be segregated into several sections. In section 2, the different compressive classification algorithms will be discussed. The theoretical proofs regarding their robustness will be provided in section 3. The experimental evaluation will be carried out in section 4. Finally in section 5, conclusions of this work will be discussed.

2. CLASSIFICATION ALGORITHMS

The classification problem is that of finding the identity of an unknown test sample given a set of training samples and their class labels. Compressive Classification addresses the case where compressive samples (random projections) of the original signals are available instead of the signal itself.

If the original high dimensional signal is 'x', then its dimensionality is reduced by

$$y = Ax$$

where A is a random projection matrix formed by normalizing the columns of an i.i.d. Gaussian matrix and y is the dimensionality reduced compressive sample. The compressive classifier has access to the compressive samples and must decide the class based on them.
Compressive Classifiers have two challenges to meet:
The classification accuracy of CC on the original signals should be at par with classification accuracy from traditional classifiers (SVM or ANN or KNN).
The classification accuracy from CC should not degrade much when compressed samples are used instead of the original signals.
Recently some classifiers have been proposed which can be employed as compressive classifiers. We discuss those classification algorithms in this section.

2.1 The Sparse Classifier

The Sparse Classifier (SC) is proposed in [2]. It is based on the assumption that the training samples of a particular class approximately form a linear basis for a new test sample belonging to the same class. If $v_{k,test}$ is the test sample belonging to the kth class then,

$$v_{k,test} = \alpha_{k,1}v_{k,1} + \alpha_{k,2}v_{k,2} + \ldots + \alpha_{k,n_k}v_{k,n_k} + \varepsilon_k = \sum_{i=1}^{n_k}\alpha_{k,i}v_{k,i} + \varepsilon_k \tag{1}$$

where $v_{k,i}$'s are the training samples of the kth class and ε_k is the approximation error (assumed to be Normally distributed).
Equation (1) expresses the assumption in terms of the training samples of a single class. Alternatively, it can be expressed in terms of all the training samples such that

$$\begin{aligned} v_{k,test} &= \alpha_{1,1} + \ldots\alpha_{k,1}v_{k,1} + \ldots + \alpha_{k,n_k}v_{k,n_k} + \ldots + \alpha_{C,n_C}v_{C,n_C} + \varepsilon \\ &= \sum_{i=1}^{n_1}\alpha_{1,i}v_{1,i} + \ldots\sum_{i=k}^{n_k}\alpha_{k,i}v_{k,i} + \ldots + \sum_{i=1}^{n_C}\alpha_{C,i}v_{C,i} + \varepsilon \end{aligned} \tag{2}$$

where C is the total number of classes.
In matrix vector notation, equation (2) can be expressed as

$$v_{k,test} = V\alpha + \varepsilon \tag{3}$$

where $V = [v_{1,1} \mid \ldots \mid v_{k,1} \mid \ldots \mid v_{k,n_k} \mid \ldots \mid v_{C,n_C}]$ and $\alpha = [\alpha_{1,1} \ldots \alpha_{k,1} \ldots \alpha_{k,n_k} \ldots \alpha_{C,n_C}]'$.
The linearity assumption in [2] coupled with the formulation (3) implies that the coefficients vector α should be non-zero only when they correspond to the correct class of the test sample.

Based on this assumption the following sparse optimization problem was proposed in [2]

$$\min \|\alpha\|_0 \text{ subject to } \|v_{k,test} - V\alpha\|_2 \le \eta,\ \eta \text{ is related to } \varepsilon \tag{4}$$

As it has already been mentioned, (4) is an NP hard problem. Consequently in [2] a convex relaxation to the NP hard problem was made and the following problem was solved instead

$$\min \|\alpha\|_1 \text{ subject to } \|v_{k,test} - V\alpha\|_2 \leq \eta \qquad (5)$$

The formulation of the sparse optimization problem as in (5) is not ideal for this scenario as it does not impose sparsity on the entire class as the assumption implies. The proponents of Sparse Classifier [2] 'hope' that the l1-norm minimization will find the correct solution even though it is not imposed in the optimization problem explicitly. We will speak more about group sparse classification later.

The sparse classification (SC) algorithm proposed in [2] is the following:

Sparse Classifier Algorithm

1. Solve the optimization problem expressed in (5).
2. For each class (i) repeat the following two steps:
3. Reconstruct a sample for each class by a linear combination of the training samples

$$v_{recon}(i) = \sum_{j=1}^{n_i} \alpha_{i,j} v_{i,j}$$

belonging to that class using.

4. Find the error between the reconstructed sample and the given test sample by $error(v_{test}, i) = \| v_{k,test} - v_{recon(i)} \|_2$.
5. Once the error for every class is obtained, choose the class having the minimum error as the class of the given test sample.

The main workhorse behind the SC algorithm is the optimization problem (5). The rest of the steps are straightforward. We give a very simple algorithm to solve this optimization problem.

IRLS algorithm for l1 minimization

Initialization – set δ(0) = 0 and find the initial $\hat{x}(0) = \min \| y - Ax \|_2^2$ by conjugate gradient method.

At iteration t – continue the following steps till convergence (i.e. either δ is less than 10^{-6} or the number of iterations has reached maximum limit)

1. Find the current weight matricex as $W_m(t) = diag(2\,| x(t-1) + \delta(t) |^{-1/2})$
2. Form a new matrix, $L = AW_m$.
3. Solve $\hat{u}(t) = \min \| y - Lu \|_2^2$ by conjugate gradient method.
4. Find x by rescaling u, $x(t) = W_m u(t)$.
5. Reduce δ by a factor of 10 if ||y-Ax||q has reduced.

This algorithm is called the Iterated Reweighted Least Squares (IRLS) algorithm [4] and falls under the general category of FOCUSS algorithms [5].

2.2 Fast Sparse Classifiers

The above sparse classification (SC) algorithm yields good classification results, but it is slow. This is because of the convex optimization (l1 minimization). It is possible to create faster versions of the SC by replacing the optimization step (step 1 of the above algorithm) by a fast greedy (suboptimal) alternative that approximates the original l0 minimization problem (4). Such greedy algorithms serve as a fast alternative to convex-optimization for sparse signal estimation problems. In this work, we apply these algorithms in a new perspective (classification).

We will discuss a basic greedy algorithm that can be employed to speed-up the SC [2]. The greedy algorithm is called the Orthogonal Matching Pursuit (OMP) [6]. We repeat the OMP algorithms here for the sake of completeness. This algorithm approximates the NP hard problem, $\min \| x \|_0$ subject to $\| y - Ax \|_2 \leq \eta$.

OMP Algorithm

```
Inputs: measurement vector y (mX1), measurement matrix A (mXn) and error tolerance η.
Output: estimated sparse signal x.
Initialize: residual $r_0=y$, the index set $\Lambda_0=\varnothing$, the matrix of chosen atoms $\Phi_0=\varnothing$, and the iteration counter t = 1.
1. At the iteration = t, find $\lambda_t = \arg\max_{j=1...n} |< r_{t-1}, \varphi_j >|$
2. Augment the index set $\Lambda_t = \Lambda_{t-1} \cup \lambda_t$ and the matrix of chosen atoms $\Phi_t = \left[ \Phi_{t-1} A_{\lambda_t} \right]$.
3. Get the new signal estimate $\min_x \| x_t - \Phi_t y \|_2^2$.
4. Calculate the new approximation and the residual $a_t = \Phi_t x_t$ and $r_t = y - a_t$.
Increment t and return to step 1 if $\| r_t \| \geq \varepsilon$.
```

The problem is to estimate the sparse signal. Initially the residual is initialized to the measurement vector. The index set and the matrix of chosen atoms (columns from the measurement matrix) are empty. The first step of each iteration is to select a non-zero index of the sparse signal. In OMP, the current residual is correlated with the measurement matrix and the index of the highest correlation is selected. In the second step of the iteration, the selected index is added to the set of current index and the set of selected atoms (columns from the measurement matrix) is also updated from the current index set. In the third step the estimates of the signal at the given indices are obtained via least squares. In step 4, the residual is updated. Once all the steps are performed for the iteration, a check is done to see if the norm of the residual falls below the error estimate. If it does, the algorithm terminates otherwise it repeats steps 2 to 4.

The Fast Sparse Classification algorithm differs from the Sparse Classification algorithm only in step 1. Instead of solving the *l1 minimization* problem, FSC uses OMP for a greedy approximation of the original *l0 minimization* problem.

2.3 Group Sparse Classifier

As mentioned in subsection 2.1, the optimization algorithm formulated in [2] does not exactly address the desired aim. A sparse optimization problem was formulated in the hope of selecting training samples of a particular (correct) class. It has been shown in [7] that *l1 minimization* cannot select a sparse group of correlated samples (in the limiting case it selects only a single sample from all the correlated samples). In classification problems, the training samples from each class are highly correlated, therefore *l1 minimization* is not an ideal choice for ensuring selection of all the training samples from a group. To overcome this problem of [2] the Group Sparse Classifier was proposed in [3]. It has the same basic assumption as [2] but the optimization criterion is formulated so that it promotes selection of the entire class of training samples.

The basic assumption of expressing the test sample as a linear combination of training samples is formulated in (3) as $v_{k,test} = V\alpha + \varepsilon$

where $V = [v_{1,1} | \ldots | v_{1,n_1} | \ldots | v_{k,1} | \ldots | v_{k,n_k} | \ldots v_{C,1} | \ldots | v_{C,n_C}]$ and

$$\alpha = [\underbrace{\alpha_{1,1},\ldots,\alpha_{1,n_1}}_{\alpha_1},\underbrace{\alpha_{2,1},\ldots,\alpha_{2,n_2}}_{\alpha_2},\ldots\underbrace{\alpha_{C,1},\ldots,\alpha_{C,n_C}}_{\alpha_C}]^T .$$

The above formulation demand that α should be 'group sparse' - meaning that the solution of the inverse problem (3) should have non-zero coefficients corresponding to a particular group of training samples and zero elsewhere (i.e. $\alpha_i \neq 0$ for only one of the αi's, i=1,…,C). This requires the solution of

$$\min_{\alpha} \|\alpha\|_{2,0} \text{ such that } \|\mathrm{v}_{test} - V\alpha\|_2 < \varepsilon \tag{6}$$

The mixed norm $\|\bullet\|_{2,0}$ is defined for $\alpha = [\underbrace{\alpha_{1,1},\ldots,\alpha_{1,n_1}}_{\alpha_1},\underbrace{\alpha_{2,1},\ldots,\alpha_{2,n_2}}_{\alpha_2},\ldots\underbrace{\alpha_{k,1},\ldots,\alpha_{k,n_k}}_{\alpha_k}]^T$ as

$\|\alpha\|_{2,0} = \sum_{l=1}^{k} I(\|\alpha_l\|_2 > 0)$, where $I(\|\alpha_l\|_2 > 0) = 1$ if $\|\alpha_l\|_2 > 0$.

Solving the *$l_{2,0}$ minimization* problem is NP hard. We proposed a convex relaxation in [3], so that the optimization takes the form

$$\min_{\alpha} \|\alpha\|_{2,1} \text{ such that } \|\mathrm{v}_{test} - V\alpha\|_2 < \varepsilon \tag{7}$$

where $\|\alpha\|_{2,1} = \|\alpha_1\|_2 + \|\alpha_2\|_2 + \ldots + \|\alpha_k\|_2$.

Solving the *$l_{2,1}$ minimization* problem is the core behind the GSC. Once the optimization problem (7) is solved, the classification algorithm is straight forward.

Group Sparse Classification Algorithm

1. Solve the optimization problem expressed in (13).
2. Find those i's for which ||αi||2 > 0.
3. For those classes (i) satisfying the condition in step 2, repeat the following two steps:
   a. Reconstruct a sample for each class by a linear combination of the training samples in that class via the equation $v_{recon}(i)=\sum_{j=1}^{n_i}\alpha_{i,j}v_{i,j}$.
   b. Find the error between the reconstructed sample and the given test sample by $error(v_{test},i)=\| v_{k,test}-v_{recon(i)}\|_2$.
4. Once the error for every class is obtained, choose the class having the minimum error as the class of the given test sample.

As said earlier the work horse behind the GSC is the optimization problem (7). We propose a solution to this problem via an IRLS method.

IRLS algorithm for *l2,1 minimization*

Initialization – set δ(0) = 0 and find the initial $\hat{x}(0)=\min\| y-Ax\|_2^2$ by conjugate gradient method.

At iteration t – continue the following steps till convergence (i.e. either δ is less than 10^{-6} or the number of iterations has reached maximum limit)

1. Find the weights for each group (i) $w_i=(\| x_i^{(k-1)}\|_2^2+\delta(t))^{-1/2}$.
2. Form a diagonal weight matrix W_m having weights w_i corresponding to each coefficient of the group x_i.
3. Form a new matrix, $L=AW_m$.
4. Solve $\hat{u}(t)=\min\| y-Lu\|_2^2$.
5. Find x by rescaling u, $x(t)=W_m u(t)$.
6. Reduce δ by a factor of 10 if ||y-Ax||q has reduced.

This algorithm is similar to the one in section 2.1 used for solving the sparse optimization problem except that the weight matrix is different.

2.4 Fast Group Sparse Classification

The Group Sparse Classifier [3] gives better results than the Sparse Classifier [2] but is slower. In a very recent work [8] we proposed alternate greedy algorithms for group sparse classification and were able to increase the operating speed by two orders of magnitude. These classifiers were named Fast Group Sparse Classifiers (FGSC).

FSC is built upon greedy approximation algorithms of the NP hard sparse optimization problem (10). Such greedy algorithms form a well studied topic in signal processing. Therefore it was straightforward to apply known greedy algorithms (such as OMP) to the sparse classification problem. Group sparsity promoting optimization however is not a vastly researched topic like sparse optimization. As previous work in group sparsity solely rely on convex optimization. We had to develop a number of greedy algorithms as (fast and accurate) alternatives to convex group sparse optimization [8].

All greedy group sparse algorithms approximate the problem $\min \| x \|_{2,0}$ subject to $\| y - Ax \|_2 \leq \eta$. They work in a very intuitive way – first they try to identify the group which has non-zero coefficients. Once the group is identified, the coefficients for the group indices are estimated by some simple means. There are several ways to approximate the NP hard problem. It is not possible to discuss all of them in this chapter. We discuss the Group Orthogonal Matching Pursuit (GOMP) algorithm. The interested reader can peruse [8] for other methods to solve this problem.

GOMP Algorithm

```
Inputs: the measurement vector y (mX1), the measurement matrix A (mXn), the group labels and the error tolerance η.
Output: the estimated sparse signal x.
Initialize: the residual $r_0 = y$, the index set $\Lambda_0 = \varnothing$, the matrix of chosen atoms $\Phi_0 = \varnothing$, and the iteration counter t = 1.
1. At iteration t, compute $\lambda(j) = |< r_{t-1}, \varphi_j >|, \forall j = 1...n$
2. Group selection – select the class with the maximum average correlation $\tau_t = \arg\max_{i=1...C} (\frac{1}{n_i} \sum_{j=1}^{n_i} \lambda(j))$, denote it by $class(\tau_t)$.
3. Augment the index set $\Lambda_t = \Lambda_{t-1} \cup class(\tau_t)$ and the matrix of the chosen atoms $\Phi_t = [\Phi_{t-1} \; A_{class(\tau_t)}]$.
4. Get the new signal estimate using $\min_x \| x_t - \Phi_t y \|_2^2$.
5. Calculate the new approximation and the residual $a_t = \Phi_t x_t$ and $r_t = y - a_t$.
Increment t and return to step 1 if $\| r_t \| \geq \varepsilon$.
```

The classification method for the GSC and the FGSC are the same. Only the convex optimization of step of the former is replaced by a greedy algorithm in the latter.

2.5 Nearest Subspace Classifier

The Nearest Subspace Classifier (NSC) [9] makes a novel classification assumption – samples from each class lie on a hyper-plane specific to that class. According to this assumption, the training samples of a particular class span a subspace. Thus the problem of classification is to find the correct hyperplane for the test sample. According to this assumption, any new test sample belonging to that class can thus be represented as a linear combination of the test samples, i.e.

$$v_{k,test} = \sum_{i=1}^{n_k} \alpha_{k,i} \cdot v_{k,i} + \varepsilon_k \tag{8}$$

where $v_{k,test}$ is the test sample (i.e. the vector of features) assumed to belong to the kth class, $v_{k,i}$ is the ith training sample of the kth class, and ε_k is the approximation error for the kth class.

Owing to the error term in equation (8), the relation holds for all the classes *k=1…C*. In such a situation, it is reasonable to assume that for the correct class the test sample has the minimum error ε_k.

To find the class that has the minimum error in equation (8), the coefficients $\alpha_{k,i}$ *k=1…C* must be estimated first. This can be performed by rewriting (8) in matrix-vector notation

$$v_{k,test} = V_k \alpha_k + \varepsilon_k \tag{9}$$

where $V_k = [v_{k,1} \mid v_{k,2} \mid ... \mid v_{k,n_k}]$ and $\alpha_k = [\alpha_{k,1}, \alpha_{k,2} ... \alpha_{k,n_k}]^T$.

The solution to (9) can be obtained by minimizing

$$\hat{\alpha}_k = \arg\min_{\alpha} \| v_{k,test} - V_k \alpha \|_2^2 \tag{10}$$

The previous work on NSC [9] directly solves (10). However, the matrix V_k may be under-determined, i.e. the number the number of samples may be greater than the dimensionality of the inputs. In such a case, instead of solving (10), Tikhonov regularization is employed so that the following is minimized

$$\hat{\alpha}_k = \arg\min_{\alpha} \| v_{k,test} - V_k \alpha \|_2^2 + \lambda \| \alpha \|_2^2 \tag{11}$$

The analytical solution of (11) is

$$\hat{\alpha}_k = (V_k^T V_k + \lambda I)^{-1} V_k^T v_{k,test} \tag{12}$$

Plugging this expression in (9), and solving for the error term, we get

$$\varepsilon_k = (V_k (V_k^T V_k + \lambda I)^{-1} V_k^T - I) v_{k,test} \tag{13}$$

Based on equations (9-13) the Nearest Subspace Classifier algorithm has the following steps.

<u>NSC Algorithm</u>

Training

1. For each class 'k', by computing the orthoprojector (the term in brackets in equation (13)).

Testing

2. Calculate the error for each class 'k' by computing the matrix vector product between the orthoprojector and $v_{k,test}$.
3. Classify the test sample as the class having the minimum error ($\| \varepsilon_k \|$).

3. CLASSIFICATION ROBUSTNESS TO DATA ACQUIRED BY CS

The idea of using random projection for dimensionality reduction of face images was proposed in [1, 2]. It was experimentally shown that the Nearest Neighbor (NN) and the Sparse Classifier (SC) are robust to such dimensionality reduction. However the theoretical understanding behind the robustness to such dimensionality reduction was lacking there in. In this section, we will prove why all classifiers discussed in the previous section can be categorized as Compressive Classifiers. The two conditions that guarantee the robustness of CC under random projection are the following:

Restricted Isometric Property (RIP) [10] – The l2-norm of a sparse vector is approximately preserved under a random lower dimensional projection, i.e. when a sparse vector x is projected by a random projection matrix A, then $(1-\delta)\|x\|_2 \leq \|Ax\|_2 \, (1+\delta)\|x\|_2$. The constant δ is a RIP constant whose value depends on the type of the matrix A and the number of rows and columns of A and the nature of x. An approximate form (without upper and lower bounds) of RIP states $\|Ax\|_2 \approx \|x\|_2$.

Generalized Restricted Isometric Property (GRIP) [11] – For a matrix A which satisfies RIP for inputs x_i, the inner product of two vectors ($<w,v>=\|w\|_2 \cdot \|v\|_2 \cos\theta$) is approximately maintained under the random projection A, i.e. for two vectors x1 and x2 (which satisfies RIP with matrix A), the following inequality is satisfied:

$$(1-\delta)\|x_1\|_2 \cdot \|x_2\|_2 \cos[(1+\sqrt{3}\delta_m)\theta] \leq \langle Ax_1, Ax_2 \rangle \leq (1+\delta)\|x_1\|_2 \cdot \|x_2\|_2 \cos[(1-\sqrt{3}\delta_m)\theta]$$

The constants δ and δm depend on the dimensionality and the type of matrix A and also on the nature of the vectors. Even though the expression seems overwhelming, it can be simply stated as: the angle between two sparse vectors (θ) is approximately preserved under random projections. An approximate form of GRIP is $\langle Ax_1, Ax_2 \rangle \approx \langle x_1, x_2 \rangle$.

RIP and the GRIP were originally proven for sparse vectors, but natural images are in general dense. We will show why these two properties are satisfied by natural images as well. Images are sparse in several orthogonal transform domains like DCT and wavelets. If I is the image and x is the transform domain representation, then

$$I = \Phi^T x \quad \text{synthesis equation}$$

$$x = \Phi I \quad \text{analysis equation}$$

where Φ is the sparsifying transform and x is sparse.

Now if the sparse vector x is randomly projected by a Gaussian matrix A following RIP, then

$$\|Ax\|_2 \approx \|x\|_2$$

$$\Rightarrow \|A\Phi I\|_2 \approx \|\Phi I\|_2 \quad \text{(by anaslysis equation)}$$

$$\Rightarrow \|A\Phi I\|_2 \approx \|I\|_2 \quad (\because \ \Phi \text{ is orthogonal})$$

$$\Rightarrow \|BI\|_2 \approx \|I\|_2, \ \mathrm{B} = \mathrm{A}\Phi$$

Since Φ is an orthogonal matrix, the matrix AΦ (=B) is also Gaussian, being formed by a linear combination of i.i.d. Gaussian columns. Thus it is seen how the RIP condition holds for dense natural images. This fact is the main cornerstone of all compressed sensing imaging applications. In a similar manner it can be also shown that the GRIP is satisfied by natural images as well.

3.1 The Nearest Neighbor Classifier

The Nearest Neighbor (NN) is a compressive classifier. It was used for classification under RP dimensionality reduction in [1]. The criterion for NN classification depends on the magnitude of the distance between the test sample and each training sample. There are two popular distance measures –

Euclidean distance ($\| v_{test} - v_{i,j} \|_2, i = 1...C \text{ and } j = 1...n_i$)

Cosine distance ($\left\langle v_{test}, v_{i,j} \right\rangle, i = 1...C \text{ and } j = 1...n_i$)

It is easy to show that both these distance measures are approximately preserved under random dimensionality reduction, assuming that the random dimensionality reduction matrix A follows RIP with the samples v. Then following the RIP approximation, the Euclidean distance between samples is approximately preserved, i.e.

$$\| Av_{test} - Av_{i,j} \|_2 = \| A(v_{test} - v_{i,j}) \|_2 \approx \| (v_{test} - v_{i,j}) \|_2$$

The fact that the Cosine distance is approximately preserved follows directly from the GRIP assumption

$$\left\langle Av_{test}, Av_{i,j} \right\rangle \approx \left\langle v_{test}, v_{i,j} \right\rangle .$$

3.2 The Sparse and the Group Sparse Classifier

In this subsection it will be shown why the Sparse Classifier and the Group Sparse Classifier can act as compressive classifiers. At the core of SC and GSC classifiers are the l1 minimization and the l2,1 minimization optimization problems respectively

$$\begin{aligned} &\text{SC-}\min \| \alpha \|_1 \text{ subject to } \| v_{k,test} - V\alpha \|_2 \le \eta \\ &\text{GSC-}\min \| \alpha \|_{2,1} \text{ subject to } \| v_{k,test} - V\alpha \|_2 \le \eta \end{aligned} \tag{14}$$

In compressive classification, all the samples are projected from a higher to a lower dimension by a random matrix A. Therefore the optimization is the following:

$$\begin{aligned} &\text{SC-}\min \| \beta \|_1 \text{ subject to } \| Av_{k,test} - AV\beta \|_2 \le \eta \\ &\text{GSC-}\min \| \beta \|_{2,1} \text{ subject to } \| Av_{k,test} - AV\beta \|_2 \le \eta \end{aligned} \tag{15}$$

The objective function does not change before and after projection, but the constraints do. We will show that the constraints of (14) and (15) are approximately the same; therefore the optimization problems are the same as well. The constraint in (15) can be represented as:

$$\begin{aligned} &\| Av_{k,test} - AV\beta \|_2 \le \eta \\ &= \| A(v_{k,test} - V\beta) \|_2 \le \eta \\ &\approx \| (v_{k,test} - V\beta) \|_2 \le \eta, \text{ following RIP} \end{aligned}$$

Since the constraints are approximately preserved and the objective function remains the same, the solution to the two optimization problems (14) and (15) will be approximately the same, i.e. $\beta \approx \alpha$.

In the classification algorithm for SC and GSC (this is also true for both the FSC, FGSC and NSC), the deciding factor behind the class of the test sample is the class-wise error

$$error(v_{test}, i) = \| v_{k,test} - \sum_{j=1}^{n_i} \alpha_{i,j} v_{i,j} \|_2, i = 1...C$$

We show why the class-wise error is approximately preserved after random projection.

$$\begin{aligned} error(Av_{test}, i) &= \| Av_{k,test} - A\sum_{j=1}^{n_i} \alpha_{i,j} v_{i,j} \|_2 \\ &= \| A(v_{k,test} - \sum_{j=1}^{n_i} \alpha_{i,j} v_{i,j}) \|_2 \\ &\approx \| (v_{k,test} - \sum_{j=1}^{n_i} \alpha_{i,j} v_{i,j}) \|_2, \text{ due to RIP} \end{aligned}$$

As the class-wise error is approximately preserved under random projections, the recognition results too will be approximately the same.

Fast Sparse and Fast Group Sparse Classifiers

In the FSC and the FGSC classifiers, the NP hard optimization problem (14) is solved greedily.

$$\text{SC-}\min \| \alpha \|_0 \text{ subject to } \| v_{k,test} - V\alpha \|_2 \le \eta$$

$$\text{GSC-}\min \| \alpha \|_{2,0} \text{ subject to } \| v_{k,test} - V\alpha \|_2 \le \eta$$

The problem (14) pertains to the case of original data. When the samples are randomly projected, the problem has the following form:

$$\begin{aligned} &\text{SC-}\min \| \beta \|_0 \text{ subject to } \| Av_{k,test} - AV\beta \|_2 \le \eta \\ &\text{GSC-}\min \| \beta \|_{2,0} \text{ subject to } \| Av_{k,test} - AV\beta \|_2 \le \eta \end{aligned} \tag{16}$$

We need to show that the results of greedy approximation to the above problems yields $\beta \approx \alpha$.

There are two main computational steps in the OMP/GOMP algorithms – i) the selection step, i.e the criterion for choosing the indices, and ii) the least squares signal estimation step. In order to prove the robustness of the OMP/GOMP algorithm to random projection, it is sufficient to show that the results from the aforesaid steps are approximately preserved.

In OMP/GOMP, the selection is based on the correlation between the measurement matrix Φ and the observations y, i.e. $\Phi^T y$. If we have $\Phi_{m\times n}$ and $y_{m\times 1}$, then the correlation can be written as inner products between the columns of Φ and the vector y i.e. $\langle \phi_i, y \rangle, i = 1...n$. After random projection, both columns of Φ and the measurement y are randomly sub-sampled by a random projection matrix A. The correlation can be calculated as $\langle A\phi_i, Ay\rangle, i = 1...n$, which by GRIP can be approximated as $\langle \phi_i, y\rangle, i = 1...n$ .n Since the correlations are approximately preserved before and after the random projection, the OMP/GOMP selection is also robust under such random sub-sampling.

The signal estimation step is also robust to random projection. The least squares estimation is performed as:

$$\min \| y - \Phi x \|_2 \tag{17}$$

The problem is to estimate the signal x, from measurements y given the matrix Φ.

Both y and Φ are randomly sub-sampled by a random projection matrix A which satisfies RIP. Therefore, the least squares problem in the sub-sampled case takes the form

$$\begin{aligned} &\min \| Ay - A\Phi x \|_2 \\ &= \min \| A(y - \Phi x) \|_2 \\ &\approx \min \| y - \Phi x \|_2, \text{ since RIP holds} \end{aligned}$$

Thus the signal estimate x, obtained by solving the original least squares problem (22) and the randomly sub-sampled problem are approximately the same.

The main criterion of the FSC and the FGSC classification algorithms is the class-wise error. It has already been shown that the class-wise error is approximately preserved after random projection. Therefore the classification results before and after projection will remain approximately the same.

3.3 Nearest Subspace Classifier

The classification criterion for the NSC is the norm of the class-wise error expressed as

$$\| \varepsilon_k \|_2 = \| (V_k (V_k^T V_k + \lambda I)^{-1} V_k^T - I) v_{k,test} \|_2$$

We need to show that the class-wise error is approximately preserved after a random dimensionality reduction. When both the training and the test samples are randomly projected by a matrix A, the class-wise error takes the form

$$\begin{aligned} &\| (AV_k((AV_k)^T(AV_k) + \lambda I)^{-1}(AV_k)^T - I)Av_{k,test} \|_2 \\ &= \| AV_k((AV_k)^T(AV_k) + \lambda I)^{-1}(AV_k)^T Av_{k,test} - Av_{k,test} \|_2 \\ &\approx \| AV_k(V_k{}^T V_k + \lambda I)^{-1} V_k{}^T v_{k,test} - Av_{k,test} \|_2, \text{ since GRIP holds} \\ &= \| A(V_k(V_k{}^T V_k + \lambda I)^{-1} V_k{}^T v_{k,test} - v_{k,test}) \|_2 \\ &\approx \| V_k(V_k{}^T V_k + \lambda I)^{-1} V_k{}^T v_{k,test} - v_{k,test} \|_2, \text{ since RIP holds} \end{aligned}$$

Since the norm of the class-wise error is approximately preserved under random dimensionality reduction, the classification results will also remain approximately the same.

4. EXPERIMENTAL RESULTS

As mentioned in section 2, compressive classifiers should meet two challenges. First and foremost it should have classification accuracy comparable to traditional classifiers. Experiments for general purpose classification are carried out on some benchmark databases from the University of California Irvine Machine Learning (UCI ML) repository [12] to compare the new classifiers (SC, FSC, GSC, FGSC and NSC) with the well known NN. We chose those databases that do not have missing values in feature vectors or unlabeled training data. The results are tabulated in Table 1. The results show that the classification accuracy from the new classifiers are better than NN.

Dataset	SC	FSC	GSC	FGSC	NSC	NN-Euclid	NN-Cosine
Page Block	94.78	94.64	95.66	95.66	95.01	93.34	93.27
Abalone	27.17	27.29	27.17	26.98	27.05	26.67	25.99
Segmentation	96.31	96.10	94.09	94.09	94.85	96.31	95.58
Yeast	57.75	57.54	58.94	58.36	59.57	57.71	57.54
German Credit	69.30	70.00	74.50	74.50	72.6	74.50	74.50
Tic-Tac-Toe	78.89	78.28	84.41	84.41	81.00	83.28	82.98
Vehicle	65.58	66.49	73.86	71.98	74.84	73.86	71.98
Australian Cr.	85.94	85.94	86.66	86.66	86.66	86.66	86.66
Balance Scale	93.33	93.33	95.08	95.08	95.08	93.33	93.33
Ionosphere	86.94	86.94	90.32	90.32	90.32	90.32	90.32
Liver	66.68	65.79	70.21	70.21	70.21	69.04	69.04
Ecoli	81.53	81.53	82.88	82.88	82.88	80.98	81.54
Glass	68.43	69.62	70.19	71.02	69.62	68.43	69.62
Wine	85.62	85.62	85.62	85.95	82.58	82.21	82.21
Iris	96.00	96.00	96.00	96.00	96.00	96.00	96.00
Lymphography	85.81	85.81	86.42	86.42	86.42	85.32	85.81
Hayes Roth	40.23	43.12	41.01	43.12	43.12	33.33	33.33
Satellite	80.30	80.30	82.37	82.37	80.30	77.00	77.08
Haberman	40.52	40.85	43.28	43.28	46.07	57.40	56.20

Table 1. Recognition Accuracy (%age)

The second challenge the Compressive Classifiers should meet is that their classification accuracy should approximately be the same, when sparsifiable data is randomly sub-sampled by RIP matrices. In section 3 we have already proved the robustness of these classifiers. The experimental verification of this claim is shown in table 2. It has already been mentioned (section 3) that images follow RIP with random matrices having i.i.d Gaussian columns normalized to unity.

The face recognition experiments were carried out on the Yale B face database. The images are stored as 192X168 pixel grayscale images. We followed the same methodology as in [2]. Only the frontal faces were chosen for recognition. Half of the images (for each individual) were selected for training and the other half for testing. The experiments were repeated 5 times with 5 sets of random splits. The average results of 5 sets of experiments are shown in table 2. The first column of the following table indicates the number of lower dimensional projections (1/32, 1/24, 1/16 and 1/8 of original dimension).

Dimensionality	SC	FSC	GSC	FGSC	NSC	NN-Euclid	NN-Cosine
30	82.73	82.08	85.57	83.18	87.68	70.39	70.16
56	92.60	92.34	92.60	91.83	91.83	75.45	75.09
120	95.29	95.04	95.68	95.06	93.74	78.62	78.37
504	98.09	97.57	98.09	97.21	94.42	79.13	78.51
Full	98.09	98.09	98.09	98.09	95.05	82.08	82.08

Table 2. Recognition Results (%) on Yale B (RP)

Table 2 shows that the new compressive classifiers are way better than the NN classifiers in terms of recognition accuracy. The Group Sparse Classifier gives by far the best results. All the classifiers are relatively robust to random sub-sampling. The results are at par with the ones obtained from the previous study on Sparse Classification [2].

The compressive classifiers have the special advantage of being robust to dimensionality reduction via random projection. However, they can be used for any other dimensionality reduction as well. In Table 3, the results of compressive classification on PCA dimensionality reduced data is shown for the Yale B database.

Dimensionality	SC	FSC	GSC	FGSC	NSC	NN-Euclid	NN-Cosine
30	83.10	82.87	86.61	84.10	88.92	72.50	71.79
56	92.83	92.55	93.40	92.57	92.74	78.82	77.40
120	95.92	95.60	96.15	95.81	94.98	84.67	82.35
504	98.09	97.33	98.09	98.09	95.66	88.95	86.08
Full	98.09	98.09	98.09	98.09	96.28	89.50	88.00

Table 3. Recognition Results (%) on Yale B (PCA)

Experimental results corroborate our claim regarding the efficacy of compressive classifiers. Results for Table 1 indicate that they can be used for general purpose classification. Table 2 successfully verifies the main claim of this chapter, i.e. the compressive classifiers are robust to dimensionality reduction via random projection. In Table 3, we show that the compressive classifiers are also applicable to data whose dimensionality has been reduced by standard techniques like PCA.

5. CONCLUSION

This chapter reviews an alternate face recognition method than those provided by traditional machine learning tools. Conventional machine learning solutions to dimensionality reduction and classification require all the data to be present beforehand, i.e. whenever new data is added, the system cannot be updated in online fashion, rather all the calculations need to be re-done from scratch. This creates a computational bottleneck for large scale implementation of face recognition systems.

The face recognition community has started to appreciate this problem in the recent past and there have been some studies that modified the existing dimensionality reduction methods for online training [13, 14]. The classifier employed along with such online dimensionality reduction methods has been the traditional Nearest Neighbour.

This work addresses the aforesaid problem from a completely different perspective. It is based on recent theoretical breakthroughs in signal processing [15, 16]. It advocates applying random projection for dimensionality reduction. Such dimensionality reduction necessitates new classification algorithms. This chapter assimilates some recent studies in classification within the unifying framework of compressive classification. The Sparse Classifier [2] is the first of these. The latter ones like the Group Sparse Classifier [3], Fast Group Sparse Classifier [8] and Nearest Subspace Classifier [9] were proposed by us. The Fast Sparse Classifier has been proposed for the first time in this chapter.

For each of the classifiers, their classification algorithms have been written concisely in the corresponding sub-sections. Solutions to different optimization problems required by the classifiers are presented in a fashion that can be implemented by non-experts. Moreover the theoretical understanding behind the different classifiers is also provided in this chapter. These theoretical proofs are thoroughly validated by experimental results.

It should be remembered that the classifiers discussed in this chapter can be used with other dimensionality reduction techniques as well such as – Principal Component Analysis, Linear Discriminant Analysis and the likes. In principle the compressive classifiers can be employed in any classification task as better substitutes for the Nearest Neighbour classifier.

6. REFERENCES

Goel, N., Bebis, G.m and Nefian, A. V., 2005. Face recognition experiments with random projections. SPIE Conference on Biometric Technology for Human Identification, 426-437.

Y. Yang, J. Wright, Y. Ma and S. S. Sastry, “Feature Selection in Face Recognition: A Sparse Representation Perspective”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 1 (2), pp. 210-227, 2009.

A. Majumdar and R. K. Ward, “Classification via Group Sparsity Promoting Regularization”, IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 873-876, 2009.

R. Chartrand, and W. Yin, "Iteratively reweighted algorithms for compressive sensing," ICASSP 2008. pp. 3869-3872, 2008.

B. D. Rao and K. Kreutz-Delgado, “An affine scaling methodology for best basis selection”, IEEE Transactions on Signal Processing, Vol. 47 (1), pp. 187-200, 1999.

Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad. “Orthogonal Matching Pursuit: Recursive function approximation with applications to wavelet decomposition”, Asilomar Conf. Sig., Sys., and Comp., Nov. 1993.

H. Zou and T. Hastie, “Regularization and variable selection via the elastic net”, Journal of Royal Statistical Society B., Vol. 67 (2), pp. 301-320.

A. Majumdar and R. K. Ward, “Fast Group Sparse Classification”, IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, Victoria, B.C., Canada, August 2009.

A. Majumdar and R. K. Ward, “Nearest Subspace Classifier” submitted to International Conference on Image Processing (ICIP09).

E. J. Cand`es and T. Tao, “Near optimal signal recovery from random projections: Universal encoding strategies?”, IEEE Trans. Info. Theory, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

Haupt, J.; Nowak, R., "Compressive Sampling for Signal Detection," Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on, vol. 3, no., pp. III-1509-III-1512, 15-20 April 2007.

http://archive.ics.uci.edu/ml/

T. J. Chin and D. Suter, “Incremental Kernel Principal Component Analysis”, IEEE Transactions on Image Processing, Vol. 16, (6), pp. 1662-1674, 2007.

H. Zhao and P. C. Yuen, “Incremental Linear Discriminant Analysis for Face Recognition”, IEEE Trans. on Systems, Man, And Cybernetics—Part B: Cybernetics, Vol. 38 (1), pp. 210-221, 2008.

D. L. Donoho, “Compressed sensing,” IEEE Transactions on Information Theory, Vol. 52 (4), pp. 1289–1306, 2006.

E. J. Cand`es and T. Tao, “Near optimal signal recovery from random projections: Universal encoding strategies?”, IEEE Transactions on Information Theory, Vol. 52 (12), pp. 5406–5425, 2006.

5

Pixel-Level Decisions based Robust Face Image Recognition

Alex Pappachen James
Queensland Micro-nanotechnology center, Griffith University
Australia

1. Introduction

Face recognition is a special topic in visual information processing that has grown to be of tremendous interest to pattern recognition researchers for the past couple of decades (Delac & Grgic, 2007; Hallinan et al., 1999; Li & Jain, 2005; Wechsler, 2006; Zhao & Chellappa, 2005). However, the methods in general faces the problem of poor recognition rates under the conditions of: (1) large changes in natural variability, and (2) limitations in training data such as single gallery per person problem. Such conditions are undesirable for face recognition as the inter-class and intra-class variability between faces become high, and the room for discrimination between the features become less.

Major methods that are employed to reduce this problem can be classified into three groups: (1) methods whose so-called gallery set consists of multiple training images per person (e.g. Etemad & Chellappa (1997); Jenkins & Burton (2008)) (2) image preprocessing techniques that aim at feature restoration (e.g. Ahlberg & Dornaika (2004)), and (3) use of geometrical transforms to form face models (e.g. Ahlberg & Dornaika (2004)). Even though they show high performance under specific conditions they lack robust performance and in many cases have proved to be computationally expensive. Being distinct from these computational schemes, the human visual system, which is the best available natural model for face recognition, uses modular approach for classification of faces (Moeller et al., 2008).

This chapter presents a method (James, 2008; James & Dimitrijev, 2008) that implements the concept of local binary decisions to form a modular unit and a modular system for face recognition. This method is applied to formulate a simple algorithm and its robustness verified against various natural variabilities occurring in face images. Being distinct from a traditional approach of space reduction at feature level or automatic learning, we propose a method that can suppress unwanted features and make useful decisions on similarity irrespective of the complex nature of underlying data. The proposed method in the process do not require dimensionality reduction or use of complex feature extraction or classifier training to achieve robust recognition performance.

2. Proposed method

Understanding vision in humans at the level of forming a theoretical framework suitable for computational theory, has opened up various disagreements about the goals of cortical processing. The works of David Marr and James Gibson are perhaps the only two major attempts to provide deeper insight. In majority of Marr's work (Marr, 1982), he assumed and believed vision in humans to be nothing more that a natural information processing mechanism, that can be modelled in a computer. The various levels for such a task would be: (1) computational model, (2) a specific algorithm for that model, and (3) a physical implementation. It is logical in this method to treat each of these level as independent components and is a way to mimic the biological vision in robots. Marr attempted to set out a computational theory for vision in a complete holistic approach. He applied the *principle of modularity* to argue visual processing stages, with every module having a function. Philosophically, this is one of the most elegant approach proposed in the last century that can suit both the paradigms of software and hardware implementations. Gibson on the other hand had an *"ecological"* approach to studying vision. His view was that vision should be understood as a tool that enables animals to achieve the basic tasks required for life: avoid obstacles, identify food or predators, approach a goal and so on. Although his explanations on brain perception were unclear and seemed very similar to what Marr explained as algorithmic level, there has been a continued interest in the *rule-based* modeling which advocates knowledge as a prime requirement for visual processing and perception.

Both these approaches have a significant impact in the way in which we understand the visual systems today. We use this understanding by applying the principles of modularity and hierarchy to focus on three major concepts: (1) spatial intensity changes in images, (2) similarity measures for comparison, and (3) decision making using thresholds. We use the following steps as essential for forming a baseline framework for the method presented in this chapter:

Step 1 Feature selection of the image data: In this step the faces are detected and localized. Spatial change detection is applied as a way to normalize the intensity features without reducing the image dimensionality.

Step 2 Local similarity calculation and Local binary decisions: The distance or similarity between the localized pixels from image to another image is determined. This results in a pixel-to-pixel similarity matrix having same size as that of the original image. Inspired from the binary nature of the neuron output we make local decisions at pixel level by using a threshold θ on the similarity matrix.

Step 3 Global similarity and decision: Aggregating all the local decisions, a global similarity score is obtained for the comparisons between a test image with different images. Based on the global similarity scores, they are ranked and the one with the highest similarity score selected as the best match.

These steps are summarised graphically in Fig. 1.

2.1 Feature Selection

The visual features mapped by the colour models used in the camera device are influenced by variations in illumination, spatial motions and spatial noise. Although noise and motion errors can be corrected at the camera itself, illumination correction or normalization is seldom done. The human eye on the other hand has inherent mechanical and functional mechanisms to form illumination invariant face images under a wide range of lighting conditions. Feature localization in humans is handled by feedback mechanisms linked to human eye and brain. However, in the case of automatic face image recognition, a perfect spatial localization of features is not possible using existing methods. Face detection methods are used to detect the face images and localize the feature with some degree of accuracy. Even after features are localised by any automatic detection methods, it is practically impossible to attain a perfect

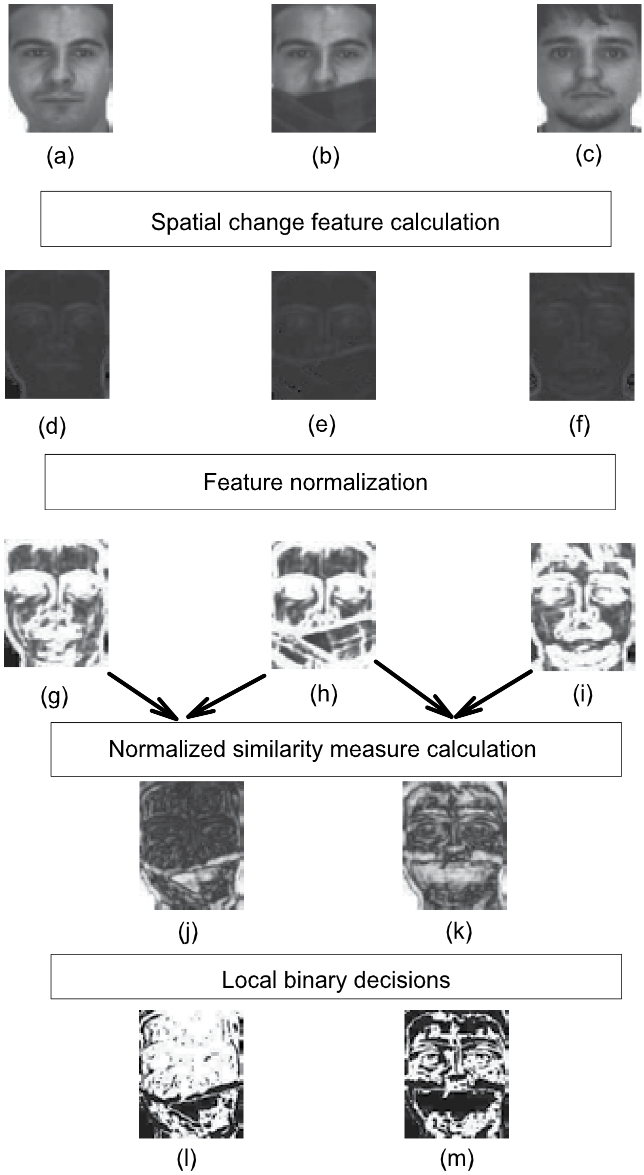


Fig. 1. An illustration of various steps in the baseline algorithm. The images labeled (a),(b), and (c) show the raw images, where (a) and (c) form the gallery images and (b) is a test image, all taken from the AR database (Martinez & Benavente, 1998). The images labeled (d), (e), and (f) show the output of a feature selection process, which corresponds to the raw images (a), (b), and (c), respectively. The normalized feature vectors are shown as the images labeled (g), (h), and (i), and are calculated from (d), (e), and (f), respectively. This is followed by comparisons of test image with gallery images. The normalized similarity measure when applied for comparing (h) with (g) and (h) with (i) results in images labeled (j) and (k), respectively. Finally, the local binary decisions when applied on (j) and (k) result in binary vectors labeled (l) and (m), respectively. Clearly, in this example, (b) is a best match to (a) due to more white areas (more similarity decisions) in (l) than in (m).

alignment due to random occlusions and natural variations that depend on environment. As a result, we need to integrate an error correction mechanism to reduce the impact of localization error by applying image perturbations. The perturbations can be applied with respect to an expected spatial coordinate such as eye coordinates. Ideally, any pixel shift from these expected coordinates results in rotation, scale or shift error. So to undo such errors, by the idea of reverse engineering, pixel shifts are applied to the expected coordinate to detect the face images. In this way any arbitrary N number of pixel shifts on an image results in N number of perturbed images, one of which will be localised the best.

After the raw features are localized, they are processed further to extract features through the detection of spatial change as an essential visual cue for recognition. Spatial change in images can be detected using spatial filtering and normalization mechanisms such as local range filtering, local standard deviation filtering, gradient filtering or gabor filtering.

The relative change of spatial intensity of a pixel in a raw image with respect to the corresponding pixels in its neighbourhood can be used to form features useful for recognition. In the baseline algorithm we can detect such features by calculating the local standard deviation on the image pixels encompassed by a window w of pixels of size $m \times n$ pixels. This type of spatial operation is known as a kernel based local spatial filtering. The local standard deviation filter is given by the following equation:

$$\sigma(i,j) = \sqrt{\frac{1}{mn} \sum_{z=-a}^{a} \sum_{t=-b}^{b} [I(i+z, j+t) - \overline{I(i,j)}]^2} \tag{1}$$

where $a = (m-1)/2$ and $b = (n-1)/2$. The local mean $\overline{I(i,j)}$ used in (1) is calculated by the following equation:

$$\overline{I(i,j)} = \frac{1}{mn} \sum_{s=-a}^{a} \sum_{t=-b}^{b} I(i+s, j+t) \tag{2}$$

In Fig. 1, the images labeled (a), (b), and (c) show the raw images, whereas the images labeled (d), (e), and (f) show the corresponding spatial change features [using Eq. (1)] respectively. The normalized spatial change features $\hat{x}$ are calculated using the following equation:

$$x(i,j) = \frac{\sigma(i,j)}{\overline{\hat{\sigma}}} \tag{3}$$

where the spatial change features σ are normalized using the global mean $\overline{\hat{\sigma}}$. The global mean is calculated by the following equation:

$$\overline{\hat{\sigma}} = \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \sigma(i,j) \tag{4}$$

In Fig. 1, the images labeled (g), (h), and (i) show the normalized spatial change features which is obtained by applying global-mean normalization on spatial change features images labeled (d), (e), and (f), respectively.

An extension to this class of filters is the V1-like features generated from Gabor filters that detect different types of spatial variations in the images. The advantage of Gabor filters for feature extraction in face recognition was evident through the works of (Zhang et al., 2005).

These suggest that like the Gradient filters Gabor filters can be used for preprocessing the images. Formally, Gabor filters are defined as:

$$\psi_{\mu,v}(z) = \frac{\| \, k_{\mu,v} \, \|^2}{\sigma^2} e^{(-\|k_{\mu,v}\|^2 \|z\|^2/2\sigma^2)} [e^{ik_{\mu,v}z} - e^{-\sigma^2/2}] \tag{5}$$

where μ defines the orientation, v defines the scale of the Gabor filters, $k_{\mu,v} = \frac{k_{max}}{\lambda^v} e^{i\frac{\pi\mu}{8}}$, λ is the spacing between the filters in frequency domain and $\| \, . \, \|$ denotes the norm operator. The phase information from these filters is not considered, and only its magnitude explored. For the experiments, we set the value of parameters as follows: $\lambda = \sqrt{2}, \sigma = 2\pi$ and $k_{max} = \pi/2$. Further by considering five scales $v \in \{0, \ldots 4\}$ and eight orientations $\mu \in \{0, \ldots, 7\}$ which on convolution result in 40 filters. Again, these class of filters work on the primary principle of local feature normalization through spatial change detection and provide a way to reduce natural variability present in intensity raw image. Following these filtering operations, the images are normalized using local mean filtering to readjust the signal strength locally.

2.2 Local similarity calculation and binary decisions

What is similarity? This question has eluded researchers from various fields for over a century. Although the idea of similarity seem simple, yet it is very different from the idea of difference. The difficulty lie in the idea of expressing similarity as a quantitative measure, for example, unlike a difference measure such as Euclidean distance there is no physical basis to similarity that can be explained. Although, perception favours similarity, the use of an exact mathematical equation dose not properly justify meaning of similarity.

Type	Equation
Min-max ratio	$min[x_g, x_t]/max[x_g, x_t]$
Difference	$\lvert x_g - x_t)\rvert \, /\gamma$
Exponential difference	$e^{-\lvert x_g - x_t\rvert/\gamma}$
	where γ is $max[x_g, x_t]$ or $[x_g + x_t]/2$ or $min[x_g, x_t]$

Table 1. Normalized similarity measures

The absolute difference between pixels is a well known distance measure used for the comparison of features and can be used to find the similarity. Further, element wise normalization of this similarity measure is done by taking the minimum of each feature within test image x_t and gallery image x_g under comparison. This feature by feature comparison results in a normalized similarity measure δ, which is given by:

$$\delta(i,j) = \frac{|x_g(i,j) - x_t(i,j)|}{min(x_g(i,j), x_t(i,j))} \tag{6}$$

Similarity measures based on this idea of measurement are shown in Table 1. However, they suffer from the inter-feature similarities being detected as true similarities from patterns involving natural variability. We find a way to get around this problem by reducing the inter-feature similarity and maintain only relevant differences through a combination of steps involving local similarity calculation and pixel-level binary decision. Inspired from the idea of ability of neurons to compare and make a binary decision at local level, we apply local similarity measures followed by a local binary decision (see Table 1). In the comparison of images this translates into pixel to pixel local similarity calculation followed by an application of a

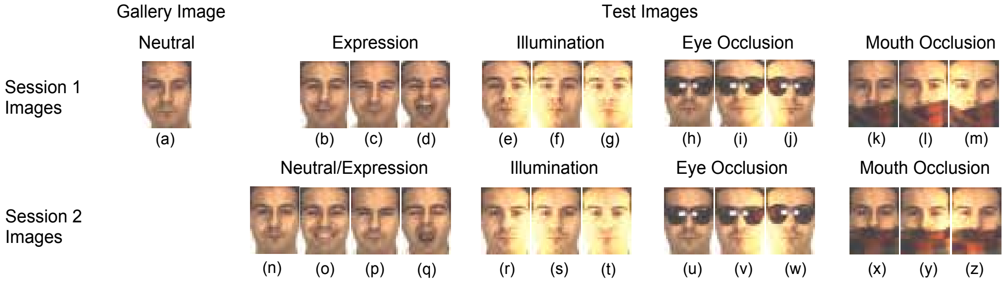


Fig. 2. The illustration shows the images of a person in the AR database (Martinez & Benavente, 1998; 2000) and its organization for the single training sample per person problem depicted in this article. The session 1 image having a neutral facial expression is selected as the gallery image. The remaining 25 images from session 1 and session 2 are used as test images.

binary decision using a threshold. The inherent ability of neurons to exhibit a logic high or logic low state based on the ionic changes occurring due to the presumably threshold limited variations in input connections inspires this idea of local decision. The resulting output for the local similarity measure $S_l(i,j)$ that is defined as to represent $S_l(i,j) = 0$ as least similar and $S_l(i,j) = 1$ as most similar, when applied on a threshold θ to form the binary decision $B(i,j)$ takes the form $B(i,j) = 1$ if $S_l(i,j) <= \theta$ and $B(i,j) = 0$ if $S_l(i,j) > \theta$. The values generated by B represents the local decision space of the image comparison.

2.3 Global similarity and decision

Local decisions on similarity give the similarity match at pixel level, this however is only useful if it can be used at a higher level of decision level abstraction. A reduction of the decision space is necessary to obtain a global value of the image comparison between a test and the gallery image. The simplest possible way to achieve this is by aggregating the local decisions to form a global score which we refer to as global similarity score S_g. The comparison of a test image with any arbitrary M number of gallery images results in M global similarity score S_g. Including the N perturbations done on the test image, this number increases to $M \times N$. These generated similarity scores are then ranked and the top rank is selected to represent the best match. This idea of ranking top rank is no different from threshold logic based decisions at global level (wherein threshold can be thought of being applied between the top rank and second most top rank). Overall, this process represents the global decision making process through a simple approach of global similarity calculation and selection.

3. Experimental Analysis

Unless specified otherwise, all the experiments presented in this section are conducted using the AR face database (See Fig. 2) with the following numerical values: 0.25 for θ, 160×120 pixels for the image size, and 7×5 pixels for the kernel window size of the standard deviation filter.

3.1 Effect of Spatial Intensity Change Used as Features

An analysis using spatial change features and raw features suggest that inter-pixel spatial change within an image is the essential photometric or geometric visual cue that contributes to the recognition of the objects in it. This can be observed from the results presented in Table 2.

The performance analysis using various features with and without mean normalization is shown in Table 2. The importance of spatial change as features for face recognition is analysed by comparing its performance with raw and edge features. For this comparison the standard nearest neighbour (NN) classifier (Cover, 1968; Cover & Hart, 1967; Gates, 1972; Hart, 1968) and the proposed classifier are used.

A raw face image in itself contains all the identity information required for face recognition. However, occurrence of external occlusions, expressions, and illumination in face images can result in loss of such identity information. Further, raw image intensities are highly sensitive to variations in illumination, which make recognition on raw images a difficult task. The comparison shown in Table 2 between spatial change features and raw image features clearly shows that spatial change features outperform the raw features significantly. This superior performance of spatial change features over raw features can be attributed to the facts that spatial change features (1) show lower local variability in the face images under various conditions such as expression, illumination, and occlusion, and (2) preserve the identity information of a face.

Most edge detection techniques are inaccurate approximations of image gradients. Spatial change detection techniques are different from standard edge detection techniques. Majority of the edge detection techniques result in the removal of medium to small texture variations and are distinct from spatial change detection techniques that preserve most of the texture details. Such variations however contain useful information for identification and show increased recognition performance. These observations are shown in Table 2. They further confirm the usefulness of spatial change features in face recognition and show the relative difference of spatial change features as opposed to the edge features.

Figure 3 is a graphical illustration of the overall impact of using spatial change features. The plot shows a normalized histogram of similarity scores S_g resulting from inter-class and intra-class comparisons. The 100 gallery images from the AR database described in the Section 3 form the 100 classes and are compared against 2500 test images in the AR database. The inter-class plots are obtained by comparing each of these test images with the gallery images belonging to a different class, whereas intra-class plots are obtained by the comparison of each test image against a gallery image belonging to its own class. Further, a comparison is done between spatial change features (Fig. 3a) and raw image features (Fig. 3b). The overlapping region of the two distributions indicates the maximum overall probability of error when using the proposed classifier. This region also shows the maximum overall false acceptance and false rejection that can occur in the system. A smaller area of overlap implies better recognition performance. Clearly, it can be seen that the use of feature vectors in Fig. 3a as opposed to the raw-image features in Fig. 3b results in a smaller region of overlap and hence better recognition performance.

An analysis is done to study the effect of using a spatial change filter window w of various sizes [w is described in Section (2.1)]. It can be observed from Fig. 4 that with an increase in resolution of the spatial change features (or the raw image) the recognition performance shows increased stability against variation in spatial change filter window size. Further, it can also be seen that higher resolution images show better recognition accuracies.

3.2 Normalization

The baseline algorithm contains two different types of normalization. They are: (1) global mean normalization of the feature vectors and (2) similarity measure normalization employed in the classifier. The relative importance of using these normalization methods is presented

		Recognition accuracy (%)[a]	
Index	Feature Type	NN Classifier	proposed Classifier
	With global mean normalization[a]		
	Raw features		
r1	Raw	46.0	63.8
	Spatial change features		
s1	Local Standard Deviation	67.6	84.9[b]
s2	Local Range	68.6	84.4
	Edge		
e1	Sobel edges	69.0	80.3
e2	Prewitt edges	69.2	80.4
	Without global mean normalization		
	Raw features		
r2	Raw	38.5	50.8
	Spatial change features		
s1	Local Standard Deviation	59.3	84.7
s2	Local Range	63.0	83.4
	Edge		
e1	Sobel edges	50.4	80.8
e2	Prewitt edges	49.4	80.8

[a] Global mean normalization is achieved using Eq. (3) and Eq. (4). While for raw features normalization is done by replacing $\sigma(i,j)$ with $I(i,j)$ in Eq. (3) and Eq. (4).

[b] proposed baseline algorithm with global mean normalization.

Table 2. Effect of global mean normalization and feature type

in Table 3. It is observed that normalization of the distance measures results in higher recognition accuracies. It can also be observed that global mean normalization shows improved recognition accuracy only when similarity measure normalization is used, which also shows that global mean normalization in isolation does not improve the recognition performance. In the following sections the effect of these two normalization is further studied and alternative methods are attempted. This is done to provide a better technical insight into the normalization methods. This also helps in understanding the unique features that contribute to the overall recognition performance.

3.3 Effect of Mean Normalization and Study of Alternative Normalization

From the experimental results obtained in Table 3, it is found that the normalization of spatial change features by a global mean is not robust against the recognition performance. Clearly, the feature normalization performed by Eq. (3) does not improve the performance considerably, which leads us to investigate alternative local mean normalization techniques. Equation (4) is now replaced by the following equation to calculate the local mean of spatial change

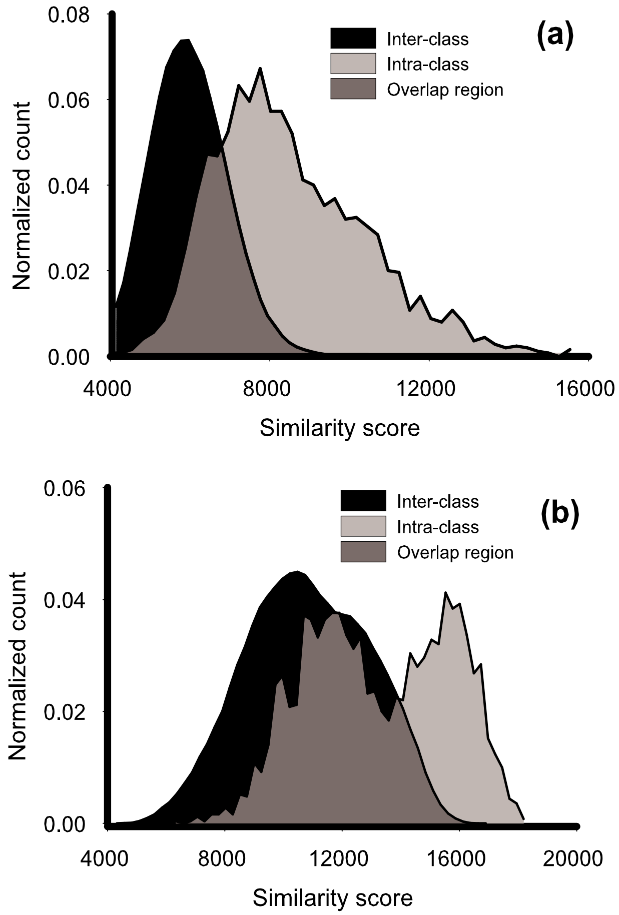


Fig. 3. Graphical illustrations showing the overall influence of using spatial change features. The graphs show a normalized frequency distribution of similarity scores S_g when using (a) spatial intensity change features (b) raw image features.

features:

$$\overline{\sigma(i,j)} = \frac{1}{kl} \sum_{s=-a1}^{a1} \sum_{t=-b1}^{b1} \sigma(i+s, j+t) \tag{7}$$

where the moving window of pixels is of size $k \times l$ pixels, $a1 = (k-1)/2$ and $b1 = (l-1)/2$. Local mean normalization is applied on spatial change features by using Eq. (7) followed by Eq. (3).

An investigation on the performance of using local mean normalization with local mean windows of different sizes is done. Figure 5 shows the effect of variation in local mean window on the recognition performance when using spatial change features and raw features. Further, the same graph shows a comparison of its performance with global mean normalization. It is observed that recognition performance increases when features are normalized using the local mean normalization described by Eq. (7) and Eq. (3). The improvement in recognition

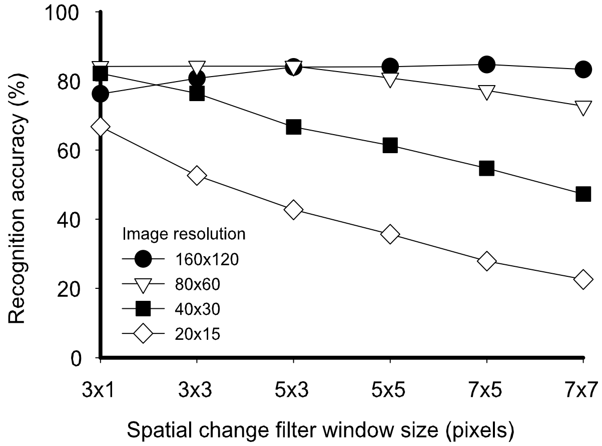


Fig. 4. A graphical illustration showing the recognition performance of the proposed algorithm under the variation of spatial change features filter window size at various image resolutions.

accuracy while using local mean normalization compared to global mean normalization is relatively large in the case of the raw features while having very little impact on spatial change features. Further, in comparison with the raw features, the spatial change features is stable for a broader range of local mean normalization filter window size. The algorithm using spatial change features provides robust performance within the local mean normalization filter window range of 80 × 60 pixels to 40 × 30 pixels as shown in Fig. 4.
Table 4 shows the effect of using local mean normalization on spatial change features. Clearly, in comparison with Table 3, the local mean normalization on spatial change features shows an increase in recognition performance when using the proposed classifier. However, the recognition performance shows no improvement when using an NN classifier. Further, Fig. 5 shows that local mean normalization improves the overall recognition performance and provides a wider stable range of threshold than when using global mean normalization [see Fig. 6 and Fig. 7]. It can be observed that in comparison with global mean normalization on similarity measure, the local mean normalization on similarity measure shows increased stability in recognition accuracy with respect to a varying threshold. All these effects make local mean normalization the preferred choice for use in a feature normalization process.

3.3.1 Effect of Similarity Measure Normalization and Study of Alternative Normalization

Normalization of the similarity measures also helps in increasing the recognition accuracy of the proposed algorithm and enables a stable threshold. This is evident from: (1) Table 3 and Table 4, showing the superiority of similarity measure normalization over mean normalization techniques and (2) Fig. 6 and Fig. 7 showing the relative importance of similarity measure normalization in stabilizing the threshold range and increasing the recognition performance. Further, the improvement of recognition performance provided by normalizing the similarity measure can be observed from Table 5. It can be observed that all of the normalized similarity measures outperform the corresponding direct similarity measures in the recogni-

	Normalization[a]		Recognition accuracy (%)	
Condition	Features	Similarity measure	NN Classifier	proposed Classifier[b]
(a)	Yes	Yes	67.6	84.9
(b)	Yes	No	67.6	76.0
(c)	No	Yes	59.3	84.7
(d)	No	No	59.3	78.4

[a] Feature extraction filter window used in Eq. (2) has a size of 7 × 5 pixels for a raw image I with a size of 160 × 120 pixels. Normalized similarity measure described using Eq. (6) is used for these simulations.

[b] The results are shown for the best accuracies by optimizing the threshold θ. The optimized values of the threshold for the condition indexes (a), (b), (c) and (d) are 0.5, 0.25, 0.35 and 0.85 respectively.

Table 3. Effect of Global Mean Normalization of Features and Similarity Measure Normalization

	Normalization[a]		Recognition accuracy (%)	
Condition	Features	Similarity measure	NN Classifier	proposed Classifier[b]
(a)	Yes	Yes	62.0	86.2
(b)	Yes	No	62.0	81.9
(c)	No	Yes	59.3	84.7
(d)	No	No	59.3	78.4

[a] Feature extraction filter window used in Eq. (2) has a size of 7 × 5 pixels for a raw image I with a size of 160 × 120 pixels. The size of local mean normalization window $w1$ used in Eq. (7) is set to 80 × 60 pixels. Normalized similarity measure described using Eq. (6) is used for these simulations.

[b] The results are shown for the best accuracies by optimizing the threshold θ. The optimized values of the threshold for the normalization conditions (a),(b),(c) and (d) are 0.5, 0.25, 0.35 and 0.85 respectively.

Table 4. Effect of Local Mean Normalization and Distance Normalization

tion accuracy. Fig. 8 shows the influence of variable threshold on the normalized and direct similarity measures. Clearly, for every threshold the normalized similarity measures show better recognition performance than those without similarity measure normalization. These results suggest that normalization of similarity measures is an important factor that helps in improving the recognition performance of the proposed algorithm.

3.3.2 Effect of Local Binary Decisions and Threshold

Binary decisions are made by transforming the normalized similarity measure to a binary decision vector by using a predefined global threshold. A threshold θ is used to set similar features to a value of one, whereas dissimilar features are set to a value of zero. The proposed classifier applies the binary decisions to individual pixels, which means that it can utilize the maximum available spatial change features in the image.

The importance of local binary decisions in the proposed classifier is shown in Fig. 9. The comparison of recognition performance with thresholding and without thresholding shows

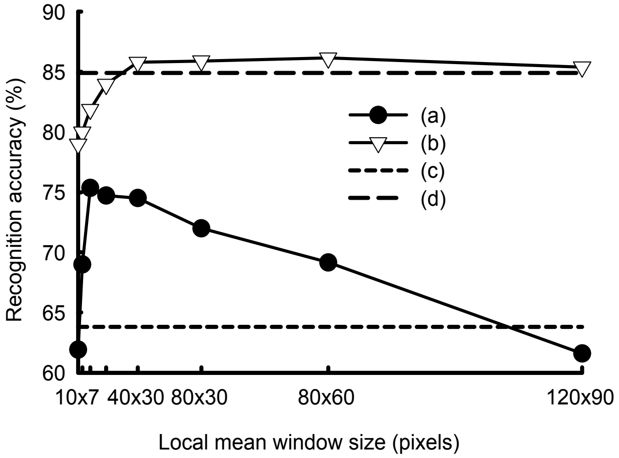


Fig. 5. Graphical illustration showing improved performance of local mean normalization compared to global mean normalization. The graph show the following conditions: **(a)** local mean normalization applied to raw features, **(b)** local mean normalization applied to spatial change features, **(c)** global mean normalization applied to raw features, and **(d)** global mean normalization applied to spatial change features. The image size is 160 × 120 pixels; w is of size 7 × 5 pixels; the local mean filter window size is varied from 10 × 7 pixels to 160 × 120 pixels; for each local mean filter window size the best recognition accuracy is selected by optimizing the threshold. Normalized similarity measure given by Eq. (6) is used for these simulations.

a very large change from 86.2% to 13.8% respectively. This shows the relative importance of local binary decisions, confirming it as the essential component of the algorithm. The local binary decisions result in the removal of noisy information associated with the natural variability. Although, it can be argued that such thresholding results in loss of information, but we find that for natural recognition problems it is the relative number of pixel information in intra-class and inter-class features that would effect the overall performance, and not the individual loss of information due threshold. For example, occlusions and facial expressions remove identity information from the face and can also add information that may seem to be relevant (false similarity) to a non-binary classifier such as the NN classifier. Without the binary decisions, the noisy information gets accumulated when forming a global similarity score (note that similarity scores are formed by adding the values of the elements in the similarity measure vector). Since the global similarity score has significant contribution of such noisy information (or false similarity), the result is a reduced recognition performance. As opposed to this, every feature is used for making local decisions in the case of the proposed classifier. In this case, the global similarity score does not accumulate the effect of less similar features, resulting in a better recognition performance.

Figure 10 shows the performance of the proposed algorithm with a change in threshold when using various normalized similarity measures. We can observe that the recognition accuracy is stable over a broad range of threshold values irrespective of the normalized similarity measures employed. The stability of the threshold and increased recognition performance can be attributed to the use of normalized similarity measures [see Fig. 8]. Further, the stability of

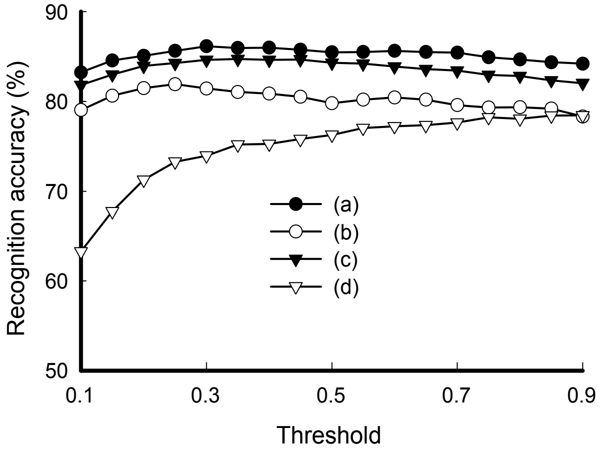


Fig. 6. Graphical illustration showing the effect of local mean normalization and similarity measure normalization on the performance of the proposed algorithm. The graph show the following conditions: **(a)** local mean normalization applied to spatial change features and with normalization similarity measure for comparison, **(b)** local mean normalization applied to spatial change features and with similarity measure without normalization for comparison, **(c)** spatial change features without normalization and with normalized similarity measure comparison, and **(d)** spatial change features without normalization and with similarity measure without normalization for comparison. Normalization of features is performed using global mean normalization of spatial change features using Eq. (4) and Eq. (3). This feature normalization is tried in combination with normalized similarity measure and the performances are compared.

the threshold enables the use of any of the possible similarity measures to form the proposed classifier. A stable threshold in turn implies that the recognition performance of the algorithm is least sensitive to threshold variation. Further, this allows for the use of a single global threshold across different databases containing images of various types of natural variability.

3.3.3 Effect of Resolution

The recognition performance with respect to variation in resolution can be studied by (1) varying the raw image resolution and (2) increasing the decision block size. In the first case, reducing the image resolution from a higher resolution will result in a smaller number of normalized spatial change features. The reduction of a higher resolution image to a lower resolution image can be achieved by averaging a block of pixels to form a single pixel. This averaging results in a loss of features and hence it is natural to expect that recognition performance will drop with lower resolution images which tends to have fewer features. We can observe from Fig. 11 that with lower resolution images the recognition performance drops considerably (this situation is labeled as *average before*).

In the second case, the resolution of spatial change features are kept to a maximum of 160 × 120 pixels, followed by the calculation of δ. The reduction in resolution is achieved by averaging on a block of elements in δ. Block by block reduction across the entire δ results in a lower resolution of δ. This situation is labeled as *average after* in Fig. 11. We can observe

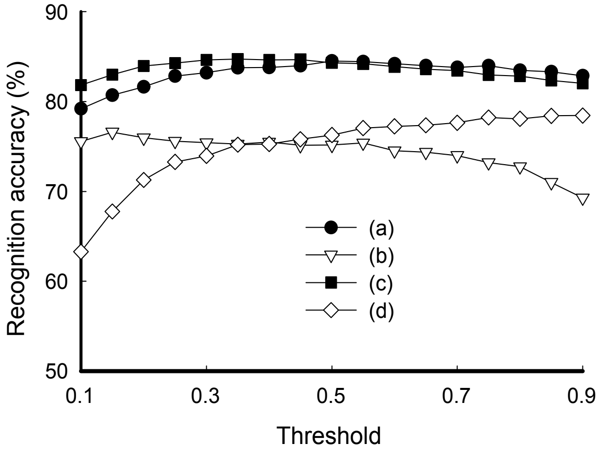


Fig. 7. Graphical illustration showing the effect of global mean normalization and similarity measure normalization on the performance of the proposed algorithm. The graph show the following conditions: **(a)** global mean normalization applied to spatial change features and with normalization similarity measure for comparison, **(b)** global mean normalization applied to spatial change features and with similarity measure without normalization for comparison, **(c)** spatial change features without normalization and with normalized similarity measure comparison, and **(d)** spatial change features without normalization and with similarity measure without normalization for comparison.Normalization of features is performed using global mean normalization of spatial change features using Eq. (4) and Eq. (3). This feature normalization is tried in combination with normalized similarity measure and the performances are compared.

from Fig. 11 that in the case of *average after*, the reduction in resolution results in a slight reduction of the recognition performance, which however, again shows that a larger number of features helps to increase the recognition performance. Further to this, Figure 11 also shows the importance of having a larger number of features irrespective of the decision block size. A larger number of features and a smaller decision block size results in increased recognition performance. Further, as observed from Fig. 4, an increased resolution of features extends the stable range of spatial change filter window size.

3.3.4 Effect of Color

Color images are formed of three channels, namely, red, green, and blue. Table 6 shows that the use of color images also helps to improve the recognition performance. Similarity scores for a comparison between a color test image and a color gallery image can be obtained by one-to-one comparison of red, green, and blue channels of one image to the other. To obtain an overall similarity score, an additive combination of the independent similarity scores observed across the red, green, and blue channels are taken. Table 6 lists some of the combinations that are used in our analysis. Table 6 further illustrates that the use of independent channels alone are not sufficient for robust performance. It can be also observed that utilizing the additive combination of similarity scores obtained from the channels of color images provides a higher recognition accuracy than when using gray images. This can be seen from the

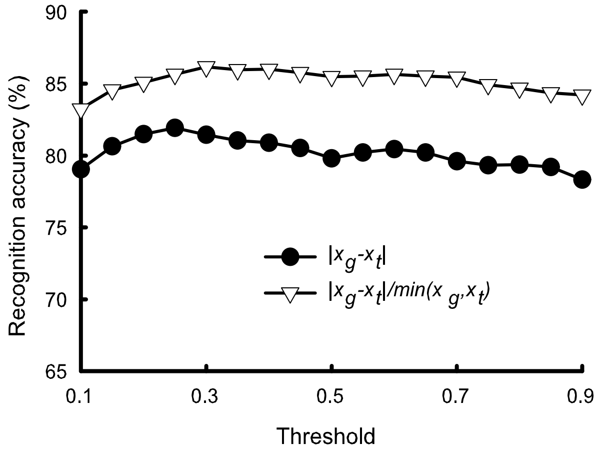


Fig. 8. Graphical illustration showing a comparison of normalized similarity measure with a direct similarity measure. The image size is 160 × 120 pixels; the size of w is 7 × 5 pixels; the size of local mean filter window $w1$ is set to 80 × 60 pixels.

recognition performance of the proposed algorithm when using the combination of the color channels (see c8 listed in Table 6). Although several other combinations can also be tried, analysis is limited to the extend to form a simple model for color, which is achieved through c8 listed in Table 6.

3.3.5 Effect of Localization

Automatic face detection and alignment is a difficult problem when natural variability in images is high. In any method that is based on pixel-by-pixel comparisons, it is essential that the features of the compared images are well aligned. Irrespective of the face detection method employed, natural variability can cause pixel-level misalignments. To compensate for the localization errors that occur after an automatic or manual alignment, we apply either test or gallery image shifts with respect to a set of registration points in the feature vectors. For example, the localization of face images can be achieved by detecting the location of eye coordinates. An error in localization means the eye coordinates are shifted. A scale error means that the eye coordinates are shifted towards each other or away from each other. A rotation error causes shifts of the two eye coordinates in opposite vertical directions. We pertubate the reference eye coordinates by applying such shifts and re-localize the face images using the shifted eye coordinates.

Using the above mentioned idea, two techniques that can be employed to reduce localization errors in the proposed algorithm are (a) application of modifications such as shift, rotation, and scaling on the test image, followed by comparison with gallery, and (b) perturbation of the eye-coordinates of the gallery images to form several sets of synthetic gallery images. In both cases, each comparison of a test image with a gallery image, results in a similarity score S_g^* for the baseline algorithm. The final similarity score S_g for the test image with a compared gallery image is found by selecting the maximum S_g^*. Table 7 shows the recognition performance using both techniques using color and gray scale images. For these simulations the values of number of perturbations is set to 15, composed of 5 horizontal, 5 vertical and 5

Index	Similarity measure[a]	Recognition accuracy (%)[b]
	Normalized	
n1	$\frac{min[x_g,x_t]}{max[x_g,x_t]}$	85.9
n2	$\frac{\lvert x_g-x_t \rvert}{max(x_g,x_t)}$	86.1
n3	$\frac{\lvert x_g-x_t \rvert}{min(x_g,x_t)}$	86.2
n4	$\frac{\lvert x_g-x_t \rvert}{mean(x_g,x_t)}$	86.1
n5	$e^{\frac{-\lvert x_g-x_t \rvert}{max(x_g,x_t)}}$	86.0
n6	$e^{\frac{-\lvert x_g-x_t \rvert}{min(x_g,x_t)}}$	86.1
n7	$e^{\frac{-\lvert x_g-x_t \rvert}{mean(x_g,x_t)}}$	86.1
	Direct	
d1	$\lvert x_g - x_t \rvert$	81.9
d2	$e^{-\lvert x_g-x_t \rvert}$	81.6

[a] Feature extraction filter window used in Eq. (2) has a size of 7 × 5 pixels for a raw image I with a size of 160 × 120 pixels. The size of local mean normalization window $w1$ used in Eq. (7) is set to 80 × 60 pixels.

[b] θ is optimised for highest accuracies on each similarity measure under consideration.

Table 5. Direct and Normalized Similarity Measures

diagonal perturbations. This performance difference is due to the fact that modification of test images is performed after cropping and results in loss of useful spatial information during comparison. This is different from the perturbation of the gallery images that preserves all the information from the original image.

4. Experimental Details

The algorithm is applied to AR (Martinez & Benavente, 1998), ORL(Samaria, 1994), YALE (Belhumeur et al., 1997), CALTECH (Lab, 1999), and FERET (Phillips et al., 2000) standard face image databases. At any specific time, illumination, occlusions, face expressions, and time gap between the gallery and test images form variabilities that make the face recognition difficult. A difficult and practically important face-recognition task is created by limiting the gallery to a single image per person. Unless otherwise specified, the results presented in this chapter are obtained by this kind of open-set testing.

For each image in the AR, YALE, and CALTECH databases, the eye coordinates of the face images are registered manually. For FERET database, the eye coordinates provided in the FERET distribution DVD is used for face alignment. The face alignment is done by rotating, shifting, and scaling the faces so that for all the faces the distance between the eyes remains constant and in fixed spatial coordinates. All the images were aligned and cropped to image size of

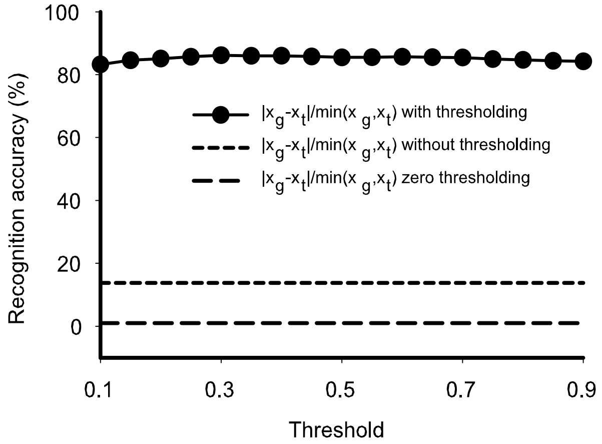


Fig. 9. Graphical illustration showing the effect of local binary decisions. "Without thresholding" is the situation when no threshold is used, which means that no local binary decisions being made. "Zero thresholding" is the situation when the threshold value is set to zero.

160×120.[1] However, as ORL images are approximately localized images, manual alignment are not done on it and are resized to 40 × 32 pixels.

Since the eye coordinates of the faces in AR, Yale, and Caltech databases are detected manually they show shift errors after processing. The eye coordinates of the faces in the gray FERET database are provided within the FERET distribution DVD, and when used, show rotation and scaling errors. Perturbation to eye coordinates are done to compensate for these localization errors. These modifications are in the range of 1 to 6 pixels.

Unless otherwise specified, the following global settings are used for the set of proposed parameters. To calculate spatial intensity change, the local standard deviation filter [see Eq. (1)] is used with optimal window size of 7 × 5 and 3 × 3 pixels when image size is 160 × 120 and 40 × 30 pixels respectively. The min-max similarity ratio shown in Table 1 is used. Finally, the value of the global threshold θ is set to 0.7 which is selected empirically. The number of perturbation used for compensating localization errors in every case is set to a value of 15.

5. Results and Discussion

The overall recognition accuracy for the 2500 gray scale test images and the gallery size of 100 in the AR database is 91%. This very high accuracy level is possible due to the consistent performance over the large number of variable conditions that are individually listed in Table 8. Similar accuracy levels are obtained for YALE, ORL and CALTECH databases as shown in Table 9. As expected, increased variations correspond to decreased recognition accuracies in all databases. The demonstrated robustness of the algorithm is consistent with the fact that the baseline algorithm does not require any prior knowledge of the specific condition that causes the dominant variations. To substantiate the claim of robustness, it is important to report the performance for a large gallery set. In practice, an increased gallery size decreases the overall

[1] This is done using the Unix script provided for face normalization in the CSU Face Identification Evaluation System, Version 5.0 (Beveridge et al. (2003)).

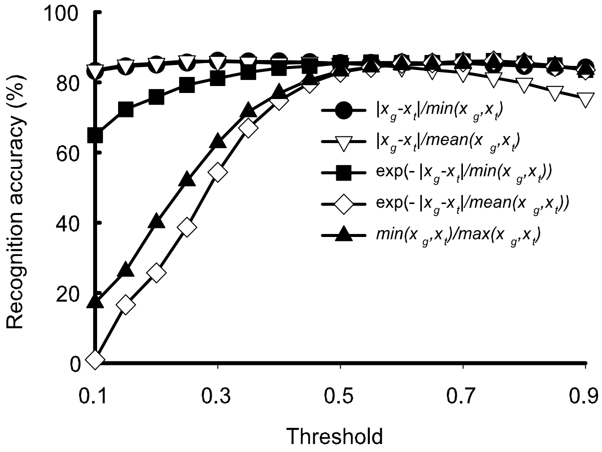


Fig. 10. Graphical illustration showing the stability of the threshold against various normalized similarity measures. The image size is 160 × 120 pixels, the size of the standard deviation filter is 7 × 5 pixels, and the value of the global threshold θ is varied from 0.1 to 0.9.

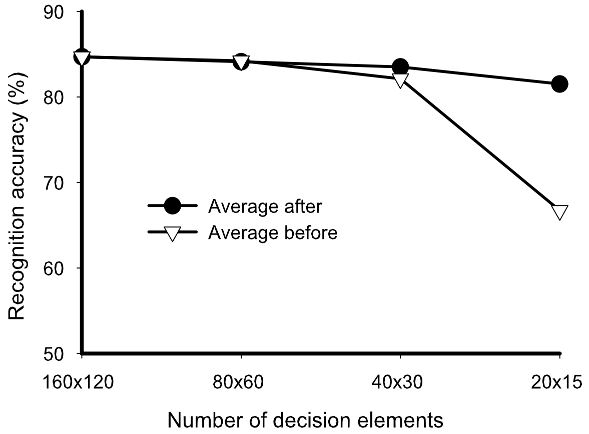


Fig. 11. Graphical illustration showing the recognition performance of the proposed algorithm with variation in resolution of the normalized similarity measure δ under comparison. Averaging is performed to reduce the resolution of δ. *Average before* shows the case when raw images at various resolutions are used, whereas *average after* shows the case when spatial change features at various resolutions are formed from a 160 × 120 pixels raw image.

	Color	Recognition accuracy (%)		
Index[a]	combination	AR (b)-(z)	AR (b)-(m)	AR (n)-(z)
c1	Gray	86.16	94.75	78.23
c2	Red	68.86	76.29	62.00
c3	Green	86.00	95.00	77.69
c4	Blue	87.64	96.33	79.61
c5	Red+Green	81.55	90.16	73.61
c6	Blue+Green	88.96	97.00	81.54
c7	Red+Blue	85.84	95.00	77.38
c8	max(c5,c6,c7)	89.60	97.00	82.76

[a] Similarity score calculated from (c1) gray images, (c2) red channel alone, (c3) green channel alone, (c4) blue channel alone, (c5) combination of scores from red and green channels, (c6) combination of scores from blue and green channels, (c7) combination of scores from red and blue channels, and (c8) the maximum of scores obtained as a result of operations c5 to c7

Table 6. Effect of color on single training samples per person scheme

Color image	Recognition Accuracy (%)		
	Perturbation		
	No	Yes	
		Test image	Gallery image
Yes	89.6	94.0	94.8
No	86.2	91.0	92.0

Table 7. Effect of Localization Error Compensation

recognition accuracy of any face recognition system. The results of testing with the FERET database, also shown in Table 9, demonstrate that the robustness is maintained under this condition.

Using the AR database, the effects of block size used to make the local binary decisions is analyzed and the results are shown in Fig. 12. The maximum recognition accuracy is achieved when the local binary decisions are made at the level of individual pixels (block size of one pixel) with a steep drop in the recognition accuracy as the block size is increased. This directly implies that larger image resolutions could further improve the recognition accuracy.

The impact of different implementations of the similarity measure is also analyzed. Using the implementations listed in Table 1, the change observed in the recognition accuracy is within 1%. Furthermore, the global threshold θ for making the local decisions is not a sensitive parameter. It is found that the recognition accuracy remains within 1% across various databases for a range of threshold values from 0.6 to 0.8. This confirms the general applicability of localised decisions on similarity as a concept.

The impact of the spatial change as features in the baseline algorithm are studied by using raw images as the feature vectors instead of spatial change feature vectors. The recognition accu-

	Recognition accuracy on AR database (%)	
	Localization error compensation	
Test conditions	Yes[a]	No
Session 1 images		
Expression	99	98
Illumination	97	94
Eye occlusion	100	100
Eye occlusion, Illumination	95	80
Mouth occlusion	97	93
Mouth occlusion, Illumination	93	86
Neutral	99	96
Expression	86	80
Illumination	85	80
Eye occlusion	90	83
Eye occlusion, Illumination	77	62
Mouth occlusion	89	74
Mouth occlusion, Illumination	78	60
Overall accuracy	91	84

[a] Proposed algorithm depicted here uses test image perturbations of ±5 pixels.

[b] Results not available from the literature.

Table 8. Recognition performance of the proposed algorithm (Single training sample per person problem) on gray scale images

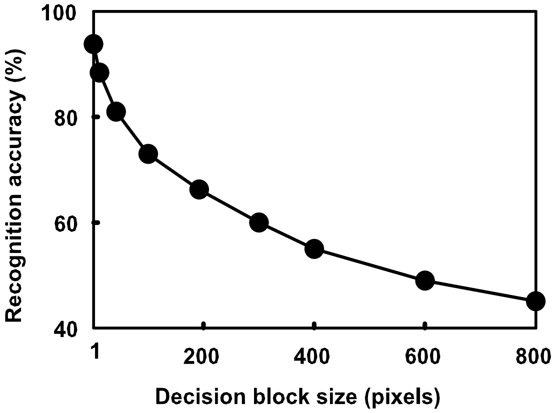


Fig. 12. The dependence of the overall recognition accuracy on the block size used to make the local binary decisions. The resolution of the images is 160×120 pixels. The window size of the standard-deviation filter is 7×5 pixels and the size of the normalization window is 80×60 pixels.

Condition index [a]	Database [b]	Top rank recognition accuracy (%)	
		Localization error compensation	
		No	Yes
(a)	CALTECH	89	95
(a)	YALE	93	95
(b)	ORL	72	84
(c)	FERET:Fb	85	96
(d)	FERET:Fc	71	90
(e)	FERET:Dup I	50	68
(f)	FERET:Dup II	40	65

[a] (a) Expression and illumination with a small gallery; (b) Small pose variation on small gallery (c) Expression on large gallery (Fb); (d) Illumination on large gallery (Fc); (e) Large gallery with mean time gap of 251 days (Dup I); (f) Large gallery with mean time gap of 627 days (Dup II).

[b] Single training image per person is used to form the gallery set. The sizes of the gallery sets are 28 in CALTECH, 15 in YALE, 40 in ORL and 1196 in FERET databases; the sizes of the test sets are 150 in the YALE database, 406 in the CALTECH database, 360 in the ORL database, 1194 in set Fb, 194 in set Fc, 722 in Dup I, and 234 in Dup II of the FERET database.

Table 9. Summary of the results on different databases

racy for the AR database dropped from 91% to 63%. Furthermore, investigation on different filters for calculating the spatial intensity changes shows that the variation of the recognition accuracy with the standard local spatial filters: standard deviation, range and gradient, is within 1%. Based on this and the clear performance difference between the use of raw images and the spatial intensity changes as the feature vectors, it is concluded that the spatial intensity change is the visual cue for face recognition.

Increased number of filters to form feature vectors can further improve the recognition accuracy. As an example, using 40 Gabor filters, the recognition performance on color images in AR database reaches around 97% from a baseline value of 91% on gray images in AR database.

6. Conclusions

In this chapter, the local binary decisions is identified an important concept that is required for recognition of faces under difficult conditions. In addition, spatial intensity changes is identified as the visual cue for face recognition. A baseline algorithm, formed by implementing the local binary decisions based classifier and the spatial intensity changes based feature extractor, shows a robust performance under difficult testing conditions. To increase the recognition performance, a baseline system is formed by including perturbation scheme for localization error compensation. Using this baseline system the effect of localization errors is analysed. Further, the analysis shows that the application of the principles of local binary decisions and modularity results in a highly accurate face recognition system. The presented algorithm does not use any known configurational information from the face images, which makes it applicable to any visual pattern classification and recognition problem. Furthermore, classifiers based on the local binary decisions on similarity can be used in other pattern recognition applications.

7. References

Ahlberg, J. & Dornaika, F. (2004). *Handbook of Face Recognition*, Springer Berlin / Heidelberg.

Belhumeur, P. N., Hespanha, J. P. & Kriegman, D. J. (1997). Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection, *IEEE Trans. Pattern Anal. Machine Intell.* **19**(7): 711–720. Special Issue on Face Recognition.

Beveridge, R., Bolme, D., Teixeira, M. & Draper, B. (2003). The csu face identification evaluation system users guide:version 5.0. Available from http://www.cs.colostate.edu/evalfacerec/algorithms5.html.

Cover, T. M. (1968). Estimation by the nearest-neighbor rule, *IEEE Transactions on Information Theory* **14**(1): 50–55.

Cover, T. M. & Hart, P. E. (1967). Nearest neighbor pattern classification, *IEEE Transactions on Information Theory* **13**(1): 21–27.

Delac, K. & Grgic, M. (2007). *Face Recognition*, I-Tech Education and Publishing, Vienna, Austria.

Etemad, K. & Chellappa, R. (1997). Discriminant analysis for recognition of human face images, *J. Opt. Soc. Am. A* **14**: 1724–1733.

Gates, G. W. (1972). The reduced nearest neighbor rule, *IEEE Transactions on Information Theory* **18**(5): 431–433.

Hallinan, P. W., Gordon, G., Yuille, A. L., Giblin, P. & Mumford, D. (1999). *Two- and Three-Dimensional Patterns of the Face*, AK Peters,Ltd.

Hart, P. E. (1968). The condensed nearest neighbor rule, *IEEE Transactions on Information Theory* **14**(5): 515–516.

James, A. P. (2008). *A memory based face recognition method*, PhD thesis, Griffith University.

James, A. P. & Dimitrijev, S. (2008). Face recognition using local binary decisions, *IEEE Signal Processing Letters* **15**: 821–824.

Jenkins, R. & Burton, A. M. (2008). 100% accuracy in automatic face recognition, *Science* **319**: 435.

Lab, C. V. (1999). Caltech face database. Available from http://www.vision.caltech.edu/html-files/archive.html.

Li, S. Z. & Jain, A. K. (2005). *Handbook of Face Recognition*, Springer Berlin / Heidelberg.

Marr, D. (1982). *Vision*, New York: W.H. Freeman and Company.

Martinez, A. M. & Benavente, R. (1998). The ar face database, CVC Technical Report 24.

Martinez, A. M. & Benavente, R. (2000). Ar face database. Available from http://rvl.www.ecn.purdue.edu/RVL/database.htm.

Moeller, S., Freiwald, W. A. & Tsao, D. Y. (2008). Patches with links: A unified system for processing faces in the macaque temporal lobe, *Science* **320**: 1355–1359.

Phillips, P. J., Moon, H., Rauss, P. J. & Rizvi, S. (2000). The feret evaluation methodology for face recognition algorithms, *IEEE Trans. Pattern Anal. Machine Intell.* **22**: 1090–1104.

Samaria, F. S. (1994). *Face Recognition Using Hidden Markov Models*, University of Cambridge.

Wechsler, H. (2006). *Reliable Face Recognition Methods*, Springer Berlin / Heidelberg.

Zhang, W., Shan, S., Gao, W., Chen, X. & Zhang, H. (2005). Local gabor binary pattern histogram sequence (lgbphs):a novel non-statistical model for face representation and recognition, *Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCVŠ05).*

Zhao, W. & Chellappa, R. (2005). *Face Processing : Advanced Modeling and Methods*, ACADEMIC PRESS.

Interest-Point based Face Recognition System

Cesar Fernandez and Maria Asuncion Vicente
Miguel Hernandez University
Spain

1. Introduction

Among all applications of face recognition systems, surveillance is one of the most challenging ones. In such an application, the goal is to detect known criminals in crowded environments, like airports or train stations. Some attempts have been made, like those of Tokio (Engadget, 2006) or Mainz (Deutsche Welle, 2006), with limited success.

The first task to be carried out in an automatic surveillance system involves the detection of all the faces in the images taken by the video cameras. Current face detection algorithms are highly reliable and thus, they will not be the focus of our work. Some of the best performing examples are the Viola-Jones algorithm (Viola & Jones, 2004) or the Schneiderman-Kanade algorithm (Schneiderman & Kanade, 2000).

The second task to be carried out involves the comparison of all detected faces among the database of known criminals. The ideal behaviour of an automatic system performing this task would be to get a 100% correct identification rate, but this behaviour is far from the capabilities of current face recognition algorithms. Assuming that there will be false identifications, supervised surveillance systems seem to be the most realistic option: the automatic system issues an alarm whenever it detects a possible match with a criminal, and a human decides whether it is a false alarm or not. Figure 1 shows an example.

However, even in a supervised scenario the requirements for the face recognition algorithm are extremely high: the false alarm rate must be low enough as to allow the human operator to cope with it; and the percentage of undetected criminals must be kept to a minimum in order to ensure security. Fulfilling both requirements at the same time is the main challenge, as a reduction in false alarm rate usually implies an increase of the percentage of undetected criminals.

We propose a novel face recognition system based in the use of interest point detectors and local descriptors. In order to check the performances of our system, and particularly its performances in a surveillance application, we present experimental results in terms of Receiver Operating Characteristic curves or ROC curves. From the experimental results, it becomes clear that our system outperforms classical appearance based approaches.

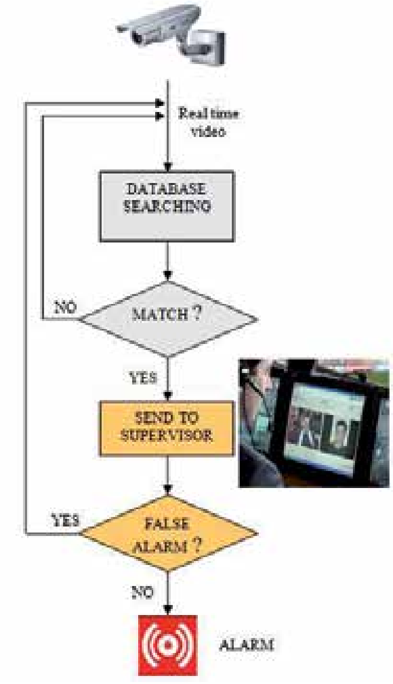


Fig. 1. Example of a supervised surveillance system.

2. Previous approaches

Classical face recognition systems are based on global appearance-based methods: PCA or Principal Component Analysis has been used by (Kirby & Sirovich, 1990) and (Turk & Pentland, 1991); ICA, or Independent Component Analysis has been used by (Bartlett et al., 2002), (Draper et al., 2003) and (Liu, 2004). Finally, LDA or Linear Discriminant Analysis has been used by (Belhumeur et al., 2002).

As an alternative to appearance-based methods, local description methods are currently an area of active research in the face recognition field. From Lowe's work on object recognition using SIFT (Scale Invariant Feature Transform) descriptors (Lowe, 2004), multiple authors have applied such descriptors in other fields, like robot navigation (Se et al., 2001), scene classification (Pham et al., 2007), and also face recognition.

Some of the main contributions using SIFT descriptors for face recognition will be briefly described: Lowe (Lowe, 2000) presents a similar scheme to that of object recognition, but does not address the problem of face authentication. Sivic (Sivic et al., 2005) combines PCA and SIFT: PCA is used to locate eyes, nose and mouth; while SIFT descriptors are used to describe fixed-sized areas around such points. Finally, Bicego (Bicego et al., 2006) measures the distance between two faces as the distance of the best matching pair of descriptors, in some cases using previous knowledge about the location of eyes and mouth.

The goal of our work is to propose a new distance measure in order to exploit the potential of SIFT descriptors in the face recognition field.

3. Interest point detection

Interest point detectors try to select the most descriptive areas of a given image. Ideally, given multiple images of the same object or person, under different lighting, scale, orientation, view angle, etc., a perfect algorithm would find exactly the same interest points across all images.

In the field of face recognition, although invariance to orientation and view angle are necessary, images that are useful for face recognition always present the user from the similar angles (usually, facing the camera) and orientations (standing up). Possible view angles and orientations are expected to be within a 30 degree range, approximately. Interest point detectors that allow much higher ranges of variation are not necessary and more simple, faster detectors would be preferred instead.

In that sense, affine invariant detectors, like those detailed in (Alvarez & Morales, 1997), (Baumberg, 2000) or (Mikolajczyk & Schmid, 2004) are not considered for our work. We have made experiments with two more simple detectors: Harris-Laplace (Mikolajczyk & Schmid, 2004) and Difference of Gaussian (Lowe, 2004).

The Harris-Laplace detector is a scale-invariant version of the well-known Harris corner detector (Harris & Stephens, 1988) and looks for corners or junctions in the images. On the other side, the Difference of Gaussian detector (DoG) is an approximation to the Laplacian of Gaussian operator, and looks for blob-like areas in images. Both detectors have been widely used in the object recognition field and they are highly reliable. In figure 2 we show the interest points found by each of these detectors over the same image (the diameter of the circle is represents the scale of the detected interest area).

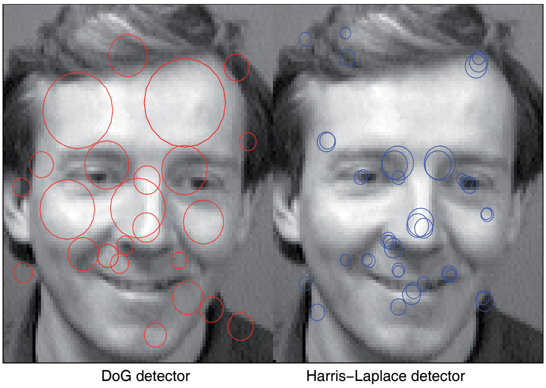


Fig. 2. Output of Harris-Laplace and DoG interest point detectors.

It becomes clear that each detector looks for specific image areas, and that, depending on the particular application, one of them should be preferred. In the case of face recognition, both sets of interest points seem to be relevant for describing faces, so our option has been to keep all interest points found by both detectors. The goal is to obtain as much information as possible from each image.

4. Interest point description

Once interest points are detected, their surrounding area must be encoded or described by a distinctive feature. Ideally, features should be invariant to lighting, scale, orientation, view angle, etc. At the same time, those features should be unique, in the sense that a different area of the object (or face), a different object, or a different person would be distinguishable.

In (Mikolajczyk & Schmid, 2005) a detailed comparison of local descriptors is carried out. The conclusion is that SIFT (Lowe, 2004) and other SIFT-like descriptors, like PCA-SIFT (Ke & Sukthankar, 2004) or GLOH (Mikolajczyk & Schmid, 2005) give the best results throughout all tests. We will briefly describe some of these descriptors.

Basically, in SIFT descriptors, the neighbourhood of the interest point, scaled accordingly to the detector information, is described as a set of orientation histograms computed from the gradient image. SIFT descriptors are invariant to scale, rotation, lighting and viewpoint change (in a narrow range). The most common implementation uses 16 histograms of 8 bins (8 orientations), which gives a 128 dimensional descriptor.

PCA-SIFT descriptor is also based on the gradient image, the main difference with SIFT being the further compression using PCA. The uncompressed dimension of the descriptor is 3042 (39x39), which is reduced to 36 after applying PCA. The authors claim improved accuracy and faster matching, but these performance improvements are not consistent throughout all tests, as it is shown in (Mikolajczyk & Schmid, 2005).

GLOH stands for Gradient Location-Orientation Histogram. It is also a SIFT-based descriptor, with modified location grids (both polar and Cartesian location grids are considered) and a further PCA compression of the information, which keeps the 128 largest eigenvectors (the dimension of the uncompressed descriptor is 272). GLOH outperforms SIFT in certain situations, with structured scenes and high viewpoint changes. However, such situations are not common in a face recognition scenario.

Recently, the SURF or Speeded Up Robust Features descriptor (Bay et al., 2006) has appeared as an alternative to SIFT. Its main advantage is its fastest computation, while keeping a high descriptive power. It is partially inspired by SIFT, but instead of using the gradient image, it computes first order Haar wavelet responses. Additionally, the use of integral images is the key factor for fast computation. So far, we have not performed tests with the SURF descriptor, so we cannot affirm its validity for face recognition applications.

Finally, LESH or Local Energy based Shape Histogram descriptor (Sarfraz & Hellwich, 2008), has been specifically designed for face recognition applications. Its goal is to encode the underlying shape present in the image. Basically, the descriptor is a concatenation of histograms obtained by accumulating local energy along several filter orientations.

However, it is focused in pose estimation, so it addresses a different problem to that of our work.

In conclusion, we decided to describe each face image with SIFT descriptors computed at all the interest points found by the Harris-Laplace and the DoG detectors.

5. Similarity between two face images

Once we have represented all face images as a set of interest points and their corresponding descriptions, the next step to be carried out is the definition of a similarity measure between two face images, in order to be able to decide whether such images correspond to the same person or not.

The simplest approach is to obtain the best possible correspondence between the interest points of both images (according to the values of their SIFT descriptors) and to compute Euclidean distances between each pair of corresponding points. However, according to Lowe's work (Lowe, 2004), SIFT descriptors must be used in a slightly different way: in order to decide whether two points in two different images correspond or not, the absolute value of the Euclidean distance is not reliable; what should be used instead is the ratio between the best match and the second best match. Briefly, for each point of the first image, the best and second best matching points of the second image must be found: if the first match is much better than the second one (as measured by the ratio between SIFT differences) the points are likely to correspond. Eq. 1 shows how to apply such condition, where points ***B*** and ***C*** in image$_2$ are the best and second best matches, respectively, for point ***A*** in image$_1$.

$$\frac{\left|\mathrm{SIFT}^{A}_{\mathrm{image}_1} - \mathrm{SIFT}^{B}_{\mathrm{image}_2}\right|}{\left|\mathrm{SIFT}^{A}_{\mathrm{image}_1} - \mathrm{SIFT}^{C}_{\mathrm{image}_2}\right|} < \text{Threshold} \qquad \mathrm{A}, \mathrm{image}_1 \text{ corresponds to } \mathrm{B}, \mathrm{image}_2 \qquad (1)$$

We have used such approach in order to compute the number of corresponding points between two images, such a number being our first measure of similarity between the images. In our notation, the number of matching points between images ***A*** and ***B***, according to the descriptor values is $\boldsymbol{MD_{AB}}$.

Even though we compute similarity according to Lowe's recommendations, the number of correct matches is not completely reliable as a measure of similarity. We have added two extra measures in order to increase the robustness of our system.

The first extra measure is obtained by computing the number of corresponding points that are coherent in terms of scale and orientation: every detected point output by the Harris-Laplace of DoG detectors has an associated scale and orientation. Scale and orientation may be different between images, even if they belong to the same person, but such difference must be coherent across all matching points. Our second measure of similarity is the number of matching points coherent in terms of scale and orientation (a simple Hough transform is used to obtain the maximum number of points fulfilling this condition). We will refer to this extra measure as $\boldsymbol{MSO_{AB}}$.

The second extra measure is obtained by imposing an additional restriction: the coherence in terms of relative location of corresponding points. Theoretically, the relative location of all matching points must be similar between two images, even if there are scale, rotation and viewpoint changes between them. We will consider a general affine transformation between images for the sake of simplicity (since faces are not planar, high viewpoint changes cannot be represented by affine transformations). The number of points coherent in the parameters of the transformation will be our third measure of similarity. We will use $\boldsymbol{MRL_{AB}}$ to refer to this measure.

Obviously, whenever an additional restriction is imposed, the robustness of the measure is increased, so the second extra measure is the most robust one, followed by the first extra measure and by the original one. In order to compare whether a certain image ***A*** is more similar to image ***B*** or to image ***C*** (i.e., we are trying to classify image ***A*** as belonging to subject ***B*** or subject ***C***), the decision tree of Fig. 3 should be used:

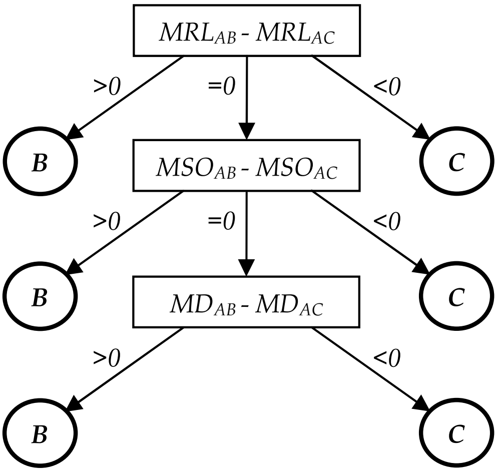


Fig. 3. Decision tree for the classification of image A as belonging to subjects B or C.

Even though a decision tree representation is valid for a classification problem, it cannot be used in an authentication application, where a threshold must be fixed. In order to cope also with such applications, we propose a simple distance measure ***M*** (see eq. 2) that combines ***MRL***, ***MSO*** and ***MD***, giving ***MRL*** a weight one order of magnitude above ***MSO*** and two orders of magnitude above ***MD***.

$$M_{AB} = MD_{AB} + 10MSO_{AB} + 100MRL_{AB} \tag{2}$$

In our experiments, such simple distance measure has shown to give the same results as the decision tree of Fig. 3.

6. Experimental results

6.1 Databases and baseline used for the evaluation

We have selected two different databases for the evaluation of our face recognition algorithm. The first one is the well-known AT&T database (Samaria, 1994)(AT&T, 2002); and the second one the LFW or Labelled Faces in the Wild database (Huang et al., 2007)(University of Massachusetts, 2007).

The AT&T database contains 40 subjects, each of one described by 10 frontal face images. All images were taken under controlled conditions of lighting, distance to the camera, etc. The main differences between shots are facial expression, and slight orientation and viewpoint changes.

The LFW database contains 5749 subjects, described by a number of images that ranges from 1 to 530. All images have been obtained from the World Wide Web, manually labelled and cropped using the Viola-Jones face detector. Variability between images of the same subject is much higher than that of the AT&T database, thus making LFW more challenging for a face recognition application. For our tests, we have selected a subset containing the 158 subjects described by at least 10 images, and we have kept only the first 10 images of each subject.

As the baseline for our evaluation, we have selected the classic PCA approach to face recognition. We have decided to use PCA because other similar approaches like ICA or LDA have not proved to perform better. In particular in one of our previous papers (Vicente et al., 2007) we showed the equivalence of PCA and ICA under restrictions such as the use of rotational invariant classifiers.

6.2 Results

As the goal of our paper is to evaluate our face recognition method for surveillance applications, we have decided to use ROC curves for showing our experimental results. The main idea is to express the relationship between false alarm rates and percentage of undetected criminals. As both databases (AT&T and the LFW subset we are using) share the same structure of 10 images per subject, in both cases we used 4 subjects for training and the remaining 6 subjects for testing. Every test image was compared to all training images of all subjects, the global distance to a subject being computed as the minimum across the 4 training images of such subject (we performed some tests using the mean distance for all training images of the subject, but the results were worse).

First, we performed some experiments in order to adjust our algorithm for the best overall results. The main parameter to tune was the threshold for accepting or rejecting matches between interest points of two different images (see Eq. 1). We carried out tests with both databases (AT&T and LFW) and with thresholds ranging from 0.60 (the most restrictive) to 1.00 (the less restrictive, all matches are accepted).

Figure 4 shows the results obtained with the AT&T database. The left plot shows the full ROC curve, where the different experiments are almost indistinguishable. All of them show close to ideal behaviours, as it was expected for such database, were all images were taken under controlled conditions. In order to show the differences between the experiments, the right plot shows a scaled detail of the upper left area of the ROC curve. However, all values of the threshold seem to perform similarly and no conclusions can be drawn.

Figure 5 shows the results obtained for the LFW database. In this case, we performed two different experiments. For the first one (left plot), we used the first four images of each subject as training images, keeping the original image order of the database. The results show clearly that the LFW database is much more challenging for a face recognition algorithm, with ROC curves far from the ideal ones. The main reason is the high variability between images of the same subject. For the second experiment (right plot) we decided to rearrange the database, so that the best four images of each subject were selected as training data. Such rearrangement makes the experiment more realistic, since in a real surveillance application training images (taken under controlled conditions) usually have higher quality than test images (taken under uncontrolled conditions, in real time). Anyway, the results are similar in both experiments.

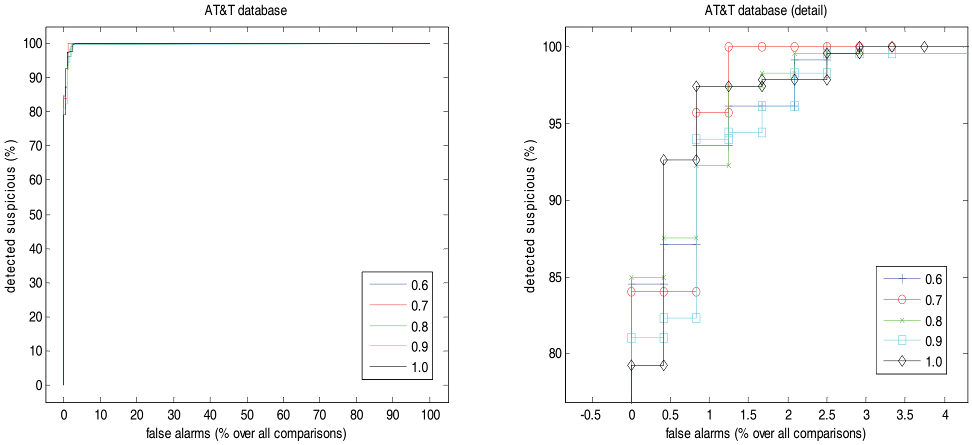


Fig. 4. AT&T database: experiments with different thresholds

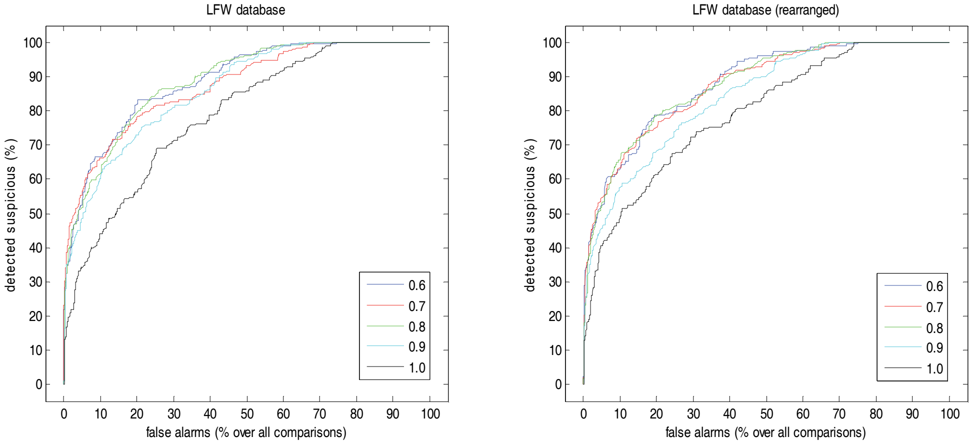


Fig. 5. LFW database: experiments with different thresholds

Concerning the threshold values, both plots of figure 5 show similar behaviours: the results improve as the threshold is reduced, up to a certain point where differences are small (in the range from 0.8 to 0.6). A threshold value of 0.6 seems to perform slightly better than the other settings, so we kept this value for further experiments.

Once the threshold was fixed, we performed several comparisons between our algorithm and the PCA baseline, working with the same databases. Figure 6 shows the results obtained for the AT&T database (left plot) and the LFW database (right plot). Our method clearly outperforms PCA throughout the ROC curve for both databases.

The left plot of figure 7 shows the comparison between PCA and our method for the rearranged version of LFW database (the 4 best images are used for training). There is a slight increase in the performances of both PCA and SIFT, but our method is still clearly superior. Finally, the right plot of figure 7 shows a further experiment: we sorted the 6 test images of LFW for each subject, so that the first 3 images were the best, easier to classify and the last 3 images were the worst, more difficult to classify, according to our opinion. The goal was to check to what extent the performances of both methods were affected by the (subjective) quality of the images: although there are not big differences, it seems that our method is more robust than the PCA approach.

Concerning the feasibility of the proposed approach for a surveillance application, our experimental results show the importance of image quality. The ROC curves obtained for the AT&T database (figure 6, left plot) are close to ideal: at a false alarm rate of 1%, it is expected that 94% of the criminals would be correctly identified (88% at a 0.5% false alarm rate). Such performances would allow us to implement the system in a real scenario. However, the ROC curves obtained for the LFW database, even if the best images are selected for training, are far from ideal: at a false alarm rate of 1%, it is expected that only 35% of the criminals would be correctly identified (32% at a false alarm rate of 0.5%). Such a system would be of little help as a surveillance tool.

As image quality is a key factor for the feasibility of the system, our recommendation is to study properly the location of the video cameras. In our opinion, if video cameras are located in relatively controlled places like walkthrough detectors, the image quality may be enough as for a successful implementation of a supervised surveillance system.

7. Conclusion

Automatic or supervised surveillance applications impose strict requirements in face recognition algorithms, in terms of false alarm rate and percentage of undetected criminals. We present a novel method, based on interest point detectors (namely, Harris-Laplace and DoG) and SIFT descriptors.

Our measure of similarity between images is based on computing the number of corresponding points that, apart from having similar values for their SIFT descriptors, fulfil scale, orientation and relative location coherence. Images with a higher number of corresponding points are likely to belong to the same subject. Such a simple similarity measure has proven to perform consistently in our tests.

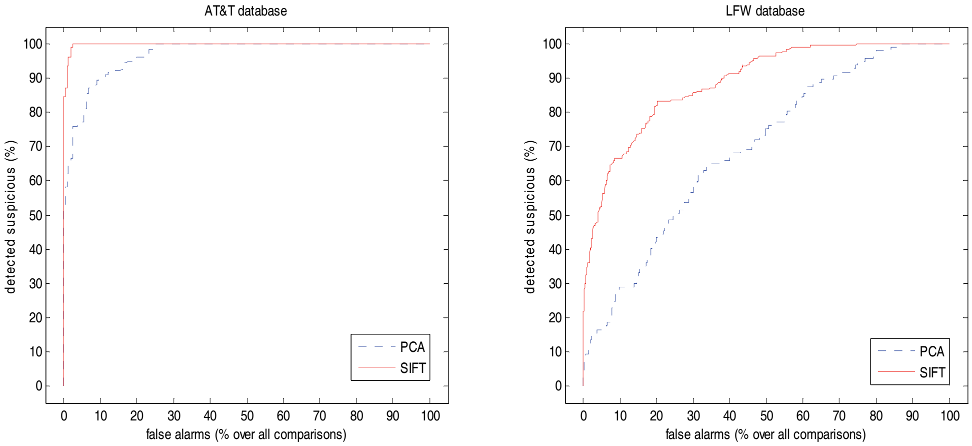


Fig. 6. AT&T and LFW databases: comparison with PCA baseline

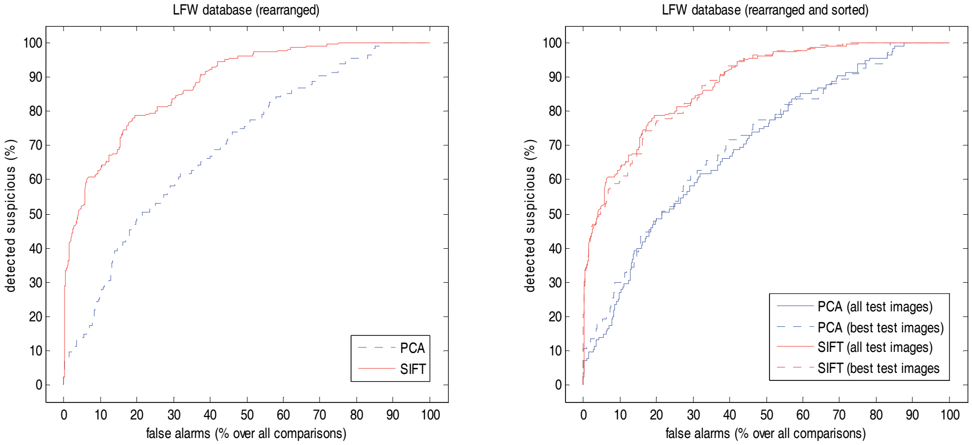


Fig. 7. LFW database reordered and sorted: comparison with PCA baseline

The results in terms of ROC curves show that our approach clearly outperforms the PCA baseline in all conditions. We have performed tests with two different databases: AT&T (not very demanding for a face recognition algorithm) and LFW (extremely demanding); and in both cases our algorithm gave much higher recognition rates than PCA.

Concerning the feasibility of a supervised surveillance system based on our face recognition algorithm, the experimental results show that the quality of the images should be comparable to that of the AT&T database. For lower quality images like those of the LFW database, high recognition rates cannot be expected.

Future work to be carried out includes the comparison of our proposal against other approaches like AAM (active appearance models) and the use of a different interest point descriptor (namely, the SURF descriptor). Another important topic for future research is an evaluation of the possible placements for surveillance cameras; such a research could give us realistic information about the feasibility of a supervised surveillance system.

8. References

Alvarez, L. & Morales, F. (1997). Affine morphological multiscale analysis of corners and multiple junctions. *International Journal of Computer Vision,* 25, 2, (November 1997) 95-107, ISSN 0920-5691.

AT&T (2002). The Database of Faces (formerly ”The ORL Database of Faces”). *www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html*

Bartlett, M.S.; Movellan, J.R. & Sejnowski, J. (2002). Face recognition by independent component analysis. *IEEE. Trans. Neural Networks,* 13, 6, (November 2002) 1450-1464, ISSN 1045-9227.

Baumberg, A. (2000). Reliable feature matching across widely separated views, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2000)*, pp. 774-781, ISBN 0-7695-0662-3, Hilton Head, SC. USA, June 2000, IEEE Computer Society.

Bay, H.; Tuytelaars, T.; Van Gool, L. (2006). SURF: Speeded Up Robust Features. *Proceedings of the 9th International Conference on Computer Vision*, pp. 404-417, ISBN 3-540-33832-2, Graz, Austria, May 2006, Springer-Verlag.

Belhumeur, P.N.; Hespanha, J.P. & Kriegman, D.J. (2002). Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, *IEEE Trans. Pattern Analysis and Machine Intelligence,* 19, 7, (August 2002) 711-720, ISSN 0162-8828.

Bicego, M.; Lagorio, A.; Grosso, E. & Tistarelli, M. (2006). On the use of SIFT features for face authentication, *Proceedings of Conf. on Computer Vision and Pattern Recognition Workshop*, pp. 35, ISBN 0-7695-2646-2, New York, NY, USA, June 2006, IEEE.

Deutsche Welle (2006). German Anti-Terrorist Technology Scans Faces in Train Stations. *www.dw-world.de/dw/article/0,2144,2222284,00.html*.

Draper, B.; Baek, K.; Bartlett, M.S. & Beveridge, R. (2003). Recognizing faces with PCA and ICA. *Computer Vision and Image Understanding,* 91, 1, (July 2003) 115-137, ISSN 1077-3142.

Engadget (2006). Tokyo train station gets facial scan payment systems. *www.engadget.com/2006/04/27/tokyo-train-station-gets-facial-scan-payment-systems/*

Harris, C & Stephens, M. (1988). A combined corner and edge detector, *Proceedings of the 4th Alvey Vision Conference*, pp. 147-151, Manchester, UK, September 1988, University of Sheffield.

Huang, B.; Ramesh, M.; Berg, T. & Learned-Miller, E. (2007). Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments, *Technical Report 07-49, University of Massachusetts*, Amherst, October 2007.

Ke, Y. & Sukthankar, R. (2004). PCA-SIFT: A more distinctive representation for local image descriptors, *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pp. 511-517, ISBN 0-7695-2158-4, Washington, USA, June 2004, IEEE Computer Society.

Kirby, M. & Sirovich, L. (1990). Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Trans. Pattern Analysis and Machine Intelligence,* 12, 1, (January 1990) 103-108, ISSN 0162-8828.

Liu, C. (2004). Face Enhanced independent component analysis and its application to content based face image retrieval. *IEEE. Trans. Systems, Man, and Cybernetics, Part B: Cybernetics,* 34, 2, (April 2004) 1117-1127, ISSN 1083-4419.

Lowe, D.G. (2000). Towards a computational model for object recognitionin IT cortex. *Lecture Notes in Computer Science,* 1811, (May 2000) 20-31, ISSN 0302-9743.

Lowe, D.G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision,* 60, 2, (November 2004) 91-110, ISSN 0920-5691.

Mikolajczyk, K. & Schmid, C. (2004). Scale and affine invariant interest point detectors. *International Journal of Computer Vision,* 60, 1, (October 2004) 63-86, ISSN 0920-5691.

Mikolajczyk, K. & Schmid, C. (2005). A Performance Evaluation of Local Descriptors. *IEEE Trans. Pattern Analysis and Machine Intelligence,* 27, 10, (October 2005) 1615-1630, ISSN 0162-8828.

Pham, T.T.; Maillot, N.E.; Lim, J.H. & Chevallet, J.P. (2007). Latent semantic fusion model for image retrieval and annotation, *Proceedings of 16th ACM Conf. on Information and Knowledge Management*, pp. 439-444, ISBN 978-1-59593-803-9, Lisbon, Portugal, November 2007, Association for Computing Machinery, Inc. (ACM).

Samaria, F. & Harter, A. (1994). Parameterisation of a Stochastic Model for Human Face Identification, *Proceedings of the 2nd IEEE Workshop on Applications of Computer Vision*, pp. 235-242, ISBN 978-989-8111-21-0, Sarasola, FL, USA, December 1994.

Sarfraz, M.S. & Hellwich, O. (2008). Head pose estimation in face recognition across pose scenarios, *Proceedings of the Third International Conference on Computer Vision Theory and Applications*, pp. 235-242, ISBN 978-989-8111-21-0, Madeira, Portugal, January 2008, INSTICC - Institute for Systems and Technologies of Information, Control and Communication.

Schneiderman, H. & Kanade, T. (2000). A Statistical Method for 3D Object Detection Applied to Faces and Cars, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2000)*, pp. 1746-1759, ISBN 0-7695-0662-3, Hilton Head, SC. USA, June 2000, IEEE Computer Society.

Se, S.; Lowe, D.G & Little, J. (2001). Vision-based mobile robot localization and mapping using scale-invariant features, *Proceedings of IEEE Int. Conf. on Robotics and Automation*, pp. 2051-2058, ISBN 0-7803-6578-X, Seoul, Korea, May 2001, IEEE.

Sivic, J.; Everingham, M. & Zisserman, A. (2005). Person spotting: Video shot retrieval for face sets. *Lecture Notes in Computer Science,* 3568, (July 2005) 226-236, ISSN 0302-9743.

Turk, M. & Pentland, A. (1991). Eigenfaces for recognition. *J.Cognitive Neuroscience,* 3, 1, (Winter 1991) 71-86, ISSN 0898-929X.

University of Massachusetts (2007). Labeled Faces in the Wild. *vis-www.cs.umass.edu/lfw/*

Vicente, M.A.; Hoyer, P.O. & Hyvarinen, A. (2007). Equivalence of Some Common Linear Feature Extraction Techniques for Appearance-Based Object Recognition Tasks. *IEEE Trans. Pattern Analysis and Machine Intelligence,* 29, 5, (May 2007) 896-900, ISSN 0162-8828.

Viola, P. & Jones, M.J. (2004). Robust Real-Tine Face Detection. *International Journal of Computer Vision,* 57, 2, (May 2004) 137-154, ISSN 0920-5691.

7

Wavelet–Based Face Recognition Schemes

Sabah A. Jassim
University of Buckingham, Buckingham MK18 1EG
United Kingdom

Abstract. The growth in direct threats to people's safety in recent years and the rapid increase in fraud and identity theft has increased the awareness of security requirements in society and added urgency to the task of developing biometric-based person identification as a reliable alternative to conventional authentication methods. In this Chapter we describe various approaches to face recognition with focus on wavelet-based schemes and present their performance using a number of benchmark databases of face images and videos. These schemes include single-stream (i.e. those using single-subband representations of face) as well as multi-stream schemes (i.e. those based on fusing a number of wavelet subband representations of face). We shall also discuss the various factors and quality measures that influence the performance of face recognition schemes including extreme variation in lighting conditions and facial expressions together with measures to reduce the adverse impact of such variations. These discussions will lead to the introduction of new innovative adaptive face recognition schemes. We shall present arguments in support of the suitability of such schemes for implementation on mobile phones and PDA's.

1. Introduction

The early part of the 21st century has ushered the shaping of a new global communication infrastructure that is increasingly dominated by new generations of mobile phones/devices including 3G and beyond devices resulting in the emergence of pervasive computing environment with less reliance on presence in specific locations or at specific times. The characteristics of such a ubiquitous environment create new security threats and the various mobile devices/nodes are expected to provide additional layers of security for online transactions and real-time surveillance. Cryptography can provide confidentiality protection mechanisms for online and mobile transactions, but authenticating/identifying the principal(s) in such virtual transactions is of utmost importance to fight crime and fraud and to establish trust between parties taking part in such transactions. Traditional authentication mechanisms are based on "something you know" (e.g. a password/PIN) or "something you own/hold" (e.g. a token/smartcard). Such authentication schemes have shown to be prone to serious threats that could have detrimental effects on global economic activities. In recent years, biometric-based authentication has provided a new approach of access control that is aimed at establishing "who you are", and research in the field of biometrics has grown rapidly. The scope of active research into biometrics has gone beyond the traditional list of single traits of fingerprint, retina, iris, voice, and face into newly proposed traits such as handwritten signature, gait, hand geometry, and scent. Moreover, the need for improved performance has lead to active research into multimodal biometrics based on fusing a number of biometrics traits at different levels of fusion including feature level, score level, and decision level. Over the past two decades significant progress has been made in developing robust biometrics that helped realising large-scale automated identification systems.

Advances in mobile communication systems and the availability of cheap cameras and other sensors on mobile devices (3G smart phones) further motivate the need to develop reliable, and unobtrusive biometrics that are suitable for implementation on mobile and constrained devices. Non-intrusive biometrics, such as face and voice are more naturally acceptable as the person's public identity. Unfortunately the performance of known face and voice biometric schemes are lower than those of the Iris or the fingerprint schemes. The processing and analysis of face image suffer from the curse of dimension problem, and various dimension reduction schemes have been proposed including PCA (principal Component analysis). In recent years a number of wavelet-based face verification schemes have been proposed as an efficient alternative to traditional dimension reduction procedures.

The Wavelet Transform is a technique for analyzing finite-energy signals at multi-resolutions. It provides an alternative tool for short time analysis of quasi-stationary signals, such as speech and image signals, in contrast to the traditional short-time Fourier transform. A wavelet-transformed image analyses the signal into a set of subbands at different resolutions each represented by a different frequency band. Each wavelet subband encapsulates a representation of the transformed images object(s), which differ from the others in scale and/or frequency content. Each wavelet subband of transformed face images can be used as a face biometric template for a face recognition scheme, and the fusion of a multiple of such schemes associated with different wavelet subbands will be termed as multi-stream face recognition scheme.

2. Face Recognition - A brief review

Automatic face based human Identification is a particularly tough challenge in comparison to identification based on other biometric features such as iris, fingerprints, or palm prints. Yet, due to its unobtrusive nature, together with voice, the face is naturally the most suitable method of identification for security related applications, ([1], [2], [3]). Recent growth in identity theft and the rise of international terrorism on one hand and the availability of high-resolution digital video-cameras at a relatively low cost is a major driving force in the surge of interest for efficient and accurate enrolment, verification schemes of face-based authentication. Moreover, there are now new opportunities, as well as tough challenges, for mass deployments of biometric-based authentications in a range of civilian and military applications. In the rest of this section we shall briefly review the main approaches to face recognition with focus on 2D schemes directly related to this chapter's aim.

2.1 Dimension reduction approach

An important part of a face recognition process is the feature extraction of a given facial image. Two current approaches to feature extraction are the geometry feature-based methods and the more common template-based methods. In the latter, sets of face images are statistically analysed to obtain a set of feature vectors that best describe face image. A typical face images is represented by a high dimensional array (e.g. 12000=120×100 pixels), the processing/analysis of which is a computationally demanding task, referred to in the literature as the "curse of dimensionality", well beyond most commercially available mobile devices. It is therefore essential to apply dimension reduction procedures that reduce redundant data without losing significant features. A common feature of dimension reducing procedures is a linear transformation of the face image into a "significantly" lower dimensional subspace from which a feature vector is extracted. The first and by far the most commonly used dimension reduction method is the Principal Component Analysis (PCA), also known as Karhunen-Love (KL) transform, [4]. In [5], M. Turk and Pentland used the PCA technique to develop the first successful and well known Eigenface scheme for face recognition. PCA requires the use of a sufficiently large training set of multiple face images of the enrolled persons, and attempts to model their significant variation from their average image, by taking a number of unit eigenvectors corresponding to the "most significant" eigenvalues (i.e. of largest absolute values). Essentially, the selected eigenvectors are used as the basis for a linear transformation that maps the original training set of face images around their mean in order to align with the directions the first few principal components which maximizes the variance as much of the as possible. The values in the remaining dimensions (corresponding to the non-significant eigenvalues), tend to be highly correlated and dropped with minimal loss of information.

Despite its success in reducing false acceptances, the PCA/Eigenface scheme is known to retain within-class variations due to many factors including illumination and pose. Moghaddam et al. [6] have demonstrated that the largest three eigen coefficients of each class overlap each other. While this shows that PCA has poor discriminatory power, it has been demonstrated that leaving out the first 3 eigenfaces (corresponding to the 3 largest eigenvalues) could reduce the effect of variations in illumination [6]. But this may also lead to loss of information that is useful for accurate identification.

An alternative approach to PCA based linear projection is Fisher's Linear Discriminant (FLD), or the Linear Discriminant Analysis (LDA) which is used to maximize the ratio of the determinant of the between class scatter to that of within-class scatter [7], [8]. The downside of these approaches is that a number of training samples from different conditions are required in order to identify faces in uncontrolled environments.

Other schemes that deal with the curse of dimension include Independent Component Analysis (ICA), or a combination of ICA and LDA/FLD, (see [1], [7], and [9]). Lack of within-class (variations in appearance of the same individual due to expression and/or lighting) information is known to hinder the performance of both PCA and ICA based face recognition schemes. Cappelli et al., [9], proposed a multi-space generalization of KL-transformation (MKL) for face recognition, in which a PCA-subspace is created for each enrolled classes. The downside of this approach is that a large number of images are required to create a subspace for each class.

All the statistical approaches above require a large number of training images to create a subspace, which in turn requires extra storage space (for the subspace and enrolled template/features), [10]. Current mobile devices (3G smart phones) and smartcards, which are widely used in commercial and military applications, have limited computing resources and it is difficult to implement complex algorithms, especially for face verification. Bicego et al. presented a face verification scheme based on Hidden Markov Models (HMM). Statistical features such as the mean and variance are obtained by overlapping sub images (of a given original face image). These features are used to compose the HMM sequence and results show that the HMM-based face verification scheme, proposed by Bicego et al., outperforms other published results, [11].

2.2 Frequency transforms based approaches

Frequency transforms provide valuable tools for signal processing and analysis. Frequency information content conveys richer knowledge about features in signals/images that should be exploited to complement the spatial information. Fourier and wavelet transforms are two examples that have been used with significant success in image processing and analysis tasks including face recognition. To some extent, such transforms reduce dimension with no or little loss of information.

The work of John Daugman, ([12], [13]) and others on generalisation of Gabor functions has led to a general compact image representation in terms of Gabor wavelets. The Gabor wavelets, whose kernels are similar to the 2D receptive field profiles of the mammalian cortical simple cells, exhibit strong characteristics of spatial locality and orientation selectivity, and are optimally localized in the space and frequency domains. The Gabor wavelet model was eventually exploited to develop new approaches to face recognition. Taking into account the link between the Gabor wavelet kernels and the receptive field profiles of the mammalian cortical simple cells, it is not unreasonable to argue that Gabor wavelet based face recognition schemes mimics the way humans recognise each others.

Lades et al. [14] demonstrated the use of Gabor wavelets for face recognition using the Dynamic Link Architecture (DLA) framework. The DLA starts by computing the Gabor jets, and then it performs a flexible template comparison between the resulting image decompositions using graph-matching. L Wiskott et al, [15], have expanded on the DLA, and developed the Elastic Bunch Graph Matching (EBGM) face recognition system, whereby individual faces were represented by a graph, each node labelled with a set of complex Gabor wavelet coefficients, called a jet. The magnitudes of the coefficients label the nodes for matching and recognition, the phase of the complex Gabor wavelet coefficients is used for location of the nodes. The nodes refer to specific facial landmarks, called fiducial points. A data structure, called the bunch graph, to represent faces by combining jets of a small set of individual faces. Originally many steps (e.g. selecting the Fiducial points) were carried out manually, but gradually these would have been replaced with a automated procedures.

Z. Zhang et al, [16], compared the performance of a Geometry-based and a Gabor wavlet-based facial expression recognition using a two-layer perceptron. The first uses the geometric positions of a set of fiducial points on a face, while the second type is a set of multi-scale and multi-orientation Gabor wavelet coefficients extracted from the face image at the fiducial points. For the comparison they used a database of 213 images of female facial expressions and their results show that the Gabor wavelet –based scheme outperforms the geometric based system.

C. Lui and H. Wechsler, [17], developed and tested an Independent Gabor Features (IGF) method for face recognition. The IGF first derives a Gabor feature vector from a set of downsampled Gabor wavelet representations of face images, then reduces the dimensionality of the vector by means of Principal Component Analysis (PCA), and finally defines the independent Gabor features based on the Independent Component Analysis (ICA). The independence property of these Gabor features facilitates the application of the

Probabilistic Reasoning Model (PRM) method for classification. The Gabor transformed face images exhibit strong characteristics of spatial locality, scale and orientation selectivity, while ICA further reduce redundancy and represent independent features explicitly.

The development of the discrete wavelet transforms (DWT), especially after the work of I. Daubechies (see e.g. [18]), and their multi-resolution properties have naturally led to increased interest in their use for image analysis as an efficient alternative to the use of Fourier transforms. DWT's have been successfully used in a variety of face recognition schemes (e.g. [10], [19], [20], [21], [22]). However, in many cases, only the approximation components (i.e. the low frequency subbands) at different scales are used either as a feature vector representation of the faces perhaps after some normalisation procedures or to be fed into traditional face recognition schemes such as the PCA as replacement of the original images in the spatial domain.

J. H. Lai et al, [23], developed a holistic face representation, called spectroface, that is based on an elaborate combination of the (DWT) wavelet transform and the Fourier transform. To make the spectroface invariant to translation, scale and on-the-plane rotation, the LL wavelet subband of the face image is subjected to two rounds of transformations. The LL wavelet subband is less sensitive to the facial expression variations while the first FFT coefficients are invariant to the spatial translation. The second round of FFT is applied after the centralised FFT in the first round is represented by polar coordinates. Based on the spectroface representation, their proposed face recognition system is tested on the Yale and Olivetti face databases. They report recognition accuracy of over 94% for rank1 matching, and over 98% for rank 3 matching.

Another wavelet-based approach for face recognition has been investigated in terms of dual-tree complex wavelets (DT-CW) techniques developed by N. G. Kingsbury, (see e.g. [24]). Y. Peng et al, [25], propose face recognition algorithm that is based on the use of an anisotropic dual-tree complex wavelet packets (ADT-CWP) for face representation. The ADT-CWP differs from the traditional DT-CW in that the decomposition structure is determined first by an average face, which is then applied to extracting feature of each face image. The performance of their scheme is compared with the traditional Gabor-based methods using a number of different benchmark databases. The AD-CWP method seems to outperform the Gabor-based schemes and it is computationally more efficient.

The rest of the chapter is devoted to DWT-based face recognition tasks. We shall first give a short description of the DWT as a signal processing and analysis tool. We then describe the most common approaches to wavelet-based multi-stream face recognition.

3. Wavelet Transforms

The Wavelet Transform is a technique for analyzing finite-energy signals at multi-resolutions. It provides an alternative tool for short time analysis of quasi-stationary signals, such as speech and image signals, in contrast to the traditional short-time Fourier transform. The one dimensional Continuous Wavelet Transform CWT of f(x) with respect to the wavelet Ψ(x) is defined as follows:

$$\psi_f(j,k) = \langle f, \psi_{j,k} \rangle = \int_{-\infty}^{\infty} f(x)\psi_{j,k}(x)dx$$

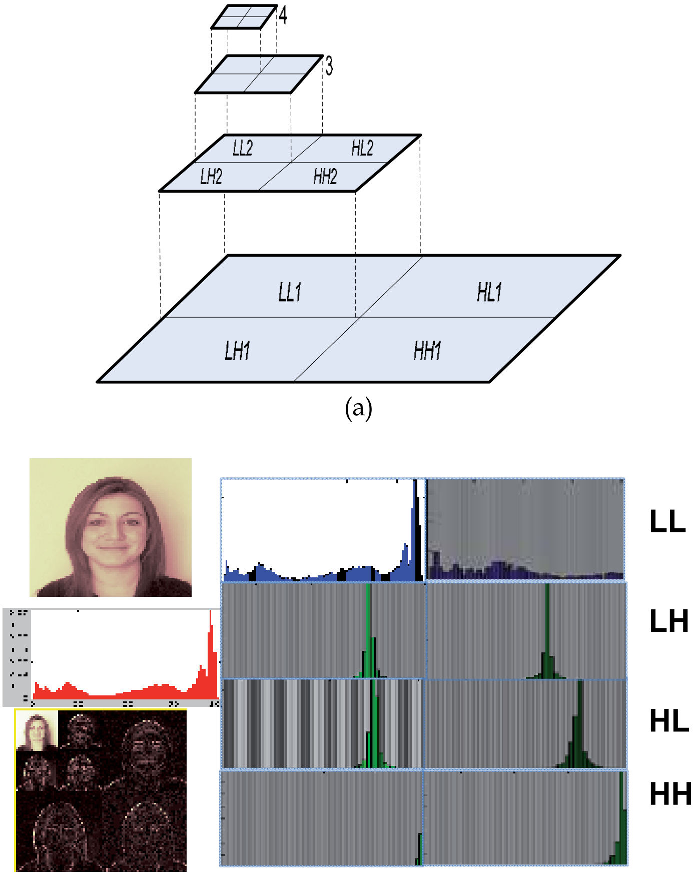
i.e. wavelet transform coefficients are defined as inner products of the function being transformed with each of the base functions $\Psi_{j,k}$. The base functions are all obtained from a single wavelet function $\Psi(x)$, called the mother wavelet, through an iterative process of scaling and shifting, i.e.

$$\psi_{j,k}(t) = 2^{\frac{j}{2}}\psi(2^{j}t - k).$$

A wavelet function is a wave function that has a finite support and rapidly diminishes outside a small interval, i.e. its energy is concentrated in time. The computation of the DWT coefficients of a signal k does not require the use of the wavelet function, but by applying two Finite Impulse Response (FIR) filters, a high-pass filter h, and a low-pass filter g. This is known as the Mallat's Algorithm. The output will be in two parts, the first of which is the detail coefficients (from the high-pass filter), and the second part is the approximation coefficients (from the low-pass filter). For more details see [26].

The Discrete Wavelet Transform (DWT) is a special case of the WT that provides a compact representation of a signal in time and frequency that can be computed very efficiently. The DWT is used to decompose a signal into frequency subbands at different scales. The signal can be perfectly reconstructed from these subband coefficients. Just as in the case of continuous wavelets, the DWT can be shown to be equivalent to filtering the input image with a bank of bandpass filters whose impulse responses are approximated by different scales of the same ***mother wavelet***. It allows the decomposition of a signal by successive highpass and lowpass filtering of the time domain signal respectively, after sub-sampling by 2. Consequently, a wavelet-transformed image is decomposed into a set of subbands with different resolutions each represented by a different frequency band. There are a number of different ways of doing that (i.e. applying a 2D-wavelet transform to an image). The most commonly used decomposition scheme is the *pyramid* scheme. At a resolution depth of k, the pyramidal scheme decomposes an image I into 3k +1 subbands, {LL_k, LH_k, HL_k, HH_k, LH_{k-1}, HL_{k-1},…, LH_1, HL_1}, with LL_k, being the lowest-pass subband, (see figure 3.1(a)).

There are ample of wavelet filters that have been designed and used in the literature for various signal and image processing/analysis. However, for any wavelet filter, the LL subband is a smoothed version of original image and the best approximation to the original image with lower-dimensional space. It also contains highest-energy content within the four subbands. The subbands LH_1, HL_1, and HH_1, contain finest scale wavelet coefficients, and the coefficients LL_k get coarser as k increases. In fact, the histogram of the LL_1-subband coefficients approximates the histogram of the original image in the spatial domain, while the wavelet coefficients in every other subband has a Laplace (also known as generalised Gaussian) distribution with ≅ 0 mean, see Figure 3.1(b). This property remains valid at all decomposition depth. Moreover, the furthest away a non-LL coefficient is from the mean in that subband, the more probable the corresponding position(s) in the original image have a significant feature, [27]. In fact the statistical properties of DWT non-LL subbands can be exploited for many image processing applications, including image/video compression, watermarking, content-based video indexing, and feature extraction.



(a)

(b)

Fig. 3.1. (a) The Multi-resolution Pyramid (b) An image, its WT to level 2 and subbands histograms.

4. DWT – based Face recognition

The LL subband of a wavelet transformed image corresponds to the low frequency components in both vertical and horizontal directions of the original image. Therefore, it is the low frequency subband of the original image. The subband LH corresponds to the low-frequency component in the horizontal direction and high-frequency components in vertical direction. Therefore it holds the vertical edge details. Similar interpretation is made on the subbands HL and HH. These remarks together with our knowledge of structure of facial features provide a strong motivation and justification for investigating wavelet-based approaches to face recognition. In fact the variety of wavelet decomposition schemes and filter banks provide a very rich and a complex source of information that could be exploited to deal with the tough challenges and difficulties associated with face recognition in the presence of expression and extreme variations in lighting conditions.

With appropriate pixel value scaling the low LL subband, displayed as an image, looks like a smoothing of the original image in the spatial domain (see Figure 3.1(b)). For efficiency purposes and for reason of normalising image sizes, non-DWT based face recognition schemes such as PCA pre-process face images first by resizing/downsampling the images. In such cases, matching accuracy may suffer as a result of loss of information. The LL subbands of the face image, does provide a natural alternative to these pre-processing procedures and this has been the motivation for the earlier work on wavelet-based face recognition schemes that have mostly combined with LDA and PCA schemes (e.g. [10], [28], [29], [30], [31]). Below, we shall describe face recognition schemes, developed by our team, that are based on the PCA in a single wavelet subband and summarise the results of performance tests by such schemes for some benchmark face databases. We will also demonstrate that the use of the LL-subband itself as the face feature vector results in comparable or even higher accuracy rate. These investigations together with the success of biometric systems that are based on fusing multiple biometrics (otherwise known as multi-modal biometrics) have motivated our work on multi-stream face recognition. This will be discussed in section 5.

4.1 PCA in the Wavelet Domain

Given the training set Γ of images, applying a wavelet transform on all images results in a set $W_k(\Gamma)$ of multi-resolution decomposed images. Let $L_k(\Gamma)$ be the set of all level-k low subbands corresponding to the set $W_k(\Gamma)$. Apply PCA on the set $L_k(\Gamma)$ whose elements are the training vectors in the wavelet domain (i.e. LL_k subbands). Note that each wavelet coefficient in the LLk subband is a function of $2^k x 2^k$ pixels in the original image representing a scaled total energy in the block. Figure 3.2, below, shows the first 4 eigenfaces obtained from a dataset of images in the spatial domain as well as in the low subbands at levels 1 and 2 using the Haar wavelet filter. There are no apparent differences between the eigenfaces in the spatial domain and those in the wavelet domain.

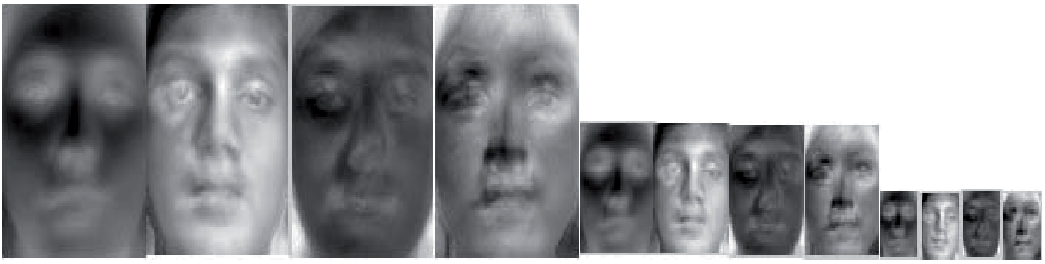


Fig. 3.2. Eigenfaces in (a) spatial domain, (b) LL1, and (c) LL2

Diagram1, below, illustrates the enrolment and matching steps which will cover face recognition in the wavelet domain with and without the application of PCA. The diagram applies equally to any wavelet subband including the high frequency ones.

There are many different wavelet filters to use in the transformation stage, and the choice of the filter would have some influence on the accuracy rate of the PCA in the wavelet domain. The experiments are designed to test the effect of the choice of using PCA or not, the choice of wavelet filter, and the depth of decomposition. The performance of the various possible schemes have been tested for a number of benchmark databases including ORL (also known as AT&T see http://www.uk.research.att.com/facedatabase.html), and the controlled section of the BANCA, [32]. These datasets of face images do not involve significant variation in illumination. The problem of image quality is investigated in section 6. Next we present a small, but representative, sample of the experimental results for few wavelet filters.

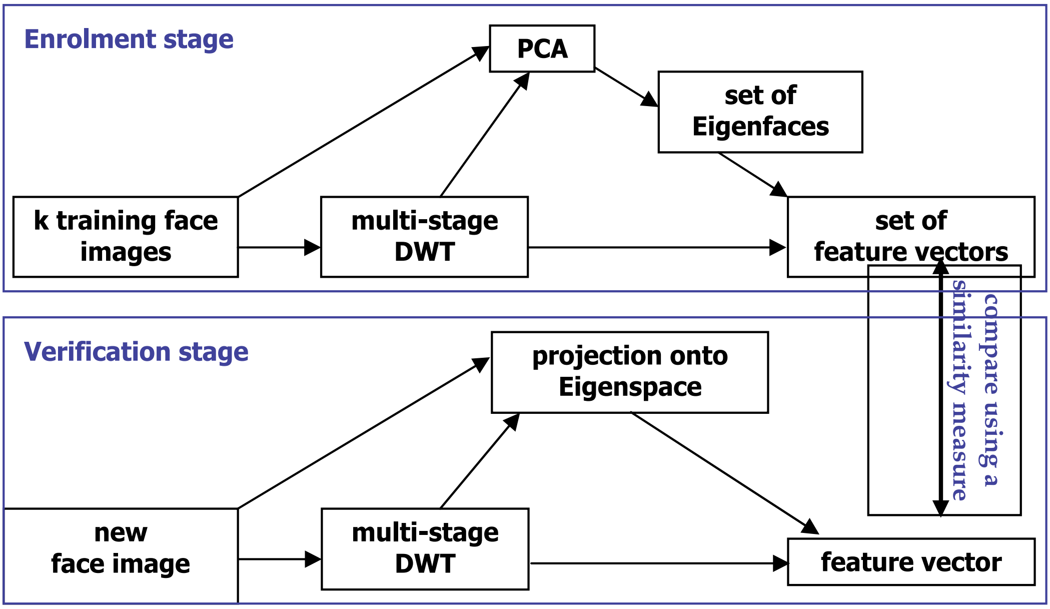


Diagram 3.1 Verification scheme.

The ORL Experiments.

- **Enrolment/Training module** There are 40 subjects in ORL. In the first instance, we split the subjects into two equal groups: Group1 and Group2. For each group we trained the system with 4 different sets of five images for each subject. These sets were, respectively, frames 1-5, frames 6-10, even indexed frames and odd indexed frames. In total, we conducted 8 different training sessions for these groups.

- ***The Testing Module.*** In each of the training sessions that consisted of 20 subjects, the remaining 100 images of the trained subjects as well as 100 impostor images (5 images per subject, selected in the same way as in the training scheme) were used to test the many-to-one identification schemes.

Chart 3.1, below, contains the test results of experiments that were designed to test the verification accuracy when the Haar wavelet filter is used to different level of decompositions. It shows the average accuracy rates for the various identification schemes measured using different number of eigenfaces (20, 30, 40 and 50) for the schemes that involve the use of PCA. The results indicate that regardless of the number of eigenvalues chosen, PCA in the LL subbands outperform the PCA in the spatial domain, and LL3 is the best performing subband in this case. Moreover, the wavelet LL3 scheme without the PCA achieves the best performance. Another interesting observation, not explicit in this chart, is that the accuracy rates for the various training schemes vary widely around the stated averages, indicating that accuracy can be improved further by making a *careful* choice of the training images for the enrolled subjects.

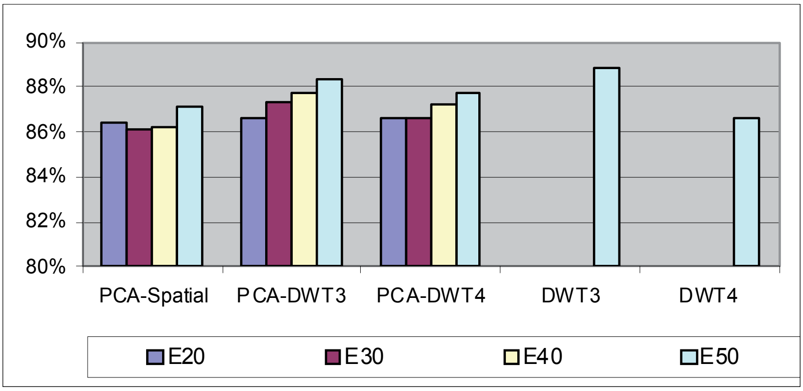


Chart 3.1. Identification accuracy for Spatial PCA, Wavelet PCA, and Wavelet-only features

The superior performance of the wavelet-only scheme compared to the other schemes, has desirable implication beyond the computational efficiency. While most conventional face recognition schemes require model/subspace training, wavelet-based recognition schemes can be developed without the need for training, i.e. adding/removing classes do not require rebuilding the model from scratch.

Jen-Tzung Chien etal ([10]) who used all the 40 subjects of ORL to test the performance of a number of recognition schemes including some of the wavelet-based ones investigated here. In those experiments, there were no impostors, i.e. untrained subjects. Thus we conducted experiments where all the 40 subjects were used for training. We trained the system 4 times each with a set of 5 different frames and in each case the remaining 200 images (5 frames for each subject) in the database were used for testing.. On average, all schemes have more or less achieved similar accuracy rate of approximately 89%. Similar experiments with 35 trained subjects, the rest being impostors, have been conducted but in all cases the results were similar to those shown above.

Chart 3.2 contains the results of verifications rather identifications. The experiments were carried out to test the performance of wavelet-based verification schemes, again with and without PCA. Here, two filters were used, the Haar as well as the Daubechies 4 wavelets, and in the case of Daubechies 4 we used two versions whereby the coefficients are scaled for normalisation in the so called Scaled D3/d4. The results confirmed again the superiority of PCA in the wavelet domain over PCA in the spatial, and the best performance was obtained when no PCA was applied. The choice of filter does not seem to make much difference at level 3, but Haar outperforms both versions of Daubechies 4.

The superiority of the PCA in the wavelet domain over the PCA in the spatial domain can be explained in terms of the poor within class variation of PCA and the properties of the linear transform defined by the low-pass wavelet filter. The low-pass filter defines a contraction mapping of the linear space of the spatial domain into the space where LL subbands resides (i.e. for any two images the distance between the LL-subbands of two images is less than that between the original images). This can easily be proven for the Haar filter. This will help reduce the within class variation.

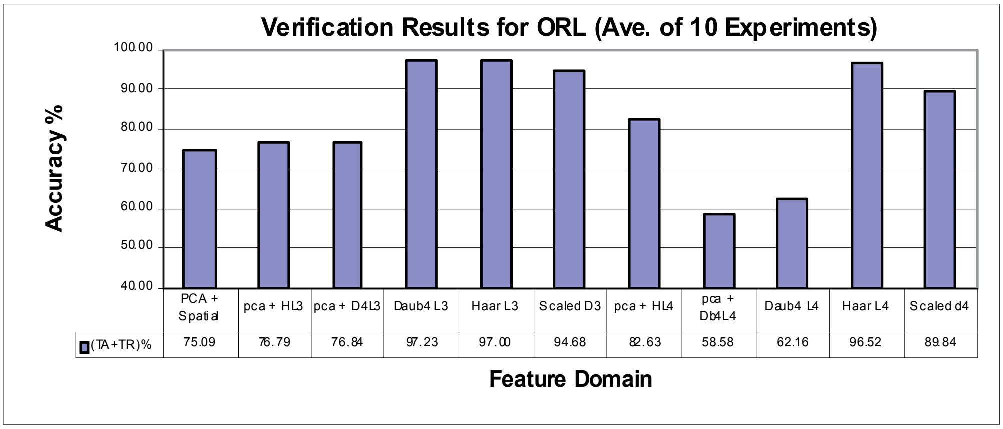


Chart 3.2.Verification accuracy for Spatial PCA, Wavelet PCA, and Wavelet-only features

The trend in, and the conclusions from these experiments are confirmed by other published data. For example, C.G.Feng et al, [33] have tested and compared the performance of PCA in the spatial domain and in wavelet subbands at different levels for the Yale database. Table 3.2, below, reports the recognition accuracy for the Daubechies 4 filter and confirms our conclusions. Note that the inter-class separation experiment in [33] can be seen to demonstrate that the contraction mapping nature of the low-pass filter transformation does not have adverse impact on the inter-class separation.

Method	PCA on original image	PCA on WT subband 1 Image	PCA on WT subband 2 Image	PCA on WT subband 3 Image	Proposed Method (PCA on WT subband 4 image)
Accuracy	78.78%	75.75%	83.03%	81.18%	85.45%

Table 3.2 [¥]. Performance comparison using Yale database

5. Multi-stream face Recognition

A wavelet transformed image is decomposed into a set of frequency subbands with different resolutions. Each frequency subband gives rise to a different feature vector representation of the face and has the potential to be used for recognition. The performances of such schemes vary significantly depending on many factors including the chosen wavelet filter, the depth of decomposition, the similarity measure, the sort of processing that the corresponding coefficients are subjected to, and the properties of subband as described at the end of section 3. The fact that identification schemes that are based on the fusion of different biometric modalities have shown to significantly outperform the best performing single modality scheme, raises the possibility of fusing different signal representing the same modality. Moreover, different subbands of wavelet-transformed face images, each representing the

[¥] Reproduced from [CGFeng].

face in different way, makes them perfect candidates for fusion without costly procedures. Diagram 2, below, depicts the stages of the wavelet based multi-stream face recognition for 3 subbands at level 1, but this could be adopted for any set of subbands at any level of decomposition.

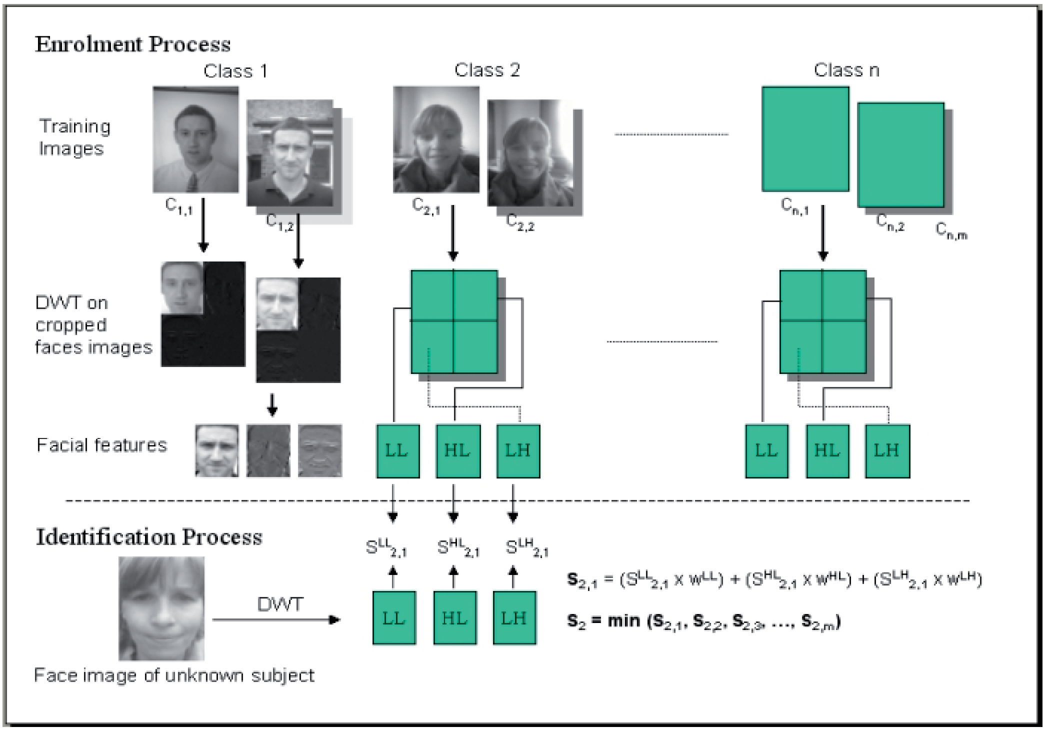


Diagram 2. Multi-stream Wavelet face recognition scheme

In this section we shall investigate the viability of fusing these streams as a way of improving the accuracy of wavelet-based face recognition. We shall establish that the fusion of multiple streams of wavelet-based face schemes does indeed help significantly improve single stream face recognition. We have mainly experimented with the score fusion of wavelet subbands at one decomposition depth. Limited experiments with other level of fusion did not achieve encouraging results.

The experiments reported here are based on the performance of the multi-stream face wavelet recognition for databases that involve face images/videos captured under varying recording conditions and by cameras of different qualities. These databases are the Yale database, and the BANCA audio-visual database. More extensive experiments have been conducted on the PDAtabase audio-visual database of videos recorded on a PDA within the SecurePhone EU-funded project (www.secure-phone.info).

5.1 The Yale database experiments

Identification experiments reported in table 5.1., below, are based on the "leave one out" protocol. The table contain the performance of 3 single wavelet–based down to level 3 (the LL3, LH3 and HH3 subbands schemes), the fusion of the three subband streams for a selection of fixed weight combinations and for comparison we include results from some of the best performing face recognition schemes reported in Yang, [34], and Belhuemer et al, [35]. These results demonstrate that among the single subband streams, the LH3 is the best performing one. The multi-stream fusion of the three subbands for all but one weight configuration outperform the best single stream scheme, illustrating the conclusion that the multi-stream approach yields improved performance. Comparing the results with those from the state of the art schemes reported in [14] and [26] shows that the multi-steam fusion of the two single streams LH3 and HL3 subbands outperform all but 3 of the SOTA schemes. One can predict with confidence that the multi-stream fusing of several subbands at different level of decomposition would result in significantly improved performance.

Method	Features/Weights			Error Rate(%)
	LL3	HL3	LH3	
Single-stream	1	0	0	23.03
	0	1	0	14.55
	0	0	1	12.73
Multi-stream	0	0.4	0.6	9.70
	0	0.3	0.7	9.09
	0	0.2	0.8	9.70
	0.1	0.3	0.6	10.91
	0.1	0.25	0.65	10.91
	0.2	0.2	0.6	12.73
	0.2	0.3	0.5	12.12
	0.2	0.4	0.4	13.33
Yang[26]	Eigenface(EF30)			28.48
	Fisherface (FF14)			8.48
	ICA			28.48
	SVM			18.18
	K.Eigenface (EF60)			24.24
	k.Fisherface (FF14)			6.06
Belhumeur et al.[14] (Full Face)	Eigenface (EF30)			19.40
	Eigenface (EF30, w/o 1st 3 EF)			10.8
	Correlation			20.00
	Linear Subspace			15.60
	Fisherface			0.60

Table 5.1. Fusion Experiments – Yale database

5.2 BANCA database experiments

Experiments reported here are only a small, but representative, sample conducted on the BANCA database and is limited to the use of the G evaluation protocol, [32]. The experiments are conducted on the English section of the database which include recording for 52 subjects. Each subject participated in 12 sessions, each consisting of two short clips uttering a true-client text while in the second clip he/she acts as an impostor uttering a text belonging to another subject. The 12 sessions are divided into 3 groups:

- the controlled group – sessions 1-4 (high quality camera, controlled environment and a uniform background)
- the degraded group – sessions 5-8 (in an office using a low quality web camera in uncontrolled environment).
- the adverse group – sessions 9-12 (high quality camera, uncontrolled environment)

For the G evaluation protocol, the true client recordings from session 1, 5, and 9 were used for enrolment and from each clip 7 random frames were selected to generate the client templates. True-client recordings from sessions 2, 3, 4, 6, 7, 8, 10, 11, and 12 (9 videos) were used for testing the identification accuracy. From each test video, we selected 3 frames and the minimum score for these frames in each stream was taken as the score of the tested video in the respective stream. In total, 468 tests were conducted. Identification accuracies of single streams (first 3 rows) and multi-stream approaches for the G protocol are shown in Table 5.2. Across all ranks the LH-subband scheme significantly outperformed all other single streams. The multi-stream fusion of the 3 streams outperformed the best single stream (i.e. the LH subband) by a noticeable percentage. The best performing multi-stream schemes are mainly the ones that give >0.5 weight to the LH subband and lower weight to the LL-subband. Again these experiments confirm the success of the multi-stream approach.

Weights			Identification Accuracy for Rank n for G test configuration									
LL	HL	LH	1	2	3	4	5	6	7	8	9	10
1.00	0.00	0.00	58.55	67.95	72.65	77.14	80.34	82.69	84.62	85.47	86.54	87.61
0.00	1.00	0.00	56.84	65.60	72.22	76.28	79.49	82.26	84.83	85.68	87.18	87.82
0.00	0.00	1.00	69.23	80.77	85.68	89.10	91.45	92.52	93.38	94.44	95.09	95.94
0.20	0.20	0.60	76.28	85.47	88.89	90.81	92.74	93.38	94.44	95.73	96.37	96.79
0.20	0.30	0.50	76.07	83.12	88.46	91.45	93.38	93.80	95.09	95.30	95.73	96.15
0.25	0.35	0.40	74.15	81.62	87.61	89.74	91.24	92.31	92.95	94.66	95.30	95.30
0.10	0.30	0.60	76.71	85.90	89.32	92.09	93.16	93.80	94.87	95.51	96.37	96.79
0.10	0.20	0.70	75.00	85.68	88.89	92.74	94.02	94.23	94.44	95.09	96.15	96.58

Table 5.2 Rank based results for single and multi-stream identification using test protocol G

6. Quality-based Adaptive face Recognition

The performance of most face recognition schemes including those mentioned earlier deteriorates when tested in uncontrolled conditions when compared to their performance under normal recording conditions. These effects are often the result of external factors such as extreme variations in illumination, expression, and occlusion. To deal with varying recording conditions, most existing schemes adopt normalization procedures that are applied irrespective of the recording conditions at the time of recording. Such strategies are known to improve accuracy in adverse conditions at the expense of deteriorated performance in somewhat normal recording conditions that generate well/reasonably lit images, and thereby yielding little or no improved overall accuracy. This section focuses on the development of adaptive approaches to deal with such variations, whereby the application of normalisation procedures will be based on certain criteria on image quality that are detected automatically at the time of recording. We shall describe some quantitative quality measures that have been incorporated in adaptive face recognition systems in the presence of extreme variation in illumination. We shall present experimental results in support of using these measures to control the application of light normalisation procedures as well as dynamic fusion of multi-stream wavelet face recognition whereby the fusion weighting become dependent on quality measures.

6.1. QUALITY ASESSMENT MEASURES

Quality measures play an important role in improving the performance of biometric systems. There has been increasing interest by researchers in using quality information to make more robust and reliable recognition systems (e.g. [36], [37], [38], [39], [40]). Quality measures can be classified as modality-dependent and modality-independent. Modality dependent measures (such as pose or expression) can be used for face biometric only, while modality-independent measures such as (contrast and sharpness) can be used for any modality because they do not need any knowledge about the specific biometric. For multi-modal and multi-streams biometrics, there is a need to combine the various trait/stream quality measures to build adaptive weighting associated with the matching scores produced by their individual matchers, ([41], [42]). Quality measures can also be classified in terms of the availability of reference information: full reference, reduced reference, and no reference quality assessment approaches, ([43]).

Face image quality measures must reflect some or all aspects variation from a "norm" in terms of lighting, expression, pose, contrast, eyes location, mouth location, ears location, blur and so forth, ([44], [45]). New quality measures based on wavelets have been developed for different biometrics, [46]. Here, we will focus on image quality measures as a result of variation in lighting conditions and its use for improving performance of face recognition and dynamic fusion of multi-streams.

UNIVERSAL IMAGE QUALITY INDEX

Illumination image quality measures must either reflect luminance distortion of any image in comparison to a known reference image, or regional variation within the image itself. The *universal image quality index* (*Q*) proposed by Wand and Bovik,[47] incorporates a number of image quality components from which one can extract the necessary ingredients an illumination image quality measure that fits the above requirements. For two signals/vectors $X=\{x_i|i=1,2,.....N\}$ *and* $Y=\{y_i|i=1,2,...N\}$, Q(X,Y) is defined as:

$$Q(X,Y) = \frac{4\,\sigma_{xy}\;\overline{X}\;\overline{Y}}{(\sigma^2{}_x+\sigma^2{}_y)[(\bar{x})^2+(\bar{y})^2]} \tag{1}$$

where,

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} X_i \quad , \quad \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i \, , \quad \sigma_x{}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2 \, ,$$

$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})$$

It models any distortion as a combination of three components: loss of correlation, luminance distortion, and contrast distortion. In fact, Q is the product of three quality measures reflecting these three components:

$$Q(X,Y) = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \bullet . \frac{2\,\overline{xy}}{(\bar{x})^2 + (\bar{y})^2} \bullet \frac{2\,\sigma_x \sigma_y}{\sigma^2{}_x + \sigma^2{}_y} . \qquad (2)$$

The luminance quality index is defined as the distortion component:

$$\mathrm{LQI} = \frac{2\,\bar{x}\,\bar{y}}{(\bar{x})^2 + (\bar{y})^2} \qquad (3)$$

In practice, the LQI of an image with respect to another reference image is calculated for each window of size 8x8 pixels in the two images, and the average of the calculated values defines the LQI of the entire image. The LQI is also referred to as the Global LQI as opposed to regional LQI, when the image is divided into regions and the LQI is calculated for each region separately, [48].

The distribution of LQI values for the images in the different subsets of the Extended Yale B database reveal an interesting, though not surprising, pattern. There is a clear separation between the images in sets 1 and 2, where all images have LQI values > 0.84, and those in sets 4 and 5 where all LQI vales < 0.78. Images in set 3 of the database have LQI values in the range 0.5 to 0.95.

The use of LQI with a fixed reference image that has a perceptually good illumination quality investigated as a pre-processing procedure prior to single-stream and multi-streams wavelet-based face recognition schemes, for adaptive face recognition schemes with improved performance over the non-adaptive schemes.

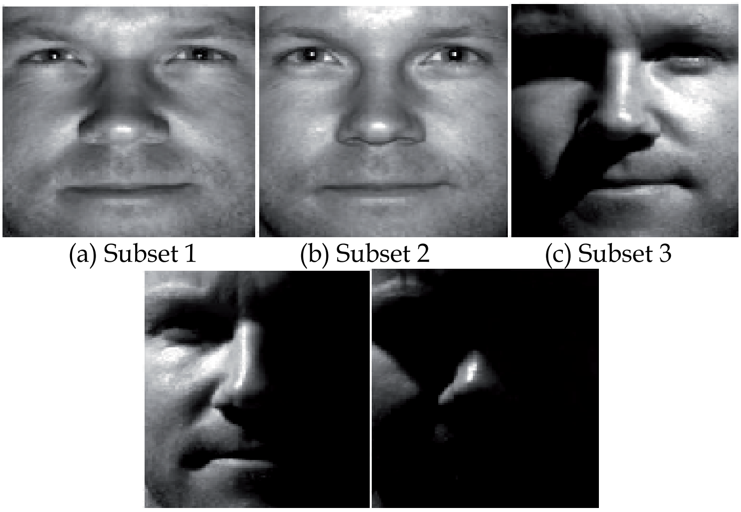
In the case of multi-streams schemes, a regional version of LQI index is used to adapt the fusion weights, [48]. A. Aboud et al, [37], have further developed this approach and designed adaptive illumination normalization without a reference image. We shall now discuss these approaches in more details and present experimental evidences on their success.

In order to test the performance of the developed adaptive schemes, the relevant experiments were conducted on the Extended Yale database, [49], which incorporates extreme variations in illumination recording condition. The cropped frontal face images of the extended Yale B database provide a perfect testing platform and framework for illumination based image quality analysis and for testing the viability of adaptive face recognition scheme. The database includes 38 subjects each having 64 images, in frontal pose, captured under different illumination conditions. In total number there are 2414 images. The images in the database are divided into five subsets according to the direction of the light-source from the camera axis as shown in Table 6.1.

Subsets	Angles	Image Numbers
1	$\theta < 12$	263
2	$20 < \theta < 25$	456
3	$35 < \theta < 50$	455
4	$60 < \theta < 77$	526
5	$85 < \theta < 130$	714

Table 6.1 Different illumination sets in the extended Yale B database

Samples of images for the same subject taken from different subsets of the Extended Yale B database are shown in Figure 6.1. LQI values are respectively 1, 0.9838, 0.8090, 0.4306, and 0.2213.



(a) Subset 1 (b) Subset 2 (c) Subset 3

(d) Subset 4 (e) Subset 5

Fig. 6.1. Sample images from different subsets in the Extended Yale B.

6.2. LQI–based Adaptive illumination normalization for face recognition.

Histogram Equalisation (HE) has been used as a mean to improve face recognition when the sample image suffers from poor illumination. In extreme cases when the presented sample is poorly illuminated HE improves the chance of recognition, but there are side effects and there are evidences that HE does reduce image quality and recognition accuracy in the cases of well lit images. An analysis of the effect of HE on the recognition accuracy of the various single-subband wavelet face recognition schemes for the different subsets of images in the Extended Yale B database has confirmed these conclusions, ([36], [50]). For the three level 2 wavelet subbands (LL2, LH2, and HL2), applying HE yields a reasonable-to-significant improvement in accuracy for sets 4 and 5; while the accuracy dropped as a result of HE application for sets 1, 2 and 3. What is also interesting, in this regard, is that as a result of the application of HE the values of LQI improved significantly for images in sets 4 and 5 but to a much lesser extent in set 3, while the LQI values in sets 1 and 2 has deteriorated greatly. The LQI of all images in sets 4 and 5 after HE became > 0.78.

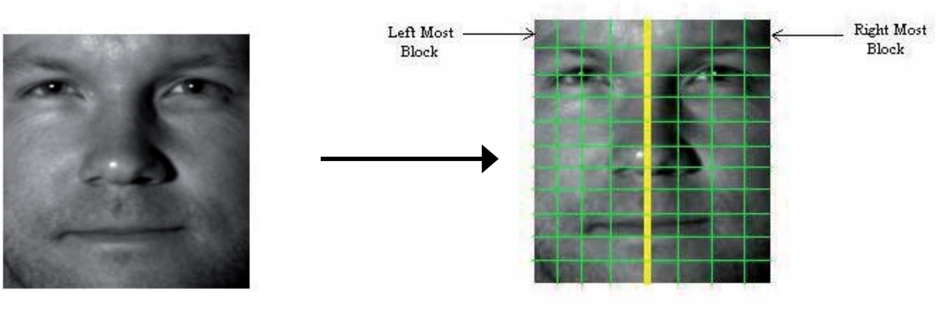
These observation and the remarks, at the end of the section 6.1 provide the perfect threshold adopted by Sellahewa et al, [36], for the first Image quality based adaptive illumination normalisation procedure and the adaptive face recognition. The use of the threshold of 0.8 for LQI below which HE is applied, has led to improved face recognition in the different single subband streams as well as in the multi-stream cases. The improvement was across all subsets but to varying degrees and more significantly in sets 4 and 5, (for more details see [36]). The identification error rates for some multi-stream wavelet schemes will be presented and discussed in the last subsection. AHE refers to this LQI-based adaptive use of HE.

6.3 No-Reference Quality Index

The choice of reference image for image quality assessment is a factor that may affect the adaptive normalisation procedures and it may not be a simple task. Defining image quality measures without a reference image is a desirable task and more so for illumination assessment in relation to face images. The frontal pose of a human face is more or less symmetric; hence it is easy to design a symmetric version of the LQI without the need for a reference image. A without a reference luminance quality measure, can be defined in a two steps process, where in the first step the LQI for the left half of a face image is measured with reference to its right half and the second step uses a form of histogram partitioning that aims to measure some aspects of distortion from normal face histogram.

Step 1. The symmetric adaptive local quality index (SALQI). For a face image (I), SALQI is defined as follows:

1. Divide I into left and right half sub-images, I_L and I_R respectively, and let I_{FR} be the horizontal flipping of I_R.
2. Starting from the top left corner, use equation (3), above, to compute LQI of the 8x8 windows in I_{FR} with respect to the corresponding windows in I_L, as indicated below



3. After calculating the quality map $\{m_i = LQI_i: i=1,\ldots.,N\}$, a pooling strategy as indicated in equations (4) and (5) to calculate the final quality-score of the image (I) as a weighted average of the m_i's:

$$Q = \frac{\sum_{i=1}^{N} m_i * w_i}{\sum_{i=1}^{N} w_i} \tag{4}$$

$$\text{where, } w_i = g(x_i, y_i) \text{ , and } \quad g(x, y) = \sigma_x^2 + \sigma_y^2 + C \tag{5}$$

Here, $x_i = I_{L,i}$ and $y_i = I_{FR,i}$, where $I_{FR,i}$ is the mirrored block of $I_{L,i}$ of a row. The C is a constant representing a baseline minimal weight. The value range of SALQI is [0, 1] and its equals 1 if and only if the image is perfectly symmetrically illuminated.

Step 2. The Middle Half index (MH). The SALQI provides an indication of how symmetrical the light is distributed, but it does not distinguish between a well-lit face images from an evenly dark image. SALQI produces high quality scores for such images. To overcome this problem we use histogram partitioning**:** A good quality image normally has a dynamic range covering the full grey scale and its histogram covers well the middle part. The MH index is thus defined as:

$$MH = \frac{Middle}{Bright + Dark} \tag{6}$$

Where, **Middle =** No. of pixels in the middle range between a Lower bound LB and an Upper bound UB,

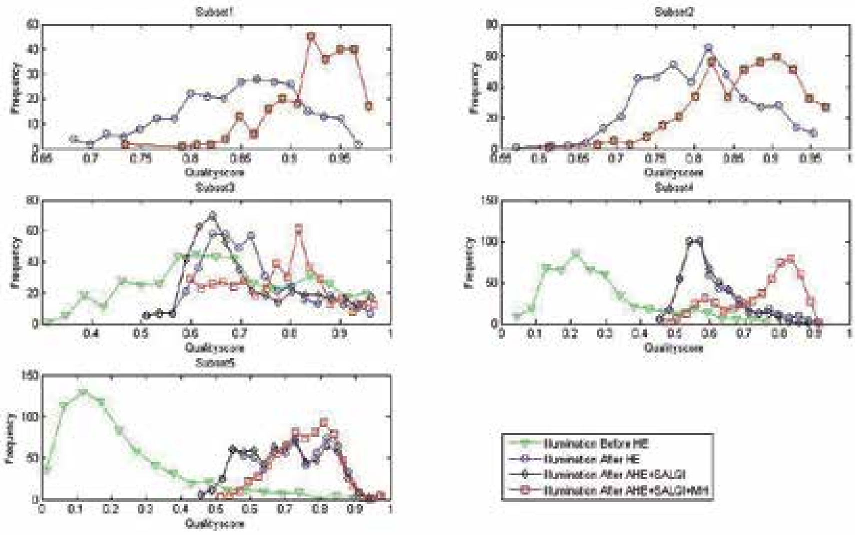
Bright = No. of pixels in the bright region of the histogram greater than UB,

Dark = No. of pixels in the dark region of the histogram less than LB,

Examining a number of so-called normal images, the LB and UB are set at 63 and 243, respectively. The MH value ranges from 0 to Max = (M/2), where M is the size of the image. The larger MH is, the better the quality is. Its maximum value depends on the image dataset.

6.4 The Symmetric Adaptive Histogram Equalisation (SAHE)

This is another adaptive scheme that uses both the SALQI and MH values to control the use of HE. Chart 6.1 displays the distribution of image LQI, SALQI and MH indices in the various subsets of the extended Yale B data base before and after the application of HE, AHE, SALQI version of AHE, and the full SAHE. For subset 1 and subset 2, we see that the application of HE results in deterioration in quality, and both AHE and MH maintain the same original quality. This confirms that for well lit images that exhibit similar illumination characteristics to those in subsets 1 and 2 (i.e. SALQI > 0.65) no normalisation is needed. The other 3 subsets benefit, to varying degrees, from pre-processing. But they benefit more from the full version of SAHE which includes the use of MH.



Charts 6.1. Distribution of for extended Yale B database before and after various normalisation.

A. Aboud et al in [37] have tested the performance of an SAHE-adaptive wavelet-based face recognition scheme in comparison with the corresponding versions with no normalization, and with the LQI-based adaptive which only used a single threshold (approx. 0.8). In the corresponding experiments, two different wavelets are used: Daubechie-1 (i.e Haar), and Daubechie-2 (also known as Daub 4), at three decomposition levels. Again the testing was based on the Extended Yale B database. The dataset are divided into two groups: training set and testing set. The training set has (38) images, one image per subject which is chosen to be (P00A+000E+00). The testing set consists of all the remaining (2394) images, i.e. 63 images per subject. Different values of SALQI and MH quality indices have been used as thresholds for SAHE approach. Recognition results, displayed in Figure 6.2, show that the LH2 subband gives the best results under varying illumination and the error rate for the SAHE with SALQI <0.6, is about 0.30% less than what was achieved by the LQI–based AHE application. However, SAHE resulted in slightly increased error rates for subset 3 images while reduced the errors of subset 4 and subset 5. The results for LL2 features are significantly better, although these error rates are much higher than the errors with LH2.

```
1. Calculate the quality scores for the image (I) using ( SALQI ) and ( MH )
2. If (SALQI < Thershold1) and (MH < Threshold 2) Then
        IF (MH < Thershold3) Then  {Apply normalization algorithm on the whole image (I)}
        Else if  (MH >= Thershold3) Then
            a.    Apply HE on the left region of image (I) and compute SALQI
            b.    Apply HE on the right region of image (I) and compute SALQI
            c.    Apply HE on left and right regions of the image (I) and compute SALQI
            Select the case that has higher SALQI value
        End if
3.  Else if  ( SALQI >= Thershold1 )  and  ( MH >= Thershold2 ) Then
        {Do not apply histogram normalization algorithm on image (I)}
4. End if
```

Fig. 6.4 Symmetrical Adaptive Histogram Equalization Algorithm

No pre-process	8.89	18.20	83.30	95.82	97.20	70.71
HE, ZN	3.11	25.88	70.99	90.11	85.57	64.52
AHE, LQI < 0.80	2.67	22.81	69.01	90.11	84.03	63.05
SAHE, SALQI < 0.60	2.67	7.89	37.8	73.76	76.61	48.36
SAHE, SALQI < 0.70	2.67	7.89	38.02	73.76	76.47	48.36
SAHE, SALQI < 0.80	2.67	20.83	40	73.76	76.47	51.22
SAHE, SALQI < 0.90	2.67	7.89	38.24	75.1	76.05	48.57

(a) Wavelet Haar, suband: LL2

No pre-process	8.00	0.00	30.55	71.10	95.24	50.97
HE, ZN	7.56	0.44	17.58	26.62	14.15	14.31
AHE, LQI < 0.80	7.11	0	11.65	20.34	11.76	10.94
SAHE, SALQI < 0.60	7.11	0	12.97	18.25	11.34	10.61
SAHE, SALQI < 0.70	7.11	0	12.97	18.63	11.48	10.73
SAHE, SALQI < 0.80	7.11	0	12.53	18.44	12.32	10.86
SAHE, SALQI < 0.90	7.11	0	12.75	18.63	11.34	10.65

(b) Wavelet Haar, suband: LH2

	Set1	Set2	Set3	Set4	Set5	All
No pre-process	8.44	14.25	80.66	95.63	97.20	69.36
HE, ZN	1.78	20.83	67.47	90.30	85.71	62.84
AHE, LQI < 0.80	0.89	17.54	64.84	90.87	84.45	61.36
SAHE, SALQI < 0.60	0.89	4.61	30.99	72.05	77.03	46
SAHE, SALQI < 0.70	0.89	4.61	31.21	71.86	76.89	45.96
SAHE, SALQI < 0.80	0.89	15.79	33.19	72.05	77.03	48.57
SAHE, SALQI < 0.90	0.89	4.61	31.43	73.38	76.47	46.21

(c) Wavelet Daub 4, suband: LL2

	Set1	Set2	Set3	Set4	Set5	All
No pre-process	14.67	0	35.60	66.35	89.64	49.83
HE, ZN	13.33	0	24.84	28.33	18.35	17.80
AHE, LQI < 0.80	13.33	0	21.76	22.24	16.11	15.19
SAHE, SALQI < 0.60	13.33	0	20.22	21.48	15.83	14.65
SAHE, SALQI < 0.70	13.33	0	20.22	21.48	15.83	14.65
SAHE, SALQI < 0.80	13.33	0	20.66	21.48	16.39	14.90
SAHE, SALQI < 0.90	13.33	0	20.22	21.29	15.69	14.56

(d) Wavelet Daub4, suband: LH2

Table 6.2 Identification error rates of wavelet-based face recognition system

6.5 Regional LQI and Adaptive fusion of multi stream face recognition

The previous parts of this section demonstrated the suitability of using the AHE and SAHE as a mean of controlling the application of illumination normalisation procedure (HE) and the benefits that this yields for single and multi-stream face recognition schemes. However, in real-life scenarios, variations in illumination between enrolled and test images could be confined to a region, rather than the whole, of the face image due to the changes in the direction of the light source or pose. Therefore, it is sensible to measure the illumination quality on a region-by-region basis. Sellahewa et al, [48], has experimented with a rather simple regional modification of the LQI, whereby we split the image into 2x2 regions of equal size, and tested the performance of the Regional AHE based adaptive multi-stream face recognition. Figure 6.5 and Figure 6.6 present that Identification error rate for the RLQI-based fusion of (LL2, LH2) and (LH2, HL2), respectively, using 10 different weighting configurations.

LL2 + LH2		Identification Error Rates (%)					
WLL	WLH	Set 1	Set 2	Set 3	Set 4	Set 5	Total
1.0	**0.0**	3.56	17.54	38.24	73.57	75.91	50.13
0.9	**0.1**	2.22	7.68	30.11	65.78	70.73	43.27
0.8	**0.2**	2.22	3.29	22.64	57.03	63.31	36.83
0.7	**0.3**	1.33	0.44	18.46	46.2	51.26	29.38
0.6	**0.4**	2.22	0.00	15.16	36.12	38.94	22.81
0.5	**0,5**	3.56	0.00	14.73	27.95	27.17	17.51
0.4	**0.6**	4.89	0.00	13.41	19.96	18.91	13.13
0.3	**0.7**	5.78	0.00	12.97	17.87	14.57	11.36
0.2	**0.8**	8.80	0.00	14.07	15.59	11.62	10.4
0.1	**0.9**	8.89	0.00	13.85	13.5	10.36	9.6
0.0	**1.0**	10.67	0.00	14.95	14.45	10.36	10.19

Fig. 6.5 Non-Adaptive Fusion (LL, LH)

LH2 + HL2		Identification Error Rates (%)					
WLH	WHL	Set 1	Set 2	Set 3	Set 4	Set 5	Total
1.0	**0.0**	10.67	0.00	14.95	14.45	10.36	10.19
0.9	**0.1**	8.44	0.00	14.73	12.55	9.52	9.26
0.8	**0.2**	7.56	0.00	12.75	12.74	9.38	8.8
0.7	**0.3**	5.78	0.00	11.65	11.41	9.8	8.25
0.6	**0.4**	5.33	0.00	9.23	10.65	11.76	8.16
0.5	**0,5**	4.44	0.00	7.47	10.65	13.17	8.16
0.4	**0.6**	3.11	0.00	7.47	11.98	18.49	9.93
0.3	**0.7**	2.22	0.00	8.13	15.78	24.37	12.58
0.2	**0.8**	5.33	0.00	9.23	19.96	36.97	17.8
0.1	**0.9**	7.11	0.00	12.75	29.09	51.68	25.08
0.0	**1.0**	9.33	0.88	15.16	39.35	61.62	31.19

Fig. 6.6 Non-Adaptive Fusion (LH, HL)

The use of RLQI has obviously resulted in further improvement in accuracy of multi-stream recognition schemes. With best overall error rate of 9.6 for the (LL2, LH2) fused scheme achieved when LL2" was given a small weight of 0.1, while best error rate for the (LH2, HL2) fused scheme is 8.16 achieved when have nearly equal weights. What is more interesting is that the best performance over the different sets is achieved with different weighting configurations in both cases. This shows that the wavelet-based multi-stream recognition scheme, developed previously, has no objective means of selecting fusion parameters and that it performed differently for face images captured with different lighting conditions has led to developing of a new adaptive approach to face recognition. This suggests a dynamic scheme of weighting that depends on image quality. Figure 6.7, below, presents the results obtained for using quality-based adaptive fusion of two or 3 subbands. In this case if the LQI of the image is>0.9 then the score for LL2 will be given a 0.7 weighting otherwise it is given a 0 weighting. The LH2 and HL2 subbands get equal proportion from the left over.

Feature	**Subband**	**Identification Error rate %**					
		Subset 1	**Subset 2**	**Subset 3**	**Subset 4**	**Subset 5**	**All**
subband	**LL2**	3.56	17.54	38.24	73.57	75.91	50.13
	LH2	10.67	0.00	14.95	14.45	10.36	10.19
	HL2	9.33	0.88	15.16	39.35	61.62	31.19
Adaptive Furion	**LL2+LH2**	2.22	0.00	14. 73	14.45	10.36	9.34
	LL2+LH2+HL2	7.47	0.22	1.78	10.65	13.17	7.95

Fig. 6.7 Adaptive Fusion

It is clear that, this dynamic choice of weighting of the scores has led to further improvement over the non-adaptive static selection of weighting.

7. CONCLUSIONS AND FUTURE WORK

In this chapter we have reviewed face recognition schemes, and in particular we advocated the use of wavelet-based face recognition. The fact that a wavelet-transform of face image into a number of different subbands representing the face at different frequency range and different scales, has been exploited to develope several single-stream face recognition schemes one for each wavelet subband. The performances of several of these were tested over a number of benchmark databases, which revealed different error rates, but achieving comparable/better results compared to PCA based schemes. This approach has the advantage of being very efficient and being scalable.

We have also shown that one mimicked the success of fusion approach to multi-modal biometric-based recognition by using multi-stream face recognition that is based on fusing a number of single streams. Even the fusion of a small (<4) number of single streams has led to significant improvement in performance.

Finally, we have demonstrated with a significant degree of success that the challenge of face recognition in the presence of extreme variation illumination can be dealt with using adaptive quality –based face recognition. The main advantages of using quality measures are the avoidance of excessive unnecessary enhancement procedures that may cause undesired artefacts, reduced computational complexity which is essential for real time applications, and improved performance.

The work on quality- based adaptive fusion and adaptive wavelet multi-stream wavelet face recognition will be expanded in the future to deal with other quality issues as well as efficiency challenges.

8. REFERENCES

1. W. Zhao, R. Chellappa, A. Rosefeld, and P. J. Phillips, "Face Recognition: A Literature Survey," Technical report, Computer Vision Lab, University of Maryland, 2000.
2. D. M. Etter. "The Link Between National Security and Biometrics", Proc. of SPIE Vol 5779, Biometric Technology for Human Identification II, pp 1-6, March 2005
3. T. Sim, R. Sukthankar, M. D. Mullin, and S. Baluja. "High-Performance Memory-based Face Recognition for Visitor Identification," *ICCV-99, Paper No. 374.*
4. M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 12, no. 1, pp. 103–108, 1990.
5. M. Turk and A. Pentland, "Eigenfaces for Recognition," Journal of Cognitive Neuroscience vol. 3, no. 1, pp. 71–86, 1991.
6. B Moghaddam, W Wahid and A pentland, Beyond eigenfaces: Probabilistic matching for face recognition, Proc. of face and gesture recognition, pp. 30 –35, 1998.
7. J. Yi, J. Kim, J. Choi, J. Han, and E. Lee, "Face Recognition Based on ICA Combined with FLD," in Biometric Authentication, M. Tistarelli and J. B. Eds, eds., Proc. Int'l ECCV Workshop, pp. 10–18, June 2002.
8. G. L. Marcialis and F. Roli, "Fushion of LDA and PCA for Face Verification," in Biometric Authentication, M. Tistarelli and J. B. Eds, eds., Proc. Int'l ECCV Workshop, pp. 30–37, June 2002.
9. R. Cappelli, D. Maio, and D. Maltoni, "Subspace Classification for Face Recognition," in Biometric Authentication, M. Tistarelli and J. B. Eds, eds., Proc. Int'l ECCV Workshop, pp. 133–141, June 2002.
10. J.-T. Chien and C.-C. Wu, "Discriminant Wavelet faces and Nearest Feature Classifiers for Face Recognition," in IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 24, no. 12, pp. 1644–1649, December 2002.
11. M. Bicego, E. Grosso, and M. Tistarelli, "Probabilistic face authentication using Hidden Markov Models" in Biometric Technology for Human Identification II, Proc. SPIE vol. 5779, pp. 299–306, March 2005.
12. J.G. Daugman, "Two-Dimensional Spectral Analysis of Cortical Receptive Field Profile," Vision Research, vol. 20, pp. 847-856, 1980.
13. J.G. Daugman, "Complete Discrete 2-D Gabor Transforms by Neural Networks for Image Analysis and Compression," IEEE Trans. Acoustics, Speech, and Signal Processing, vol. 36, no. 7, pp. 1,169-1,179, 1988.

14. M. Lades, J.C. Vorbrüggen, J. Buhmann, J. Lange, C. von der Malsburg, R.P. Würtz, and W. Konen, "Distortion Invariant Object Recognition in the Dynamic Link Architecture," IEEE Trans. Computers, vol. 42, no. 3, pp. 300–311, 1993.
15. L. Wiskott, J-M Fellous, N. Krüger, and C. Malsburg, "Face Recognition by Elastic Bunch Graph Matching", IEEE Tran. On Pattern Anal. and Mach. Intell., Vol. 19, No. 7, pp. 775-779, 1997.
16. Z. Zhang, M. Lyons, M. Schuster, and_ S. Akamatsu, "Comparison Between Geometry-Based and Gabor-Wavelets-Based Facial Expression Recognition Using Multi-Layer Perceptron", Proc. 3rd IEEE Int. Conf. on Automatic Face and Gesture Recognition, Nara Japan, IEEE Computer Society, pp. 454-459 (1998).
17. Chengjun Liu and Harry Wechsler, "Independent Component Analysis of Gabor features for Face Recognition", IEEE Trans. Neural Networks, vol. 14, no. 4, pp. 919-928, 2003.
18. I. Daubechies, "The Wavelet Transform, Time-Frequency Localization and Signal Analysis," IEEE Trans. Information Theory, vol. 36, no. 5, pp. 961-1004, 1990.
19. K. Etemad and R. Chellappa, "Face Recognition Using Discriminant Eigenvectors," Proc. IEEE Int'l Conf. Acoustic, Speech, and Signal Processing, pp. 2148-2151, 1996.
20. Dao-Qing Dai, and P. C. Yuen. "Wavelet-Based 2-Parameter Regularized Discriminant Analysis for Face Recognition," Proc. AVBPA Int'l Conf. Audio-and Video-Based Biometric Person Authentication, pp. 137-144, June, 2003.
21. D. Xi, and Seong-Whan Lee. "Face Detection and Facial Component Extraction by Wavelet Decomposition and Support Vector Machines," Proc. AVBPA Int'l Conf. Audio-and Video-Based Biometric Person Authentication, pp. 199-207, June, 2003.
22. F. Smeraldi. "A Nonparametric Approach to Face Detection Using Ranklets," Proc. AVBPA Int'l Conf. Audio-and Video-Based Biometric Person Authentication, pp. 351-359, June, 2003.
23. J.H. Lai, P. C. Yuen!, G. C. Feng," Face recognition using holistic Fourier invariant features", Pattern Recognition 34, pp. 95-109, (2001)
24. N. G. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals", J.of Appl. And Comp. Harmonic Analysis, 01 (3), pp. 234-253, May 2001.
25. Y. Peng, X. Xie, W. Xu, and Q. Dai, "Face Recognition Using Anistropic Dual-Tree Complex Wavelet Packets", Proc IEEE Inter. Conf. on Pattern Recognition, 2008.
26. Sidney Burrus Ramesh, C, A Gopinath, and Haittao Guo. Introduction to Wavelet and Wavelet Transforms A Primer. Prentice Hall. Inc., 1998.
27. Naseer AL-Jawad, "Exploiting Statistical Properties of Wavelet Coefficients for Image/Video Processing and Analysis", DPhil Thesis, University of Buckingham, 2009.
28. D. Xi, and Seong-Whan Lee. "Face Detection and Facial Component Extraction by Wavelet Decomposition and Support Vector Machines," Proc. AVBPA Int'l Conf. Audio-and Video-Based Biometric Person Authentication, pp. 199-207, June, 2003.
29. A. Z. Kouzani, F. He, and K. Sammut. "Wavelet Packet Face Representation and Recognition," Proc IEEE Conf. Systems, Man, and Cybernetics, pp. 1614-1619, 1997.
30. Dao-Qing Dai, and P. C. Yuen. "Wavelet-Based 2-Parameter Regularized Discriminant Analysis for Face Recognition," ProcComputer vol. 33, no. 2, pp. 50–55, February 2000.
31. H. Sellahewa,"Wavelet–based Automatic Face Recognition for Constrained Devices", Ph.D. Thesis, University Of Buckingham, (2006).

32. E. Bailly-Bailli´ere, S. Bagnio, F. Bimbot, M. Hamouz, J. Kittler, J. Mari´ethoz, J. Matas, K. Messer, V. Popovici, F. Por´ee, B. Ruiz, and J. Thiran, “The BANCA Database Evaluation Protocol,” in Audio-and Video-Based Biometric Person Authentication, Proc. AVBPA Int'l Conf, pp. 625–638, June 2003.
33. G C Feng a b, P C Yuen b and D Q Dai, “Human Face Recognition Using PCA on Wavelet Subband”, SPIE Journal of Electronic Imaging 9 (2), pp: 226–233, (2000).
34. M-H. Yang, “Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition Using Kernel methods” in Automatic Face and Gesture Recognition, Proc. IEEE Int'l Conf, pp. 215–220, 2002.
35. P. N. Belhumeur and D. J. Kriegman, “What is the set of images of an object under all possible lighting conditions?,” in Proc. IEEE Conf
36. Harin Sellahewa, Sahah A. Jassim, "Illumination and Expression Invariant Face Recognition: Toward Sample Quality –based Adaptive Fusion", in Biometrics: Theory, Applications and Systems, Proc. Second IEEE Int'l Conference 10414693, 1-6 (2008).
37. A. Aboud, H.Sellahewa, and S. A. Jassim, “Quality Based Approach for Adaptive Face Recognition”, in Proc. SPIE, Vol. 7351, Mobile Multimedia/image processing, Security, and Applications, Florid, April 2009.
38. Li Ma, Tieniu Tan,” Personal Identification Based on Iris Texture Analysis”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(12), 20-25 (2003).
39. Oriana Yuridia Gonzalez Castillo, "Survey about Facial Image Quality", Fraunhofer Institute for Computer Graphics Research, 10-15 (2005).
40. Yi Chen, Sarat C. Dass, and Anil K. Jain, " Localized Iris Image Quality using 2-D Wavelets ", Proc. Advances in Biometrics Inter. Conf., ICB 2006, 5-7 (2006).
41. Sarat C. Dass, Anil K. Jain, "Quality-based Score Level Fusion in Multibiometric Systems", Proc. 18th Inter. Conf. on Pattern Recognition 4(1), 473 - 476 (2006).
42. Jonas Richiardi, Krzysztof Kryszczuk, Andrzej Drygajlo, "Quality Measures in Unimodal and Multimodal Biometric Verification ", Proc. 15th European Conference on Signal Processing EUSIPCO,2007.
43. Hanghang Tong, Mingjing Li, Hong-Jiang Zhang, Changshui Zhang, Jingrui He. "Learning No-Reference Quality Metric by Examples", Proc. The 11th International Multi-Media Modelling Conferene 1550, 247- 254 (2005)
44. Robert Yen, "A New Approach for Measuring Facial Image Quality", Biometric Quality Workshop II, Online Proc. National Institute of Standards and Technology, 7-8, (2007).
45. Krzysztof Kryszczuk, Andrzej, "Gradient–based Image Segmentation for Face Recognition Robust to Directional Illumination", Proc. SPIE, 2005.
46. Azeddine Beghdadi, "A New Image Distortion Measure Based on Wavelet Decomposition", Proceeding of IEEE ISSPA2003, 1-4 (2003).
47. Alan C. Bovik, Zhou Wang",A Universal Image Quality Index", IEEE Signal Processing Letters. 9(.3), 81-84 (2002).
48. Harin Sellahewa, Sahah A. Jassim, “Image Quality-based Face Recognition” To appear in IEEE Trans. On Instrumentation And Measurements, Biometrics, 2009.
49. Georghiades, A.S. and Belhumeur, P.N. and Kriegman, D. J, "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose", IEEE Transactions on Pattern Analysis and Machine Intelligence"23(6), 643-660 (2001).
50. Harin Sellahewa, Sahah A. Jassim, “Image Quality-based Adaptive Illumination Normalisation for Face Recognition” in Proc. SPIE Conf. on Biometric Technology for Human Identification, Florid, April 2009.

8

Face Recognition in Ideal and Noisy Conditions Using Support Vector Machines, PCA and LDA

Miloš Oravec[1], Ján Mazanec[1], Jarmila Pavlovičová[2],
Pavel Eiben[1] and Fedor Lehocki[1]
[1] Dept. of Applied Informatics and Information Technology, [2] Dept. of Telecommunications, Faculty of Electrical Engineering and Information Technology,
Slovak University of Technology, Ilkovičova 3, 812 19 Bratislava
Slovak Republic

1. Introduction

In this chapter, we consider biometric recognition based on human face. Biometrics became frequently used in automated systems for identification of people (Jain et al., 2004) and huge interest is devoted to the area of biometrics at present (Jain et al., 2008; Shoniregun & Crosier, 2008; Ross et al, 2006).

Along with well-known methods such as fingerprint or DNA recognition, face image already opened new possibilities. Face recognition has been put into real life by many companies. It is already implemented in image organizing software (e.g. Google's Picasa: http://www.deondesigns.ca/blog/picasa-3-5-adds-face-recognition/), web applications (e.g. web photo albums http://picasa.google.com/intl/en_us/features-nametags.html) and even in commercial compact cameras (e.g. Panasonic Lumix). Passports contain face biometric data since 2006 (EU – Passport Specification, 2006).

In the area of face recognition, a class represents all images of the same subject (person). The goal is to implement an automated machine supported system that recognizes well the identity of a person in the images that were not used in a training phase (an initialization and training by representative sample of images precede an evaluation phase). Various applications are possible, e.g. automated person identification, recognition of race, gender, emotion, age etc. The area of face recognition is well described at present, e.g. starting by conventional approaches (PCA, LDA) (Turk & Pentland1991; Marcialis & Roli, 2002; Martinez & Kak, 2001), and continuing at present by kernel methods (Wang, et al., 2008; Hotta, 2008; Wang et al., 2004; Yang, 2002; Yang et al., 2005). Advances in face recognition are summarized also in books (Li & Jain, 2005; Delac et al., 2008) and book chapters (Oravec et al., 2008).

Our aim is to present complex view to biometric face recognition including methodology, settings of parameters of selected methods (both conventional and kernel methods), detailed recognition results, comparison and discussion of obtained results using large face database.

The rest of this chapter is organized as follows: In section 2, we present theoretical background of methods used for face recognition purposes - PCA (Principal Component

Analysis), LDA (Linear Discriminant Analysis) and SVM (Support Vector Machines). Section 3 provides information about FERET database (FERET Database, 2001), since large image set from this database including total 665 images is used in our experiments. The face images are first preprocessed (normalization with respect to size, position and rotation and also contrast optimization and face masking). In Section 4, face recognition methods that are used in the rest of the chapter are discussed. We also propose methods utilizing PCA and LDA for extracting the features that are further classified with SVM and compare them to usual approaches with conventional classifiers. Section 5 presents results of recognition systems in ideal conditions. We show that proposed methods result in excellent recognition rate and robustness. Also behavior of presented methods is analyzed in detail and best settings for these methods are proposed. Section 6 is devoted to the influence of input image quality to face recognition accuracy. For this purpose, we use best parameter settings we obtained running 600 tests in ideal conditions. Gaussian noise, salt & pepper noise and speckle noise with various intensities are included. This enables to get insight into face recognition system robustness. Also equivalence of different types of noise from the recognition point of view is discussed.

2. Face Recognition Methods and Algorithms

We use different methods in our single-stage and two-stage face recognition systems: PCA (Principal Component Analysis), LDA (Linear Discriminant Analysis) and SVM (Support Vector Machines). The role of PCA and LDA falls into feature extraction. We use different classifiers that are in the form of both simple metrics and more complex SVMs.

2.1 Principal Component Analysis PCA

This standard statistical method can be used for feature extraction. Principal component analysis PCA (Turk & Pentland, 1991; Marcialis & Roli, 2002; Martinez & Kak, 2001; Haykin, 1994; Bishop, 1995) reduces the dimension of input data by a linear projection that maximizes the scatter of all projected samples. Let $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N\}$ be a set of N sample images of dimensionality n belonging to one of c classes $\{X_1, X_2, ..., X_c\}$. Its covariance (total scatter) matrix is

$$\mathbf{S}_T = \sum_{k=1}^{N} (\mathbf{x}_k - \mu)(\mathbf{x}_k - \mu)^T \tag{1}$$

PCA transforms input images to new feature vectors

$$\mathbf{y}_k = \mathbf{W}^T \mathbf{x}_k \qquad k = 1,2,...,N, \tag{2}$$

where $\mathbf{W} \in \Re^{n \times m}$ is a transform matrix with orthonormal columns and $\mu \in \Re^n$ is the mean image of all sample images. This yields also in dimensionality reduction ($m < n$). The scatter of the transformed feature vectors $\{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_N\}$ is $\mathbf{W}^T \mathbf{S}_T \mathbf{W}$. In PCA, the projection $\mathbf{W}_{opt}$ maximizes the determinant of the total scatter matrix of the projected samples

$$\mathbf{W}_{opt} = \arg\max_{\mathbf{W}}\left(\det\left(\mathbf{W}^T \mathbf{S}_T \mathbf{W}\right)\right) = \left[\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\right], \tag{3}$$

where $\left[\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_N\right]$ is the set of *n*-dimensional eigenvectors (called eigenfaces when applying PCA to face images) of $\mathbf{S}_T$ corresponding to the *m* largest eigenvalues $\left[\lambda_1, \lambda_2, ..., \lambda_m\right]$. Thus, PCA maximizes the total scatter - this is the disadvantage of this method.

2.2 Fisher's Linear Discriminant FLD, Linear Discriminant Analysis LDA

Fisher's Linear Discriminant (FLD) (Marcialis & Roli, 2002; Martinez & Kak, 2001; Bishop, 1995; Belhumeur et al., 1997; Oravec & Pavlovičová, 2004; Duda & Hart, 1973) shapes the scatter with the aim to make it more suitable for classification. A computation of the transform matrix results in maximization of the ratio of the between-class scatter and within-class scatter.

Between-class scatter matrix $\mathbf{S}_B$ and within-class scatter matrix $\mathbf{S}_W$ are defined by

$$\mathbf{S}_B = \sum_{i=1}^{c} N_i \left(\mu_i - \mu\right)\left(\mu_i - \mu\right)^T \tag{4}$$

$$\mathbf{S}_w = \sum_{i=1}^{c} \sum_{\mathbf{x}_k \in X_i} \left(\mathbf{x}_k - \mu_i\right)\left(\mathbf{x}_k - \mu_i\right)^T \tag{5}$$

respectively, where N_i is the number of samples in class X_i and μ_i is the mean image of class X_i. The transform matrix $\mathbf{W}_{opt}$ maximizes the ratio of the determinant of the between-class scatter matrix of the projected samples to the determinant of the within class scatter matrix of the projected samples:

$$\mathbf{W}_{opt} = \arg\max_{\mathbf{W}} \frac{\det\left(\mathbf{W}^T \mathbf{S}_B \mathbf{W}\right)}{\det\left(\mathbf{W}^T \mathbf{S}_W \mathbf{W}\right)} = \left[\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\right] \tag{6}$$

where $\left[\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_m\right]$ is the set of generalized eigenvectors of $\mathbf{S}_B$ and $\mathbf{S}_W$ corresponding to the *m* largest generalized eigenvalues $\left\{\lambda_1, \lambda_2, ..., \lambda_m\right\}$:

$$\mathbf{S}_B \mathbf{w}_i = \lambda_i \mathbf{S}_W \mathbf{w}_i \quad i = 1, 2, \ldots, m \tag{7}$$

There are at most $c-1$ nonzero generalized eigenvalues, i.e. the upper bound of *m* is $c-1$ (Belhumeur et al., 1997; Duda & Hart, 1973).

In (Marcialis & Roli, 2002), the eigenvectors of $\mathbf{S}_W^{-1} \mathbf{S}_B$ are the columns of $\mathbf{W}_{opt}$ and the authors show that this choice maximizes the ratio $\det\left(\mathbf{S}_B\right) / \det\left(\mathbf{S}_W\right)$.

In face recognition, the number of sample images *N* is typically much smaller than the number of pixels *n* in each image (so called small sample size problem). This is why $\mathbf{S}_W \in \Re^{n \times n}$ can be singular. The rank of $\mathbf{S}_W$ is at most $N-c$. In (Belhumeur et al., 1997), authors solve the problem of singular $\mathbf{S}_W$ by proposal of alternative criterion to that of (6). At first, sample images are projected into lower dimensional space using PCA. This results in nonsingular $\mathbf{S}_W$. PCA reduces the dimension of the feature space to $N-c$, and then standard FLD (6) is applied to reduce the dimension to $c-1$. This method is called Fisherfaces. Then $\mathbf{W}_{opt}$ can be computed as follows:

$$\mathbf{W}_{opt} = \mathbf{W}_{FLD}^T \mathbf{W}_{PCA}^T \tag{8}$$

where

$$\mathbf{W}_{PCA} = \arg\max_{\mathbf{W}} \left(\det\left(\mathbf{W}^T \mathbf{S}_T \mathbf{W}\right)\right) \tag{9}$$

$$\mathbf{W}_{FLD} = \arg\max_{\mathbf{W}} \frac{\det\left(\mathbf{W}^T \mathbf{W}_{PCA}^T \mathbf{S}_B \mathbf{W}_{PCA} \mathbf{W}\right)}{\det\left(\mathbf{W}^T \mathbf{W}_{PCA}^T \mathbf{S}_W \mathbf{W}_{PCA} \mathbf{W}\right)} \tag{10}$$

Optimization for $\mathbf{W}_{PCA}$ is performed over $N \times (N-c)$ matrices and optimization for $\mathbf{W}_{FLD}$ is performed over $(N-c) \times m$ matrices. The smallest $c-1$ principal components are discarded in PCA computation.

It is often said that algorithms based of LDA outperform those based on PCA. LDA is insensitive to significant variation in lighting direction (Marcialis & Roli, 2002; Belhumeur et al., 1997), and facial expression (Belhumeur et al., 1997). However in (Martinez & Kak, 2001), authors show that when the training data set is small, PCA achieves better results compared to LDA and that PCA is less sensitive to different training data sets.

2.3 Support Vector Machines SVM

Support vector machines SVM belong to kernel methods (Muller et al., 2001; Hofmann et al., 2008) and play a major role in present machine learning algorithms.

Kernel algorithms map data $[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N] \in \Re^p$ from an original space x into a higher dimensional feature space F using a nonlinear mapping Φ (Muller et al., 2001)

$$\phi : \Re^p \to F, \mathbf{x} \to \phi(\mathbf{x}) \tag{11}$$

An original learning algorithm from original space is used in the feature space. High-dimensional space increases complexity of a problem; fortunately, it can be solved. Computation of a scalar product between two feature space vectors can be done using kernel function k

$$\phi(\mathbf{x}) \cdot \phi(\mathbf{y}) = k(\mathbf{x}, \mathbf{y}) \tag{12}$$

Thus, using kernel functions, the feature space does not need to be computed explicitly, only inner products in the kernel feature space are taken into account. Gaussian radial basis function, polynomial, sigmoidal, and inverse multiquadrics function are used in a role of kernel functions. Every linear algorithm that uses scalar products only can implicitly be executed in high-dimensional feature space by using kernels. Nonlinear versions of linear algorithms can be constructed in this way (Muller et al., 2001).

The basic principle of data separation by SVM is demonstrated on a simplified example in Fig. 1.

SVM accomplishes the task of finding the optimal separating hyperplane by maximizing the margin between the hyperplane and the support vectors. Dashed lines in Fig. 1 containing support vectors are parallel with the separating hyperplane and they run through the samples that are nearest to the separating hyperplane.

The separating hyperplane is defined as

$$\mathbf{w}^T\mathbf{x}+b=0 \tag{13}$$

where $\mathbf{w}$ is vector of weight coefficients and $\mathbf{b}$ is bias. The task of finding optimal separating hyperplane is accomplished by minimizing

$$\mathbf{w}^T\mathbf{w}+C\sum_i \xi_i \tag{14}$$

according to

$$y_i\left(\mathbf{w}^T\mathbf{x}_i+b\right)\geq 1-\xi_i \tag{15}$$

where ξ_i is a slack variable that defines tolerance band around support vector and thus creates so called soft margin. The *C* variable controls the influence of this tolerance band.

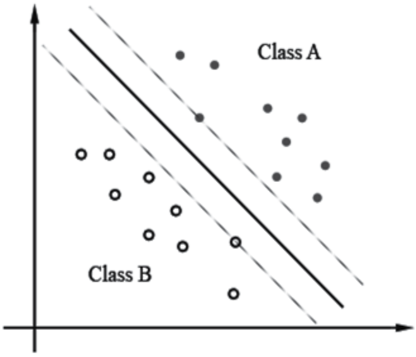


Fig. 1. Separation of data using SVM

Large amount of available papers, e.g. (Wang, et al., 2008; Hotta, 2008; Wang et al., 2004; Yang, 2002; Yang et al., 2005) indicates intensive use of SVMs and other kernel methods (kernel principal component analysis, kernel linear discriminant analysis, kernel radial basis function networks) also in face recognition area.

2.4 Metrics

Mahalinobis (also called Mahalanobis) Cosine (MahCosine) (Beveridge et al., 2003) is defined as the cosine of the angle between the image vectors that were projected into the

PCA feature space and were further normalized by the variance estimates. Let vectors $\mathbf{w}_i$ and $\mathbf{w}_j$ be image vectors in the unscaled PCA space (eigenvectors) and vectors $\mathbf{s}$ and $\mathbf{t}$ their projections in the Mahalinobis space. Using the fact that variance σ_i^2 of the PCA projections of input vectors to vector $\mathbf{w}_i$ equals to eigenvalue λ_i ($\lambda_i = \sigma_i^2$, where σ_i is the standard deviation), the relationships between the vectors are then defined as:

$$s_i = \frac{w_{ii}}{\sigma_i},\ t_i = \frac{w_{ji}}{\sigma_i} \tag{16}$$

The Mahalinobis Cosine is

$$\begin{aligned} S_{MahCosine}(\mathbf{w}_i, \mathbf{w}_j) &= \cos(\theta_{\mathbf{st}}) = \frac{|\mathbf{s}||\mathbf{t}|\cos(\theta_{\mathbf{st}})}{|\mathbf{s}||\mathbf{t}|} = \frac{\mathbf{s}\cdot\mathbf{t}}{|\mathbf{s}||\mathbf{t}|} \\ D_{MahCosine}(\mathbf{w}_i, \mathbf{w}_j) &= -S_{MahCosine}(\mathbf{w}_i, \mathbf{w}_j) \end{aligned} \tag{17}$$

(this is the covariance between the images in Mahalinobis space).

LDASoft (Beveridge et al., 2003) is LDA specific distance metric. It is similar to the Euclidean measure computed in Mahalinobis space with each axis weighted by generalized eigenvalue λ (also used to compute LDA basis vectors) raised to the power *0.2* (Zhao et al., 1999):

$$D_{LDAsoft}(\mathbf{w}_i, \mathbf{w}_j) = \sum_i \lambda_i^{0.2}(w_{ii} - w_{ji})^2 \tag{18}$$

3. Image database

For our tests, we used images selected from FERET image database (Phillips et al., 1998; Phillips et al., 2000). We worked with grayscale images from Gray FERET (FERET Database, 2001). FERET face images database is de facto standard database in face recognition research. It is a complex and large database which contains more than 14126 images of 1199 subjects of dimensions 256 x 384 pixels. Images differ in head position, lighting conditions, beard, glasses, hairstyle, expression and age of subjects. Fig. 2 shows some example images from FERET database.

We selected image set containing total 665 images from 82 subjects. It consists of all available subjects from whole FERET database that have more than 4 frontal images containing also corresponding eyes coordinates (i.e. we chose largest possible set fulfilling these conditions from FERET database). The used image sets are visualized in Fig. 3.

Recognition rates are significantly influenced by size of a training set. We used 3 different sets of images for training – i.e. two, three and four images per subject in the training set. Two, three or four images for training were withdrawn from FERET database according to their file name, while all remaining images from the set were used for testing purposes.

Prior to feature extraction, all images were preprocessed. Preprocessing eliminates undesirable recognition based on non-biometric data (e.g. "T-shirts recognition" or "haircut recognition"). Preprocessing includes following basic steps of converting original FERET image to a normalized image:

- Geometric normalization – aligning image according to available coordinates of eyes.
- Masking – cropping the image using an elliptical mask and image borders. In our experiments we tried two different maskings:
    - "face" - such that only the face from forehead to chin and cheek to cheek is visible
    - "BIGface" – leaving more of face surrounding compared to "face" – more potentially useful information is kept.
- Histogram equalization – equalizes the histogram of the unmasked part of the image.

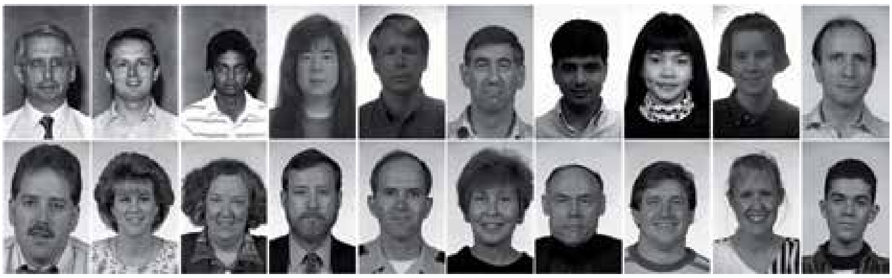


Fig. 2. Example of images from FERET database

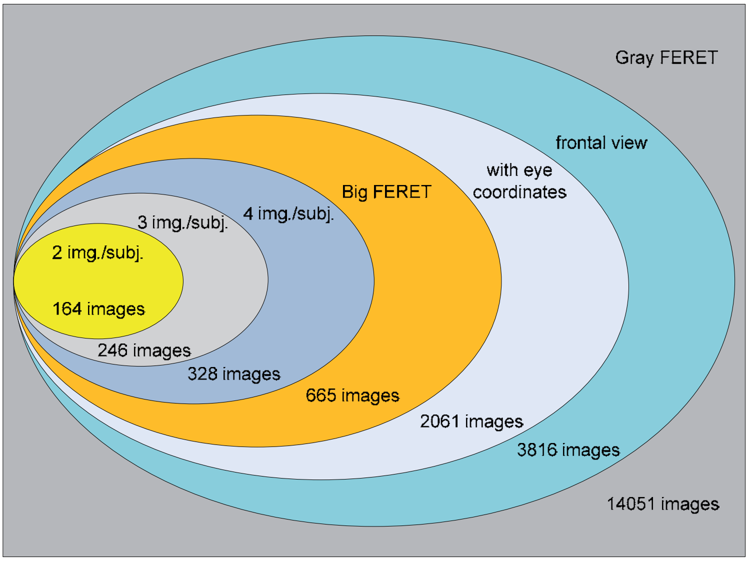


Fig. 3. Visualization of subset of images from FERET database used in our experiments

After preprocessing, the image size was 65x75 pixels. Fig. 4 shows an example of the original image, the image after "face" preprocessing and the image after "BIGface" preprocessing. All images from Fig. 2 preprocessed by "BIGface" preprocessing are shown in Fig. 5.

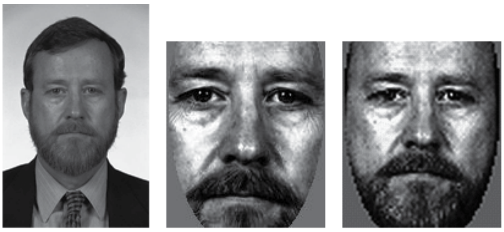


Fig. 4. Example of original image, image after "face" preprocessing and image after "BIGface" preprocessing

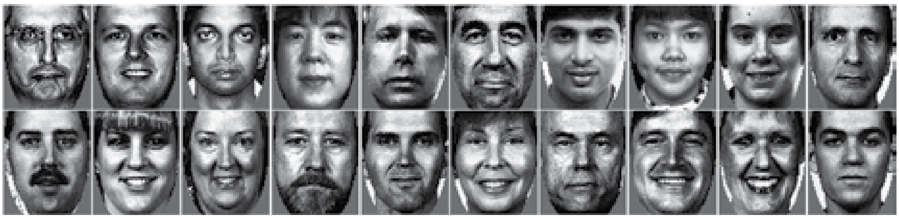


Fig. 5. Images from Fig. 2 preprocessed by "BIGface" preprocessing

4. Examined Methods for Face Recognition

We examined five different setups of face recognition experiments. They contain both single-stage and two-stage recognition systems as shown in Fig. 6:

- In single-stage face recognition (Fig. 6a), SVM is used for classification directly (i.e. there is no feature extraction performed).
- For two-stage face recognition setups including both feature extraction and classification (Fig. 6b - Fig. 6e), we used PCA with MahCosine metrics, LDA with LDASoft metrics and our proposed methods utilizing both PCA and LDA for feature extraction followed by SVM for classification. We propose also optimal parameter setups for the best performance of these methods.

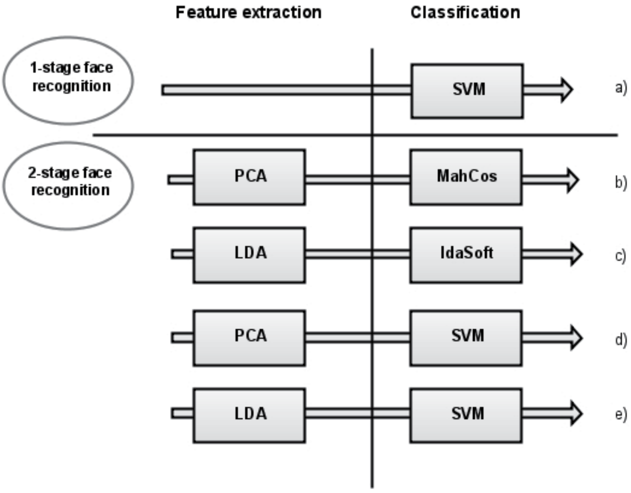


Fig. 6. Methods and classifiers used in our experiments

Last two setups (Fig. 6d) and e)) are our proposed combinations of efficient feature extraction combined with strong classifier. Fist three setups (Fig. 6a)-c)) are the conventional methods, presented for comparison with proposed approaches.

All five setups are significantly influenced by different settings of parameters of the examined methods (i.e. PCA, LDA or SVM). This is the reason we present serious analysis and proposal of parameter settings in following chapters.

We used CSU Face Identification Evaluation System (csuFaceIdEval) (Beveridge et. al., 2003) and libsvm - A Library for Support Vector Machines (LIBSVM, web) that implement mentioned algorithms.

5. Face Recognition Experiments and Results in Ideal Conditions

5.1 Single-Stage Recognition

SVM was directly used for recognizing faces without previous feature extraction from the images (see Fig. 6a)). Input images were of size 65x75 pixels.

In our tests we used SVM with the RBF (radial basis function) kernel

$$k\left(\mathbf{x}_{\mathrm{i}},\mathbf{x}_{\mathrm{j}}\right)=\exp\left(-\gamma\left\|\mathbf{x}_{\mathrm{i}}-\mathbf{x}_{\mathrm{j}}\right\|^{2}\right),\qquad \gamma>0 \tag{19}$$

where $\mathbf{x}_{\mathrm{i}},\mathbf{x}_{\mathrm{j}}$ are data points (face images) from original space.

It is important to find optimal parameters γ (gamma) and C, because different parameter setups are suitable for solving different problems. $C>0$ is the penalty parameter of the error term used in a determination of a separating hyperplane with the maximal margin in higher dimensional space by SVM. We used methodology from (Hsu et al., 2008), i.e. parameters search where the best *v*-fold cross-validation rate performed on training data suggests also the best parameter setup. *v*-fold cross-validation divides the training set into *v* subsets of equal size, and sequentially one subset is tested using the classifier that was trained on the remaining *v*-1 subsets. Fig. 7 shows example of the graph we used for parameter search – the dependence of cross validation rate on the parameters *C* and *gamma*. The best found parameters setups for all training sets and the results are shown in Table 1.

More images per subject in the training set result in better cross-validation rate and also better recognition rate. Difference between face recognition rate using "face" and "BIGface" preprocessing is noticeable only with 2 images per subject, where the result with "BIGface" preprocessing is approx. 5,6% worse than with "face" preprocessing.

It is important to point out that it is not possible to find "universal" values of parameters *C* and *gamma* that would lead to the best recognition rates independent of used training set and preprocessing type.

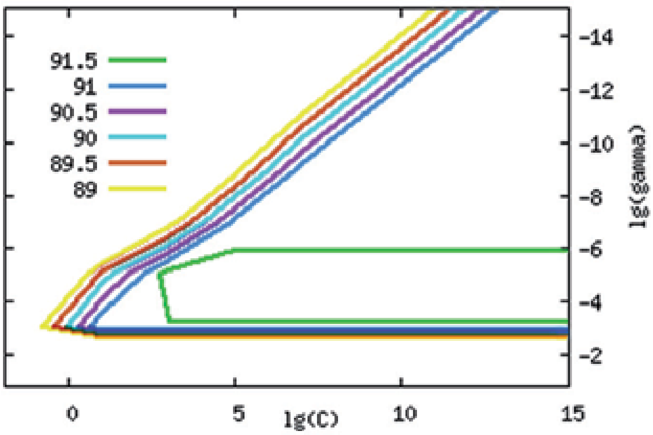


Fig. 7. Example of the output graph – dependence of cross validation rate on the parameters *C* and *gamma* for training set with 3 images per subject

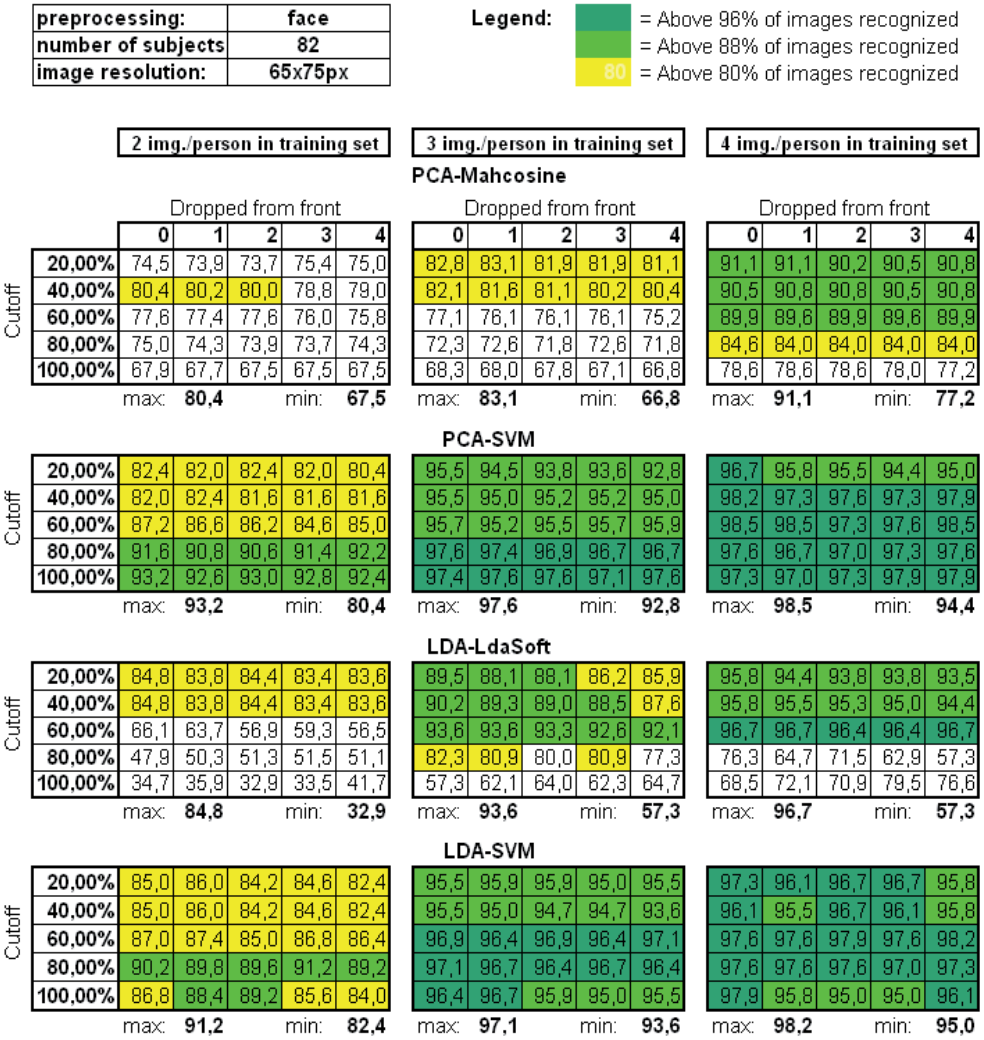
training set	*C*	γ	cross-valid.	rec. rate
face, 2img/pers.	0,03125	0,0078125	51,22%	80,04%
face, 3img/pers.	128	3,05176E-05	78,86%	93,79%
face, 4img/pers.	128	3,05176E-05	86,59%	96,74%
BIGface, 2img/pers.	0,03125	0,0078125	64,63%	74,45 %
BIGface, 3img/pers.	8	0,00012207	83,33%	93,56%
BIGface, 4img/pers.	128	3,05176E-05	89,63%	96,74%

Table 1. Recognition rate and optimal SVM parameter setups for used training sets

5.2 Two-Stage Recognition Systems

PCA and LDA algorithms are used to reduce the dimension and extract the features from face images. Using the training set, they produce a transform matrix. For face recognition purposes, we do not need the whole transform matrix and therefore we truncate first or last vectors from the transform matrix. The results of recognition are significantly influenced by parameters "Dropped from front" and "CutOff".

- **Dropped from front (DPF)** – denotes number of eigenvectors cut from the beginning of transform matrix (first vectors - vectors belonging to the largest eigenvalues). These vectors will not be used by image projection to PCA (or LDA) feature space. Reason to truncate these vectors is based on the assumption that these vectors do not correspond to useful information such as lighting variations (Beveridge et. al., 2003). Our tests were performed for "Dropped from front" values 0, 1, 2, 3, and 4.
- **CutOff (CO)** – represents how many vectors remain in the transform matrix. Reason to truncate last basis vectors (vectors corresponding to the smallest eigenvalues) is to lower the computation requirements and to eliminate unnecessary information that correlates with noise – and as such is meaningless for recognizing faces (Beveridge et. al., 2003). Our tests were performed for CutOff parameter set to 20%, 40%, 60%, 80% and 100%.



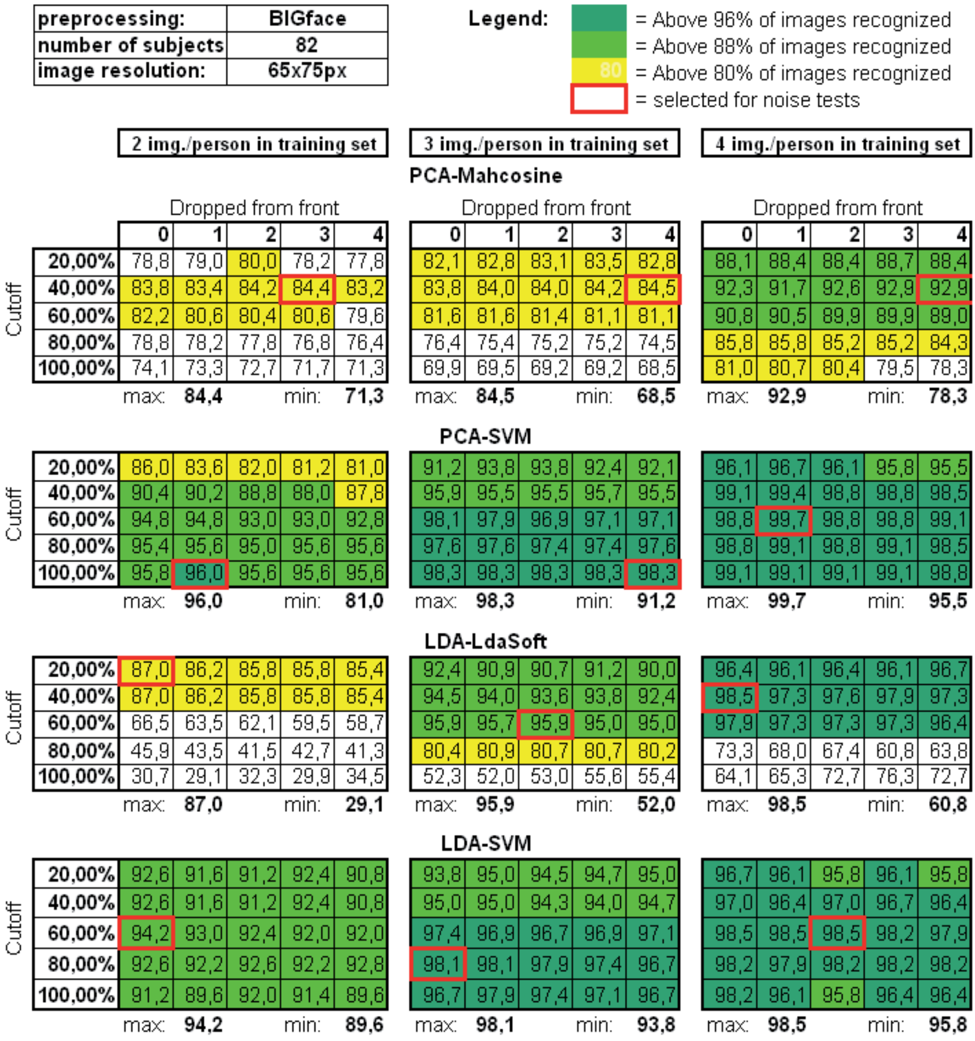
preprocessing:	face
number of subjects	82
image resolution:	65x75px

Legend:
🟩 = Above 96% of images recognized
🟩 = Above 88% of images recognized
80 = Above 80% of images recognized

	2 img./person in training set					3 img./person in training set					4 img./person in training set				
PCA-Mahcosine															
	Dropped from front					Dropped from front					Dropped from front				
Cutoff	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
20,00%	74,5	73,9	73,7	75,4	75,0	82,8	83,1	81,9	81,9	81,1	91,1	91,1	90,2	90,5	90,8
40,00%	80,4	80,2	80,0	78,8	79,0	82,1	81,6	81,1	80,2	80,4	90,5	90,8	90,8	90,5	90,8
60,00%	77,6	77,4	77,6	76,0	75,8	77,1	76,1	76,1	76,1	75,2	89,9	89,6	89,9	89,6	89,9
80,00%	75,0	74,3	73,9	73,7	74,3	72,3	72,6	71,8	72,6	71,8	84,6	84,0	84,0	84,0	84,0
100,00%	67,9	67,7	67,5	67,5	67,5	68,3	68,0	67,8	67,1	66,8	78,6	78,6	78,6	78,0	77,2
	max:	80,4		min:	67,5	max:	83,1		min:	66,8	max:	91,1		min:	77,2
PCA-SVM															
Cutoff															
20,00%	82,4	82,0	82,4	82,0	80,4	95,5	94,5	93,8	93,6	92,8	96,7	95,8	95,5	94,4	95,0
40,00%	82,0	82,4	81,6	81,6	81,6	95,5	95,0	95,2	95,2	95,0	98,2	97,3	97,6	97,3	97,9
60,00%	87,2	86,6	86,2	84,6	85,0	95,7	95,2	95,5	95,7	95,9	98,5	98,5	97,3	97,6	98,5
80,00%	91,6	90,8	90,6	91,4	92,2	97,6	97,4	96,9	96,7	96,7	97,6	96,7	97,0	97,3	97,6
100,00%	93,2	92,6	93,0	92,8	92,4	97,4	97,6	97,6	97,1	97,6	97,3	97,0	97,3	97,9	97,9
	max:	93,2		min:	80,4	max:	97,6		min:	92,8	max:	98,5		min:	94,4
LDA-LdaSoft															
Cutoff															
20,00%	84,8	83,8	84,4	83,4	83,6	89,5	88,1	88,1	86,2	85,9	95,8	94,4	93,8	93,8	93,5
40,00%	84,8	83,8	84,4	83,4	83,6	90,2	89,3	89,0	88,5	87,6	95,8	95,5	95,3	95,0	94,4
60,00%	66,1	63,7	56,9	59,3	56,5	93,6	93,6	93,3	92,6	92,1	96,7	96,7	96,4	96,4	96,7
80,00%	47,9	50,3	51,3	51,5	51,1	82,3	80,9	80,0	80,9	77,3	76,3	64,7	71,5	62,9	57,3
100,00%	34,7	35,9	32,9	33,5	41,7	57,3	62,1	64,0	62,3	64,7	68,5	72,1	70,9	79,5	76,6
	max:	84,8		min:	32,9	max:	93,6		min:	57,3	max:	96,7		min:	57,3
LDA-SVM															
Cutoff															
20,00%	85,0	86,0	84,2	84,6	82,4	95,5	95,9	95,9	95,0	95,5	97,3	96,1	96,7	96,7	95,8
40,00%	85,0	86,0	84,2	84,6	82,4	95,5	95,0	94,7	94,7	93,6	96,1	95,5	96,7	96,1	95,8
60,00%	87,0	87,4	85,0	86,8	86,4	96,9	96,4	96,9	96,4	97,1	97,6	97,6	97,9	97,6	98,2
80,00%	90,2	89,8	89,6	91,2	89,2	97,1	96,7	96,4	96,7	96,4	97,6	97,6	97,6	97,0	97,3
100,00%	86,8	88,4	89,2	85,6	84,0	96,4	96,7	95,9	95,0	95,5	97,9	95,8	95,0	95,0	96,1
	max:	91,2		min:	82,4	max:	97,1		min:	93,6	max:	98,2		min:	95,0

Table 2. Results of experiments for methods PCA+MahCosine, PCA+SVM, LDA+LDASoft, LDA+SVM with "face" preprocessing (total 300 tests)

Methods utilizing PCA or LDA (Fig. 6b - Fig. 6e) were tested using three training sets with 2, 3 and 4 images per subject. For each method, we tested 25 different parameters DPF and CO setups on three different training sets, what gives total 75 tests per each method and per each type of preprocessing (600 tests in total). Results of these tests are shown in Table 2 and Table 3. The maximal recognition rates are summarized in Fig. 8 and Fig. 9.



preprocessing:	BIGface
number of subjects	82
image resolution:	65x75px

Legend:
🟩 = Above 96% of images recognized
🟩 = Above 88% of images recognized
🟨 80 = Above 80% of images recognized
🟥 = selected for noise tests

	2 img./person in training set					3 img./person in training set					4 img./person in training set				
	Dropped from front					Dropped from front					Dropped from front				
Cutoff	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
PCA-Mahcosine															
20,00%	78,8	79,0	80,0	78,2	77,8	82,1	82,8	83,1	83,5	82,8	88,1	88,4	88,4	88,7	88,4
40,00%	83,8	83,4	84,2	84,4	83,2	83,8	84,0	84,0	84,2	84,5	92,3	91,7	92,6	92,9	92,9
60,00%	82,2	80,6	80,4	80,6	79,6	81,6	81,6	81,4	81,1	81,1	90,8	90,5	89,9	89,9	89,0
80,00%	78,8	78,2	77,8	76,8	76,4	76,4	75,4	75,2	75,2	74,5	85,8	85,8	85,2	85,2	84,3
100,00%	74,1	73,3	72,7	71,7	71,3	69,9	69,5	69,2	69,2	68,5	81,0	80,7	80,4	79,5	78,3
	max:	84,4		min:	71,3	max:	84,5		min:	68,5	max:	92,9		min:	78,3
PCA-SVM															
20,00%	86,0	83,6	82,0	81,2	81,0	91,2	93,8	93,8	92,4	92,1	96,1	96,7	96,1	95,8	95,5
40,00%	90,4	90,2	88,8	88,0	87,8	95,9	95,5	95,5	95,7	95,5	99,1	99,4	98,8	98,8	98,5
60,00%	94,8	94,8	93,0	93,0	92,8	98,1	97,9	96,9	97,1	97,1	98,8	99,7	98,8	98,8	99,1
80,00%	95,4	95,6	95,0	95,6	95,6	97,6	97,6	97,4	97,4	97,6	98,8	99,1	98,8	99,1	98,5
100,00%	95,8	96,0	95,6	95,6	95,6	98,3	98,3	98,3	98,3	98,3	99,1	99,1	99,1	99,1	98,8
	max:	96,0		min:	81,0	max:	98,3		min:	91,2	max:	99,7		min:	95,5
LDA-LdaSoft															
20,00%	87,0	86,2	85,8	85,8	85,4	92,4	90,9	90,7	91,2	90,0	96,4	96,1	96,4	96,1	96,7
40,00%	87,0	86,2	85,8	85,8	85,4	94,5	94,0	93,6	93,8	92,4	98,5	97,3	97,6	97,9	97,3
60,00%	66,5	63,5	62,1	59,5	58,7	95,9	95,7	95,9	95,0	95,0	97,9	97,3	97,3	97,3	96,4
80,00%	45,9	43,5	41,5	42,7	41,3	80,4	80,9	80,7	80,7	80,2	73,3	68,0	67,4	60,8	63,8
100,00%	30,7	29,1	32,3	29,9	34,5	52,3	52,0	53,0	55,6	55,4	64,1	65,3	72,7	76,3	72,7
	max:	87,0		min:	29,1	max:	95,9		min:	52,0	max:	98,5		min:	60,8
LDA-SVM															
20,00%	92,6	91,6	91,2	92,4	90,8	93,8	95,0	94,5	94,7	95,0	96,7	96,1	95,8	96,1	95,8
40,00%	92,6	91,6	91,2	92,4	90,8	95,0	95,0	94,3	94,0	94,7	97,0	96,4	97,0	96,7	96,4
60,00%	94,2	93,0	92,4	92,0	92,0	97,4	96,9	96,7	96,9	97,1	98,5	98,5	98,5	98,2	97,9
80,00%	92,6	92,2	92,6	92,2	92,8	98,1	98,1	97,9	97,4	96,7	98,2	97,9	98,2	98,2	98,2
100,00%	91,2	89,6	92,0	91,4	89,6	96,7	97,9	97,4	97,1	96,7	98,2	96,1	95,8	96,4	96,4
	max:	94,2		min:	89,6	max:	98,1		min:	93,8	max:	98,5		min:	95,8

Table 3. Results of experiments for methods PCA+MahCosine, PCA+SVM, LDA+LDASoft, LDA+SVM with "BIGface" preprocessing (total 300 tests)

5.3 Evaluation of Simulation Results

Based on presented experiments, we can formulate several conclusions:

1. More images in the training stage cause better performance of all methods.
2. LDA+LDASoft performs better than PCA+MahCosine, but PCA+SVM is slightly better than LDA+SVM.
3. The best performing setups of parameters CO and DPF differ using different preprocessing and number of images per subject in training set. Generally PCA+MahCosine and LDA+LDASoft perform better for truncating 0-4 first vectors and leaving 20%-60% of the vectors in transform matrix.

4. The recognition rate is most significantly affected by setting of the CO parameter – for PCA+MahCosine and LDA+LDASoft it is better to truncate vectors from the end of the transform matrix leaving only 20% - 60% of the vectors. Methods PCA+SVM and LDA+SVM perform better when leaving more (60% - 100%) vectors of the transform matrix.
5. Results of LDA+LDASoft are more influenced by setting the CO parameter compared to PCA+MahCosine – especially with only 2 images per subject in the training set, where the worst recognition rate is around 30% (see Table 2 and Table 3).
6. Using SVM for classification (methods PCA+SVM and LDA+SVM) makes the recognition rates more stable and less influenced by setting the CO and DPF parameters (see Table 2 and Table 3) and these methods perform better compared to simple PCA+MahCosine and LDA+LDASoft – see Fig. 8 and Fig. 9.

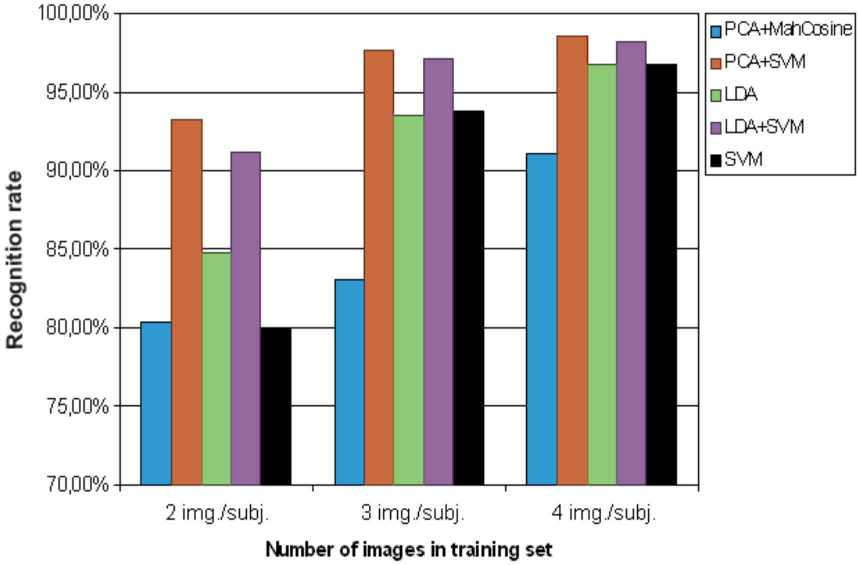


Fig. 8. Graph of maximum recognition rates for methods PCA+MahCosine, PCA+SVM, LDA+LDASoft, LDA+SVM, SVM (left to right) with "face" preprocessing

6. Face Recognition Experiments and Results in Noisy Conditions

In this part of the chapter, we concentrate on the influence of input image quality to face recognition accuracy. Noise and distortions in face images can seriously affect the performance of face recognition systems. Analog or digital capturing the image, image transmission, image copying or scanning can suffer from noise. This is why we study behaviour of discussed methods in the presence of noise.

We include Gaussian noise, salt & pepper noise and speckle noise. Huge effort in removing these types of noise from static or dynamic images in the area of face recognition is documented in the literature, e.g. (Uglov et al., 2008; Reda, & Aoued, 2004; Wheeler et al., 2007). We use these types of noise with various intensities (various parameters).

6.1 Types of Noises

Each image capturing generates digital or analog noise of diverse intensity. The noise is also generated while transmitting and copying analog images. Noise generation is a natural property for image scanning systems. Diverse types of noises exist. Herein we use three different types: Gaussian (Truax, 1999), salt & pepper (Chan et al., 2005), and speckle (Anderson & Trahey, 2006) noises.

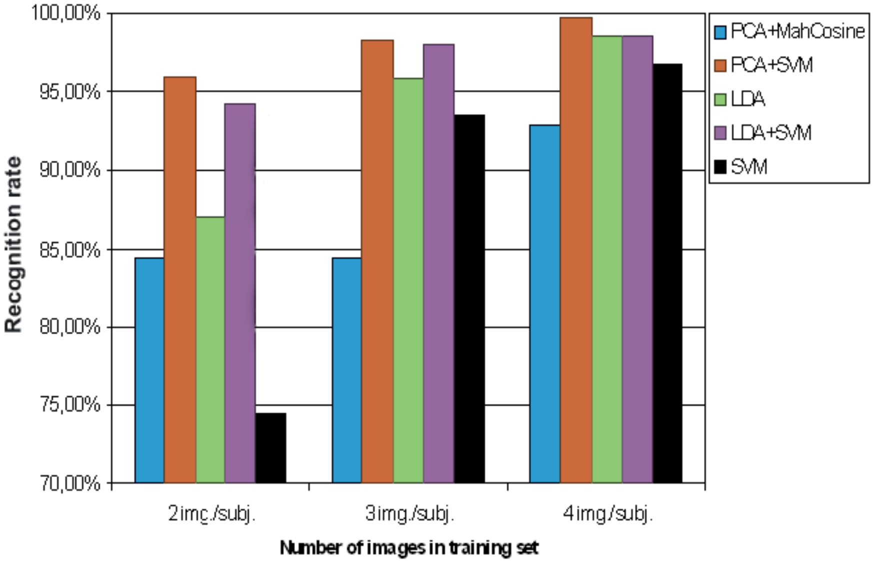


Fig. 9. Graph of maximum recognition rates for methods PCA+MahCosine, PCA+SVM, LDA+LDASoft, LDA+SVM, SVM (left to right) with "BIGface" preprocessing

Gaussian Noise

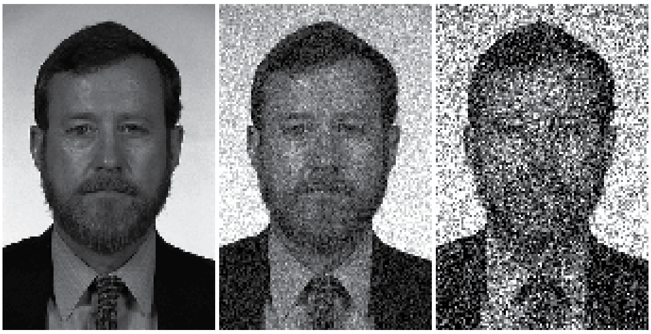
Gaussian noise is the most common noise occurring in everyday life. The Gaussian noise can be detected in free radio waves or in television receivers. Gaussian noise is produced in analog images that are stored for a long time.

We studied face recognition with different Gaussian noise intensity. Gaussian noise was generated with Gaussian normal distribution function which can be written as:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad (20)$$

where μ is the mean value of the required distribution and σ^2 is a variance (Truax, 1999; Chiodo, 2006).

Noise parameters settings for our simulations were determined empirically. The mean of Gaussian distribution was set to 0 and we changed the variance. Examples of images corrupted by Gaussian noise can be seen in Fig. 10. The label g0.01 means that the Gaussian noise of variance 0.01 was applied on the image. The same notation is used also in presented graphs.



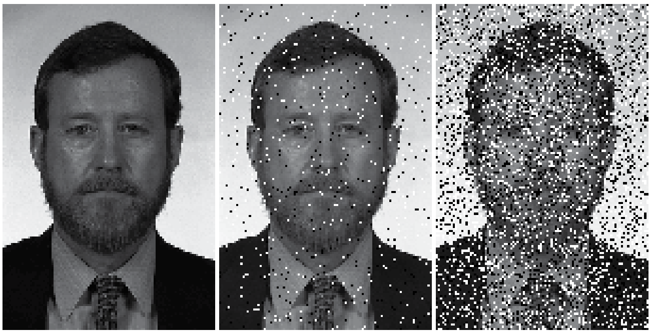
Original g 0.01 g 0.09

Fig. 10. Examples of images corrupted by Gaussian noise

Salt & Pepper Noise

Salt & pepper noise is perceived as a random occurrence of black and white pixels in a digital image. It can be caused by incorrect data transmission or by a damage of already received data. In CCD and CMOS sensors or LCD displays, the salt & pepper noise can be caused by permanently turned-on or turned-off pixels. Remaining pixels are unchanged. Usually, the intensity (frequency of the occurrence) of this noise is quantified as a percentage of incorrect pixels (Fisher et al., 2003). The median filtering (as a specific case of order-statistic filtering) is used as an effective method for elimination of salt & pepper noise from digital images (Chan et al., 2005).

Noise parameter settings for our simulations vary from 4% of noise intensity (0.04) up to the 30% of damaged pixels. The label sp0.04 means, that the salt & pepper noise of intensity 4% was applied on the image. Examples of images corrupted by salt & pepper noise are shown in Fig. 11.

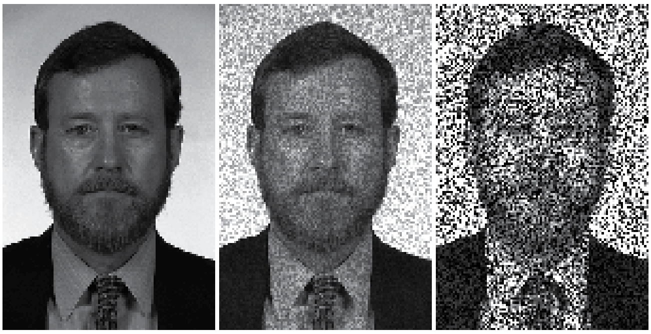


Original sp 0.04 sp 0.3

Fig. 11. Examples of images corrupted by 4% and 30% salt & pepper noise

Speckle Noise

This granular noise occurs in ultrasound, radar and X-ray images and images obtained from the magnetic resonance (Chaillan et al., 2007). The multiplicative signal dependent noise is generated by constructive and destructive interference of detected signals. The wave interference is a reason of multiplicative noise occurrence in the scanned image. The speckle noise is image dependent. Therefore it is very hard (if possible) to find a mathematical model that describes the removal of this noise, especially if we expect the randomness of the input data (Fisher et al., 2003).



Original s 0.03 s 0.7

Fig. 12. Examples of images corrupted by speckle noise

The values which determined intensity of noise in our tests were set empirically. The noise was applied according to the following equation

$$S = I + n * I \tag{21}$$

where I is the original human face image and n is the uniform distribution of the noise with zero mean value and variance σ^2 . For our simulations, variance varied from 0.03 to 0.7. The label s0.03 means that the speckle noise of variance 0.03 was applied on the image. Presence of speckle noise in the face image is illustrated in Fig. 12.

For simulation of methods in presence of noise, we use the best parameter settings we obtained running 600 tests in Section 5, i.e. when the methods worked in ideal conditions.
In order to mimic real-world conditions, we use images not distorted by noise for training purposes whilst noisy images are used for testing. Such scenario simulates real-world face recognition conditions.
We concentrate on "BIGface" preprocessed images only, since this preprocessing gives better results compared to "face" preprocessing (this can be seen when comparing Tables 2 and 3). Parameters for settings of the algorithms (CO and DPF) were empirically obtained from Table 3. We selected and used only those parameters for which the recognition experiments were most successful (they are marked by red in Table 3). This was necessary in order to reduce the number of experiments. Using all possible settings from simulations in ideal conditions and combining them with three types of noises with all selected parameters would lead to total 13500 results. Selecting best parameters only lead us to total 540 results. Obtained results are shown in Fig. 13 – 21 along with brief comments.

6.2 Simulation Results for Face Images Corrupted by Gaussian Noise

Simulation results for face images corrupted by Gaussian noise are summarized in Fig. 13 – 15. PCA-MahCosine method is most influenced by increasing the intensity of Gaussian noise. Results for training sets with 2 and 3 img./subj. look alike – recognition rates decrease with higher noise. The effect of the noise for training set containing 4 img./subj. is not so noticeable. Worst results are achieved by PCA-MahCosine method. For training set with 4 img./subj., the results of other 3 methods are almost equal and the recognition rates are surprisingly high even for higher noise intensities and they do not decrease. For 3 img./subj., the best results come from LDA-SVM method, followed by LDA-LDASoft (from intensity of noise >0.01). For training set containing 2 img./subj. only, both SVM methods result in best recognition rates and LDA-SVM is slightly better than PCA-SVM. It is also interesting to notice that there are some cases, when consecutive increase of noise levels resulted in better recognition rates.

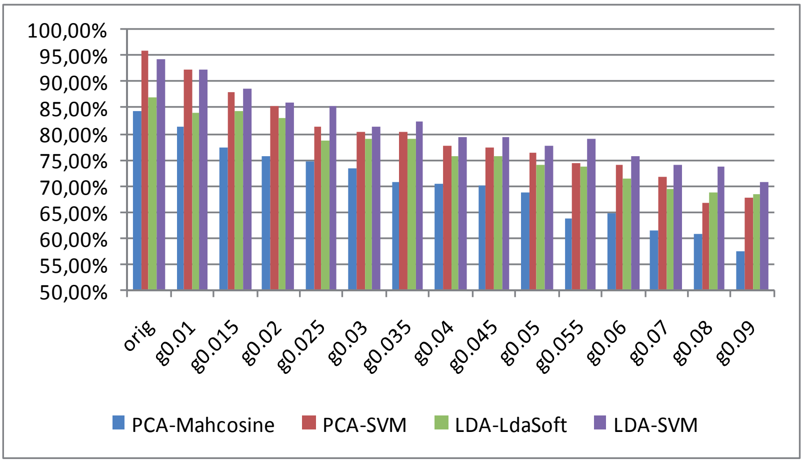


Fig. 13. Recognition rates of examined methods, Gaussian noise, training set 2 img./subj.

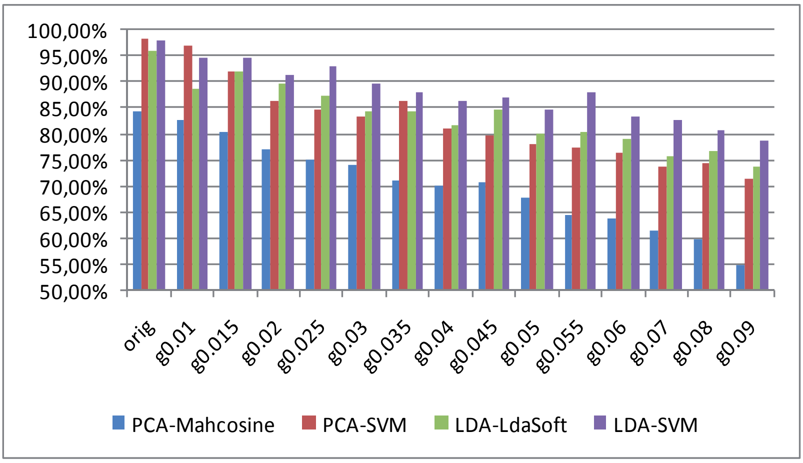


Fig. 14. Recognition rates of examined methods, Gaussian noise, training set 3 img./subj.

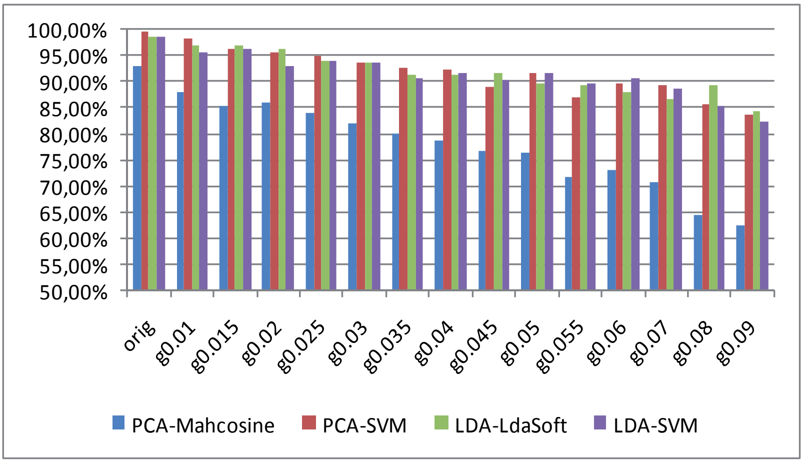


Fig. 15. Recognition rates of examined methods, Gaussian noise, training set 4 img./subj.

6.3 Simulation Results for Face Images Corrupted by Salt & Pepper Noise

Fig. 16 – 19 show results for face images corrupted by salt & pepper noise. Increasing the noise level does not have significant effect till intensity 0.2. Decrease of the recognition rate while increasing the noise intensity is most noticeable for results with 2 img./subj. in the training set. PCA-MahCosine is again the worst method. Best recognition rates are achieved by the methods that use SVM and they both achieved almost equal results. For 3 img./subj., LDA-SVM was slightly better than PCA-SVM. One can again notice, that in some cases consecutive increase of noise levels resulted in better recognition rates.

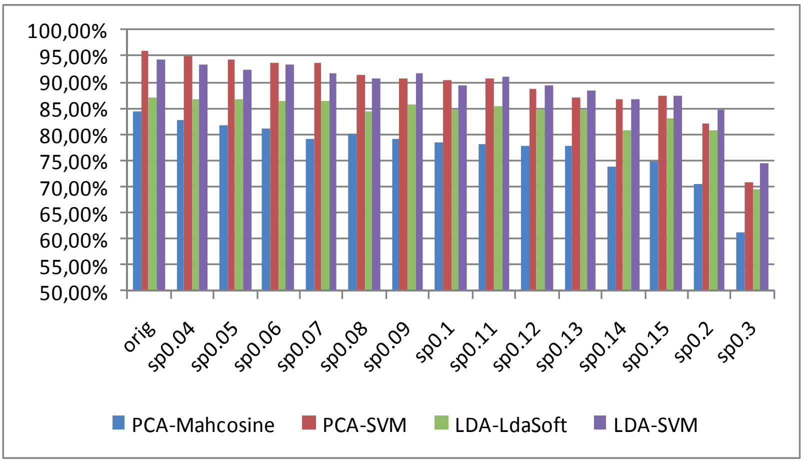


Fig. 16. Recognition rates of examined methods, salt &pepper noise, training set 2 img./subj.

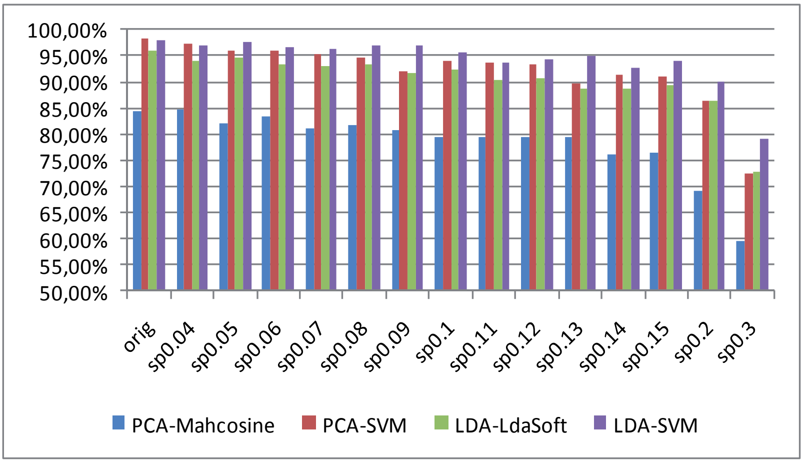


Fig. 17. Recognition rates of examined methods, salt &pepper noise, training set 3 img./subj.

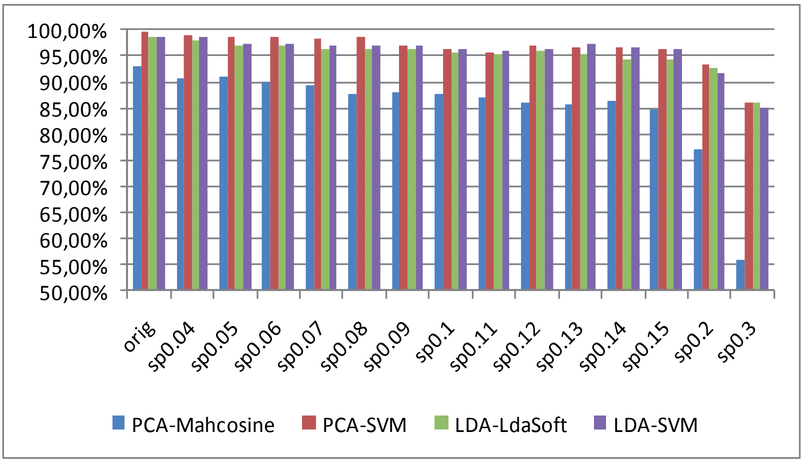


Fig. 18. Recognition rates of examined methods, salt &pepper noise, training set 4 img./subj.

6.4 Simulation Results for Face Images Corrupted by Speckle Noise

Fig. 19 – 21 contains simulation results for face images corrupted by speckle noise. PCA-MahCosine method achieves worst results. Best results can be achieved by LDA-SVM; this is more noticeable for higher noise intensities. For 4 img./subj., the PCA+SVM, LDA+LDASoft and LDA+SVM methods have almost equal recognition rates. For 3img./subj., the LDA+LDASoft method is better than PCA+SVM, for 2 img./subj., the PCA+SVM is better than LDA+LDASoft. For speckle noise, there are not cases when higher noise levels result in better recognition rates. There was an exception for speckle noise of intensity 0.03 for training set 3 img./subj., because recognition by PCA-MahCosine method gives better rate for corrupted images (84.73%) than recognition using the original images (84.5%).

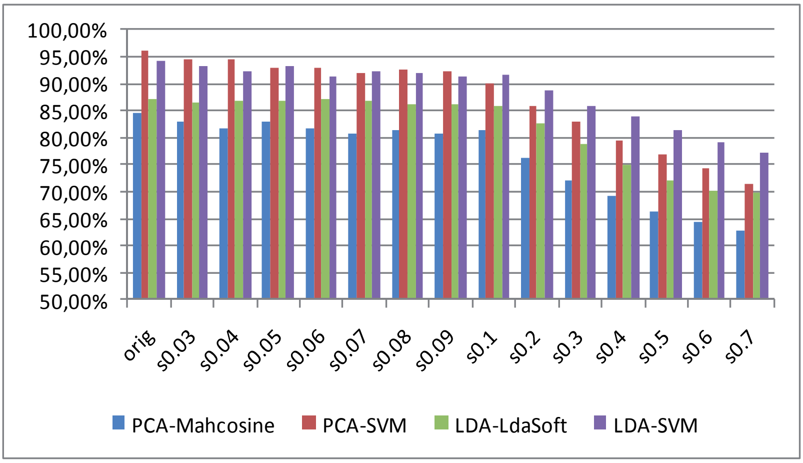


Fig. 19. Recognition rates of examined methods, speckle noise, training set 2 img./subj.

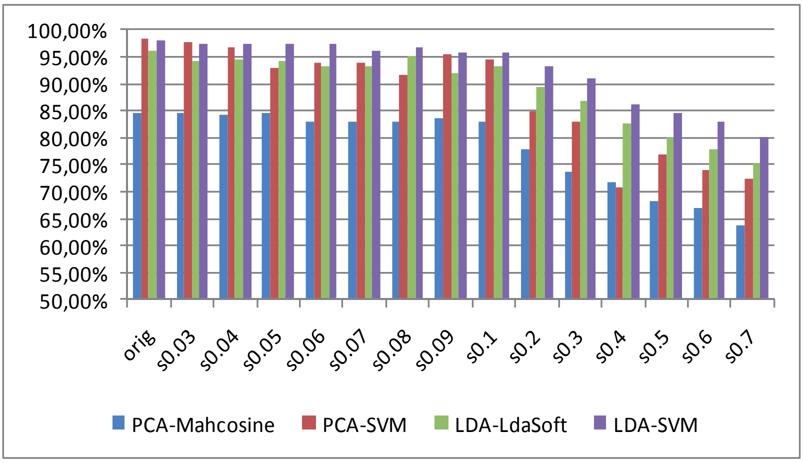


Fig. 20. Recognition rates of examined methods, speckle noise, training set 3 img./subj.

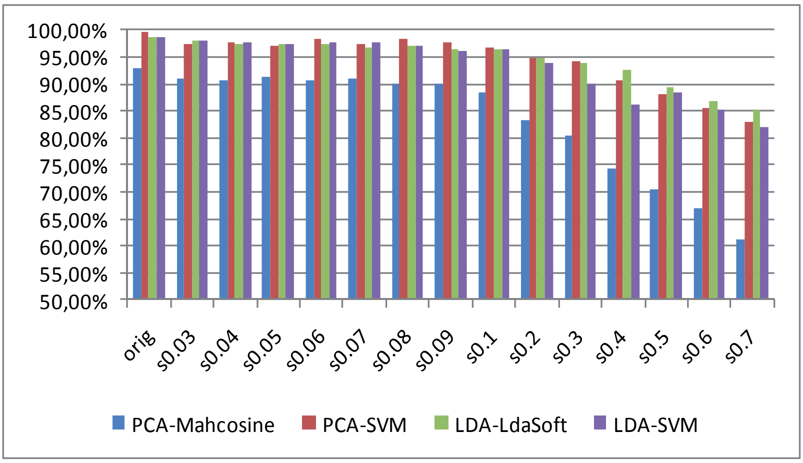
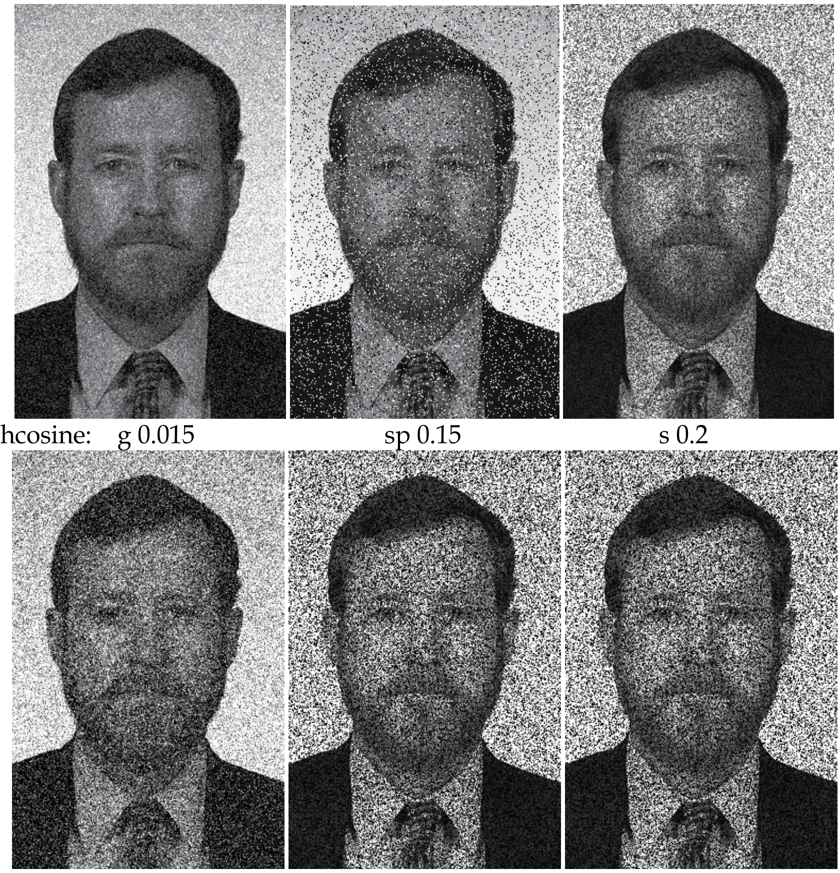


Fig. 21. Recognition rates of examined methods, speckle noise, training set 4 img./subj.

6.5 Equivalence of Different Types of Noise from the Recognition Point of View

After presenting recognition results for different types of noise an interesting question arises: What is the relationship among different noise types? The concrete values of noise parameters do not give the answer – a comparison cannot be based on non-related parameters.



PCA-Mahcosine: g 0.015 sp 0.15 s 0.2

LDA-SVM: g 0.08 sp 0.3 s 0.6

Fig. 22. Example of the subject, for who all the studied methods (here shown PCA-MahCosine and LDA-SVM) result in recognition accuracy about 85 % (see Table 4 for exact noise type and intensity)

One possible solution can be based exactly on results of machine face recognition. This approach is illustrated in Fig. 22 and in corresponding Table 4. Fig. 22 shows images of the subject corrupted by different types of noises. The noise parameters are chosen in such manner that all studied methods (PCA-MahCosine, PCA-SVM, LDA-LDASoft, LDA-SVM) result in recognition accuracy near 85 %. Table 4 specifies each noise type and its corresponding parameter. PCA-MahCosine and LDA-SVM methods are included in Fig. 22, since PCA-SVM and LDA-LDASoft methods are visually similar to LDA-SVM. Fig. 22 thus shows equivalence of different types of noise from the face recognition point of view of

PCA-MahCosine and LDA-SVM methods. But this is equivalency of noise types from machine point of view. It should be even more interesting to compare recognition ability of machine learning methods and humans.

method	Gaussian noise	Recognition rate in %	Salt&pepper noise	Recognition rate in %	Speckle noise	Recognition rate in %
PCA-Mahcosine*	g0.015	85,16%	sp0.15	84,87%	s0.2	83,38%
PCA-SVM	g0.08	85,76%	sp0.3	86,05%	s0.6	85,46%
LDA-LdaSoft	g0.09	84,27%	sp0.3	86,05%	s0.7	85,16%
LDA-SVM*	g0.08	85,16%	sp0.3	85,16%	s0.6	85,16%

Table 4. Types and intensity of noise resulting in recognition rate about 85 % (for training set 4img./subj.).
* included in Fig. 22

7. Conclusion

We examined different scenarios of face recognition experiments. They contain both single-stage and two-stage recognition systems. Single-stage face recognition uses SVM for classification directly. Two-stage recognition systems include PCA with MahCosine metric, LDA with LDASoft metric and also methods utilizing both PCA and LDA for feature extraction followed by SVM for classification. All methods are significantly influenced by different settings of parameters that are related to the algorithm used (i.e. PCA, LDA or SVM). This is the reason we presented serious analysis and proposal of parameter settings for the best performance of discussed methods.

For methods working in ideal conditions, the conclusions are as follows: When comparing non-SVM based methods, higher maximum recognition rate is generally achieved by method LDA+LDASoft compared to PCA+MahCosine; on the other hand LDA+LDASoft is more sensitive to method settings. Using SVM in classification stage (PCA+SVM and LDA+SVM) produced better maximum recognition rate than standard PCA and LDA methods.

Experiments with single-stage SVM show that this method is very efficient for face recognition even without previous feature extraction. With 4 images per subject in training set, we reached 96.7% recognition rate.

The experiments were made with complex image set selected from FERET database containing 665 images. Such number of face images entitles us to speak about general behavior of presented methods. Altogether more than 600 tests were made and maximum recognition rates near 100% were achieved.

It is important to mention that the experiments were made with "closed" image set, so we did not have to deal with issues like detecting people who are not in the training set. On the other hand, we worked with real-world face images; our database contains images of the same subjects that often differ in face expressions (smiling, bored, …), with different hairstyles, with or without beard, or wearing glasses and that were taken in different session after longer time period (i.e. we did not work with identity card-like images).

We also presented recognition results for noisy images and graphically compared them to results for non-distorted images. In this way, the insight on face recognition system robustness is obtained.

Independently on noise type or its parameter, the PCA-MahCosine method gives the lowest success in face recognition compared to all tested methods. Using other methods, the results were significantly better. Methods that use SVM classifier achieve globally better results for each training set. On the other hand, SVM-based methods need a lot of time to search for optimal parameters, while PCA-MahCosine method is the fastest.

By our work, we continue in our effort to offer complex view to biometric face recognition. In (Oravec et al., 2008) besides detection of faces and facial features, we presented feature extraction methods from face images (linear and nonlinear methods, second-order and higher-order methods, neural networks and kernel methods) and relevant types of classifiers. Face recognition in ideal conditions using FERET database is contained partly in (Oravec et al., 2009) and in this chapter.

Our work on presented methods now further continues in evaluating their sensitivity and behavior in non-ideal conditions. First our contribution to this area which includes presence of noise is covered in this chapter. Our future work will comprise partially occluded faces and also faces extracted from static images and/or video streams transmitted with errors or loss of data, where some parts of face image are missing (block or blocks of pixels) or an error-concealment mechanism is applied prior to recognition (Pavlovičová et al., 2006; Polec et al., 2009; Marchevský & Mochnáč, 2008).

Our future work will also be focused on a psychological experiment trying to find relationship for mentioned types of distortions from the point of view of recognition ability of humans and machines (as an extension of the aspect of noise for machine recognition that is outlined in section 6.5).

8. Acknowledgements

Research described in this paper was done within the grants 1/0718/09 and 1/0883/08 of the Slovak Grant Agency VEGA.

Portions of the research in this paper use the FERET database of facial images collected under the FERET program, sponsored by the DOD Counterdrug Technology Development Program Office.

9. References

Anderson, M. E. & Trahey, G. E. (2006). A beginner's guide to speckle. *A seminar on k-space applied to medical ultrasound* [online], Available on: http://dukemil.bme.duke.edu/Ultrasound/k-space/node5.html, Dept. of Biomedical Engineering, Duke University

Belhumeur, P. N.; Hespanha, J. P. & Kriegman, D. J. (1997). Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection, *IEEE Trans. Pattern Analysis and Machine Intelligence*, pp. 711-720, Vol. 19, No. 7, July 1997

Beveridge, R.; Bolme, D.; Teixeira, M. & Draper, B. (2003). The CSU Face Identification Evaluation System User's Guide, Version 5.0, Technical Report., Colorado State University, May 2003, http://www.cs.colostate.edu/evalfacerec/algorithms/version5/faceIdUsersGuide.pdf

Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*, Oxford University Press, Inc., ISBN 0 19 853864 2, New York

Chaillan, F.; Fraschini, C. & Courmontagne, P. (2007). Speckle Noise reduction in SAS imagery, *Signal Processing*, Vol. 87, No. 4., page numbers 762-781, April 2007, ISSN 0165-1684

Chan, R. H., Ho, C. W., Nikolova, M. (2005). Salt-and-pepper noise removal by median-type noise detector and edge-preserving regularization, *IEEE Trans. Image Processing*, Vol. 14, page numbers 1479-1485

Chiodo, K. (2006). Normal Distribution, *NIST/SEMATECH e-Handbook of Statistical Methods*, http://www.itl.nist.gov/div898/handbook/eda/section3/eda3661.htm

Delac, K.; Grgic, M. & Bartlett, M. S. (2008). Recent Advances in Face Recognition, *IN-TECH*, 2008, http://intechweb.org/book.php?id=101

Duda, R.O. & Hart, P.E. (1973). *Pattern Classification and Scene Analysis*, John Wiley & Sons, Inc., ISBN-10: 0471223611, New York

EU – Passport Specification (2006), Biometrics Deployment of EU-Passports, Working document (EN) – 28/06/2006, http://ec.europa.eu/justice_home/doc_centre/freetravel/ documents/doc/c_2006_2909_prov_en.pdf

FERET Database (2001). http://www.itl.nist.gov/iad/humanid/feret/, NIST

Fisher, R., Perkins, S., Walker, A. & Wolfart, E. (2003). Noise Generation, *Hypermedia Image Processing Reference*, http://homepages.inf.ed.ac.uk/rbf/HIPR2/noise.htm

Haykin, S. (1994). *Neural Networks - A Comprehensive Foundation*, Macmillan College Publishing Company, ISBN-10: 0139083855, New York

Hofmann, T.; Schölkopf, B. & Smola, A. J. (2008). Kernel Methods in Machine Learning, *Annals of Statistics*, Vol. 36, No. 3, 2008, page numbers 1171-1220

Hotta, K. (2008). Robust Face Recognition under Partial Occlusion Based on Support Vector Machine with Local Gaussian Summation Kernel, *Image and Vision Computing*, Vol. 26, Issue 11, November 2008, page numbers 1490-1498

Hsu, C. W.; Chang, C.C. & Lin, C.J. (2008). A Practical Guide to Support Vector Classification, *Technical report*, Department of Computer Science, National Taiwan University. July, 2003, last updated October 2, 2008, http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf

Jain, A. K.; Flynn, P. & Ross, A. A. (2008), (Eds). *Handbook of Biometrics*, Springer, ISBN: 978-0-387-71040-2

Jain, A. K.; Ross, A. & Prabhakar, S. (2004). An Introduction to Biometric Recognition. *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 14, No. 1, January 2004, page numbers 4-20

Li, S. Z. & Jain, A. K. (2005) (Eds.). *Handbook of Face Recognition*, Springer, ISBN# 0-387-40595-x New York

LIBSVM - *A Library for Support Vector Machine*, http://www.csie.ntu.edu.tw/~cjlin/libsvm/

Marchevský, S. & Mochnáč, J. (2008). Hybrid Concealment Mechanism, *Acta Electrotechnica et Informatica*, Vol. 8, No. 1, 2008, page numbers 11-15, ISSN 1335-8243

Marcialis, G. L. & Roli, F. (2002). Fusion of LDA and PCA for Face Recognition, *Proc. of the Workshop on Machine Vision and Perception, 8th Meeting of the Italian Association of Artificial Intelligence (AI*IA)*, Siena, Italy, September 2002

Martinez, A. M. & Kak, A. C. (2001). PCA versus LDA. *IEEE Trans. Pattern Analysis and Machine Intelligence,* Vol. 23, No. 2, February 2001, page numbers 228-233

Mazanec, J.; Melišek, M.; Oravec, M. & Pavlovičová, J. (2009). Face Recognition Using Support Vector Machines, PCA and LDA with the FERET Database, Acta Polytechnica Hungarica, submitted for publication in 2009

Muller, K.; Mika, S.; Ratsch, G.; Tsuda, K. & Scholkopf, B. (2001). An Introduction to Kernel-Based Learning Algorithms. *IEEE Transactions on Neural Networks*, Vol. 12, No. 2, March 2001, page numbers 181—201

Oravec, M. & Pavlovičová, J. (2004). Linear Discriminant Analysis for Face Images, *Proc. of the 27th Int. Conf. Telecommunications and Signal Processing TSP-2004*, pp. 146-149, Brno, Czech Republic, September 2004

Oravec, M.; Beszédeš, M. & Rozinaj, G.(2008). Detection and Recognition of Human Faces and Facial Features, book chapter In: *Speech, Audio, Image and Biomedical Signal Processing Using Neural Networks,* Editors: Bhanu Prasad and S. R. Mahadeva Prasanna, page numbers 283-306, Springer-Verlag, Germany, ISBN 978-3-540-75397-1

Oravec, M.; Mazanec, J.; Melišek, M. & Pavlovičová, J. (2009). Face Recognition Using Support Vector Machines, PCA and LDA with the FERET Database, Acta Polytechnica Hungarica. Submitted for publication in 2009.

Pavlovičová, J.; Oravec, M.; Polec, J.; Keleši, M. & Mokoš, M. (2006). Error Concealment using Neural Networks for Block-based Image Coding. *Radioengineering*, Brno University of Technology, Dept. of Radio Electronics, Vol. 15, No. 2, June 2006, page numbers 30–36, ISSN 1210-2512

Phillips, P. J.; Moon, H.; Rizvi, S. A. & Rauss, P. J. (2000). The FERET Evaluation Methodology for Face Recognition Algorithms, *IEEE Trans. Pattern Analysis and Machine Intelligence,* Vol. 22, page numbers 1090-1104

Phillips, P. J.; Wechsler, H.; Huang, J. & Rauss, P. (1998). The FERET database and evaluation procedure for face recognition algorithms, *Image and Vision Computing*, Vol. 16, No. 5, page numbers 295-306

Polec, J.; Pohančeník, M.; Ondrušová, S.; Kotuliaková, K. & Karlubíková, T. (2009). Error Concealment for Classified Texture Images, *IEEE Region 8 EUROCON 2009*, pp.1348-1353, ISBN 978-1-4244-3861-7, St. Petersburg, Russia, May 2009, Piscataway : IEEE

Reda, A.; Aoued, B. (2004). Artificial Neural Network-based Face Recognition, *Proc. of First International Symposium on Control, Communications and Signal Processing*, pp. 439-442

Ross, A. A.; Nandakumar,K. & Jain, A. K. (2006). *Handbook of Multibiometrics*, Springer, ISBN: 978-0-387-22296-7

Shoniregun, C. A. & Crosier,S. (2008). *Securing Biometrics Applications*, Springer, ISBN: 978-0-387-69932-5

Truax, B. (1999). (Ed.) Gaussian noise In: *Handbook for Acoustic Ecology*, Available on: http://www.sfu.ca/sonic-studio/handbook/Gaussian_Noise.html Cambridge Street Publishing

Turk, M. & Pentland, A. (1991). Eigenfaces for Recognition, *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, 1991, page numbers 71-86.

Uglov, J.; Jakaite, L.; Schetinin, V. & Maple, C. (2008). Comparing Robustness of Pairwise and Multiclass Neural-Network Systems for Face Recognition, *EURASIP Journal on Advances in Signal Processing*, Vol. 2008, Article ID 468693, doi:10.1155/2008/468693 http://www.hindawi.com/journals/asp/2008/468693.html

Wang, J.; Lin, Y.; Yang, W. & Yang, J. (2008). Kernel Maximum Scatter Difference Based Feature Extraction and its Application to Face Recognition, *Pattern Recognition Letters,* Vol. 29, Issue 13, October 2008, page numbers 1832-1835

Wang, Y.; Jiar, Y.; Hu, C. & Turk, M. (2004). Face Recognition Based on Kernel Radial Basis Function Networks, *Asian Conference on Computer Vision,* pp. 174-179, ISBN 10: 8995484209, Korea, January 2004, Asian Federation of Computer Vision Soc.

Wheeler, F. W.; Liu, X.; Tu, P. H. & Hoctor, R. T. (2007). Multi-Frame Image Restoration for Face Recognition, *Proc. of IEEE Workshop on Signal Processing Applications for Public Security and Forensics*, SAFE '07, pp. 1-6, April 2007

Yang, M. (2002). Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods, *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 215-220, Mountain View, California

Yang, M.; Frangi, A. F.; Yang, J. Y.; Zhang, D. & Jin, Z.(2005). KPCA Plus LDA: A Complete Kernel Fisher Discriminant Framework for Feature Extraction and Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 2, February 2005, page numbers 230-244, ISSN: 0162-8828

Zhao, W.; Chellappa, R. & Phillips, P. J. (1999). Subspace Linear Discriminant Analysis for Face Recognition, *Technical Report CAR-TR-914,* Center for Automation Research, University of Maryland, College Park

Pseudo 2D Hidden Markov Model and Neural Network Coefficients in Face Recognition

Domenico Daleno[1,2], Lucia Cariello[1,2],
Marco Giannini[1,2] and Giuseppe Mastronardi[1,2]
[1]Department of Electrotecnical and Electronics, Polytechnic of Bari,
Via Orabona, 4 – 70125 Bari – Italy
[2] e.B.I.S. s.r.l. (electronic Business In Security), Spin-Off of Polytechnic of Bari,
Via Pavoncelli, 139 – 70125 Bari – Italy

1. Personal Identification

For thousands of years, humans have instinctively used some physical characteristics (such as face, voice, posture, etc.) to recognize each other. About half the 800, A. Bertillon, chief of criminal identification section of the Paris police, plans to use some measures of the human body (height, length of arms, feet, fingers, etc.) to identify those responsible crimes. Towards the end of the nineteenth century, this original idea was further developed through the discovery (due studies F. Galton and E. Henry) the distinctiveness of fingerprints: they uniquely identify a person. Today, in full digital era, huge numbers of people use individual recognition techniques based on the identification of human characteristics, not only in justice but in civil and military applications. In fact, the only way to conclusively identify an individual is to recognize the personal characteristics. These are defined biometric features and, the technology behind this identification is called Biometric. The term Biometric, from the greek bios (life) and meters (measure), in computer sense, means the automatic identification or verification of the identity of a person based on physical characteristics and/or behavioral (CNIPA, 2004).

Biometric feature is described as a physiological or behavioral characteristic that can be measured and subsequently identified to confirm the identity of a person. We can then divide the biometrics in:

- physical biometric: it is that based on data derived from measurements made on a person's physical characteristics such as iris, fingerprint, facial features, hand or other;
- behavioral biometric: it is that based on aspects linked to behavioral characteristics such as, for example, the issue of voice, dynamic signing, or the type of gait.

As each biometric process starts with a preliminary phase called "enrollment" in which, generally, the person must provide the biometric system, through a sensor, its characteristic physical and behavioral, which is then converted into a mathematical model (template), two operating modes of biometrics are:

- 1:1 (one-to-one) for which the generated data by the biometric sensor are compared with a single template, creating, so, a verification process (Verification);
- 1: N (one-to-many) for which data are compared with a set of templates contained in a file, realizing, so, a process of identification (identification).

It is essential to emphasize that in biometric field two are terms usually used:

- Physical access: procedure for establishing ownership of the person entering a room, building area or area;
- Logical access: procedure for establishing ownership of the subject to make use of a computer resource.

For example, an employee of a firm could enter the office (physical access) via a biometric check between his physical characteristic (such as a fingerprint) and that deposited on a smart-card (process 1:1). To gain access to his computer (logical access) the same employee's fingerprint could be compared with that of authorized users, stored in archive (1: N) (Bossi, 2002; Maio, 2004).

1.1 Biometric Process

Biometric systems are characterized by a process of using that, in principle, it can be traced to the comparison operation of a physical characteristic or behavioral acquired by a person, with one or more of the same samples previously recorded. Both the recording that the comparison is made according to the following sequence of steps (CNIPA, 2004):

- ✓ *Stage of Registration (Enrollment):* in the process of enrollment, the user provides the biometric system a physical or behavioral feature by a capture device (such as a fingerprint scanner or a video camera). The sample is processed to extract the distinctive informations, which form the so-called template that can be defined as a mathematical representation of biometric data. The template consists essentially of a sequence of numbers from which it is generally impractical his reconstruction and it is, theoretically, comparable to a user's "physical password".

  At the end of the enrollment process, the template is registered. The registration is the most difficult step because of the importance of the choices to be made. First is necessary to identify as to save the template: because of the sensitivity of data and the possible impact on privacy, the information should be encrypted. Second is indispensable determined where to store and where to save the model, for example on a chip card in a database, a local workstation or directly on the capture device.

  The different possibilities lead to restrictions: if a system that must handle a large number of users is used, the latter two types are not applicable to matters concerning the physical size and required computing power. By using a database, is important to consider that the data could be stolen and used in a manner not acceptable. Saving in a chip can be a good solution; however, is necessary to sign digitally the saved template and to apply security techniques which take into account the fault-based attacks (Bossi, 2002);
- ✓ *Verification step:* During the verification process, the acquisition of the sample and extraction of the template are made as before. The latter is compared with that already acquired to obtain both an authentication and recognition.
- ✓ *Authentication phase:* if the objective is the subject's authentication , the biometric system attempts to provide an answer to the question "The person is who he claimed to be?",

making a comparison 1 to 1 between the template of the subject and the reference template stored in the archive (or on a smart card). Authentication requires that the identity is provided, for example, typing a username or a pin and the output of the comparison algorithm is a score, which is positive if it occurs above a certain threshold, and negative if below this threshold. The threshold for comparison is an adjustable parameter of the system (CNIPA, 2004).

- ✓ *Recognition/identification phase:* in this case, the system determines the user's identity, or attempts to provide an answer to the question "Who is the user?", making a lot of confrontations with the biometric data models registered in its archives. When the search algorithm produces as output a score higher than the so-called "threshold", is reported a match (called "matching" or "hit") (CNIPA, 2004). Authentication is generally a cooperative process (ouvert), while identification may also be a poster or even hidden from users (covert). While in the cooperative process the subject voluntarily manifest his own identity, usually to go to a place (physical access) or use a service (logical access); in the case of hidden biometrics, the physical and/or behavioral characteristics are matched, without the person knows, with those stored in an archive.
- ✓ *Performance Mesurement:* in this performance of a biometric system are evaluated according to three parameters: size, speed and accuracy (Bossi, 2002). The size of the model have relevance to extract device storage used, consider, for example to smart-card having a memory limited. The speed with which gives a positive or negative response is discriminating about the possible use in identification rather than verification. Accuracy is a rather critical parameter to determine because of the approach probabilistic biometric systems adopted in the choice. The types of errors that can make a biometric system are essentially two: False acceptances, an unauthorized user is authenticated by the system because its footprint is quite similar to a model previously filed; False discards, an authorized user is rejected by the system because its footprint is not sufficiently similar to the model with which it was compared.

1.2 Biometric Tecniques

Currently the efforts of the scientific community and industrial research are oriented to the study of those variables that permit reliable identification of individuals. Biometric identification techniques are indeed aimed at identifying a person based on its unique physiological or behavioral characteristics, difficult to alter or simulate. The most common evaluate the follow features:

- Fingerprints;
- Iris
- Retina vasculature
- Dynamics of attaching the signature
- Face
- Hand geometry
- Vocal timbre
- Multiple biometrics

1.3 Face Recognition

The recognition of facial features is perhaps one of biometric technologies more fascinating and that users consider psychologically less repulsive. System for face recognition is based on the physical characteristics of the face and is then the closest, in theory, to the human concept of "personal recognition". The enrollment usually takes a few seconds that are required to frame more static images of the face. Some systems can classify the user from multiple angles obtaining a three dimensional model of the face. This last, according to the different acquisition modes, varying in size from 100 to 3,500 byte (CNIPA, 2004). The user's acceptance of the feature based biometric recognition is generally high, since the natural and not invasive nature of the acquisition method. The sensor's prices may also be in the range of the hundreds euro's for logical access and personal computer systems, but they can remarkably increase for more sophisticated systems. Moreover the face recognition biometric technique has the advantage to be low invasiveness (no physical contact) and to provide the possibility of acquiring a distance a subject to recognize. Usually the first step of any fully automatic system that analyzes the information contained in faces, e.g. identity verfication, is the Face Detection. Face detection is concerned with finding whether or not there are any faces in a given image (usually in gray scale) and, if present, return the image location and content of each face.

1.3.1 Face Recognition Phases

In general, facial recognition can be decomposed into four phases (Medugno et al., 2007):

- *Pre-processing*: This means ensuring that the image which is applied to the recognition process meets certain required standards: for such that the face is located in the center of the image and provided part of the same; that the background satisfies certain constraints, and so on. Usually this phase is done by sampling equipment designed to image through mechanisms that tend to prevent the user from providing distorted images: an example may be the sensors necessary to capture the image when the subject is an acceptable distance.
- *Phase segmentation or localization*: is the exact location of the face or certain parts of it. This phase arises from the need to characterize, through some characteristic features, the face of a subject.
- *Feature Extraction Phase*: maybe it is the core of the whole face recognition process. A feature it's a characteristic useful for distinguish a face from another. It can be extracted from the image through different kind of processes. Usually, higher is amount of extracted features, the higher is the capacity of discrimination between similar faces. Some interesting features are, for example, the eyes or hairs color, the nose or the mouth shape. Those features are usually referred as locals because they refer to a particular and restricted area of the image.
- *Recognition Phase*: once the image is associated with an array of values, the recognition problem reduces itself to a widely studied problem in the past literature: the main part of those is then mainly related to the features extraction. The recognition problem can be divided into three phases: deciding over which features the recognition will be done; automatic extracting the chosen parameters from the face digitalized image; classifying the faces over the acquired parameters base.

2 State of art of Two - Dimensional Face Recognition

The efforts of researchers over the past 30 years have resulted in many sophisticated and mature 2D face recognition algorithms. In this section it is represented a brief description recurring methods existing in literature for face recognition.

The Principal Component Analysis (PCA) is one of the most successful techniques that have been used in image recognition and compression. PCA is a statistical method under the broad title of *factor analysis*. The purpose of PCA is to reduce the large dimensionality of the data space (observed variables) to the smaller intrinsic dimensionality of feature space (independent variables), which are needed to describe the data economically. The PCA techniques consist of: eigenfaces in which the face images are projected onto a features space that best encodes the variation among known faces images, that is use a nearest neighbor classifier (Turk & Pentland, 1991), (Craw & Cameron, 1996); feature-line based methods, which replace the point-to-point distance with the distance between a point and the feature line linking two stored sample points (Li & Lu, 1999); Fisherfaces (Swets & Weng, 1996; Belhumeur et al., 1997; Zhao et al., 1998), which use Linearr/Fisher Discriminant Analysis (FLD/LDA) (Liu & Wechsler, 2000); Bayesian methods, which use a probabilistic, distance metric (Moghaddam & Pentland, 1997); and SVM methods, which use a support vector machine as the classifier (Phillips, 1998). Utilizing higher-order statistics, Independent Component Analysis (ICA) is argued to have more representative power than PCA, and hence may provide better recognition performance than PCA (Bartlett et al., 1998). Being able to offer potentially greater generalization through learning, neural networks/learning methods have also been applied to face recognition. One example is the Probabilistic Decision-Based Neural Network (PDBNN) method (Lin et al., 1997) and the other is the evolution pursuit (EP) method (Etemad & Chellappa, 1997).

The category of feature based (structural) matching methods, using the width of the head, the distances between the eyes and from the eyes to the mouth, etc. (Kelly, 1970), or the distances and angles between eye corners, mouth extrema, nostrils, and chin top (Kanade, 1973). More recently, a mixture-distance based approach using manually extracted distances was reported (Manjunath et al., 1992; Cox et al., 1996). Without finding the exact locations of facial features, Hidden Markov Model (HMM) based methods use strips of pixels that cover the forehead, eye, nose, mouth, and chin (Samaria & Young, 1994), (Samaria, 1994; Nefian & Hayes III, 1998). (Nefian & Hayes III, 1998) reported better performance than (Samaria, 1994) by using the KL projection coefficients instead of the strips of raw pixels. One of the most successful systems in this category is the graph matching system (Wiskott et al., 1997), (Okada et al., 1998) which is based on the Dynamic Link Architecture (DLA). Using an unsupervised learning method based on a Self Organizing Map (SOM), a system based on a convolutional neural network (CNN) has been developed (Lawrence et al., 1997).

Moreover, in the hybrid method category, we will briefly review the modular eigenface method (Pentland et al., 1994), an hybrid representation based on PCA and Local Feature Analysis (LFA) (Penev & Atick, 1996), a flexible appearance model based method (Lanitis et al., 1995), and a recent development (Huang et al., 2003) along this direction. In (Samaria, 1994), the use of hybrid features by combining eigenfaces and other eigenmodules is explored: eigeneyes, eigenmouth, and eigen-nose. Though experiments show slight improvements over holistic eigenfaces or eigenmodules based on structural matching, we believe that these types of methods are important and deserve further investigation. Perhaps many relevant problems need to be solved before fruitful results can be expected, e.g., how to optimally arbitrate the use of holistic and local features.

Many types of systems have been successfully applied to the task of face recognition, but they all have some advantages and disadvantages. Appropriate schemes should be chosen starting from the specific requirements of a given task. Most of the systems reviewed here focus on the subtask of recognition, but others also include automatic face detection and feature extraction, making them fully automatic systems (Moghaddam & Pentland, 1997; Wiskott et al., 1997; Lin et al., 1997).

3. Artificial Neural Network

An artificial neural network is a system that tries to reproduce the operation of biological neural networks or, in other words, is an emulation of the biological neural system. This approach give the chance of performing tasks that a linear program is not able to do exploiting its capability of learning with no needs of writing new code lines. These advantages have a cost. They need a good training to operate correctly and their computing time can be high for large Neural Networks. According to what has been said, an Artificial Neural Network is an adaptive nonlinear system that learns to perform a function from data. Adaptive means that the system parameters are changed during operation, normally called the training phase. After the training phase the Artificial Neural Network parameters are fixed and the system is deployed to solve the problem at hand (the testing phase). The Artificial Neural Network is built with a systematic step-by-step procedure to optimize a performance criterion or to follow some implicit internal constraint, which is commonly referred to as the learning rule. The input/output training data are fundamental in neural network technology, because they convey the necessary information to "discover" the optimal operating point. The nonlinear nature of the neural network processing elements (PEs) provides the system with lots of flexibility to achieve practically any desired input/output maps. In the case of supervised Neural Networks, in order to train an ANN, an input is presented to the neural network and a corresponding result is set at the output. The error is the difference between the desired response and the actual system output. The error information is fed back to the system so that it can adjust its parameters in a systematic way, following the adopted learning rule.

The process is repeated until the performance is acceptable. It comes clear that the performances of the trained Neural Network would be heavily influenced by the dataset that was used for the training phase. If it does not cover a significant portion of the operating conditions or if they are ambiguous, then neural network technology is probably not the right solution. On the other hand, if there is plenty of data and the problem is poorly understood to derive an approximate model, then neural network technology is a good choice. This operating procedure should be contrasted with the traditional engineering design, made of exhaustive subsystem specifications and intercommunication protocols. In artificial neural networks, the designer chooses the network topology, the performance function, the learning rule, and the criterion to stop the training phase, while the system automatically adjusts the parameters.

Thus it is difficult to bring a priori information into the design and, when the system does not work properly, it is also hard to refine the solution in a following step. At the same time, ANN-based solutions are extremely efficient in terms of development time and resources, and in many difficult problems artificial neural networks provide performance that is difficult to match with other technologies. At present, artificial neural networks are emerging as the technology of choice for many applications, such as pattern recognition, prediction, system identification, and control.

When creating a functional model of the biological neuron, there are three basic components of importance. First, the synapses of the neuron are modelled as weights. Operating in this way, the strength of the connection between an input and a neuron is noted by the value of the weight. Inhibitory connection will have negative weight values, while positive values designate excitatory connections. The next two components model the actual activity within the neuron cell. An adder sums up all the inputs modified by their respective weights. This activity is referred to as linear combination. Finally, an activation function controls the amplitude of the output of the neuron. An acceptable range of output is usually between 0 and 1, or -1 and 1.

Mathematically, this process is described in the Fig. 1:

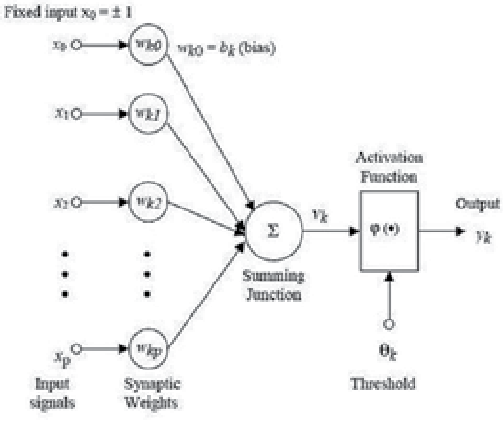


Fig. 1. Artificial Neural Network Process

From this model the interval activity of the neuron can be represented as:

$$v_k = \sum_{j=1}^{p} w_{kj} x_j \tag{1}$$

The output of the neuron, yk, would therefore be the outcome of some activation function on the value of vk.

The activation function is functions that compel the output of a neuron in a neural network inside certain values (usually 0 and 1, or -1 and 1). In general, there are three types of activation functions, denoted by Φ(). First, there is the **Threshold Function** which takes on a value of 0 if the summed input is lower than a certain threshold value (v), and the value 1 if the summed input is greater than or equal to the threshold value.

$$\varphi(v) = \begin{cases} 1 & if\ v \geq 0 \\ 0 & if\ v < 0 \end{cases} \tag{2}$$

Secondly, there is the **Piecewise-Linear function**. This function too admits values of 0 or 1 as input, but can also take on values belonging to that interval, depending on the amplification factor in a certain region of linear operation.

$$\varphi(v) = \begin{cases} 1 & v \geq \frac{1}{2} \\ v & -\frac{1}{2} > v > \frac{1}{2} \\ 0 & v \leq -\frac{1}{2} \end{cases} \tag{3}$$

Thirdly, there is the **sigmoid function**. This function can range between 0 and 1, but it is also sometimes useful to use the -1 to 1 range. An example of the sigmoid function is the hyperbolic tangent function.

$$\varphi(v) = \tanh\left(\frac{v}{2}\right) = \frac{1-\exp(-v)}{1+\exp(-v)} \tag{4}$$

The pattern of connections between the units and the propagation of data are clustered into two main class:

- **Feed-forward neural networks**, where the data flow from input to output units is strictly feedforward. The data processing can extend over multiple layers of units, but no feedback connections are present, that is, connections extending from outputs of units to inputs of units in the same layer or previous layers.
- **Recurrent neural networks** that do contain feedback connections. Contrary to feed-forward networks, the dynamical properties of the network are important. In some cases, the activation values of the units undergo a relaxation process such that the neural network will evolve to a stable state in which these activations do not change anymore. In other applications, the change of the activation values of the output neurons are significant, such that the dynamical behaviour constitutes the output of the neural network.

4. Hidden Markov Models

The Hidden Markov Models are stochastic models which provide a high level of flexibility for modelling the structure of an observation sequence. They allow for recovering the (hidden) structure of a sequence of observations by pairing each observation with a (hidden) state. Hidden Markov Models (HMMs) represent a most famous statistical pattern recognition technique and can be considered as the state-of-the-art in speech recognition. This is due to their excellent time warping capabilities, their effective self organizing learning capabilities and their ability to perform recognition and segmentation in one single step. They are used not only for speech and handwriting recognition but they are involved in modelling and processing images too. This is the case of their use in the face recognition field.

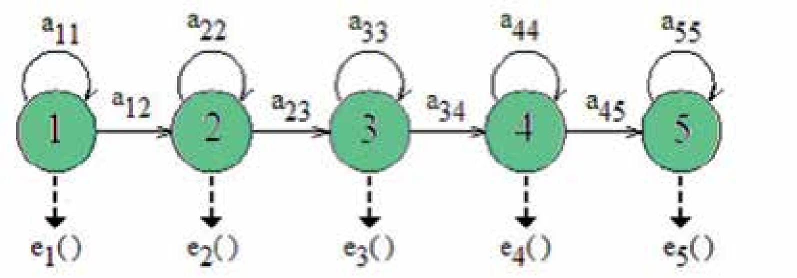


Fig. 2. "*Left to Right*" Hidden Markov Model – 5 state

4.1 One - Dimensional Hidden Markov Models

The HMM are characterised by two interrelated processes (Samaria & Young, 1994):

1. An unobservable Markov chain with a finite number of states, a state transition probability matrix and an initial state probability distribution.
2. A set of probability density functions for each state.

The elements that characterised a HMMs are:

- $N = |S|$ which represent the number of states of the model. Where S is the set of the states and can be shown as $S = \{s_1, s_2, \ldots, s_n\}$, where s_i is one of the states that can be employed by the model. To describe the system, T observation sequences are used, where T is the number of observations. The state of the model at time t is given by q_t *in* S, $1 < t < T$;
- $M = |V|$ is the number of different observation symbols. If V is the set of all possible observation symbols (also called the codebook of the model), then $V = \{v_1, v_2, \ldots, v_M\}$;
- $A = \{a_{ij}\}$ is the state transition probability matrix, where a_{ij} is the probability that the state i became the state j:

$$a_{ij} = p(q_t = s_j \mid q_{t-1} = s_i) \tag{5}$$

where $1 \le i\,; j \le N$, with constraint $0 \le a_{i,j} \le 1$, *and* $\sum_{j=1}^{N} a_{ij} = 1$, $1 \le i \le N$

- $B=\{b_j(k)\}$ the observation symbol probability matrix, $b_j(k)$ is the probability to have the observation k when the state is j:

$$b_j(k) = p\,(O_t = v_k \mid q_t = S_j) \tag{6}$$

where $1 \le j \le N$; $1 \le k \le M$; and O_t is the observation symbol at time t.

- $\Pi = \{\pi_1, \pi_2, \ldots \pi_N\}$ is the initial state distribution:

$$\pi_i = p\,(q_j = S_i) \tag{7}$$

where $1 \le j \le N$.

Using a shorthand notation, a HMM is defined by the following expression:

$$\lambda = (A,B,\pi). \tag{8}$$

The training of the model, given a set of sequences $\{O_i\}$, is usually performed by means of the standard Baum-Welch re-estimation, which determines the parameters (A, B, π) that maximize the probability $P(\{O_i\} \mid \lambda)$.

4.2 Pseudo Two-Dimensional Hidden Markov Models

Pseudo Two-Dimensional Hidden Markov Models (P2D-HMMs) are an extension of the one-dimensional HMMs, applied to two-dimensional data. Fig. 3 shows a general schema of P2D-HMMs, which are also known as planar HMMs, that are stochastic automata with a two-dimensional arrangement of the states. A planar HMM can be seen as a vertical One-dimensional HMM that links together an indefinite number of super-states.

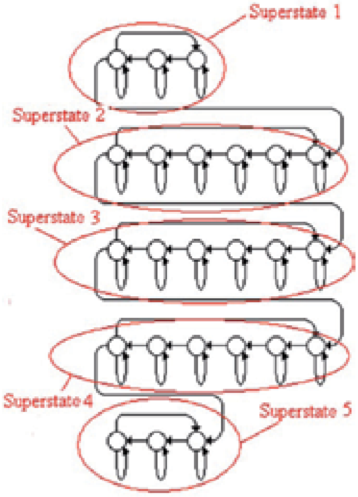


Fig. 3. General schema of a Pseudo Two-Dimensional Hidden Markov Model

Considering that a facial image can be subdivided into stripes, thus allowing the implementation of P2D-HMMs for modelling this kind of elaboration. Each stripe is aligned to one of the super-states of the P2D-HMMs, resulting in a horizontal warping of the pattern. Furthermore, the stripes can be vertically disposed, within the super-state, in a manner that the pattern related to the stripe result to be aligned to the vertical HMM states. In a similar way, it is possible to model any kind of data that can be considered represented by means horizontal stripes. The recognition process achieved by means of P2D-HMMs is pretty similar to the recognition process made with one-dimensional HMM as it was showed by Samaria (Samaria, 1994). The P2D-HMMs can be trained using the standard Baum-Welch algorithm and the recognition step can be carried out with the standard Viterbi algorithm.

The super-states is the model of the sequence of rows in the image and the linear 1D-HMMs, which are inside the super-states, are used to model each row (Nefian, 1998). The states sequence in each rows is independent from the states sequences of neighbouring rows.

Figure 3 shows the particular structure of the P2D-HMM that we use: the schema is 3-6-6-6-3, where the 1st and the 5th super-states are constituted by a left to right 1D-HMM with 3 states, while the 2nd, the 3rd and the 4th super-states are constituted by a left to right 1D-HMM with 6 states.

The formal representation of a Pseudo Two-Dimensional Hidden Markov Models can be given by the expression $\Lambda = \{\lambda, A, B, \Pi\}$ where,

- $\lambda = \{\lambda^{(1)}, \lambda^{(2)}, \dots, \lambda^{(N)}\}$ is the set of *N* possible super-states in the model.
- λ^i is a 1DHMM super-state, whose parameters are $\lambda^i = \{s^i, V, A^i, \pi^{(i)}\}$

In different words,

- $s = \{s_1{}^i, s_2{}^i, \dots, s_N{}^i\}$ is the set of N^i possible states of super-state λ^i.
- $V = \{V_1, V_2, \dots, V_L\}$ is the output alphabet (common to all super-states). In other words, for any *t*, there exist *l* such that $o_t = v_l$.
- $A^i = \{a_{kl}{}^i\}_{kl=1\dots N^i}$ is the set of transition probabilities within super-state λ^i.
- $B^i = \{b_k{}^i(l)\}_{k=1\dots N^i, l=1\dots L}$ is the set of output probabilities of super-state λ^i.
- $\pi = \{\pi_1{}^i, \pi_2{}^i, \dots, \pi_N{}^i\}$g is the set of initial state probabilities of super-state λ^i.
- $A = \{a_{ij}\}_{ij=1\dots N}$ is the set of transition probabilities through the states of the P2DHMM.
- $\Pi = \{\pi_1, \pi_2, \dots, \pi_N\}$ is the set of initial super-state probabilities of the P2DHMM.

Similarly to the one-dimensional model, the Pseudo two-dimensional Hidden Markov Models will associate a state sequence *Q* to an observation sequence $O = \{o_{xy}\}_{x=1\dots X, y=1\dots Y}$. The state sequence *Q* will consist of two levels. *Q* is primarily a super-state sequence $Q = \{Q_1, Q_2, \dots, Q_N\}$ indicating the super-state corresponding to the sequence of lines of observation $O = \{O_1, O_2, \dots, O_N\}$. Each state line Q_y is composed itself of states $q_{xy}\left(Q_y = \{q_{1y}, q_{2y}, \dots, q_{Xy}\}\right)$, each of them indicating the state of the corresponding 1DHMM at a position *(x; y)*.

A formal expression of the parameters of a P2DHMM can be given as follows:

- Super-state transition probability: $a_{ij} = P\left[q_y = \lambda^j \mid q_{y-1} = \lambda^i\right]$.
- Initial super-state probability: $\pi_i = P\left[Q_1 = \lambda^j \mid \Lambda\right]$.
- State transition probability of super-state $\lambda^i : a^i{}_{kl} = P\left[q_{xy} = s_l^i \mid q_{x-1y} = s_k^i\right]$.
- State output probability of super-state $\lambda^i : b^i{}_j(l) = P\left[o_{xy} = v_l \mid q_{xy} = s_j^i\right]$.
- Initial state probability: $\pi_j^i = P\left[q_{1y} = s_j^i \mid \lambda^i\right]$.

4.3 Hidden Markov Models applied to Face Recognition

The HMM can be applied to image processing. In consideration of the fact that the image can be seen as a two dimension matrix of data, according to Samaria, space sequences must be considered (Samaria, 1992). The idea is again to exploit the vertical sequential structure of a human face. A sequence of overlapping horizontal stripes are built on the image and the sequence of these stripes is labeled by means of a 1DHMM. Considering frontal face images, the facial region can be considered as the sum of 5 regions: forehead, eyes, nose, mouth and chin (Nefian & Monson, 1998).

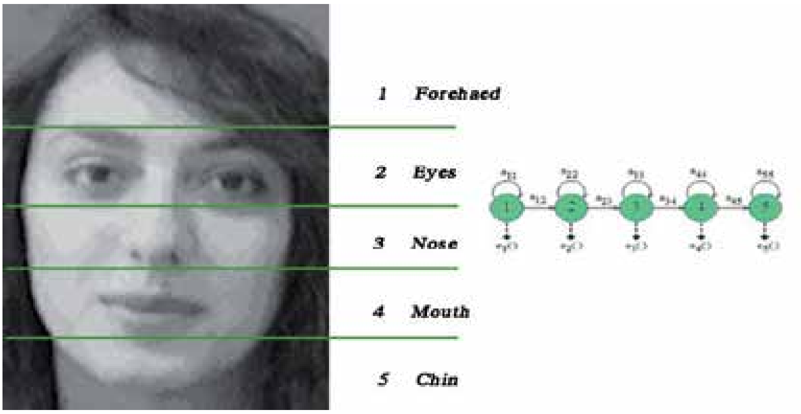


Fig. 4. The significant facial regions

Each of these facial regions (facial band) will correspond to a state in a left to right 1D continuous HMM. The Left-to-right HMM used for face recognition is shown in the previous figure. To recognize the face *k* the following HMM has been trained:

$$\boldsymbol{\lambda(k) =(A(k), B(k), p(k))} \tag{9}$$

The HMM should be trained for each person that we want to recognize subsequently. The HMM training, that equals to an enrolment operation for every subject of the database, requires a grey scale image of the face of each person. Each image of width *X* and height *Y* is divided into overlapping blocks of height *L* and width *W*. The amount of overlap between bounding blocks is *M*.

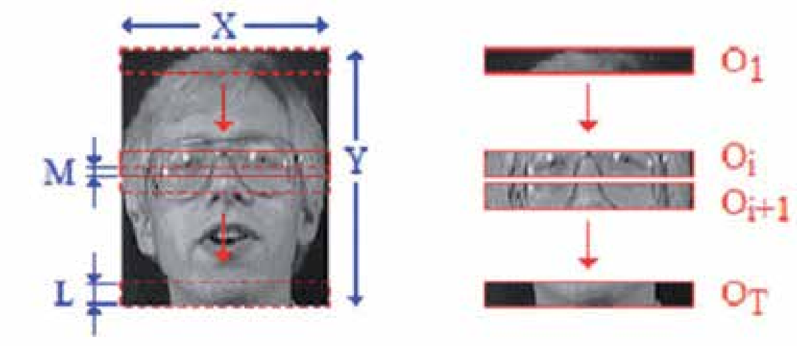


Fig. 5. Extraction of overlapping blocks from the face

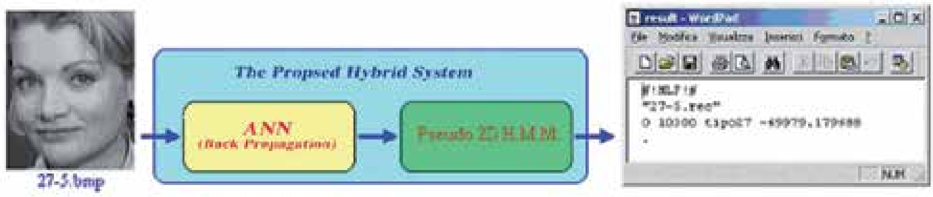
The number of blocks extracted from each face image and the number of observation vectors *T* are the same and are given by:

$$T = \frac{(Y-L)}{(L-M)} + 1 \tag{9}$$

The system recognition rate is significantly affected by the parameters M and L that, for this reason, should be chosen accurately. Increasing the overlap area M can significantly increase the recognition rate because it allows the features to be captured in a manner that is independent of their position along the vertical axis. The choice of parameter L is more delicate. An insufficient amount of information about the observation vector could arise from a small value of the parameter L while, on the contrary, large values of L are dangerous as the probability of cutting across the features increase. However, as the system recognition rate is more sensitive to the variations in M than in L, $M \leq (L - 1)$ is used.

5. The System Proposed

The system for face recognition proposed, showed in the figure below, is an hybrid system as showed built as a cascade connection of two different systems: an Artificial Neural Network, existing in literature (Bevilacqua et al., 2006), and different representation of P2D-HMMs.


27-5.bmp
The Proposed Hybrid System
ANN
(Back Propagation)
Pseudo 2D H.M.M.
result - WordPad
File Modifica Visualizza Inserisci Formato ?
NUM

```
#!MLF!#
"27-5.rec"
0 10300 tipo27 -49979.179688
.
```

Fig. 6. The proposed hybrid system.

The system's input is an image of a person that must be recognised and the output is its identification with the corresponding rate of recognition. The experiments will be performed on a database obtained by the combination of the Olivetti Research Laboratory database (Samaria & Harter, 1994), and other profiles photos of persons disguised with dark glasses or bandage. These images are ".bmp" files in grey scales of 92x112 pixels.

The hybrid schema was built executing the following steps:

1. Training and saving of Artificial Neural Network;
2. Transformation of photos in HTK format;
3. Training of different P2D-HMM structures, and identification of the Validation Set subjects, for control a proper training of system;

5.1 Training and Saving of Artificial Neural Network

The considered faces are sequenced in observation windows, according to the Samaria model already described in the previous section, where the number of blocks extracted from each face image equals the number of observation vectors T, and is obtained from Eq. 9.

Table 1 collects the values of the parameters for the observation windows after the manipulation operated by this system.

X = width photo = 92 pixels	T = number of blocks for photos = 103
Y = height photo = 112 pixels	XxY = photo dimension =10304 pixels
L = height block = 10 pixels	XxL = block dimension = 920 pixels
M = blocks overlapping = 9 pixels	XxM = overlapping dimension = 828 pixels

Table 1. Characteristic parameters of photos and blocks.

The Artificial Neural Network utilized in this system uses the EBP (Error Back Propagation) algorithm and its task is to extract the main features from the image and then store them in a sequence of 50 bits, reducing the complexity of the problem and compressing the bitmap images in order to represent them with a number of coefficients smaller than pixels. The image is a facial feature of a face image; from this area we consider 103 segments of 920 pixels that represent the observable states of the model (Bevilacqua et al, 2006).

Now all of these sections are divided into features of 230 pixels, that are the input of the Artificial Neural Network. The ANN is composed of three layers where the first layer is formed by 230 neurons, one neuron per each pixel, the hidden layer is composed by 50 units and the last layer by 230 neurons. After the training, the ANN is able to work as a pure linear function, the input of the first layer must be the same of the output of the last layer. The "compressed image" is described by 50 bits that are the outputs of an hidden layer consisting of an Heaviside function processing elements. For any window of 230 pixels we have an array of 50 elements, this means that a section of 920 pixels is compressed in a 4 sub-windows of 50 binary values array each. The matrix weights, referred to the connections between the inputs and the hidden layer, codifies the image bitmap, while the matrix weights associated to the connections between the hidden layer and the outputs, decodes the sequence of bits. Each of the 103 blocks of 920 pixels (4x230) gives 103 observation vectors with 200 coefficients (4x50) and the compression rate equals to

$$\frac{(103\text{x}920)}{(103\text{x}200)} = 4{,}6 \tag{10}$$

By observing the schema of Fig. 7 it is possible to note that the "Training Set" used for ANN is composed by 300 photos: 10 face images for each of the first 30 samples of the database. The training function is iterated 200 times for each photo and, at the end of the training phase, the neuron weights are saved in a ".bin" file. Finally the ANN is tested with other images, of the same size of the training images, representing the same subject used for the training, but, of course, different from those belonging to the training set features.

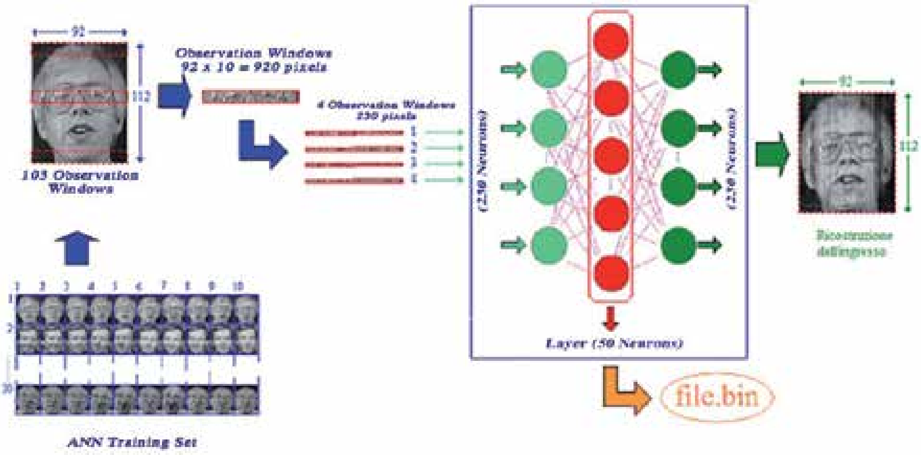


Fig. 7. Schema of the ANN training phase

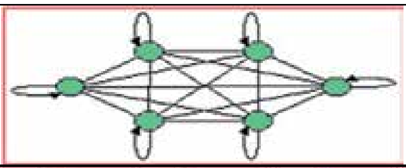
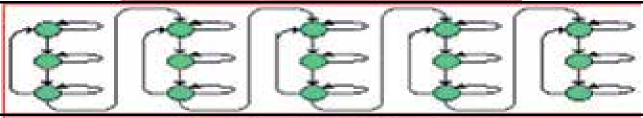
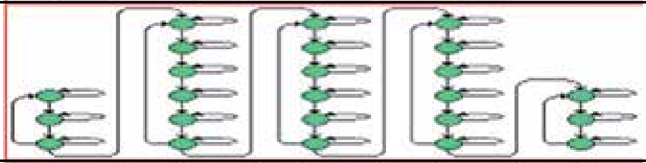
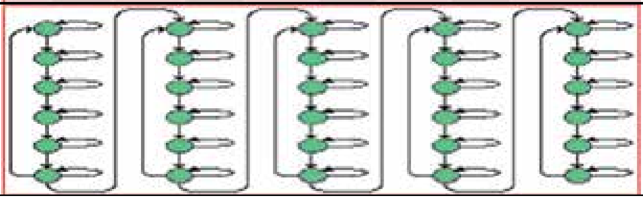
5.2 Transformation of Photos in HTK Format

After compressing the image containing the face into an observation vector of 103 elements of 200 binary (1/0) values, it will be computed by the Pseudo 2D Hidden Markov Models. The operations of building and manipulating the Hidden Markov models has been computed by the Hidden Markov Model ToolKit (HTK) (Young and Young, 1994). The HTK supports HMMs using both continuous density mixture Gaussians and discrete distributions and can be used to build complex HMM systems.

Finally, is necessary to transform the ANN output ".bin" file into another binary file in HTK format. The HTK like binary file has got an header, that should accomplish the HTK syntax, and 20600 coefficients (103x200), according the "Little Endian" data storage, which is commonly used by Motorola processors, IBM and Sun. Little Endian format provides the least significant byte is stored in the first memory location while the most significant byte is the last memory location.

5.3 Training of Different P2D-HMM Structures and Identification of the Validation Set subjects

Every subject populating the database was used to train the Pseudo 2D Hidden Markov Model and a Markov Model was associated to each of them. The different Hidden Markov Model structures were then trained. The table below reports the training results of one Ergodic HMM with 5 state and four Pseudo 2D Hidden Markov Model structures, that differs one by the others for the number of states in a super-state. The Table helps the comparison between the different performance and the choice of the structure that gives the best recognition rate. In table 2 are represented different Hidden Markov Model structures.

HMM **5Ergodic**	
Pseudo *2D HMM* **3-3-3-3-3**	
Pseudo *2D HMM* **3-6-6-6-3**	
Pseudo *2D HMM* **6-6-6-6-6**	

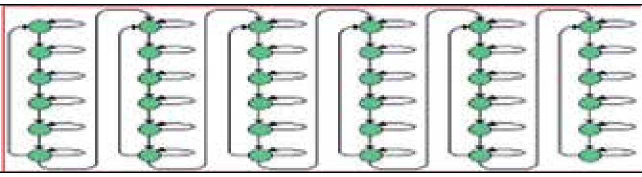
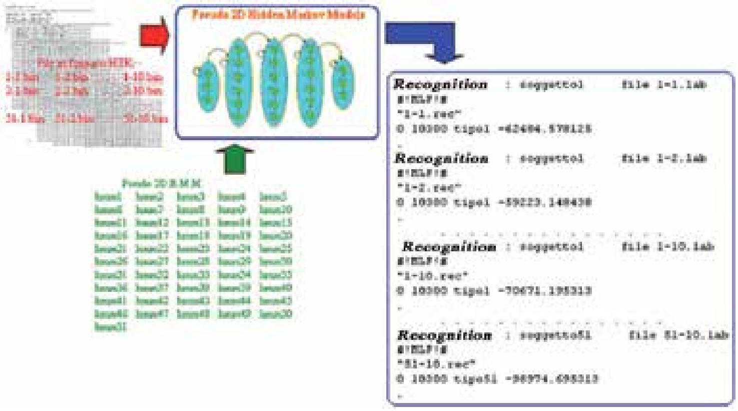
Pseudo *2D HMM* **6-6-6-6-6-6**	

Table 2. Different Hidden Markov Models structures.

After the P2D-HMM training process was completed, it was possible to proceed with the recognition phase, according the schema shown in Fig. 8.


Pseudo 2D Hidden Markov Models
File in Formato HTK
1-1.bin 1-2.bin 1-10.bin
2-1.bin 2-2.bin 2-10.bin
51-1.bin 51-2.bin 51-10.bin
Pseudo 2D H.M.M.
hmm1 hmm2 hmm3 hmm4 hmm5
hmm6 hmm7 hmm8 hmm9 hmm10
hmm11 hmm12 hmm13 hmm14 hmm15
hmm16 hmm17 hmm18 hmm19 hmm20
hmm21 hmm22 hmm23 hmm24 hmm25
hmm26 hmm27 hmm28 hmm29 hmm30
hmm31 hmm32 hmm33 hmm34 hmm35
hmm36 hmm37 hmm38 hmm39 hmm40
hmm41 hmm42 hmm43 hmm44 hmm45
hmm46 hmm47 hmm48 hmm49 hmm50
hmm51

```
Recognition  : soggetto1     file 1-1.lab
#!MLF!#
"1-1.rec"
0 10000 tipo1 -62484.578125
.
Recognition  : soggetto1     file 1-2.lab
#!MLF!#
"1-2.rec"
0 10000 tipo1 -59223.148438
.
           . . . . . . . . . . . . . . . .
 Recognition : soggetto1     file 1-10.lab
#!MLF!#
"1-10.rec"
0 10000 tipo1 -70671.195313
.
           . . . . . . . . . . . . . . . .
Recognition  : soggetto51     file 51-10.lab
#!MLF!#
"51-10.rec"
0 10000 tipo51 -98974.695313
.
```

Fig. 8. Schema of recognition phase.

The Viterbi algorithm is applied to each of the P2D-HMMs, built after the training phase, using the same HTK file (26-5.bmp in Fig.9). Each of the P2D-HMMs returns a logarithmic probability value. The highest probability value identifies the P2D-HMM and so the corresponding recognised sample as showed in Figure 9.

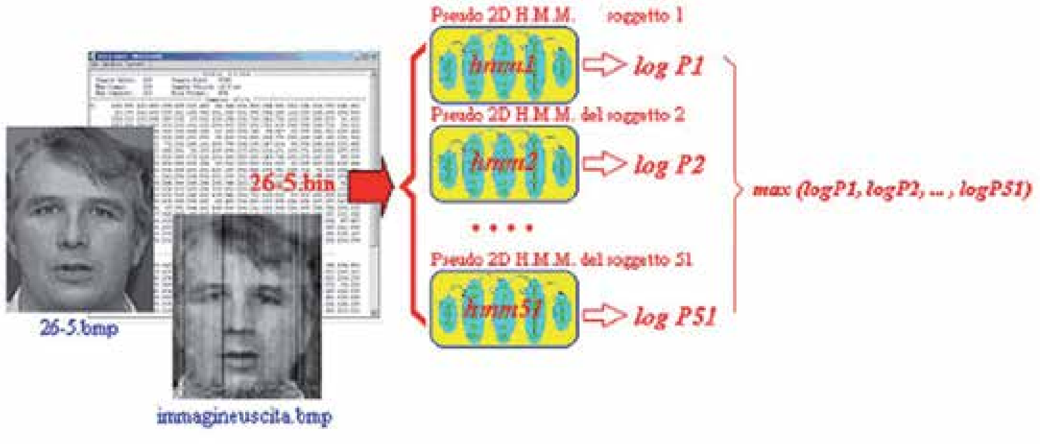

Pseudo 2D H.M.M. soggetto 1
hmm1
log P1
Pseudo 2D H.M.M. del soggetto 2
hmm2
log P2
26-5.bin
• • • •
Pseudo 2D H.M.M. del soggetto 51
hmm51
log P51
max (logP1, logP2, ... , logP51)
26-5.bmp
immagineuscita.bmp

Fig. 9. Example of identification by logarithmic probability.

At the end of the process the final outcome of the identification is the recognised person and the logarithmic probability value of his similarity to the template.

6. Experimental Result

As said in the preceding paragraphs, different Hidden Markov Model structure was tested on a database obtained as a combination of the Olivetti Research Laboratory database together with other photos of persons camouflaged wearing dark glasses, scarf or bandage, in order to check system reliability. The results are shown in the Table 3, here below.

Hidden Markov **Models**	**The exact identification** $100 - \frac{n^{\circ} errors}{5,1}$
Pseudo 2D **3-3-3-3-3**	99.80 % (1 error on 510 photo)
Pseudo 2D **3-6-6-6-3**	**100 %**
Pseudo 2D **6-6-6-6-6**	99.80 % (1 error on 510 photo)
Pseudo 2D **6-6-6-6-6-6**	99.80 % (1 error on 510 photo)
5-Ergodic	98.82 % (6 error on 510 photo)

Table 3. Rates of recognition obtained from the different implemented P2D-HMMs

The recognition rate was satisfying for all the HMM tested structures, but the system using the HMM structure 3-6-6-6-3 gave a percentage of identification of 100%, that is to say that any of the 510 photo tested were properly recognized.

Subsequently was made an experimental comparison of the results obtained with the hybrid system ANN-P2DHMM (using an HMM with structure 3-6-6-6-3) with the most important face recognition algorithms proposed in the literature when applied to the ORL images.

Methods	**Recognition Rate**	**Reference**
Eigenface	90.5%	Samaria, 1994
Pseudo 2D HMM feature: gray values	94.5%	Samaria, 1994
Convolutional Neural Network	96.2%	Lawrence et al., 1997
Pseudo 2D HMM feature: DCT Coefficients	99.5%	Eickeler, 1998
Ergodic HMM + DCT	99.5%	Kohir & Desai, 1998
Pseudo 2D HMM + **Neural Network Coefficients**	**100%**	**This work.**

Table 4. Comparative results on ORL database.

Table 4 resumes the results obtained and highlights that the hybrid system that combines Artificial Neural Networks and Pseudo 2D Hidden Markov Model produced the best Recognition Rate.

This result encourages the prosecution of the research to obtain a fundamental surplus to enhance the P2D-HMMs potentiality, allowing an efficient and sure personal identification process.

7. Reference

Bartlett, M.S.; Lades, H.M. & Sejnowski, T. (1998). Independent Component Representation for Face Recognition, *Proceedings of SPIE Symposium on Electronic Imaging: Science and Technology*, pp. 528-539, 1998

Belhumeur, P.N.; Hespanha, J.P. & Kriegman, D.J. (1997). Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 19, pp. 711-720, 1997

Bevilacqua, V., Mastronardi, G., Pedone, G., Romanazzi, G. & Daleno, D. (2006). Hidden Markov Model for Recognition using Artificial Neural Networks. Springer-Verlag, Heidelberg, New York, 2006.

Bossi, F. (2002). Biometrica per la sicurezza. http://microservo.altervista.org/Collaborazioni/biometrica%20per%20la%20sicurezza.pdf, March 2002

CNIPA. (2004). "Brevi note sulle tecnologiche biometriche". http://www.privacy.it/cnipabiometria.html, January 2004

Cox, I.J.; Ghosn, J. & Yianilos, P.N. (1996). Feature-based Face Recognition Using Mixture-Distance, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 209-216, 1996

Craw, I. & Cameron, P. (1996). Face Recognition by Computer, *Proceedings of British Machine Vision Conference* , pp. 489-507, 1996

Etemad, K. & Chellappa, R. (1997). Discriminant Analysis for Recognition of Human Face Images. *Journal of the Optical Society of America A*, Vol. 14 (1997), pp. 1724-1733

Kanade, T. (1973). *Computer Recognition of Human Faces*, Basel and Stuttgart: Birkhauser

Kelly, M.D. (1970). *Visual Identification of People by Computer, Technical Report AI-130, Stanford AI Project*, Stanford, CA

Lanitis, A.; Taylor, C.J. & Cootes, T.F. (1995). Automatic Face Identification System Using Flexible Appearance Models. *Image and Vision Computing*, Vol. 13 (1995), pp. 393-401.

Lawrence, S.; Giles, C.L.; Tsoi, A.C. & Back, A.D. (1997). Face Recognition: A Convolutional Neural-Network Approach. *IEEE Transaction on Neural Networks*, Vol. 8 (1997), pp. 98-113

Li, S.Z. & Lu, J. (1999). Face Recognition Using the Nearest Feature Line Method. *IEEE Transaction on Neural Networks*, Vol. 10 (1999), pp. 439-443

Lin, S.H.; Kung, S.Y. & Lin, L.J. (1997). Face Recognition/Detection by Probabilistic Decision-Based Neural Network. *IEEE Transaction on Neural Networks*, Vol. 8 (1997), pp. 114-132

Liu, C. & Wechsler, H. (2000). Evolutionary Pursuit and its Application to Face Recognition. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 22 (2000), pp. 570-582

Maio, D. (2004). Introduzione ai sistemi biometrici, http://www.cnipa.gov.it/site/files/introduzione%20ai%20sistemi%20biometrici%20prof%20Maio.pdf , 2004

Manjunath, B.S.; Chellappa, R. & Malsburg, C.v.d. (1992). A Feature Based Approach to Face Recognition, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 373-378, 1992

Medugno, V., Valentino, S. & Acampora, G. (2007). Sistemi biometrici http://www.dia.unisa.it/professori/ads/corsosecurity/www/CORSO9900/biometria /index.htm

Moghaddam, B. & Pentland, A. (1997). Probabilistic Visual Learning for Object Representation. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 19 (1997), pp. 696-710

Nefian, A. V. & Monson, H. (1998). Hidden Markov Models for Face Recognition, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP98), 2721-2724, 1998.

Nefian, A. V. (1998). A Hidden Markov Model-based approach for face detection and recognition. *Proposal for a Doctoral Dissertation*. 1998.

Okada, K.; Steffans, J.; Maurer, T.; Hong, H.; Elagin, E.; Neven, H. & Malsburg, C.v.d. (1998). The Bochum/USC Face Recognition System and how it Fared in the FERET Phase III Test, In: *Face Recognition: From Theory to Applications,* Wechsler, H.; Phillips, P.J.; Bruce, V.; Soulie, F.F. & Huang, T.S., (Eds.), pp. 186-205, Springer-Verlag, Berlin

Pentland, A.; Moghaddam, B. & Starner, T. (1994). View-based and Modular Eigenspaces for Face Recognition, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1994

Penev, P. & Atick, J. (1996). Local Feature Analysis: A General Statistical Theory for Object Representation. *Network: Computation in Neural Systems*, Vol. 7 (1996), pp. 477-500

Phillips, P.J. (1998). Support Vector Machines Applied to Face Recognition. *Advances in Neural Information Processing Systems*, Vol. 11 (1998), pp. 803-809

Samaria, F. (1992). Face identification using Hidden Markov Model. In *1st Year Report Cambridge University Engineering Department*, London, UK, 1992.

Samaria, F. (1994). Face Recognition Using Hidden Markov Models. *PhD Thesis, University of Cambridge*, 1994.

Samaria, F. & Young, S. (1994). Face HMM based architecture for face identification. In *Image and Computer Vision*, pages 537–583, 1994.

Young, S.J. & Young, Sj. (1994). The HTK Hidden Markov Model Toolkit: Design and Philosophy. *Entropic Cambridge Research Laboratory*, Ltd, Vol., 2, p.p. 2-44

Samaria, F. & Harter, A. (1994). Parameterisation of a Stochastic Model for Human Face Identification. *Proceedings of 2nd IEEE Workshop on Applications of Computer Vision,* Sarasota FL, December, 1994.

Samaria, F. & Young, S. (1994). HMM Based Architecture for Face Identification. *Image and Vision Computing*, Vol. 12 (1994), pp. 537-583

Samaria, F. (1994). *Face Recognition Using Hidden Markov Models*, PhD Thesis, University of Cambridge, UK

Swets, D.L. & Weng, J. (1996). Using Discriminant Eigenfeatures for Image Retrieval. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 18 (1996), pp. 831-836

Turk, M. & Pentland, A. (1991). Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, Vol. 3 (1991), pp. 72-86

Wiskott, L.; Fellous, J.M. & Malsburg, C.v.d. (1997). Face Recognition by Elastic Bunch Graph Matching. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 19 (1997), pp. 775-779

Zhao, W.; Chellappa, R. & Krishnaswamy, A. (1998). Discriminant Analysis of Principal Components for Face Recognition, *Proceedings of International Conference on Automatic Face*

10

VG-RAM Weightless Neural Networks for Face Recognition

Alberto F. De Souza, Claudine Badue, Felipe Pedroni, Stiven Schwanz Dias,
Hallysson Oliveira and Soterio Ferreira de Souza

Departamento de Informática
Universidade Federal do Espírito Santo
Av. Fernando Ferrari, 514, 29075-910 - Vitória-ES
Brazil

1. Introduction

Computerized human face recognition has many practical applications, such as access control, security monitoring, and surveillance systems, and has been one of the most challenging and active research areas in computer vision for many decades (Zhao et al., 2003). Even though current machine recognition systems have reached a certain level of maturity, the recognition of faces with different facial expressions, occlusions, and changes in illumination and/or pose is still a hard problem.

A general statement of the problem of machine recognition of faces can be formulated as follows: given an image of a scene, (i) identify or (ii) verify one or more persons in the scene using a database of faces. In identification problems, given a face as input, the system reports back the identity of an individual based on a database of known individuals; whereas in verification problems, the system confirms or rejects the claimed identity of the input face. In both cases, the solution typically involves segmentation of faces from scenes (face detection), feature extraction from the face regions, recognition, or verification. In this chapter, we examine the recognition of frontal face images required in the context of identification problems.

Many approaches have been proposed to tackle the problem of face recognition. One can roughly divide these into (i) *holistic* approaches, (ii) *feature-based* approaches, and (iii) *hybrid* approaches (Zhao et al., 2003). Holistic approaches use the whole face region as the raw input to a recognition system (a classifier). In feature-based approaches, local features, such as the eyes, nose, and mouth, are first extracted and their locations and local statistics (geometric and/or appearance based) are fed into a classifier. Hybrid approaches use both local features and the whole face region to recognize a face.

Among holistic approaches, eigenfaces (Turk & Pentland, 1991) and fisher-faces (Belhumeur et al., 1997; Etemad & Chellappa, 1997) have proved to be effective in experiments with large databases. Feature-based approaches (Gao & Leung, 2002; Lee & Seung, 1999; Li et al., 2001) have also been quite successful and, compared to holistic approaches, are less sensitive to facial expressions, variations in illumination and occlusion. Some of the hybrid approaches include the modular eigenface approach (Martinez, 2002), the Flexible Appearance Model approach (Lanitis et al., 1995), an approach that combines component-based recognition with 3D morphable models (Huang et al., 2003), and an approach that encodes geometric and structural information extracted from the face image in attributed relational graphs (ARG) and matches Face-ARG's for recognition (Park et al., 2005). Experiments with hybrid approaches showed slight improvements over feature-based approaches.

Recently, Wright et al. (2009) proposed a new approach to face recognition named Sparse Representation-based Classification (SRC). SRC is based on the compressive sampling theory (Candès & Wakin, 2008) and can use the whole face, a combination of features, or both features and the whole face for recognition. In SRC, the recognition problem is casted as one of classifying among multiple linear regression models. Wright et al. (2009) argue that compressive sampling offers the key to address this problem and, based on a sparse representation computed by ℓ^1-minimization, they propose a general classification algorithm for face recognition that provides new insights into what kind of transformation one should perform on face images to extract data to use as the input of the classifier of the recognition system. They showed that, if sparsity in the recognition problem is properly harnessed, the choice of transformation is no longer critical. What they found that is critical is whether the size of the data vector extracted is sufficiently large and whether the sparse representation is properly selected. They discovered that unconventional image transformations such as downsampling and random projections perform just as well as conventional ones such as eigenfaces, as long as the dimension of the data vector extracted surpasses certain threshold, predicted by the theory of sparse representation (Wright et al., 2009).

Virtual Generalizing Random Access Memory Weightless Neural Networks VG-RAM WNN (Aleksander, 1998) is an effective machine learning technique that offers simple implementation and fast training and test. In this chapter, we evaluated the performance of VG-RAM WNN on face recognition using the well known AR Face Database (Martinez & Benavente, 1998) and Extended Yale Face Database B (Georghiades et al., 2001; Lee et al., 2005). We examined two VG-RAM WNN architectures, one holistic and the other feature-based, each implemented with different numbers of neurons and synapses per neuron. Using the AR Face Database, we compared the best VG-RAM WNN performance with that of: (i) a holistic approach based on principal component analysis (PCA) (Turk & Pentland, 1991); (ii) feature-based approaches based on non-negative matrix factorization (NMF) (Lee & Seung, 1999), local non-negative matrix factorization (LNMF) (Li et al., 2001), and line edge maps (LEM) (Gao & Leung, 2002); and (iii) hybrid approaches based on weighted eigenspace representation (WER) (Martinez, 2002) and attributed relational graph (ARG) matching (Park et al., 2005). In addition, using both the AR Face Database and the Extended Yale Face Database B, we compared the best VG-RAM WNN performing architecture (feature-based) with that of SRC. We selected these approaches for comparison because they are representative of some of the best techniques for face recognition present in the literature. Our results showed that, even training with a single face image per person, VG-RAM WNN outperformed PCA, NMF, LNMF, LEM, WER, and ARG approaches under all face conditions tested. Also, training and testing in the same conditions as those employed by Wright et al. (2009) (downsampled face images), VG-RAM WNN outperformed SRC. These results show that VG-RAM WNN is a powerful technique for tackling this and other important problems in the pattern recognition realm.

This chapter is organized as follows. Section 2 introduces VG-RAM WNN and Section 3 describes how we have used them for face recognition. Section 4 presents our experimental methodology and experimental results. Our conclusions follows in Section 5.

2. VG-RAM WNN

RAM-based neural networks, also known as n-tuple classifiers or weightless neural networks, do not store knowledge in their connections but in Random Access Memories (RAM) inside the network's nodes, or neurons. These neurons operate with binary input values and use RAM as lookup tables: the synapses of each neuron collect a vector of bits from the network's inputs that is used as the RAM address, and the value stored at this address is the neuron's output. Training can be made in one shot and basically consists of storing the desired output in the address associated with the input vector of the neuron (Aleksander, 1966) (see Figure 1).

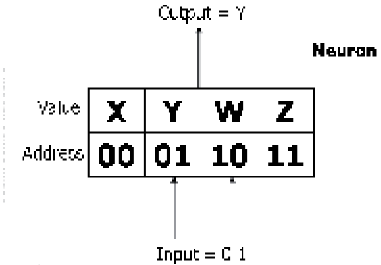
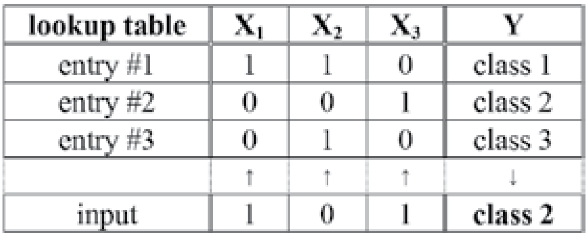


Fig. 1. Weightless neural network.

In spite of their remarkable simplicity, RAM-based neural networks are very effective as pattern recognition tools, offering fast training and test, in addition to easy implementation (Aleksander, 1998). However, if the network input is too large, the memory size becomes prohibitive, since it must be equal to 2^n, where n is the input size. Virtual Generalizing RAM (VG-RAM) Weightless Neural Networks (WNN) are RAM-based neural networks that only require memory capacity to store the data related to the training set (Ludermir et al., 1999). In the neurons of these networks, the memory stores the input-output pairs shown during training, instead of only the output. In the test phase, the memory of VG-RAM WNN neurons is searched associatively by comparing the input presented to the network with all inputs in the input-output pairs learned. The output of each VG-RAM WNN neuron is taken from the pair whose input is nearest to the input presented—the distance function employed by VG-RAM WNN neurons is the *hamming distance*. If there is more than one pair at the same minimum distance from the input presented, the neuron's output is chosen randomly among these pairs.

Figure 2 shows the lookup table of a VG-RAM WNN neuron with three synapses (X_1, X_2 and X_3). This lookup table contains three entries (input-output pairs), which were stored during the training phase (entry #1, entry #2 and entry #3). During the test phase, when an input vector (input) is presented to the network, the VG-RAM WNN test algorithm calculates the distance between this input vector and each input of the input-output pairs stored in the lookup table. In the example of Figure 2, the *hamming distance* from the input to entry #1 is two, because both X_2 and X_3 bits do not match the input vector. The distance to entry #2 is one, because X_1 is the only non-matching bit. The distance to entry #3 is three, as the reader



lookup table	X_1	X_2	X_3	Y
entry #1	1	1	0	class 1
entry #2	0	0	1	class 2
entry #3	0	1	0	class 3
	↑	↑	↑	↓
input	1	0	1	**class 2**

Fig. 2. VG-RAM WNN neuron lookup table.

may easily verify. Hence, for this input vector, the algorithm evaluates the neuron's output, Y, as class 2, since it is the output value stored in entry #2.

3. Face Recognition with VG-RAM WNN

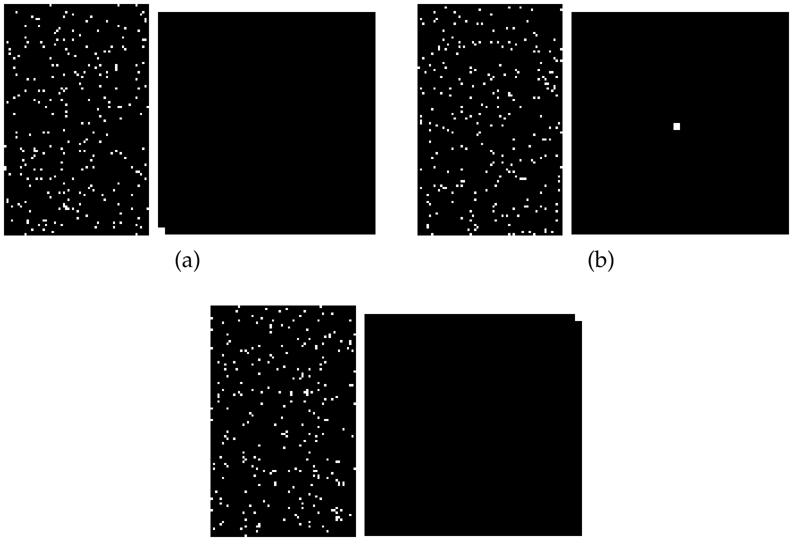
We examined the recognition part of the face identification problem only. That is, in the experiments described in this chapter, the segmentation of faces from images (face detection) is performed semi-automatically. Also, thanks to the properties of the VG-RAM WNN architectures employed, explicit feature extraction (e.g., line edge extraction; eye, nose, or mouth segmentation; etc.) is not required, even though in one of the two VG-RAM WNN architectures studied some neurons specializes in specific regions of the faces and, because of that, we say it is feature-based. The other VG-RAM WNN architecture studied is holistic.

3.1 Holistic Architecture

The holistic architecture has a single bidimensional array of $m \times n$ VG-RAM WNN neurons, N, where each neuron, $n_{i,j}$, has a set of synapses $W = \{w_1, \ldots, w_{|W|}\}$, which are randomly connected to the network's bidimensional input, Φ, of $u \times v$ inputs, $\varphi_{k,l}$ (see Figure 3 and Figure 4). The random synaptic interconnection pattern of each neuron $n_{i,j}$, $\Omega_{i,j}(W)$, is created when the network is built and does not change afterwards.

VG-RAM WNN synapses can only get a single bit from the input. Thus, in order to allow our VG-RAM WNN to deal with images, in which a pixel may assume a range of different values, we use *minchinton cells* (Mitchell et al., 1998). In the proposed VG-RAM WNN architectures, each neuron's synapse, w_t, forms a minchinton cell with the next, w_{t+1} ($w_{|W|}$ forms a minchinton cell with w_1). The type of the minchinton cell we have used returns 1 if the synapse w_t of the cell is connected to an input element, $\varphi_{k,l}$, whose value is larger than the value of the element $\varphi_{r,s}$ to which the synapse w_{t+1} is connected, i.e., $\varphi_{k,l} > \varphi_{r,s}$; otherwise, it returns zero (see the synapses w_1 and w_2 of the neuron $n_{m,n}$ of Figure 4).

The input face images, I, of $\xi \times \eta$ pixels (Figure 4) must be transformed in order to fit into the network's input, Φ. In the case of the AR Face Database, the images are rotated, scaled and cropped (Figure 5); the rotation, scaling and cropping are performed semi-automatically, i.e., the position of the eyes are marked manually and, based on this marking, the face in the image is computationally adjusted to fit into Φ. Before being copied to Φ, the transformed image is filtered by a Gaussian filter to smooth out artifacts produced by the transformations (Figure 5(c)). In the case of the Extended Yale Face Database B, only scaling and filtering are



(a)

(b)

(c)

Fig. 3. The synaptic interconnection pattern of the holistic architecture. (a) Left, input Φ: in white, the elements $\varphi_{k,l}$ of the input Φ that are connected to neuron $n_{1,1}$ of N via $\Omega_{1,1}(W)$. Right, neuron array N: in white, the neuron $n_{1,1}$ of N. (b) Left: in white, the elements $\varphi_{k,l}$ of Φ connected to $n_{\frac{m}{2},\frac{n}{2}}$ via $\Omega_{\frac{m}{2},\frac{n}{2}}(W)$. Right: in white, the neuron $n_{\frac{m}{2},\frac{n}{2}}$ of N. (c) Left: in white, the elements of Φ connected to $n_{m,n}$ via $\Omega_{m,n}(W)$. Right: in white, the neuron $n_{m,n}$.

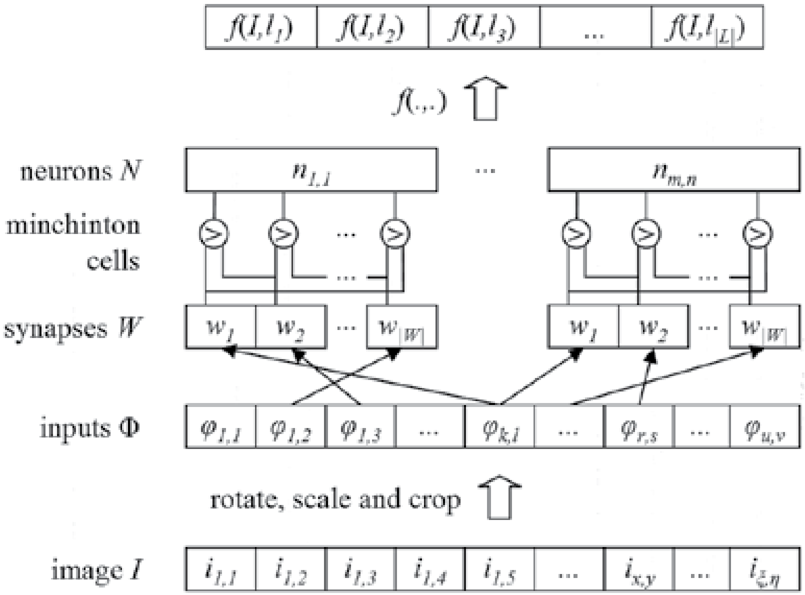


Fig. 4. Schematic diagram of the holistic and feature-based VG-RAM WNN architectures.

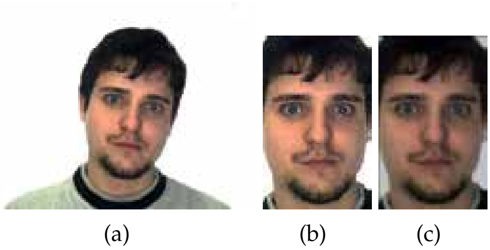


Fig. 5. Face image and its preprocessing. (a) Original image; (b) rotated, scaled and cropped image; and (c) filtered image.

necessary, since this database includes versions of the images already properly cropped (Lee et al., 2005).

During training, the face image I_x of a person p is transformed and filtered, and its pixels are copied to the VG-RAM WNN's input Φ and all $n_{i,j}$ neurons' outputs are set to the value of the label $l_p \in L = \{l_1, \ldots, l_{|L|}\}$, associated with the face of the person p ($|L|$ is equal to the number of known persons). All neurons are then trained to output this label with this input image. This procedure is repeated for all images I_x of the person p and, likewise, for all persons in the training data set. During testing, each face image I_y is also transformed, filtered, and copied to the VG-RAM WNN's input Φ. After that, all neurons' outputs are computed and the number of neurons outputting each label is counted by a function $f(I_y, l_p)$ for all $l_p \in L = \{l_1, \ldots, l_{|L|}\}$. The network's output is the label with the largest count.

3.2 Feature-Based Architecture

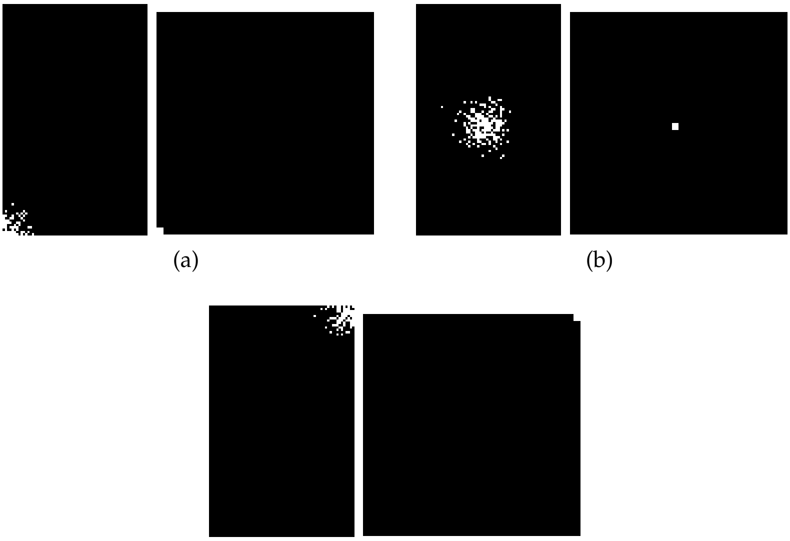
As the holistic architecture, the feature-based architecture has a single bidimensional array of $m \times n$ VG-RAM WNN neurons, N, where each neuron, $n_{i,j}$, has a set of synapses, $W = \{w_1, \ldots, w_{|W|}\}$, which are connected to the network's bidimensional input, Φ, of $u \times v$ inputs. The synaptic interconnection pattern of each neuron $n_{i,j}$, $\Omega_{i,j,\sigma}(W)$, is, however, different (Figure 6). In the feature-based architecture, $\Omega_{i,j,\sigma}(W)$ follows a bidimensional Normal distribution with variance σ^2 centered at φ_{μ_k,μ_l}, where $\mu_k = \frac{i.u}{m}$ and $\mu_l = \frac{j.v}{n}$; i.e., the coordinates k and l of the elements of Φ to which $n_{i,j}$ connects via W follow the probability density functions:

$$\omega_{\mu_k,\sigma^2}(k) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(k-\mu_k)^2}{2\sigma^2}} \tag{1}$$

$$\omega_{\mu_l,\sigma^2}(l) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(l-\mu_l)^2}{2\sigma^2}} \tag{2}$$

where σ is a parameter of the architecture. This synaptic interconnection pattern mimics that observed in many classes of biological neurons (Kandel et al., 2000), and is also created when the network is built and does not change afterwards.

A comparison between Figure 3 and Figure 6 illustrates the difference between the interconnection patterns of the holistic and feature-based architectures. In the feature-based architecture (Figure 6), each neuron $n_{i,j}$ monitors a region of the input Φ and, therefore, specializes in the face features that are mapped to that region. On the other hand, each neuron $n_{i,j}$ of the holistic architecture monitors the whole face (Figure 3).



(a)

(b)

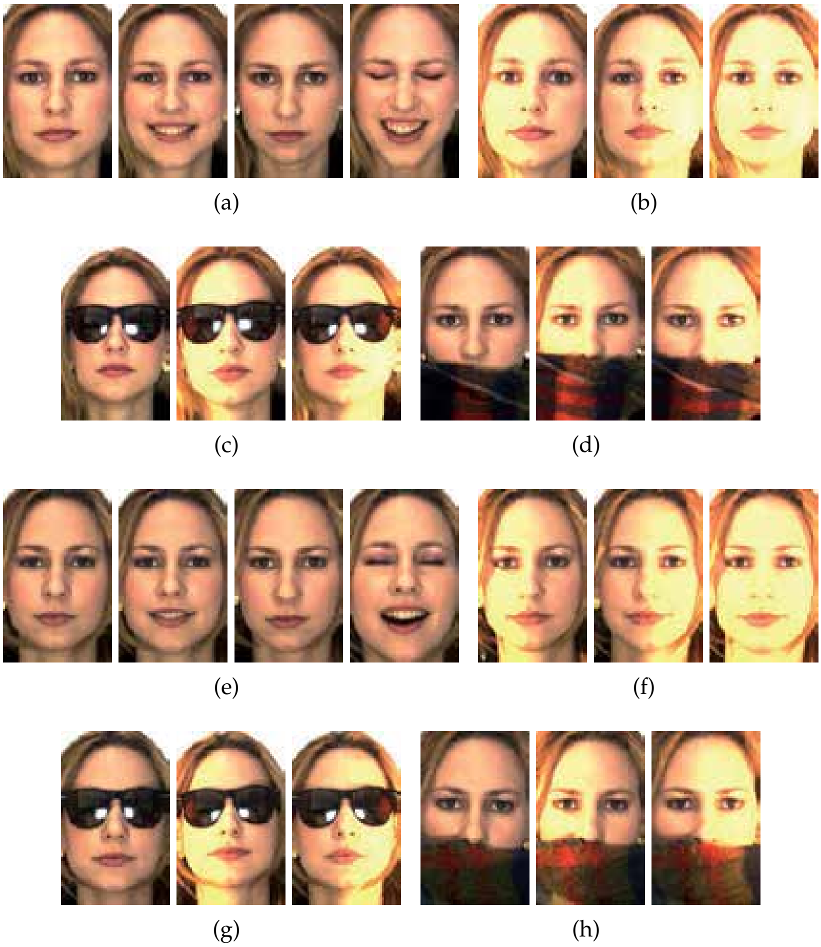
(c)

Fig. 6. The synaptic interconnection pattern of the feature-based architecture. (a) Left, input Φ: in white, the elements $\varphi_{k,l}$ of the input Φ that are connected to neuron $n_{1,1}$ of N via $\Omega_{1,1,\sigma}(W)$. Right, neuron array N: in white, the neuron $n_{1,1}$ of N. (b) Left: in white, the elements $\varphi_{k,l}$ of Φ connected to $n_{\frac{m}{2},\frac{n}{2}}$ via $\Omega_{\frac{m}{2},\frac{n}{2},\sigma}(W)$. Right: in white, the neuron $n_{\frac{m}{2},\frac{n}{2}}$ of N. (c) Left: in white, the elements of Φ connected to $n_{m,n}$ via $\Omega_{n,n,\sigma}(W)$. Right: in white, the neuron $n_{m,n}$.

As in the holistic architecture, in the feature-based architecture each neuron's synapse, w_t, forms a minchinton cell with the next, w_{t+1}, and, before training or testing, the input face images, I, are transformed and only then copied to the VG-RAM WNN input Φ. Training and testing are performed the same way as in the holistic architecture.

4. Experimental Evaluation

We used the AR Face Database (Martinez & Benavente, 1998) and the Extended Yale Face Database B (Georghiades et al., 2001; Lee et al., 2005) to evaluate the performance of VG-RAM WNN on face recognition. The AR Face Database contains over 4,000 color images corresponding to 135 people's faces (76 men and 59 women). Images feature frontal view faces with different facial expressions, illumination conditions, and occlusions (sun glasses and scarf). Its 768×576 pixels pictures were taken under strictly controlled conditions, but no restrictions on wear (clothes, glasses, etc.), make-up, hair style, etc. were imposed to participants. Each person participated in two sessions, separated by two weeks (14 days) time. On each of those sessions, thirteen images of each person were taken: four with variations of expression (neutral expression, smile, anger and scream—first session Figure 7(a) and second session Figure 7(e)), three with different illumination conditions (left light on, right light on, all side lights on—Figure 7(b) and Figure 7(f)), three wearing large sun glasses in different illumination conditions (Figure 7(c) and Figure 7(g)), and three wearing scarf in different illumination conditions (Figure 7(d) and Figure 7(h)).



(a) (b)

(c) (d)

(e) (f)

(g) (h)

Fig. 7. Rotated, scaled and croped images of one person of the AR Face Database.

The Extended Yale Face Database B consists of 2,414 frontal-face images of 38 individuals (Georghiades et al., 2001). The manually cropped and 192 × 168 sized face images were captured under 64 different laboratory-controlled lighting conditions (Lee et al., 2005). Figure 7 shows the 64 face images of one person of the Extended Yale Face Database B.

We used these face databases to perform two sets of experiments. In the first set, we used the AR Face Database to compare the performance of VG-RAM WNN with that of: (i) a holistic method based on principal component analysis (PCA) (Turk & Pentland, 1991); (ii) feature-based methods based on non-negative matrix factorization (NMF) (Lee & Seung, 1999), local non-negative matrix factorization (LNMF) (Li et al., 2001), and line edge maps (LEM) (Gao & Leung, 2002); and (iii) hybrid methods based on weighted eigenspace representation (WER) (Martinez, 2002) and attributed relational graph (ARG) matching (Park et al., 2005). In the second set of experiments, we compared the performance of VG-RAM WNN with that of Sparse Representation-based Classification (SRC) (Wright et al., 2009) using both

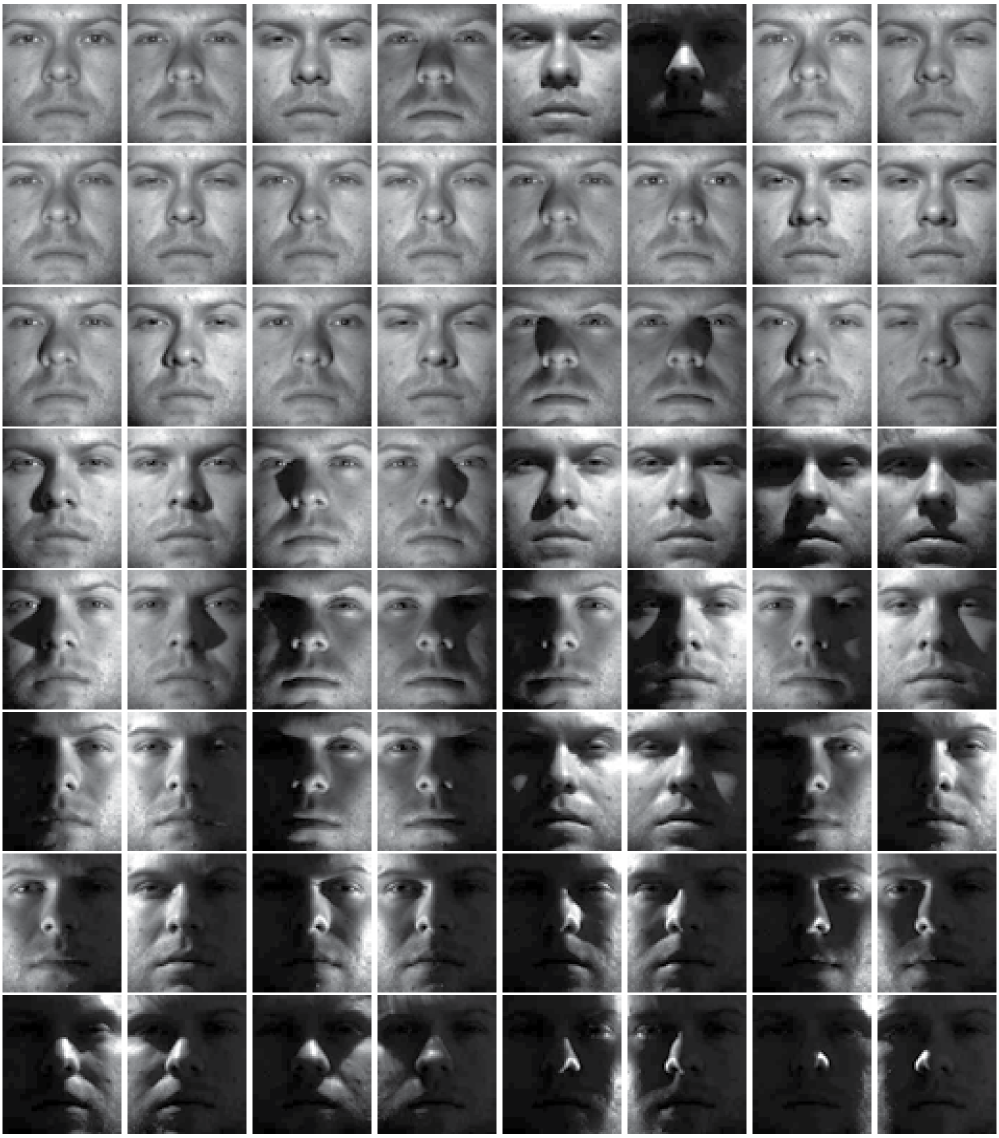


Fig. 8. Images of one person of the Extended Yale Face Database B.

the AR Face Database and the Extended Yale Face Database B. In the following sections we present these experiments.

4.1 VG-RAM WNN versus PCA, NMF, LNMF, LEM, WER, and ARG

In order to allow the comparison of VG-RAM WNN with that of PCA, NMF, LNMF, LEM, WER, and ARG, we used an experimental setup equivalent to that of Park et al. (2005). Park et al. (2005) proposed the ARG approach and compared it with PCA, NMF, LNMF, LEM, and

WER. By using an equivalent experimental setup, we can compare VG-RAM WNN with his approach and the others mentioned.

As Park et al. (2005), we used only the following subset of image types of the AR Face Database: neutral expression, smile, anger, scream, left light on, right light on, all side lights on, wearing sun glasses (with a single illumination condition), and wearing scarf (with a single illumination condition). These can be divided into four groups: (i) normal (neutral expression); (ii) under expression variation (smile, anger, scream); (iii) under illumination changes (left light on, right light on, all side lights on); and (iv) with occlusion (wearing sun glasses, wearing scarf). We took these types of 768 × 576 sized face image of all persons in the AR Face Database and rotated, scaled, cropped and filtered them to obtain 128 × 200 face images that we used as the input Φ of our VG-RAM WNN. Figure 9 shows a set of transformed images of one subject of the AR Face Database (rotated, scaled and cropped to 128 × 200 sized images).

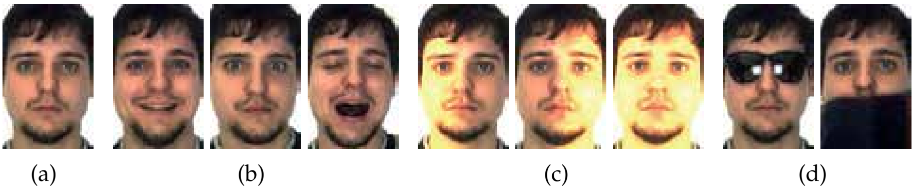


Fig. 9. The AR face database: (a) normal (neutral expression); (b) under expression variation (smile, anger, scream); (c) under illumination changes (left light on, right light on, all side lights on); and (d) with occlusion (wearing sun glasses, wearing scarf).

We randomly selected 50 people from the database to tune the parameters of the VG-RAM WNN architectures (25 men and 25 women). We used one normal face image of each person to train (50 images), and the smile, anger, wearing sun glasses, and wearing scarf to evaluate the architectures (200 images) while varying their parameters. Below, we describe the experiments we performed to tune the parameters of the architectures.

4.1.1 Holistic Architecture Parameter Tunning

The holistic architecture has three parameters: (i) the number of neurons, $m \times n$; (ii) the number of synapses per neuron, $|W|$; and (iii) the size of the network input, $u \times v$. We tested networks with: $m \times n$ equal to 2×2, 4×4, 16×16, 32×32 and 64×64; number of synapses per neuron equal to 32, 64, 128 and 256; and $u \times v$ equal to 128×200 (we did not vary $u \times v$ to reduce the parameter search space). Figure 10(a) presents the results of the experiments we carried out to tune the parameters of the holistic architecture.

As Figure 10(a) shows, the performance, i.e., the percentage of correctly recognized faces (recognition rate) of the holistic architecture grows with the number of neurons and synapses per neuron; however, as these numbers increase, the gains in performance decrease forming a plateau towards the maximum performance. The simplest configuration in the plateau has around 16×16 neurons and 64 synapses.

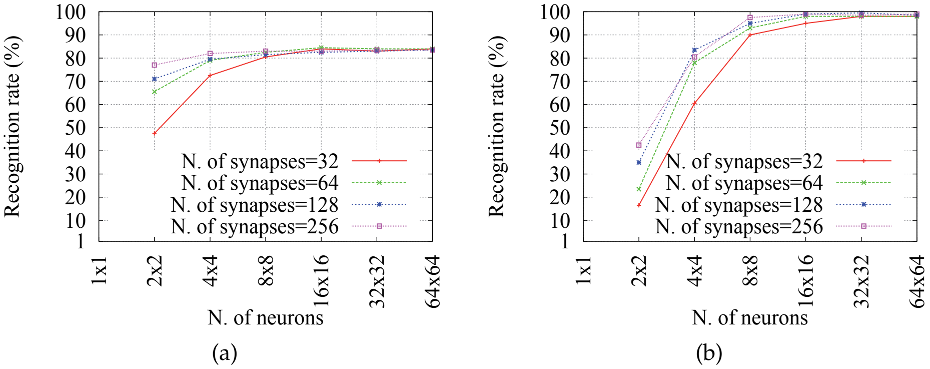


Fig. 10. Performance tunning: (a) holistic architecture and (b) feature-based architecture.

4.1.2 Feature-Based Architecture Parameter Tunning

The feature-based architecture has four parameters: (i) the number of neurons; (ii) the number of synapses per neuron; (iii) the size of the network input; and (iv) σ (see Section 3.2). We tested networks with: $m \times n$ equal to 2×2, 4×4, 16×16, 32×32 and 64×64; number of synapses per neuron equal to 32, 64, 128 and 256; $u \times v$ equal to 128×200; and σ equal to 10 (we did not vary $u \times v$ and σ to reduce the parameter search space).

Figure 10(b) presents the results of the experiments we conducted to tune the parameters of the feature-based architecture. As Figure 10(b) shows, the performance of the feature-based architecture also grows with the number of neurons and synapses per neuron, and again reaches a plateau at about 32×32 neurons and 128 synapses. However, it is important to note that, in this case, the plateau is very close to a recognition rate of 100%—the best performing configuration achieved a recognition rate of 99.5%.

4.1.3 Performance Comparison

We compared the performances of the holistic and feature-based VG-RAM WNN architectures with that of PCA, NMF, LNMF, LEM, WER, and ARG approaches. For that, we took the best VG-RAM WNN architectures configurations (holistic: 16×16 neurons and 64 synapses per neuron; feature-based: 32×32 neurons and 128 synapses per neuron), trained them with the normal face image of all people in the database (135 images), and tested them with the remaining face image categories of Figure 9 of all people in the database (135 images of each face image category). Table 1 summarizes this comparison, showing one technique on each line, grouped by type, and the corresponding performance for each face image category on each column.

As the results in Table 1 show, the VG-RAM WNN holistic (VWH) architecture outperformed all holistic and feature-based techniques examined (except the VG-RAM WNN feature-based architecture - VWF) in all face image categories. It also performed better than the hybrid techniques, except for the categories with occlusion and single side illumination. That was expected, since occlusions and single side illumination compromise eventual similarities between the input patterns learned by the VWH neurons and those collected by its synapses from a partially occluded or illuminated face. Nevertheless, it is important to note the overall

Table 1. Comparison of the performance on the AR Face Database of the holistic (VWH) and feature-based (VWF) VG-RAM WNN architectures with that of: (i) PCA: principal component analysis (Turk & Pentland, 1991) (results obtained from (Park et al., 2005)); NMF: non-negative matrix factorization (Lee & Seung, 1999) (results from (Park et al., 2005)); LNMF: local non-negative matrix factorization (Li et al., 2001) (results from (Park et al., 2005)); LEM: line edge maps (Gao & Leung, 2002) (results from (Gao & Leung, 2002) with only 112 people of the AR Face Database); WER: weighted eigenspace representation (Martinez, 2002) (results from (Martinez, 2002) with only 50 people of the AR Face Database); and ARG: attributed relational graph matching (Park et al., 2005) (results from (Park et al., 2005)).

Type	Technique	Category							
		Smile	Anger	Scream	Glasses	Scarf	Left light	Right light	All side lights
HOL[a]	PCA	94.1%	79.3%	44.4%	32.9%	2.2%	7.4%	7.4%	2.2%
	VWH	**98.5%**	**97.8%**	**91.1%**	**66.7%**	**25.2%**	**97.8%**	**95.6%**	**95.6%**
FBA[b]	NMF	68.1%	50.4%	18.5%	3.7%	0.7%	N/A[d]	N/A	N/A
	LNMF	94.8%	76.3%	44.4%	18.5%	9.6%	N/A	N/A	N/A
	LEM	78.6%	92.9%	31.3%	N/A	N/A	92.9%	91.1%	74.1%
	VWF	**99.3%**	**99.3%**	**93.3%**	**85.2%**	**98.5%**	**99.3%**	**98.5%**	**99.3%**
HYB[c]	WER	84.0%	94.0%	32.0%	80.0%	82.0%	N/A	N/A	N/A
	ARG	97.8%	96.3%	66.7%	80.7%	85.2%	98.5%	96.3%	91.1%

[a]HOL: holistic techniques. [b]FBA: feature-based techniques. [c]HYB: hybrid techniques. [d]N/A: not available.

performance achieved by VWH, which is better than that of several other relevant techniques from literature.

As Table 1 also shows, the VG-RAM WNN feature-based (VWF) architecture performed better than all other techniques examined in all categories and, in many cases, by a large margin.

4.2 VG-RAM WNN versus SRC

The central point of the SRC approach is that there are a lot of redundancy of information in face images, i.e., for the purpose of face recognition, the dimensionality of face images is typically too large because they are frequently oversampled. One can appreciate this by reasoning about the fact that images can be compacted; i.e., images sampled (or either oversampled) with 8 megapixels—which would result in a file with 8 megapixels × 3 bytes, one byte for each color, that is, 3 × 8 megabytes—can typically be compacted into a file of about one megabyte.

In the work of Wright et al. (2009), they studied several methods to reduce the dimensionality of the information extracted from face images for being used as input of the face recognition systems' classifiers. Therefore, in order to allow the comparison of VG-RAM WNN with that of SRC, we used an experimental setup equivalent to that of Wright et al. (2009).

We compared the best VG-RAM WNN performing architecture (feature-based) with that of SRC. For the experiments with the AR Face Database, as Wright et al. (2009) did, we rotated, scaled, and cropped the 768 × 576 sized face images to 120 × 165 sized images and, after that, downsampled the images at a ratio of 1/6. The downsampled images have size 20 × 27, or 540 dimensions, which was the same used by Wright et al. (2009). After downsampling the images, we rescaled them back to 120 × 165 to use as the input Φ of the VG-RAM WNN (about the same size we used in the previous experiments, 128 × 200). Note that this does not add any information to the images; we did that in order to not change the parameters we have found in the tuning of the VG-RAM WNN feature-based architecture. After rescaling the images, we filtered them with a Gaussian filter to smooth out artifacts produced by the transformations. Again, it is important to note that this does not add any information to the images; it is required only for the proper work of our VG-RAM WNN. Figure 11(a) shows a transformed face image (rotated, scaled, and cropped), the downsampled version of this image, and the filtered version of this same image.

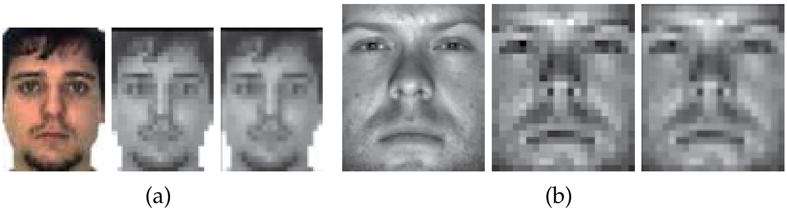

(a) (b)

Fig. 11. Face image subsampling. (a) AR Face Database. (b) Extended Yale Face Database B.

For the experiments with the Extended Yale Face Database B, also used by Wright et al. (2009), only scaling and filtering were necessary, since this database includes versions of the images already properly cropped. The image sizes in the Extended Yale Face Database B are different from those in the AR Face Database; we downsampled the 168 × 192 sized face images to 21 × 24 and, as we did with the AR Face Database, we rescaled these images back to 168 × 192 and filtered them. Figure 11(b) shows an original face image, the downsampled version of this image, and the filtered version of this same image.

In the case of the AR Face Database, following the same procedure of Wright et al. (2009), we randomly selected 50 men and 50 women. For each person, fourteen images with variations of expression and different illumination conditions were selected; the seven images from the first session were used for training and the other seven from the second session for testing. In the case of the Extended Yale Face Database B, for each person, we randomly selected half of the face images for training (i.e., about 32 images for each of the 38 people) and the other half for testing.

Table 2 summarizes the comparison of the performance of the VG-RAM WNN feature-based architecture with that of SRC, following the same format of the Table 1. The kind of face recognition system of Wright et al. (2009) is a holistic type.

As the results in Table 2 show, VG-RAM WNN feature-based (VWF) architecture outperformed SRC for both databases and, in the case of the AR Face Database, by a large margin. The VWF superior performance, shown in both the Table 1 and Table 2, is the result of two factors. First, each VWF (or VWH) synapse collects the result of a comparison between two pixels, executed by its corresponding minchinton cell. Our best VWF has 128 synapses per neuron and 32×32 neurons. Therefore, during test, 131072 (128×32×32) such comparisons are executed on an input face image and the results are checked against equivalent results learned from training images. This amount of pixel comparisons allows not only high discrimination

Table 2. Comparison of the performance on the AR Face Database and the Extended Yale Face Database B of the VG-RAM WNN feature-based (VWF) architecture with that of the Sparse Representation-based Classification (SRC). Results for SRC were obtained from Wright et al. (2009).

Type	Technique	Database	
		AR Face Database	Extended Yale Face Database B
HOL[a]	SRC (Random)	94.70%	98.1%
FBA[b]	**VWF**	**98.86%**	**99.34%**

[a]HOL: holistic techniques. [b]FBA: feature-based techniques.

capability but also generalization. Second, thanks to the characteristics of the VWF architecture, i.e., its synaptic interconnection pattern, each VWF neuron monitors a specific region of the face only, which reduces the overall impact of occlusions and varying illumination conditions on recognition performance.

5. Conclusions and Future Work

In this work, we presented an experimental evaluation of the performance of Virtual Generalizing Random Access Memory Weightless Neural Networks (VG-RAM WNN Aleksander (1998)) on face recognition. We presented two VG-RAM WNN face recognition architectures, one holistic and the other feature-based, and examined its performance with two well known face database: the AR Face Database and the Extended Yale Face Database B. The AR Face Database is challenging for face recognition systems because it has images with different facial expressions, occlusions, and varying illumination conditions. The best performing architecture (feature-based) showed robustness in all image conditions and better performance than many other techniques from literature, even when trained with a single sample per person.

In future works, we will examine the performance of VG-RAM WNN with other databases and use it to tackle other problems associated with face recognition systems, such as face detection, face alignment, face recognition in video, etc.

6. Acknowledgments

We would like to thank *Receita Federal do Brasil*, *Conselho Nacional de Desenvolvimento Científico e Tecnológico*—CNPq-Brasil (grants 308207/2004-1, 471898/2004-0, 620165/2006-5), *Fundação Espírito Santense de Tecnologia* — FAPES-Brasil (grant 37711393/2007), and *Financiadora de Estudos e Projetos*—FINEP-Brasil (grants CT-INFRA-PRO-UFES/2005, CT-INFRA-PRO-UFES/2006) for their support to this research work.

7. References

Aleksander, I. (1966). Self-adaptive universal logic circuits, *IEE Electronic Letters* **2**(8): 231–232.

Aleksander, I. (1998). *RAM-Based Neural Networks*, World Scientific, chapter From WISARD to MAGNUS: a Family of Weightless Virtual Neural Machines, pp. 18–30.

Belhumeur, P. N., Hespanha, J. P. & Kriegman, D. J. (1997). Eigenfaces vs. fisherfaces: Recognition using class specific linear projection, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(7): 711–720.

Candès, E. & Wakin, M. (2008). An introduction to compressive sampling, *IEEE Signal Processing Magazine* **2**(25): 21–30.

Etemad, K. & Chellappa, R. (1997). Discriminant analysis for recognition of human face images, *Journal of the Optical Society of America A* **14**(8): 1724–1733.

Gao, Y. & Leung, M. K. (2002). Face recognition using line edge map, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(6): 764–779.

Georghiades, A. S., Belhumeur, P. N. & Kriegman, D. J. (2001). From few to many: Illumination cone models for face recognition under variable lighting and pose, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23**(6): 643–660.

Huang, J., Heisele, B. & Blanz, V. (2003). Component based face recognition with 3d morphable models, *Proceedings of the International Conference on Audio- and Video-Based Person Authentication*, pp. 27–34.

Kandel, E. R., Schwartz, J. H. & Jessell, T. M. (2000). *Principles of Neural Science*, 4th edn, Prentice-Hall International Inc.

Lanitis, A., Taylor, C. J. & Cootes, T. F. (1995). Automatic face identification system using flexible appearance models, *Image Vision Computing* **13**(5): 393–401.

Lee, D. D. & Seung, S. H. (1999). Learning the parts of objects by non-negative matrix factorization, *Nature* **401**: 788–791.

Lee, K. C., Ho, J. & Kriegman, D. (2005). Acquiring linear subspaces for face recognition under variable lighting, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**(5): 684–698.

Li, S. Z., Hou, X. W., Zhang, H. & Cheng, Q. (2001). Learning spatially localized, parts-based representation, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 207–212.

Ludermir, T. B., Carvalho, A. C. P. L. F., Braga, A. P. & Souto, M. D. (1999). Weightless neural models: a review of current and past works, *Neural Computing Surveys* **2**: 41–61.

Martinez, A. & Benavente, R. (1998). The AR face database, *Technical Report #24*, Computer Vision Center (CVC), Universitat Autònoma de Barcelona.

Martinez, A. M. (2002). Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(6): 748–763.

Mitchell, R. J., Bishop, J. M., Box, S. K. & Hawker, J. F. (1998). *RAM-Based Neural Networks*, World Scientific, chapter Comparison of Some Methods for Processing Grey Level Data in Weightless Networks, pp. 61–70.

Park, B.-G., Lee, K.-M. & Lee, S.-U. (2005). Face recognition using face-arg matching, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**(12): 1982–1988.

Turk, M. & Pentland, A. (1991). Eigenfaces for recognition, *Journal of Cognitive Neuroscience* **3**: 71–86.

Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S. & Ma, Y. (2009). Robust face recognition via sparse representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **31**(2): 210–227.

Zhao, W.-Y., Chellappa, R., Phillips, P. J. & Rosenfeld, A. (2003). Face recognition: A literature survey, *ACM Computing Surveys* **35**(4): 399–458.

11

Illumination Processing in Face Recognition

Yongping Li, Chao Wang and Xinyu Ao
Shanghai Institute of Applied Physics, Chinese Academy of Sciences
China

1. Introduction

Driven by the demanding of public security, face recognition has emerged as a viable solution and achieved comparable accuracies to fingerprint system under controlled lightning environment. In recent years, with wide installing of camera in open area, the automatic face recognition in watch-list application is facing a serious problem. Under the open environment, lightning changes is unpredictable, and the performance of face recognition degrades seriously.

Illumination processing is a necessary step for face recognition to be useful in the uncontrolled environment. NIST has started a test called FRGC to boost the research in improving the performance under changing illumination. In this chapter, we will focus on the research effort made in this direction and the influence on face recognition caused by illumination.

First of all, we will discuss the quest on the image formation mechanism under various illumination situations, and the corresponding mathematical modelling. The Lambertian lighting model, bilinear illuminating model and some recent model are reviewed. Secondly, under different state of face, like various head pose and different facial expression, how illumination influences the recognition result, where the different pose and illuminating will be examined carefully. Thirdly, the current methods researcher employ to counter the change of illumination to maintain good performance on face recognition are assessed briefly. The processing technique in video and how it will improve face recognition on video, where Wang's (Wang & Li, 2009) work will be discussed to give an example on the related advancement in the fourth part. And finally, the current state-of-art of illumination processing and its future trends will be discussed.

2. The formation of camera imaging and its difference from the human visual system

With the camera invented in 1814 by Joseph N, recording of human face began its new era. Since we do not need to hire a painter to draw our figures, as the nobles did in the middle age. And the machine recorded our image as it is, if the camera is in good condition.

Currently, the imaging system is mostly to be digital format. The central part is CCD (charge-coupled device) or CMOS (complimentary metal-oxide semiconductor). The CCD/CMOS operates just like the human eyes. Both CCD and CMOS image sensors operate in the same manner -- they have to convert light into electrons. One simplified way to think about the sensor used in a digital camera is to think of it as having a 2-D array of thousands or millions of tiny solar cells, each of which transforms the light from one small portion of the image into electrons. The next step is to read the value (accumulated charge) of each cell in the image. In a CCD device, the charge is actually transported across the chip and read at one corner of the array. An analog-to-digital converter turns each pixel's value into a digital value. And the value is mapping to the pixel value in the memory, thus forming the given object image. Although they shared lots of similarity as human eyes, however, the impression is different. One of the advantage of human visual system is the human eye could view color constantly regardless of the luminance value in the surrounding. People with normal visual capabilities could recall the leave of tree is always green either in the morning, at the noon, or in the dust of sunset. Color constancy is subjective constancy, it remains relatively constant under normal variation situation. This phenomena was explained by N. Daw (Conway & Livingstone, 2006) using the Double-opponent cells, later E. land developed retinex theory to explain it (Am. Sci., 1963). However, for the CCD/CMOS, the formed color of the leave of the tree is related to the surrounding luminance value greatly. Thus, the difference between them is the reason that there should be some difference in the face recognition between human and machine. Machine could not take it for granted the appearance has some ignorance of its surrounding luminance value.

Human gets the perception of objects from the radiance reflected by the objects. Usually, the reflection from most objects is scattered reflection. Unlike reflected by smooth surface, the ray is deflected in random directions by irregularities in the propagation medium.

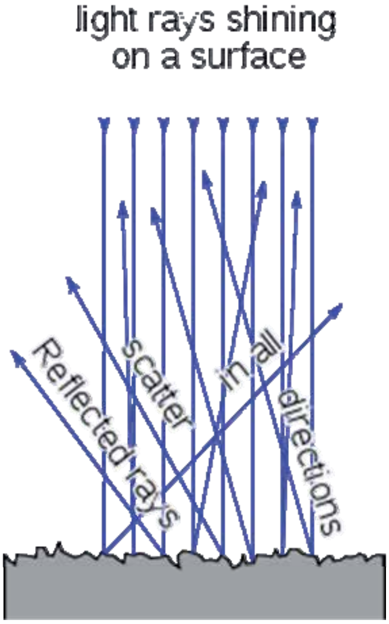


Fig. 1. Diagram of diffuse reflection (taken from the wikipedia.org)

If it is captured by eyes of human being, then perception could be fulfilled. Illumination independent image representation is an important research topic for face recognition. The face images were recorded under tightly controlled condition, where different pose, various distance, and different facial expression were presented. The edge maps, Gabor-filtered images, and the derivative of gray image were tried, but none of them could achieve the goal to be illumination independent, and none of these works provided a good enough framework to overcome the influence of various lighting condition.

3. Models for illumination

To overcome the problem, some mathematical models describing the reflectance of object in computer graphics were utilized to recover the facial image under various lighting condition. Image of a human face is the projection of its three-dimensional head on a plane, the important factors influencing the image representation is the irradiance. In computer graphics, Lambertian surface (Angel, 2003) is used to model the object surface's irradiance. The surface is called Lambertian surface if light falling on it is scattered such that the apparent brightness of the surface to an observer is the same regardless of the observer's angle of view. It could be modeled mathematically as in the following equation (1), where $I(x, y)$ is the image irradiance, ρ is the surface reflectance of the object, $n(x, y)$ is the surface normal vector of object surface, and s is the incidence ray.

$$I(x, y) = \rho(x, y)n(x, y)^T \cdot s \tag{1}$$

The Lambertian surface luminance could be called to be isotropic technically. Recently, Shashua and Riklin-Raviv (Shashua & Riklin-Raviv, 2001) proposed a method to extract the object's surface reflectance as an illumination invariant description. The method is called quotient image, which is extracted from several sample image of the object. The quotient image is defined as shown in equation 2, using the quotient image, it could recover image under some different lighting condition. It is reported outperformed the PCA. However, it works in very limited situation.

$$Q_y = \frac{I_y}{I_a} = \frac{\rho_y(x,y)n^T s}{\rho_a(x,y)n^T s} = \frac{\rho_y(x,y)}{\rho_a(x,y)} \tag{2}$$

Basri and Jacobs (Basri & Jacobs, 2003) illustrated that the illumination cone of a convex Lambertian surface could be represented by a nine-dimensional linear subspaces. In some limited environment, it could achieve some good performance. Further, Gross et al. (Gross et al., 2002) proposed a similar method called Eigen light-fields. This method claimed to only have one gallery and one probe image to estimate the light-field of the subject head, there is none further requirement on the subject pose and illumination value. And the authors declared that the performance of the proposed method on the CMU PIE database (Sim et al., 2002) is much better than that of other related algorithm.

The assumption of Lambertian model requires perfect situation, E. Nicodemus (Nicodemus, 1965) put forward a theory called BRDF (bidirectional reflectance distribution function) later. The BRDF is a four-dimensional function that defines how light is reflected at an opaque surface. The function takes an incoming light direction ω_i, and outgoing direction ω_o, both defined with respect to the surface normal n, and returns the ratio of reflected radiance exiting along ω_o to the irradiance incident on the surface from direction ω_i, Note that each direction ω is itself parameterized by azimuth angle φ and zenith angle θ, therefore the BRDF as a whole is 4-dimensional. BRDF is used in the field of modelling the reflectance on an opaque surface. These parameters could be illustrated in Fig. 2.

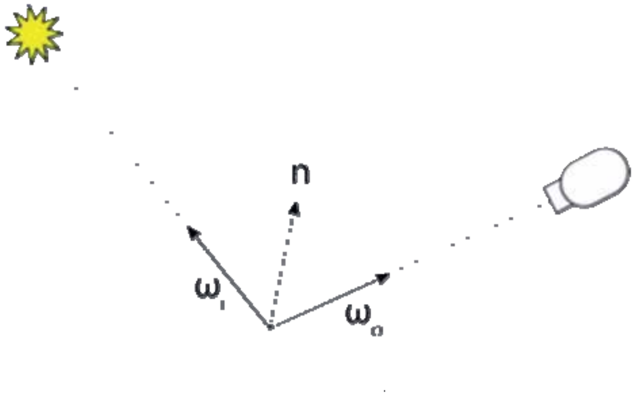


Fig. 2. Diagram showing BRDF, ω_i points toward the light source. ω_o points toward the viewer (camera). n is the surface normal

BRDF model is extensively used in the rendering artificial illuminating effects in computer graphics. To counter the effect of illumination variation, we could artificial render the different lighting situation by using this model. Comparing with Lambertain model, BRDF is of 4 dimensions, the complexity of related computation process is very large. Also, inverting the rendering situation is an ill posed problem, the equation must try some assumptions in serial to solve this problem. Thus, the efforts to employ BRDF model to attack the illumination is not successful currently.

The above models are general approaches to illumination invariant presentation; they have no requirement on the content of the image. However, recently years, there is lots of work towards to make human face image independent of illuminance, and it will be discussed thoroughly in the next section.

4. Current Approaches of Illumination Processing in Face Recognition

Many papers have been published to study on illumination processing in face recognition in the recent years. By now, these approaches can be divided into two categories: passive approaches and active approaches (Zou et al., 2007, a).

4.1 Passive Approaches

The idea of passive approaches: attempt to overcome illumination variation problem from images or video sequences in which face appearance has been altered due to environmental illumination change. Furthermore, this category can be subdivided into three classes at least, described as follows.

4.1.1 Photometric Normalization

Illumination variation can be removed: the input face images can be normalized to some state where comparisons are more reliable.

Mauricio and Roberto (Villegas & Paredes, 2005) divided photometric normalization algorithms into two types: global normalization methods and local normalization methods. The former type includes gamma intensity correction, histogram equalization, histogram matching and normal distribution. The latter includes local histogram equalization, local histogram matching and local normal distribution. Each method was tested on the same face databases: the Yale B (Georghiades et al., 2000) and the extended Yale B (Georghiades et al., 2001) face database. The results showed that local normal distribution achieves the most consistent result. Short. et al. (Short et al., 2004) compared five classic photometric normalization methods: a method based on principal component analysis, multiscale retinex (Rahman et al., 1997), homomorphic filtering, a method using isotropic smoothing to estimate the luminance function and one using anisotropic smoothing (Gross & Brajovic, 2003). The methods were tested extensively across the Yale B, XM2VTS (Messer et al., 1999) and BANCA (Kittler et al., 2000) face databases using numerous protocols. The results showed that the anisotropic method yields the best performance across all three databases. Some of photometric normalization algorithms are illuminated in detail as follows.

4.1.1.1 Histogram Equalization

Histogram equalization (HE) is a classic method. It is commonly used to make an image with a uniform histogram, which is considered to produce an optimal global contrast in the image. However, HE may make an image under uneven illumination turn to be more uneven.

S.M. Pizer and E.P. Amburn (Pizer & Amburn, 1987) proposed adaptive histogram equalization (AHE). It computes the histogram of a local image region centered at a given pixel to determine the mapped value for that pixel; this can achieve a local contrast enhancement. However, the enhancement often leads to noise amplification in "flat" regions, and "ring" artifacts at strong edges. In addition, this technique is computationally intensive.

Xudong Xie and Kin-Man Lam (Xie & Lam, 2005) proposed another local histogram equalization method, which is called block-based histogram equalization (BHE). The face image can be divided into several small blocks according to the positions of eyebrows, eyes, nose and mouth. Each block is processed by HE. In order to avoid the discontinuity between adjacent blocks, they are overlapped by half with each other. BHE is simple so that the computation required of BHE is much lower than that of AHE. The noise produced by BHE is also very little.

4.1.1.2 Gamma Intensity Correction

Shan et al. (Shan et al., 2003) proposed Gamma Intensity Correction (GIC) for illumination normalisation. The gamma transform of an image is a pixel transform by:

$$G(x,y) = I(x,y)^{1/\gamma} \tag{3}$$

where $G(x,y)$ is the output image; $I(x,y)$ is the input image; γ is the Gamma coefficient. With the value γ varying, the output image is darker or brighter. In GIC, the image $G(x,y)$ is transformed as to best match a canonically illuminated image $I_C(x,y)$. To find the best optimal γ, the value should be subject to:

$$\gamma = \arg\min_{\gamma^*} \sum_{x,y} \left[I(x,y)^{1/\gamma^*} - I_C(x,y)\right]^2 \tag{4}$$

4.1.1.3 LogAbout

To solve illumination problem, Liu et al. (Liu et al., 2001) proposed the LogAbout method which is an improved logarithmic transformations as the following equation:

$$g(x,y) = a + \frac{\ln(f(x,y)+1)}{b\ln c} \tag{5}$$

where $g(x,y)$ is the output image; $f(x,y)$ is the input image; a, b and c are parameters which control the location and shape of the logarithmic distribution.

Logarithmic transformations enhance low gray levels and compress the high ones. They are useful for non-uniform illumination distribution and shadowed images. However, they are not effective for high bright images.

4.1.1.4 Sub-Image Homomorphic Filtering

In Sub-Image Homomorphic filtering method (Delac et al., 2006), the original image is split vertically in two halves, generating two sub-images from the original one (see the upper part of Fig. 3). Afterwards, a Homomorphic Filtering is applied in each sub-image and the resultant sub-images are combined to form the whole image. The filtering is subject to the illumination reflectance model as follows:

$$I(x,y) = R(x,y) \cdot L(x,y) \tag{6}$$

where $I(x,y)$ is the intensity of the image; $R(x,y)$ is the reflectance function, which is the intrinsic property of the face; $L(x,y)$ is the luminance function.

Based on the assumption that the illumination varies slowly across different locations of the image and the local reflectance changes quickly across different locations, a high-pass filtering can be performed on the logarithm of the image $I(x,y)$ to reduce the luminance part, which is the low frequency component of the image, and amplify the reflectance part, which corresponds to the high frequency component.

Similarly, the original image can also be divided horizontally (see the lower part of Fig. 3), and the same procedure is applied. But the high pass filter can be different. At last, the two resultant images are grouped together in order to obtain the output image.

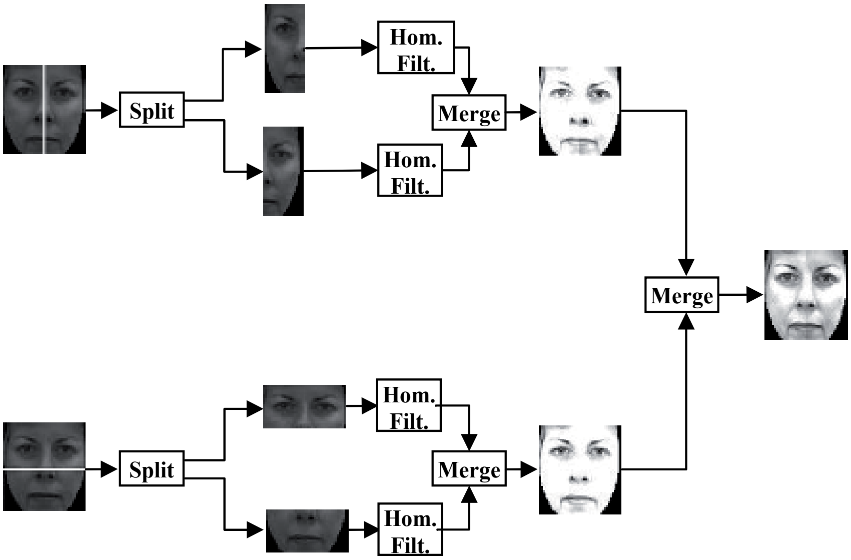


Fig. 3. Sub-image Homomorphic filtering

4.1.2 Illumination Variation Modeling

Some papers attempt to model the variation caused by changes in illumination, so as to generate a template that encompasses all possible environmental changes. The modeling of faces under varying illumination can be based on a statistical model or a physical model. For statistical model, no assumption concerning the surface property is needed. Statistical analysis techniques, such as PCA and LDA, are applied to the training set which contains faces under different illuminations to achieve a subspace which covers the variation of possible illumination. For physical model, the model of the process of image formation is based on the assumption of certain object surface reflectance properties, such as Lambertian reflectance (Basri & Jacobs, 2003). Here we also introduce some classic algorithms on both aspects.

4.1.2.1 Illumination Cone

Belhumeur and Kriegman (Belhumeur & Kriegman, 1998) proposed a property of images called the illumination cone. This cone (a convex polyhedral cone in IRn and with a dimension equal to the number of surface normals) can be used to generate and recognize images with novel illumination conditions.

This illumination cone can be constructed from as few as three images of the surface, each under illumination from an unknown point source. The original concept of the illumination cone is based on two major assumptions: a) the surface of objects has Lambertian reflectance functions; b) the object's surface is convex in shape.

Every object has its own illumination cone, the entirety of which is a set of images of the object under all possible lighting conditions, and each point on the cone is an image with a unique configuration of illumination conditions. The set of n-pixel images of any object seen under all possible lighting conditions is a convex cone in IRn.

Georghiades et al. (Georghiades et al., 1998; Georghiades et al., 2001) have used the illumination cone to further show that, using a small number of training images, the shape and albedo of an object can be reconstructed and that this reconstruction can serve as a model for recognition or generation of novel images in various illuminations. The illumination cone models the complete set of images of an object with Lambertian reflectance under an arbitrary combination of point light sources at infinity. So for a fixed pose, an image can be generated at any position on the cone which is a superposition of the training data (see Fig. 4).

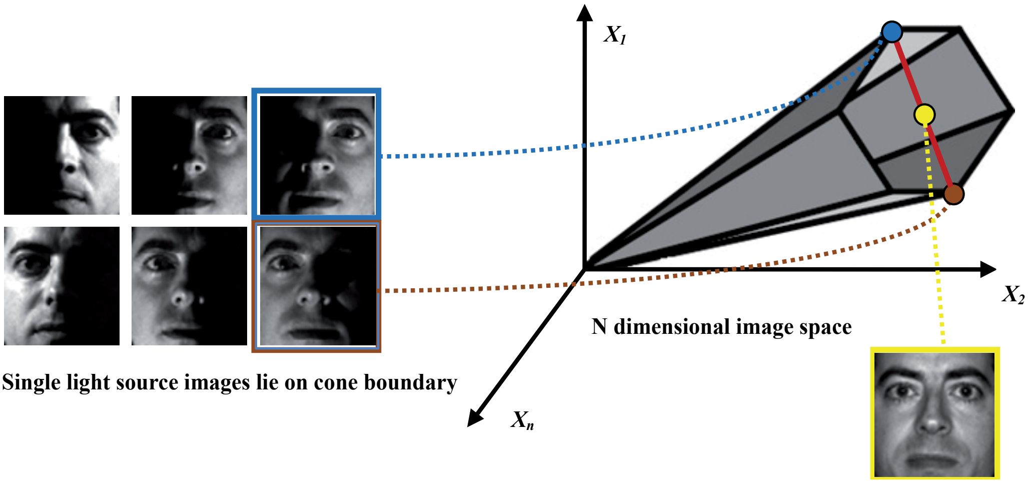

X_1
X_2
X_n
N dimensional image space
Single light source images lie on cone boundary

Fig. 4. An example of the generation of novel data from an illumination curve

4.1.2.2 3D Linear Subspace

Belhumeur et al. (Belhumeur et al., 1997) presented 3D linear subspace method for illumination invariant face recognition, which is a variant of the photometric alignment method. In this linear subspace method, three or more images of the same face under different lighting are used to construct a 3D basis for the linear subspace. The recognition proceeds by comparing the distance between the test image and each linear subspace of the faces belonging to each identity.

Batur and Hayes (Batur & Hayes, 2001) proposed a segmented linear subspace model to generalize the 3D linear subspace model so that it is robust to shadows. Each image in the training set is segmented into regions that have similar surface normals by K-Mean clustering, then for each region a linear subspace is estimated. Any estimation only relies on a specific region, so it is not influenced by the regions in shadow.

Due to the complexity of illumination cone, Batur and Hayes (Batur & Hayes, 2004) proposed a segmented linear subspace model to approximate the cone. The segmentation is based on the fact that the success of low dimensional linear subspace approximations of the illumination cone increases if the directions of the surface normals get close to each other.

The face image pixels are clustered according to the angles between their normals and apply the linear subspace approximation to each of these clusters separately. They also presented a way of finding the segmentation by running a simple K-means algorithm on a few training images, without ever requiring to obtain a 3D model for the face.

4.1.2.3 Spherical Harmonics

Ravi Ramamoorthi and Pat Hanrahan (Ramamoorthi & Hanrahan, 2001) presented spherical harmonics method. Basri and Jacobs (Basri & Jacobs, 2003) showed that, a low-dimensional linear subspace can approximate the set of images of a convex Lambertian object obtained under a wide variety of lighting conditions which can be represented by Spherical Harmonics.

Zhang and Samaras (Zhang & Samaras, 2004) combined the strengths of Morphable models to capture the variability of 3D face shape and a spherical harmonic representation for the illumination. The 3D face is reconstructed from one training sample under arbitrary illumination conditions. With the spherical harmonics illumination representation, the illumination coefficients and texture information can be estimated. Furthermore, in another paper (Zhang & Samaras, 2006), 3D shape information is neglected.

4.1.3 Illumination Invariant Features

Many papers attempt to find some face feature which is insensitive to the change in illumination. With the feature, the varying illumination on face cannot influence the recognition result. In other words, we can eliminate the illumination factor from the face image. The best way is to separate the illumination information from the identity information clearly. Here some algorithms are listed as follows.

4.1.3.1 Edge-based Image

Gao and Leung (Gao & Leung, 2002) proposed the line edge map to represent the face image. The edge pixels are grouped into line segments, and a revised Hausdorff Distance is designed to measure the similarity between two line segments. In the HMM-based face recognition algorithms, 2D discrete cosine transform (DCT) is often used for generating feature vectors. For eliminating the varying illumination influence, Suzuki and Shibata (Suzuki & Shibata, 2006) presented a directional edge-based feature called averaged principal-edge distribution (APED) to replace the DCT feature. APED feature is generated from the spatial distributions of the four directional edges (horizontal, $+45^o$, vertical, and -45^o).

4.1.3.2 Gradient-based Image

Given two images I and J of some plannar Lambertian object taken under the same viewpoint, their gradient-based image ∇I and ∇J must be parallel at every pixel where they are difined. Probabilistically, the distribution of pixel values under varying illumination may be random, but the distribution of image gradients is not.

Chen et al. (Chen et al., Chen) showed that the probability distribution of the image gradient is a function of the surface geometry and reflectance, which are the intrinsic properties of the face. The direction of image gradient is revealed to be insensitive to illumination change. S.

Samsung (Samsung, 2005) presented integral normalized gradient image for face recognition. The gradient is normalized with a smoothed version of input image and then the result is integrated into a new greyscale image. To avoid unwanted smoothing effects on step edge region, anisotropic diffusion method is applied.

4.1.3.3 Wavelet-based Image

Gomez-Moreno et al. (Gomez-Moreno et al., 2001) presented an efficient way to extract the illumination from the images by exploring only the low frequencies into them jointly with the use of the illumination model from the homomorphic filter. The low frequencies where the illumination information exists can be gained by the discrete wavelet transform. In another point of view, Du and Ward (Du & Ward, 2005) performed illumination normalization in the wavelet domain. Histogram equalization is applied to low-low sub-band image of the wavelet decomposition, and simple amplification is performed for each element in the other 3 sub-band images to accentuate high frequency components. Uneven illumination is removed in the reconstructed image obtained by employing inverse wavelet transform on the modified 4 sub-band images.

Gudur and Asari (Gudur & Asari, 2006) proposed a Gabor wavelet based Modular PCA approach for illumination robust face recognition. In this algorithm, the face image is divided into smaller sub-images called modules and a series of Gabor wavelets at different scales and orientations. They are applied on these localized modules for feature extraction. A modified PCA approach is then applied for dimensionality reduction.

4.1.3.4 Quotient Image

Due to the varying illumination on facial appearance, the appearances can be classified into four components: diffuse reflection, specular reflection, attached shadow and cast shadow. Shashua et al. (Shashua & Riklin-Raviv, 2001) proposed quotient image (QI), which is the ratio of albedo between a face image and linear combination of basis images for each pixel. This ratio of albedo is illumination invariant. However, the QI assumes that a facial appearance includes only diffuse reflection. Wang et al. (Wang et al., 2004) proposed self quotient image (SQI) by using only single image. The SQI was obtained by using the Gaussian function as a smoothing kernel function. The SQI however is neither synthesized at the boundary between a diffuse reflection region and a shadow region, nor at the boundary between a diffuse reflection region and a specular reflection region. Determining the reflectance type of an appearance from a single image is an ill-posed problem.

Chen et al. (Chen et al., 2005) proposed total variation based quotient image (TVQI), in which light estimated by solving an optimal problem so-called total variation function. But TVQI requires complex calculation. Zhang et al. (Zhang et al., 2007) presented morphological quotient image (MQI) based on mathematical morphological theory. It uses close operation, which is a kind of morphological approach, for light estimation.

4.1.3.5 Local Binary Pattern

Local Binary Pattern (LBP) (Ojala et al., 2002) is a local feature which characterizes the intensity relationship between a pixel and its neighbors. The face image can be divided into some small facets from which LBP features can be extracted. These features are concatenated into a single feature histogram efficiently representing the face image. LBP is unaffected by any monotonic grayscale transformation in that the pixel intensity order is not changed after such a transformation. For example, Li et al. (Li et al., 2007) used LBP features to compensate for the monotonic transform, which can generate an illumination invariant face representation.

4.1.3.6 3D Morphable Model

The 3D Morphable model is based on a vector space representation of faces. In this vector space, any convex combination of shape and texture vectors of a set of examples describes a realistic human face. The shape and texture parameters of the model can be separated from the illumination information.

Blanz and Vetter (Blanz & Vetter, 2003) proposed a method based on fitting a 3D Morphable model, which can handle illumination and viewpoint variations, but they rely on manually defined landmark points to fit the 3D model to 2D intensity images.

Weyrauch et al. (Weyrauch et al., 2004) used a 3D Morphable model to generate 3D face models from three input images of each person. The 3D models are rendered under varying illumination conditions to build a large set of synthetic images. These images are then used to train a component-based face recognition system.

4.2 Active Approaches

The idea of active approaches: apply active sensing techniques to capture images or video sequences of face modalities which are invariant to environmental illumination.

Here we introduce two main classes as follows.

4.2.1 3D Information

3D face information can be acquired by active sensing devices like 3D laser scanners or stereo vision systems. It constitutes a solid basis for face recognition, which is invariant to illumination change. Illumination is extrinsic to 3D face intrinsic property. Humans are capable to recognize some person in the uncontrolled environment (including the varying illumination), precisely because they learn to deal with these variations in the real 3D world.

3D information can be represented in different ways, such as range image, curvature features, surface mesh, point set, and etc. The range image representation is the most attractive. Hesher et al. (Hesher et al., 2003) proposed range image to represent 3D face information. Range images have the advantage of capturing shape variation irrespective of illumination variability. Because the value on each point represents the depth value which does not depend on illumination.

Many surveys (Kittler et al., 2005; Bowyer et al., 2006; Abate et al., 2007) on 3D face recognition have been published. However, the challenges of 3D face recognition still exist (Kakadiaris et al., 2007): ⑴ 3D capture creates larger data files per subject which implies significant storage requirements and slower processing. The conversion of raw 3D data to efficient meta-data must thus be addressed. ⑵ A field-deployable system must be able to function fully automatically. It is therefore not acceptable to assume user intervention for locating key landmarks in a 3D facial scan. ⑶ Actual 3D capture devices have a number of drawbacks when applied to face recognition, such as artifacts, small depth of field, long acquisition time, multiple types of output, high price, and etc.

4.2.2 Infrared Spectra Information

Infrared (IR) image represents a viable alternative to visible imaging in the search for a robust and practical face recognition system.

According to astronomy division scheme, the infrared portion of the electromagnetic spectrum can be divided into three regions: near-infrared (Near-IR), mid-infrared (Mid-IR) and far-infrared (Far-IR), named for their relation to the visible spectrum. Mid-IR and Far-IR belong to Thermal-IR (see Fig. 5). These divisions are not precise. There is another more detailed division (James, 2009).

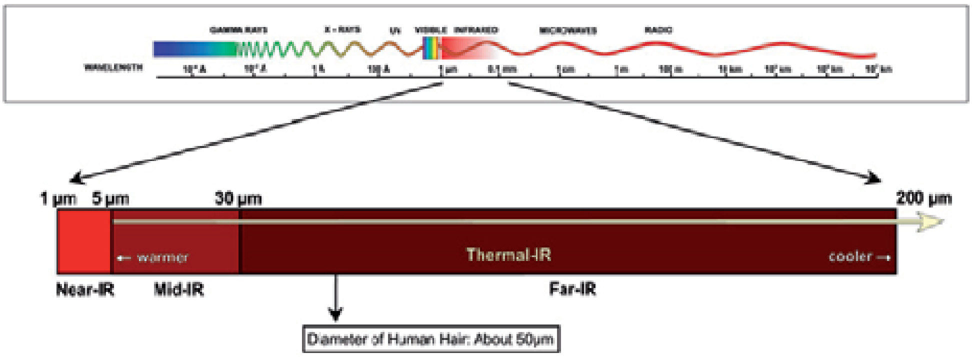


Fig. 5. Infrared as Part of the Electromagnetic Spectrum

Thermal-IR directly relates to the thermal radiation from object, which depends on the temperature of the object and emissivity of the material. For Near-IR, the image intensifiers are sensitive.

4.2.2.1 Thermal-IR

Thermal IR imagery has been suggested as an alternative source of information for detection and recognition of faces. Thermal-IR cameras can sense temperature variations in the face at a distance, and produce thermograms in the form of 2D images. The light in the thermal IR range is emitted rather than reflected. Thermal emissions from skin are an intrinsic property, independent of illumination. Therefore, the face images captured using Thermal-IR sensors will be nearly invariant to changes in ambient illumination (Kong et al., 2005).

Socolinsky and Selinger (Socolinsky & Selinger, 2004, a) presented a comparative study of face recognition performance with visible and thermal infrared imagery, emphasizing the influence of time-lapse between enrollment and probe images. They showed that the performance difference between visible and thermal face recognition in a time-lapse scenario is small. In addition, they affirmed that the fusion of visible and thermal face recognition can perform better than that using either alone. Gyaourova et al. (Gyaourova et al., 2004) proposed a method to fuse the both modalities of face recognition. Thermal face recognition is not perfect enough. For example, it is opaque to glass which can lead to facial occlusion caused by eyeglasses. Their fusion rule is based on the fact that the visible-based recognition is less sensitive to the presence or absence of eyeglasses. Socolinsky and Selinger (Socolinsky & Selinger, 2004, b) presented visible and thermal face recognition results in an operational scenario including both indoor and outdoor settings. For indoor settings under controlled illumination, visible face recognition performs better than that of thermal modality. However, Outdoor recognition performance is worse for both modalities, with a sharper degradation for visible imagery regardless of algorithm. But they showed that fused of both modalities performance outdoors is nearing the levels of indoor visible face recognition, making it an attractive option for human identification in unconstrained environments.

4.2.2.2 Near-IR

Near-IR has advantages over both visible light and Thermal-IR (Zou et al., 2005). Firstly, since it can be reflected by objects, it can serve as active illumination source, in contrast to Thermal-IR. Secondly, it is invisible, making active Near-IR illumination friendly to client. Thirdly, unlike Thermal-IR, Near-IR can easily penetrate glasses.

However, even though we use the Near-IR camera to capture face image, the environmental illumination and Near-IR illumination all exist in the face image. Hizem et al. (Hizem et al., 2006) proposed to maximize the ratio between the active Near-IR and the environmental illumination is to apply synchronized flashing imaging. But in outdoor settings, the Near-IR energy in environmental illumination is strong. Zou et al. (Zou et al., 2005) employed a light emitting diode (LED) to project Near-IR illumination, and then capture two images when the LED on and off respectively. The difference between the two images can be independent of the environment illumination. But when the face is moving, the effect is not good. To solve this problem, Zou et al. (Zou et al., 2007, b) proposed an approach based on motion compensation to remove the motion effect in the difference face images.

Li et al. (Li et al., 2007) presented a novel solution for illumination invariant face recognition based on active Near-IR for indoor, cooperative-user applications. They showed that the Near-IR face images encode intrinsic information of the face, which is subject to a monotonic transform in the gray tone. Then LBP (Ojala et al., 2002) features can be used to compensate for the monotonic transform so as to derive an illumination invariant face representation.

Above active Near-IR face recognition algorithms need that both the enrollment and probe samples are captured under Near-IR conditions. However, it is difficult to realize in some actual applications, such as passport and driver license photos. In addition, due to the distance limitation of Near-IR, many face images can only be captured only under visible lights. Chen et al. (Chen et al., 2009) proposed a novel approach, in which the enrollment samples are visual light images and probe samples are Near-IR images. Based on learning the mappings between images of the both modalities, they synthesis visual light images from Near-IR images effectively.

5. Illumination Processing in Video-based Face Recognition

Video-based face recognition is being increasingly discussed and occasionally deployed, largely as a means for combating terrorism. Unlike face recognition in still, it has its own unique features, such as temporal continuity and dependence between two neighboring frames (Zhou et al., 2003). In addition, it requires high real time in contrast to face recognition in still. Their differences are compared in Table 1.

Face Recognition in Video	Face Recognition in Still
Low resolution faces	High resolution faces
Varying illumination	Even illumination
Varying pose	Frontal pose
Varying expression	Neutral expression
Video sequences	Still image
Continuous motion	Single motion

Table 1. The comparison between face recognition in video and in still

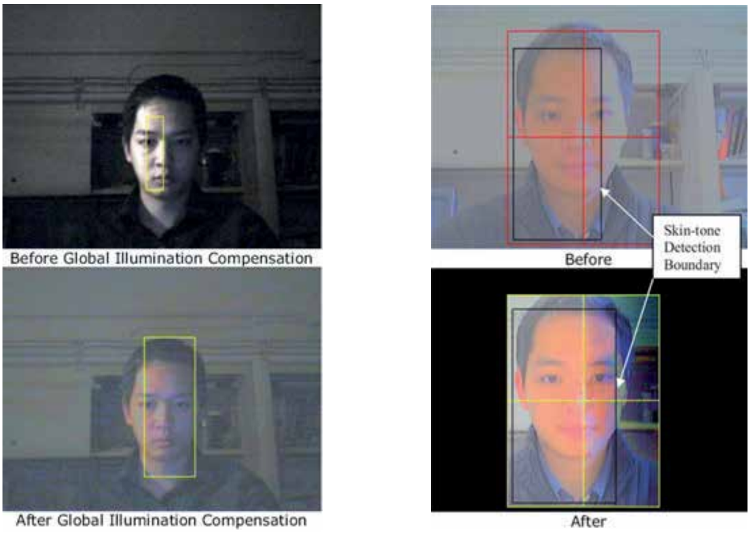
Most existing video-based face recognition systems (Gorodnichy, 2005) are realized in the following scheme: the face is first detected and then tracked over time. Only when a frame satisfying certain criteria (frontal pose, neutral expression and even illumination on face) is acquired, recognition is performed using the technique of face recognition in still. However, maybe the uneven illumination on face always exists, which lead that we cannot find a suitable time to recognize the face.

Using the same algorithms, the recognition result of video-based face recognition is not satisfying like face recognition in still. For example, the video-based face recognition systems were set up in several airports around the United States, including Logan Airport in Boston, Massachusetts; T. F. Green Airport in Providence, Rhode Island; San Francisco International Airport and Fresno Airport in California; and Palm Beach International Airport in Florida. However, the systems have never correctly identified a single face in its database of suspects, let alone resulted in any arrests (Boston Globe, 2002). Some illumination processing algorithms mentioned in Section 3 can be applied for video-based face recognition, but we encounter three main problems at least: (1) Video-based face recognition systems require higher real-time performance. Many illumination processing algorithms can achieve a very high recognition rate, but some of them take much more computational time. 3D face modeling is a classic one. Building a 3D face model is a very difficult and complicated task in the literature even though structure from motion has been studied for several decades. (2) In video sequences, the direction of illumination on face is not single. Due to the face moving or the environmental illumination changing, the illumination on face is in dynamic change. Unlike illumination processing for face recognition in still, the algorithms need more flexible. If the light source direction cannot change suddenly, the illumination condition on face only depend on the face motion. The motion and illumination are correlative. (3) In contrast to general high resolution still image, video sequences often have low resolution (less than 80 pixels between two eyes). For illumination processing, it would be more difficulty. According to the three problems, we introduced some effective algorithms for video-based face recognition.

5.1 Real-time Illumination Processing

Unlike the still image, the video sequences are displayed at a very high frequency (about 10 – 30 frames/second). So it's important to improve the real-time performance of illumination processing for video-based face recognition.

Chen and Wolf (Chen & Wolf, 2005) proposed a real-time pre-processing system to compensate illumination for face processing by using scene lighting modeling. Their system can be divided into two parts: global illumination compensation and local illumination compensation (see Fig. 6). For global illumination compensation, firstly, the input video image is divided into four areas so as to save the processing power and memory. And then the image histogram is modified to a pre-defined luminance level by a non-linear function. After that, the skin-tone detection is performed to determine the region of interest (ROI) and the lighting update information for the following local illumination compensation. The detection is a watershed between global illumination compensation and local illumination compensation. For local illumination compensation, firstly, the local lighting is estimated within the ROI determined from the previous stage. After obtaining the lighting information, a 3D face model is applied to adjust the luminance of the face candidate. The lighting information is not changed if there is no update request sent from the previous steps.


Before Global Illumination Compensation
After Global Illumination Compensation
Before
Skin-tone Detection Boundary
After

(a) Global illumination (b) Local illumination

Fig. 6. Global and local illumination compensation

Arandjelović and Cipolla (Arandjelović & Cipolla, 2009) presented a novel and general face recognition framework for efficient matching of individual face video sequences. The framework is based on simple image processing filters that compete with unprocessed greyscale input to yield a single matching score between individuals. It is shown how the discrepancy between illumination conditions between novel input and the training data set can be estimated and used to weigh the contribution of two competing representations. They found that not all the probe video sequences should be processed by the complex algorithms, such as a high-pass (HP) filter and SQI (Wang et al., 2004). If the illumination difference between training and test samples is small, the recognition rate would decrease with HP or SQI in contrast to non-normalization processing. In other words, if the illumination difference is large, normalization processing is the dominant factor and recognition performance is improved. If this notation is adopted, a dramatic performance improvement would be offered to a wide range of filters and different baseline matching algorithms, without sacrificing their online efficiency. Based on that, the goal is to implicitly learn how similar the probe and training samples illumination conditions are, to appropriately emphasize either the raw input guided face comparisons or of its filtered output.

5.2 Illumination change relating to face motion and light source

Due to the motion of faces or light sources, the illumination conditions on faces can vary over time. The single and changeless illumination processing algorithms can be unmeaning. The best way is to design an illumination compensation or normalization for the specific illumination situation. There is an implicit problem in this work: how to estimate the illumination direction. If the accuracy of the illumination estimation is low, the same to the poor face detection, the latter work would be useless. Here we will introduce several illumination estimation schemes as follows.

Huang et al. (Huang et al., 2008) presented a new method to estimate the illumination direction on face from one single image. The basic idea is to compare the reconstruction residuals between the input image and a small set of reference images under different illumination directions. In other words, the illumination orientation is regard as label information for training and recognition. The illumination estimation is to find the nearest illumination condition in the training samples for the probe. The way to estimate illumination of an input image adopted by the authors is to compute residuals for all the possible combinations of illumination conditions and the location of the minimal residual is the expectation of illumination.

Wang and Li (Wang & Li, 2008) proposed an illumination estimation approach based on plane-fit, in which environmental illumination is classified according to the illumination direction. Illumination classification can help to compensate uneven illumination with pertinence. Here the face illumination space is expressed well by nine face illumination images, as this number of images results in the lowest error rate for face recognition (Lee et al., 2005). For more accurate classification, illumination direction map, which abides by Lambert's illumination model, is generated. BHE (Xie & Lam, 2005) can weaken the light contrast in the face image, whereas HE can enhance the contrast. The difference between the face image processed by HE and the same one processed by BHE, which can reflect the light variance efficiently, generates the illumination direction map (see Fig. 7).

In order to make the direction clearer in the map, the Laplace filter and Gaussian low pass filter are also applied. In order to estimate the illumination orientation, a partial least square plane-fit is carried out on the current pixel of the illumination direction map. In actual, $I(x, y)$ is the fitted value. Suppose $f(x, y)$ is the observed value at (x, y). Then the least square between $I(x, y)$ $(I(x, y) = ax + by + c)$ and $f(x, y)$ is shown in Eq. (7).

$$S = \sum_{x,y}[f(x, y) - (ax + by + c)]^2 \tag{7}$$

note: x, y, $f(x, y)$ are known, so S is the function of a, b and c.

The illumination orientation can be defined as the value as follows:

$$\beta = \frac{\partial f(x,y)}{\partial y} \Big/ \frac{\partial f(x,y)}{\partial x}_{x=y=0} = \frac{b}{a} = \frac{\sum_{x,y} f(x,y)y}{\sum_{x,y} f(x,y)x} \tag{8}$$

where β denotes the illumination orientation on the illumination direction map.

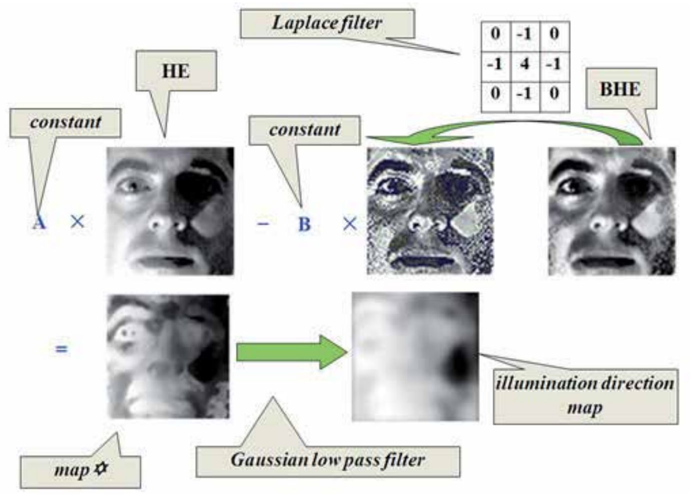


Fig. 7. Generation of illumination direction map

For the same person, the value of β is greatly different with illumination orientation variations; for different persons, the value of β is similar with the same illumination orientation. β can be calculated to make the lighting category determined.

Supposing that the light source direction is fixed, the surface of a moving face cannot change suddenly over a short time period. So the illumination varying on face can be regarded as a continuous motion. The face motion and illumination are correlative.

Basri and Jacobs (Basri & Jacobs, 2003) analytically derived a 9D spherical harmonics based on linear representation of the images produced by a Lambertian object with attached shadows. Their work can be extended from the still image to video sequences, where the video sequences can be only regarded as some separate frames, but it is inefficient. Xu and Roy-Chowdhury (Xu & Roy-Chowdhury, 2005; Xu & Roy-Chowdhury, 2007) presented a theory to characterize the interaction of face motion and illumination in generating video sequences of a 3D face. The authors showed that the set of all Lambertian reflectance functions of a moving face, illuminated by arbitrarily distant light sources, lies "close" to a bilinear subspace consisting of 9 illumination variables and 6 motion variables. The bilinear subspace formulation can be used to simultaneously estimate the motion, illumination and structure from a video sequence. The problem, how to deal with both motion and illumination, can be divided into two stages: □ the face motion is considered, and the change in its position from one time instance to the other is calculated. The change of position can be referenced as the coordinate change of the object. □ the effect of the incident illumination ray, which is projected onto face, and reflected conform to the Lambert's cosine law. For the second stage, incorporating the effect of the motion, Basri and Jacob's work is used.

However, the idea, supposing that the illumination condition is related to the face motion, has a certain limitation. If the environment illumination varies suddenly (such as a flash) or illumination source occultation, the relation between motion and illumination is not credible. All approaches conforming to the supposition would not work.

5.3 Illumination Processing for Low Resolution faces

As a novel input, it is difficult to capture a high resolution face in an arbitrary position of the video. But we can obtain a single high quality video of a person of interest, for the purpose of database enrolment. This problem is of interest in many applications, such as law enforcement. For low resolution faces, it is harder to adopt illumination processing, especially pixel-by-pixel algorithms.

However, it clearly motivates the use of super-resolution techniques in the preprocessing stages of recognition. Super-resolution concerns the problem of reconstructing high-resolution data from a single or multiple low resolution observations. Formally, the process of making a single observation can be written as the following generative model:

$$\mathrm{x} = \downarrow [\mathrm{t}(\hat{\mathrm{x}}) + \mathrm{n}] \tag{9}$$

where $\hat{\mathrm{x}}$ is the high-resolution image; $\mathrm{t}(\cdot)$ is an appearance transformation (e.g. due to illumination change, in the case of face images); n is additive noise; $\downarrow$ is the downsampling operator.

Arandjelović and Cipolla (Arandjelović & Cipolla, 2006) proposed the Generic Shape-Illumination (gSIM) algorithm. The authors showed how a photometric model of image formation can be combined with a statistical model of generic face appearance variation, learnt offline, to generalize in the presence of extreme illumination changes. gSIM performs face recognition by extracting and matching sequences of faces from unconstrained head motion videos and is robust to changes in illumination, head pose and user motion pattern. For the form of gSIM, a learnt prior is applied. The prior takes on the form of an estimate of the distribution of non-discriminative, generic, appearance changes caused by varying illumination. It means that unnecessary smoothing of person-specific, discriminative information is avoided. In the work, they make a very weak assumption on the process of image formation: the intensity of each pixel is a linear function of the albedo $\mathrm{a(j)}$ of the corresponding 3D point:

$$\mathrm{X(j)} = \mathrm{a(j)} \cdot \mathrm{s(j)} \tag{10}$$

where s is a function of illumination parameters , which is not modeled explicitly. Lambertian reflectance model is a special case.

Given two images X_1 and X_2, which are both the same person under the same pose, are of different illuminations.

$$\Delta \log \mathrm{X(j)} = \log \mathrm{s}_2\mathrm{(j)} - \log \mathrm{s}_1\mathrm{(j)} \equiv \mathrm{d}_\mathrm{s}\mathrm{(j)} \tag{11}$$

So the difference between these logarithm-transformed images is not relative to the face albedo. Under the very general assumption that the mean energy of light incident on the camera is proportional to the face albedo at the corresponding point, d_s is approximately generic i.e. not dependent on the person's identity.

However, this is not the case when dealing with real images, as spatial discretization differently affects the appearance of a face at different scales. In another paper (Arandjelović & Cipolla, 2007) of the authors, they proposed not to explicitly compute super-resolution face images from low resolution input; rather, they formulated the image formation model in such a way that the effects of illumination and spatial discretization are approximately mutually separable. Thus, they showed how the two can be learnt in two stages: (1) a generic illumination model is estimated from a small training corpus of different individuals in varying illumination. (2) a low-resolution artifact model is estimated on a person-specific basis, from an appearance manifold corresponding to a single sequence compounded with synthetically generated samples.

6. Recent State-of-art Methods of Illumination Processing in Face Recognition

How to compensate or normalize the uneven illumination on faces is still a puzzle and hot topic for face recognition researchers. There are about 50 IEEE papers on illumination processing for face recognition within past 12 months. Here we illuminated some excellent papers published on the important conferences (e.g. CVPR and BTAS) or journals (such as IEEE Transactions on Pattern Analysis and Machine Intelligence) since 2008. Many papers, which have been introduced in the former sections, are not restated.

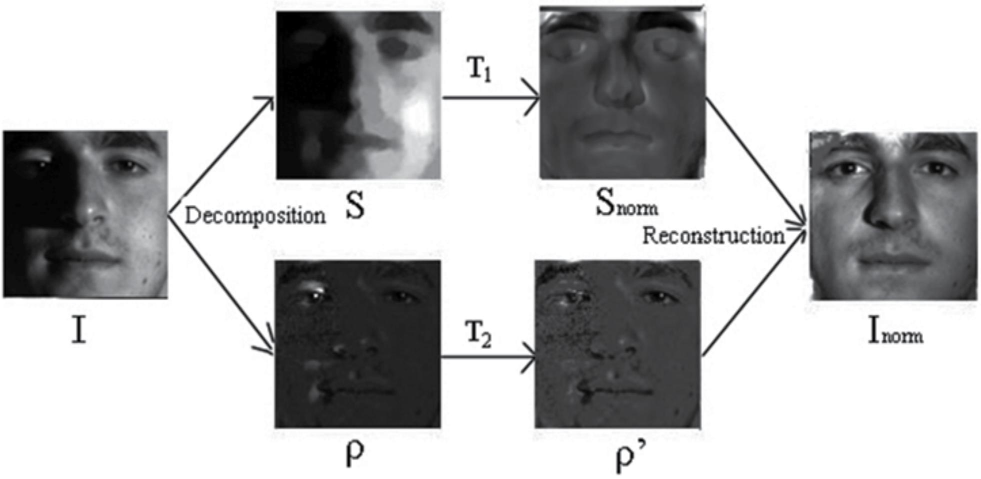


Fig. 8. Illumination Normalization Framework for Large-scale Features

Xie et al. (Xie et al., 2008) proposed a novel illumination normalization approach shown in Fig. 8. In the framework, illumination normalization whereas small-scale features (high frequency component) are only smoothed. Their framework can be divided into 3 stages: (1) Adopt an appropriate algorithm to decompose the face image into 2 parts: large-scale features and small-scale features. Methods in this category include logarithmic total variation (LTV) model (Chen et al., 2006), SQI (Wang et al., 2004) and wavelet transform (Gomez-Moreno et al., 2001) based method. However, some of the methods discard the large-scale features of face images. In this framework, the authors use LTV. (2) Eliminate the illumination information from the large-scale features by some algorithms, such as HE, BHE (Xie & Lam, 2005) and QI (Shashua & Riklin-Raviv, 2001) etc. In addition, these methods also distort the small-scale features simultaneously during the normalization process. (3) a normalized face image is generated by combination of the normalized large-scale feature image and smoothed small-scale feature image.

Holappa et al. (Holappa et al., 2008) presented an illumination processing chain and optimization method for setting its parameters so that the processing chain explicitly tailors for the specific feature extractor. This is done by stochastic optimization of the processing parameters using a simple probability value derived from intra- and inter-class differences of the extracted features as the cost function. Moreover, due to the general 3D structure of faces, illumination changes tend to cause different effects at different parts of the face image (e.g., strong shadows on either side of the nose, etc.). This is taken into account in the processing chain by making the parameters spatially variant. The processing chain and optimization method can be general, not for any specific face descriptor. To illuminate the chain and optimization method, the authors take LBP (Ojala et al., 2002) for example. LBP descriptor is relatively robust to different illumination conditions but severe changes in lighting still pose a problem. To order to solve this problem, they strive for a processing method that explicitly reduces such intra-class variations that the LBP description is sensitive to. Unlike other slowly processed interactive methods, the authors use only logarithmic transformation of pixel values and convolution of the input image region with small sized filter kernels, which makes the method very fast. The complete preprocessing and feature extraction chain is presented in Fig. 9. For the optimization method, the scheme adopted by the authors is to maximize the probability that the features calculated from an image region, that the filter to be optimized is applied to, are closer to each other in the intra class case than in the extra class case.

Face recognition in uncontrolled illumination experiences significant degradation in performance due to changes in illumination directions and skin colors. The conventional color CCD cameras are not able to distinguish changes of surface color from color shifts caused by varying illumination. However, multispectral imaging in the visible and near infrared spectra can help reduce color variations in the face due to changes in illumination source types and directions. Chang et al. (Chang et al., 2008) introduced the use of multispectral imaging and thermal infrared imaging as alternative means to conventional broadband monochrome or color imaging sensors in order to enhance the performance of face recognition in uncontrolled illumination conditions. Multispectral imaging collects reflectance information at each pixel over contiguous narrow wavelength intervals over a wide spectral range, often in the visible and Near-IR spectra. In multispectral imaging, narrowband images provide spectral signatures unique to facial skin tissue that may not be detected using broadband CCD cameras. Thermal-IR imagery is less sensitive to the variations in face appearance caused by illumination changes. Because the Thermal-IR sensors only measure the heat energy radiation, which is independent of ambient lighting. Fusion techniques have been exploited to improve face recognition performance.

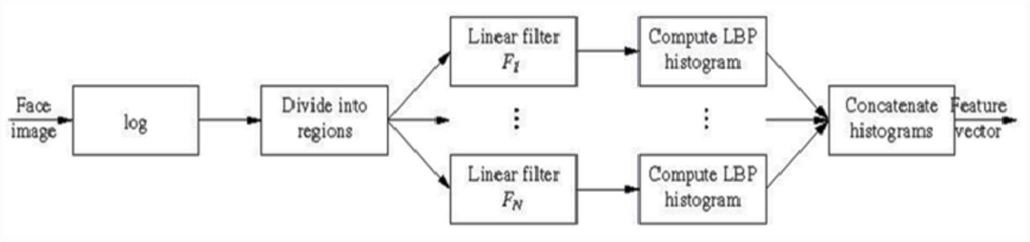
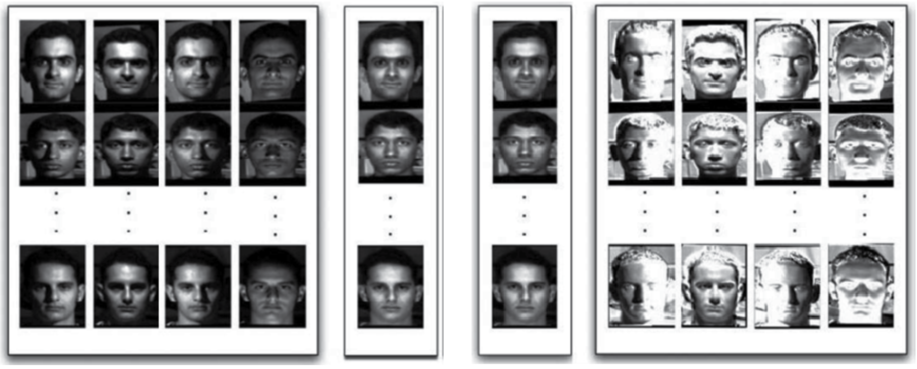


Fig. 9. Illumination Normalization Framework for Large-scale Features

The fusion of Thermal-IR and visible sensors is a popular solution to illumination-invariant face recognition (Kong et al., 2005). However, face recognition based on multispectral image fusion is relatively unexplored. The image based fusion rule can be divided into two kinds: pixel-based and feature-based fusion. The former is easy to implement but more sensitive to registration errors than the latter. Feature based fusion methods are computationally more complex but robust to registration errors.



(a) (b)

Fig. 10. (a) Example de-illumination training data for the Small Faces. Each column represents a source training set for a particular illumination model. In this case: illumination from the right; illumination from the top; illumination from the left; illumination from the bottom. The far right column is the uniformly illuminated target training data from which the derivatives are generated. (b) Example re-illumination training data for the Small Faces. The far left column is the uniformly illuminated source training data. Each remaining column represents the quotient image source training set for a particular illumination model. In this case: illumination from the right; illumination from the top; illumination from the left; illumination from the bottom.

Moore et al. (Moore et al., 2008) proposed a machine learning approach for estimating intrinsic faces and hence de-illuminating and re-illuminating faces directly in the image domain. For estimation of an intrinsic component, the local linear constraints on images are estimated in terms of derivatives using multi-scale patches of the observed images, comprising from a three-level Laplacian Pyramid. The problem of decomposing an observed face image into its intrinsic components (i.e. reflectance and albedo) is formulated as a nonlinear regression problem. For de-illuminating faces (see Fig. 10(a)), with the non-linear regression, the derivatives of the face image are estimated from a given class as it would appear with a uniform illumination. The uniformly illuminated image can then be reconstructed from these derivatives. So the de-illumination step can be regarded as an estimation problem. For re-illuminating faces (see Fig. 10(b)), it is just like an adverse stage of de-illuminating faces. The goal has changed from calculating the de-illuminated face to calculating new illuminations and the input images are de-illuminated faces. Besides these differences, the illumination estimation involves the same basic steps of estimating derivative values and integrating them to form re-illuminated images.

Most public face databases lack images with a component of rear (more than 90 degrees from frontal) illumination, either for training or testing. Wagner et al. (Wagner et al., 2009) made an experiment (see Fig. 11) which showed that training faces with the rear illumination can help to improve the face recognition. The experiment is that the girl should be identified among 20 subjects, by computing the sparse representation (Wright et al., 2009) of her input face with respect to the entire training set. The absolute sum of the coefficients associated with each subject is plotted on the right. The figure also show the faces reconstructed with each subject's training images weighted by the associated sparse coefficients. The red line corresponds to her true identity, subject 12. For the upper row of the figure, the input face is well-aligned (the white box) but only 24 frontal illuminations are used in the training for recognition. For the lower row of the figure, informative representation is obtained by using both well-aligned input face and sufficient (all 38) illuminations in the training. A conclusion can be drawn that illuminations from behind the face are also needed to sufficiently interpolate the illumination of a typical indoor (or outdoor) environment in the training. If not have, the representation will not necessarily be sparse or informative.

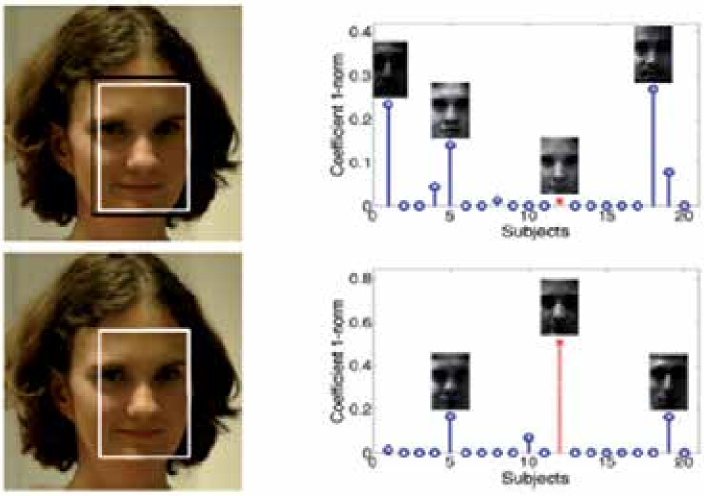
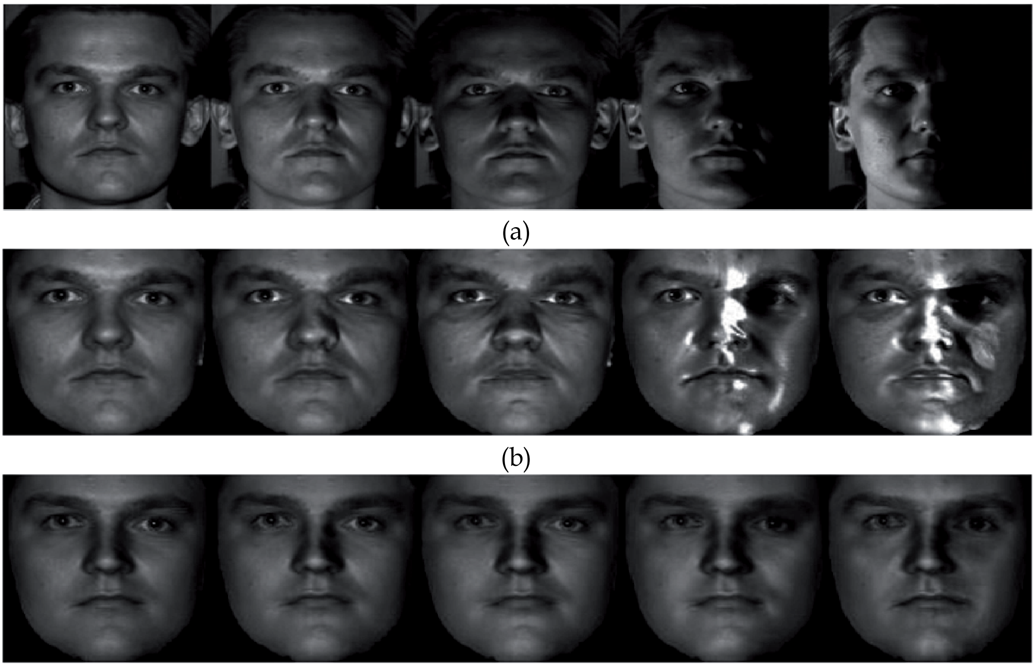


Fig. 11. Recognition Performance with and without rear illumination on faces for training

In order to solve the problem, the authors designed a training acquisition system that can illuminate the face from all directions above horizontal. The illumination system consists of four projectors that display various bright patterns onto the three white walls in the corner of a dark room. The light reflects off of the walls and illuminates the user's head indirectly. After taking the frontal illuminations, the chair is rotated by 180 degrees and then pictures are taken from the opposite direction. Having two cameras speeds the process since only the chair needs to be moved in between frontal and rear illuminations. The experiment results are satisfying. However, it is impossible to obtain samples of all target persons using the training acquisition system, such as law enforcement for terrorists.

Wang et al. (Wang et al., 2008) proposed a new method to modify the appearance of a face image by manipulating the illumination condition, even though the face geometry and albedo information is unknown. Besides, the known information is only a single face image under arbitrary illumination condition, which makes face relighting more difficult.



(a)

(b)

(c)

Fig. 12. (a) Examples of Yale B face database. (b) SHBMM method. (c) MRF-based method

According to the illumination condition on face, the authors divided their methods into two parts: face relighting under slightly uneven illumination and face relighting under extreme illumination. For the former one, they integrate spherical harmonics (Zhang & Samaras, 2004; Zhang & Samaras, 2006) into the morphable model (Blanz & Vetter, 2003; Weyrauch et al., 2004) framework by proposing a 3D spherical harmonic basis morphable model (SHBMM), which modulates the texture component with the spherical harmonic bases. So any face under arbitrary unknown lighting and pose can be simply represented by three low-dimensional vectors, i.e., shape parameters, spherical harmonic basis parameters, and illumination coefficients, which are called the SHBMM parameters. As shown in Fig. 12 (b), SHBMM can perform well for face image under slightly uneven illumination. However, the performance decreases significantly in extreme illumination. The approximation error can be large, thus making it difficult to recover albedo information. This is because the representation power of SHBMM model is inherently limited by the coupling of texture and illumination bases. In order to solve this problem, the authors presented the other sub-method – a subregion-based framework, which uses a Markov random field (MRF) to model the statistical distribution and spatial coherence of face texture. So it can be called MRF-based framework. Due to MRF, an energy minimization framework was proposed to jointly recover the lighting, the geometry (including the surface normal), and the albedo of the target face (see Fig. 12 (c)).

Gradient-based image has been proved to insensitive to illumination. Based on that, Zhang et al. (Zhang et al., 2009) proposed an illumination insensitive feature called Gradientfaces for face recognition. Gradientfaces is derived from the image gradient domain such that it can discover underlying inherent structure of face images since the gradient domain explicitly considers the relationships between neighboring pixel points. Therefore, Gradientfaces has more discriminating power than the illumination insensitive measure extracted from the pixel domain.

Given an arbitrary image $I(x, y)$ under variable illumination conditions, the ratio of y-gradient of $I(x, y)$ $(\frac{\partial I(x,y)}{\partial y})$ to $I(x, y)$ $(\frac{\partial I(x,y)}{\partial x})$ is an illumination insensitive measure. Then Gradientfaces (G) of image I can be defined as

$$G = \arctan\left(\frac{I_{y-\text{gradient}}}{I_{x-\text{gradient}}}\right),\ G \in [0, 2\pi). \tag{12}$$

where $I_{x-\text{gradient}}$ and $I_{y-\text{gradient}}$ are the gradient of image I in the x, y direction, spectively.

7. Acknowledgement

This research work is partially funded by the National High Technology Research and Development Program (863) of China under Grant # 2008AA01Z124.

8. References

Abate, A.F.; Nappi, M.; Riccio, D. & Sabatino, G. (2007). 2D and 3D Face Recognition: A Survey. Pattern Recognition Letters, 28(14), pp. 1885-1906.

American Science, (1963). The Retinex, 3, pp. 53-60, Based on William Proctor Prize address, Cleveland, Ohio.

Arandjelović, O. & Cipolla, R. (2006). Face Recognition from Video Using the Generic Shape-Illumination Manifold, In: Proc. European Conference on Computer Vision (ECCV), 4, pp. 27–40.

Arandjelović, O. & Cipolla, R. (2007). A Manifold Approach to Face Recognition from Low Quality Video across Illumination and Pose using Implicit Super-Resolution, ICCV.

Arandjelović, O. & Cipolla, R. (2009). A Methodology for Rapid Illumination-invariant Face Recognition Using Image Processing Filters, Computer Vision and Image Understanding, 113(2), pp.159-171.

Basri, R. & Jacobs, D.W. (2003). Lambertian Reflectance and Linear Subspaces, IEEE Trans. PAMI, 25(2), pp. 218-233.

Batur, A.U. & Hayes, M.H. (2001). Linear Subspaces for Illumination Robust Face Recognition, CVPR.

Batur, A.U. & Hayes, M.H. (2004). Segmented Linear Subspaces for Illumination-Robust Face Recognition, International Journal of Computer Vision, 57(1), pp. 49-66.

Belhumeur, P.N. & Kriegman, D.J. (1998). What is the Set of Images of an Object under All Possible Illumination Conditions? Int. J. Computer Vision, 28(3), pp. 245–260.

Belhumeur, P.N.; Hespanha, J.P. & Kriegman, D.J. (1997). Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection, IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7), pp. 711-720.

Blanz, V. & Vetter, T. (2003). Face Recognition Based on Fitting a 3D Morphable Model, IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(9), pp. 1063-1074.

Boston Globe, (2002). Face Recognition Fails in Boston Airport.

Bowyer, K.W.; Chang, K. & Flynn, P. (2006). A Survey of Approaches and Challenges in 3D and Multi-Modal 3D+2D Face Recognition, In CVIU.

Chang, H.; Koschan, A.; Abidi, M.; Kong, S.G. & Won, C.H. (2008). Multispectral Visible and Infrared Imaging for Face Recognition, CVPRW, pp.1-6.

Chen, C.Y. & Wolf, W. (2005). Real-time Illumination Compensation for Face Processing in Video Surveillance, AVSS, pp.293-298.

Chen, H.F.; Belhumeur, P.N. & Jacobs, D.W. (2000). In Search of Illumination Invariants, CVPR.

Chen, J.; Yi, D.; Yang, J; Zhao, G.Y.; Li, S.Z. & Pietikainen, M. (2009). Learning Mappings for Face Synthesis from Near Infrared to Visual Light Images, IEEE Computer Society Conference, pp. 156 – 163.

Chen, T.; Yin, W.T.; Zhou, X.S.; Comaniciu, D. & Huang, T.S. (2005). Illumination Normalization for Face Recognition and Uneven Background Correction Using Total Variation Based Image Models, CVPR, 2, pp.532-539.

Chen, T.; Yin, W.T.; Zhou, X.S.; Comaniciu, D. & Huang, T.S. (2006). Total Variation Models for Variable Lighting Face Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(9), pp. 1519-1524.

Conway B.R. & Livingstone M.S. (2006). Spatial and Temporal Properties of Cone Signals in Alert Macaque Primary Visual Cortex (V1), Journal of Neuroscience.

Delac, K.; Grgic, M. & Kos, T. (2006). Sub-Image Homomorphic Filtering Techniques for Improving Facial Identification under Difficult Illumination Conditions. International Conference on System, Signals and Image Processing, Budapest.

Du, S. & Ward, R. (2005). Wavelet-Based Illumination Normalization for Face Recognition, Proc. of IEEE International Conference on Image Processing, 2, pp. 954-957.

Edward, A. (2003). Interactive Computer Graphics: A Top-Down Approach Using OpenGL (Third Ed.), Addison-Wesley.

Fred, N. (1965). Directional reflectance and emissivity of an opaque surface, Applied Optics, 4 (7), pp. 767–775.

Gao, Y. & Leung, M. (2002). Face Recognition Using Line Edge Map, IEEE Trans. on PAMI.

Georghiades, A.S., Belhumeur, P.N. & Kriegman, D.J. (2001). From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose, IEEE Trans. Pattern Anal. Mach. Intelligence, 23(6), pp. 643-660.

Georghiades, A.S.; Belhumeur, P.N. & Kriegman, D.J. (2000). From Few to Many: Generative Models for Recognition under Variable Pose and Illumination, In IEEE Int. Conf. on Automatic Face and Gesture Recognition, pp. 277-284.

Georghiades, A.S.; Kriegman, D.J. & Belhumeur, P.N. (1998). Illumination Cones for Recognition under Variable Lighting: Faces, CVPR, pp.52.

Gomez-Moreno, H.; Gómez-moreno, H.; Maldonado-Bascon, S.; Lopez-Ferreras, F. & Acevedo-Rodriguez, F.J. (2001). Extracting Illumination from Images by Using the Wavelet Transform. International Conference on Image Processing, 2, pp. 265-268.

Gorodnichy, D.O. (2005). Video-based Framework for Face Recognition in Video, Second Workshop on Face Processing in Video, pp. 330-338, Victoria, BC, Canada.

Gross, R. & Brajovic, V. (2003). An Image Preprocessing Algorithm, AVBPA, pp. 10-18.

Gross, R.; Matthews, I. & Baker, S. (2002). Eigen Light-Fields and Face Recognition Across Pose, International Conference on Automatic Face and Gesture Recognition.

Gudur, N. & Asari, V. (2006). Gabor Wavelet based Modular PCA Approach for Expression and Illumination Invariant Face Recognition, AIPR.

Gyaourova, A.; Bebis, G. & Pavlidis, I. (2004). Fusion of Infrared and Visible Images for Face Recognition, European Conference on Computer Vision, 4, pp.456-468.

Hesher, C.; Srivastava, A. & Erlebacher, G. (2003). A Novel Technique for Face Recognition Using Range Imaging. Seventh International Symposium on Signal Processing and Its Applications, pp. 201- 204.

Hizem, W.; Krichen, E.; Ni, Y.; Dorizzi, B. & Garcia-Salicetti, S. (2006). Specific Sensors for Face Recognition, International Conference on Biometrics, 3832, pp. 47-54.

Holappa, J.; Ahonen, T. & Pietikainen, M. (2008). An Optimized Illumination Normalization Method for Face Recognition, 2nd IEEE International Conference on Biometrics: Theory, Applications and Systems, pp. 1-6.

Huang, X.Y.; Wang, X.W.; Gao, J.Z. & Yang, R.G. (2008). Estimating Pose and Illumination Direction for Frontal Face Synthesis, CVPRW, pp.1-6.

James, B. (2009). Unexploded Ordnance Detection and Mitigation, Springer, pp. 21-22, ISBN 9781402092527.

Kakadiaris, I.A.; Passalis, G.; Toderici, G.; Murtuza, M.N.; Lu, Y.; Karampatziakis, N. & Theoharis, T. (2007). Three-Dimensional Face Recognition in the Presence of Facial Expressions: An Annotated Deformable Model Approach, IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(4), pp. 640-649.

Kittler, J.; Hilton, A.; Hamouz, M. & Illingworth, J. (2005). 3D Assisted Face Recognition: A Survey of 3D Imaging, Modelling and Recognition Approaches. CVPRW.

Kittler, J.; Li, Y.P. & Matas, J. (2000). Face Verification Using Client Specific Fisher Faces, The Statistics of Directions, Shapes and Images, pp. 63-66.

Kong, S.G.; Heo, J.; Abidi, B.R.; Paik, J. & Abidi, M.A. (2005). Recent Advances in Visual and Infrared Face Recognition: A Review, Computer Vision and Image Understanding, 97(1), pp. 103-135.

Lee, K.; Jeffrey, H. & Kriegman, D. (2005). Acquiring Linear Subspaces for Face Recognition under Variable Lighting, IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(5), pp. 1–15.

Li, S.Z.; Chu, R.F.; Liao, S.C. & Zhang, L. (2007). Illumination Invariant Face Recognition Using Near-Infrared Images, IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(4), pp. 627-639.

Liu, H.; Gao, W.; Miao, J.; Zhao, D.; Deng, G. & Li, J. (2001). Illumination Compensation and Feedback of Illumination Feature in Face Detection, Proc. IEEE International Conferences on Info-tech and Info-net, 23, pp. 444-449.

Messer, M.; Matas, J. & Kittler, J. (1999). XM2VTSDB: The Extended M2VTS Database, AVBPA.

Moore, B.; Tappen, M. & Foroosh, H. (2008). Learning Face Appearance under Different Lighting Conditions, 2nd IEEE International Conference on Biometrics: Theory, Applications and Systems, pp. 1-8.

Ojala, T.; Pietikäinen, M. & Mäenpää, T. (2002). Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns, IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(7), pp. 971-987.

Pizer, S.M. & Amburn, E.P. (1987). Adaptive Histogram Equalization and Its Variations. Comput. Vision Graphics, Image Process, 39, pp. 355-368.

Rahman, Z.; Woodell, G. & Jobson, D. (1997). A Comparison of the Multiscale Retinex with Other Image Enhancement Techniques, Proceedings of the IS&T 50th Anniversary Conference.

Ramamoorthi, R. & Hanrahan, P. (2001). On the Relationship between Radiance and Irradiance: Determining the Illumination from Images of a Convex Lambertian Object, Journal of Optical Society of American, 18(10), pp. 2448-2459.

Samsung, S. (2005). Integral Normalized Gradient Image a Novel Illumination Insensitive Representation, CVPRW, pp.166.

Shan, S.; Gao, W.; Cao, B. & Zhao, D. (2003). Illumination Normalization for Robust Face Recognition against Varying Lighting Conditions, Proc. of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures, 17, pp. 157-164.

Shashua, A. & Riklin-Raviv, T. (2001). The Quotient Image: Class Based Re-rendering and Recognition with Varying Illuminations, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 23(2), pp. 129-139.

Short, J.; Kittler, J. & Messer, K. (2004). A Comparison of Photometric Normalisation Algorithms for Face Verification, Proc. Automatic Face and Gesture Recognition, pp. 254-259.

Sim, T.; Baker, S. & Bsat, M. (2002). The CMU Pose, Illumination, and Expression (PIE) Database, Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition.

Socolinsky, D.A. & Selinger, A. (2004), a. Thermal Face Recognition over Time, ICPR, 4, pp.187-190.

Socolinsky, D.A. & Selinger, A. (2004), b. Thermal Face Recognition in an Operational Scenario. CVPR, 2, pp.1012-1019.

Suzuki, Y. & Shibata, T. (2006). Illumination-invariant Face Identification Using Edge-based Feature Vectors in Pseudo-2D Hidden Markov Models, In Proc. EUSIPCO, Florence, Italy.

Villegas, M. & Paredes, R. (2005). Comparison of Illumination Normalization Methods for Face Recognition, In COST275, pp. 27–30, University of Hertfordshire, UK.

Wagner, A.; Wright, J.; Ganesh, A.; Zhou, Z.H. & Ma, Y. (2009). Towards A Practical Face Recognition System: Robust Registration and Illumination by Sparse Representation, IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp. 597-604..

Wang, C & Li, Y.P. (2008). An Efficient Illumination Compensation based on Plane-Fit for Face Recognition, The 10th International Conference on Control, Automation, Robotics and Vision, pp. 939-943.

Wang, C. & Li, Y.P. (2009). Combine Image Quality Fusion and Illumination Compensation for Video-based Face Recognition, Neurocomputing (To appear).

Wang, H.; Li, S.Z. & Wang, Y.S. (2004). Generalized Quotient Image, CVPR, 2, pp.498-505.

Wang, Y.; Zhang, L.; Liu, Z.C.; Hua, G.; Wen, Z.; Zhang, Z.Y. & Samaras, D. (2008). Face Re-Lighting from a Single Image under Arbitrary Unknown Lighting Conditions, IEEE Transactions on Pattern Analysis and Machine Intelligence

Weyrauch, B.; Heisele, B.; Huang, J. & Blanz, V. (2004). Component-based Face Recognition with 3D Morphable Models. CVPRW.

Wright, J.; Ganesh, A.; Yang, A. & Ma, Y. (2009). Robust Face Recognition via Sparse Representation, IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(2), pp. 210-227.

Xie, X. & Lam, K.M. (2005). Face Recognition under Varying Illumination based on a 2D Face Shape Model, Pattern Recognition, 38(2), pp. 221-230.

Xie, X.H.; Zheng, W.S.; Lai, J.H. & Yuen, P.C. (2008). Face Illumination Normalization on Large and Small Scale Features. CVPR, pp.1-8.

Xu, Y.L. & Roy-Chowdhury, A.K. (2005). Integrating the Effects of Motion, Illumination and Structure in Video Sequences, ICCV, 2, pp.1675-1682.

Xu, Y.L. & Roy-Chowdhury, A.K. (2007). Integrating Motion, Illumination, and Structure in Video Sequences with Applications in Illumination-Invariant Tracking, IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(5), pp. 793-806.

Zhang, L. & Samaras, D. (2004). Pose Invariant Face Recognition under Arbitrary Unknown Lighting using Spherical Harmonics, In Proceedings, Biometric Authentication Workshop (in conjunction with ECCV 2004).

Zhang, L. & Samaras, D. (2006). Face Recognition from a Single Training Image under Arbitrary Unknown Lighting Using Spherical Harmonics, IEEE Trans. PAMI, 29.

Zhang, T.P.; Tang, Y.Y.; Fang, B.; Shang, Z.W. & Liu, X.Y. (2009). Face Recognition under Varying Illumination Using Gradientfaces, IEEE Transactions on Image Processing.

Zhang, Y.Y.; Tian, J.; He, X.G. & Yang, X. (2007). MQI based Face Recognition under Uneven Illumination, ICB, pp. 290-298.

Zhou, S.H.; Krueger, V. & Chellappa, R. (2003). Probabilistic Recognition of Human Faces from Video, Computer Vision and Image Understanding, 91(1-2), pp. 214-245.

Zou, X.; Kittler, J. & Messer, K. (2005). Face Recognition Using Active Near-IR Illumination, Proc. British Machine Vision Conference.

Zou, X.; Kittler, J. & Messer, K. (2007), a. Illumination Invariant Face Recognition: A Survey, First IEEE International Conference on Biometrics: Theory, Applications, and Systems, pp. 1-8.

Zou, X.; Kittler, J. & Messer, K. (2007), b. Motion Compensation for Face Recognition Based on Active Differential Imaging, ICB, pp. 39-48.

12

From Gabor Magnitude to Gabor Phase Features: Tackling the Problem of Face Recognition under Severe Illumination Changes

Vitomir Štruc and Nikola Pavešić
Faculty of Electrical Engineering, University of Ljubljana
Slovenia

1. Introduction

Among the numerous biometric systems presented in the literature, face recognition systems have received a great deal of attention in recent years. The main driving force in the development of these systems can be found in the enormous potential face recognition technology has in various application domains ranging from access control, human-machine interaction and entertainment to homeland security and surveillance (Štruc et al., 2008a).

While contemporary face recognition techniques have made quite a leap in terms of performance over the last two decades, they still struggle with their performance when deployed in unconstrained and uncontrolled environments (Gross et al., 2004; Phillips et al., 2007). In such environments the external conditions present during the image acquisition stage heavily influence the appearance of a face in the acquired image and consequently affect the performance of the recognition system. It is said that face recognition techniques suffer from the so-called PIE problem, which refers to the problem of handling Pose, Illumination and Expression variations that are typically encountered in real-life operating conditions. In fact, it was emphasized by numerous researchers that the appearance of the same face can vary significantly from image to image due to changes of the PIE factors and that the variability in the images induced by the these factors can easily surpass the variability induced by the subjects' identity (Gross et al., 2004; Short et al., 2005). To cope with image variability induced by the PIE factors, face recognition systems have to utilize feature extraction techniques capable of extracting stable and discriminative features from facial images regardless of the conditions governing the acquisition procedure. We will confine ourselves in this chapter to tackling the problem of illumination changes, as it represents the PIE factor which, in our opinion, is the hardest to control when deploying a face recognition system, e.g., in access control applications.

Many feature extraction techniques, among them particularly the appearance based methods, have difficulties extracting stable features from images captured under varying illumination conditions and, hence, perform poorly when deployed in unconstrained environments. Researchers have, therefore, proposed a number of alternatives that should compensate for the illumination changes and thus ensure stable face recognition performance.

Sanderson and Paliwal (Sanderson & Paliwal, 2003), for example, proposed a feature extraction technique called DCT-mod2. The DCT-mod2 technique first applies the Discrete Cosine Transform (DCT) to sub-regions (or blocks) of facial images to extract several feature sets of DCT coefficients, and then compensates for illumination induced appearance changes by replacing the coefficients most affected by illumination variations with the corresponding vertical and horizontal delta coefficients. The authors assessed the technique on images rendered with an artificial illumination model as well as on (real-life) images captured under varying illumination conditions. Encouraging results were achieved on both image types.

Another technique was proposed by Gao and Leung in (Gao & Lung, 2002). Here, the authors argue that the so-called Line Edge Maps (LEM) represent useful face representations for both image coding and recognition and that, moreover, they also exhibit insensitiveness to illumination variations. With this technique a given face image is first processed to extract edge pixels, which are then combined into line segments that constitute the LEMs. The authors showed that their technique successfully outperformed popular feature extraction approaches on various databases.

Liu and Wechsler (Liu & Wechsler, 2002) use the Gabor wavelet (or filter) representation of face images to achieve robustness to illumination changes. Their method - the Gabor Fisher Classifier (GFC), adopts a filter bank of forty Gabor filters (featuring filters of five scales and eight orientations) to derive an augmented feature vector of Gabor magnitude features and then applies a variant of the multi-class linear discriminant analysis called the Enhanced Fisher discriminant Model (EFM) to the constructed Gabor feature vector to improve the vector's compactness.

While all of the presented techniques ensure some level of illumination invariance, Gabor wavelet based methods received the most attention due to their effectiveness and simplicity.

The feasibility of Gabor wavelet based methods for robust face recognition is not evidenced solely by the large number of papers following up on the work of Liu and Wechsler (e.g., Shen et al., 2007; Štruc & Pavešić, 2009b), but also by the results of several independent evaluations (and competitions) of face recognition technology, where the techniques utilizing Gabor wavelets regularly resulted in the best performance (Messer et al., 2006; Poh et al. 2009).

It has to be noted that the Gabor face representation as proposed by Liu and Wechsler does not represent an illumination invariant face representation, but rather exhibits robustness to illumination changes due to the properties of the deployed Gabor filter bank. Since Gabor filters represent band limited filters, the filter bank is usually constructed in such a way that it excludes the frequency bands most affected by illumination variations. Furthermore, the Gabor magnitude features that constitute the augmented Gabor feature vector are local in nature, which again adds to the illumination insensitiveness of the computed Gabor face representation.

While the existing Gabor based methods are among the most successful techniques for face recognition, they still exhibit some shortcomings, which, when properly solved, could result in an improved recognition performance. These shortcomings can be summarized into the following main points:

- most of the existing techniques rely solely on Gabor magnitude information while discarding the potentially useful Gabor phase information (e.g., Liu & Wechsler, 2002; Liu, 2006; Shen et al., 2007; Štruc & Pavešić, 2009b),
- the deployment of a large filter bank (usually comprising 40 filters) results in an inflation of the data size by a factor equaling the number of filters in the filter bank triggering the need for down-sampling strategies, which often discard information that could prove useful for the recognition task and
- Gabor magnitude features ensure only partial insensitiveness to illumination changes resulting in the necessity for additional (robust) face descriptors.

To tackle the above issues, we propose in this chapter a novel face representation called the oriented Gabor phase congruency pattern (OGPCP), which, as the name suggests, is derived from the Gabor phase congruency model presented in (Kovesi, 1999). The proposed face representation is based on the phase responses of the Gabor filter bank rather than the magnitude responses and as such offers an alternative to the established Gabor magnitude based methods. As we will show in this chapter, the feature vector constructed from the OGPCPs is more compact (i.e., less dimensional) than the traditional Gabor magnitude representation of face images and also exhibits robustness to illumination changes. Thus, it represents a novel robust face representation capable of substituting or complementing the existing Gabor magnitude based recognition techniques.

The rest of the chapter is structured as follows: In Section 2, a brief review of the Gabor filter based methods is given. In Section 3, the novel face representation, i.e., the oriented Gabor phase congruency pattern is presented and the augmented Gabor phase congruency feature vector introduced. In Section 4, the classification rule for the experiments is highlighted, while the experimental databases are described in Section 5. The feasibility of the proposed features is assessed in Section 6. The chapter concludes with some final remarks and directions for future work in Section 7.

2. Review of Gabor filters for face recognition

2.1 Gabor filter construction

Gabor filters (sometimes also called Gabor wavelets or kernels) have proven to be a powerful tool for facial feature extraction. They represent band-limited filters with an optimal localization in the spatial and frequency domains. Hence, when used for facial feature extraction, they extract multi-resolutional, spatially local features of a confined frequency band. In the spatial domain, the family of 2D Gabor filters can be defined as follows (Lades et al., 1993; Liu & Wechsler, 2002; Liu, 2006; Shen & Bai, 2006; Štruc & Pavešić, 2009b):

$$\psi_{u,v}(x,y) = \frac{f_u^2}{\pi\gamma\eta} e^{-(\frac{f_u^2}{\gamma^2}x'^2 + \frac{f_u^2}{\eta^2}y'^2)} e^{j2\pi f_u x'}, \tag{1}$$

where $x' = x\cos\theta_v + y\sin\theta_v$, $y' = -x\sin\theta_v + y\cos\theta_v$, $f_u = f_{max}/2^{(u/2)}$, and $\theta_v = v\pi/8$. Each filter represents a Gaussian kernel function modulated by a complex plane wave whose centre frequency and orientation are defined by the parameters f_u and θ_v, respectively. The parameters γ and η determine the ratio between the centre frequency and the size of the Gaussian envelope and, when set to a fixed value, these parameters ensure that Gabor filters of different scales and a given orientation behave as scaled versions of each other. It has to be noted at this point that with fixed values of the parameters γ and η the scale of the given Gabor filter is defined uniquely by its centre frequency f_u. Commonly the values of the parameters γ and η are set to $\gamma = \eta = \sqrt{2}$. The last parameter of the Gabor filters f_{max} denotes the maximum frequency of the filters and is commonly set to $f_{max} = 0.25$. When employed for facial feature extraction, researchers typically use Gabor filters with five scales and eight orientations, i.e., $u = 0, 1, \ldots, p-1$ and $v = 0, 1, \ldots, r-1$, where $p = 5$ and $r = 8$, which results in a filter bank of 40 Gabor filters (Liu, 2006; Shen et al. 2007; Štruc et al., 2008a).

It should be noted that Gabor filters represent complex filters which combine an even (cosine-type) and odd (sine-type) part (Lades et al., 1993). An example of both filter parts in 3D is shown on the left side of Fig. 1, while the real parts of the entire filter bank (commonly comprising 40 Gabor filters) are presented in 2D on the right hand side of Fig. 1.

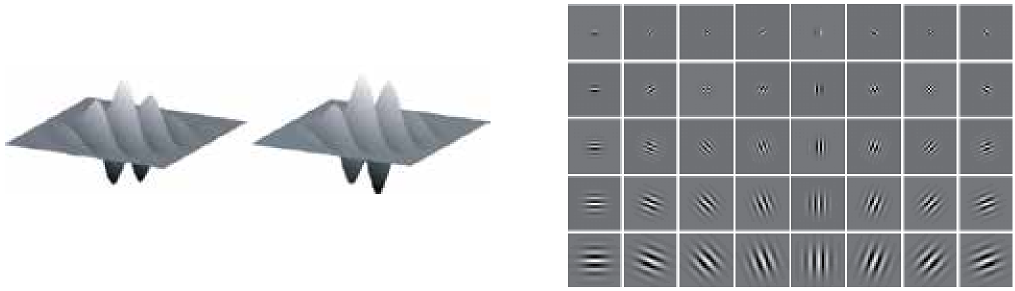


Fig. 1. Examples of Gabor filters: the real and imaginary part of a Gabor filter in 3D (left), the real part of the commonly employed Gabor filter bank of 40 Gabor filters in 2D (right)

2.2 Feature extraction with Gabor filters

Let $I(x, y)$ denote a grey-scale face image of size $a \times b$ pixels and let $\psi_{u,v}(x, y)$ represent a Gabor filter defined by its centre frequency f_u and orientation θ_v. The filtering operation or better said the feature extraction procedure can then be written as the convolution of the face image $I(x, y)$ with the Gabor wavelet (filter, kernel) $\psi_{u,v}(x, y)$, i.e. (Štruc & Pavešić, 2009b),

$$G_{u,v}(x, y) = I(x, y) * \psi_{u,v}(x, y). \qquad (2)$$

In the above expression, $G_{u,v}(x, y)$ represents the complex convolution output which can be decomposed into its real (or even) and imaginary (or odd) parts as follows:

$$E_{u,v}(x, y) = Re[G_{u,v}(x, y)] \text{ and } O_{u,v}(x, y) = Im[G_{u,v}(x, y)]. \qquad (3)$$

Based on these results, both the phase ($\phi_{u,v}(x, y)$) as well as the magnitude responses ($A_{u,v}(x, y)$) of the filter can be computed, i.e.:

$$
\begin{aligned}
A_{u,v}(x,y) &= \sqrt{E_{u,v}^2(x,y) + O_{u,v}^2(x,y)} \\
\phi_{u,v}(x,y) &= \arctan(O_{u,v}(x,y)/E_{u,v}(x,y))
\end{aligned}
\qquad (4)
$$

As already stated in the previous section, most of the techniques found in the literature discard the phase information of the filtering output and retain only the magnitude information for the Gabor face representation. An example of this information (in image form) derived from a sample face image is shown in Fig. 2.

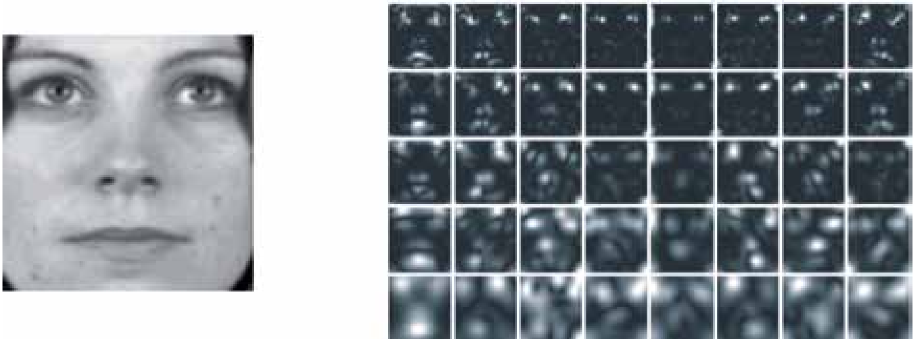


Fig. 2. An example of the Gabor magnitude output: a sample image (left), the magnitude output of the filtering operation with the entire Gabor filter bank of 40 Gabor filters (right)

2.3 The Gabor (magnitude) face representation

The first step when deriving the Gabor (magnitude) face representation of facial images is the construction of the Gabor filter bank. As we have pointed out several times in this chapter, most existing techniques adopt a filter bank comprised of 40 Gabor filters, i.e., Gabor filters with 5 scales ($u = 0, 1, \ldots, 4$) and 8 orientations ($v = 0, 1, \ldots, 7$).

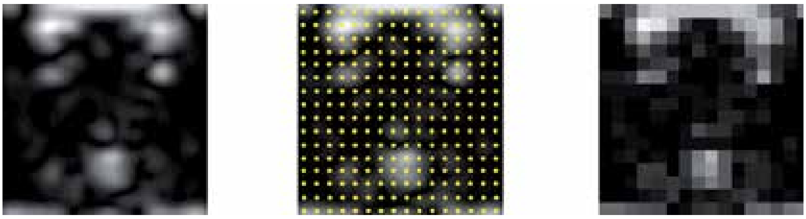


Fig. 3. Down-sampling of a magnitude filter response: an example of a magnitude response (left), an example of the magnitude response with a superimposed sampling grid (middle), a down-sampled magnitude response (right)

To obtain the Gabor (magnitude) face representation, a given face image is filtered with all 40 filters from the filter. However, even for a small image of 128 × 128 pixels, the magnitude responses of the filtering outputs comprise a pattern vector with 655360 elements, which is far too much for efficient processing and storage. To overcome this problem, down-sampling strategies are normally exploited. The down-sampling techniques reduce the dimensionality of the magnitude responses, unfortunately often at the expense of potentially useful discriminatory information. A popular down-sampling strategy is to employ a rectangular sampling grid (as shown in Fig. 3) and retain only the values under the grid's nodes. This down-sampling procedure is applied to all magnitude responses, which are then normalized to zero mean and unit variance and ultimately concatenated into the final Gabor (magnitude) face representation or, as named by Liu and Wechsler (Liu & Wechsler, 2002), into the augmented Gabor feature vector.

If we denote the down-sampled magnitude response (in column vector form) of the Gabor filter of scale u and orientation v as $\boldsymbol{g}_{u,v}$, we can define the augmented Gabor (magnitude) feature vector $\boldsymbol{x}$ as follows:

$$\boldsymbol{x} = \left(\boldsymbol{g}_{0,0}^{\boldsymbol{T}}, \boldsymbol{g}_{0,1}^{\boldsymbol{T}}, \boldsymbol{g}_{0,2}^{\boldsymbol{T}}, \dots, \boldsymbol{g}_{4,7}^{\boldsymbol{T}}\right)^{\boldsymbol{T}}. \tag{5}$$

It has to be noted that in the experiments presented in Section 6 of this chapter, we use images of 128×128 pixels and a rectangular down-sampling grid with 16 horizontal and 16 vertical lines, which corresponds to a down-sampling factor of 64. Nevertheless, even after the down-sampling, the augmented Gabor (magnitude) feature vector still resides in a very high-dimensional space (Shen et al., 2007) - in our case the dimensionality of the vectors still equals 10240. To make the processing more efficient, researchers commonly turn to so-called subspace projection techniques, e.g. (Liu, 2006; Shen et al. 2007; Štruc & Pavešić, 2009b). Two of these techniques, namely, the Principal Component Analysis (PCA) and the Linear Discriminant analysis (LDA), will also be adopted for our experiments. The description of these techniques is beyond the scope of this chapter, the reader is, however, referred to (Turk & Pentland, 1991) and (Belhumeur et al. 1997) for more information on PCA and LDA, respectively.

3. The Gabor (phase) face representation

This section introduces a novel face representation called oriented Gabor phase congruency pattern (OGPCP) and, consequently, the augmented Gabor phase congruency feature vector.

3.1 Background

Before we turn our attention to the novel representation of face images, i.e., to the oriented Gabor phase congruency pattern, let us take a closer look at why the Gabor phase information is commonly discarded when deriving the Gabor face representation.

Unlike the (Gabor) magnitude, which is known to vary slowly with the spatial position, the (Gabor) phase can take very different values even if it is sampled at image locations only a few pixels apart. This fact makes it difficult to extract stable and discriminative features from the phase responses of Eq. (2) and is the primary reason that most of the existing methods use only the (Gabor) magnitude to construct the Gabor feature vector (Zhang et al., 2007; Štruc et al., 2008a).

To the best of our knowledge, there are only a few studies in the literature that successfully derived useful features from Gabor phase responses for the task of face recognition, e.g., (Zhang et al., 2007; Bezalel & Efron, 2005; Gundimada & Asari, 2006; Gundimada et al., 2009). A common characteristic of these methods is the fact that they use features or face representations derived from the Gabor phase information rather than the "raw" phase responses themselves or combine the phase information with other face descriptors to compensate for the variability of the Gabor phase.

Zhang et al. (Zhang et al., 2007), for example, adopt local histograms of the phase responses encoded via local binary patterns (LBPs) as face image descriptors and show that over small image regions the Gabor phase patterns exhibit some kind of regularity (in terms of histograms) and, hence, contain useful information for the task of face recognition. Other authors (e.g., Bezalel & Efron, 2005; Gundimada & Asari, 2006; Gundimada et al., 2009) incorporate the Gabor phase information by adopting the 2D phase congruency model of Kovesi (Kovesi, 1999) to detect edges in a given face image and deploy the resulting “edge” image for detection of interest points that are used with other image descriptors, such as Gabor magnitude features.

The face representation proposed in this chapter is related to the work of (e.g., Bezalel & Efron, 2005; Gundimada & Asari, 2006; Gundimada et al., 2009) only as far as it uses the concept of phase congruency for encoding the Gabor phase information. However, unlike previous work on this subject it proposes a face representation that is only partially based on Kovesi’s 2D phase congruency model and employs the proposed representation for recognition rather than solely for feature selection. As will be shown in the next section, the proposed face representation exhibits several desirable properties which overcome most of the shortcomings of the existing Gabor magnitude based methods.

3.2 The oriented Gabor phase congruency patterns

The original 2D phase congruency model as proposed by Kovesi in (Kovesi, 1999) was developed with the goal of robust edge and corner detection in digital images. However, as we will show, it can (though with a few modifications) also be used to encode phase information of the Gabor filter responses in a way that is useful for the task of face recognition.

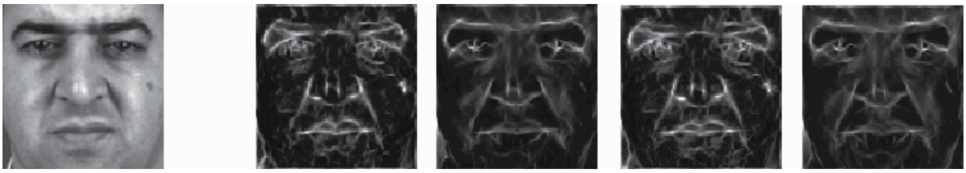


Fig. 4. Examples of phase congruency images (from left to right): the original image, the PCI for $p = 3$ and $r = 6$, the PCI for $p = 5$ and $r = 6$, the PCI for $p = 3$ and $r = 8$, the PCI for $p = 5$ and $r = 8$

Kovesi's original phase congruency model searches for points in an image where the log-Gabor filter responses over several scales and orientations are maximally in phase (Kovesi, 1999; Štruc & Pavešić, 2009a). Thus, a point in an image is of significance only if the phase responses of the log-Gabor filters over a range of scales (i.e., frequencies) display some kind of order. In the original approach, phase congruency is first computed for each of the employed filter orientations, while the results are then combined to form the final phase congruency image (PCI). Some examples of such images obtained with log-Gabor filters with p scales and r orientations are shown in Fig. 4. Note that the code used to produce the presented phase congruency images was provided by P. Kovesi and can be found at his homepage: http://www.csse.uwa.edu.au/~pk/Research/MatlabFns/index.html

While the presented approach is suitable for robust (in terms of noise, illumination variations and image contrast) edge and corner detection, its usefulness for facial feature extraction is questionable. As it was emphasized by Liu in (Liu, 2006), a desirable characteristic of any feature extraction procedure is the capability of extracting multi-orientational features. Rather than combining phase congruency information computed over several orientations, and using the result for construction of the facial feature vector, we therefore propose to compute an oriented Gabor phase congruency pattern (OGPCP) for each of the employed filter orientations and to construct an augmented (Gabor) phase congruency feature vector based on the results. Note that, differently from the original 2D phase congruency model proposed in (Kovesi, 1999), we use conventional Gabor filters as defined by Eq. (1) rather than log-Gabor filters (Štruc et al., 2008a).
Taking into account the original definition of the phase congruency, we can derive an oriented version of the phase congruency, which, when presented in image form, reveals the oriented Gabor phase congruency patterns (OGPCPs) for the v-th orientation:

$$OGPCP_v(x,y) = \frac{\sum_{u=0}^{p-1} A_{u,v}(x,y)\Delta\phi_{u,v}(x,y)}{\sum_{u=0}^{p-1}(A_{u,v}(x,y)+\epsilon)}, \tag{6}$$

where ϵ denotes a small constant that prevents division by zero and $\Delta\phi_{\mathrm{u,v}}(x,y)$ stands for the phase deviation measure of the following form:

$$\Delta\phi_{u,v}(x,y) = \cos\big(\phi_{u,v}(x,y) - \bar{\phi}_v(x,y)\big) - \big|\sin(\phi_{u,v}(x,y) - \bar{\phi}_v(x,y))\big|. \tag{7}$$

In the above expression $\phi_{u,v}(x,y)$ denotes the phase angle of the Gabor filter (with a centre frequency f_u and orientation θ_v) at the spatial location (x,y), while $\bar{\phi}_v(x,y)$ represents the mean phase angle at the v-th orientation. Several examples of the OGPCPs for a sample image are shown in Fig. 5.

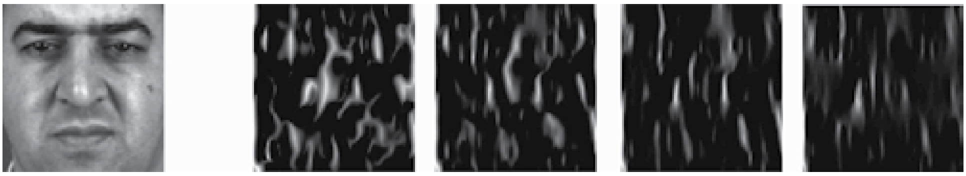


Fig. 5. Examples of OGPCPs (from left to right): the original image, the OGPCP for $\theta_v = 0°$ and $p = 2$, the OGPCP for $\theta_v = 0°$ and $p = 3$, the OGPCP for $\theta_v = 0°$ and $p = 4$, the OGPCP for $\theta_v = 0°$ and $p = 5$

Kovesi showed that the expression given in (6) can be computed directly from the filter outputs defined by (3); however, for details on computing the OGPCPs the reader should refer to the original paper (Kovesi, 1999).
It should be noted that the OGPCPs as defined by Eq. (6) represent illumination invariant (and contrast independent) face representations, since they do not depend on the overall magnitude of the filter responses. This property makes the OGPCPs a very useful image representation for face recognition.

3.3 The augmented Gabor phase congruency feature vector

The OGPCPs presented in the previous section form the foundation for the augmented Gabor phase congruency feature vector, which is computed from a given face image by the following procedure:

- for a given face image all r OGPCPs are computed for the chosen number of filter scales p (an example of all OGPCPs for a sample image with $r = 8$ and $p = 2$ is presented in Fig. 6.)
- the OGPCPs are down-sampled by a down-sampling factor ρ (in a similar manner as shown in Fig. 3),
- the down-sampled OGPCPs are normalized using the selected normalization procedure (zero mean and unit variance, histogram equalization, …), and
- the down-sampled and normalized OGPCPs in column vector form (denoted as $\boldsymbol{D}_v$) are concatenated to form the augmented Gabor phase congruency feature vector $\boldsymbol{x}$.

Formally, the augmented Gabor phase congruency feature vector is defined as follows:

$$\boldsymbol{x} = \left(\boldsymbol{D}_0^{\boldsymbol{T}}, \boldsymbol{D}_1^{\boldsymbol{T}}, \boldsymbol{D}_2^{\boldsymbol{T}}, \dots, \boldsymbol{D}_{r-1}^{\boldsymbol{T}}\right)^{\boldsymbol{T}}, \tag{8}$$

where T denotes the transform operator and $\boldsymbol{D}_v$, for $v = 0, 1, 2, \dots, r - 1$, stands for the vector derived from the OGPCP at the v-th orientation.

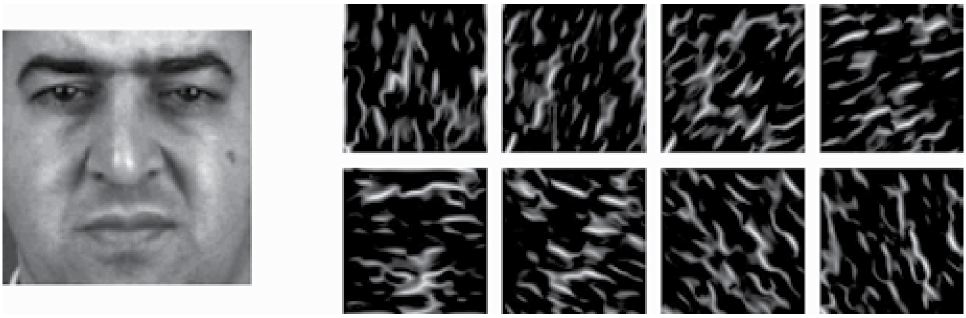


Fig. 6. An example of all OGPCPs: the original image (left), the OGPCPs (for $r = 8$) that form the foundation for construction of the augmented Gabor phase congruency feature vector

Note that in the experiments presented in Section 6 a down-sampling factor of $\rho = 16$ was used for the OGPCPs, as opposed to the Gabor magnitude responses, where a down-sampling factor of $\rho = 64$ was employed. This setup led to similar lengths of the constructed (Gabor) feature vectors of both methods and thus enabled a fair comparison of their face recognition performances. Furthermore, as the smaller down-sampling factor was used for the OGPCPs, less potentially useful information is discarded when oriented Gabor phase congruency patterns are employed for the face representation rather than the Gabor magnitude features.

As with the augmented Gabor magnitude feature vector, the Gabor phase congruency feature vector also resides in high-dimensional space and, hence, requires additional processing with, for example, subspace projection techniques to further reduce its dimensionality. In the experimental section, PCA and LDA are evaluated for this purpose.

4. Classification rule

In general, a face recognition system can operate in one of two modes: in the verification or identification mode (Štruc et al., 2008b).

When operating in the verification mode, the goal of the system is to determine the validity of the identity claim uttered by the user currently presented to the system. This is achieved by comparing the so-called "live" feature vector $\boldsymbol{y}$ extracted from the given face image of the user with the template corresponding to the claimed identity. Based on the outcome of the comparison, the identity claim is rejected or accepted. The verification procedure can formally be written as follows: given the "live" feature vector $\boldsymbol{y}$ and a claimed identity C_i associated with a user-template $\boldsymbol{y}_i$, where $i \in \{1,2,\dots,N\}$ and N stands for the number of enrolled users, determine the validity of the identity claim by classifying the pair $(\boldsymbol{y}, C_i)$ into one of two classes w_1 or w_2 (Jain et al., 2004):

$$(\boldsymbol{y}, C_i) \in \begin{cases} w_1, & if\ d(\boldsymbol{y}, \boldsymbol{y}_i) \geq \Delta \\ w_2, & otherwise \end{cases}, \tag{9}$$

where w_1 stands for the class of genuine identity claims, w_2 denotes the class of impostor identity claims, $d(\cdot,\cdot)$ denotes a function measuring the similarity of its arguments. In our case the similarity function takes the form of the cosine similarity measure, i.e.,

$$d(\boldsymbol{y}, \boldsymbol{y}_i) = \frac{\boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{y}_{\boldsymbol{i}}}{\sqrt{\boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{y}\boldsymbol{y}_{\boldsymbol{i}}^{\boldsymbol{T}}\boldsymbol{y}_{\boldsymbol{i}}}} \tag{10}$$

and Δ represents a predefined decision threshold.

In a face recognition system operating in the identification mode the problem statement is different from that presented above. In case of the identification task we are not interested whether the similarity of the "live" feature vector with a specific user template is high enough; rather, we are looking for the template in the database which best matches the "live" feature vector. This can be formalized as follows: given a "live" feature vector $\boldsymbol{y}$ and a database containing N templates $\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_N$ of the enrolled users (or identities) $C_1, C_2, \dots, C_N$, determine the most suitable identity, i.e., (Jain et al., 2004):

$$\boldsymbol{y} \in \begin{cases} C_i, & if\ \max_{\mathrm{i}} d(\boldsymbol{y}, \boldsymbol{y}_i) \geq \Delta,\ i = 1,2,\dots.,N \\ C_{N+1}, & otherwise \end{cases}, \tag{11}$$

where $d(\boldsymbol{y}, \boldsymbol{y}_i)$ again denotes the cosine similarity measure and C_{N+1} stands for the case, where no appropriate identity from the database can be assigned to the "live" feature vector $\boldsymbol{y}$. The presented expression postulates that, if the similarity of the "live" feature vector and the template associated with the i-th identity is the highest among the similarities with all user templates in the system, then the i-th identity is assigned to the "live" feature vector $\boldsymbol{y}$. It should be noted that, in the experiments presented in the experimental section, the user templates are constructed as the mean vectors of the feature vectors extracted from the enrollment face images of the users.

5. The databases and experimental configurations

This section presents the experimental databases used to determine the feasibility of the proposed augmented phase congruency feature vectors for face recognition. It commences by describing the two face databases, namely, the XM2VTS (Messer et al., 1999) and the Extended YaleB database (Georghiades et al., 2001; Lee et al., 2005), and proceeds by presenting the pre-processing procedure applied to the experimental images prior to the actual experiments.

5.1 The XM2VTS database

The XM2VTS database comprises a total of 2360 facial images that correspond to 295 distinct subjects (Messer et al., 1999). The images were recorded in controlled conditions (in front of a homogenous background, with artificial illumination, with frontal pose and a neutral facial expression, etc.), during four recording sessions and over a period of approximately five months. At each of the recording session, two recordings were made resulting in eight facial images per subject that are featured in the database. Since the time elapsed between two successive sessions was around one month, the variability in the images is mainly induced by the temporal factor. Thus, images of the same subject differ in terms of hairstyle, presence or absence of glasses, make-up and moustaches, etc. Some examples of the images from the XM2VTS database are shown in Fig. 7.

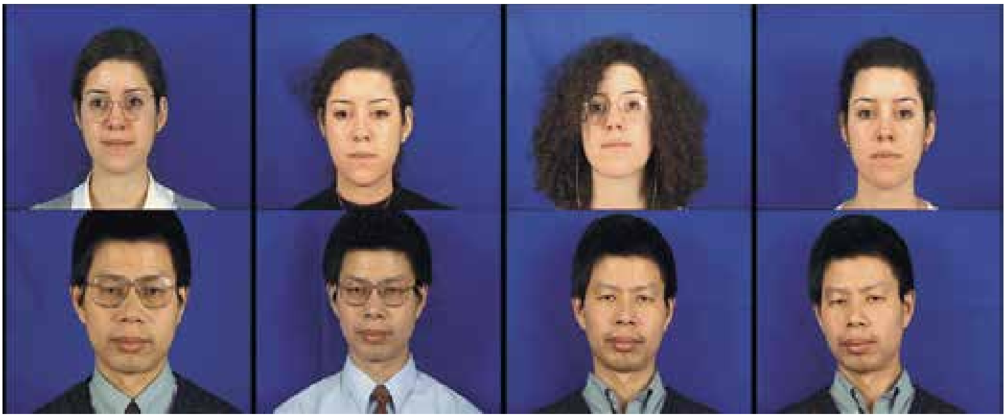


Fig. 7. Sample images from the XM2VTS database

The face verification experiments on the XM2VTS were conducted in accordance with the first configuration of the experimental protocol associated with the database, known also as the Lausanne protocol (Messer et al., 1999). Following the protocol, the subjects of the database were divided into groups of 200 clients and 95 impostors. Images corresponding to the subjects in these two groups were then partitioned into image sets used for:

- training and enrolment (3 images for each of the 200 clients) – this image set was used for training of the feature extraction techniques and for building client models/templates in the form of mean feature vectors,
- evaluation (3 images for each of the 200 clients and 8 images for each of the 25 evaluation impostors) – this image set was employed to determine the decision threshold for a given operating point of the face verification system and to estimate any potential parameters of the feature extraction techniques, and
- testing (2 images for each of the 200 clients and 8 images for each of the 70 test impostors) – this image set was used to determine the verification performance in real operating conditions (i.e., with predetermined parameters)

While the first image set featured only images belonging to the client group, the latter two image sets comprised images belonging to both the client and the impostor groups. The client images were employed to assess the first kind of error a face verification system can make, namely, the false rejection error, whereas the impostor images were used to evaluate the second type of possible verification error, namely, the false acceptance error. The two errors are quantified by two corresponding error rates: the false rejection and false acceptance error rates (FRR and FAR), which are defined as the relative frequency with which a face verification system falsely rejects a client- and falsely accepts an impostor-identity-claim, respectively. To estimate these error rates each feature vector extracted from an image of the client group was matched against the corresponding client template, while each of the feature vectors extracted from an impostor image was matched against all client templates in database. The described setup resulted in the following verification experiments: 600 client verification attempts in the evaluation stage, 40000 impostor verification attempts in the evaluation stage, 400 client verification attempts in the test stage and 112000 impostor verification attempts in the test stage (Messer et al., 1999; Štruc et al. 2008).

It has to be noted that there is a tradeoff between the FAR and FRR. We can select an operating point (determined by the value of the decision threshold) where the FAR is small and the FRR is large or vice versa, we can choose an operating point with a small FRR but at the expense of a large FAR. To effectively compare two face verification systems, an operating point that ensures a predefined ratio of the two error rates has to be selected on the evaluation image set or the values of the error rates must be plotted against various values of the decision threshold, resulting in the so-called performance curves. In this chapter we choose the latter approach and present our results in terms of two kinds of performance curves, namely, the Detection Error Tradeoff (DET) curves and the Expected Performance Curves (EPC), which plot the FAR against the FRR at different values of the decision threshold on the evaluation and test sets, respectively.

5.2 The Extended YaleB database

The Extended YaleB database was recorded at the Yale University and comprises 2432 frontal face images of 38 distinct subjects (Georghiades et al., 2001; Lee et al., 2005). It exhibits large variations in illumination, which is also the main source of variability in the images of the Extended YaleB database employed in our experiments. Some examples of these images are shown in Fig. 8.

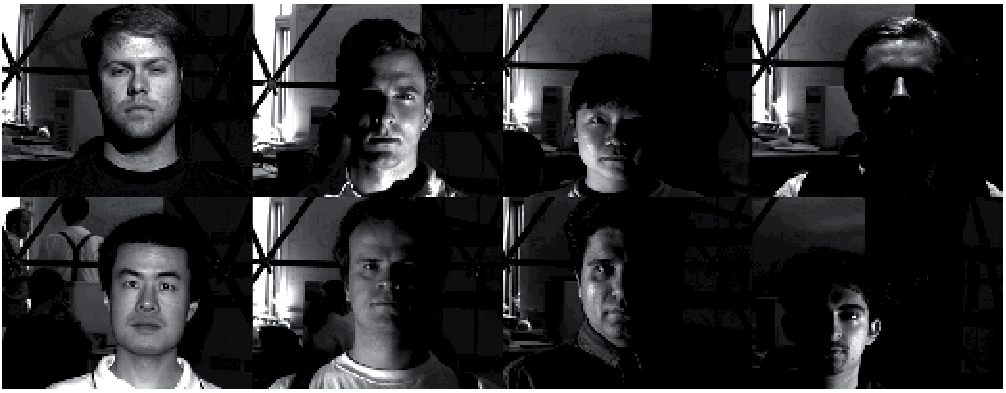


Fig. 8. Sample images from the Extended YaleB database

After removing a number of corrupt images from the database, a total of 2414 frontal face images with variable lighting were available for our experiments with each subject of the database being accounted for with a little more than 60 images. These images were then partitioned into 5 image subsets according to the extremity of illumination in the images, as proposed by Georghiades et al. in (Georghiades et al., 2001). The reader is referred to the original publication for more information on the partitioning.

The first image subset (denoted as S1 in the remainder) featured images captured in relatively good illumination conditions, while the conditions got more extreme for the image subsets two (S2) to five (S5). It should also be noted that the subsets did not contain the same number of images. The first subset, for example, contained 263 images, which corresponds to approximately 7 images per subject. The second subset contained 456 images, the third 455 images, the fourth 526 images and finally the fifth subset contained 714 facial images.

For our experiments we adopted the first subset for the training of the feature extraction techniques as well as for creating the user models/templates, and employed all remaining subsets for testing. Such an experimental setup resulted in highly miss-matched conditions for the recognition technique, since the test subsets featured images captured under varying illumination conditions, while the training images were acquired in controlled illumination conditions. Clearly, for a feature extraction technique to be successful, it has to extract stable features from the images regardless of the conditions present during the image acquisition stage. Furthermore, the experimental configuration is also in accordance with real life settings, as the training and enrollment stages are commonly supervised and, hence, the training and/or enrollment images are usually of good quality. The actual operational conditions, on the other hand, are typically unknown in advance and often induce severe illumination variations.

The results of our experiments on the Extended YaleB database are reported in terms of the rank one recognition rate, which corresponds to the relative frequency with which the test images from a given subset are recognized correctly.

5.3 Data pre-processing

Prior to the experiments, we subjected all images from both databases to a pre-processing procedure comprised of the following steps:

- a conversion of the (colour) face images to 8-bit monochrome (grey-scale) images – applicable only for the XM2VTS database,
- a geometric normalization procedure, which, based on the manually marked eye coordinates, rotated and scaled the images in such a way that the centres of the eyes were aligned and, thus, located at predefined positions,
- a cropping procedure that cropped the facial region of the images to a standard size of 128 × 128 pixels,
- a photometric normalization procedure that first equalized the histogram of the cropped facial images and then further normalized the results to zero-mean and unit-variance.

It should be noted that manual labelling of the facial landmarks is the only way to achieve a fair comparison of the recognition techniques, as it ensures that the differences in the observed recognition performances are only a consequence of the employed feature extraction techniques and not other influencing factors. Some examples of the pre-processed images (prior to photometric normalization) from the two databases are shown in Fig. 9.

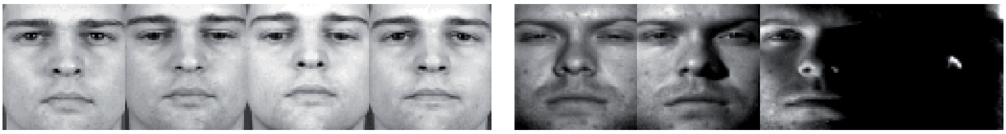


Fig. 9. Examples of pre-processed images from the XM2VTS (left quadruple of images) and Extended YaleB (right quadruple of images) databases

6. Experiments, results and discussion

6.1 Baseline performance

In the first series of our recognition experiments, we aimed at determining the performance of some baseline face recognition techniques on the two test databases. To this end, we implement the popular Principal Component Analysis (PCA) (Turk & Pentland, 1991) and Linear Discriminant Analysis (LDA) (Belhumeur et al., 1997) techniques, also known as the Eigenface and Fisherface methods, and assess the techniques for different lengths (i.e., different Number Of Features - NOF) of the PCA and LDA feature vectors. The results of this assessment are presented in Fig. 10 for the XM2VTS database in the form of DET curves and in Table 1 for the Extended YaleB database (EYB) in the form of rank one recognition rates (in %). Considering the number of subjects and images in the databases, the maximum length of the feature vector for the PCA technique equals 599 for the XM2VTS database and 262 for the EYB database, while the maximum length for the LDA technique is 199 for the XM2VTS database and 37 for the EYB database.

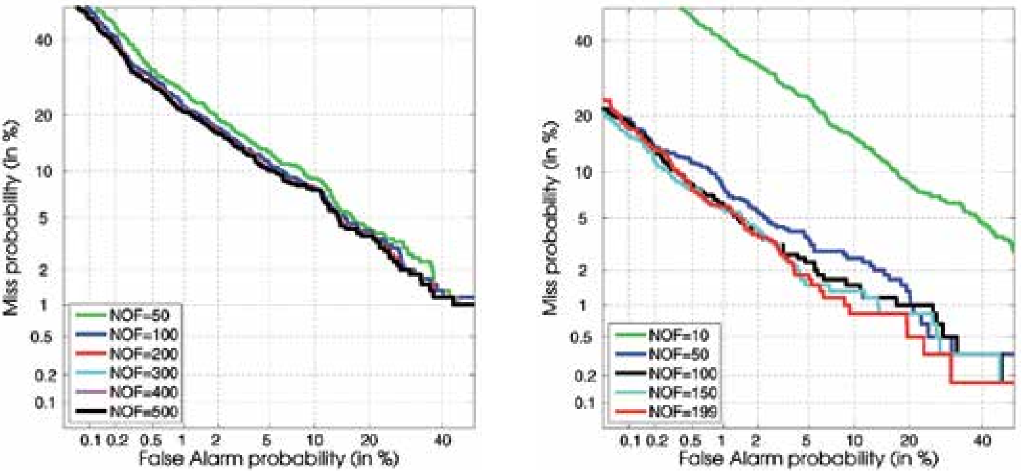


Fig. 10. DET curves of the baseline experiments on the evaluation image sets of the XM2VTS database: for the PCA technique (left), for the LDA technique (right)

NOF	PCA				NOF	LDA			
	S2	S3	S4	S5		S2	S3	S4	S5
10	56.6	29.5	11.2	15.6	5	98.3	56.9	9.9	13.6
50	93.4	54.7	16.7	21.9	10	100	85.3	27.2	29.7
100	93.6	54.9	16.7	22.0	20	100	97.8	47.0	43.7
150	93.6	55.0	16.7	22.0	30	100	99.3	53.6	47.6
200	93.6	55.0	16.7	22.0	37	100	99.8	56.3	51.0

Table 1. Rank one recognition rates (in %) for different lengths of the PCA and LDA feature vectors obtained on different subsets of the EYB database

Note that for the PCA technique the performance on the XM2VTS saturates when 200 features are used in the feature vectors. Similar results are also observed for the EYB database, where the performance on all subsets peaks with 150 dimensional feature vectors. For the LDA technique the best performance on both databases is achieved with the maximum number of features, i.e., 199 for the XM2VTS database and 37 for the EYB database. The presented experimental results provide a baseline face recognition performance on the two databases for the following comparative studies of the techniques using the augmented phase congruency feature vectors.

6.2 Baseline performance with the augmented phase congruency feature vector

In our second series of face recognition experiments we evaluate the performance of the PCA and LDA techniques in conjunction with the augmented phase congruency feature vectors and assess the relative usefulness of additional normalization techniques applied to the augmented feature vectors prior to the deployment of the subspace projection techniques PCA and LDA. We use five filter scales ($p = 5$) and eight orientations ($r = 8$) to construct the oriented phase congruency patterns, which in their down-sampled form constitute the augmented phase congruency feature vectors, and apply the following normalization schemes to these vectors:

- after the down-sampling of the oriented Gabor phase congruency patterns, each down-sampled OGPCP is normalized to zero mean and unit variance prior to concatenation into the final augmented feature vector (denoted as ZMUV) – Fig. 11 (upper left corner),
- after the down-sampling of the oriented Gabor phase congruency patterns, each down-sampled OGPCP is first subjected to histogram equalization and then to zero mean and unit variance normalization prior to concatenation into the final augmented feature vector (denoted as oHQ) - Fig. 11 (upper right corner), and
- after the down-sampling of the oriented Gabor phase congruency patterns, the down-sampled OGPCPs are concatenated into a bigger image, which is subjected to the histogram equalization procedure and then to zero mean and unit variance normalization (denoted as HQ) - Fig. 11 (lower row).

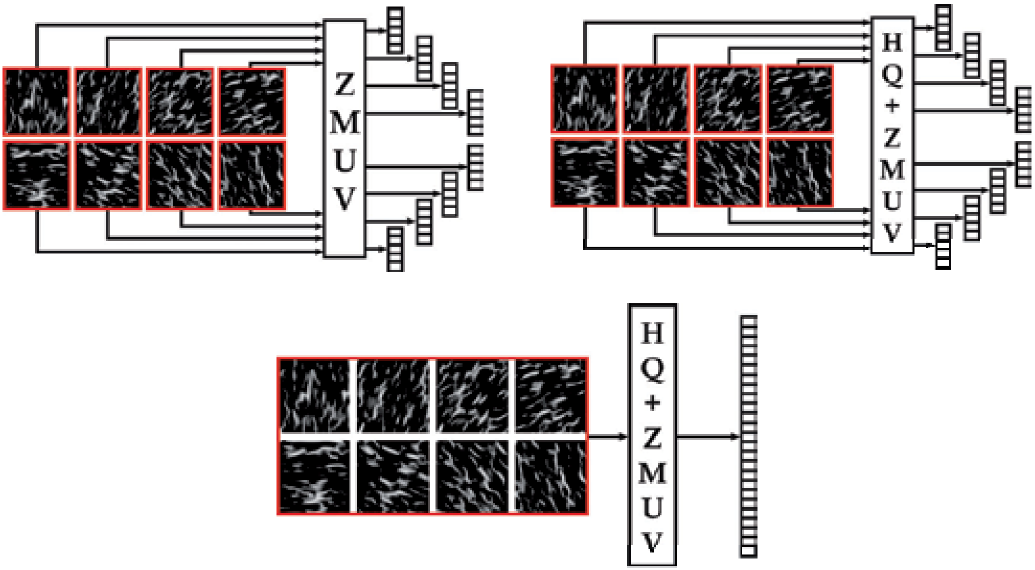


Fig. 11. Diagrams of the employed normalization procedures: ZMUV (upper left corner), oHQ (upper right corner), and HQ (lower row)

It should be noted that the number of scales and orientations that were used in our experiments, i.e., $p = 5$ and $r = 8$, was chosen based on other Gabor filter based methods presented in the literature – see, for example, (Liu & Wechsler, 2002; Shen et al., 2007; Štruc & Pavešić, 2009b). For the implementation of the subspace projection techniques the following feature vector lengths were chosen: 37 for LDA on EYB, 199 for LDA on XM2VTS, 150 for PCA on EYB and 200 for PCA on XM2VTS. These lengths were selected based on the baseline results from the previous series of experiments. However, since the number of features in the feature vectors is not the primary concern of this section, it could also be set differently.

The results of the experiments are again presented in the form of DET curves for the XM2VTS database in Fig. 12 and in the form of rank one recognition rates for the EYB database in Table 2.

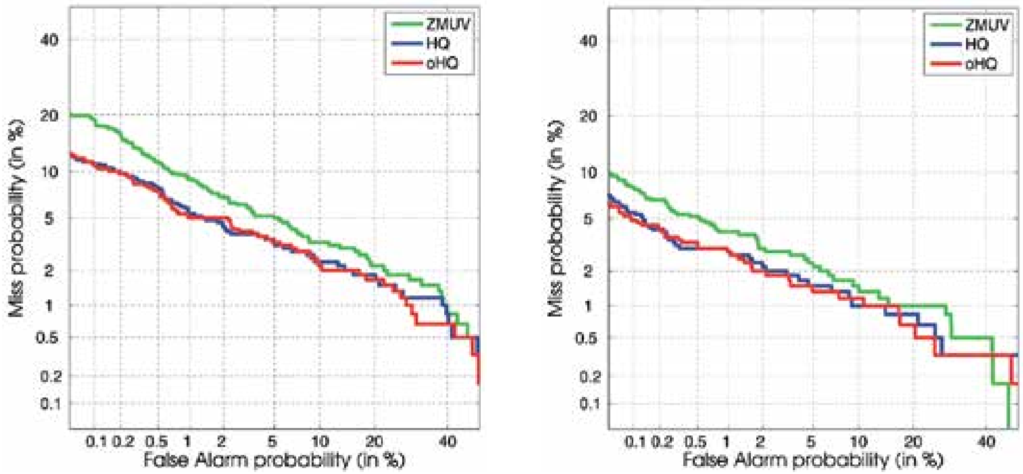


Fig. 12. DET curves of the comparative assessment of the normalization techniques on the evaluation image sets of the XM2VTS database: for the PCA technique (left), for the LDA technique (right)

Norm	PCA				Norm	LDA			
	S2	S3	S4	S5		S2	S3	S4	S5
ZMUV	100	99.1	83.4	92.7	ZMUV	100	99.8	88.8	93.8
HQ	100	99.1	81.6	89.8	HQ	100	100	86.1	94.8
oHQ	100	99.3	84.6	92.7	oHQ	100	100	87.1	94.8

Table 2. Rank one recognition rates (in %) for different normalization schemes of the augmented phase congruency vector prior to PCA and/or LDA deployment on different subsets of the EYB database

From the experimental results we can see that the traditional ZMUV technique resulted in the worst performance, while both the HQ and oHQ techniques achieved similar recognition rates on both databases. While the difference in their performance is statistically not significant, we nevertheless chose the oHQ technique for our following comparative assessments due to better results on the EYB database. Furthermore, if we compare the results obtained with the PCA and LDA techniques on the raw pixel data (Table 1) and the results obtained with the augmented feature vectors, we can see that the performance has improved significantly.

6.3 Impact of filter scales

In the third series of face recognition experiments, we assess the impact of the number of filter scales p in the Gabor filter bank on the performance of the PCA and LDA techniques applied to the augmented phase congruency feature vectors. We fix the angular resolution of the filter bank to $r = 8$ and vary the value of the filter scales from $p = 2$ to $p = 5$. In all of the performed experiments we use the same dimensionality of the PCA and LDA feature vectors as in the previous section and adopt the oHQ technique for the normalization of the augmented feature vectors. We once more present the results of the described experiments in form of the DET curves for the XM2VTS database (Fig. 13) and in form of rank one recognition rates for the EYB database (Table 3).

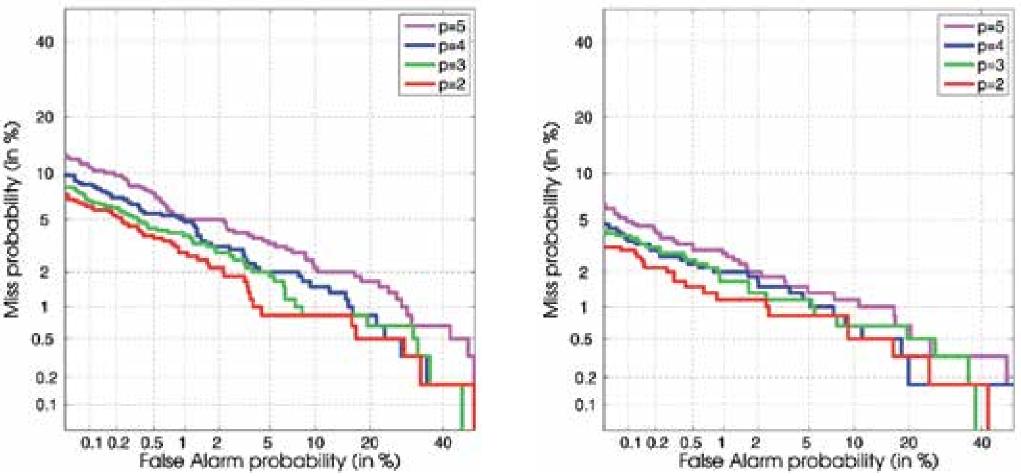


Fig. 13. DET curves generated for different numbers of filter scales employed during construction of the OGPCPs. The results were obtained on the evaluation image sets of the XM2VTS database: for the PCA technique (left), for the LDA technique (right)

p	PCA				p	LDA			
	S2	S3	S4	S5		S2	S3	S4	S5
5	100	99.3	84.6	92.7	5	100	100	87.1	94.8
4	100	100	91.8	94.8	4	100	100	94.5	94.4
3	100	100	93.4	95.2	3	100	100	96.4	96.4
2	100	100	93.0	92.3	2	100	100	94.7	96.6

Table 3. Rank one recognition rates (in %) on the EYB database for different numbers of filter scales employed during construction of the OGPCPs.

We can notice that on the XM2VTS database the verification performance steadily improves when the number of filter scales employed for the construction of the augmented phase congruency feature vector decreases. Thus, the best performance for the PCA (Fig. 13 left) as well as for the LDA (Fig. 13 right) techniques is observed with two filter scales, i.e., $p = 2$. Here, equal error rates of 2.15% and 1.16% are achieved for the PCA and LDA techniques, respectively.

Similar results are obtained on the EYB database. Here, the performance also increases with the decrease of used filter scales. However, the performance peaks with $p = 3$ filter scales. Since the improvements with the EYB database are not as pronounced as with the XM2VTS database, we chose to implement the construction procedure of the augmented phase congruency feature vector with 2 filter scales for the final comparative assessment.

6.4 Comparative assessment

In our last series of recognition experiments, we compare the performance of the PCA and LDA techniques on the proposed augmented phase congruency feature (PCF) vector with that of several established face recognition techniques from the literature. Specifically, we implement the following techniques for our comparative assessment: the Eigenface technique (PCA) (Turk & Pentland, 1991), The Fisherface technique (LDA) (Belhumeur et al., 1997), and the LDA and PCA techniques applied to the Gabor face representation (GF) proposed in (Liu & Wechsler, 2002).

All experiments on the XM2VTS database presented so far have been performed on the evaluation image sets, while the test image sets were not used. In this series of experiments we employ the test image sets for our assessment and implement all recognition techniques with all parameters (such as decision thresholds, feature vector lengths, number of employed filter scales, etc.) predefined on the evaluation image sets. Differently from the experiments presented in the previous sections, we do not present the results in the form of DET curves, but rather use the EPC curves. The choice of these performance curves is motivated by the work presented in (Bengio & Marithoz, 2004). Here, the authors argue that two recognition techniques cannot be compared fairly using DET curves, as in real life operating conditions a decision threshold has to be set in advance. In these situations the actual operating point may differ from the operating point the threshold was set on. To overcome this problem, the authors proposed the EPC curves, which plot the half total error rate (HTER=0.5(FAR+FRR)) against the parameter α, which controls the relative importance of the two error rates FAR and FRR in the expression: α FAR + $(1 - \alpha)$FRR. To produce the EPC curves, an evaluation image set and a test image set are required. For each α the decision threshold that minimizes the weighted sum of the FAR and FRR is computed on the evaluation image set. This threshold is then used on the test images to determine the value of the HTER used for the EPC curves.

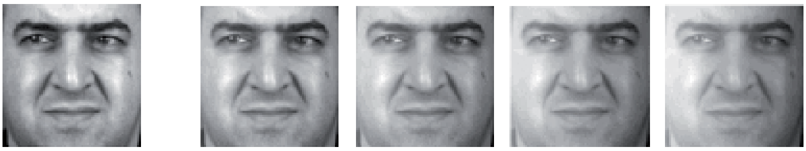


Fig. 14. Examples of modified face images (from left to right): the original image, the modified image for $\tau = 40$, the modified image for $\tau = 80$, the modified image for $\tau = 120$, the modified image for $\tau = 160$

To make the final assessment more challenging, we introduce an artificial illumination change to the test sets of the XM2VTS database. To this end, we adopt the model previously employed in (Sanderson & Paliwal, 2003), which simulates different illumination conditions during the image acquisition stage by modifying the pre-processed face images $I(x, y)$, i.e.,

$$\tilde{I}(x,y) = I(x,y) + mx + \tau, \tag{12}$$

where $x = 0, 1, \dots, a - 1$; $y = 0, 1, \dots, b - 1$; $m = -2\tau/(b - 1)$ and τ denotes the parameter that controls the "strength" of the introduced artificial illumination change. Sanderson and Paliwal (Sanderson & Paliwal, 2003) emphasized that this model does not cover all illumination effects possible in real life settings, but is nevertheless useful for providing suggestive results. Some examples of the modified face images $\tilde{I}(x,y)$ obtained with the presented model for different values of the parameter τ are shown in Fig. 14.

The results of the final assessment are presented in Fig. 15 for the XM2VTS database and in Table 4 for the EYB database.

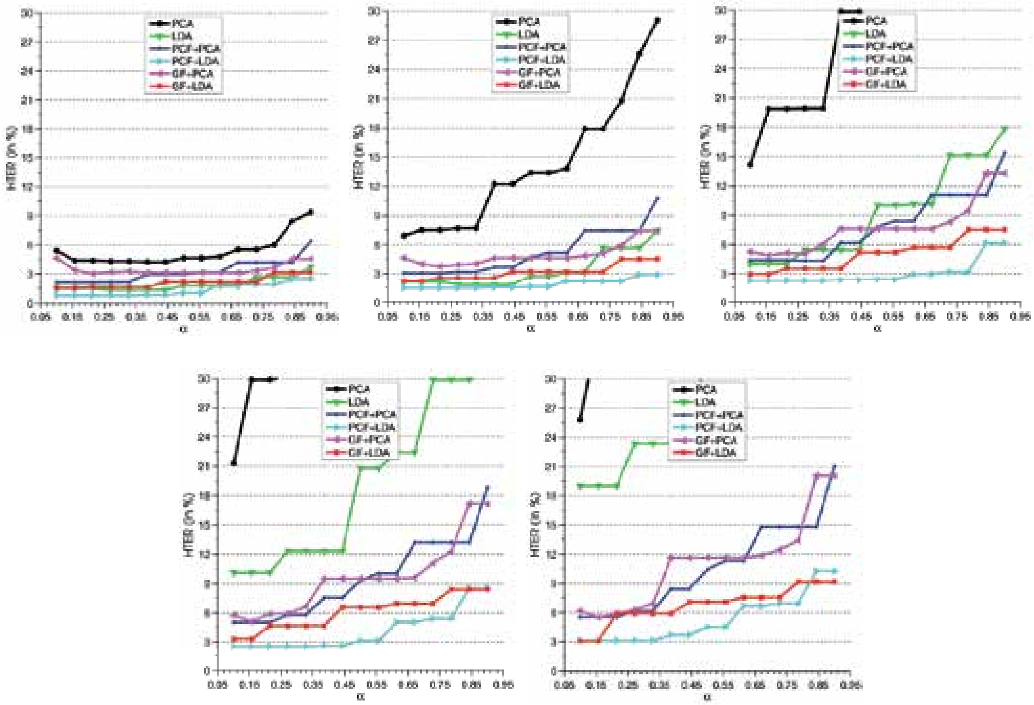


Fig. 15. EPC curves obtained on the test sets of the XM2VTS database for different values of the parameter τ (from left to right starting in the upper left row): with the original images, with the modified images with τ = 40, with the modified images with τ = 80, with the modified images with τ = 120, with the modified images with τ = 160

Method	S2	S3	S4	S5
PCA	93.6	55.0	16.7	22.0
LDA	100	99.8	56.3	51.0
GF+PCA	100	97.8	77.9	85.2
GF+LDA	100	100	83.2	89.1
PCF+PCA	100	100	93.0	92.3
PCF+LDA	100	100	94.5	94.4

Table 4. Rank one recognition rates (in %) on the EYB for the comparative assessment

The first thing to notice from the presented results is that both the Gabor magnitude as well as the Gabor phase congruency features result in a significant improvement in the recognition performance when compared to the raw pixel data and, furthermore, that both types of features result in a more robust performance in the presence of illumination changes. This fact is best exemplified by the recognition rates on subsets S4 and S5, where the increase in performance from the pixel data to the Gabor-based features is more than 60% (in absolute terms) for the PCA-based techniques and more than 25% (in absolute terms) for the LDA-based techniques.

In general, the augmented Gabor phase congruency feature vectors resulted in better performance in difficult illumination conditions than the Gabor magnitude features. While this improvement was only minimal on the XM2VTS database and its synthetically degraded versions, the results on the EYB database show improvements (on the image subset S4) of around 10% in absolute terms.

6.5 Discussion

From the experimental results presented in previous sections, we found that amongst the tested feature extraction techniques, the LDA technique combined with the Gabor magnitude and Gabor phase congruency features ensured the best recognition performance on all experimental databases (i.e., on the XM2VTS, the EYB and on the degraded versions of the XM2VTS databases). While both feature types significantly improved upon the techniques baseline performance with "raw" pixel intensity values, there are several differences in both feature types, which affect their usability in real-life face recognition systems.

First of all, as stated in a number of studies from the literature (e.g., Liu & Wechsler, 2002; Shen & Bai, 2006; Shen et al. 2007; Štruc & Pavešić, 2009b), the Gabor magnitude based methods require 40 Gabor filters, i.e., filters with five scales and eight orientations, to achieve their optimal performance. The same number of filters was also used in our experiments to obtain the performance presented in previous sections. The Gabor phase congruency features based methods presented in this chapter, on the other hand, require only 16 Gabor filters, filters with two scales and eight orientations, for an optimal performance. This fact makes the Gabor phase congruency methods significantly faster than the Gabor magnitude based methods.

Second of all, since there is only one output per employed filter orientation for the Gabor phase congruency based methods and not five, as it is the case with the Gabor magnitude based techniques, the increase in data is not that extreme for the proposed face representation.

Last but not least, we have to emphasize that in its optimized form (with two filter scales and eight orientations) the Gabor phase congruency techniques operate on a much narrower frequency band than the Gabor magnitude methods. Based on the experimental results presented in previous sections, we can in fact conclude that most of the discriminatory Gabor-phase information is contained in the OGPCPs obtained with Gabor filters of high frequencies ($u = 0, 1$). In addition to the high frequency filters, the Gabor magnitude methods effectively also use the low frequency Gabor filters. This finding suggests that the Gabor phase congruency and Gabor magnitude features represent feature types with complementary information and could therefore be combined into a unified feature extraction technique which uses Gabor magnitude as well as Gabor phase information for face recognition.

7. Conclusion and future work

In this chapter we have proposed a novel face representation derived from the Gabor filter outputs. Unlike popular Gabor filter based methods, which mainly use only Gabor magnitude features for representing facial images, the proposed feature extraction technique exploits the Gabor phase information and derives the novel face representation named the Oriented Gabor Phase Congruency Pattern or OGPCP. This representation forms the foundation for the construction of the augmented Gabor phase congruency feature vector, which, similar to the established Gabor magnitude representations, can be combined with subspace projection techniques to form powerful and efficient feature extraction approaches. The feasibility of the proposed face representation (or features) was assessed on two publicly available datasets, namely, on the XM2VTS and on the Extended YaleB dataset. On both datasets, the proposed features resulted in a promising face recognition performance and outperformed popular face recognition techniques, such as PCA, LDA, the Gabor-Fisher classifier and others. The proposed features were shown to ensure robust recognition performance in the presence of severe illumination changes as well.

The future work with respect to the proposed Gabor phase congruency face representation, i.e., the OGPCP, will be focused on evaluating different strategies to combine the traditional Gabor magnitude face representation with the proposed Gabor phase congruency patterns of facial images.

8. References

Belhumeur, B.; Hespanha, J. & Kriegman, D. (1997). Eigenfaces vs. Fisherfaces: Recognition using Class specific Linear Projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 19, No. 7, 711-720, 0162-8828

Bengio, S. & Marithoz, J. (2004). The Expected Performance Curve: A New Assessment Measure for Person Authentication. *Proceedings of the Speaker and Language Recognition Workshop Oddyssey*, pp. 279-284

Bezalel, E. & Efron, U. (2005). Efficient Face Recognition Method Using a Combined Phase Congruency/Gabor Wavelet Technique. *Proceedings of the SPIE Conference on Optical Information Systems III*, doi:10.1117/12.618226, San Diego, CA, USA, August 2005, SPIE

Gao, Y. & Leung, M.K.H. (2002). Face Recognition Using Line Edge Map. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 24, No. 6, 764-779, 0162-8828

Georghiades, A.S.; Belhumeur, P.N. & Kriegman, D. (2001). From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 23, No. 6, 643-660, 0162-8828

Gross, R.; Baker, S.; Metthews, I. & Kanade, T. (2004). Face Recognition Across Pose and Illumination, In: *Handbook of Face Recognition*, Li, S.Z. & Jain, A.K. (Ed.), 193-216, Springer, 0-387-40595-X, New York, USA

Gundimada, S. & Asari, V.K. (2006). A Novel Neighborhood Defined Feature Selection on Phase Congruency Images for Recognition of Faces with Extreme Variations. *International Journal of Information Technology*, Vol. 3, No. 1, 25-31, 2070-3961

Gundimana, S.; Asari, V.K. & Gudur N. (2009). Face Recognition in Multi-sensor Images Based on a Novel Modular Feature Selection Technique. *Information Fusion*, article in press, 1566-2535

Jain, A.K.; Ross, A. & Prabhakar, S. (2004). An Introduction to Biometric Recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 14, No. 1, 4-20, 1051-8215

Kovesi, B. (1999). Image Features from Phase Congruency. *Videre: Journal of Computer Vision Research*, Vol. 1, No. 3, 1-26

Lades, M.; Vorbruggen, J.; Buhmann, J.; Lange, J.; Malsburg, C. von der; Wurtz, R. & Konen, W. (1993). Distortion Invariant Object Recognition in the Dynamic Link Architecture. *IEEE Transactions on Computers*, Vol. 42, No. 3, 300-311, 0018-9340

Lee, K.C.; Ho, J. & Kriegman, D. (2005). Acquiring Linear Subspaces for Face Recognition under Variable Lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 27, No. 5, 684-698, 0162-8828

Liu, C. & Wechsler, H. (2002). Gabor Feature Based Classification using the Enhanced Fisher Linear Discriminant Model for Face Recognition. *IEEE Transactions on Image Processing*, Vol. 11, No. 4, 467-476, 1057-7149

Liu, C. (2006). Capitalize on Dimensionality Increasing Techniques for Improving Face Recognition Grand Challenge Performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 28, No. 5, 725-737, 0162-8828

Messer, K.; Matas, J.; Kittler, J.; Luettin, J. & Maitre, G. (1999). XM2VTSDB: The Extended M2VTS Database. *Proceedings of the 2nd International Conference on Audio- and Video-Based Person Authentication*, pp. 72-77, Washington D.C., USA

Messer, K.; Kittler, J.; Short, J.; Heusch, G.; Cardinaux, F.; Marcel, S.; Rodriguez, Y.; Shan, S.; Su, Y.; Gao, W. & Chen, X. (2006). Performance Characterisation of Face Recognition Algorithms and Their Sensitivity to Severe Illumination Changes. *Proceedings of the IAPR International Conference on Biometrics*, pp. 1-11, 978-3-540-31111-9, Hong Kong, China, January 2006, Springer, Berlin/Heidelberg

Phillips, P.J.; Scruggs, W.T.; O'Toole, A.J.; Flynn, P.J.; Bowyer, K.W.; Schott, C.L. & Sharpe, M. (2007). FRVT 2006 and ICE 2006 Large-Scale Results, *NISTIR 7408*

Poh, N.; Chan, C.H.; Kittler, J.; Marcel, S.; McCool, C.; Argones-Rua, E.; Alba-Castro, J.L.; Villegas, M.; Paredes, R.; Štruc, V.; Pavešić, N.; Salah, A.A.; Fang, H. & Costen, N. (2009). Face Video Competition. *Proceedings of the IAPR International Conference on Biometrics*, pp. 715-724, 978-2-642-01792-6, Alghero, Italy, June 2009, Springer, Berlin/Heidelberg

Sanderson, C. & Paliwal, K. (2003). Fast Features for Face Authentication under Illumination Direction Changes. *Pattern Recognition Letters*, Vol. 24, No. 14, 2409-2419, 0167-8655

Shen, L. & Bai, L. (2006). A Review of Gabor Wavelets for Face Recognition. *Pattern analysis and applications*, Vol. 9, No. 2, 273-292, 1433-7541

Shen, L.; Bai, L. & Fairhurst, M. (2007). Gabor Wavelets and General Discriminant Analysis for Face Identification and Verification. *Image and Vision Computing*, Vol. 25, No. 5, 553-563, 0262-8856

Short, J.; Kittler, J. & Messer, K. (2005). Photometric Normalisation for Face Verification, *Proceedings of the 5th International Conference on Audio- and Video-Based Person Authentication*, pp. 617-626, 978-3-540-27887-0, New York, USA, July 2005, Springer, Berlin/Heidelberg

Štruc, V.; Vesnicer, B. & Pavešić, N. (2008a). The Phase-based Gabor Fisher Classifier and its Application to Face Recognition under Varying Illumination Conditions. *Proceedings of the 2nd International Conference on Signal Processing and Communication Systems*, pp. 1-6, 978-1-4244-4242-3, Gold Coast, Australia, IEEE, NJ

Štruc, V.; Mihelič, F. & Pavešić, N. (2008b). Face Recognition using a Hybrid Approach. *Journal of Electronic Imaging*, Vol. 17, No. 1, 1-11, 1017-9909

Štruc, V. & Pavešić, N. (2009a). Phase-congruency Features for Palm-print Verification. *IET Signal Processing*, Vol. 3, No. 4, 258-268, 1751-9675

Štruc, V. & Pavešić, N. (2009b). Gabor-based Kernel-partial-least-squares Discrimination Features for Face Recognition. *Informatica (Vilnius)*, Vol. 20, No. 1, 115-138, 0868-4952

Turk, M. & Pentland, A. (1991). Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, 71-86, 0898-929X

Zhang, B; Shan, S., Chen, X. & Gao, W. (2007). Histogram of Gabor Phase Patterns (HGPP): A Novel Object Representation approach for Face Recognition. *IEEE Transactions on Image Processing*, Vol. 16, No. 1, 57-68, 1057-7149

Robust Face Alignment for Illumination and Pose Invariant Face Recognition

Fatih Kahraman[1], Binnur Kurt[2], Muhittin Gökmen[2]
Istanbul Technical University, [1]Informatics Institute, [2]Computer Engineering Department
Turkey

1. Introduction

Face recognition systems have matured from the systems working only in highly controlled indoor environments to the systems capable of identifying individuals in indoor or outdoor environments under severe conditions though some problems still remain, constraining their success to a limited degree. Illumination and pose variations are mostly responsible for dramatic changes on the face appearance. They produce such complex effects on the appearance of the acquired face that the face image pertains little to the actual identity. So any improvement in face appearance will enhance the recognition performance. Face recognition systems are usually required to handle highly varying illumination and pose conditions and more advanced techniques are needed to eliminate the undesired effects of variations from any sources. Research on face recognition has focused on solving issues arising from illumination and pose variations in one or more shots.

Lighting direction changes alter the relative gray scale distribution of face image and those changes due to lighting are larger than the one due to different personal identities. Consequently, illumination normalization is required to reach acceptable recognition rates. Varying illumination is a difficult problem and has received much attention in recent years. Several recent studies are centered around this issue: symmetric shape from shading (Zhao & Chellappa, 2000), for the illumination cones method (Georghiades & Belhumeur, 2001) theoretically explained the property of face image variations due to light direction changes. In this algorithm, both self shadow and cast-shadow were considered and its experimental results outperformed most existing methods. The main drawbacks of the illumination cone model are the computational cost and the strict requirement of seven input images per person.

Other directions of the photometric stereo in face recognition include introducing a more general illumination model, (Ramamoorthi, 2002) proposed a spherical harmonic representation for face images under various lighting conditions. (Basri & Jacobs, 2001) represent lighting using a spherical harmonic basis wherein a low-dimensional linear subspace is shown to be quite effective for recognition. The harmonic images can easily be computed analytically given surface normals and the albedos. (Zhou et al., 2007) extended a photometric stereo approach to unknown light sources. (Lee et al., 2005) empirically found a set of universal illumination directions, images under which can be directly used as a basis for the 9 dimensional illumination subspaces.

(Shashua & Riklin-Raviv, 2001) employ a very simple and practical image ratio method to map the face images into different lighting conditions. This method is suitable for modelling the variation in facial appearance caused by diffuse reflection and the proposed method is simply the ratio of albedo between a face image and linear combination of basis images, for each pixel. (Wang et al. , 2004) developed a reflectance estimation methods by using the idea of the ratio of the original image and its smooth version. In the same research direction (Zhang et al., 2007) and (An et al., 2008) proposed new methods to extract an illumination invariant representation of a face images from a raw facial images. Even though the proposed photometric normalization based representations increase the recognition performance, it is not suitable to say that these representations provide complete invariance against illumination. There are many recent works on illumination invariant face recognition. An extensive review of illumination invariant face recognition approaches is given by (Zou et al., 2007) and (Zhao & Chellappa, 2006).

There are several recent image-based studies on illumination invariant face recognition. Image-based methods are known to be robust to illumination variations. Main drawback of the image-based methods is that they always assume the face image is already aligned. Usually it is not an easy assumption to be satisfied especially when the input image is poorly illuminated. Appearance-based methods require training images of individuals taken under different illumination conditions. A method proposed by (Sim & Kanade, 2001) overcomes this restriction by using a statistical shape-from-shading model. Using this method they generate images for each of the individuals under different lighting conditions to serve as database images in a recognizer.

Face alignment is a crucial step to extracting good facial features and obtaining high performance in face recognition, expression analysis and face animation applications. Several face alignment methods were proposed by Active Shape Models (ASM) (Cootes et al., 1995) and Active Appearance Models (AAM) (Cootes et al., 2001; Stegmann et al., 2003), by Cootes *et al* which are two successful models for object localization. ASM utilizes local appearance models to find the candidate shape and global model to constrain the searched shape. AAM combines the constraints on both shape and texture variations in its characterization of facial appearance. In searching for a solution, it assumes linear relationships between appearance variation and texture variation and between texture variation and position variation. In this study, we have used AAM to solve the pose-invariant face alignment problem.

AAM is known to be very sensitive to illumination, particularly if the lighting conditions during testing are significantly different from the lighting conditions during training. Several variations of AAM appear in the literature to improve the original algorithm, namely Direct Appearance Models (Hou et al., 2001) and view-based AAM (Cootes et al., 2002). Cootes et al constructed three AAMs which are called as View-based AAMs. These models are linear model of frontal, profile and half profile views of faces. They also show how to estimate the pose from the model parameters. The approach in this study differs from their method in the way that only one AAM is constructed rather than three models. The motivation here is to reduce the three separate searching procedures to just one fitting procedure based on one linear statistical model. The model has better generalization performance in capturing pose variations than the one using three separate linear models. In order to construct the one linear model, a training dataset comprised of 8 different poses of 3 individuals captured under similar illumination conditions is used. Despite the success of these methods, problems still remain to be solved. Moreover, under the presence of partial occlusion, the PCA-based texture model of AAM causes the reconstruction error to be globally spread over the image, thus degrading alignment. In this paper, we propose an approach based on histogram-fitting to overcome the problem explained above. A detailed explanation of the proposed approach is given in Section 2.

Yet another issue related to face recognition is to recognize different poses of the same person. Pose-invariant face recognition requires pose alignment where images are captured either by multiple cameras or by a single camera at different time instances. There are several works related to pose normalization. (Blanz & Vetter, 2003) use a statistical 3D morphable model (3DMM) to tackle with pose and illumination variations. Since their method requires textured 3D scans of heads, it is computationally expensive. The vertices in a 3DMM shape are much denser than an AAM shape. 3DMM achieved promising results for illumination invariant face recognition. However, fitting a dense model requires much higher computational effort, which is not suitable for real-time face recognition systems.

In Section 3, we will study the proposed AAM based approach capable of producing different poses of unseen person and explain how a non-frontal face is projected to a frontal face in detail. In this paper, we have focused on the problems induced by varying illumination and poses in face recognition. Our primary goal is to eliminate the negative effect of challenging conditions, especially illumination and pose, on the face recognition system performance through illumination and pose-invariant face alignment based on Active Appearance Model. The rest of the paper is structured as follows: Section 2 introduces Active Appearance Model (AAM) and Section 3 introduces illumination normalization inserted into the searching procedure of AAM. Section 4 is for the proposed pose invariant combined active appearance model. The experimental results and the conclusion are presented in Section 5 and 6, respectively.

2. Active Appearance Model

Active Appearance Models are generative models capable of synthesizing images of a given object class. By estimating a compact and specific basis from a training set, model parameters can be adjusted to fit unseen images and hence perform image interpretation. The modeled object properties are usually shape and pixel intensities (here denoted texture). AAM aims to find the optimal model parameters to represent the target image that belongs to the same object class by using an iterative scheme.

Training objects are defined by marking up each image with points of correspondence. Relying upon the landmarks, a triangulated mesh is produced for the reference position and orientation of the object. Before modeling variations, all shape vectors are normalized to a common reference shape frame by using Procrustes Analysis (Goodall, 1991). After obtaining the reference shape vector, all of the training images are warped to the reference shape by using a piecewise affine warping (Glasbey & Mardia, 1998), which is defined between corresponding triangles to obtain normalized texture vectors.

Using prior knowledge of the optimization space, AAMs can rapidly be fitted to unseen images with a reasonable initialization given. AAM uses principal component analysis (PCA) to model the variations of the shapes and textures of the images. Usage of PCA representation allows AAM to model and represent a certain image with a small set of parameters.

AAM works according to the following principle: An image is marked with n landmark points. The content of the marked object is analyzed based on a Principal Component Analysis (PCA) of both texture and shape. The shape is defined by a triangular mesh and the vertex locations of the mesh. Mathematically the shape model is represented as follows:

$$x_0 = \left(\left(x_1, x_2, x_3, \ldots, x_n\right), \left(y_1, y_2, y_3, \ldots, y_n\right)\right) \in R^{2n} \tag{1}$$

Shape is reduced to a more compact form through PCA such that,

$$x = \overline{x} + \Phi_s b_s. \tag{2}$$

In this form, x is the synthesized shape in the normalized frame, Φ_s is a matrix that contains the t-eigenvectors corresponding to the largest eigenvalues and b_s is a t-dimensional vector of shape coefficients. By varying the parameters in b_s, the synthesized shape can be varied.

In the texture case one needs a consistent method for collecting the texture information (intensities) between the landmarks, i.e. an image warping function needs to be established. This can be done in several ways. Here, we used a piece-wise affine warp (Glasbey & Mardia, 1998) based on the Delaunay triangulation (Shewchuk, 1996).

All training images are warped to the reference shape and are sampled into a vector to obtain the texture vectors represented as g. Prior to the PCA modeling of the texture, we need to normalize all texture vectors. So, a photometric normalization of the texture vectors of the training set is done to avoid the side effects of global linear changes in pixel intensities. The aim of this normalization is to obtain texture vectors with zero mean and unit variance. Texture model can now be obtained by applying PCA to the normalized textures,

$$g = \overline{g} + \Phi_g b_g \tag{3}$$

where g is the synthesized texture, $\overline{g}$ is the mean texture and b_g is a k-dimension vector of texture parameters. In the linear model of texture, Φ_g is a set of orthogonal modes of variation.

To remove the correlation between shape and texture model parameters, a third PCA is applied on the combined model parameters, giving a further model,

$$b = Q\ c \tag{4}$$

where Q is the eigenvectors and c is a vector of appearance parameters controlling both the shape and the texture of the model. Note we do not use a mean vector in this model since the shape, texture and appearance model parameters need to have zero mean. Due to the linear nature of the model, the shape and texture vectors can be expressed in terms of the appearance parameters c,

$$x = \overline{x} + \Phi_s W_s^{-1} Q_s c \tag{5}$$

$$g = \overline{g} + \Phi_g Q_g c, \tag{6}$$

where b=[$W_s b_s b_g$]^T and Q=[Q_s Q_g]^T. In this form, W_s is a diagonal matrix of weights for each shape parameter, allowing for the difference in units between the shape and the grey models. Generally W_s is the square root of the ratio of the total intensity variation to the total shape variation. An example image can be synthesized for a given c appearance vector by generating the shape normalized image from the vector g and warping it using the control points described by x vector. Appearance parameters vector, c, controls both the shape and the grey-levels of the model. Q_s and Q_g are the eigenvectors of the shape and texture models respectively. An image can be represented by a vector p which is written in terms of x, g and c as p=[x g c]^T. It is possible to synthesize a new image by changing the parameter p.

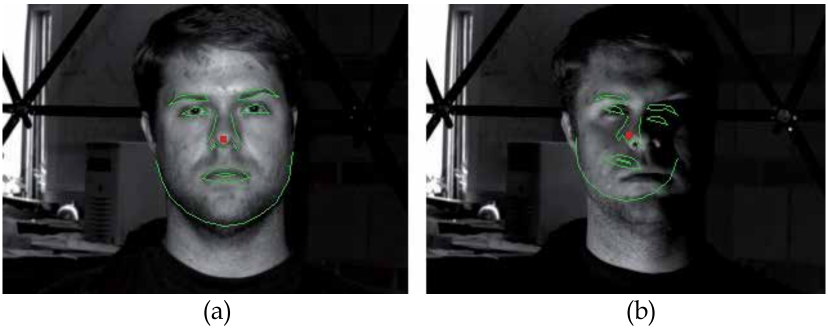


Fig. 1. Face alignment using standard AAM under good and extreme illumination.
(a) Normal illumination, (b) Extreme illumination.

The underlying problem with the classical AAM is demonstrated in Fig.1. In Fig.1 (a) a correct AAM search result is shown where the input image contains a frontal face which is also illuminated frontally. Since the model is constructed from a database containing frontally illuminated faces, the standard AAM searching procedure cannot converge to a meaningful solution for an extremely illuminated frontal face given in Fig.1 (b). We propose an illumination normalization method explained in Section 3 and insert it into the standard AAM searching procedure applied to the faces captured under different illumination conditions. The inserted normalization module guides the AAM to converge to a meaningful solution and also enhances the accuracy of the solution.

3. Illumination Normalization

We discuss here two light normalization methods and analyze their behavior in AAM searching. The first proposed method is ratio-image face illumination normalization method (Liu et al., 2005). Ratio-image is defined as the quotient between an image of a given face whose lighting condition is to be normalized and an image of the reference face. These two images are blurred using a Gaussian filter, and the reference image is then updated by an iterative strategy in order to improve the quality of the restored face. Using this illumination restoration method, a face image with arbitrary illumination can be restored to a face having frontal illumination.

The second normalization method discussed in this study is based on image histogram techniques. The global histogram equalization methods used in image processing for normalization only transfers the holistic image from one gray scale distribution to another. This processing ignores the face-specific information and cannot normalize these gray level distribution variations. To deal with this problem, researchers have made several improvements in recent years. The underlying problem is that well-lit faces do not have a uniform histogram distribution and this process gives rise to an unnatural face illumination. As suggested in (Jebara, 1996), it is possible to normalize a poorly illuminated image via histogram fitting to a similar, well illuminated image.

In this study, a new histogram fitting algorithm is designed for face illumination normalization taking the structure of the face into account. The algorithm is explained over poorly illuminated frontal face image where one side of the face is dark and the other side is bright. The main idea here is to fit the histogram of the input face image to the histogram of the mean face. The face is first divided into two parts (left/right) and then the histogram of each window is independently fitted to the histogram of mean face. For these two histograms, namely the histogram of the left window denoted as $H_{LEFT}(i)$ and the histogram of the right window denoted as $H_{RIGHT}(i)$, two mapping functions are computed: $f_{H_{LEFT} \to G}$ and $f_{H_{RIGHT} \to G}$ corresponding to the left and right windows, respectively. Here $G(i)$ is the histogram of the reference image which is also called as *mean face* in AAM. An artifact introduced by this mapping is the sudden discontinuity in illumination as we switch from the left side of the face to the right side. The problem is solved by averaging the effects of the two mapping functions with a linear weighting that slowly favors one for the other as we move from the left side to the right side of the face. This is implemented with the mapping function $f_{H_{TOTAL} \to G}$ defined as bellow:

$$f_{H_{TOTAL} \to G}(i) = leftness \times f_{H_{LEFT} \to G}(i) + (1 - leftness) \times f_{H_{RIGHT} \to G}(i) \tag{7}$$

Illumination normalization result is shown in Fig. 2 obtained by using the histogram fitting method explained above. As it can be seen from the figure the normalization method can produce more suitable images to be used in AAM search mechanism. The classical AAM search fails in all images given in the first row of Fig. 2. We will show in the next section that AAM search procedure can now converge to the correct shape for the restored image both in point-to-point error and point-to-curve error senses.

Fig. 3 presents several results obtained for Set 4 (left) and Set 3 (right) faces of different individuals having extremely dark and bright regions. A significant amount of improvement in quality can be easily verified from the experimental results. The dark parts now become somehow noisy whereas there are still some very bright areas.

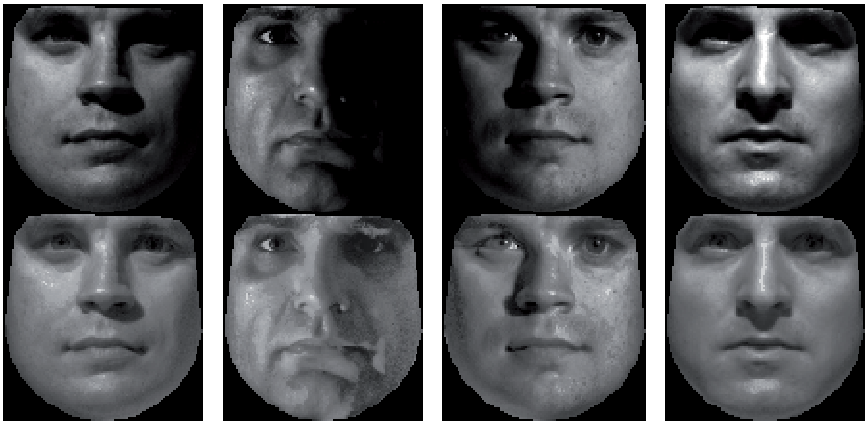


Fig. 2. Illumination normalization using histogram fitting: On the top the input images are given, and on the bottom the normalized images are shown.

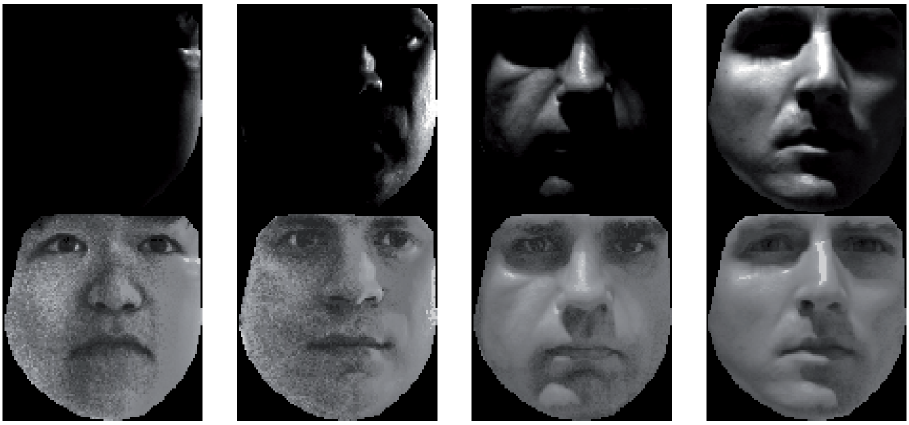


Fig. 3. Illumination normalization results for extreme cases: On the top the input images are given, and on the bottom the normalized images are shown.

4. Pose Normalization

Pose normalization is required before recognition in order to reach acceptable recognition rates. There are several studies related to pose normalization. (Blanz & Vetter, 2003) use a statistical 3D morphable model to tackle with pose and illumination variations. Since their method requires textured 3D scans of heads, it is computationally expensive. Cootes et al constructed three AAMs which are called as View-based AAMs (Cootes et al., 2002). We developed AAM based pose normalization method which uses only one AAM. There are two important contributions to the previous studies. By using the proposed method:

- One can synthetically generate appearances for different poses when a single frontal face image is available.
- One can generate frontal appearance of the face if only non-frontal face image is available.

Next section explains the proposed pose normalization and generation method in detail.

4.1 Pose Generation from 2D Images

The same variation in pose imposes similar effect on the face appearance for all individuals. Fig.4.a demonstrates how face texture and shape are affected by pose. Deformation mostly occurs on the shape whereas the texture is almost constant. Since the number of landmarks in AAM is constant, the wireframe triangles are translated or scaled as pose changes. Therefore, as we change pose, only wireframe triangles undergo affine transformation but the gray level distribution within these translated and rotated triangles remains the same. One can easily generate frontal face appearance if AAM is correctly fitted to any given non-frontal face of the same individual provided that there is no self-occlusion on face. Self-occlusion usually is not a problem for angles less than ±*45*.

For 2D pose generation, we first compute how each landmark point translates and scales with respect to the corresponding frontal counterpart landmark point for 8 different poses, and obtain a ratio vector for each pose. We use the ratio vector to create the same pose variation over the shape of another individual. In Fig.4.b, two examples are given where the landmark points of unseen individuals are synthetically generated using the ratio vector obtained from that different person.

Appearances are also obtained through warping in AAM framework, using synthetically generated landmarks given in Fig.4.b. These are shown in Fig.5. First column in Fig.5 shows the frontal faces and the second column shows appearances for various poses. It is important to note that the generated faces contain no information about the individual used in building the ratio matrix.

4.2 Training AAM for Pose Normalization

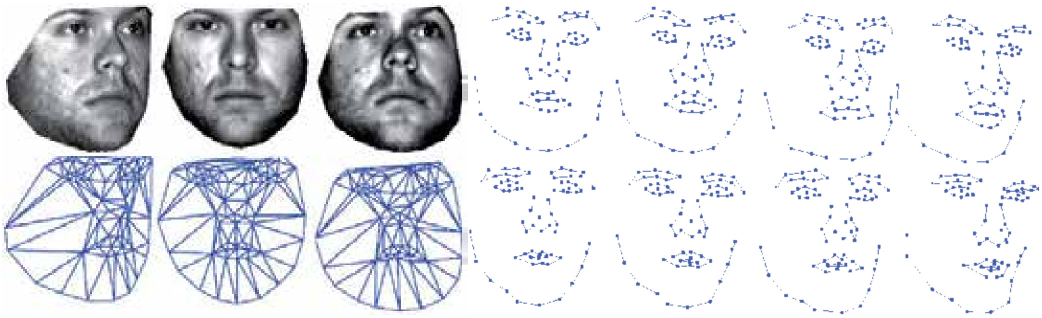
An AAM model trained by using only frontal faces can only fit into frontal faces well and fail to fit into non-frontal faces. Our purpose here is to enrich the training database by inserting synthetically generated faces at different poses so that AAM model trained by frontal faces can now converge to images at any pose.

We manually labeled 73 landmarks on 4920 images. Let us denote the landmark points on i^{th} frontal image as $S_i^0 = \big((x_{i,1}, y_{i,1}), (x_{i,2}, y_{i,2}), \ldots, (x_{i,K}, y_{i,K})\big) \in R^{2K}$ where $i = 1,2,\ldots,N$. N is *4920* and K=*73* in our database. The shape-ratio vector explained in the previous subsection (3.1) is defined between the p-posed shape and the frontal shape as

$$r_p(S^p, S^0) = \left(\left(\frac{x_{p,1}}{x_{0,1}}, \frac{y_{p,1}}{y_{0,1}} \right), \ldots, \left(\frac{x_{p,K}}{x_{0,K}}, \frac{y_{p,K}}{y_{0,K}} \right) \right) \tag{8}$$

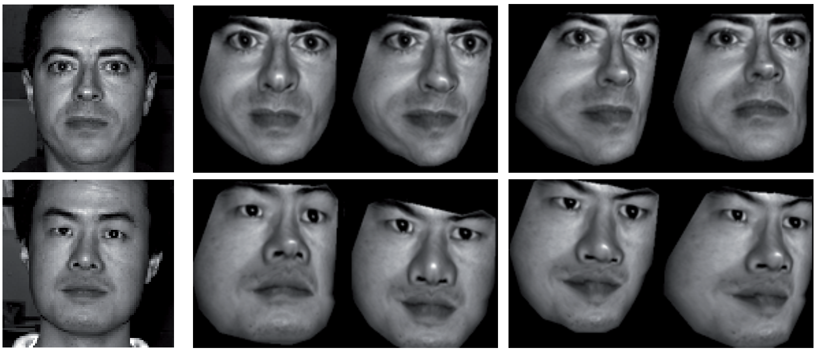
Shape of any unseen individual at pose p can now be easily obtained from frontal shape using shape-ratio vector r_p as

$$\hat{S}^p_{unseen} = r_p S^0_{unseen}. \tag{9}$$



(a) (b)

Fig. 4. Pose variations and synthetically generated landmarks. a) Texture and shape triangles variations due to pose variations, b) Synthetically generated landmark points given in the first and the second rows are generated by using the ratio matrix obtained from the landmarks in the database.



(a) (b)

Fig. 5. Synthetic pose generation from frontal face: a) Frontal face, b) Synthetically generated non-frontal faces.

Shapes in the database for $p=8$ different poses can be synthesized from frontal-view images as,

$$\hat{S}_i^p = r_p S_i^0 \text{, } i=1,2,\ldots,10, \text{ and } p=1,2,..,8. \tag{10}$$

AAM shape component is constructed from these aggregated shapes, $\hat{S}_i^p$ and S_i^0, by applying principal component analysis as $S=\overline{S}+Q_s s$ where $\overline{S}$ is the mean shape, Q_s contains k eigenvector of the covariance matrix corresponding to the highest k eigenvalues. Next step is to wrap each face in the training database to mean shape ($\overline{S}$) and apply the principal component analysis to the texture, this time as $T=\overline{T}+Q_t t$ where $\overline{T}$ is called as mean face. Any shape (S) and texture (T) can be steadily mapped to the AAM subspace as $s=Q_s^T\left(S-\overline{S}\right)$ and $t=Q_t^T\left(T-\overline{T}\right)$.

AAM is comprised of both shape (Q_s) and texture (Q_t) subspaces. Any change in face shape leads to a change in face texture and vice versa. Face appearance (A) is dependent on shape and textures. This dependency is expressed as A=[Λs t]^T. In order to exploit the dependency between shape and texture modeled by the diagonal matrix (Λ), one further PCA is applied to the shape and texture components collectively and we obtained the combined model called appearance model as A=Q_aA. Any appearance is obtained by a simple multiplication as $a = Q_a^T A$.

In order to show how rich representation AAM provides us, we used the first 5 coefficients and select random points in 5-dimensional space. The corresponding faces are plotted in Fig.6. Even this simple experiment proves that AAM trained as explained above can generate pose variations not governed by any shape ratio vector (r_p).

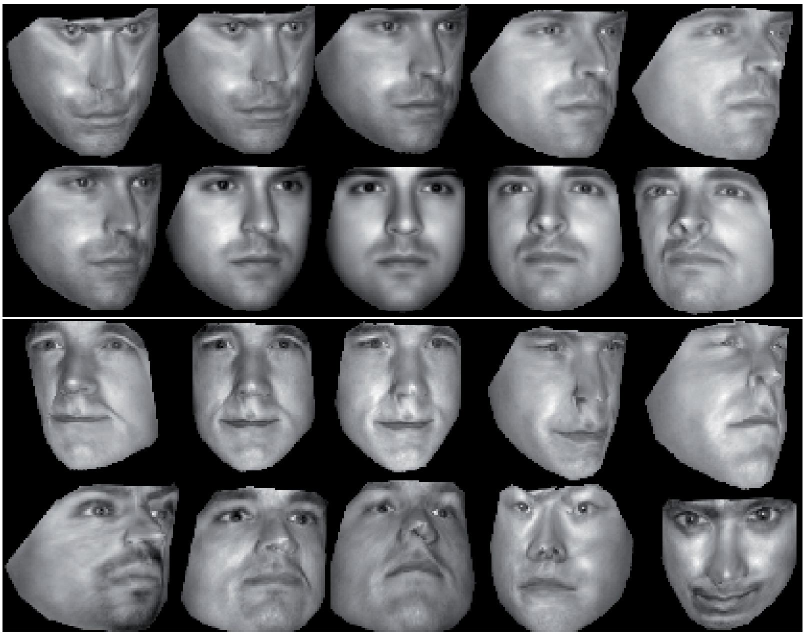


Fig. 6. Randomly synthesized faces from leading 5 AAM parameters.

5. Experimental Results

AAM combines the shape and texture model in one single model. The alignment algorithm (also called AAM searching) optimizes the model in the context of a test image of a face. The optimization criterion is the error occurring between a synthesized face texture and the corresponding texture of the test image.

Due to the illumination problems the error can be high and the classic searching algorithm fails. In the proposed approach, we normalize the corresponding texture in the test image just before we compute the error. We tested the proposed method on the Yale-B face dataset (Georghiades et al., 2001). The total number of images under different lighting conditions for each individual is 64. The database is portioned into four sets identified as Set 1-4. Set 1 contains face images whose light direction is less than ±20 degrees. Set 2 contains face images whose light directions are between ±20 and ±50 degrees. Set 3 contains face images whose light directions are between ±50 and ±70 degrees. Set 4 contains face images whose light directions are greater than ±70 degrees. All details about the Yale B dataset are given in (Georghiades et al., 2001). We manually labeled 4920 images. To establish the models, 73 landmarks were placed on each face image; 14 points for mouth, 12 points for nose, 9 points for left eye, 9 points for right eye, 8 points for left eyebrow, 8 points for right eyebrow and 11 points for chin. The warped images have approximately 32533 pixels inside the facial mask. We constructed a shape space to represent 95% of observed variation. Then we warped all images into the mean shape using triangulation. Using normalized textures, we constructed a 21-dimensional texture space to represent 95% of the observed variation in textures and for shapes we constructed a 12-dimensional shape space to represent 95% of the observed variation in shapes. Finally, we constructed a 15-dimensional appearance space to represent 95% of the total variation observed in the combined (shape and texture) coefficients.

Using a ground truth given by a finite set of landmarks for each example, performance can be easily calculated. A distance measure $D(x_{gt}, x)$ is computed in a leave-one-out setting, and it gives a scalar interpretation of the fit between the two shapes, i.e. the ground truth (x_{gt}) and the optimized shape (x). Two distance measures defined over landmarks are used to obtain the convergence performance. The first one is called point-to-point error, defined as the Euclidean distance between each corresponding landmark:

$$D_{pt.pt.} = \Sigma \sqrt{\left(x_i - x_{gt,i}\right)^2 + \left(y_i - y_{gt,i}\right)^2} \tag{11}$$

The other distance measure is called point-to-curve error, defined as the Euclidean distance between a landmark of the fitted shape (x) and the closest point on the border given as the linear spline, $r(t)=(r_x(t),r_y(t)),t\in[0,1]$, of the landmarks from the ground truth (x_{gt}):

$$D_{pt.crv.} = \frac{1}{n}\sum_{i=1}^{n} \min_t \sqrt{\left(x_i - r_x(t)\right)^2 + \left(y_i - r_y(t)\right)^2} \tag{12}$$

We calculated these errors on all images in the datasets (from Set 1 to Set 4). We conducted an experiment to see how close we fit into unseen faces at different poses.

To match a given face image with the model, an optimal vector of parameters are searched by minimizing the difference between synthetic model image and input image. Fig.7 illustrates the optimization and search procedures for fitting the model to input image. The first column of the figure is the arbitrarily illuminated unseen image from test dataset and the remaining images (columns) are the steps of the optimization. The fitting results are rendered at each iteration for classical AAM (the first row) and the proposed method (the second row).

The AAM searching is known to be very sensitive to the selection of initial configuration. We tested the proposed method against the selection of initial configuration. We translate, rotate and scale initial configurations and see how the proposed method can handle the poor initialization. We made 10 experiments for each test image with different initializations

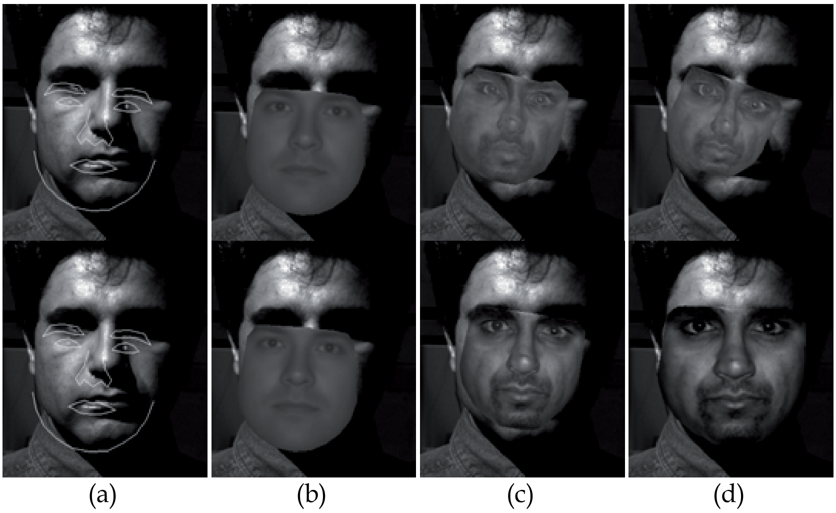


Fig. 7. Searching results: The first row is the classical AAM searching results and the second row is the proposed method. (a) Initial configuration (b) Mean face (c) Searching result obtained in the third iteration (d) Searching result obtained in the sixth iteration.

and took the average error. These experiments include mean-shape configuration, ±5 degrees rotation, scaling by 0.85 and 0.95, translation by 10% in x and y directions.

Table.1 summarizes the averages of point-to-point and point-to-curve errors when classical AAM search is used without any illumination normalization. Point-to-point and point-to-curve errors obtained by the proposed illumination normalization method are much less than the errors obtained by the classical AAM (Table.2).

	Yale B face database subsets			
	Subset$_1$	Subset$_2$	Subset$_3$	Subset$_4$
Pt.-pt.	4.9±0.2	11.4±0.5	19.4±0.58	36.6±1.6
Pt.-crv.	2.9±0.1	6.8±0.33	12.9±0.36	33.2±1.4

Table 1. Standard AAM fitting performance.

	Yale B face database subsets			
	Subset$_1$	Subset$_2$	Subset$_3$	Subset$_4$
Pt.-pt.	4.1±0.12	8.06±0.3	13.03±0.4	21.3±0.5
Pt.-crv.	2.4±0.08	5.24±0.23	8.76±0.3	14.7±0.4

Table.2 Proposed AAM fitting performance.

Ratio-image method is not suitable for AAM searching, at least for the first iterations of the algorithm. Let's suppose that we start searching in a position far away from the ground truth location. The model synthesizes a face that best fits the current location. Then the textures of the synthesized face and corresponding part in the test image are analyzed and an error coefficient is computed, reflecting the similarity degree of the two textures. We normalize the corresponding texture in the test image before computing the error. The main problem with the ratio-image method is that when it is applied to a region of an image that is not face-like, the normalization result will include a lot of information of the mean face, in other words, it will be mean-face-like. Thus the error will be much smaller than the real one, and it will introduce false alarm in the searching process creating additional local minima. On the other hand, the histogram based normalization method will never change the general aspect of an image, only the pixel intensities follow a different distribution. Thus the chances of introducing false alarms are reduced using this normalization method. The ratio-image can produce very good results provided that the shape is already aligned. But this is not the case in AAM searching. We assume that the best fit returned by the searching algorithm using histogram-based normalization is a good approximation of the real face, and thus the alignment requirement is satisfied. Fig.8 summarizes the alignment results for these unseen faces.

We also analyze how the proposed alignment method affects the recognition performance. We used the following feature spaces in our experiments: PCA and LDA. Randomly selected *25* images of each person from Set 1 dataset are used in training. All datasets (Set 1 through Set 4) contain faces of all poses. The remaining faces in Set 1 dataset are used as test data. Recognition rates for two feature spaces (i.e. PCA and LDA) in Set 1-4 are plotted in Fig.10 for increasing dimensions. The recognition rates obtained when the original images are used as input to the classifier are denoted as ORG-PCA and ORG-LDA. The recognition rates obtained when the images restored by RI are used as input and are denoted as RI-PCA and RI-LDA. Finally, the recognition rates obtained when the images restored by HF are used as input and are denoted as HF-PCA and HF-LDA. PCA is known to be very sensitive to misalignment in faces. Our experimental studies also verify this behavior. When the original images are used, the PCA recognition rates for all sets are poor. LDA is more successful if dimension is closer to 9. ORG-PCA reaches to 74.36% at most, while ORG-LDA reaches to 91.26% at most in Set 1. This performance drops to 30.99% for ORG-PCA and to 41.13% for ORG-LDA in Set 4.

One important observation is that AAM alignment with histogram fitting always leads to better recognition rates in all test sets (Set 1- 4) compared to the case where original faces are used and ratio-image normalization is used right after the AAM alignment. Another advantage of the proposed method is that similar recognition performance is obtained at lower dimensions. Recognition rate for ORG-LDA is just 32.81% while LDA performance for the proposed approach (called HF-LDA) is 83.38% when the dimension is set to 3. ORG-LDA catches this rate when the dimension is set to 5.

For the challenging test set, i.e. Set 4, both ORG-LDA and ORG-PCA fail. The recognition rate is at most 30.99% for ORG-PCA and 41.13% for ORG-LDA. On the other hand, HF-PCA reaches to 76.20% at most and HF-LDA reaches to 82.68% at most. This is a significant improvement when compared to the results obtained without applying any preprocessing (41%). Note that all test sets include faces of 8 different poses selected from Yale B dataset.

6. Conclusion

In this study we developed AAM based on face alignment method which handles illumination and pose variations. The classical AAM fails to model the appearances of the same identity under different illuminations and poses. We solved this problem by inserting histogram fitting into the searching mechanism and inserting synthetically generated poses of the same identity into the training set. From the experimental results, we showed that the proposed face restoration scheme for AAM provides higher accuracy for face alignment in point-to-point error sense. Recognition results based on PCA and LDA feature spaces showed that the proposed illumination and pose normalization outperforms the standard AAM.

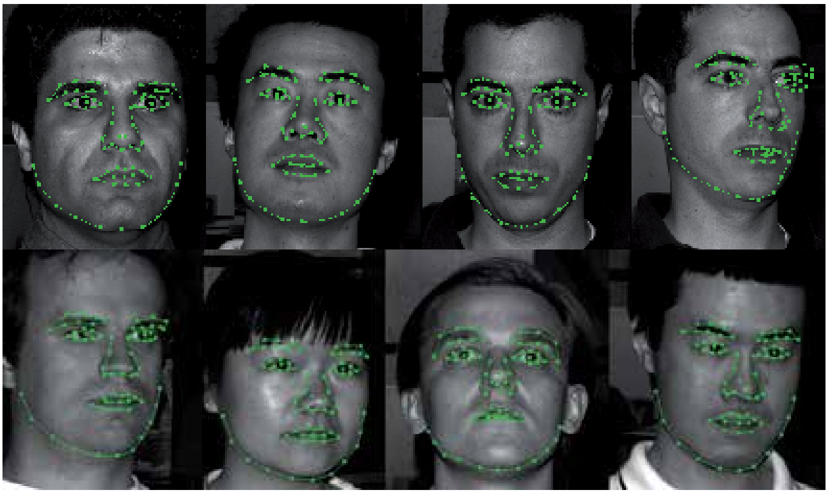


Fig. 8. Face alignment results for unseen faces.

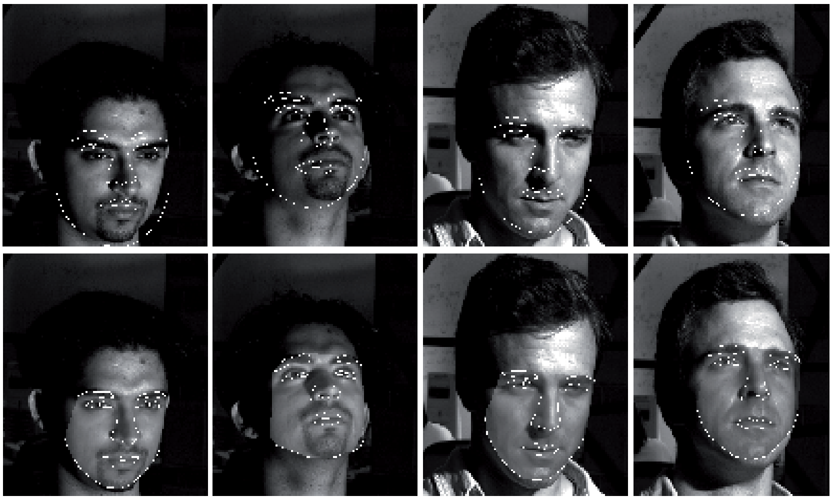
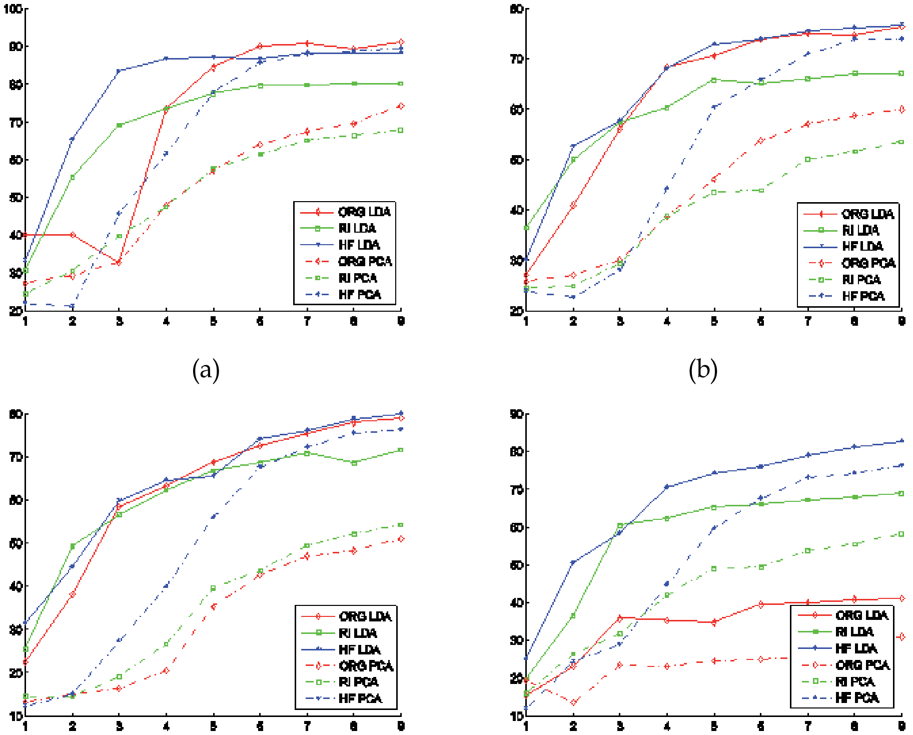


Fig. 9. Initialization (the first row) and alignment/restoration results of the proposed method (the second row) for different pose and illumination variations.

7. Acknowledgement

This work was partly supported by the National Scientific and Research Council of Turkey; project no: EEEAG-104E121 and the State Planning Agency (DPT) of Turkey. We would like to thank Florin S. Telcean for his initial contributions to this work.



(a) (b)

(c) (d)

Fig. 10. PCA and LDA recognition rates for Set 1 (a), Set 2 (b), Set 3 (c), and Set 4 (d) when original face (ORG), Ratio Image (RI) and the proposed restoration (HF) are used.

8. References

Zhao, W. & Chellappa R. (2000). SFS Based View Synthesis for Robust Face Recognition. *Proc. 4th Conf. on Automatic Face and Gesture Recognition*, 2000.

Georghiades, A.S.; Belhumeur, P. N., & Kriegman, D. J. (2001). From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no.6, pp.643–660, 2001.

Ramamoorthi R. (2002). Analytic PCA Construction for Theoretical Analysis of Lighting Variability in Images of a Lambertian Object", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 10, 2002.

Basri, R. & Jacobs, D. (2001). Photometric Stereo with General, Unknown Lighting. *Proceedings of IEEE Conf. on Computer Vision and Pattern Recognition* (CVPR), vol. 2, pp. 374-381, 2001.

Zhou S.; Aggarwal G.; Chellappa R. & Jacobs D. (2007). Appearance characterization of linear lambertian objects, generalized photometric stereo, and illumination invariant face recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 230-245, 2007.

Lee J.; Moghaddam B.; Pfister H. & Machiraju R. (2005). A bilinear illumination model for robust face recognition. IEEE Conf. on ICCV, pp. 1177-1184, 2005.

Shashua A. & Riklin-Raviv T. (2001). The Quotient Image: Class-Based Re-Rendering and Recognition With Varying Illuminations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 129-139, 2001.

Wang H.; Li S. Z & Wang Y (2004). Generalized quotient image, *IEEE Proceeding Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 498-505, 2004.

Zhang Y.; Tian J.; He X. & Yang X. (2007). MQI Based Face Recognition Under Uneven Illumination, *Advances in Biometrics*, vol. 4642, pp. 290-298, 2007.

An G.; Wu J. & Ruan Q. (2008). Kernel TV-Based Quotient Image Employing Gabor Analysis and Its Application to Face Recognition", IEICE *Transactions on Information and Systems*, vol. E91-D, no. 5, pp. 1573-1576, 2008.

Sim T. & Kanade T. (2001). Combining models and exemplars for face recognition: An illuminating example, *IEEE Proceeding Conference on Computer Vision and Pattern Recognition, Workshop on Models versus Exemplars in Computer Vision*, 2001.

Cootes, T.F.; Taylor, C.J.; Cooper, D.H. & Graham, J. (1995). Active Shape Models-their training and application. *Computer Vision and Image Understanding*, 61(1), pp. 38-59, 1995.

Cootes, T.F.; Edwards, G. & Taylor, C.J. (2001). Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681-685, 2001.

Stegmann, M.B.; Ersboll, B.K. & Larsen, R. (2003). FAME - A Flexible Appearance Modeling Environment. *IEEE Trans. on Medical Imaging*, vol. 22, no.10, pp.1319-1331, 2003.

Hou, X.; Li, S.; Zhang, H. & Cheng, Q. (2001). Direct appearance models. *Proceedings of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 828–833, 2001.

Cootes, T.F.; Wheeler G.; Walker, K. & Taylor, C. (2002). View based active appearance models. *Image and Vision Computing*, vol. 20, pp.657–664, 2002.

Blanz V. & Vetter T. (2003). Face recognition based on fitting a 3D morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence, v*ol. 25, vo.9, pp. 1063-1074, 2003.

Goodall, C. (1991). Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society*, vol.B53, no.2, pp. 285-339, 1991.

Glasbey, C.A. & Mardia, K.V. (1998). A review of image warping methods. Journal of Applied Statistics, Vol. 25 (2), 155-171, 1998.

Jebara, T. (1996). 3D Pose Estimation and Normalization for Face Recognition, B. Thesis, McGill Centre for Intelligent Machines, 1996.

Shewchuk J.R. (1996). Triangle: engineering a 2D quality mesh generator and Delaunay triangulator. *Workshop on Applied Computational Geometry. Toward Geometric Engineering.*, 203–222. Springer-Verlag, 1996.

Zhao, W. & Chellappa R. (2006). *Face Processing: Advanced Modeling and Methods*. Academic Press, Elsevier, 2006.

Liu, D.H.; Lam, K.M. & Shen, L.S. (2005). Illumination invariant face recognition. *Pattern Recognition*, vol. 38, no.10, pp. 1705-1716, 2005.

14

Eye Movements in Face Recognition

Janet H. Hsiao
University of Hong Kong
Hong Kong

1. Introduction

In human vision, visual acuity falls off rapidly from the center of fixation to the periphery. Hence, in visual perception, we actively change our gaze directions in order to bring relevant information into foveal vision, where the highest quality visual information can be obtained. In recent years, researchers have shown that eye movements during visual perception are linked to the underlying cognitive processes (e.g., Hayhoe & Ballard, 2005; Henderson, 2003; Kowler, 1990). This phenomenon has especially been extensively demonstrated in the research on reading and word recognition (Rayner, 1998; Sereno & Rayner, 2003). For example, the existence of preferred landing positions (PLP, Rayner, 1979) in sentence reading and optimal viewing positions (OVP, O'Regan et al., 1984) in isolated word recognition has been consistently reported. The preferred landing location refers to the location where eye fixations fall the most often during reading, whereas the optimal viewing position refer to the location where the initial eye fixation is directed to when the best recognition performance is achieved. For English words, both the preferred landing location and the optimal viewing position have shown to be to the left of the center of the words. These locations have been argued to reflect an interplay among several different variables, including difference in visual acuity between foveal and peripheral vision, information profile of the words, influence of perceptual learning, and hemispheric asymmetry (Brysbaert & Nazir, 2005; Rayner, 1998).

Similar to word recognition, the recognition of faces is an over-learned skill that we have constantly performed since birth, even earlier than the time we started to read. Nevertheless, in contrast to research on reading, the understanding of the role of eye movements during face recognition remains limited. Recent research on face recognition has suggested a dissociation between face and object recognition. Faces have been argued to be represented and recognized holistically; the recognition of faces has been shown to involve relatively less part-based shape representation compared with the recognition of objects (e.g., Farah et al., 1995; Tanaka & Farah, 1993). Since we process faces holistically, there is a concern whether we need eye movements at all during face recognition; we may just need a single eye fixation to recognize a faces, and its location may not influence our performance since faces are represented and recognized holistically. This statement has recently been shown to be wrong. Some studies have suggested that performance in face recognition is related to eye movement behavior. For example, Henderson et al. (2005) examined the influence of restricting eye fixations during face learning on the performance of the subsequent face recognition task, and showed that eye movements generated during face recognition have a functional role and are not just recapitulation of those produced during learning (see also Mäntylä & Holm, 2006). In my recent study (Hsiao & Cottrell, 2008), we restricted the number of fixations that participants were allowed to make during face recognition, and showed that the optimal recognition performance was achieved with two fixations, with the first fixation just to the left of the center of the nose and the second on the center of the nose. These studies suggest that eye fixations during face recognition do have functional roles; they also reflect the underlying cognitive processes involved in face recognition. I will review these studies in the following section.

In recent years, some researchers have proposed computational models of face recognition that incorporate eye fixations, in order to more realistically account for the cognitive processes involved in face recognition. For example, Lacroix et al. (2006) proposed the Natural Input Memory (NIM) model that uses image patches at eye fixation points as the internal representation for modeling face recognition memory. Barrington et al. (2008) later proposed a Bayesian version of the NIM model (the NIMBLE model, NIM with Bayesian Likelihood Estimation). In section three, I will review these two models and discuss their cognitive plausibility in addressing human face recognition behavior.

In face perception, a left side bias has been consistently reported, in both perceptual judgments (e.g. Gilbert & Bakan, 1973) and eye movements (e.g., Everdell et al., 2007). This phenomenon has been argued to be an indicator of right hemisphere (RH) involvement in the perception of faces (e.g., Burt & Perrett, 1997; Rossion et al., 2003). A recent study of mine suggests a link between this left side bias and visual expertise (Hsiao & Cottrell, 2009). In section four, I will review these studies about the left side bias in face perception, and discuss their implications for visual expertise processing.

Eye movements during face recognition have also been used to examine the processing differences between familiar and unfamiliar face recognition. Previous research has suggested that internal facial features (e.g. eyes and nose), as opposed to external features (e.g. hair and facial shape), are more important in the recognition of familiar faces compared with unfamiliar faces (e.g. Ellis & Shepherd, 1992). Nevertheless, Althoff and Cohen (1999) compared eye movements during familiar and unfamiliar face recognition in a familiarity judgment task, and showed that there was no difference between familiar and unfamiliar faces in the number of fixations falling into the internal face region. In a more recent study, Stacey et al. (2005) showed that participants made more fixations on the internal features when viewing familiar faces compared with unfamiliar faces only in a face matching task, but not in a familiarity judgment task or a standard recognition memory task. In section five, I will review these studies and discuss the processing differences between familiar and unfamiliar face recognition. Finally, in the conclusion section, I will give a summary of the chapter and some perspectives for future research directions.

2. Functional Roles of Eye Movements in Face Recognition

Henderson et al. (2005) examined whether eye movements in face recognition have a functional role or just a recapitulation of those produced during face learning. Participants performed a standard face recognition memory task while their eye movements were recorded. During the learning phase, they were presented with face images one at a time and asked to remember them; after a short break, during the recognition phase, they were presented with face images one at a time and asked to judge whether they saw the face image during the learning phase. They manipulated participants' eye movements during the learning phase: in the free-viewing learning condition, participants were allow to move their eyes naturally; in the restricted-viewing learning condition, participants had to remain looking at a single central location when they viewed the faces. They examined the influence of this eye fixation restriction during the learning phase on their performance in the subsequent face recognition task. Their results showed that this eye fixation restriction during face learning significantly impaired participants' recognition performance compared with the free-viewing learning condition. Also, they showed that the fixation locations participants made during the recognition phase following the free-viewing and the restricted-viewing conditions were similar to each other. Their results thus suggest that eye movements generated during face recognition have a functional role and are not just recapitulation of those produced during learning.

Mäntylä and Holm (2006) conducted a similar face recognition study in which they restricted participants' eye movements during either the learning phase or the recognition phase, or both phases. In the trials in which participants had correct responses, they asked participants to report whether they indeed remembered seeing the face during the learning phase (i.e. recollection), or they only knew that the face was presented earlier (i.e. familiarity). They found that the restriction of eye movements impaired participants' explicit recollection, but not familiarity-based recognition.

Henderson et al.'s (2005) and Mäntylä and Holm's (2006) studies demonstrate that eye movements in face recognition have a functional role, since restricting eye movements during face recognition significantly impair participants' recognition performance. However, it remains unclear what the function roles of eye movements in face recognition are. For example, do we have preferred eye fixations during face recognition? What is the nature of these preferred fixations? Does the number of eye fixations we make during face recognition influence our performance? How many fixations do we need to recognize a face? Do we have better recognition performance with more fixations?

In a recent study (Hsiao & Cottrell, 2008), we aimed to answer these questions regarding the functional roles of eye movements in face recognition. Instead of restricting eye fixation locations during either face learning or recognition phases, we restricted the number of eye fixations participants were allowed to make during the recognition phase. We recruited Caucasian participants and had them perform a face recognition memory task with Caucasian face images (i.e. own-race face recognition). During the learning phase, they viewed 32 face images one at a time, each for three seconds. During the recognition phase, they viewed the same 32 face images and another set of 32 new face images one at a time, and asked to judge whether they saw the face during the learning phase (i.e. old/new judgments). Four different restriction conditions were created for the recognition phase: in the unrestricted condition, participants viewed the face image for three seconds at most or until their response if they responded within the three seconds; in the one, two, or three fixation conditions, they were only allowed to make one, two, or three fixations on the face image; the face image was covered by an average face image, a pixel-wise average of all face images in the materials, when their eyes moved away from the last permissible fixation (Fig. 1). We used the average face image as a mask when participants reached their maximum number of fixations allowed to create a smooth transition between the target face image and the mask; the average face image did not contain any identity information. During the experiment, participants were not aware of the relationship between the number of fixations they made and the time the average-face mask appeared; even if they were aware of this manipulation in the middle of the experiment, they were not able to anticipate the number of permissible fixations in each trial since the condition order was randomized during the experiment.

A major difference between our study and Henderson et al.'s (2005) and Mäntylä and Holm's (2006) studies was that the restriction was on the number of fixations as opposed to fixation locations; in our experiment, during both phases, they were allowed to move their eyes naturally. Thus we were able to examine the functional roles of eye fixations during face recognition and answer questions such as how many fixations we need in order to recognize a face and where they are located. Another difference between our study and previous examinations of eye movements in face recognition was that, in order to examine participants' preferred landing positions in face recognition, instead of having participants start a trial from the face centre, we started each trial with a central fixation, followed by the target face image presented either above or below the central fixation (randomly determined; Fig. 1). Thus, in each trial, participants had to make a saccade from the centre of the screen onto the target face image. If participants started a trial from the centre of the face (e.g., Henderson et al., 2005), their first fixation would always be away from the face centre and it would be impossible to know where the preferred landing position of their first fixation was when they viewed a face. In addition, in each trial in order to prevent participants from obtaining face identity information from parafoveal vision before they made their first fixation, the average-face mask was first presented and then replaced by the target face image after participants made a saccade from the initial central fixation (Fig. 1).

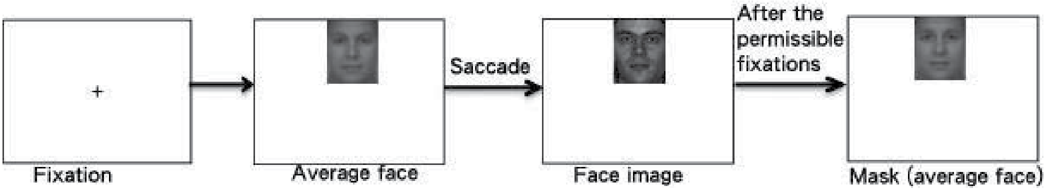


Fig. 1. Flow chart of a test trial during the recognition phase (Hsiao & Cottrell, 2008).

We used A′, a bias-free nonparametric measure of sensitivity, as the measure of participants' discrimination sensitivity in the study. Our results showed that participants were able to recognize a face with a single fixation (the average A′ was 0.63; A′ at the chance level is 0.5). Participants had better performance when they were allowed to make two fixations; there was no further performance improvement when they were allowed to make more than two fixations (Table 1). This result suggests that two fixations suffice in face recognition. These two fixations were at the centre of the nose, with the first fixation slightly to the left of the centre (Fig. 2), suggesting that the centre of the nose is the preferred landing location in face recognition.

In this experiment participants had better performance in the two-fixation condition compared with the one-fixation condition. This performance difference may be due to a longer total face viewing time in the two-fixation condition compared with the one-fixation condition. To address this issue, we conducted another experiment in which the total face viewing time was fixed to be 610 ms, the sum of the average durations of the first two fixations in the previous experiment. In the one-fixation condition, after participants made the first fixation, the face image moved with their gaze (i.e., the display became gaze contingent); thus, they would keep looking at the same location as their first fixation on the face image. In the two-fixation condition, the face image became gaze contingent after they made a second fixation. Participants were unaware of the gaze contingent design during the experiment. Our results showed that given the same total face viewing time, participants still had better performance when they were allowed to make two fixations compared with a single fixation (Table 2). This result suggests that the second fixation has functional significance: to obtain more information from a different location, but not just to increase the total face viewing time.

	One fixation	Two fixations	Three fixations	Unrestricted
Mean A′	0.625	0.826	0.812	0.799
Standard Deviation	0.125	0.083	0.140	0.127

Table 1. Mean A′ (a discrimination sensitivity measure) and standard deviations in the four fixation conditions (Hsiao & Cottrell, 2008).

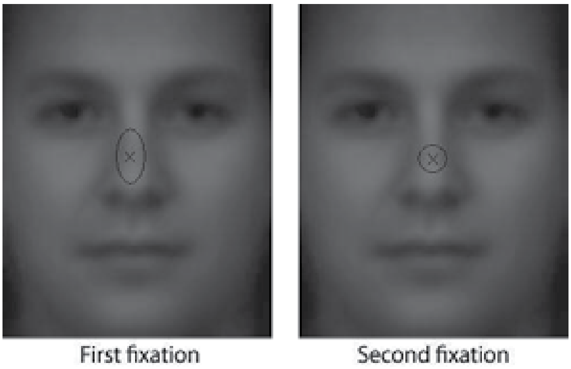


Fig. 2. Average eye fixation locations of the first two fixations during the recognition phase. The radii of the ellipses show standard deviations of the locations. The background shows the average face (Hsiao & Cottrell, 2008).

	One fixation	Two fixations
Mean A′	0.787	0.853
Standard Deviation	0.108	0.074

Table 2. Mean A′ (a discrimination sensitivity measure) and standard deviations in the one- and two- fixation conditions in the second experiment (Hsiao & Cottrell, 2008).

Previous studies examining the diagnostic features in face recognition using the Bubbles procedure (e.g., Gosselin & Schyns, 2001, Schyns et al., 2002, Vinette et al., 2004) showed that the most diagnostic features for face identification are the eyes. Standard approaches to modeling human eye fixations and visual attention usually use a saliency map that is calculated according to biologically motivated feature selection or information maximization (e.g., Itti et al., 1998; Bruce & Tsotsos, 2005; Yamada & Cottrell, 1995). These models would predict fixations on the eyes when we view face images. Our results (Hsiao & Cottrell, 2008) showed that this was not the case: the preferred landing position was not located on the most informative position on a face (i.e., the eyes). This phenomenon suggests that eye fixation behavior in face recognition is different from that during scene viewing or visual search tasks. Indeed, recent research on face recognition suggests that the recognition of faces is holistic and involves relatively less part-based shape representation compared with the recognition of objects (e.g., Farah et al., 1995). Some argue that this face-specific effect is in fact expertise-specific (e.g., Gauthier & Tarr, 1997, 2002; Gauthier et al., 1998; Gauthier et al., 1999). Our result is consistent with this view. It is possible that, due to our familiarity with the information structure of faces, fixations at each individual feature fragments only generate redundant processes and increase the processing time; instead, a more efficient strategy is to get as much information as possible with a single fixation. Given a perceptual window large enough to cover a whole face and the fact that visual acuity drops dramatically from the fovea to the periphery, the fixation from which the most information can be obtained should be at the "center of the information", where the information is balanced in all directions. This location may also be the optimal viewing position for the recognition of faces. This claim is consistent with the observation that face recognition tends to be more holistic compared with the recognition of objects.

In summary, in this study (Hsiao & Cottrell, 2008), we showed that two fixation suffice in face recognition. The distributions of these two fixations were around the centre of the nose, suggesting that this location is the preferred landing position in face recognition. We argue that this location may be the centre of the information for face recognition; it may also be the optimal viewing position for face recognition. Further research is required to examine these hypotheses.

3. Incorporating Eye Fixations in Computational Models of Face Recognition

In computational modelling of cognitive processes, several models have been proposed to address human behaviour in face recognition. Recently researchers started to incorporate eye fixations into their computational models of face recognition in order to more accurately modelling the cognitive processes involved in face recognition. For example, Lacroix et al. (2006) proposed the Natural Input Memory (NIM) model to account for human behaviour in face recognition memory. The model uses fixation-based face fragments and transforms these fragments into a feature-vector representation as the internal representation of recognition memory. Thus, memories can be stored as points in a vector space, and recognition processes can be modelled as comparing the currently perceived fragments to the stored fragments: the larger the distance between the two representations in the vector space, the harder the memory can be successfully recollected. The NIM model can be considered as an exemplar model of memory (Raaijmakers & Shiffrin, 2002). However, the NIM model differs from standard mathematical psychology models in that (1) it uses actual facial images as input, and (2) it is based on the idea of storing fixation-based face fragments, rather than whole face exemplars (e.g., Dailey & Cottrell, 1999; O'Toole et al., 1988). In accounting for human behaviour, Lacroix et al. (2006) showed that the NIM model was able to simulate experimentally obtained human similarity ratings and recognition memory for individual faces.

Barrington et al. (2008) further proposed a Bayesian version of the NIM model, (which was referred to as NIMBLE, NIM with Bayesian Likelihood Estimation), as a general framework that is able to handle multi-class problems. The model was able to achieve human-level performance on standard face recognition tasks and also performed multi-class face and object identification tasks with high accuracy. The Bayesian combination of individual fragment likelihoods in the NIMBLE model outperformed the combination method in the original NIM model in most cases; in addition, if using new kernels for density estimation, the NIMBLE model was shown to far outperform the NIM model.

In accounting for human behaviour, the model was able to achieve correct face recall and identification with a very small number of fixations; on average, consistent with our human data (Hsiao & Cottrell, 2008), a single fixation was enough to recognize a face. Nevertheless, inconsistent with the human data, the probability of successful recognition increased with an increasing number of fixations; in contrast, the human performance levelled off after two fixations (Hsiao & Cottrell, 2008). This inconsistency may be due to the difference in the choice of eye fixation locations between the model and the human data: the NIMBLE model implemented a model of visual saliency (Yamada & Cottrell, 1995) as the way to select fixation points; as the result, the eyes, instead of the centre of the nose shown in the human data, were usually selected as the first fixation points. Thus, in order examine the cognitive plausibility of the NIMBLE model in modelling recognition memory alone, in another examination we directly used human eye fixation locations obtained in Hsiao and Cottrell (2008) in the NIMBLE model. The results showed that by using human fixation locations, with a single fixation the NIMBLE model already achieved a similar performance level (with ROC area between 0.8 and 0.9) to the best performance in the human data (i.e., with two fixations), and more fixations did not further improve the model's performance. In other words, the NIMBLE model achieved maximum performance using just the first human fixation. This result is consistent with our claim that the first fixation location chosen by humans (i.e., the preferred landing position, the center of the nose) may be the optimal viewing position for face recognition (Hsiao & Cottrell, 2008).

A possible explanation for the discrepancy between the human data (Hsiao & Cottrell, 2008) and the NIMBLE model's behaviour using the same fixations is that, as shown in Fig. 2, the first and second fixations in the human data tended to be in very similar locations (around the centre of the nose). Recall that in the human data, the participants achieved their maximum performance with these two fixations. This phenomenon suggests that all of the information required for face recognition may be obtainable by looking at the centre of the nose, but perhaps the participants were not able to obtain all of the information required during the duration of a typical fixation. Since we move our eyes about three times per second (Henderson, 2003) (in our human data, the average first fixation duration was 295 ms and the second was 315 ms on average), it may be that a second fixation in a nearby location is required to accumulate more information to achieve the best face recognition performance. This limitation in human vision may be explained by a task-switching cost from localizing to exploring for recognition; that is, in the experiment, participants had to plan a localizing saccade from the centre of the screen to the target face stimulus before the first fixation, and then switch the task from localizing to exploring for recognition afterwards (Tatler, 2007). In contrast to human vision, the NIMBLE model does not have this limitation. In addition, this localizing fixation has been shown to have a central bias (e.g., Renninger et al., 2007; Tatler, 2007), regardless of image feature distribution (Tatler, 2007). Thus, the first fixation in face recognition may be influenced by both this central bias and the tendency to look at the optimal viewing position due to our expertise.

In summary, the results of the NIMBLE model, which incorporates eye fixations in modelling cognitive processes involved in face recognition, support our hypothesis that the preferred landing position in face recognition (the centre of the nose; Hsiao & Cottrell, 2008) may also be the optimal viewing position for face recognition. The modelling results also suggest some possible limitations of the human visual system; further research is required to examine these hypotheses.

4. Left Side Bias in Face Perception and Its Link to Visual Expertise

In face recognition, a left side bias effect has been consistently reported. For example, a chimeric face made from two left half faces from the viewer's perspective has been reported to be judged more similar to the original face than one made from two right half faces (Gilbert & Bakan, 1973), especially for highly familiar faces (Brady et al., 2005). This phenomenon has been argued to be an indicator of right hemisphere involvement in the perception of faces (Burt & Perrett, 1997; Rossion et al., 2003).

This left side bias effect has also been shown to be reflected in eye movement behaviour. For example, in an eye movement study of face recognition, Mertens et al. (1993) reported an asymmetry in gaze-movement strategies for faces in a visual memory task: the overall time that the centre of gaze remained in the left side of the stimulus was longer than the right side; this asymmetry was not observed for vases. Leonards & Scott-Samuel (2005) showed that participants tended to have their initial saccade direction to the left side for face stimuli, but not for landscapes, fractals, or inverted faces. They hence attributed the observed initial saccade direction bias to internal cognition-related factors, i.e., familiarity of the stimuli. Their results also showed significantly shorter initial saccade latencies to the left half-field compared with those to the right half-field for participants who had the leftward bias, suggesting higher effectiveness through automatization. Vinette et al. (2004) used the Bubbles procedure (Gosselin & Schyns, 2001) to examine the timing of different face features used during face identification. They showed that the left eye was diagnostic between 47 ms to 94 ms after the stimulus onset, and both eyes became informative after 94 ms. Joyce (2001) also showed that during the first 250 ms in face recognition, participants' eye fixations tended to be on the left half-face. Consistent with these results, in a recent study (Hsiao & Cottrell, 2008), we also showed that during face recognition participants' first fixation tended to be slightly to the left of the centre (Fig. 2). These findings suggest that the left side of a face (from the viewer's perspective) may be more informative in face recognition. This hypothesis is consistent with the diagnostic face images obtained from the Bubbles procedure (Gosselin & Schyns, 2001, Schyns et al., 2002, Vinette et al., 2004), showing that the left eye is the most diagnostic point at an early stage of face recognition.

Why is the left side of a face from the viewer's perspective more diagnostic than the right side of a face in face recognition? Researchers have shown that low spatial frequency information is important for face recognition (e.g., Whitman & Konarzewski-Nassau, 1997). Also, the right hemisphere has been shown to have an advantage over the left hemisphere in tasks requiring low spatial frequency information processing (Sergent, 1982). Ivry and

Robertson (1999) proposed the Double Filtering by Frequency (DFF) theory, in which they argue that information coming into the brain goes through two frequency filtering stages: the first stage involves selection of a task-relevant frequency range; at the second stage, the right hemisphere biases information to low frequency ranges, whereas the left hemisphere biases information to high frequency ranges. Consistent with these findings, the right hemisphere has been shown to be important for face processing. For example, fMRI studies have shown that an area inside the fusiform gyrus (fusiform face area, FFA) responds selectively to faces (although some argue that FFA is an area for expertise in subordinate level visual processing instead of selective for faces, e.g., Tarr & Gauthier, 2000), with larger activation in the right hemisphere compared with the left hemisphere (e.g. Kanwisher et al., 1997). Electrophysiological data show that faces elicit larger Event Related Potential (ERP) N170 than other types of objects, especially in the right hemisphere (e.g., Rossion et al., 2003). Neuropsychological data also suggest a link between right hemisphere damage and deficits in face recognition and perception (e.g., Young et al., 1990; Evans et al., 1995)

Also, because of the partial decussation of the optic nerves, our visual field is vertically split and the two visual hemifields are initially contralaterally projected to the two hemispheres. Thus, when we are viewing a face and looking at the centre of the face, the left side of the face from our perspective has direct access to the right hemisphere. It has been shown that, when a stimulus is centrally fixated, each of the two hemispheres plays a dominant role in the processing of the half of the stimulus to which it has direct access (e.g., Lavidor et al., 2004; Lavidor & Walsh, 2004; Hsiao et al., 2006; although these are based on research on visual word recognition). Hence it is possible that the representation of the left side of a face is most often encoded by and processed in the right hemisphere, making it more informative than the right side of the face, which is usually processed in the left hemisphere. As the result, we may gradually direct our fixations more to the left, because the internal representation of the left stimulus-half is more informative and attracts our attention.

In addition to face processing, recent research on visual expertise has suggested that this left side bias may be related general visual expertise processing. For example, it has been shown that the increase of holistic processing effect for artificial objects after expertise training was correlated with right fusiform area activity (Gauthier & Tarr, 2002; Gauthier et al., 1999), suggesting that the low spatial frequency biased representation developed in the right hemisphere (Ivry & Robertson, 1998; Sergent, 1982; Hsiao et al., 2008) may be crucial for the development of visual expertise. In our recent study that examines visual expertise in Chinese character recognition (Hsiao & Cottrell, 2009), we showed that Chinese readers (i.e., experts) had a left side bias in the perception of mirror-symmetric characters, whereas non-Chinese readers (i.e., novices) did not; this effect was also reflected in participants' eye fixation behaviour: the distribution of Chinese readers' fixations when viewing the characters was significant to the left of the distribution of non-Chinese readers' fixations. In another study (Hsiao et al., in preparation), we trained participants to recognize other-race faces and a novel type of objects, Greebles (Gauthier & Tarr, 1997), through either individual-level (i.e. recognize them by individual names) or categorical-level (i.e. recognize their races/categories) recognition training; participants performed a standard recognition memory task once before and once after the training, while their eye movements were recorded. Our results showed that during the recognition phase of the recognition memory task, after training participants' second and third fixations significantly shifted leftwards in both face and Greeble recognition, compared with their pre-training behaviour. These findings thus suggest a link between the left side bias effect and visual expertise. In addition, in our study (Hsiao et al., in preparation), whereas in other-race face recognition this leftward shift in eye fixations was observed after both individual- and categorical-level training, in Greeble recognition it was only observed after the individual-level training. This phenomenon may suggest a more effective learning (i.e. learning to use the more informative, right hemisphere/low spatial frequency biased representation) in a domain in which participants already had prior perceptual knowledge (i.e. other-race face recognition) as opposed to a completely novel domain (i.e. Greeble recognition); it could also be because faces were automatically and unconsciously recognized at the individual level during the categorical-level training.

In addition to hemisphere asymmetry in face processing, there may be other factors that also account for this left side bias in face recognition. For example, Heath et al. (2005) showed that the left side bias effect in the perception of facial affect was influenced by both laterality and script reading direction. They showed that right-handed readers of Roman script demonstrated the greatest mean leftward bias, Arabic script readers demonstrated a mixed or weak rightward bias, and illiterates showed a slight leftward bias (see also Vaid & Singh, 1989). In Hsiao and Cottrell (2008), in both the learning and recognition phases of the face recognition task, participants scanned from left to right. This direction was consistent with their reading direction, since all participants were native English speakers and English is a script read from left to right. Further research is required to examine whether Arabic script readers (i.e., for scripts read from right to left) have a different scan path from English readers in face perception and recognition. This left side bias in face recognition may also be due to fundamental differences in the amount of information normally portrayed relevant to face recognition between the two sides of a face, although so far there has not been any evidence suggesting this may be the case.

5. Eye Movements in Familiar and Unfamiliar Face Recognition

Previous face recognition research has shown that as we get to know a person better, the more expressive internal features of his/her face become more important in our mental representation of the face, as opposed to features in the external part such as hair and facial shape (e.g., Ellis & Shepherd, 1992). For example, Ellis et al. (1979) showed that there was an advantage of identifying famous people from face internal features compared with external features; this advantage was also found in a face recognition task with famous faces. In contrast, no difference was found between internal and external features when identifying unfamiliar faces. Young et al. (1985) showed that in a face matching task, in which participants were asked to match a face image that contained only either external or internal features and another complete face image and decide whether the two face images were from the same person, participants were significantly faster in matching internal features when faces were familiar compared with when faces were unfamiliar; in contrast, there was not difference between familiar and unfamiliar faces in matching external features. This advantage of familiar faces in matching internal features was held when the pair of face images did not have the same orientation or expression. In addition, they showed that this advantage disappeared if the pair of face images being matched was from the same photograph so that participants could simply match the photographs instead of face features, suggesting that this advantage of familiar faces was due to structural properties of the faces instead of pictorial codes of the face images. Bonner et al. (2003) trained participants to be familiarized with a set of unfamiliar faces; they showed that in a face matching task, participants' performance on matching internal features was improved gradually during training and eventually became equivalent to their performance on matching external features; in contrast, their performance on matching external features remained at a constant level over the training days. These results suggest that we use more internal facial features in processing familiar faces compared with unfamiliar faces.

Later research showed that these internal features form a "configuration" that is crucial for the recognition of both familiar and unfamiliar faces (e.g., Farah, 1991; Tanaka & Farah, 1993). More specifically, the processing of this configuration involves "the ability to identify by getting an overview of an item as a whole in a single glance" (Farah, 1991); in particular, the spatial organization of features relative to each other (i.e. second order relationships) have been shown to be important for face recognition and distinguish it from object recognition (Farah, 1991; Farah et al., 1995). Some have argued that this ability to identify faces according to the spatial organization of internal features is due to our expertise in subordinate level discrimination/individualization of faces (e.g. Bukach et al., 2006; Gauthier & Bukach, 2007). Consistent with this view, the recognition of unfamiliar, other-race faces has been shown to involve less holistic processing compared with own-race faces (Tanaka et al., 2004); this phenomenon suggests difficulty in integrating internal features to form a configuration for holistic processing in the recognition of unfamiliar, other-race faces.

This difference between familiar and unfamiliar face recognition has also been demonstrated in eye movement behaviour, although the effect seems to be task-dependent. For example, Althoff and Cohen (1999) presented famous (familiar) and unfamiliar faces to participants and asked them to make familiarity decisions while their eye movements were recorded; they found that although most of the fixations fell in the internal face region, there was no difference between familiar and unfamiliar faces. Stacey et al. (2005) replicated Althoff and Cohen's results (1999) in a familiarity judgment task and a standard face recognition memory task; in contrast, in a face-matching task, in which participants were presented with two face images simultaneously and asked to judge whether the two face images were from the same person, they found that participants made more fixations in the internal face region when matching familiar faces compared with unfamiliar faces. Their result thus is consistent with earlier behavioural studies showing the importance of internal facial features in matching familiar faces compared with unfamiliar faces. This effect was also observed in face recall (identification) tasks: in a recent study, Heizs and Shore (2008) showed that when a face became more familiar, participants made more fixations to the eye region compared with other regions such as the nose, mouth, forehead, chin, and cheek regions in a face recall (identification) task, but not in a face recognition memory task.

In a recent study (Hsiao et al., in preparation), we trained participants to recognize other-race faces at either the individual level (i.e. identify faces by individual names) or the categorical level (i.e. identify faces by their races). Participants performed a standard face recognition memory task before and after this training. Our results showed that participants' saccade lengths significantly decreased in the face recognition task after individual level training, but not after categorical level training; this decrease in saccade length may have reflected more local, finer-grain perceptual processing after the participants' learned to individualize other-race faces. This result thus is consistent with the previous studies showing the importance of internal features in familiar face processing, and suggest that this shift from processing external to internal facial features when we get to know a person better may be due to our experience in individualizing faces in daily life as opposed to basic-level categorization or mere exposure.

6. Conclusion and Future Perspectives

In this chapter, I have reviewed the current literature on eye movements in face recognition, in particular trying to answer questions that are frequently raised in the literature, such as whether eye movements are required in face recognition since faces are processed holistically; if they are, what their functional roles are in face recognition, and what they can tell us about the cognitive processes involved in face recognition. We have seen that eye movements during face recognition have a functional role and are not just a recapitulation of those produced during the learning phase (Henderson et al., 2005); eye movements are especially important for explicit face recollection, as opposed to familiarity-based recognition, since restriction of eye movements during either the learning or the recognition phase impairs explicit recollection but not familiarity-based recognition (Mäntylä and Holm's, 2006). These results suggest that eye movements have a functional role and are required in face recognition.

As for what their functional roles are, we have shown that two fixations suffice in face recognition, and these two fixations are just around the centre of the nose, with the first fixation slightly to the left of the centre (Hsiao & Cottrell, 2008). We argue that this location may be the optimal viewing position for face recognition because of our expertise in face processing and knowledge about the information structure of faces relevant to face recognition. This hypothesis has been supported by our computational modelling study (Barrington et al., 2008), which shows that by using only the first fixation in the human data (i.e. around the centre of the nose), the model already achieves its best face recognition performance, as opposed to using the first fixation selected according to computational programs that calculate visual salience (i.e. usually fixations on the eyes are selected first). The observation that humans nevertheless require two fixations around this optimal viewing position suggests a limitation of human vision. It may be that the duration of a typical fixation (about 300 ms) is not long enough to allow us to obtain all the information required for face recognition. Also, in addition the tendency to look at the optimal viewing location, the location of our first fixation may be influenced by the central bias that is usually observed in localizing fixations (Tatler, 2007), and thus a second fixation in a different location is usually required to accumulate more information. Our human data are consistent this hypothesis (Hsiao & Cottrell, 2008): given the same amount of total face viewing time, participants had better performance when they were allowed to make two fixations compared with a single fixation. An important direction of future work is to examine whether the centre of the nose is indeed the optimal viewing position for face recognition, whether our first fixation is influenced by both the localizing central bias and the tendency to look at the optimal viewing position, and whether we require two fixations to recognize a face because of a task-switching cost from localizing to exploring.

Eye movements in face recognition also reflect hemispheric asymmetry in face processing. It has been shown that we have a preference of looking at the left side of a face from the viewer's perspective when we view faces, and this phenomenon has been linked to the right hemisphere dominance in face processing (e.g., Hsiao & Cottrell, 2008). This phenomenon may be because the low-spatial-frequency biased representation developed in the right hemisphere is more informative in face processing compared with that in the left hemisphere (i.e. high-spatial-frequency biased; Ivry & Robertson, 1998). Because of the contralateral projection from our visual hemfields to the two hemispheres, when we learn to recognize faces, the left side of a face is most often initially projected to and processed in the right hemisphere, making its internal representation more informative than the right side of the face. As the result, we tend to direct our eye fixations to the left side of a face since it is more informative and thus attracts our attention. Consistent with this hypothesis, it has been shown that this left side bias may be related to visual expertise (e.g., Hsiao & Cottrell, 2009). Possible future work is to investigate other factors that may also influence eye fixation behaviour in face recognition, such as script reading directions, and to examine whether the left side bias is a general visual expertise marker or specific to certain expertise domains.

Eye movements in face recognition also help us understand cognitive processes involved in familiar and unfamiliar face recognition. They reflect that we use more internal facial features as opposed to external features in matching familiar faces compared with matching unfamiliar faces (Stacey et al., 2005); internal facial features are also more important when we identify familiar faces compared with unfamiliar faces (Ellis et al., 1979; Heizs & Chore, 2008). This difference between familiar and unfamiliar face processing has also been reflected in participants' saccade lengths in our recent eye movement study of visual expertise training in face and object recognition (Hsiao et al., in preparation): after expertise training, participants' saccade lengths significantly decreased in standard recognition memory tasks of faces and objects; this decrease in saccade lengths may reflect a finer-grained perceptual processing on the internal features. In addition, this effect was observed in face recognition after either categorical or individual level training, but was only observed in object recognition after individual level training. This result thus suggests that the shift from processing external to internal features in face recognition when we become familiar with a face may be related to visual expertise. Further research is required to examine this potential link between the shift from external to internal feature processing and the development of visual expertise.

In conclusion, although using eye movements to study cognitive processes involved in face recognition is a relatively new approach in the face recognition literature, researchers have obtained several important findings about the dynamics of cognitive processes in face recognition that would not have been possible without the eye tracking technology. In the future, eye tracking technology will keep contributing to the research on face recognition and visual expertise to promote our understanding of how we recognize faces, how the brain processes faces, how we learn to recognize a face or develop expertise in a visual domain, and also, more generally, how we direct our eye gaze to obtain useful information in the environment to perform cognitive tasks.

7. References

Althoff, R. R & Cohen, N. J. (1999). Eye-movement-based memory effect: A reprocessing effect in face perception. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *25*, 997-1010.

Barrington, L., Marks, T., Hsiao, J. H., & Cottrell, G. W. (2008). NIMBLE: A kernel density model of saccade-based visual memory. *Journal of Vision*, *8(14):7*, 1-14.

Bonner, L., Burton, A. M., & Bruce, V. (2003). Getting to know you: How we learn new faces. *Visual Cognition*, *10*, 527–536.

Brady, N., Campbell, M., & Flaherty, M. (2005). Perceptual asymmetries are preserved in memory for highly familiar faces of self and friend. *Brain and Cognition*, *58*, 334-342.

Bruce, N. D . B. & Tsotsos, J. K. (2005). Saliency Based on Information Maximization. *Advances in Neural Information Processing Systems*, *18*, 155-162.

Brysbaert, M. & Nazir, T. (2005) Visual constraints in written word recognition: evidence from the optimal viewing-position effect. *Journal of Research in Reading*, *28*, 216-228.

Bukach, C.M., Gauthier, I., & Tarr, M.J. (2006). Beyond faces and modularity: The power of an expertise framework. *Trends in Cognitive Sciences*, *10*, 159–166.

Burt, D.M., & Perrett, D.I. (1997). Perceptual asymmetries in judgments of facial attractiveness, age, gender, speech and expression. *Neuropsychologia*, *35*, 685–693.

Dailey, Matthew N. & Cottrell, Garrison W. (1999) Organization of face and object recognition in modular neural networks. *Neural Networks*, *12(7-8)*, 1053-1074.

Ellis, H. D. & Shepherd, J. W. (1992). Face memory – Theory and practice. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues*, *Vol. 1*. Chichester, UK: Wiley.

Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1979). Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception*, *8*, 431-439.

Evans, J. J., Heggs, A. J., Antoun, N., & Hodges, J. R. (1995). Progressive prosopagnosia associated with selective right temporal lobe atrophy: a new syndrome? *Brain*, *118*, 1-13.

Everdell, I. T., Marsh, H., Yurick, M. D., Munhall, K. G., & Paré, M. (2007). Gaze behaviour in audiovisual speech perception: Asymmetrical distribution of face-directed fixations. *Perception*, *36(10)*, 1535 – 1545.

Farah, M. J. (1991). Patterns of co-occurrence among the associative agnosias. *Cognitive Neuropsychology*, *8*, 1-19.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1995). What is "special" about face perception? *Psychological Review*, *105*, 482-498.

Gauthier, I., & Bukach, C. (2007). Should we reject the expertise hypothesis? *Cognition*, *103*, 322–330.

Gauthier, I., & Tarr, M. J. (1997). Becoming a "greeble" expert: exploring face recognition mechanisms. *Vision Research*, *37*, 1673–1682.

Gauthier, I. & Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: bridging brain activity and behavior. *Journal of Experimental Psychology: Human Perception & Performance*, *28*, 431–446.

Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. (1998). Training Materials"greeble" experts: a framework for studying expert object recognition processes. *Vision Research*, *38*, 2401–2428.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform "face area" increases with expertise in recognizing novel objects. *Nature Neuroscience*, *2*, 568–573.

Gilbert, C. & Bakan, P. (1973). Visual asymmetry in perception of faces. *Neuropsychologia*, *11*, 355-362.

Gosselin, F. & Schyns, P.G. (2001) Bubbles: a technique to reveal the use of information in recognition tasks. *Vision Research*, *41*, 2261-2271.

Hayhoe, M. & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, *9*, 188-193.

Heath, R. L., Rouhana, A., & Ghanem, D. A. (2005). Asymmetric bias in perception of facial affect among Roman and Arabic script readers. *Laterality*, *10*, 51-64.

Heizs, J. J. & Shore, D. I. (2008). More efficient scanning for familiar faces. *Journal of Vision*, *8(1):9*, 1–10.

Henderson, J. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*, 498–504.

Henderson, J. M., & Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Memory & Cognition*, *33*, 98-106.

Hsiao, J. H. & Cottrell, G. W. (2008). Two fixations suffice in face recognition. *Psychological Science*, *9(10)*, 998-1006.

Hsiao, J. H. & Cottrell, G. W. (2009). Not all expertise is holistic, but it may be leftist: The case of Chinese character recognition. *Psychological Science*, *20(4)*, 455-463.

Hsiao, J. H., Shieh, D., & Cottrell, G. W. (2008). Convergence of the visual field split: hemispheric modeling of face and object recognition. *Journal of Cognitive Neuroscience, 20*, 2298-2307.

Hsiao, J. H., Shillcock, R., & Lavidor, M. (2006). A TMS examination of semantic radical combinability effects in Chinese character recognition. *Brain Research*, *1078*, 159-167.

Hsiao, J. H., Tanaka, J. W., & Cottrell, G. W. (in preparation). How does learning experience influence performance and eye movement behavior in face and object recognition?

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine*, *20*, 1254 – 1259.

Ivry, R. & Robertson, L. C. (1998). *The Two Sides of Perception*. Cambridge: MIT Press.

Joyce, C.A. (2001). Saving faces: Using eye movement, ERP, and SCR measures of face processing and recognition to investi-gate eyewitness identification. *Dissertation Abstracts International B: The Sciences and Engineering*, 61 (08), 4440. (UMI No. 72775540).

Kanwisher, N., McDermott, J. & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*, 4302–4311.

Kowler, E. (1990). The role of visual and cognitive processes in the control of eye movements. In *Eye Movements and Their role in Visual and Cognitive Processes*, E. Kowler (Ed). Elsevier Science Ltd., pp. 1-70.

Lacroix, J. P. W., Murre, J. M. J., Postma, E. O., & Van den Herik, H. J. (2006). Modeling recognition memory using the similarity structure of natural input. *Cognitive Science*, *30*, 121-145.

Lavidor, M., Hayes, A., Shillcock, R. & Ellis, A. W. (2004). Evaluating a split processing model of visual word recognition: Effects of orthographic neighborhood size. *Brain & Language*, *88*, 312-320.

Lavidor, M. & Walsh, V. (2004). The nature of foveal representation. *Nature Reviews Neuroscience*, *5*, 729-735.

Leonards, U. & Scott-Samuel N.E. (2005). Idiosyncratic initiation of saccadic face exploration in humans. *Vision Research*, *45*, 2677-2684.

Mäntylä, T. & Holm, L. (2006). Gaze control and recollective experience in face recognition. *Visual Cognition*, *13*, 365-386.

Mertens, I., Siegmund, H., & Grusser, O. J. (1993). Gaze motor asymmetries in the perception of faces during a memory task. *Neuropsychologia*, *31*, 989-998.

O'Regan, J.K., Lévy-Schoen, A., Pynte, J., & Brugaillère, B. (1984). Convenient fixation location within isolated words of different length and structure. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 250-257.

O'Toole, A. J., Millward, R. B., & Anderson, J. A. (1988). A physical system approach to recognition memory for spatially transformed faces. *Neural Networks*, *1*, 179-199.

Raaijmakers, J. & Shiffrin, R. (2002). Models of memory. In H. Pashler & D. Medin (Eds.), *Stevens' handbook of experimental psychology* (Third ed., Vol. 2: Memory and Cognitive Processes, p. 43-76). Wiley.

Rayner, K. (1979). Eye guidance in reading: Fixation locations within words. *Perception, 8,* 21-30.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422.

Renninger, L., Verghese, P., & Coughlan, J. (2007). Where to look next? Eye movements reduce local uncertainty. *Journal of Vision*, *7(3)*, 1-17.

Rossion, B., Joyce, C. A., Cottrell, G. W., & Tarr, M. J. (2003). Early lateralization and orientation tuning for face, word, and object processing in the visual cortex. *Neuroimage*, *20*, 1609-1624.

Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychological Science*, *13*, 402-409.

Sereno, S. & Rayner, K. (2003). Measuring word recognition in reading: eye movements and event-related potentials. *Trends in Cognitive Sciences*, *7*, 489–493.

Sergent, J. (1982). The cerebral balance of power: Confrontation or cooperation? *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 253–272.

Stacey, P. C., Walker, S., & Underwood, J. D. M. (2005). Face processing and familiarity: Evidence from eye-movement data. *British Journal of Psychology*, *96*, 407-422.

Tanaka, J. W. & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology*, *46*, 225-245.

Tanaka, J. W., Kiefer, M., & Bukach, C. M. (2004). A holistic account of the own-race effect in face recognition: evidence from a cross-cultural study. *Cognition*, *93(1)*, B1-B9.

Tarr, M. J., & Gauthier, I. (2000). FFA: A flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience*, *3*, 764-769.

Tatler, B. W. (2007, 11). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, *7(14)*, 1-17.

Vaid, J. & Singh, M. (1989). Asymmetries in the perception of facial affect: is there an influence of reading habits? *Neuropsychologia*, *27*, 1277-1287.

Vinette, Gosselin, & Schyns. (2004). Spatio-temporal dynamics of face recognition in a flash: it's in the eyes. *Cognitive Science*, *28*, 289-301.

Whitman, D. & Konarzewski-Nassau, S. (1997). Lateralized Facial Recognition: Spatial Frequency and Masking Effects. *Archives of Clinical Neuropsychology*, *12*, 428-428(1).

Yamada, K. & Cottrell. G. W. (1995). A model of scan paths applied to face recognition. *Proceedings of the Seventeenth Annual Cognitive Science Conference, Pittsburgh, PA*, pp. 55-60, Mahwah: Lawrence Erlbaum.

Young, A. W., de Haan, E. H. F., Newcombe, F., & Hay, D. C. (1990). Facial neglect. *Neuropsychologia*, *28*, 391-415.

Young, A. W., Hay, D. C., McWeeny, K. H., Flude, B. M., & Ellis, A.W. (1985). Matching familiar and unfamiliar faces on internal and external features. *Perception*, *14*, 737–746.

15

Surface representations for 3D face recognition

Thomas Fabry*, Dirk Smeets* and Dirk Vandermeulen
Katholieke Universiteit Leuven
Belgium

1. Introduction

For long, face recognition has been a 2D discipline. However, 2D face recognition has shown to be extremely difficult to be robust against a.o. lighting conditions and pose variations (Phillips et al., 2003). At the same time, technological improvements are making 3D surface capturing devices affordable for security purposes. As a result of these recent developments face recognition shifts from 2D to 3D. This means that in the current state-of-the-art face recognition systems the problem is no longer the comparison of 2D color photos, but the comparison of (textured) 3D surface shapes.

With the advent of the third dimension in face recognition, we think it is necessary to investigate the known surface representations from this point of view. Throughout recent decades, a lot of research focused on finding an appropriate digital representation for three dimensional real-world objects, mostly for use in computer graphics (Hubeli & Gross, 2000; Sigg, 2006). However, the needs for a surface representation in computer graphics, where the primary concerns are visualization and the ability to process it on dedicated computer graphics hardware (GPUs), are quite different from the needs of a surface representation for face recognition.

Another motivation for this work is the non-existence of an overview of 3D surface representations, altough the problem of object representation is studied since the birth of computer vision (Marr, 1982).

With this in mind, we will, in this chapter, try to give an overview of surface representations for use in biometric face recognition. Also surface representations that are not yet reported in current face recognition literature, but we consider to be promising for future research – based on publications in related fields such as 3D object retrieval, computer vision, computer graphics and 3D medical imaging – will be discussed.

What are the desiderata for a surface representation in 3D face recognition? It is certainly useful for a surface representation in biometric applications, to be accurate, usable for all sorts of 3D surfaces in face recognition (open, closed...), concise (efficient in memory usage), easy to acquire/construct, intuitive to work with, have a good formulation, be suitable for computations, convertible in other surface representations, ready to be efficiently displayed and useful for statistical modelling. It is nevertheless also certainly necessary to look further than a list of desiderata. Herefore, our approach will be the following: we make a taxonomy of all surface representations within the scope of 3D face recognition. For each of the of the representations in this taxonomy, we will shortly describe the mathematical theory behind it. Advantages and disadvantages of the surface representation will be stated. Related research using these representations will be discussed and directions for future research will be indicated.

*The first two authors have equal contribution in this work .

The structure of this chapter follows the taxonomy of Fig. 1. First we discuss the explicit (meshfree) surfaces in section 2, followed by the implicit surfaces in section 3. We end with some conclusions regarding surface representations for 3D face recognition.

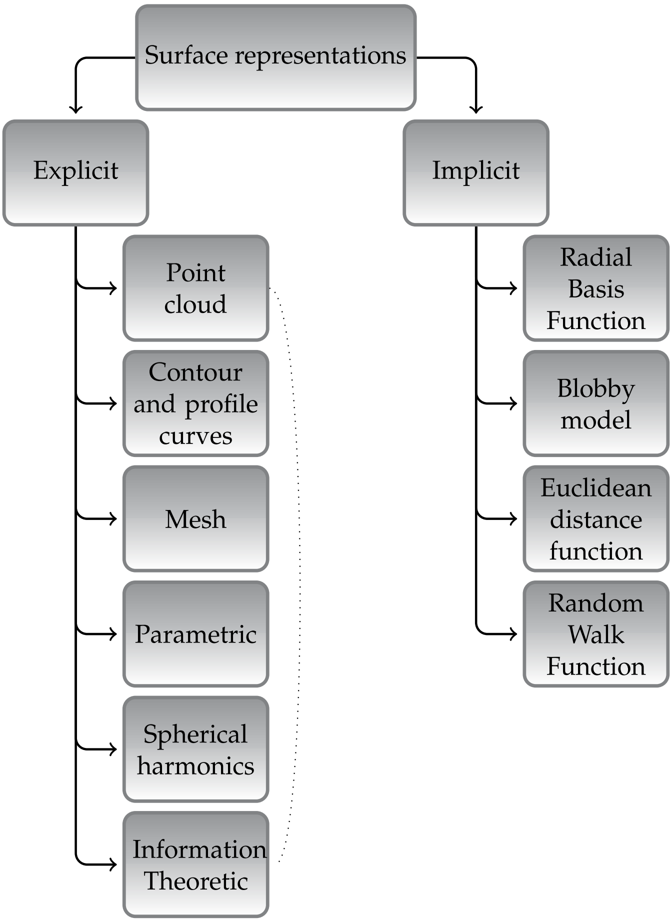


Fig. 1. Overview of surface representations

2. Explicit surface representations

In this section, several explicit surface representations are discussed. Strictly speaking, explicit functions $f(\vec{x}) : \mathbb{R}^m \rightarrow \mathbb{R}^n$ are functions in which the n dependent variables can be written explicitly in terms of the m independent variables. Simple shapes (spheres, ellipsoids,...) can be described by analytic functions. Unfortunately, it is mostly not possible to represent real world objects by analytical surfaces. Therefore, this section mainly focusses on discretised surface representations.

2.1 Point clouds

The point cloud is without doubt the simplest surface representation. It consists of an unordered set of points that lie on the surface, in the 3D case an unordered set of x, y, and

z-coordinates. While a point cloud is not a real surface representation, but a (sometimes very) sparse approximation of the surface at certain well-defined points, we consider this representation for a number of reasons. Firstly, point clouds are most often the output created by 3D scanners. Secondly, the point cloud can be the base of most of the following surface representations. Another reason to incorporate the point cloud in this document is its increased popularity because of the ever-increasing memory capacities in today's computers. Earlier the amount of stored information was to be minimized, so a minimal amount of points were stored. As these point clouds were very sparse and as such a very coarse approximation of the surface, they were then interpolated using, for instance, tensor-product splines. Today, memory shortage is of less concern, so more points can be stored, making point clouds approximate other surface representations on finer and finer levels of detail.
Figure 2 gives an example of a 3D face represented as a point cloud, containing approximately 2000 points.

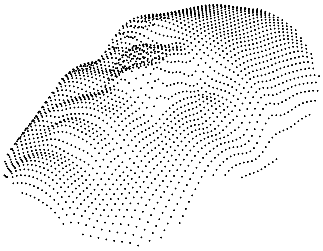


Fig. 2. An example of a 3D face represented as a point cloud.

One big advantage of the point cloud surface representation is the easy editing of point clouds: because of the lack of a global connectivity graph or parameterization, point insertion, deletion, repositioning,…is trivial. Another important advantage is the large amount of algorithms developed for point clouds. A very popular method for 3D surface alignment, the Iterative Closest Point (ICP) algorithm (Besl & McKay, 1992), uses only points on both surfaces and iterates between closest points search for correspondence finding and transformation calculation and application. Afterwards, many variants of the original algorithm were developed (Rusinkiewicz & Levoy, 2001). The main drawback of point clouds is the incompleteness of the surface description: only at some sparse point locations the surface is known. Therefore, by representing a surface by a point cloud, a trade-off has to be made between accuracy and amount of stored information. Also, rendering a set of points as a smooth surface needs special processing, as explained in (Fabio, 2003).
The earliest use of point clouds was as a rendering primitive in (Levoy & Whitted, 1985). Point clouds have also been used for shape and appearance modeling in e.g. (Kalaiah & Varshney, 2003; Pauly et al., 2003).
In 3D face recognition, point clouds are frequently used. Mostly for surface registration with ICP (Alyüz et al., 2008; Amberg et al., 2008; Chang et al., 2006; Faltemier et al., 2008; Kakadiaris et al., 2007; Lu & Jain, 2006; Maurer et al., 2005; Russ et al., 2006; Wang et al., 2006). Bronstein et al. (2005) represent faces by an expression-invariant canonical form, i.e. a point cloud where

point wise Euclidean distances approximately equal the point wise geodesic distances[1] in the original face representation. Three-dimensional face shapes are often modeled with a point cloud based statistical model, mostly a PCA model as in (Al-Osaimi et al., 2009; Amberg et al., 2008; Lu & Jain, 2006; Russ et al., 2006). In Mpiperis et al. (2008), a bilinear model, based on point clouds, is used to seperate intra-class and inter-class variations. In Fabry et al. (2008), point clouds are treated in an information theoretic way, leading to probability density function as surface representation, which is discussed in more detail in section 2.6.

2.2 Contour and profile curves

Contour and profile curves are also very sparse surface representations. They can even be sparser than point clouds, but can also be made to approximate the surface as good as wanted. The main idea is to represent shapes by the union of curves. The curves itself can be represented by a set of connected points or as a parametric curve.

Contour curves are closed, non intersecting curves on the surface, mostly of different length. Depending on the extraction criterion, different types of contour curves can defined. Iso-depth curves are obtained by translating a plane through the 3D face in one direction and considering n different intersections of the plane and the object. n is the number of contours that form the surface representation. Mostly the plane is positioned perpendicular to and translated along the gaze direction, which is hereby defined as the z-axis. Then iso-depth curves have equal z-values. Iso-radius curves are contours, obtained as an intersection of the object with a cylinder with radius $r = \sqrt{x^2 + y^2}$, or as an intersection with a sphere with radius $r = \sqrt{x^2 + y^2 + z^2}$, with the z-axis parallel to the gaze direction and the y-axis parallel to the longitudinal axis of the face. An iso-geodesic curve, or iso-geodesic, is a contour with each part of the curve on an equal geodesic distance to a reference point, i.e. the distance of the shortest path on the full surface between the part of the curve and the reference point. The calculation of geodesic distances is mostly done using a polygon mesh (see section 2.3).

Examples of iso-depth, iso-radius and iso-geodesic curves are given in Fig. 3. Only points lying on those curves are shown.

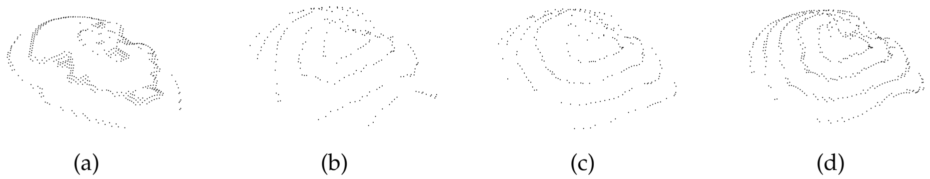

(a) (b) (c) (d)

Fig. 3. Points lying on iso-depth curves (a), on iso-radius curves obtained by intersection with a cylinder (b) or with a sphere (c) and iso-geodesics with respect to the nose tip (d).

Profile curves on the contrary have a starting and an end point. For 3D faces, the starting point is most frequently a point in the middle of the face, mostly the nose tip, while the end point is often at the edge of the face. There exist an infinite number of profile curves in between those points. Figure 4(a) shows an example with each part on the curve having the same angle with respect to a line in the xy-plane through the central point (nose tip). Figure 4(b) shows points

[1] The geodesic distance is the length of the shortest path *on the object surface* between two points on the object.

on lines with equal x-value, again with the z-axis parallel to the gaze direction and the y-axis parallel to the longitudinal axis.

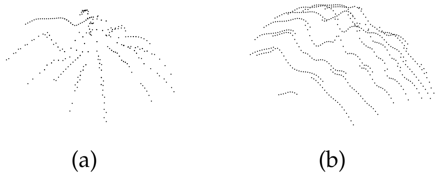


Fig. 4. Points on profile curves with curve parts under the same angle (a) or with the same x-value (b).

Curves are non-complete surface representations, implying that the surface is only defined on the curves. On the one hand, this implies a loss of information, on the other hand lower storage requirements. In order to construct contour curves, a reference point is needed. In 3D face recognition mostly the nose is used, which infers the manual or automatic extraction of this landmark. For extraction of iso-depth and cylinder based iso-radius curves and most types of profile curves, even more information is required: the gaze direction and/or longitudinal axis of the face. When done this, it is more easy to set correspondences between faces based on corresponding curves.

Contour and profile curves are frequently used in 3D face recognition. Iso-geodesics are popular because of their lower sensitivity to expression variation, based on the hypothesis that expression-induced surface variations can approximately be modeled by isometric transformations. Those transformations keep geodesic distances between every point pair on the surface. Berretti et al. (2006) use the spatial relationship between the intra-subject iso-geodesics as a subject specific shape descriptor. Feng et al. (2007) divide the iso-geodesics in segments that form the basis of trained face signatures. Mpiperis et al. (2007) map the iso-geodesic curves to concentric circles on a plane using a piecewise linear warping transformation. Jahanbin et al. (2008) extract five shape descriptors from each iso-geodesic: convexity, ratio of principal axes, compactness, circular and elliptic variance. These features are trained with Linear Discriminant Analysis (LDA). In Li et al. (2008), LDA is also used for training of texture intensities sampled at fixed angles on iso-geodesic curves. Pears & Heseltine (2006) use sphere based iso-radius curves. Due to the infinite rotational symmetry of a sphere, the representation is invariant to pose variations. Using this representation, registration can be implemented using a simple process of 1D correlation resulting in a registration of a comparable accuracy to ICP, but fast, non iterative, and robust to the presence of outliers. Samir et al. (2006) compare faces using dissimilarity measures extracted from the distances between iso-depth curves. Jahanbin et al. (2008) use iso-depth curves in the same way as they use iso-geodesics. Profile curves are used by ter Haar & Veltkamp (2008), where comparison is done using the weighted distance between corresponding sample points on the curves, and in Feng et al. (2006) where the curves are used similar as in Feng et al. (2007) (see above).

2.3 Polygon meshes

In 3D research in general, and a fortiori in 3D face recognition, the vast majority of researchers represent 3D object surfaces as meshes. A mesh is in essence an unordered set of vertices (points), edges (connection between two vertices) and faces (closed set of edges) that together

represent the surface explicitly. Mostly, the faces consist of triangles, quadrilaterals or other simple convex polygons, since this simplifies rendering. Figure 5 shows the triangular mesh corresponding with the point cloud of Fig. 2.
The problem of constructing a mesh given a point cloud is commonly called the surface reconstruction problem, although this might also incorporate reconstruction of other complete surface representations. The most powerful algorithm to deal with mesh construction given a point cloud is the *power crust* algorithm, described in Amenta et al. (2001). Other algorithms that deal with this problem are, a.o., the algorithms of Hoppe et al. (1992), Edelsbrunner & Mücke (1994) and Curless & Levoy (1996). A short overview of methods for triangulation can be found in the paper of Varshosaz et al. (2005).

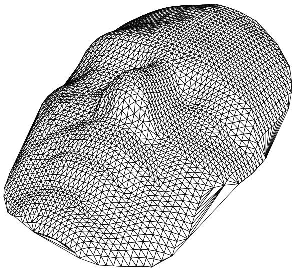


Fig. 5. An example of a 3D face represented as a mesh.

Probably the main benefit of a polygon mesh is its ease of visualization. Many algorithms for ray tracing, collision detection, and rigid-body dynamics are developed for polygon meshes. Another advantage of meshes, certainly in comparison to point clouds, is the explicit knowledge of connectivity, which is useful for the computation of the geodesic distance between two points. This is particularly useful in face recognition because geodesic distances between points on the surface are often used in 3D expression-invariant face recognition, because they seem to vary less than Euclidean distances. The use of this was introduced by Bronstein et al. (2003), using a fast marching method for triangulated domains (Kimmel & Sethian, 1998) for geodesic distance calculation. Afterwards, many other researchers used the concept of invariance of geodesic distances during expression variations, by directly comparing point wise distances (Gupta et al., 2007; Li & Zhang, 2007; Smeets et al., 2009) or using iso-geodesic curves (see section 2.2). On the other hand, mesh errors like cracks, holes, T-joints, overlapping polygons, dupplicated geometry, self intersections and inconsistent normal orientation can occur as described by Veleba & Felkel (2007).

2.4 Parametric surface representations

A generic parametric form for representing 3D surfaces is a function with domain $\mathbb{R}^2$ and range $\mathbb{R}^3$ (Campbell & Flynn, 2000):

$$\mathcal{S}(u,v) = \begin{cases} x = f_1(u,v) \\ y = f_2(u,v) \\ z = f_3(u,v) \end{cases} \tag{1}$$

where u and v are the two parametric variables.

Amongst the advantages of parametric representations in general we can count the simple but general and complete mathematical description, the easiness to handle and the readiness of technology to visualise these representations. General disadvantages are that only functions can be represented, which cause problems in, for instance, the ear and nose regions.
The class of parametric surface representations is very general, but the following three deserve particular attention.

2.4.1 Height maps

A height map is a special form of a parametric surface with $x = u$ and $y = v$ and is also often referred as depth map, range image or graph surface. A height map represents the height of points along the z-directions in a regular sampling of the x, y image axes in a matrix. An example of a face represented as a depth map can be seen in figure 6. A big advantage of this representation is that many 3D laser scanners produce this kind of output. Because mostly the x and y values lay on a regular grid, the surface can be descibed by a matrix and 2D image processing techniques can be applied on it. The most prominent disadvantage of height maps is the limited expressional power: only what is 'seen' when looking from one direction (with parallel beams) can be represented.This causes problems in the representation of, for instance, the cheeks and ears of human faces. The use of height maps in face recognition has already been discussed by Akarun et al. (2005). One very specific example of a 3D face recognition method using height maps is the method of Samir et al. (2006), because here height maps are extracted from triangulated meshes in order to represent the surface by level curves, which are iso-contours of the depth map. Colbry & Stockman (2007) extend the definition of depth map leading to the canonical face depth map. This is obtained by translating a parabolic cylinder or a quadratic, instead of a plane, along the z-direction. Because of this alternative definition, it could also belong to section 2.4.2.

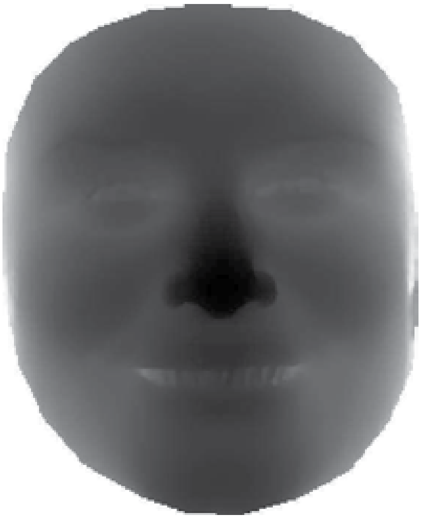


Fig. 6. An example of a height map surface representation.

2.4.2 Geometry images

Another often used parametric surface representation is the geometry image, a regularly sampled 2D grid representation, but here not representing the distance to the surface along a viewing direction. The directions are 'adaptive' in the sense that they are conveyed so to be able to represent the whole surface, thus also regions that would not be representable in a height map due to the directionality of 'height'. Gu et al. (2002) describe an automatic system for converting arbitrary meshes into geometry images: the basic idea is to slice open the mesh along an appropriate selection of cut paths in order to unfold the mesh. Next the cut surface is parametrized onto a square domain containing this opened mesh creating an $n \times n$ matrix of $[x, y, z]$ data values. The geometry image representation has some advantages, the biggest being that the irregular surface is represented on a completely regular grid, without loosing information. As with the height map, this structure is easy to process, both for graphics applications and for recognition applications. A big disadvantage is the computational and technical complexity of the generation of geometry images.

Geometry images have already been used for 3D face recognition. In (Kakadiaris et al., 2007; Perakis et al., 2009), a geometry image maps all vertices of the face model surface from $\mathbb{R}^3$ to $\mathbb{R}^2$. This representation is segmented to form the Annotated Face Model (AFM) which is rigidly aligned with the range image of a probe. Afterwards, the AFM is fitted to each probe data set using the elastically adapted deformable model framework, described by Metaxas & Kakadiaris (2002). The deformed model is then again converted to a geometry image and a normal map. Those images are analysed using the Haar and pyramid wavelet transform. The distance metric that is used to compare different faces, uses the coefficients of the wavelet transforms.

2.4.3 Splines

Spline surfaces are piecewise polynomial parametric surface representations that are very popular in the field of Computer Aided Design and Modelling (CAD/CAM) because of the simplicity of their construction, including interpolation and approximation of complex shapes, and their ease and accuracy of evaluation. The basis of splines are control points, which are mostly lying on a regular grid.

One well-known type of spline surfaces are Bézier surfaces. These are represented as:

$$\mathcal{S}(u,v) = \sum_{i=0}^{n} \sum_{j=0}^{m} B_i^n(u)\ B_j^m(v)\ \vec{x}_{i,j} \tag{2}$$

where B_i^n are Bernstein polynomials. Another well-known spline surface type are nonuniform rational B-splines (NURBS), defined as

$$\mathcal{S}(u,v) = \sum_{i=1}^{k} \frac{N_{i,n} w_i}{\sum_{j=1}^{k} N_{j,n} w_j} \vec{x}_i, \tag{3}$$

with k the number of control points $\vec{x}_i$ and w_i the corresponding weights. An example of face representation by Bézier curves can be found in (Yang et al., 2006) and of NURBS in (Liu et al., 2005; Yano & Harada, 2007). Although spline surfaces have been considered for representing faces, they have not found widespread use in 3D face recognition. Most probably because, as is stated by Besl (1990), a.o.: "it is difficult to make a surface defined on a parametric rectangle fit an arbitrary region on the surface". Also, the control points are not easily detectable. Another disadvantage of splines is the uselessness of the spline parameters for recognition.

2.5 Spherical harmonics surface representations

Spherical harmonics are mathematical functions that can be used for representing spherical objects (sometimes also called star-shaped objects). The surface first has to be represented as a function on the unit sphere: $f(\theta, \phi)$. This function can then be decomposed (spherically expanded) as:

$$f(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} a_{lm} Y_l^m(\theta, \phi). \tag{4}$$

The spherical harmonics $Y_l^m(\theta, \phi) : |m| \in \mathbb{N}$ are defined on the unit sphere as:

$$Y_l^m(\theta, \phi) = k_{l,m} P_l^m(\cos\theta) e^{im\phi}, \tag{5}$$

where $\theta \in [0, \pi], \phi \in [0, 2\pi[, k_{l,m}$ is a constant, and P_l^m is the associated Legendre polynomial. The coëfficients a_{lm} are uniquely defined by:

$$a_{lm} = \int_0^{2\pi} \int_0^{pi} f(\theta, \phi) \overline{Y_l^m}(\theta, \phi) sin(\theta) d\theta d\phi. \tag{6}$$

While this theory is stated for spherical objects, it is important to mention that spherical harmonics have already been used for non spherical objects by first decomposing the object into spherical subparts (Mousa et al., 2007) using the volumetric segmentation method proposed by Dey et al. (2003).

Advantages of the spherical harmonics surface representation include the similarity to the Fourier transform, which has already proven to be a very interesting technique in 1D signal processing. Also, because of the spectral nature of this surface representation, it can lead to large dimensionality reductions, leading to decreases in computation time and efficient storage. Other advantages are the rotational invariance of the representation and the ability to cope with missing data (occlusions and partial views).

The spherical harmonics surface representation has some drawbacks as well. One of them has already been mentioned and is not insuperable: the need for spherical surfaces. Other disadvantages include the unsuitability for intuitive editing, the non-trivial visualisation and the global nature of the representation.

Spherical harmonics surface representations have already been used in a number of applications that bear a close relation to face recognition. Kazhdan et al. (2003) used it as a 3D shape descriptor for use in searching a 3D model database. Mousa et al. (2007) use the spherical harmonics for reconstruction of 3D objects from point sets, local mesh smoothing and texture transfer. Dillenseger et al. (2006) have used it for 3D kidney modelling and registration. To the best of our knowledge, the only use of this representation in face recognition so far is in (Llonch et al., 2009). In this work, a similar transformation with another (overcomplete) basis is used as a surface representation for the 3D faces in the face recognition experiments, where this representation is also submitted to linear discriminant analysis (LDA). The performance reported is better than PCA on depth maps, which the authors consider as baseline.

Further applications in face recognition are not yet widely explored, but we see a great potential in the method of Bülow & Daniilidis (2001), who combine the spherical harmonics representation with Gabor wavelets on the sphere. In this way, the main structure of the 3D face is represented globally, while the (person-specific) details are modelled locally (with wavelets). This solves the drawback of the global nature of the representation and could as such be used for multiscale progressive 3D face recognition.

2.6 Point cloud-based surface representations: information theoretic measures

Recently, some researchers have proposed to work directly on the point cloud, without using any real surface representation, but instead use information theoretic measures defined directly on the raw point cloud surface representation. Some of these methods do nevertheless implicitly use some kind of kernel surface representation, which can also be viewed as density estimation (Silverman, 1986), and although the density estimation itself is explicit, the surface can be thought of as implicitly present. This is also the reason why this surface reprsentation was not included in section 2.1 (which would also make sense) but is treated as a link between the explicit and implicit surface representations. Density estimation is a fundamental concept in statistics, the search for an estimate of the density from a given dataset. In the case of surface representations, this dataset is a point cloud. This estimation can be nonparametric, which can be considered as an advantage of this method. Also the generality, sound statistical base and low data requirements are advantages. Disadvantages include the difficulties in visualising the surface representation,

The most used density estimation technique is kernel density estimation KDE, introduced by Parzen (1962). Here the density is computed as

$$\widehat{f}_h(\vec{x}) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{\vec{x}-\vec{x}_i}{h}\right) \tag{7}$$

where K is some kernel with parameter h. A volume rendering of a KDE of a 3D face surface can be seen in figure 7.

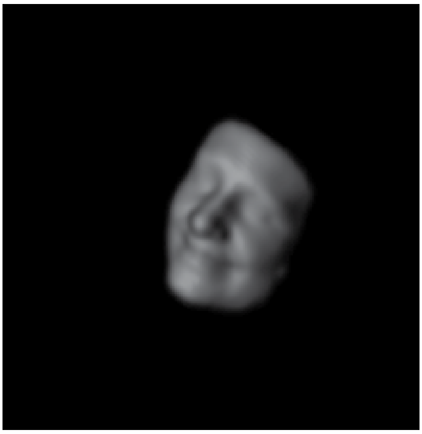


Fig. 7. Volume rendering of a kernel density estimation of a human face.

The information theoretic methods have already been proven to be useful in 3D registration and recognition. Tsin & Kanade (2004) proposed the *kernel correlation* of two point clouds, an entropy-related measure expressing the compatibility of two point clouds, and used this for robust 3D registration. This measure has later been used in 3D face recognition by Fabry et al. (2008). A related technique, which has until now not been applied to face recognition, is found in (Wang et al., 2008). Here, groupwise registration between point clouds is performed by minimizing the Jensen-Shannon divergence between the Gaussian mixture representations of the point clouds.

3. Implicit surface representations

In general, implicit functions are defined as the iso-level of a scalar function $\phi : \mathbb{R}^n \to \mathbb{R}$. A 3D implicit surface $\mathcal{S}$ is then mathematically defined as

$$\mathcal{S} = \{\vec{x} \in \mathbb{R}^3 | \phi(\vec{x}) = \rho\}. \tag{8}$$

We call this the *iso-surface* of the implicit function. The iso-surface at $\rho = 0$ is sometimes referred to as the *zero contour* or *zero surface*. As such, implicit surfaces are 2D geometric shapes that exist in 3D space (Bloomenthal & Wyvill, 1997). The iso-surface partitions the space into two regions: interior of the surface, and exterior of the surface. Mostly, the convention is followed that inside the surface, the function returns negative values and outside the surface, the function returns positive values. The inside portion is referred as Ω^-, while points with positive values belong to the outside portion Ω^+. The border between the inside and the outside is called the interface $\partial\Omega$.

The simplest surfaces (spheroids, ellipsoids,...) can be described by analytic functions and are called algebraic surfaces. The surface is the set of roots of a polynomial $\phi(\vec{x}) = \rho$. The degree of the surface n is the maximum sum of powers of all terms. The general form of a linear surface ($n = 1$), or plane, is

$$\phi(x, y, z) = ax + by + cz - d = \rho, \tag{9}$$

while the general form for a quadratic surface ($n = 2$) is:

$$\begin{aligned} &\phi(x, y, z) \\ &= ax^2 + bxy + cxz + dx + ey^2 + fyz + gy + hz^2 + iz + j \\ &= \rho. \end{aligned} \tag{10}$$

Superquadrics ($n > 2$) provide more flexibility by adding parameters to control the polynomial exponent, allowing to describe more complex surfaces. Nevertheless, analytic functions are designed to describe a surface globally by a single closed formula. In reality, it is mostly not possible to represent a whole real-life object by an analytic function of this form.

3.1 Radial Basis Functions

Radial Basis Functions (RBFs) are another type of implicit functions that have been proven to be a powerful tool in interpolating scattered data of all kinds, including 3D point clouds representing 3D objects. A RBF is a function of the form

$$\mathcal{S}(\vec{x}) = \sum_{i=1}^{N} \lambda_i \Phi(\|\vec{x} - \vec{x}_i\|) + p(\vec{x}), \tag{11}$$

with λ_i the *RBF-coefficients*, Φ a *radial basic function*, $\vec{x}_i$ the *RBF centra* and $p(\vec{x})$ a polynomial of low degree.

As can be seen from equation (11), the RBF consists of a weighted sum of radially symmetric basic functions located at the RBF-centra x_i and a low degree polynomial p. For surface representation, the RBF-centra x_i are simply a subset of points on the surface. Finding the appropriate RBF-coefficients for implicitly representing a surface is done by solving:

$$\forall x_i : s(x_i) = f_i, \tag{12}$$

For a surface representation, we want the surface to be the zero-contour of the implicit surface $s(\vec{x})$ and hence $f_i = 0, \forall i$. To prevent the interpolation to lead to the trivial solution, $s(\vec{x}) = 0$ everywhere, we have to add additional constraints. This is done by adding *off-surface points*: points at a distance of the surface, whose implicit function value is different from zero and mostly equal to the euclidean distance to the surface. Figure 8 gives an example of a RBF interpolation with zero iso-surface.

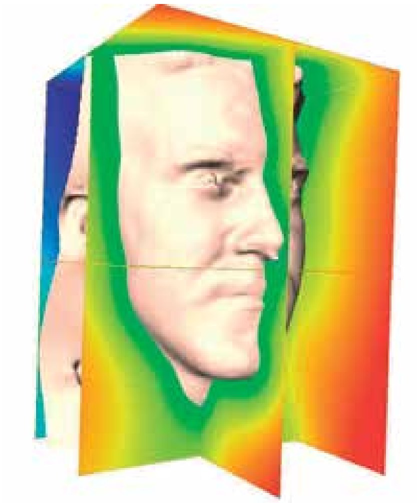


Fig. 8. An example of a RBF interpolation with zero iso-surface.

A very clear introduction to the RBF-theory, and info about a fast commercial RBF-implementation can be found in (Far, 2004). A mathematically very complete reference book about Radial Basis Functions is (Buhmann, 2003).

The biggest advantage of radial basis function interpolation is the absence of the need for point connectivity. Other advantages include the low input data requirements (bare point clouds), and the possibility to insert smoothness constraints when solving for the RBF. A disadvantage of RBFs is the computational complexity of the problem. This problem can however be alleviated by specific mathematical algorithms (Fast Multipole Methods (Beatson & Greengard, 1997)), or compactly supported basis functions (Walder et al., 2006). Because of this computational complexity, also the editing of the surface is not trivial.

In Claes (2007), a robust framework for both rigid and non-rigid 3D surface representation is developed to represent faces. This application can be seen as 3D face biometrics in the wide sense: representing and distinguishing humans by measuring their *face geometry*. This is used for craniofacial reconstruction.

Thin Plate Splines, one particular kind of RBF basic function, are popular in non-rigid registration of face models. Surface registration is an important step in some model-based 3D face recognition methods, but then the RBF is not used as the surface representation method but merely as a preprocessing technique (Irfanoglu et al., 2004; Lu & Jain, 2005).

Another application of RBFs in face recognition can be found in (Pears, 2008), where the RBF is sampled along concentric spheres around certain landmarks to generate features for face recognition.

3.2 Blobby Models

The blobby model is another kind of implicit surface representation introduced by Blinn (1982). It was originally perceived as a way to model molecular models for display, and is, as such, tightly related to the quantum mechanical representation of an electron: a density function of the spatial location. This way, the molecule surface can be thought of as the ρ iso-contour of the sum of atom contributions

$$D(x, y, x) = \sum_i b_i \exp(-a_i r_i^2), \tag{13}$$

where r_i are distances to the atom locations. Various variants of the original blobby models exist, which can also be called *metaballs* or *soft objects*, and instead of the exponential, one can also use polynomials (Nishita & Nakamae, 1994) or ellipsoids (Liu et al., 2007) to represent the blobs.

An advantage of the blobby model surface representation is the apparent possibility for huge data reduction without loosing much detail. However, the efficient construction of blobby models is still a problem under research (Liu et al., 2007).

Maruki Muraki (1991) used this blobby model to describe a surface originally represented by range data with normals. He does this by solving an optimization problem with parameters x_i, y_i, z_i, a_i, b_i with x_i, y_i, z_i the locations of the blobs and a_i, b_i the blob parameters. Interestingly, the examples shown in this 1991 paper are representations of faces. It seems that a face can reasonably well be represented with about 250 blobs, making this representation promising for 3D face recognition.

Nevertheless, there are not yet applications of this method in 3D face recognition. It has however been used in the related problem of *natural object recognition*, where 2D contours were represented as blobby models, and these blobby models were then used for classification of the contours (Jorda et al., 2001).

3.3 Euclidean distance functions

A special class of scalar functions are distance functions. The *unsigned distance function* yields the distance from a point $\vec{p}$ to the closest point on the surface $\mathcal{S}$ (Jones et al., 2006):

$$dist_{\mathcal{S}}(\vec{p}) = \inf_{\vec{x} \in \mathcal{S}} ||\vec{x} - \vec{p}||, \tag{14}$$

while *signed distance functions* represent the same, but have a negative sign in Ω^-, inside the object. The signed distance function is constructed by solving the Eikonal equation:

$$||\nabla \phi(x, y, z)|| = 1, \tag{15}$$

together with the boundary condition $\phi|_{\mathcal{S}} = 0$. At any point in space, ϕ is the Euclidean distance to the closest point on $\mathcal{S}$, with a negative sign on the inside and a positive on the outside (Sigg, 2006). The gradient is orthogonal to the iso-surface and has a unit magnitude (Jones et al., 2006). An example of a distance function is given in figure 9. The signed distance function can also be approximated using Radial Basis Functions (Far, 2004), as shown in figure 8.

One advantage of a surface represented by a distance function is that the surface can easily be evolved using a level set method. In those methods, also other implicit surface representations are possible, but distance transforms have nice numerical properties (Osher & Fedkiw, 2003).

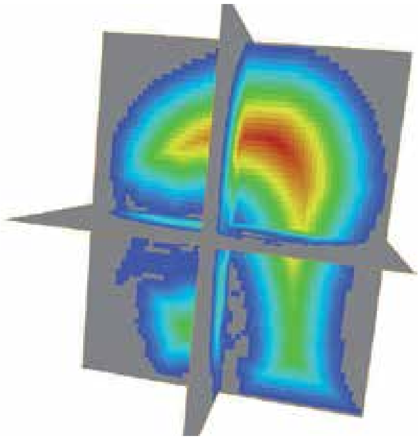


Fig. 9. An example of a distance function.

An interesting application in face recognition (in 2D though) is given in (Akhloufi & Bendada, 2008) where a distance transform is used to get an invariant representation for face recognition, using thermal face images. After extraction of the face region, a clustering technique constructs the facial isotherm layers. Computing the medial axis in each layer provides an image containing physiological features, called face print image. A Euclidean distance transform provides the necessary invariance in the matching process. Related to the domain of face recognition, the signed distance function is used in craniofacial reconstruction (Vandermeulen et al., 2006). A reference skull, represented as distance maps, is warped to all target skulls and subsequently these warps are applied to the reference head distance map.

Signed distance maps are also interesting for aligning surfaces, as described in Hansen et al. (2007). Symmetric registration of two surfaces, represented as signed distance maps, is done by minimizing the energy functional:

$$\begin{aligned} F(\vec{p}) = & \sum_{\vec{x} \in U_x^r} (\phi_y(W(\vec{x};\vec{p})) - \phi_x(\vec{x}))^2 \\ & + \sum_{\vec{y} \in U_y^r} (\phi_x(W(\vec{y};\vec{p})) - \phi_y(\vec{y}))^2, \end{aligned} \tag{16}$$

with $W(-;\vec{p})$ the warp function, U_x^r and U_y^r the narrow bands around the surfaces $\mathcal{S}_x$ and $\mathcal{S}_y$ and ϕ the signed distance map. The width of the narrow band r should be larger than the width of the largest structure. Hansen et al. (2007) state that the level set registration performs slightly better than the standard ICP algorithm (Besl & McKay, 1992).

3.4 Random walk functions

This scalar surface representation gives at a point in space a value that is the average time of a random walk to reach the surface starting from that point. This scalar function is the result of solving the Poisson equation:

$$\Delta\phi(x,y,z) = -1, \tag{17}$$

again subject to the boundary condition $\phi|_{\mathcal{S}} = 0$ and with $\Delta\phi = \frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + \frac{\partial^2\phi}{\partial z^2}$. For every internal point in the surface, the function assigns a value reflecting the mean time required for a random walk beginning at the boundaries and ending in this particular point. The level sets of ϕ represent smoother versions of the bounding surface. A disadvantage of this function is that a unique solution of equation (17) only exists within a closed surface. An example of a random walk function is given in figure 10.

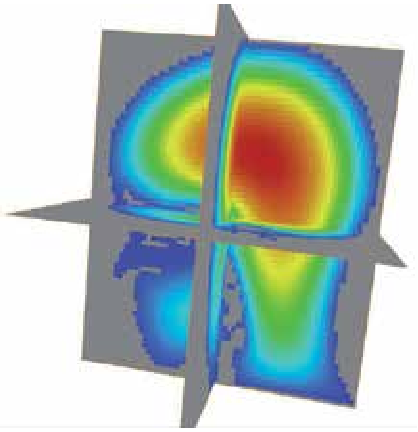


Fig. 10. An example of a random walk function.

To the best of our knowledge, this scalar function is not yet used in face recognition. However, it has already been proven to be useful in 2D object classification which makes it for auspicious for use in biometrics (Gorelick et al., 2006).

4. Conclusions

We can conclude that, although many representations in biometrics are based on meshes, a number of interesting alternatives exist. We have given a systematic discussion of the different three-dimensional surface representations that seem to be promising for use in 3D biometrics and, if known, their already existing use. While we are aware of the non-exhaustive nature of this work, we hope to have given to the face recognition and other related research communities (computer graphics, mathematics of surfaces,...) some interesting ideas.

We have paid attention to the advantages and disadvantages of the different surface representations throughout the whole text. The main advantages and disadvantages of the different surface representations are summarized in table 1. From this we can conclude that many of the advantages of the different surface representations have not yet been taken advantage of in current face recognition research. This could thus be very interesting for future research.

Other interesting conclusions can be drawn from the preceeding text and table. First of all, we see that the left branch of our taxonomy, the explicit representations, is much more frequently used in todays face recognition research, opposed to the other branch, the implicit representations. This can be explained by the fact that explicit representations have been very much a topic of interest in computer graphics because of hardware requirements and are as such also the first to be considered in 3D face recognition.

Furthermore: although the polygonal mesh is often used in 3D face recognition and certainly has some advantages, we think it is more important to keep the other surface representations in mind for doing face recognition research. Moreover, we already see a gain in importance of meshfree methods in the field of numerical analysis, where meshfree methods are used

	Accuracy	Conciseness	Acquiering	Intuitiveness	Parameterization	Computability	Displayability	Main advantage	Main disadvantage	Previous use in face recognition
Point Cloud	±	±	+	+	±	+	+	Simplicity	Sparseness	many (see text)
Contour and profile curves	−	+	±	+	+	±	+	Invariance	Sparseness	many (see text)
Polygonal Mesh	+	±	±	+	±	±	+	Many tools	Discrete	many (see text)
Parametric										
• Height map	±	±	+	±	+	+	±	Aquiring	Self Occlusion	many (see text)
• Geometry image	+	±	−	−	+	±	±	Parameterization	Construction	Metaxas & Kakadiaris (2002)
• Splines	+	+	−	±	+	+	+	Parameterization	Computability	not yet…
Spherical harmonics	+	+	−	−	+	−	−	Multiresolution	Globallity	Llonch et al. (2009)
Point-Cloud based	±	±	−	±	±	−	−	Sound theory	Computability	Fabry et al. (2008)
Radial Basis Functions	+	±	−	±	+	−	−	Completeness	Computability	not yet…
Blobby Models	+	+	−	±	+	−	−	Compression	Construction	not yet…
Distance Functions	+	−	−	±	±	−	−	Evolution	Memory req.	not yet…
Random walk function	+	−	−	±	±	−	−	Evolution	Memory req.	not yet…

Table 1. Summary of advantages and disadvantages of the discussed surface representations.

in finite element modelling, for solving partial differential equations or for approximating functions without using classic mesh discretisations. These kind of methods have many advantages: easy handling of large deformations because of the lack of connectivity between points, good handling of topological changes, ability to include prior knowledge, support for flexible adaptable refinement procedures, the ability to support multiscale,... (Li & Liu, 2004). Also in computer graphics, meshfree surface representations are gaining importance, especially in the physically based deformable models (Nealen et al., 2006), but also in many other computer graphics subfields (Dyn et al., 2008; Sukumar, 2005). In 3D face recognition, some progressive methods make use of some of these interesting meshfree surface representations as explained in this chapter.

5. References

Akarun, L., Gokberk, B. & Salah, A. A. (2005). 3D face recognition for biometric applications, *EUSIPCO '05: Proceedings of the 13th European Signal Processing Conference*, Antalya.

Akhloufi, M. & Bendada, A. H. (2008). Infrared face recognition using distance transforms, *ICIVC 2008: International Conference on Image and Vision Computing*, Vol. 30 of *Proceedings of World Academy of Science, Engineering & Technology*, Paris, France, pp. 160–163.

Al-Osaimi, F., Bennamoun, M. & Mian, A. (2009). An expression deformation approach to non-rigid 3D face recognition, *International Journal of Computer Vision* **81**(3): 302–316.

Alyüz, N., Gökberk, B. & Akarun, L. (2008). A 3D face recognition system for expression and occlusion invariance, *BTAS '08: Proceedings of the IEEE Second International Conference on Biometrics Theory, Applications and Systems*, Arlington, Virginia, USA.

Amberg, B., Knothe, R. & Vetter, T. (2008). Expression invariant 3D face recognition with a morphable model, *FG '08: Proceedings of the 8th IEEE International Conference on Automatic Face and Gesture Recognition*, IEEE Computer Society, Amsterdam, The Netherlands.

Amenta, N., Choi, S. & Kolluri, R. K. (2001). The power crust, *SMA '01: Proceedings of the sixth ACM symposium on Solid modeling and applications*, ACM, New York, NY, USA, pp. 249–266.

Beatson, R. & Greengard, L. (1997). A short course on fast multipole methods, *Wavelets, Multilevel Methods and Elliptic PDEs*, Oxford University Press, pp. 1–37.

Berretti, S., Bimbo, A. D. & Pala, P. (2006). Description and retrieval of 3D face models using iso-geodesic stripes, *MIR '06: Proceedings of the 8th ACM international workshop on Multimedia information retrieval*, ACM, New York, NY, USA, pp. 13–22.

Besl, P. (1990). The free-form surface matching problem, *Machine vision for three-dimensional scenes* pp. 25–71.

Besl, P. J. & McKay, N. D. (1992). A method for registration of 3-D shapes, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **14**(2): 239–256.

Blinn, J. F. (1982). A generalization of algebraic surface drawing, *ACM Trans. Graph.* **1**(3): 235–256.

Bloomenthal, J. & Wyvill, B. (eds) (1997). *Introduction to Implicit Surfaces*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Bronstein, A. M., Bronstein, M. M. & Kimmel, R. (2003). Expression-invariant 3D face recognition, *in* J. Kittler & M. Nixon (eds), *AVBPA '03: Proceedings of the 4th International Conference on Audio and Video-based Biometric Person Authentication*, Vol. 2688 of *Lecture Notes in Computer Science*, Springer, pp. 62–69.

Bronstein, A. M., Bronstein, M. M. & Kimmel, R. (2005). Three-dimensional face recognition, *International Journal of Computer Vision* **64**(1): 5–30.

Buhmann, M. (2003). *Radial Basis Functions: Theory and Implementations*, Cambridge University Press.

Bülow, T. & Daniilidis, K. (2001). Surface representations using spherical harmonics and gabor wavelets on the sphere, *Technical Report MS-CIS-01-37*, University of Pennsylvania, Department of Computer and Information Science.

Campbell, R. J. & Flynn, P. J. (2000). Survey of free-form object representation and recognition techniques, *Computer Vision and Image Understanding* **81**: 166–210.

Chang, K. I., Bowyer, K. W. & Flynn, P. J. (2006). Multiple nose region matching for 3D face recognition under varying facial expression, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(10): 1695–1700.

Claes, P. (2007). *A robust statistical surface registration framework using implicit function representations - application in craniofacial reconstruction*, PhD thesis, Katholieke Universiteit Leuven, Leuven, Belgium.
URL: *http://www.medicalimagingcenter.be/PhD/PeterClaes/Thesis.pdf*

Colbry, D. & Stockman, G. C. (2007). Canonical face depth map: A robust 3D representation for face verification, *CVPR '07: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, Minneapolis, Minnesota, USA.

Curless, B. & Levoy, M. (1996). A volumetric method for building complex models from range images, *SIGGRAPH '96: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, ACM, New York, NY, USA, pp. 303–312.

Dey, T., Giesen, J. & Goswami, S. (2003). Shape segmentation and matching with flow discretization, *Algorithms and Data Structures* pp. 25–36.
URL: *http://www.springerlink.com/content/6qwfh2adqmm60n30*

Dillenseger, J.-L., Guillaume, H. & Patard, J.-J. (2006). Spherical harmonics based intrasubject 3-d kidney modeling/registration technique applied on partial information, *Biomedical Engineering, IEEE Transactions on* **53**(11): 2185–2193.

Dyn, N., Iske, A. & Wendland, H. (2008). Meshfree thinning of 3d point clouds, *Foundations of Computational Mathematics* **8**(4): 409–425.

Edelsbrunner, H. & Mücke, E. P. (1994). Three-dimensional alpha shapes, *ACM Trans. Graph.* **13**(1): 43–72.

Fabio, R. (2003). From point cloud to surface: the modeling and visualization problem, *Workshop on Visualization and Animation of Reality-based 3D Models*, Tarasp-Vulpera, Switzerland.

Fabry, T., Vandermeulen, D. & Suetens, P. (2008). 3D face recognition using point cloud kernel correlation, *BTAS '08: Proceedings of the IEEE Second International Conference on Biometrics Theory, Applications and Systems*, Arlington, Virginia, USA.

Faltemier, T., Bowyer, K. W. & Flynn, P. J. (2008). A region ensemble for 3-D face recognition, *IEEE Transactions on Information Forensics and Security* **3**(1): 62–73.

Far (2004). *FastRBF MATLAB Toolbox Manual*.

Feng, S., Krim, H., Gu, I. & Viberg, M. (2006). 3D face recognition using affine integral invariants, *ICASSP '06: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 2, Toulouse, France, pp. 189–192.

Feng, S., Krim, H. & Kogan, I. A. (2007). 3D face recognition using euclidean integral invariants signature, *SSP '07: IEEE/SP 14th Workshop on Statistical Signal Processing*, Madison, WI, USA, pp. 156–160.

Gorelick, L., Galun, M. & Brandt, A. (2006). Shape representation and classification using the poisson equation, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(12): 1991–2005. Member-Eitan Sharon and Member-Ronen Basri.

Gu, X., Gortler, S. J. & Hoppe, H. (2002). Geometry images, *ACM Trans. Graph.* **21**(3): 355–361.

Gupta, S., Aggarwal, J. K., Markey, M. K. & Bovik, A. C. (2007). 3D face recognition founded on the structural diversity of human faces, *CVPR '07: Proceedings of the International Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, Minneapolis, Minnesota, USA.

Hansen, M. F., Erbou, S. G. H., Vester-Christensen, M., Larsen, R., Ersboll, B. K. & Christensen, L. B. (2007). Surface-to-surface registration using level sets, *in* B. K. Ersboll & K. S. Pedersen (eds), *SCIA '07: Proceedings of the 15th Scandinavian Conference on Image Analysis*, Vol. 4522 of *Lecture Notes in Computer Science*, Springer, pp. 780–788.

Hoppe, H., DeRose, T., Duchamp, T., McDonald, J. & Stuetzle, W. (1992). Surface reconstruction from unorganized points, *SIGGRAPH '92: Proceedings of the 19th annual conference on Computer graphics and interactive techniques*, ACM, New York, NY, USA, pp. 71–78.

Hubeli, A. & Gross, M. (2000). A survey of surface representations for geometric modeling, *Technical Report 335*, ETH Zürich.

Irfanoglu, M., Gokberk, B. & Akarun, L. (2004). 3D shape-based face recognition using automatically registered facial surfaces, *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on* **4**: 183–186 Vol.4.

Jahanbin, S., Choi, H., Liu, Y. & Bovik, A. C. (2008). Three dimensional face recognition using iso-geodesic and iso-depth curves, *BTAS '08: Proceedings of the IEEE Second International Conference on Biometrics Theory, Applications and Systems*, Arlington, Virginia, USA.

Jones, M. W., Baerentzen, J. A. & Sramek, M. (2006). 3D distance fields: A survey of techniques and applications, *IEEE Transactions on Visualization and Computer Graphics* **12**(4): 581–599.

Jorda, A. R., Vanacloig, V. A. & García, G. A. (2001). A geometric feature set based on shape description primitives to recognize natural objects, *Proceedings of the IX Spanish Symposium on Pattern Recognition and Image Analysis*, Benicasim, Spain.

Kakadiaris, I. A., Passalis, G., Toderici, G., Murtuza, M. N., Lu, Y., Karampatziakis, N. & Theoharis, T. (2007). Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **29**(4): 640–649.

Kalaiah, A. & Varshney, A. (2003). Statistical point geometry, *SGP '03: Proceedings of the 2003 Eurographics/ACM SIGGRAPH symposium on Geometry processing*, Eurographics Association, pp. 107–115.

Kazhdan, M., Funkhouser, T. & Rusinkiewicz, S. (2003). Rotation invariant spherical harmonic representation of 3d shape descriptors, *SGP '03: Proceedings of the 2003 Eurographics/ACM SIGGRAPH symposium on Geometry processing*, Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, pp. 156–164.

Kimmel, R. & Sethian, J. A. (1998). Computing geodesic paths on manifolds, *Proceedings of the National Academy of Sciences of the United States of America* **95**: 8431–8435.

Levoy, M. & Whitted, T. (1985). The use of points as a display primitive, *Technical report*, Department of Computer Science, University of North Carolina at Chappel Hill.

Li, L., Xu, C., Tang, W. & Zhong, C. (2008). 3d face recognition by constructing deformation invariant image, *Pattern Recognition Letters* **29**(10): 1596–1602.

Li, S. & Liu, W. (2004). *Meshfree particle methods*, Springer.

Li, X. & Zhang, H. (2007). Adapting geometric attributes for expression-invariant 3D face recognition, *SMI '07: Proceedings of the IEEE International Conference on Shape Modeling and Applications*, IEEE Computer Society, Washington, DC, USA, pp. 21–32.

Liu, D., Shen, L. & Lam, K. (2005). Image Synthesis and Face Recognition Based on 3D Face Model and Illumination Model, *ICNC 2005*, Springer, p. 7.

Liu, S., Jin, X., Wang, C. & Hui, K. (2007). Ellipsoidal-blob approximation of 3D models and its applications, *Computers & Graphics* **31**(2): 243–251.

Llonch, R. S., Kokiopoulou, E., Tosic, I. & Frossard, P. (2009). 3D face recognition with sparse spherical representations, *Pattern Recognition* .

Lu, X. & Jain, A. K. (2005). Deformation analysis for 3d face matching, *WACV-MOTION '05: Proceedings of the Seventh IEEE Workshops on Application of Computer Vision (WACV/MOTION'05) - Volume 1*, IEEE Computer Society, Washington, DC, USA, pp. 99–104.

Lu, X. & Jain, A. K. (2006). Deformation modeling for robust 3D face matching, *CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, Washington, DC, USA, pp. 1377–1383.

Marr, D. (1982). *Vision: a computational investigation into the human representation and processing of visual information*, W. H. Freeman, San Francisco.

Maurer, T., Guigonis, D., Maslov, I., Pesenti, B., Tsaregorodtsev, A., West, D. & Medioni, G. (2005). Performance of Geometrix ActiveID^{TM} 3D face recognition engine on the FRGC data, *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*, IEEE Computer Society, Washington, DC, USA, p. 154.

Metaxas, D. N. & Kakadiaris, I. A. (2002). Elastically adaptive deformable models, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(10): 1310–1321.

Mousa, M.-H., Chaine, R., Akkouche, S. & Galin, E. (2007). Efficient spherical harmonics representation of 3d objects, *PG '07: Proceedings of the 15th Pacific Conference on Computer Graphics and Applications*, IEEE Computer Society, Washington, DC, USA, pp. 248–255.

Mpiperis, I., Malasiotis, S. & Strintzis, M. G. (2007). 3-D face recognition with the geodesic polar representation, *IEEE Transactions on Information Forensics and Security* **2**(3): 537–547.

Mpiperis, I., Malassiotis, S. & Strintzis, M. G. (2008). Bilinear models for 3-D face and facial expression recognition., *IEEE Transactions on Information Forensics and Security* **3**(3): 498–511.

Muraki, S. (1991). Volumetric shape description of range data using "blobby model", *SIGGRAPH Comput. Graph.* **25**(4): 227–235.

Nealen, A., Mueller, M., Keiser, R., Boxerman, E. & Carlson, M. (2006). Physically based deformable models in computer graphics, *Computer Graphics Forum* **25**(4): 809–836.
URL: *http://dx.doi.org/10.1111/j.1467-8659.2006.01000.x*

Nishita, T. & Nakamae, E. (1994). A method for displaying metaballs by using bezier clipping, *Computer Graphics Forum* **13**: 271–280.

Osher, S. & Fedkiw, R. (2003). *Level Set Methods and Dynamic Implicit Surfaces*, Vol. 153 of *Applied Mathematical Sciences*, Springer-Verlag New York.

Parzen, E. (1962). On estimation of a probability density function and mode, *The Annals of Mathematical Statistics* **33**(3): 1065–1076.
URL: *http://www.jstor.org/stable/2237880*

Pauly, M., Keiser, R. & Gross, M. (2003). Multi-scale feature extraction on point-sampled surfaces, *Computer Graphics Forum*, Vol. 22, Blackwell Publishing, Inc, pp. 281–289.

Pears, N. (2008). Rbf shape histograms and their application to 3d face processing, pp. 1–8.

Pears, N. & Heseltine, T. (2006). Isoradius contours: New representations and techniques for 3D face registration and matching, *3DPVT'06: Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission*, IEEE Computer Society, Washington, DC, USA, pp. 176–183.

Perakis, P., Passalis, G., Theoharis, T., Toderici, G. & Kakadiaris, I. (2009). Partial matching of interpose 3d facial data for face recognition, *BTAS 2009: Proceedings of the IEEE Third International Conference on Biometrics: Theory, Applications and Systems*, Washington DC.

Phillips, P., Grother, P., Micheals, R., Blackburn, D., Tabassi, E. & Bone, J. (2003). FRVT 2002: Evaluation report, *Technical Report NISTIR 6965*, NIST.

Rusinkiewicz, S. & Levoy, M. (2001). Efficient variants of the ICP algorithm, *3DIM '01: Proceedings of the Third International Conference on 3D Digital Imaging and Modeling*, pp. 145–152.

Russ, T., Boehnen, C. & Peters, T. (2006). 3D face recognition using 3D alignment for PCA, *CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, Washington, DC, USA, pp. 1391–1398.

Samir, C., Srivastava, A. & Daoudi, M. (2006). Three-dimensional face recognition using shapes of facial curves, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(11): 1858–1863.

Sigg, C. (2006). *Representation and Rendering of Implicit Surfaces*, PhD thesis, ETH Zürich.

Silverman, B. (1986). *Density estimation for statistics and data analysis*, Chapman & Hall/CRC.

Smeets, D., Fabry, T., Hermans, J., Vandermeulen, D. & Suetens, P. (2009). Isometric deformation modeling using singular value decomposition for 3d expression-invariant face recognition, *Biometrics: Theory, Applications, and Systems, 2009. BTAS '09. IEEE 3rd International Conference on*, pp. 1–6.

Sukumar, N. (2005). Maximum entropy approximation, *Bayesian Inference and Maximum Entropy Methods in Science and Engineering* **803**: 337–344.

ter Haar, F. B. & Veltkamp, R. C. (2008). SHREC'08 entry: 3D face recognition using facial contour curves, *SMI '08: Proceedings of the IEEE International Conference on Shape Modeling and Applications*, Stony Brook, NY, USA, pp. 259–260.

Tsin, Y. & Kanade, T. (2004). A correlation-based approach to robust point set registration, *ECCV (3)*, pp. 558–569.

Vandermeulen, D., Claes, P., Loeckx, D., De Greef, S., Willems, G. & Suetens, P. (2006). Computerized craniofacial reconstruction using CT-derived implicit surface representations, *Forensic Science International* (159): S164–S174.

Varshosaz, M., Helali, H. & Shojaee, D. (2005). The methods of triangulation, *Map Middle East '05: Proceedings of the 1st Annual Middle East Conference and Exhibition on Geospatial Information, Technology and Applications*, Dubai, UAE.

Veleba, D. & Felkel, P. (2007). Survey of errors in surface representation and their detection and correction, *WSCG '07: Proceedings of the 15th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, Plzen-Bory, Czech Republic.

Walder, C., Schölkopf, B. & Chapelle, O. (2006). Implicit surface modelling with a globally regularised basis of compact support, *Computer Graphics Forum* **25**(3): 635–644.

Wang, F., Vemuri, B., Rangarajan, A. & Eisenschenk, S. (2008). Simultaneous nonrigid registration of multiple point sets and atlas construction, *IEEE Transactions onPattern Analysis and Machine Intelligence* **30**(11): 2011–2022.

Wang, Y., Pan, G., Wu, Z. & Wang, Y. (2006). Exploring facial expression effects in 3D face recognition using partial ICP, *in* P. Narayanan (ed.), *Computer Vision Ű ACCV 2006*, Vol. 3851 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, pp. 581–590.

Yang, Y., Yong, J., Zhang, H., Paul, J. & Sun, J. (2006). Optimal parameterizations of bézier surfaces, pp. I: 672–681.

Yano, K. & Harada, K. (2007). Single-patch nurbs face, *Signal-Image Technologies and Internet-Based System, International IEEE Conference on* **0**: 826–831.

16

An Integrative Approach to Face and Expression Recognition from 3D Scans

Chao Li
Florida A&M University
USA

1. Introduction

Face recognition, together with fingerprint recognition, speaker recognition, etc., is part of the research area known as 'biometric identification' or 'biometrics', which refers to identifying an individual based on his or her distinguishing characteristics. More precisely, biometrics is the science of identifying, or verifying the identity of, a person based on physiological or behavioral characteristics (Bolle et al., 2003). Biometric characteristics include something that a person is or produces. Examples of the former are fingerprints, the iris, the face, the hand/finger geometry or the palm print, etc. The latter include voice, handwriting, signature, etc. (Ortega-Garcia et al., 2004).

Face recognition is a particularly compelling biometric approach because it is the one used every day by nearly everyone as the primary means for recognition of other humans. Because of its natural character, face recognition is more acceptable than most other biometric methods. Face recognition also has the advantage of being noninvasive.

Face recognition has a wide range of potential applications for commercial, security, and forensic purposes. These applications include automated crowd surveillance, access control, mug shot identification (e.g., for issuing driver licenses), credit card authorization, ATM machine access control, design of human computer interfaces (HCI), etc. Especially, the surveillance systems rely on the noninvasive property of face recognition systems.

According to the different purposes in applications, face recognition scenarios can be classified into the following two:

- Face verification: ("Am I who I say I am?") is a one-to-one match that compares a query face image against a gallery face image whose identity is being claimed.
- Face identification: ("Who am I?") is a one-to-many matching process that compares a query face image against all the gallery images in a face database to determine the identity of the query face. In the identification task, it is assumed that the person is in the database. The identification of the query image is done by choosing the image in the database that has the highest similarity score with the query image.

According to the format of the data, analyzed face recognition methods can be classified as 2D face recognition, 3D face recognition and infrared face recognition modalities. The infrared face recognition is commonly combined with other biometrics technologies.

Most of the face recognition methods developed until recently use 2D intensity images obtained by photographic cameras as the data format for processing. There are two reasons for the interest in 2D face recognition. First of all, human beings can recognize a person from his or her picture alone, which means that the picture contains enough information about a person's identity for a human being. Second, a picture is very easy to obtain in terms of acquisition and cost. This was even more important decades ago, when other imaging techniques, such as 3D imaging, were not well developed.

Varying levels of success have been achieved in 2D face recognition research. A short overview is given in the following paragraph. More detailed and comprehensive surveys can be found in (Chellappa et al., 1995; Zhao et al., 2003). Most proposed techniques fall into two lines of research. The first one is appearance based (view based) face recognition. This approach represents a face image as a high dimensional vector, i.e., a point in a high dimensional vector space. Statistical techniques are then used to analyze the distribution of the face image in the vector space, and features are extracted in the vector space using linear decomposition like Principal Component Analysis (PCA) (also called 'the Eigen Face' method proposed in (Turk and Pentland, 1991)), Independent Component Analysis (ICA) (Bartlett, et al., 1998) or non linear methods like Kernel Principal Component Analysis (KPCA) (Cholkopf et al., 1998). The second group of approaches is model-based. The model-based face recognition scheme is aimed at constructing a model of the human face, which can capture the facial variations of individuals. Exemplar approaches in this category include Feature-based Elastic Bunch Graph Matching (Wiskott et al. 1997) and Active Appearance Model (AAM).(Edwards et al., 1998)

In 2002, the Face Recognition Vendor Test (FRVT 2002) was held. It was an independently administered technology evaluation sponsored by the Defense Advanced Research Projects Agency (DARPA), the National Institute of Standards and Technology (NIST) and other agencies. The primary objective of FRVT 2002 was to provide performance measures for assessing the ability of automatic face recognition systems to meet real-world requirements. FRVT 2002 measured the performance of the core capabilities of face recognition technology, all based on 2D face recognition. Ten participants were evaluated. (Phillips et al., 2002)

FRVT 2002 showed that one of the most challenging tasks for modern face recognition systems is recognizing faces in non-frontal imagery. Most face recognition systems performed well when all of the images were frontal. But, as a subject became more and more off angle (both horizontally and vertically), performance decreased. Additionally FRVT 2002 also showed that the variation in the structure of lighting had a great effect on performance (Phillips et al., 2002).

From these observations, the following conclusions can be drawn. Although 2D face recognition has achieved considerable success, certain problems still exist. Because the 2D face images used not only depend on the face of a subject, but also depend on the imaging factors, such as the environmental illumination and the orientation of the subject. These two sources of variability in the face image often make the 2D face recognition system fail. That is the reason why 3D face recognition is believed to have an advantage over 2D face recognition. Although the first main rationale for 2D face recognition mentioned above, i.e. "a picture contains enough information about the identity of a person", may be true for a human being, for an artificial system, without the biological knowledge of how the human vision system works, 2D face recognition is still a very difficult task.

With the development of 3D imaging technology, more and more attention has been directed to 3D face recognition. In (Bowyer et al., 2004), Bowyer et al. provide a survey of 3D face recognition technology. Some of the techniques are derived from 2D face recognition, such as the use of PCA in (Hesher et al., 2003; Chang et al. 2003) to extract features from faces. Some of the techniques are unique to 3D face recognition, such as the geometry matching method in (Gordon, 1991) and the profile matching proposed in (Cartoux et al. 1989; Nagamine et al. 1992).

Most of the 3D face recognition systems treat the 3D face surface as a rigid surface. But actually, the face surface is deformed by different expressions of the subject. So, systems which treat the face as a rigid surface are significantly challenged when dealing with faces with expressions. In (Chang et al., 2005), experiments using Iterative Closest Point (ICP) and PCA methods were performed on the recognition of faces with expression. The authors found that expression changes do cause performance declines in all the methods. (Chang et al., 2005)

Therefore, expression has become a big challenge in 3D face recognition systems. Up to now, only some methods address the facial expression issue in face recognition. In (Bronstein et al., 2003), the authors present a 3D face recognition approach based on a representation of the facial surface, invariant to isometric deformation by facial expression. In (Lu and Jain, 2005), both rigid registration and non-rigid deformations caused by expression were calculated. Iterative Closest Point (ICP) was used to perform rigid registration. For non-rigid deformation, the thin plated spline (TPS) was applied. For the purpose of face matching, the non-rigid deformations from different sources were identified, which was formulated as a two-class classification problem: intra-subject deformation vs. inter-subject deformation. The deformation classification results were integrated with the matching distance of rigid registration to make the final decision. In (Chang et al., 2005), the author tried to extract the area that deforms least with different facial expressions and used this area as the feature for every subject. Then ICP and PCA methods were applied for the matching.

In our research, we want to tackle the expression challenge in 3D face recognition from a different point of view. Because the deformation of the face surface is always related with different expressions, an integrated expression recognition and face recognition system is proposed. In section 2, a model on the relationship between expression and face recognition is introduced. Based on this model, the framework of integrated expression recognition and face recognition is proposed. Section 3 explains the acquisition of the experimental data used and preprocessing performed. Section 4 outlines our approach to 3D face expression recognition. Section 5 explains the process used for 3D face recognition. Section 6 describes the experiments and the results obtained. Section 7 presents our discussion and conclusion.

2. Relationship between expression recognition and face recognition

From the psychological point of view, it is still not known whether facial expression recognition information directly impacts the face recognition process in human beings. Some models suggest there is not relationship between face recognition and facial expression recognition (Bruce and Young, 1986). Other models support the opinion that a connection exists between the two processes (Hansch and Pirozzolo, 1980).

One of the experiments that support the existence of the connection between facial expression recognition and face recognition was reported in (Etcoff and Magee, 1992). The authors found that people are slower in identifying happy and angry faces than they are in identifying faces with neutral expression. Also, in (Hay et al., 1991) experiments show that people are slower in identifying pictures of familiar faces when they exhibit uncommon facial expressions.

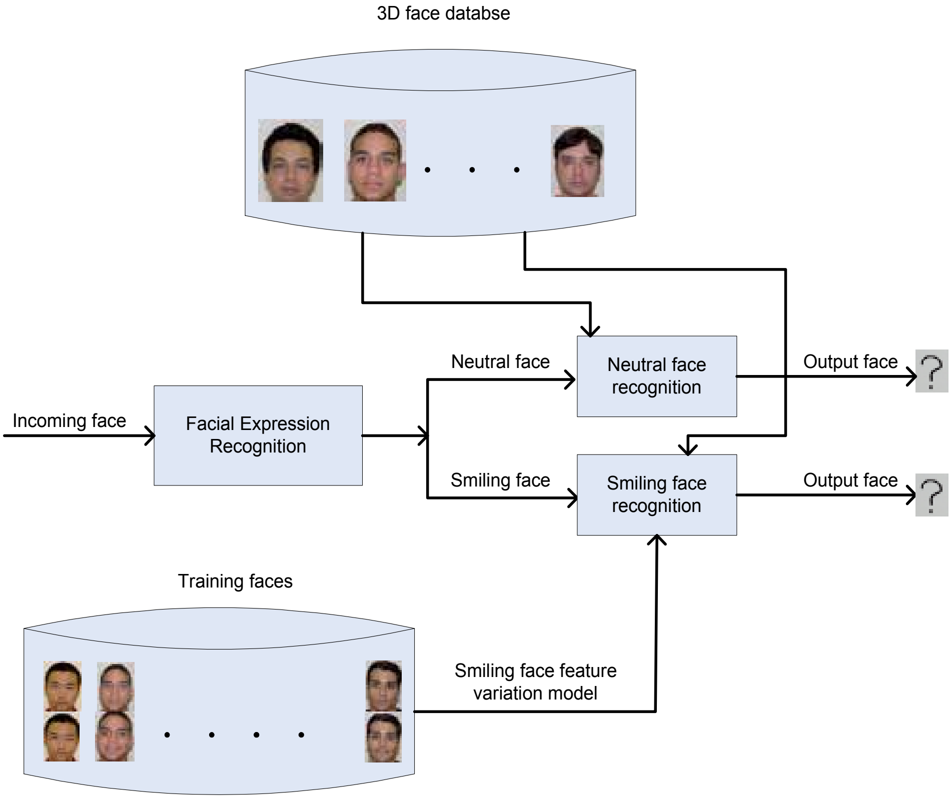


Fig. 1. Simplified framework of 3D face expression

Our proposed framework is based on the assumption that the identification of the facial expression of a query face will aid an automated face recognition system achieve its goal.

The incoming 3D range image is first processed by an expression recognition system to find the most appropriated expression label for it. The expression label could be one of the six prototypical expressions of the faces, which are happiness, sadness, anger, fear, surprise and disgust (Ekman and Friesen, 1971). In addition, the face could also be labeled as 'neutral'. Therefore, the output of the expression recognition system will be one of the seven expressions. Our framework proposes that a different face recognition approach be used for each type of expression. If the expression label determined is neutral expression, then the incoming 3D range image is directly passed to a neutral expression face recognition system, which uses the features of the probe image to match those of the gallery images and get the closest match. If the expression label determined is other than neutral expression, then for each of the six prototypical expressions, a separate face recognition subsystem should be used. The system will find the right face by modeling the variations of the face features between the neutral face and the expressional face. Because recognition through modeling is a more complex process than the direct matching for the neutral face, our framework aligns with the view that people will be slower in identifying happy and angry faces than they will be in identifying faces with neutral expression. Figure 1 shows a simplified version of this framework, it only deals with happy (smiling) expressions in addition to neutral. Smiling is the most common (non-neutral) expression displayed by people in public.

3. Data acquisition and preprocessing

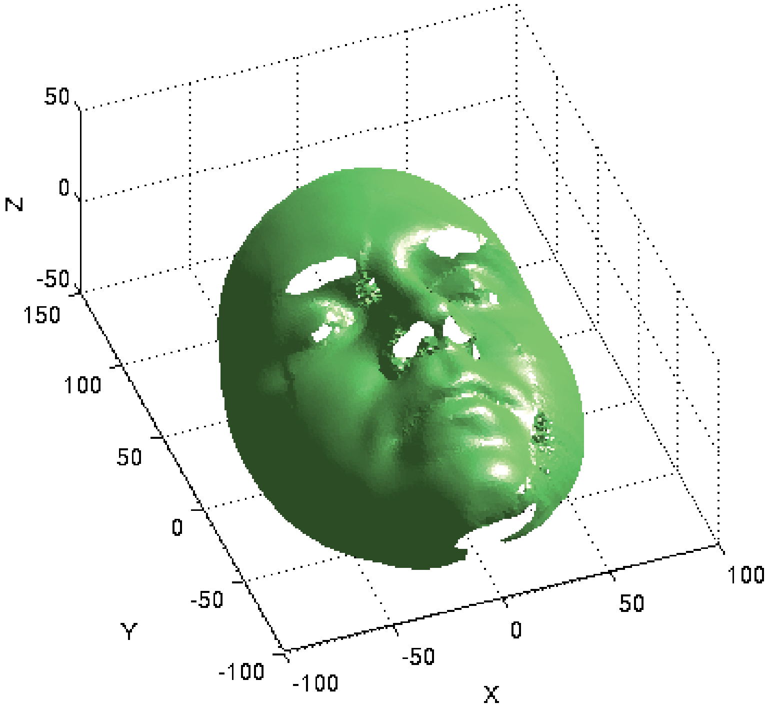


Fig. 2. 3D face surface acquired by the 3D scanner

To test the performance of our framework, a database including images from 30 subjects, was built. In this database, we included faces with the most common expression i.e., smiling, as well as neutral faces from the same subjects. Each subject participated in two data acquisition sessions, which took place in two different days. In each session, two 3D scans were acquired. One was neutral expression; the other was a happy (smiling) expression. The 3D scanner used was a Fastscan 3D scanner from Polhemus Inc. The accuracy of this scanner is specified as 1mm. The resulting database contains 60 3D neutral scans and 60 3D smiling scans of 30 subjects. Figure 2 shows an example of the 3D scans obtained using this scanner.

In 3D face recognition, registration is a key pre-processing step. It may be crucial to the efficiency of matching methods. In our experiment, a method based on the symmetric property of the face is used to register the face image. In converting the 3D scan from triangulated mesh format to a range image with a sampling interval of 2.5 mm(e.g., Fig 2), trilinear interpolation was used (Li and Barreto, 2005). Unavoidably, the scanning process will result in face surfaces containing unwanted holes, especially in the area covered by dark hair, such as the eye brows. To circumvent this problem, the cubic spline interpolation method was used to patch the holes. An example of the resulting 3D range image is shown in Fig 3.

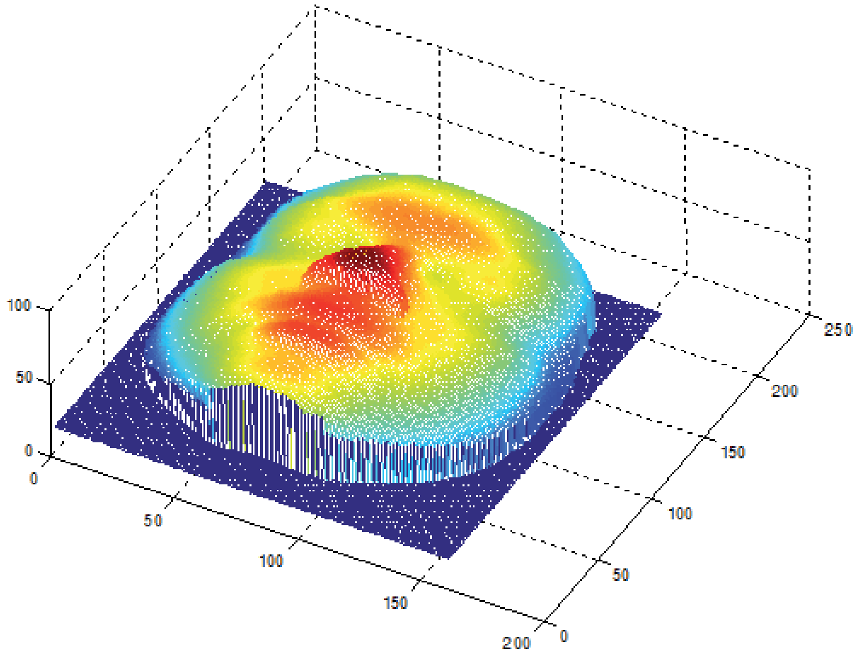


Fig. 3. Mesh plot of the converted range image

4. 3D expression recognition

Facial expression of the face is a basic mode of nonverbal communication among people. The facial expression of another person is often the basis on which we form significant opinions on such characteristics as friendliness, trustworthiness, and status. The facial expressions convey information about emotion, mood and ideas.

In (Ekman and Friesen, 1971), Ekman and Friesen proposed six primary emotions. Each possesses a distinctive content together with a unique facial expression. These prototypical emotional displays are also referred to as basic emotions. They seem to be universal across human ethnicities and cultures. These six emotions are happiness, sadness, fear, disgust, surprise and anger. Together with the neutral expression, these seven expressions also form the basic prototypical facial expressions.

Facial expressions are generated by contractions of facial muscles, which result in temporally deformed facial features such as eye lids, eye brows, nose, lips and skin textures, often revealed by wrinkles and bulges. Typical changes of muscular activities for spontaneous expressions are brief, usually between 250ms and 5s. Three stages have been defined for each expression, which are onset (attack), apex (sustain) and offset (relaxation). In contrast to these spontaneous expressions, posed or deliberate expressions can be found very commonly in social interactions. These expressions typically last longer than spontaneous expressions.

Automatic facial expression recognition has gained more and more attention recently. It has various potential applications in improved intelligence for human computer interface, image compression and synthetic face animation. “Face expression recognition deals with the classification of facial motion and facial feature deformation into abstract classes that are purely based on visual information.” (Fasel and Luettin, 2003).

As in face recognition, most contemporary facial expression recognition systems use two-dimensional images or videos as data format. Therefore, the same challenges exist for the face expression recognition as for face recognition, i.e. 2D formats are dependent on the pose of the subjects and on the illumination of the environment. In this respect this paper fills the gap by proposing a facial expression recognition system that uses three dimensional images or range images. 3D range images have the advantage of invariance with respect to subject alignment and illumination. In addition, the deformed features resulting from expressions are easy to extract from 3D range images.

In our experiment, we sought to recognize social smiles, which were posed by each subject, in their apex period. Smiling is the easiest of all expressions to find in photographs and is readily produced by people on demand. The smile is generated by the contraction of the Zygomatic Major muscle. The Zygomatic Major originates in the cheek bone (Zygomatic arch) and inserts in muscles near the corner of the mouth. This muscle lifts the corner of the mouth obliquely upwards and laterally, producing a characteristic “smiling expression”. So the most distinctive features associated with a smile are the bulge of the cheek muscle and the uplift of the corner of the mouth, as can be seen in Fig 4. The line on the face generated by a smiling expression is called the nasal labial fold (smile line).

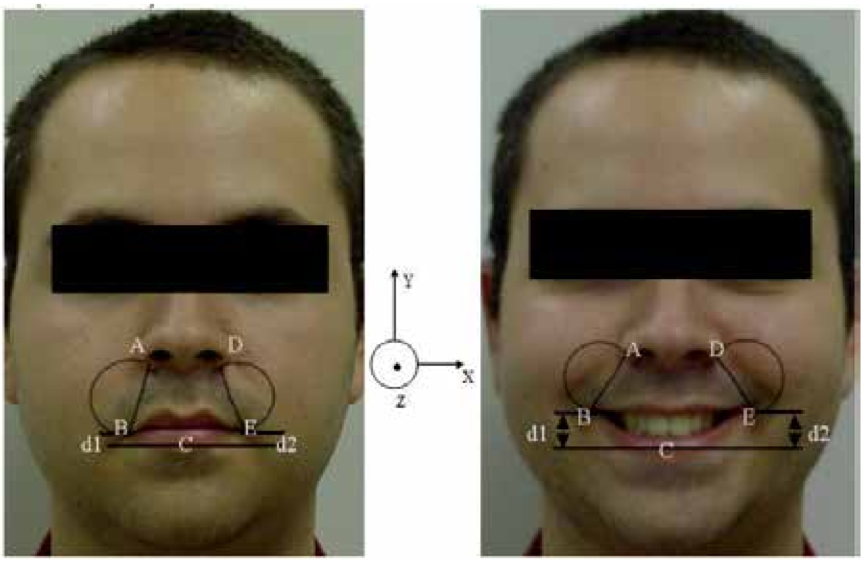


Fig. 4. Illustration of the features of a smiling face

The following steps are followed to extract the features for the smiling expression:

1. An algorithm is developed to obtain the coordinates of five characteristic points A, B, C, D and E in the face range image as shown in Figure 4. A and D are at the extreme points of the base of the nose. B and E are the points defined by the corners of the mouth. C is in the middle of the lower lip.
2. The first feature is the width of the mouth BE normalized by the length of AD. Obviously, while smiling the mouth becomes wider. The first feature is represented by *mw.*
3. The second feature is the depth of the mouth (The difference between the Z coordinates of points BC and EC) normalized by the height of the nose to capture the fact that the smiling expression pulls back the mouth. This second feature is represented by *md.*
4. The third feature is the uplift of the corner of the mouth, compared with the middle of the lower lip d1 and d2, as shown in the figure, normalized by the difference of the Y coordinates of points AB and DE, respectively and represented by *lc.*
5. The fourth feature is the angle of AB and DE with the central vertical profile, represented by *ag*.
6. The last two features are extracted from the semicircular areas shown, which are defined by using AB and DE as diameters. The histograms of the range (Z coordinates) of all the points within these two semicircles are calculated.

The following figure shows the histograms for the smiling face and the neutral face of the above subject.

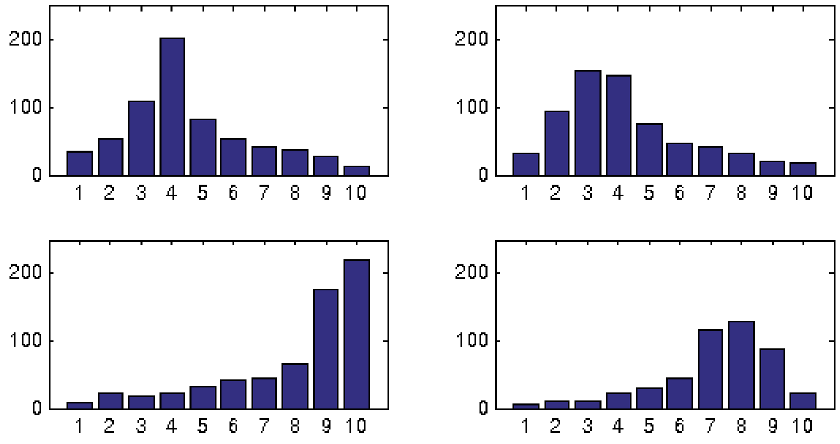


Fig. 5. Histogram of range of cheeks for neutral and smiling face

The two figures in the first row are the histograms of the range values for the left cheek and right cheek of the neutral face image; the two figures in the second row are the histograms of the range values for the left cheek and right cheek of the smiling face image.

From the above figures, we can see that the range histograms of the neutral and smiling expressions are different. The smiling face tends to have large values at the high end of the histogram because the bulge of the cheek muscle. On the other hand, a neutral face has large values at the low end of the histogram distribution. Therefore two features can be obtained from the histograms: One is called the 'histogram ratio', represented by *hr*, the other is called the 'histogram maximum', represented by *hm*.

$$hr = \frac{h6 + h7 + h8 + h9 + h10}{h1 + h2 + h3 + h4 + h5} \tag{1}$$

$$hm = i \quad ; \qquad i = \arg\{\max(h(i))\} \tag{2}$$

In summary, six features, i.e. *mw*, *md*, *lc*, *ag*, *hr* and *hm* are extracted from each face for the purpose of expression recognition.

After the features have been extracted, this becomes a general classification problem. Two pattern classification methods are applied to recognize the expression of the incoming faces.

1. Linear discriminant classifier: (Linear Discriminant Analysis-LDA)

LDA tries to find the subspace that best discriminates different classes by maximizing the between-class scatter matrix S_b, while minimizing the within-class scatter matrix S_w in the projective subspace. S_w and S_b are defined as follows,

$$S_w = \sum_{i=1}^{L} \sum_{\vec{x}_k \in X_i} (\vec{x}_k - \vec{m}_i)(\vec{x}_k - \vec{m}_i)^T \tag{3}$$

$$S_b = \sum_{i=1}^{L} n_i (\vec{m}_i - \vec{m})(\vec{m}_i - \vec{m})^T \tag{4}$$

Where $\vec{m}_i$ is the mean vector for the individual class X_i, and n_i is the number of samples in class X_i, $\vec{m}$ is the mean vector of all the samples. L is the number of classes.

The LDA subspace is spanned by a set of vectors W, satisfying

$$W = \arg\max \left| \frac{W^T S_b W}{W^T S_w W} \right| \tag{5}$$

2. Support Vector Machine (SVM):

Support vector machine is a relatively new technology for classification. It relies on preprocessing the data to represent patterns in a high dimension, typically much higher than the original feature space. With an appropriate nonlinear mapping to a sufficiently high dimension, data from two categories can always be separated by a hyperplane (Duda et al., 2001). In our research, the Libsvm program package (Chang and Lin, 2001) was used to implement the support vector machine.

5. 3D face recognition

5.1 Neutral face recognition

In previous related work, we have found that the central vertical profile and the face contour are both discriminant features for every person (Li et al., 2005a). Therefore, for neutral face recognition, the same method as in (Li et al., 2005b) is used. In this approach, the results of central vertical profile matching and contour matching are combined. The combination of the two classifiers improves the performance noticeably. The final similarity score for the probe image is the product of ranks for each of the two classifiers. The image which has the smallest score in the gallery will be chosen as the matching face for the probe image.

5.2 Smiling face recognition

For the recognition of faces labeled as 'smiling' by the expression recognition module, the probabilistic subspace method proposed by (Moghaddam and Pentland, 1995) is used. The following paragraphs provide an outline of this method and the relative principal component analysis (PCA).

Subspace methods are commonly used in computer vision, including face recognition. A raw 2D image can be represented as a vector in a high dimensional space. In most cases, however, the information which needs to be extracted has a much lower dimension. That is where subspace methods such as *principal component analysis* (PCA), or the previously introduced *linear discriminant analysis* (LDA), can be applied to cope with the problem of reducing excessive dimensionality in the data to be analyzed.

PCA

Unlike LDA, which seeks a set of features that results in the best separation of each class, PCA seeks a projection that best represents the data in a least-square sense. In PCA, a set of vectors are computed from the eigenvectors of the sample covariance matrix C,

$$C = \sum_{i=1}^{M} (\vec{x}_i - \vec{m})(\vec{x}_i - \vec{m})^T \tag{6}$$

where $\vec{m}$ is the mean vector of the sample set. The eigen space Y is spanned by k eigenvectors $u_1 \ u_2 \ldots\ldots u_k$, corresponding to the k largest eigen values of the covariance matrix C.

$$\vec{y}_i = (\vec{x}_i - \vec{m})^T \left[\vec{u}_1 \vec{u}_2 \ldots\ldots \vec{u}_k\right] \tag{7}$$

The dimensionality of vector $\vec{y}_i$ is *K (K<<M).*

Probabilistic subspace method

In (Moghaddam and Pentland, 1995) (Moghaddam and Pentland, 1997), B. Moghaddam et al. presented an unsupervised technique for visual learning, which is based on density estimation in high dimensional spaces using an eigen decomposition. The probability density is used to formulate a maximum-likelihood estimation framework for visual search, target detection and automatic object recognition. Using the probabilistic subspace method, a multi-class classification problem can be converted into a binary classification problem.

Let Δ represents the difference between two vectors in a high dimensional subspace.

$$\Delta = I1 - I2 \tag{8}$$

Δ belongs to the intrapersonal space in the high dimensional subspace if I1 and I2 are two different instances of the same subject; Δ belongs to the interpersonal or extrapersonal space if I1 and I2 are instances from different subjects. $S(\Delta)$ is defined as the similarity between I1 and I2. Using Bayes Rule,

$$S(\Delta) = P(\Omega_I \mid \Delta) = \frac{P(\Delta \mid \Omega_I)P(\Omega_I)}{P(\Delta \mid \Omega_I)P(\Omega_I) + P(\Delta \mid \Omega_E)P(\Omega_E)} \tag{9}$$

$P(\Delta \mid \Omega_I)$ and $P(\Delta \mid \Omega_E)$ are the likelihoods of intrapersonal space and extrapersonal space. The likelihood function can be estimated by traditional means, i.e. maximum likelihood estimation or Parzen window estimation if there are enough data available. In most cases, because of the high dimensionality of the subspace, training data are not sufficient. Subspace density estimation is another choice, which is the case in our experiment. $P(\Omega_I)$ and $P(\Omega_E)$ are *a priori* probabilities for intrapersonal and extrapersonal subspace. Thus, according to the maximum *a posteriori* (MAP) rule, if $P(\Omega_I \mid \Delta)$ is greater than $P(\Omega_E \mid \Delta)$, the two images are considered to be different instances of the same subject, otherwise, they belong to two subjects.

Another method based only on Ω_I can be used to simplify the computation. This maximum-likelihood (ML) similarity measure ignores extrapersonal variations.

$$S^{'}(\Delta) = P(\Delta \mid \Omega_I) \tag{10}$$

In (B.Moghaddam 1995), it was found that the Ω_I density in (10) carries greater weight in modeling the posterior similarity used for MAP recognition. The extrapersonal Ω_E, on the other hand serves a secondary role and its accurate modeling is less critical. By dropping the Ω_E likelihood in favor of an ML similarity, the results typically suffer only a minor deficit in accuracy as compared to $S(\Delta)$.

Subspace density estimation

Given the high dimensionality of Δ, traditional methods are not suitable for the purpose of probability density estimation. An efficient subspace density estimation method proposed in (B.Moghaddam 1995; B.Moghaddam 1997) was used. The vector space of R^N is divided into two complementary subspaces: DIFS (Difference in Feature Space), F, and DFFS (Difference from Feature Space), $\overline{F}$,as show in the figure.

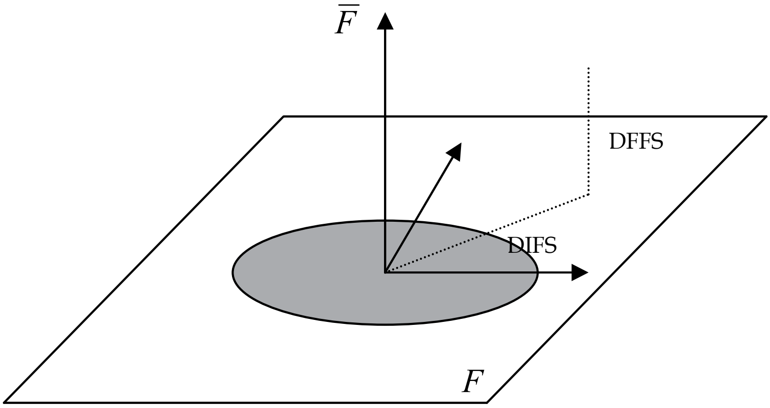


Fig. 6. The principal subspace F and its orthogonal complement $\overline{F}$ for a Gaussian density

F is spanned by the first M (M<<N) eigen vectors corresponding to the largest M eigen values of principal component decomposition results.
As derived in(B. Moghaddam 1995), the complete likelihood estimate can be written as the product of two independent marginal Gaussian densities

$$\hat{P}(\Delta\,|\,\Omega)=\left[\frac{\exp\left(-\frac{1}{2}\sum_{i=1}^{M}\frac{y_i^2}{\lambda_i}\right)}{(2\pi)^{\frac{M}{2}}\prod_{i=1}^{M}\lambda_i^{1/2}}\right]\left[\frac{\exp\left(-\frac{\varepsilon^2(\Delta)}{2\rho}\right)}{2\pi\rho^{(N-M)/2}}\right]=P_F(\Delta\,|\,\Omega)\hat{P}_{\bar{F}}(\Delta\,|\,\Omega;\rho) \tag{11}$$

where $P_F(\Delta\,|\,\Omega)$ is the true marginal density in F, $\hat{P}_{\bar{F}}(\Delta\,|\,\Omega;\rho)$ is the estimated marginal density in the orthogonal complement $\bar{F}$, y_i are the principal components and $\varepsilon^2(\Delta)$ is the PCA residual . From (B.Moghaddam 1995), the optimal value for ρ is the average of the $\bar{F}$ eigen values.

$$\rho=\frac{1}{N-M}\sum_{i=M+1}^{N}\lambda_i \tag{12}$$

In the experiment for smiling face expression recognition, because of the limited number of subjects (30), the central vertical profile and the contour are not used directly as vectors in a high dimensional subspace. Instead, they are down sampled to a dimension of 17 for the analysis. The dimension of subspace F is set to be 10, which contains approximately 97% of the total variance. The dimension of complementary subspace $\bar{F}$ is 7. In this case also, independent ranks are computed for the central profile and the contour of each gallery face. The overall rank is found by sorting the product of these two ranks and is used to determine the final recognition result.

6. Experiments and Results

One gallery and three probe databases are formed for the evaluation of our methods. The gallery database has 30 neutral faces, one for each subject, acquired in the first data acquisition session. Three probe sets are constituted as follows and used in experiments 2 and 3.

Probe set 1: 30 neutral faces acquired in the second session.
Probe set 2: 30 smiling faces acquired in the second session.
Probe set 3: 60 faces, constituted by probe set 1 and probe set 2.
The following experiments are undertaken:

Experiment 1: Testing the expression recognition module

The leave-one-out cross validation method is used to test the expression recognition classifier. Every time, the faces collected from 29 subjects in both data acquisition sessions are used to train the classifier and the four faces of the remaining subject collected in both sessions are used to test the classifier. The results shown below are the average of the 30 recognition rates. Two classifiers are used. One is the linear discriminant classifier; the other is a support vector machine classifier. They have similar performance of over 90% recognition rate.

Methods	LDA	SVM
Expression recognition rate	90.8%	92.5%

Table 1. Expression recognition rate

Experiment 2: Testing the neutral and smiling recognition modules separately

In the first two sub experiments, probe faces are directly fed to the neutral face recognition module. In the third sub experiment leave-one-out cross validation is used to verify the performance of the smiling face recognition module. In each cycle, 29 subjects' faces from both acquisition sessions are used for the training and the remaining subject's smiling face from the second session is used as testing face.

2.1 Neutral face recognition: probe set 1 is used. (neutral face recognition module is used)
2.2 Neutral face recognition: probe set 2 is used. (neutral face recognition module is used)
2.3 Smiling face recognition: probe set 2 is used. (smiling face recognition module is used)

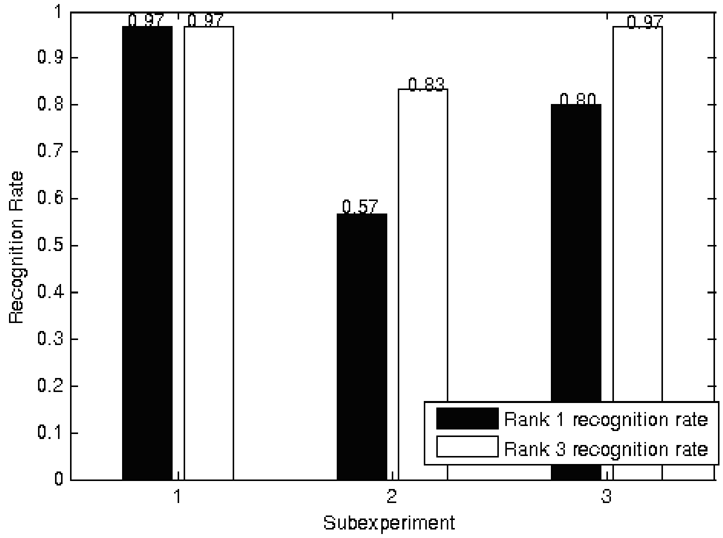


Fig. 7. Results of Experiment 2

From Figure 7, it can be seen that when the incoming faces are all neutral, the algorithm which treats all the faces as neutral achieves a very high recognition rate. On the other hand, if the incoming faces are smiling faces, then the neutral face recognition algorithm does not perform well, only 57% rank one recognition rate is obtained. In contrast, when the smiling face recognition algorithm is used to deal with smiling faces, the recognition rate can go back as high as 80%.

Experiment 3: Testing a realistic scenario

This experiment emulates a realistic situation in which a mixture of neutral and smiling faces (probe set 3) must be recognized. Sub experiment 1 investigates the performance obtained if the expression recognition front end is bypassed, and the recognition of all the probe faces is attempted with the neutral face recognition module alone. The last two sub experiments implement the full framework shown in Figure 1. (Faces are first sorted according to expression and then routed to the appropriate recognition module.) In 3.2 the expression recognition is performed with the linear discrimant classifier, while in 3.3 it is implemented through the support vector machine approach.

3.1 Neutral face recognition: probe 3 is used. (Probe 3 is treated as neutral faces.)
3.2 Integrated expression and face recognition: probe 3 is used. (Linear discriminate classifier for expression recognition.)
3.3 Integrated expression and face recognition: probe 3 is used. (Support vector machine for expression recognition.)

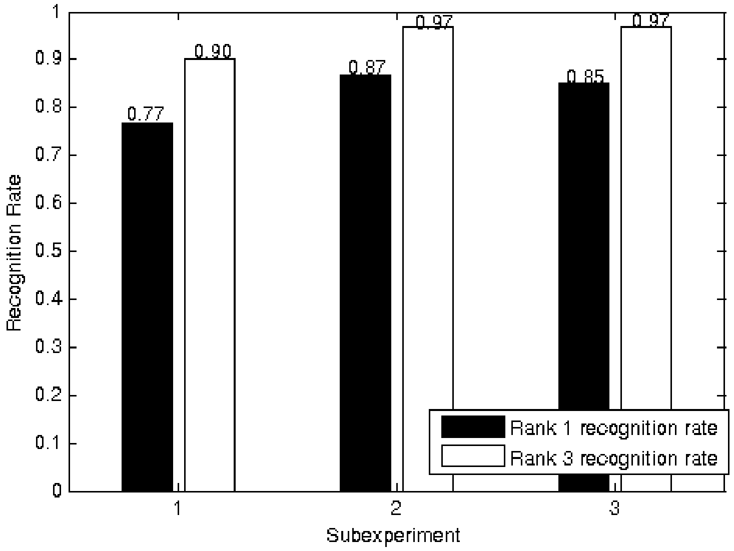


Fig. 8. Results of Experiment 3

It can been seen in Figure 8, that if the incoming faces include both neutral faces and smiling faces the recognition rate can be improved about 10 percent by using the integrated framework proposed here.

7. Discussion and Conclusion

7.1 Discussion

Experiment 1 was aimed at determining the level of performance of the Facial Expression Recognition Module, by itself. Using the leave-one-out cross validation approach, 30 different tests were carried out (Each using 29 x 2 neutral faces and 29 x 2 smiling faces for training). The average success rate in identifying the expressions of the face belonging to the subject not used for training, in each case, was 90.8% with LDA and 92.5% when SVM was used. This confirms the capability of this module to successfully sort these two types of faces (neutral vs. smiling). Both algorithms were applied on the six facial features obtained from the range images (*mw*, *md*, *lc*, *ag*, *hr* and *hm*). Using these features, the actual choice of algorithm used to separate neutral from smiling faces did not seem to be critical.

Experiment two was carried out to test one of the basic assumptions behind the framework proposed (Figure 1). That is, a system meant to recognize neutral faces may be successful with faces that are indeed neutral, but may have much less success when dealing with faces displaying an expression, e.g., smiling faces. This differentiation was confirmed by the high rank-one recognition (97%) achieved by the Neutral Face Recognition Module for neutral faces (probe set 1) in subexperiment 1, which was in strong contrast with the much lower rank-one recognition rate (57%) achieved by this same module for smiling faces (probe set 2), in subexperiment 2. On the other hand, in the third subexperiment we confirmed that a module that has been specifically developed for the identification of individuals from smiling probe images (probe set 2) is clearly more successful in this task (80% rank-one recognition).

Finally, Experiment 3 was meant to simulate a more practical scenario, in which the generation of probe images does not control the expression of the subject. Therefore for all three subexperiments in Experiment 3 we used the comprehensive probe set 3, including one neutral range image and one smiling range image from each of the subjects. In the first subexperiment we observe the kind of results that could be expected when these 60 probe images are processed by a “standard” Neutral Face Recognition Module alone, which is similar to several of the contemporary approaches used for 3D face recognition. Unfortunately, with a mix of neutral and smiling faces this simple system only achieves a 77% rank-one face recognition (much lower than the 97% obtained for probe set 1, made up of just neutral faces, in Experiment 2). This result highlights the need to account for the possibility of a non-neutral expression in 3D face recognition systems. On the other hand, in sub experiments two and three we apply the same mixed set of images (Probe set 3) through the complete process depicted in our proposed framework (Figure 1). That is, every incoming image is first sorted by the Facial Expression Recognition Module and accordingly routed to either the Neutral Face Recognition Module or the Smiling Face Recognition Module, where the identity of the subject is estimated. The right-most four columns in Figure 8 show that, whether using the linear discriminant analyzer or the support vector machine for the initial expression sorting, the rank-one face recognition levels achieved by the overall system are higher (87%, 85%).

In reviewing the results of these experiments, it should be noted that all the experiments involving smiling faces are done using the leave-one-out cross validation method because of the size of the database. Therefore the results displayed are the average, not the best one. For simplicity of implementation, the training samples for the expression recognition system and the smiling face recognition systems are the same faces. In a real application, we would select the training samples to make the best classifier for expression recognition and the identification of faces with a type of expression separately. Considerable performance improvement might be achieved in this way.

7.2 Conclusion

In this paper we have presented an alternative framework proposed to enhance the performance of 3D face recognition algorithms, by acknowledging the fact that the face of a subject is a deformable surface that undergoes significant changes when the subject displays common expressions. Our main proposition is that, instead of ignoring the possibility of significant facial changes due to expressions, 3D face recognition systems should account for the potential presence of an expression in their probe images. In particular our suggested framework requires the development of two new functional modules, in addition to a "standard" face recognition module for neutral faces:

- A Facial Expression Recognition Module, capable to "tag" an incoming probe image with an appropriate expression label, and route it to an appropriate "specialized" face recognition classifier (matched to the expression found in the probe face), where the identity of the subject will be estimated.
- "Specialized" Face Recognition Classifiers that are trained to identify faces with expressions other than "neutral"

In this work we have developed the framework for the simplest case in which we consider only neutral and "smiling" faces, as one very common form of expression, frequently displayed by people in public.

Our experimentation with this implementation of the framework, using 3 sets of probe images has revealed that:

- It is possible to implement an appropriate module for the sorting of neutral vs. smiling 3D face images, based on classification of six facial features we have defined, and utilizing either the linear discriminant analysis or the support vector machine approaches (Experiment 1).
- While a contemporary "neutral" face classifier is capable of achieving a good performance, (97% rank-one recognition), when identifying neutral 3D faces, the performance of this same classifier is much weaker (57% rank-one recognition) when dealing with otherwise comparable "smiling" faces. (Experiment 2).
- It is feasible to develop a "specialized" classifier that will identify "smiling" faces with a reasonable level of success (80% rank-one recognition), which is clearly higher than the performance of the "neutral" face classifier for the same scenario (Experiment 2).
- A system that follows the complete framework proposed (Figure 1) is better able to identify subjects from a mixture of neutral and smiling 3D faces (87% and 85% rank-one recognition) than a standard 3D face recognition system (77% one-rank

recognition) that relies on the assumption that the subjects are expressionless during the capture of the probe images (Experiment 3).

The work reported in this paper represents an attempt to acknowledge and account for the presence of expression on 3D face images, towards their improved identification. In comparison with other methods that pursue similar goals [1, 19, 32], the method introduced here is computationally efficient. Furthermore, this method also yields as a secondary result the information of the expression found in the faces.

Based on these findings we believe that the impact of expression on 3D face recognition and the development of systems that account for it, such as the framework introduced here, will be keys to future enhancements in the field of 3D Automatic Face Recognition.

8. References)

Bartlett, M., Lades, H., Sejnowski, T., 1998. Independent Component Representations for Face Recognition. Proceedings of SPIE.

Bolle, R., Connell, J., Pankanti, S., Ratha, N., Senior, A., 2003. Guide to Biometrics, Springer.

Bowyer, K., Chang, K., Flynn, P., 2004._A Survey of Approaches to 3D and Multi-Modal 3D+2D Face Recognition. Proceedings of IEEE International Conference on Pattern Recognition. pp. 358-361.

Bronstein, A., Bronstein, M., Kimmel, R., 2003. Expression-invariant 3D face recognition. Proceedings of Audio & Video-based Biometric Person Authentication (AVBPA), pp. 62-69.

Bruce, V., Young, A., 1986. Understanding face recognition. British Journal of Psychology **77**: 305-327.

Cartoux, J., LaPreste, J., Richetin, M.,1989. Face authentication or recognition by profile extraction from range images. Proceedings of the Workshop on Interpretation of 3D Scenes. pp.194-199.

Chang, C, Lin, C., 2001. LIBSVM: a library for support vector machines.

Chang, K., Bowyer, K., Flynn, P., 2003. Multimodal 2D and 3D biometrics for face recognition., 2003. IEEE International Workshop on Analysis and Modeling of Faces and Gestures. pp.187-194.

Chang, K., Bowyer, K., Flynn, P., 2005. Effects of facial expression in 3D face recognition. SPIE 5779: Biometric Technology for Human Identification II.

Chellappa, R., Wilson, C., Sirohey, S., 1995. Human and Machine Recognition of Faces: A Survey. Proceedings of the IEEE **83(5)**: 705-740.

Cholkopf, B., Smola, A., Muller, K., 1998. Nonlinear Component Analysis as a Kernel Eigenvalue Problem. Neural Computation **10**(5): 1299-1319.

Duda, R., Hart, P., Stork D., 2001 Pattern Classification.

Edwards, G., Cootes, T., Taylor, C., 1998. Face Recognition Using Active Appearance Models. Proceedings of ECCV, pp. 581-695.

Ekman, P., Friesen, W., 1971. Constants across cultures in the face and emotion. Journal of Personality and Social Psychology **17**(2): 124-129.

Etcoff, N., Magee, J., 1992. Categorical perception of facial expressions. Cognition **44**: 227-240.

Fasel, B., Luettin, J., 2003. Automatic facial expression analysis: a survey. Pattern Recognition **36**: 259-275.

Gordon, G., 1991. Face recognition based on depth maps and surface curvature. Geometric Methods in Computer Vision, SPIE 1570: pp. 1-12.

Hansch, E., Pirozzola, F., 1980. Task Relevant Effects on the Assessment of Cerebral Specialization for Facial Emotions. Brain and Language **10**: 51-59.

Hay, D., Young A., Ellis, A.,1991. Routes through the face recognition system. Q. J. Exp. Psychol A-Human Exp. Psy. **43**: 761-791.

Hesher, C., Srivastava, A., Erlebacher G., 2003. A novel technique for face recognition using range images. Seventh International Symposium on Signal Processing and Its Application.

Li, C., Barreto, A., 2005. Profile-Based 3D Face Registration and Recognition. Lecture Notes on Computer Science **3506**: 484-494.

Li, C., Barreto, A., Zhai, J., Chin C., 2005. Exploring Face Recognition Using 3D Profiles and Contours. Proceedings of IEEE SoutheastCon 2005. pp.576-579.

Lu, X., Jain, A., 2005. Deformation Analysis for 3D Face Matching. 7th IEEE Workshop on Applications of Computer Vision (WACV'05).

Moghaddam, B., Pentland, A., 1995. Probabilistic Visual Learning for Object Detection. Proceedings of International Conferrence of Computer Vision (ICCV' 95), pp.786-793.

Moghaddam, B., Pentland, A., 1997. Probabilistic Visual Learning for Object Representation. IEEE Trans. on Pattern Analysis and Machine Intelligence **19**(7): pp. 696-710.

Nagamine, T., Uemura, T., Masuda, I., 1992. 3D facial image analysis for human identification. Internaitonal Conference on Pattern Recognition (ICPR,1992).pp. 324-327.

Ortega-Garcia, J., Bigun J., Reynolds, D., Gonzales-Rodriguez, J., 2004. Authentication gets personal with biometrics. IEEE Signal Processing **21**(No.2): 50-61.

Phillips, J., Grother, P., Blackburn, D., Bone, M., Michaels, R., 2002. Face Recognition Vendor Test 2002.

Turk, M., Pentland, A., 1991. Eigenfaces for Recognition. Jounal of Cognitive Neuroscience **3**(1): 71-86.

Wiskott, L., Fellous, J., Krger, N., Von der Malsburg, C., 1997. Face Recognition by Elastic Bunch Graph Matching. IEEE Trans. on Pattern Analysis and Machine Intelligence **19**(7): 775-779.

Zhao, W., Chellappa, R., Rosenfeld, A., 2003. Face recognition: a literature survey. ACM Computing Survey **35**: 399-458.

17

Template Protection For 3D Face Recognition

Xuebing Zhou, Arjan Kuijper and Christoph Busch
Fraunhofer Institute for Computer Graphics Research IGD
Germany

Abstract

The human face is one of the most important biometric modalities for automatic authentication. Three-dimensional face recognition exploits facial surface information. In comparison to illumination based 2D face recognition, it has good robustness and high fake resistance, so that it can be used in high security areas. Nevertheless, as in other common biometric systems, potential risks of identity theft, cross matching and exposure of privacy information threaten the security of the authentication system as well as the user's privacy. As a crucial supplementary of biometrics, the template protection technique can prevent security leakages and protect privacy.

In this chapter, we show security leakages in common biometric systems and give a detailed introduction on template protection techniques. Then the latest results of template protection techniques in 3D face recognition systems are presented. The recognition performances as well as the security gains are analyzed.

1. Introduction

Biometrics is technique to automatically recognize a person based on his/her physiological or behavior characteristics. Since the characteristics used are unique to each individual, biometrics can create a direct link between users and their identity. From this point of view, it can provide more secure authentication in comparison to password and token based methods. Moreover, it is very convenient to use.

Applications of biometrics have spread rapidly in last decade and are still growing. In European e-passports, images of faces and fingerprints are stored. Many countries employ biometric information in citizen cards. In the US visit program, 10 fingers and the facial image are also acquired to support visa application and border control. It is efficient to prevent identity fake and increase the security against terrorists. Additionally, biometrics are widely used in access control, payment, banking and so on.

As biometrics markets are blooming, novel security and privacy leakages, such as exposure of private sensitive information, unchangeability and impersonate of biometric identities, and profiling, attract a lot of attention from public sectors, data protection officers, service providers and end users. After carefully analyzing the weakness of biometrics and summarizing requirements for secure biometric authentication, template protection techniques have been developed as an important supplement to common biometric system with improved security and enhanced privacy protection.

In this chapter we address the template protection techniques for 3D face recognition systems. Among fingerprints and iris, face is one of the most popular biometric modalities. So far, facial information is mainly acquired as 2D illumination based images, 3D surfaces, or 2D near infra-red images. In comparison with other methods, 3D face recognition utilizes rich geometric information of facial surfaces, is less sensitive to ambient light conditions and robust to face variation due to different poses. Especially due to its resistance to fake attack it is very attractive in high security applications. Recently, 3D scanners become more and more efficient and economically priced. A sufficiently precise 3D face scan can be accomplished in several milliseconds. 3D face recognition shows a better recognition performance than 2D face recognition and other kinds of biometrics modalities. By implementing template protection techniques for 3D facial information not only the 3D facial information itself is protected. Also potential risks are avoided and high secure authentication systems become realizable.

2. Motivation

As biometrics is applied in manifold areas and users enjoy its benefits, vulnerabilities of biometric system and potential security risks cannot be neglected or underestimated as shown in following example.

Bob is a frequent traveller. He does not like waiting in long queues including the queue in front of a border control desk. Thus, he registers at the airport for the automatic verification system that is based on 3D face recognition. When crossing the border, the system reads his travel document, performs a 3D scan of his face and compares it with the stored information. This saves the busy businessmen Bob a lot of time. He gets enthusiastic about biometrics and enrols himself in 3D face recognition in the bank for cash points. Later he visits a casino and leaves his 3D face information to access the casino.

In this example, he used the same biometric information in different areas. Some of them are trustworthy. But can he really trust all the different service providers? What will happen, if his biometric information is compromised? In the following we elaborate the privacy and security issues in biometrics.

Aggravation of identity theft/fraud Biometric characteristics can not be stolen or handed over like token or password, but they can be faked. For example, it is shown in (2; 20) how effortless it is to make a gummy or laminate finger using a trace left on a glass. With one more surveillance camera, facial information can be completely exposed even without knowledge or consent of victims. Also, a synthetic artifact can be created with stored biometric templates (4; 10; 15). Remote authentication systems based on digital transmitted biometric data can be even attacked without reconstructing biometric modalities. As a consequence, on the one hand integrating liveness detection techniques in sensors is essential to prevent impersonation. On the other hand, protection of stored and transmitted biometric data is significant.

3D face recognition has advantages in comparison with other modalities from a security point of view. It is hard to obtain 3D face information, since it can not leave a trace like fingerprints. Besides the up-to-now minor risk of deriving a 3D face surface from photos, there is the risk to reconstruct a 3D face surface from stored biometric templates. [1]

Unchangeability On the one hand, one of the advantages using biometrics is that users and their identity are linked together with their personal unique biometric characteristics. On the other hand, the exposure of biometric data is critical since it can not be easily revoked or renewed as in common password or toked based authentication. One

[1] It is difficult to keep some biometric modalities such as fingerprint, 2D face images, ''secret". In applications requiring high security, using these modalities should be avoided.

can only choose another biometric modality or try to modify the exposed one. Unfortunately, both are not suitable solutions: we only have a limited number of biometric modalities, e.g. ten fingers, one face and two irises. And alteration of our biometric modalities is only possible with very complicated methods such as transplantation or cosmetic surgery. In the report of defense science board (3) , it is also emphasized that revocation of biometric keys used for identity or privilege is indispensable.

Cross matching As the same biometric modality is adopted in multiple applications, all these applications are potentially coupled. A untrustworthy data collector can track the activities of a subject in external applications and misuse these informations. Additionally, if the biometric identity is compromised in one application, all the others get in danger.

Privacy Biometric data is derived from human bodies or the behavior of a person. Per-se this personal information is sensitive information. For example, in (30) it is shown that some diseases and sexual orientation have influence on fingerprint. The disease such as free-floating iris cyst, diffuse Iris Melanoma, can change iris pattern. From a face photo, gender and race of users can be recognized. DNA-analysis can expose sensitive genetic information. 3D face information is widely used in medical analysis. Moreover, many genetic syndromes can effect the facial trait. Therefore, 3D face scans is used to analyze facial morphology (13) as well as to diagnose some gene syndromes (29). Such private information is not relevant for the authentication purpose. But it is contained in every 3D face scan and therefore retrievable from the scan results.

Legislation Since biometrics belongs to personal information, usage of biometrics including storage and collection is very critical from the legislation point of view. In the European data protection directive (1), it is emphasized to protect individuals regarding to the processing of their personal data. In the white paper of TeleTrustT Deutschland e.V. (7), it is elaborated the importance of data protection in biometrics in compliance with the legislation.

Hill climbing Decision of biometric authentication rests on the similarity (or distance) measurement between stored templates and a live derived template. A feedback of a comparison can be obtained, e.g. directly from the authentication response, or from a Trojan horse embedded in a computer. An attacker can exploit such information to reconstruct the stored biometric template or the biometric sample recursively (33). It is so called hill climbing attack

The above issued security and privacy problems arise from the uncertainty of stored or transmitted biometric data. The Information and Privacy Commissioner/Ontario also dwelt on the problems of biometrics and drew the conclusions that applying biometrics causes ''a zero-sum game'' for this reason: biometrics provide additional security guarantees, meanwhile, it brings also new leakages. Therefore techniques to protect biometric data is necessary(11). In the next session we introduce the possible solutions to overcome these drawbacks.

3. Template Protection Techniques

Recently template protection techniques – also known as biometric encryption, untraceable biometrics, cancelable or revocable biometrics – have been developed in order to meet the requirements of protecting stored biometric data. These methods convert biometric data elements into multiple (ideally) uncorrelated references, from which it is infeasible to retrieve the original information. Template protection is a generalized and efficient method to preserve privacy and to enhance the security of biometric data by limiting the exposure of template data which must not be revoked. They have the following key properties:

One-Way and Robustness The computational complexity of deriving a secure reference from a biometric datum (template) is limited, while it is either computationally hard or impossible to deduce the template from such a reference. The derivative references can be compared to a biometric datum under similarity metrics for the underlying biometric template. This allows the successful comparison of measurements exhibiting small variations or measurement errors to a derivative reference.

Diversity and Randomness Template protection can create numerous secure references from one biometric feature with the references independent of each other, i.e. knowledge of one reference does not yield information on other references derived from the same template.

The resulting various references are also called pseudo identities (9). Different methods to protect the biometrics data exist. They can be classified into four categories: cancelable biometrics, biometric salting, fuzzy encryption and biometric hardening passwords (36). *Cancelable biometrics* modifies the original biometric features or samples with "non-invertible" functions such as scrambling, morphing so that the original data is no longer to be recognized (26), (8), (27). *Biometric salting* randomizes biometric data with random patterns. For example, in biometric encryption (28) , biometric data is convoluted with randomly generated code and in the biohashing algorithm (16), biometric features are projected into different orthogonal spaces generated from large amount random sequences. *Biometric hardening passwords* fuses password-based authentication with biometrics (22), (21). Finally, *fuzzy encryption* combines cryptographic hashing or secret sharing protocol with error correction codes (ECC) (17; 18). Among these methods, fuzzy commitment is one of the most successful algorithm. In the next section, we give detailed introduction on the helper data scheme, a practical realization of fuzzy commitment.

4. Template Protection with Helper Data Scheme

In 1999, Juels et al. proposed the "fuzzy commitment" to protect biometric information by using an existing cryptographic scheme (18). It is similar to password authentication system on Unix computers. There, passwords are never stored or compared in a plain text form but in an encrypted form. The system can authenticate a user without knowing the original password. Accordingly, a fuzzy template can be generated from biometric information and can be safeguarded with cryptographic functions. Yet, the protection scheme must have tolerance to variation of biometric data due to measurement noise or alteration of modalities. To overcome this problem, error correction coding is used.

Later, the Helper Data Scheme (HDS) has been developed to make the idea of fuzzy commit feasible for biometric systems. HDS can extract secure templates from biometric data. This secure template is stable to biometric variation and it is impossible to retrieve original biometric information from it. The mathematical formulation of these properties is summarized as delta-contracting and epsilon-revealing by J. P. Linnartz et al. (19). The block diagram of the HDS is depicted in figure 1.

In figure 1 M is a biometric template extracted from a biometric measurement. In the enrollment process, the binarization converts the biometric template M into a binary vector Q. Ideally, the binarization results in a binary string that is uniformly distributed for different users and invariant for an identical user. The detailed description of binarization is given

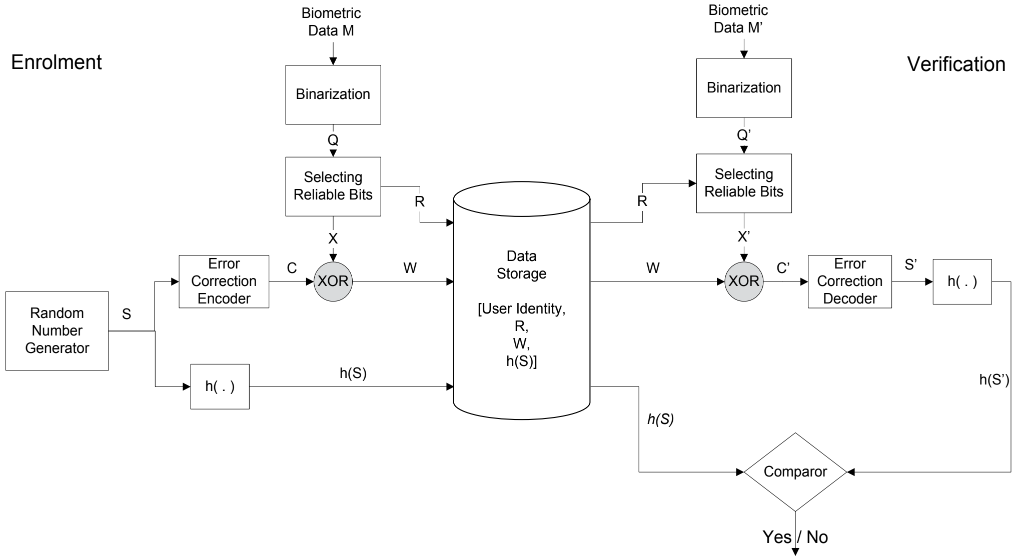


Fig. 1. Block diagram of the helper data scheme

in section 4.1. In parallel, a random number generator creates a secret code S. First, S is hashed and stored. Thus, it enables randomness in the system so that distinct references can be created from the same biometric characteristics for different applications. Second, a error correction encoder adds redundancy in the secret S. As a consequence, the resulting codeword C is longer than S. Depending on the characteristics of the bit errors, different error correction codes can be adopted. Foe example, when bit errors are uniformly distributed in the codeword, a BCH-code, which has a codeword length of $2^L - 1$[2], can be employed. If the length of the binarized vector Q extends the length of the codeword C, then the most reliable bits in Q are selected so that the resulting binary string X is as long as the codeword C and robustness is improved. R indicates the position of reliable bits. W, the result of the bitwise XOR-function of X and C, is so called helper data. With help of W and for a suitable input, the secret S can be recovered in the verification process. Instead of storing the secret S, the position vector R, the helper data W, the hashed secret code $h(S)$ and user identity information are stored in data storage. It can be proved that both W and $h(S)$ reveal little information about S, X as well as the biometric template M (32).
During the verification process, with claimed identity, R, W and $h(S)$ are retrieved from the data storage. The binary string Q' is extracted from biometric template M', which is M potentially distorted by noise. The binary string X' is estimated with Q' and R. A potentially distorted codeword C' can be acquired from W and X'. The following error correction decoder removes errors in C' and results in the reconstructed secret code S'. By comparing $h(S)$ with $h(S')$, a positive or negative response for a verification query can be given. In contrast to common biometrics system, only a "hard decision" (rejected or accepted) is given and no similarity score is available in the comparator of the template protection system due to the applied hash function. The previously described hill climbing attack, which iteratively reconstructs biometrics using matching scores (5)' (31), is not applicable.

[2] L is a natural number

The length of the secret code is one security issue. If the length of the codeword is fixed, the length of the secret code is restricted by the error correction ability. The maximum length of the codeword relies on the entropy for the considered biometric characteristics. Obviously, the processes of binarization and selection of reliable bits strongly affects the performance of the template protection scheme. In the following sections we introduce their functionalities and construction.

4.1 Binarization

Binarization is the core component of the helper data scheme. The requirements of its output binary vector can be summarized as follows: binarized vectors of different users should be uniformly and independently distributed, and the binary vector of a specific user should be robust to variation of biometric data. It guarantees that no prediction of a binary vector is possible and the discriminability of binary vector is optimized. And no information of a user can be obtained using binary vectors of other users. The binarized features have certain resilience to noise.

Moreover, binarization tries to extract a long binary vector from biometric template without any degradation of authentication performance. The construction of binarization depends on the statistical analysis of the input biometric templates. Assuming that a training template set contains N users and each user has K samples and $M_{n,k} = \left[m_{n,k,1}, m_{n,k,2}, \cdots, m_{n,k,T}\right]$ is the template with T components extracted from the k-th samples of the user n with $k \in \{1, \cdots, K\}$ and $n \in \{1, \cdots, N\}$. If each component is statistically independent and at least one bit can be extracted from each component, the binarization function can be defined as:

$$q_{n,t} = B\left\{m_{n,k,t} | k \in [1, \cdots, K]\right\} = \begin{cases} 1 & \text{if} \quad \mu_{n,t} \geq \mu_t \\ 0 & \text{if} \quad \mu_{n,t} < \mu_t \end{cases} \tag{1}$$

where $\mu_{n,t}$ is an estimation of the real template for user n and μ_t is the threshold of binarization. In order to achieve uniform distribution of the binary vector, μ_t could be the median of $\mu_{n,t}$ of all the users. Instead of the median, the mean can also be adopted. If the training data set is large enough, there is no significant difference between median and mean. In practice, we suggest to use the median, which is resistant to extreme values caused by measure errors.

4.2 Selecting Reliable Bits

Selecting reliable bits contributes to the robustness of the system. It is based on the estimation of the error probability for each bit. Only the bits with the lowest error probability are selected. In the previously presented binarization method, the error probability depends on the distance between $\mu_{n,t}$ and μ_t as shown in equation 1. $\mu_{n,t}$ of a relative stable bit should derive from μ_t. On the other hand, intra-class variation also affects the error probability. The smaller the intra-class variation is, the more reliable the corresponding bit is.

Statistical analysis of intra class characteristics for each user has a major effect on the performance of selecting reliable bits. If biometric templates are Gaussian distributed, then:

$$\mu_{n,t} = E\left\{m_{n,k,t} | k \in [1, \cdots, K]\right\} \tag{2}$$

$$p_{n,t} \propto \frac{|\mu_{n,t} - \mu_t|}{\sigma_{n,t}} \tag{3}$$

where E is the function calculating the expected value, $p_{n,t}$ is the error probability of the t-th component of user n, and $\sigma_{n,t}$ is the standard deviation of $m_{n,k,t}$ for $k \in [1, \cdots, K]$ (see also (34)).

If biometric templates are not Gaussian distributed or it is impossible to estimate intra class variation, then:

$$\mu_{n,t} = MEDIAN_{k=1}^{K}\left\{m_{n,k,t}\right\} \quad (4)$$

$$p_{n,t} \propto \left|\mu_{n,t} - \mu_t\right| \quad (5)$$

Actually, the reliable estimation of error probabilities can only be achieved with a sufficient number of samples. In the next section we show how the template protection using the helper data scheme is integrated in 3D face recognition system.

5. An Example of Secure 3D Face Recognition System

The above sections stressed the importance of protecting biometric data and introduced to the details of template protection systems. In this section, we show how such a method can be integrated in a 3D face recognition system. Our experimental results show the effect on the performance. At first, we will give an introduction on 3D face recognition algorithm.

5.1 3D Face Recognition Algorithm

In a 3D face recognition system, a 3D face image can be acquired by using a structured light projection approach. To compensate pose variation during acquisition, the 3D face images are normalized in a pre-processing step to a frontal view. The normalized facial image represents the face geometry and can be used as a biometric feature. For example, the normalized images can be compared using the Hausdorff distance classifier ((24), (23)). This normalized data, however, cannot directly be utilized in the template protection, since these features are strongly correlated and very sensitive to noise. A process to extract compact and robust features is required. The Eigenface and Fisherface feature extraction algorithms (e.g. (12), (14) and (6)) are widely used to reduce dimensions of the original data. These statistics-based algorithms achieve a good verification performance, however, the size of the features is strongly reduced and it is difficult to extract binary vectors of sufficient length, which is required for an input for the helper data scheme.

In our experiments, we use a histogram-based feature extraction algorithm. It is based on the distribution of depth-values of the face region to characterize facial geometry. In this algorithm, a three dimensional rectangular region of a normalized image is used to limit the considered facial surface points. In Figure 2, the intermediate processes of the proposed algorithm is depicted. The algorithm consists of the following processing steps:

1. The facial surface points to be evaluated are selected from a normalized range image as shown in the dark area of the image at the lower left of Figure 2.
2. The selected facial region is further divided into J disjunct horizontal stripes S_j, where $j \in [1, \cdots, J]$ (see the image at the lower right of Figure 2). By this, the algorithm evaluates local geometric surface information. Due to the symmetric properties of a human face, the stripes are perpendicular to the symmetry plane.
3. The distribution of the facial points p_i in stripe S_j is counted. If $\{d_0, \cdots, d_L\}$ is a vector with $L+1$ elements. This vector is used to partition the surface points in the horizontal stripes according to their depth-value. d_0 and d_L indicate the upper band and lower band of depth limit, the l-th feature of the stripe S_j is given as follows:

$$f_{l,j} = \frac{\left|\left\{p_i = (x_i, y_i, z_i) | p_i \in S_j, d_{l-1} < z_i < d_l\right\}\right|}{\left|S_j\right|}, \quad (6)$$

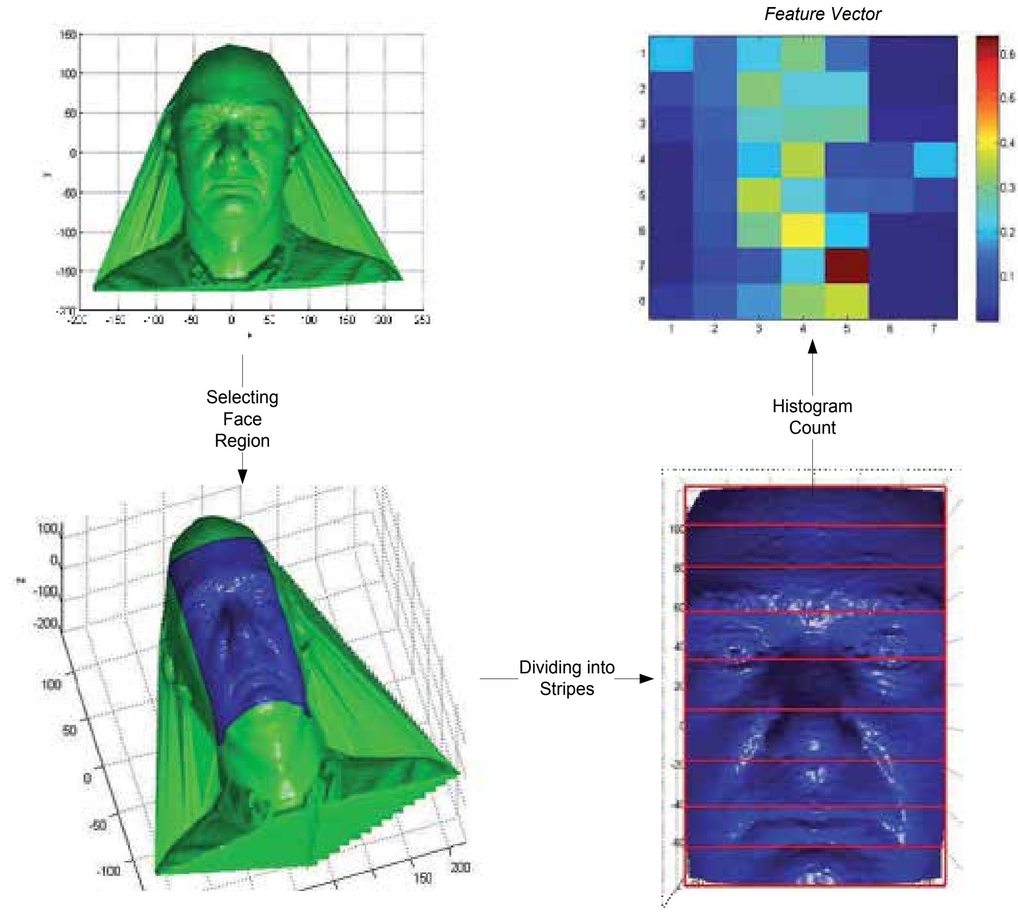


Fig. 2. An overview of the histogram-based face recognition algorithm

where $l \in [1, \cdots, L]$, $j \in [1, \cdots, J]$, z_i is the depth value (z-value) of point p_i, $\left|S_j\right|$ is the number of the facial points in S_j. $f_{l,j}$ represents the proportions of the points in S_j, whose z-values are located in the region $[d_{l-1}, d_l]$.

The resulting feature corresponds to the histogram count of the stripe. Therefore the proposed algorithm is called histogram-based face recognition algorithm. An example of feature values is shown in the image at the top right of Figure 2, where the feature vector corresponding to each stripe is represented as a row in the image and the color indicates their absolute feature values.

The algorithm adopts a simple statistical analysis to describe the geometrical character of a facial surface. This algorithm efficiently filters noise and reduces the correlation in the range image. The resulting feature vectors can be used as an input to the quantizer preceding the template protection scheme. More details about this algorithm are shown in (35).

5.2 Experimental Results

We have implemented HDS in the 3D face recognition system. The 3D facial images of face recognition grant challenge (FRGC) (25) database version 2 are used as testing data. The 3507 samples from 454 subjects are correctly normalized. During the test, only the users, who have at least 4 samples, are chosen. Three samples per user are chosen as enrollment data and one sample as verification data. A different sample for the verification is chosen for each test and the tests are repeated 4 times.

The preciously described feature extraction process is applied with the following parameters. The 3D facial data is normalized to compensate the pose variation in the acquisition. The normalized 3D facial data is projected into regular grids. Then a fixed face region is selected for each resulting range image. The selected face region is divided into 68 sub-areas. A histogram-based extraction algorithm is applied in each sub-area. A feature vector containing $68 \times 6 = 408$ real values is obtained. The false acceptance rate (FAR) and the false rejection rate (FRR) using the correlation classifier is plotted in figure 3. The equal error rate (EER) is equal to 3.38%.

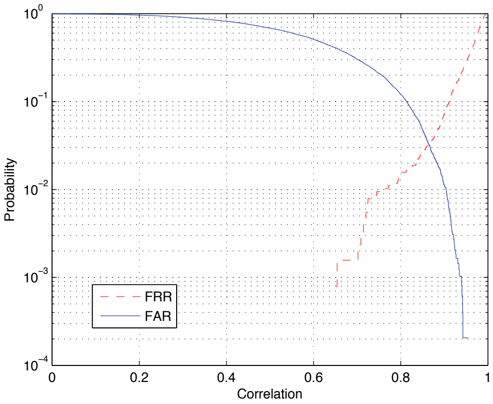


Fig. 3. Classification results of the histogram-based face recognition algorithm

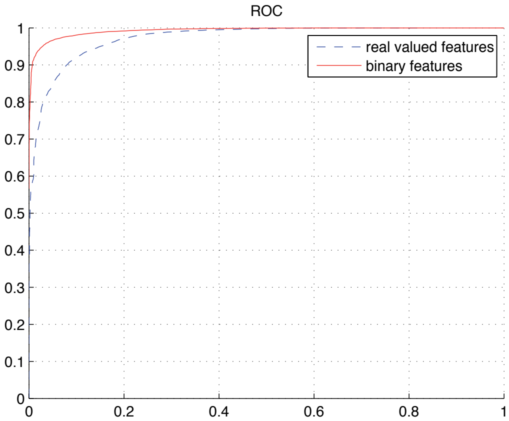


Fig. 4. ROC curves of real-valued feature vectors and binary feature vectors

Then, we use the above mentioned binarization function to convert the extracted feature vectors into binary strings. To compare the authentication performance before and after binarization, we show the receiver operation characteristic (ROC) curves in figure 4. The solid line of the binary feature vectors is obviously above the dashed line of the real-valued feature vector. That is to say, binarization function improves slightly the authentication performance. Generally, a good binarization can be applied with acceptable changing on the authentication performance.

If we compare the distribution of interclass and intraclass distance before and after binarization process as shown in figure 5 and figure 6, the binarization has much stronger influence on the distribution of the inter-class distance than on the intra-class distance. After binarization, the inter-class distance becomes more symmetrical and concentrates on 50%.

In the above binarization process, the median was adopted to calculate the binarization threshold. If we compare the FAR and FRR curves of the binarization using median (figure 7) and mean (figure 8), there is no significant difference regarding authentication performance. Both

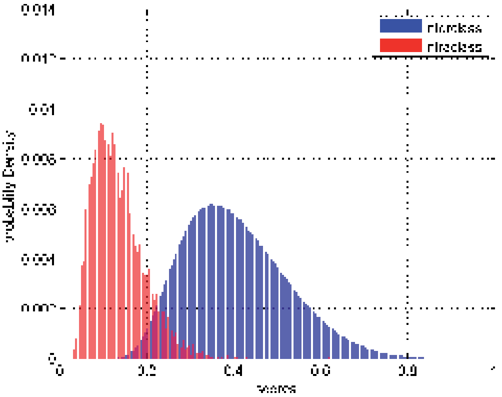


Fig. 5. Probability density of interclass and intraclass distance of real valued features

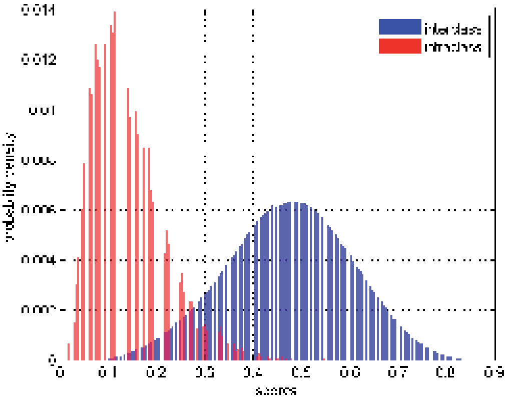


Fig. 6. Probability density of interclass and intraclass distance of binary features

EERs are around 3%. However, the FRR-curve of the mean-based binary vectors deviates from the probability-axis in comparison with the one of the median-based binary vectors. The median-based binarization has higher robustness to noise. This is an advantage over the mean-based binarization, since the performance of template protection is restricted by errors occurring in the binary feature vectors.

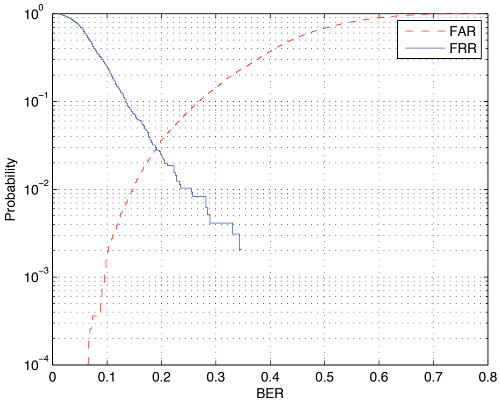


Fig. 7. The classification results for the binary vectors using the median-based binarization

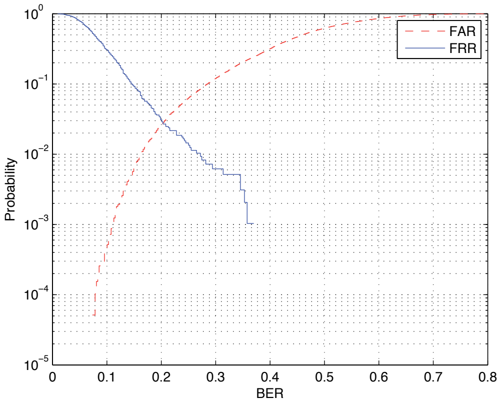


Fig. 8. The classification results for the binary vectors using the mean-based binarization

In the implemented scheme, a BCH- code is chosen as error correction code. The binary features have the length of 408. The maximum length of a codeword under 408 is 255. The 255 most reliable bits is chosen from the 408-bits long binary vector. The classification results under the assumption of uniquely distributed templates and Gaussian distributed templates are shown in Figure 9 and in Figure 10. Both classification results are similar. Under the assump-

tion of uniquely distributions, the robustness is better than under the assumption of Gaussian distributions, however, the discriminative power is slightly worse.
With codeword of 255 bits, only a discrete set of the secret code length *s* and the correctable errors length *e* is possible. Several examples and their corresponding bit error rate (BER), FRR and FAR are given in Table 1. The FRR under the assumption of a uniquely distribution is significantly better than under the assumption of a Gaussian distribution, while its FAR decreases slightly.

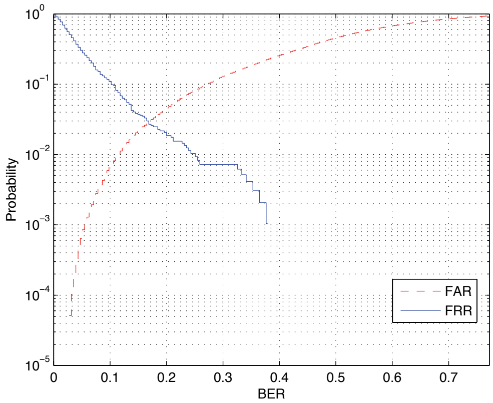


Fig. 9. The classification results for the selected binary vectors under the assumption of uniquely distributed templates

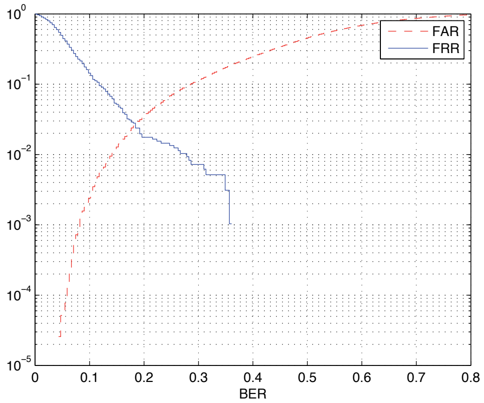


Fig. 10. The classification results for the selected binary vectors under the assumption of Gaussian distributed templates

BCH $(c/s/e)$	Correctable BER	Results for uniquely distribution	Results for Gaussian distribution
255/107/22	8.6%	FRR=12%; FAR=0.4%	FRR=21%; FAR≈ 0
255/91/25	9.8%	FRR=11%; FAR=0.6%	FRR=16%; FAR=0.2%
255/79/27	10.5%	FRR=10%; FAR=0.7%	FRR=13%; FAR=0.3%

Table 1. Examples of possible BCH codes and the corresponding FRR and FAR

6. Discussion

Template protection techniques can derive different secure independent references from biometric data. Secure references can not reveal information of biometric data. No biometric related information is available in authentication systems. Not only biometric data is protected, security leakages such as crossing match, impersonation or cross matching are also avoided. Moreover, they enable revocation and renewing of templates, which are crucial functionalities for authentication process.
Our implementation of a 3D face recognition system shows the general feasibility of a template protection technique. A minor performance degradation is observed in the experimental

results. This might originate in the binarization process and feature selection process: Converting continuously distributed feature vectors into binary vectors can result in information lost.

Although the performance after binarization is improved in our experiments, this can not be generalized for other cases. If feature extraction and the corresponding comparator are optimal, using the biometric features directly has the best performance characteristics. Similarly, the performance degradation for an optimal binarization method is expected to be very small. The selection of reliable components is the requirement of the coding schemes. Additionally, the error probabilities of biometric features normally are not equal. The filtering of unreliable features is helpful to increase the code rate of ECC and secret length. The reduced feature vector, however, may lose discriminative power in comparison to the unabridged feature vector. Although the performance degradation after integrating template protection in the 3D face recognition exist, the resulting performance is still acceptable and comparable with the system without template proction. Moreover, there is potential for improvement: data-source specific adaptation of the coding scheme or the binarization method. Furthermore, the fusion of different modalities or different feature extraction algorithms – in other words deriving more biometric features for one user – can enhance the security and performance.

As the security of template protection schemes is crucial, their evaluation and analysis should not be limited to their performance. In the given experiment, the security evaluation is based on the length of the secret. However, different security levels can be defined depending on the information that is available and accessible to the attacker. If the attacker has only access to individual entries in database, the only way to obtain the secret or the biometric related information is the brute-force attack for the desired hash value. The length of secret is representative to security. However, if they know the details of the template protection algorithm and also distribution of biometric features, the risk of tracking system with much lower complexity is possible. In the second case, the security is over-estimated if security is simply defined by the secret length. Even worse, an attacker with a sufficiently large biometric database can exploit the false acceptance. By doing this, he can identify individual users which share similar biometric data, i.e. biometric twins.

7. Conclusions

In this chapter we show the privacy and security of common biometric systems, which can not be neglected. Template protection techniques are introduced. They can safeguard biometric data and prevent exposing user's private information. They can stop crossing matching, impersonation and hill climbing problems, meanwhile they enable renewing and revocation of identities. They are an very important supplementary to biometric technique.

An implementation in 3D face recognition is demonstrated. 3D face recognition has good resistance to counterfeit and is widely used in high security area. The experimental results show feasibility of template protection technique. The system performance after integration is comparable with the one without template protection. High security can be achieved with sufficient length of the secret. However, it is possible to improve performance and security with optimized binarization and coding methods. Hopeful this work can draw more attention of security enhancement in biometrics and motivate more research in this area.

8. References

[1] Directive 95/46/ec of the european parliament and of the council. *Offical Journal of the European Communities*, L 281 (1995).

[2] How to fake fingerprints? *Chaos Computer Club e.V.* (2004).

[3] Report of the defense science board task force on defense biometrics. Tech. Rep. 20301-3140, Office of the Under Secretary of Defense For Acquisition, Technology, and Logistics, Washington, D.C., March 2007.

[4] ADLER, A. Sample images can be independently restored from face recognition templates. In *Proceedings of Canadian Conference on Electrical and Computer Engineering* (Montreal, Canada, 2003), pp. 1163–1166.

[5] ADLER, A. Reconstruction of source images from quantized biometric match score data. In *In Biometrics Conference, Washington, DC* (September 2004).

[6] BAI, X.-M., YIN, B.-C., AND SUN, Y.-F. Face recognition using extended fisherface with 3d morphable model. In *Proceedings of the Fourth International Conference on Machine Learning and Cybernetics* (2005), pp. 4481–4486.

[7] BIERMANN, H., BROMBA, M., BUSCH, C., HORNUNG, G., MEINTS, M., AND QUIRING-KOCK, G. White paper zum datenschutz in der biometrie. TELETRUST Deutschland e.V., March 2008.

[8] BOLLE, R., CONNELL, J. H., AND RATHA, N. System and method for distorting a biometric for transactions with enhanced security and privacy. US 6836554 B1, Dec 2004.

[9] BREEBAART, J., BUSCH, C., GRAVE, J., AND KINDT, E. A reference architecture for biometric template protection based on pseudo identities. In *BIOSIG 2008: Biometrics and Electronic Signatures* (2008).

[10] BROMBA, M. On the reconstruction of biometric raw data from template data. *Bromba Biometrics* (2006).

[11] CAVOUKIAN, A., AND STOIANOV, A. Biometric encryption: A positive-sum technology that achieves strong authentication, security and privacy. Tech. rep., Information and Privacy Commissioner/Ontario, March 2007.

[12] CHANG, K., BOWYER, K., AND FLYNN., P. Face recognition using 2d and 3d facial data. *In IEEE International Workshop on Analysis and Modeling of Faces and Gestures, Nice, France.* (2003).

[13] HAMMOND, P., HUTTON, T. J., ALLANSON, J. E., CAMPBELL, L. E., HENNEKAM, R. C., HOLDEN, S., MURPHY, K. C., PATTON, M. A., SHAW, A., TEMPLE, K., TROTTER, M., AND WINTER, R. M. 3d analysis of facial morphology. *American Journal of Medical Genetics Part A*, 126(4) (2004), 339–348.

[14] HESELTINE, T., PEARS, N., AND AUSTIN, J. Three-dimensional face recognition: A fishersurface approach. Springer Belin / Heidelberg.

[15] HILL, C. J. Risk of masquerade arising from the storage of biometrics. Master's thesis, The Department of Computer Science, Australian National University, November 2001.

[16] JIN, A. T. B., LING, D. N. C., AND GOH, A. Biohashing: two factor authentication featuring fingerprint data and tokenised random number. *Pattern Recognition Issue 11 37* (November 2004), 2245–2255.

[17] JUELS, A., AND M.SUDAN. A fuzzy vault scheme. In *IEEE International Symposium on Information Theory* (2002).

[18] JUELS, A., AND WATTENBERG, M. A fuzzy commitment scheme. In *6th ACM Conference on Computer and Communications Security* (1999), pp. 28–36.

[19] LINNARTZ, J. P., AND TUYLS, P. New shielding functions to enhance privacy and prevent misuse of biometric templates. In *4th international conference on audio- and video-based biometric person authentication* (2003).

[20] MATSUMOTO, T., MATSUMOTO, H., YAMADA, K., AND HOSHINO, S. Impact of artificial "gummy" fingers on fingerprint systems. In *Optical Security and Counterfeit Deterrence Techniques IV* (2002), vol. SPIE Vol. 4677, pp. 275–289.

[21] MONROSE, F., REITER, M. K., LI, Q., LOPRESTI, D. P., AND SHIH, C. Toward speech-generated cryptographic keys on resource constrained devices. In *in Proc. 11th USENIX Security Symp* (2002), p. 283ñ296.

[22] MONROSE, F., REITER, M. K., AND WETZE, S. Password hardening based on keystroke dynamics. In *International Journal on Information Security, Springer* (2002), vol. 1, pp. 69–83.

[23] PAN, G., WU, Y., WU, Z., AND LIU, W. 3D face recognition by profile and surface matching. In *Proc. International Joint Conference on Neural Networks* (Portland, Oregon, 2003), pp. 2168–2174.

[24] PAN, G., AND WU, Z. Automatic 3d face verification from range data. In *ICASSP* (2003), pp. 193–196.

[25] PHILLIPS, P. J., FLYNN, P. J., SCRUGGS, T., BOWYER, K. W., CHANG, J., HOFFMAN, K., MARQUES, J., MIN, J., AND WOREK, W. Overview of the face recognition grand challenge. In *In IEEE CVPR* (http://face.nist.gov/frgc/, June 2005), vol. 2, pp. 454–461.

[26] RATHA, N., CONNELL, J., AND BOLLE, R. Enhancing security and privacy of biometric-based authentication systems. *IBM Systems Journal 40*, 3 (2001), 614–634.

[27] RATHA, N. K., CHIKKERUR, S., CONNELL, J. H., AND BOLLE, R. M. Generating cancelable fingerprint templates. In *IEEE Transactions on Pattern Analysis and Machine Intelligence* (April 2007), vol. 29.

[28] ROBERGE, C. S. D., STOIANOV, A., GILROY, R., AND KUMAR, B. V. Biometric encryption. *ICSA Guide to Cryptography, Chapter 2* (1999).

[29] SCIENCE, R. M. 3d face scans spot gene syndromes. *BBC News* (September 2007), http://news.bbc.co.uk/2/hi/science/nature/6982030.stm.

[30] SEIDEL, J. Zusatzinformationen in fingerbildern. Master's thesis, Hochschule Darmstadt, 2006.

[31] SOUTAR, C. Biometric system security. In *Information Technology Security Symposium* (2002).

[32] TUYLS, P., AND GOSELING, J. Capacity and examples of template protecting biometric authentication systems. In *Biometric authentication workshop (BioAW 2004)* (Prague, 2004), LNCS, Ed., no. 3087, pp. 158–170.

[33] ULUDAG, U., AND JAIN, A. K. Attacks on biometric systems: a case study in fingerprints. In *SPIE-EI 2004, Security, Seganography and Watermarking of Multimedia Contents VI* (San Jose, CA, 2004), pp. 622–633.

[34] VAN DER VEEN, M., KEVENAAR, T., SCHRIJEN, G.-J., AKKERMANS, T. H., AND ZUO, F. Face biometrics with renewable templates. In *Security, Steganography, and Watermarking of Multimedia Contents VIII* (San Jose, California, USA).

[35] ZHOU, X., SEIBERT, H., BUSCH, C., AND FUNK, W. A 3d face recognition algorithm using histogram-based features. In *Eurographics Workshop on 3D Object Retrieval* (Crete, Greece, 2008), pp. 65–71.

[36] ZHOU, X., WOLTHUSEN, S. D., BUSCH, C., AND KUIJPER, A. A security analysis of biometric template protection schemes. In *International Conference on Image Analysis and Recognition ICIAR 2009* (2009), L. 5627, Ed., pp. 29–38.

18

Geodesic Distances and Hidden Markov Models for the 3D Face Recognition

Giuseppe Mastronardi[1,2], Lucia Cariello[1,2],
Domenico Daleno[1,2] and Marcello Castellano[1,2]
[1]*Department of Electrotecnical and Electronics, Polytechnic of Bari,*
Via Orabona, 4 – 70125 Bari – Italy
[2] *e.B.I.S. s.r.l. (electronic Business In Security), Spin-Off of Polytechnic of Bari,*
Via Pavoncelli, 139 – 70125 Bari – Italy

1. Personal Identification Technologies

The citizen security is a problem that in the recent years, as the daily crime news show, has grown in importance. In particular, the world research in this area, has spent his energy into the producing of a personal identification system as most possible secure. In the last years the Biometrics methods have imposed itself more and more in personal recognition field.
The word "Biometrics" is used to refer to the study of the automatic methods of identification or authorization of people which entail the use of physiologic and behavioural characteristics (Ashbourn, 2000).
The main technologies used for the identification of people and based on biometric recognition (Cascetta & De Luccia, 2004) can be devided in

- personal identification founded on the biometric recognition of "static features": capturing and processing of human anatomic characteristics such as fingerprints, geometry and "footprint" of the hand vascular, facial, iris and retinal geometry;
- personal identification founded on the recognition of "dynamic features": vocal timbre, speech peculiarities (spectral analysis of the sound field), dynamic signature (pressure), digitization (pressure), gait (steps in walking).

There are also other recognition techniques among which DNA matching that are not widely employed due to their complexity and the impossibility to wiork in real time (Jain & Bolle et al., 1999)(Jain & Halici et al., 1999).

1.1 Biometric Identification Systems

The biometrical personal recognition techniques require the use of expert systems, neural networks, fuzzy logic systems and the development of sophisticated computing. These methods offer the main advantage over traditional ones to be able to remember and learn.
For a long time, the aim of international researchers and scientists has been to create machines and systems capable of imitating certain human abilities, among which the identification based on biometric recognition or the identification through the acquisition and subsequent processing of images.

The main areas of interest of biometric technologies are:

- direct authentication and verification of personal identity, proof of the identity stated by the individual;
- indirect identification of a person through the available biometric characteristics.

The main physiologic or behavioural characteristics that may be used for personal identification must meet the following essential requirements:

- universality (each individual must have the characteristics that have been defined);
- uniqueness (it is not possible for two people to be the same in terms of the characteristics);
- permanence/invariance (biometric characteristics must remain unchanged in time or their change must be very slow and independent from changes in weight of the subjects and from the acquisition process);
- collectability (in the sense that biometric characteristic can be measured quantitatively);
- acceptability (the acquisition should not result, in case of invasive methods, not tolerated from the subjects);
- reliability (detection must be reproducible);
- reducibility (the data must be compressed and organized into easily accessible files for following use);

In the follow list are described the main "intelligent" methods of biometric recognition (Prabhakar et al., 2003) (Zhang, 2000):

a. Hand recognition: hand geometry recognition systems measure the physical characteristics (geometry) of an individual's hand that is palm and fingers. The uniqueness and permanence feature are questioned as there may be many causes of insidious instability, and furthermore changes may intervene over time (age, diseases, accidents, etc). This method is acceptable (when excluding the hygiene factors for the hand positioning on the capture plate), requires easy reproducible acquisition methods, does not affect the privacy, but is used only when there are minimal identification requirements;
b. Fingerprint recognition: the fingerprint is represented as a series of 'ridges' and 'valleys' with local discontinuities in the ridge flow pattern called minutiae. The map of ridges and minutiae is unique for each individual and may change only with the destruction of the skin;
c. Vein recognition: the map of veins on the back of the hand varies for each individual, sufficiently to secure recognition from the comparison of the respective maps.The non-invasive method has encouraged the search for the realization of sensors for recognition based on this principle: the map of veins is acquired by an infrared optical systems by temperature sensors;
d. Face recognition: the identification of an individual through the face is based on the extraction of physiognomic parameters and metric measurements. The methodologies differ greatly depending on the type of recognition: they can be fully automatic and delegated to the appropriate software and hardware, or involving the operator for off-line identification;
e. Digitization (pressure): Gaines and his colleagues at the Rand Corporation have demonstrated the potential of dynamic typing in 1980 by measuring the amount of latency in the pressure of two consecutive letters in the text beaten by seven professional secretaries. The applicability of this methodology requires a large set of input data. A

research of 1986 showed that the typing of own name provides reliable performance and easier applicability. Alternative methods include the use of special keyboards that also measure the pressure exerted on the letters when typing;

f. Retina recognition: is based on retinal scanning, a special scanner illuminates the retina through the pupil with an infrared (IR) light and memorises the information from the reflection of vascular contrast. The vascular map is stable and unambiguous. The biggest problem with this methodology relates to the difficulty of acquisition that requires extensive collaboration of the user. Moreover, the costs are high, so the method is confined in applications that require secure identification, such as research facilities, military and detention;

g. Iris recognition: iris contains parameters of high power of discrimination for who does not use drugs, does not wear corrective contact lenses or colored and is not suffering from diseases (such as glaucoma). The acquisition is performed through scanning with high definition laser light. The method has limits of applicability similar to those mentioned for the retina recognition;

h. Vocal timbre: these systems, based on the recognition of vocal timbre, are divided into two classes according to whether approval is given by depending on the text pronounced or independent of it. In the first case the personal recognition is carried out by asking to pronounce a "password" with which the various models are loaded. To avoid the use of recordings that can "circumvent" the system, it is preferred to use more words of which, the person must be recognized, they are asked to say a few. In the case of systems independent from the text, the system does not know in advance what will be pronounced and therefore for the enrollment must be loaded many words. For correct enrollment, therefore, it requires a large quantity of recordings.

1.2 Biometric Recognition Applications

In the traditional personal identification systems, the individual declares his identity through an personal identification code, numeric or alphanumeric code (PIN, login or userID). The weakness of these methods of personal identification is their fraudability due to the fact that the access code or password can be stolen or deducted, and then used fraudulently by others, obviously malicious. In fact, to limit and mitigate the fraudability, and then to improve security at physical access (of people) or in the remote systems (Internet, Intranet, corporate computer networks, etc..) is possible to replace alphanumeric access codes with the biometric features of the subject.

This Section briefly reviews the main biometric recognition applications (Buciu, 2008):

Security in the control of the physical and informatic accesses (banks, courts, judicial police and military facilities, strategic sectors of industry, patent offices, Research & Development, etc.): can be realized accesses with turnstiles or gates (sliding doors automatically opening), by inserting the magnetic badge in appropriate readers and by validating the identity through the extraction of a biometric characteristic (fingerprint, iris, hand geometry).

Accreditation at institutions or services (digital signature and biometric signature). *Counterfeiting of identity documents*: the growing need for security has prompted many states, including Italy, to realize electronic identity documents (identity cards, passports, etc.). The smart cards can store, into the chip, much more information than traditional anagraphic data: the inclusion of fingerprints or images of iris/retina, making it very efficient and secure the recognition of the subject, traditionally given to the photograph that accompanies the anagraphic data.

Face identification was widely used for identifying driver licenses, in immigration programs, passports, or welfare registration.

Access control deals with border – crossing, vehicle access, ATM, computer access, computer network access, online transaction access, online database access.

Security refers to the terrorist alert issue in airports, secure boarding systems, file encryption, intranet and internet security, or medical records. Many airports have been adopted face recognition technology for improving security.

Surveillance is another application area where face recognition plays a major part, including video surveillance, CCTV control or portal control.

Multimedia management deals with face-based searching information, face-based video segmentation and summarization or event detection. Human faces are frequently seen in news, sports, films, home video, and other multimedia content. Indexing this multimedia content by face detection, face tracking, face recognition, and face change detection is important to generate segments of coherent video content for video browsing, skimming, and summarization.

Low enforcement is closely related to suspect tracking and investigation, identifying cheats in casinos, criminal face retrieval and recognition.

Human – computer interaction refers to interactive gaming and proactive computing.

Table 1 lists some of the applications of face recognition that is the scgoal of this chapter (Zhao et al., 2003).

Areas	Specific Applications
Entertainment	Video Game, Virtual Reality, Training Programs
	Human-Robot-Interaction, Human-Computer-Interaction
Smart Cards	Drivers' Licenses, Entitlement Programs
	Immigration, National ID, Passports, Voter Registration
	Welfare Fraud
Information Security	TV Parent Control, Personal Device Logon, Desktop Logon
	Application Security, Database Security, File Encryption
	Intranet Security, Internet Access, Medical Records
	Secure Tracing Terminals
Law Enforcement and Surveillance	Advanced Video Surveillance, CCTV Control
	Portal Control, Post-Event Analysis
	Shoplifting, Suspect Tracking and Investigation

Table 1. Typical applications of face recognition

1.3 Face Recognition

Face recognition is an innate method used by humans to recognize one another. Face recognition techniques have an advantage over the other biometric techniques in that they are non invasive and require little or no cooperation from the (passive) individual subjected to recognition as they are not susceptible to behavioural modifications (voluntary or involuntary).

The main technologies used for facial recognition are:

- Principal Component Analysis (PCA);
- Local Feature Analysis (LFA);
- Neural Networks.

Recognition systems should be designed according to the type of application and to the attitude of the individual. The latter may be of three types:
1. Cooperative: the subject is motivated to use the system in order to be recognised so that s/he can access portals and gates to certain areas;
2. Uncooperative: the subject does not help or hinder the recognition process;
3. Hostile or reticent: the subject tries to avoid recognition and shows evasive behaviour.
The human face is made up of a "multidimensional set of images". From a biometric point of view, face recognition is not characterized by high permanence: the numerous facial expressions, age, the radical changes which can have certain features as hair, hairstyle, beard, moustache, etc; the presence of eye glasses, changes in the use of cosmetics or facial color, variation of the user pose and change the direction of the incident light, are examples of exterior features that can change in time and make facial recognition difficult. The 'impermanent' features of the face add to the complexity of the technical problems to solve. Nevertheless, successful techniques for personal identification have been developed at reasonable prices.

1.4 3D Face Recognition Systems

The difficulty of two-dimensional system of recognition (photographs, videotapes, etc. ...), comes from the type of data used to verify the similarity between two faces. This is because these devices are working on a two-dimensional representation in a three-dimensional scene.

The 3D Face Recognition Systems have the ability to reproduce in 3D an image, capturing the smallest detail with outstanding precision. The promises of 3D face recognition are the high accuracy of recognition necessary for high-security applications, reducing the problems of pose and lighting, the best location of facial features.

The advantages of this technology are in fact:

- The 3D technology is more efficient than two dimensions because it is able to analyze much more information because you have access to new information;
- A system of identifying 3D is less sensitive to the illumination conditions;
- The problem of the pose can be solved with the realignment of the faces;
- The occlusion can easily be found through a process of segmentation;
- The automatic generation of synthetic facial expressions;

conversely, there is that:

- The 3D acquisition hardware is expensive (the cost increases with the precision of acquisition);
- Some 3D acquisition systems are invasive;
- Long acquisition times;
- Some scanner systems can even be dangerous for the retina (laser);
- Replace 2D devices (cameras, videocameras, etc..) with 3D with new equipment is a process that requires time and high costs.

A first important classification of the techniques used is derived directly from the type of images that are presented as input to the process of recognition by dividing it into two broad categories: still images and video clips. In particular, in general applications of the automatic recognition are generally used still facial images of the subject to recognize or identify; these pictures are taken in layers of gray, as the color, almost anything used in the algorithms, presented a decidedly unfavorable report between the information conveyed

and the additional computational load required for its preparation. In some special applications, however, requires the face recognition in a complex scene and time variable, usually by video sequences taken by cameras (which would be that of people in an airport).
Face recognition systems currently available are based on following approaches: *Stereoscopy/photogrammetry* is the technique that detects the position, shape and size of an object starting from its photographs. The method performs the steps without coming into contact with the object, and is based on the principle of triangulation. The subject is photographed through a pair of stereoscopic cameras, at least two different angles and from each camera is projected a "line of view" to the points of the object. The intersection of these lines provides the 3D coordinates of points of interest. The basic concept is that knowing the location of the cameras in relation to the same points in two different images, you can calculate the 3D depth by triangulation. The reconstruction algorithms are divided in two main stage: matching (search for corresponding points between images) and reconstruction (from a match is possible to obtain the 3D point associated). As regards reconstruction, there are two modes: classic stereo (with two fixed cameras) and multiple views (with a camera in motion).

Time of flight acquisition is the technique that derive the distance to the object framed by measuring the time that a spot light takes to reach the subject framed and come back to the receiver placed in the device itself, from time of flight can be calculated immediately the distance traveled. In particular, they are optomechanic devices capable of emitting a laser, or an electromagnetic pulse that is distorted by the surface on which an impact, and receiving the reflected signal. From latter the time interval (time of flight) and then the distance between the instrument and the point found are measured.

Laser scanner is based on optical-mechanical devices capable of emitting an electromagnetic pulse (the laser) and receiving the reflected signal, measuring the time interval and therefore the distance between the instrument and the point found. This system operates by measuring thousands of points in the second form of "point clouds". For each measurement (x, y, z), the system provides the intensity of the return signal describing the surface of the object scanned. These systems, also known as active methods, project on the object a specific pattern of light and obtain the depth image by analyzing the deformation that the pattern projected undergoes. The patterns of structured light that can be used are different and it is also possible to combine different patterns in the same device. The main patterns are: binary, n-ary, gray code and phase-shift. These are among the most used because they allow also the reconstruction of subject in the slight movement. The quality of the extracted model depends on the conditions of global illumination, which limits the use of these systems to environments with controlled lighting. These systems are conceptually very similar to those in time of flight, but with the important difference that the first image is acquired in a single pulse, while the scanners require several acquisitions to reconstruct the complete 3D model of the object.

2. Face Recognition Methods

The Face recognition has involved during the past years several research areas: psychology, pattern recognition, neural networks, computer vision and computer graphics. It is due to this fact that the literature on face recognition is vast and diverse. In fact, many methods of face recognition are based on different principles, algorithms and a mixture of techniques.

The usage of a mixture of techniques makes it difficult to classify these systems based purely on what types of techniques they use for feature representation or classification. We propose the following categorization (Zhao et al., 2003):

- **Holistic matching methods.** These methods use the whole face region as the raw input to a recognition system. One of the most widely used representations of the face region is eigenpictures, which are based on principal component analysis.
- **Feature based (structural) matching methods.** Typically, in these methods, local features such as the eyes, nose, and mouth are first extracted and their locations and local statistics (geometric, and/or appearance) are fed into a structural classifier.
- **Hybrid methods.** Just as the human perception system uses both local features and the whole face region to recognize a face, a machine recognition system should use both. One can argue that these methods could potentially offer the best of the two types of methods.

Table 2 lists some of the techniques of face recognition.

Approach	Representative Work
Holistic methods	
Principal Component Analysis (PCA)	
Eigenfaces	Direct application of PCA
Probabilistic Eigenfaces	Two-class problem with prob. measure
Fisherfaces/Subspace LDA	FLD on eigenspace
SVM	Two-class problem based on SVM
Evolution Pursuit	Enhanced GA learning
Feature Lines	Point-to-line distance based
ICA	ICA-based feature analysis
Other Representations	
LDA/FLD	FLD/LDA on raw image
PDBNN	Probabilistic decision based NN
Feature-based methods	
Pure geometry methods	Earlier methods; Recent methods
Dynamic Link Architecture	Graph matching methods
Hidden Markov Model	HMM methods
Convolution Neural Network	SOM learning based CNN methods
Hybrid methods	
Modular Eigenfaces	Eigenfaces and eigenmodules
Hybrid LFA	Local feature method
Shape-normalized	Flexible appearance models
Component-based	Face region and components

Table 2. Categorization of still face recognition techniques

2.1 Three - Dimensional Face Recognition Methods

In the recent years, different 3D face recognition methods have been proposed, a possible categorization can include the following four groups (Pan et al., 2005):

The 3D face recognition methods that can be categorized into four groups (Pan et al., 2005):

Curvature analysis-based: Curvature is the intrinsic local property of curved surface. The local shape could be determined by its primary curvature and direction (Trucco & Verri, 1998). Therefore, most of the early studies used curvature to analyze the features of 3D facial data (Lee & Milios, 1990; Gordon a & b, 1991; Yacoob & Davis, 1994). Gordon presented a template-based recognition system involving descriptors based on curvature calculations from range image data. The sensed surface regions were classified as convex, concave and saddle by calculating the minimum and maximum normal curvatures. Then locations of nose, eyes, mouth and other features were determined, which were used for depth template comparison. Yacoob proposed an approach to label the components of human faces. Qualitative reasoning about possible interpretations of the components was performed, followed by consistency of hypothesized interpretations. (Zhang et al., 2002) proposed an efficient method based on Gaussian curvature analysis. It consists of three major steps, Gaussian curvature estimation, boundary detection, and region growing. Moreover, (Song et al., 2006) presented a curvature analysis based method to estimate the principal freeform feature in a specified region of interest for template fitting. Based on the minimal, maximal, mean or Gaussian curvature computing, the geometric information is transferred to the curvature domain. Using a variant of Laplacian smoothing methods, the high frequency noises and interferences in the curvature domain are suppressed and the principal feature is addressed. By feeding back the extracted feature information to the geometric shape, the geometry of the template is estimated based on the feature analysis.

Spatial matching based: Recognizing was performed via matching facial surface or profile directly in 3D Euclidean space (Beumier & Acheroy, 2000; Pan et al., 2003; Wu et al., 2003; Pan & Wu, 2005; Lu et al., 2004). This kind of approaches generally assume that the facial surface is a rigid object so that are not competent for the recognition among the models with expressions.

Shape descriptor based: It exploits the shape representation or shape descriptor to achieve the recognition task in the representation domain. For instance, (Chua et al., 2000) used point signature - a representation for free-form surfaces for 3D face recognition, in which the rigid parts of the face of one person are extracted to deal with different facial expressions. (Wang et al., 2004) used a new shape representation called Local Shape Map for 3D face recognition. This kind of technique is simple and somewhat robust to small perturbations. However, its classification rate is considered to be low. (Berretti et al., 2008) presented an approach based on integral geometric shape information of a face. The method extracts salient face information by jointly considering metric and spatial properties between points of 3D face scans. First, the face surface is partitioned into a set of iso-geodesic surfaces, centered on the nose tip and characterized by increasing values of geodesic distance. Facial information captured by the iso-geodesic surfaces is then represented in a compact form by extracting their basic 3D shape and evaluating the spatial relationships between every pairs of surfaces. This is accomplished by a modeling technique capable to quantitatively measure the spatial relationships between 3D entities. Finally, surfaces and their relationships are cast to a graphlike representation, where graph nodes are the representations of the iso-geodesic surfaces, and graph edges are their spatial relationships. In this way, the similarity between two 3D face models can be estimated extracting their graph representation and combining distances between arc labels of the two graphs.

Recover-and-synthesis based: For this kind of methods, their probes still are 2D images but not 3D data. The partial 3D information was recovered from the 2D image, then the facial image in virtual view was synthesized for recognition (Zhao, 1999), (Lee & Ranganath, 2003), (Hu et al., 2004) or the recognition task was accomplished with the recovered parameters (Blanz et al., 2002). This type of approaches do not make full use of those online-available 3D information.

3. The Proposed System

In this chapter the authors present a 3D face recognition system to personal identification based on Pseudo 2D Hidden Markov Models training by expression-invariant representation (*Canonical Form*) of the databases faces.

In the section 1 has been emphasized that the main problems encountered in 3D face recognition are related to deformations of the face due to different facial expressions. Fortunately, the class of transformations that can suffer the surfaces representing the human faces are not arbitrary and can be modelled as isometric transformations that preserve the lengths. One way to find a representation which is the same for all isometric surfaces was proposed by Elad and Kimmel, named "*bending invariant canonical form*" (Elad & Kimmel, 2001). In fact, from the consideration that the identity of a person is associated to the intrinsic geometry of the surface of the face, while facial expressions are associated with extrinsic geometry, begin the idea underlying this method: to represent the intrinsic geometry of the surface of the face in a form that is identical for different postures of the face.

This invariant representation is an embedding of the intrinsic geodesic structure of the surface in a finite dimensional Euclidean space, in which geodesic distances are approximated by Euclidean ones: the *Canonical Forms*.

To construct a Canonical Form it is necessary, by first, a three-dimensional surface of face of subject that is in recognition phase, called Mesh; next, to measure the geodesic distances between points on the surfaces by Fast Marching on triangulated curved surfaces, and finally the Multi-Dimensional Scaling (MDS) algorithm is applied to extract a finite dimensional flat space in which geodesic distances are represented as Euclidean ones. The first phase of system is summarized in Fig. 1.

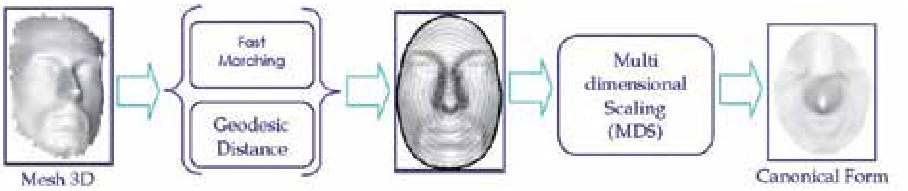


Fig. 1. Canonical Form Computation

In the second phase, the obtained expression-invariant representation of the face (*Canonical Form*) is putting into the Pseudo 2-D Hidden Markov Model with 3-6-6-6-3 structure to perform the facial recognition (see Fig. 2).

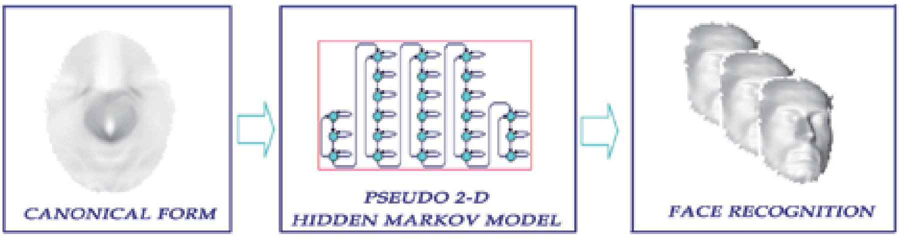


Fig. 2. Face Recognition Process

4. Fast Marching Methods

The Fast Marching Method is a computational technique that approximate the solution of the "Eikonal equation" non-linear that is part of the broader class of Hamilton-Jacobi equations.
The Eikonal equations are in form:

$$|\nabla u(x)| = F(x)\ in\ \Omega, \qquad F(X) > 0 \tag{1}$$

$$u = g(x)\ on\ \Gamma \tag{2}$$

where Ω is a domain on $\Re^n$ The right side is generally known as it is the boundary condition that u equals a given function *g(x)* along the curve or surface Γ in Ω .
One of the main difficulties in solving these equations is that the solution should not be differentiable even with smooth boundary data. The techniques based on Fast Marching Method are founded on two components. First, exploring the upwind "viscosity schemes", they automatically select solutions that include non-differentiability in natural ways. The second component matches the causality of these patterns with fast sorting methods borrowed from the problems of discrete networks, making the Fast Marching Method computationally efficient. Considering a domain Ω with N total points, the complexity of these algorithms is the order of O (N log N).
Many applications make use of the solution of non-linear Eikonal equation such as the problem of computing accurately the distance to a curve or a surface.
For example, consider the case of $|\nabla u| = 1$ outside the unit circle u= 0 on the unit circle it is verified that the function $u(x, y) = (x^2 + y^2)^{1/2} - 1$ corresponding to the distance function from the unit circle solves the given Eikonal equation. By specifying the right side F(x) = 1 and assuming zero the boundary condition, the solution is precisely the distance from the initial curve Γ . The solution is shown in Fig. 3, as a set of concentric circles, and you can verify that is differentiable everywhere.

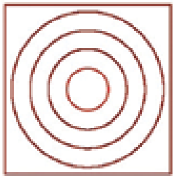


Fig. 3. Distance function to Eikonal equation $|\nabla u| = 1$

Another perspective is to imagine a disturbance propagating with unit speed away from the initial curve, and to calculate the propagation delay "first arrival time" at each point in the domain Ω. The points where the solution is not differentiable are placed between two points on the boundary curve at the same distance. By implementing the same calculation on the surface and back (backtrack) in a way orthogonal to the curves of propagation (curves in which the propagation delay is equal for all points) we can calculate the shortest path on the manifold.

4.1 Upwind Schemes

To deduce the reason why to use the upwind schemes (forward) for approximating the gradient operator, consider the Eikonal equation given by

$$\begin{gathered} \sqrt{u_x^2} = F(x) \quad u(0) = 0 \\ u(0) = 0 \end{gathered} \tag{3}$$

The right side $F(x) > 0$ is given, the objective is to calculate $u(x)$ away from the boundary condition that $u(0) = 0$. It possible to notice that the solution to this problem is not unique because if $v(x)$ solves the problem, then does it also $-v(x)$, and then attention will focus on non-negative solutions u.

It is feasible to imagine building the solution "outwards" along the positive and negative x-axis from the origin by solving each problem separately. Considering the following ordinary differential equations:

$$\begin{gathered} \frac{du}{dx} = F(x) \\ u(0) = 0 \quad x \geq 0 \end{gathered} \tag{4}$$

$$\begin{gathered} \frac{du}{dx} = -F(x) \\ u(0) = 0 \quad x \leq 0 \end{gathered} \tag{5}$$

them right-hand side is only a function of x: a numerical quadrature is performing. Using the standard finite difference notion that $u_i \approx u(i\Delta x)$ and $F_i = F(i\Delta x)$, you can approximate each of these two solutions using Euler's method,

$$\frac{u_{i+1} - u_i}{\Delta x} = F_i \qquad i > 0 \tag{6}$$

$$\frac{u_i - u_{i-1}}{\Delta x} = -F_i \qquad i \leq 0 \tag{7}$$

where $u_0 = 0$. This is an upwind scheme: it is possible to compute derivatives using points "upwind" or towards the boundary condition.

4.2 The Fast Marching Method on Orthogonal Grids

In the previous section has been told that the Eikonal equation is non-differentiable and that the Fast Marching Method objective is to create an appropriate weak solution that arise from satisfying the entropy condition. The aim in this numerical scheme is the correct direction of the upwinding and treatment of sonic points. One particular upwind approximation to the gradient, is the following

$$\begin{bmatrix} \max(D_{ij}^{-x}T, -D_{ij}^{+x}T, 0)^2 + \\ \max(D_{ij}^{-y}T, -D_{ij}^{+y}, 0)^2 \end{bmatrix}^{1/2} = F_{ij} \tag{8}$$

The Fast Marching Methods progress in an upwind wise to find the solution T. It is fundamental observe that in Eq. 8 the information propagates from smaller to larger values of T, so the algorithm is based on solving the Eq. 8 by building the solution outwards starting from the smallest value of T.

The Fast Marching algorithm is initialized as follows:

- The points are tagged as Alive
- All points one considered grid point away are tagged as Close
- All remaining grid points are tagged as Far

Then the algorithm proceeds as follow. The adjective "fast" is due to explorating of the points is done only for those places in a narrow strip outside of the front. The front is then moved forward in an upwind wise considering a set of points belonging to a band around the front stand; then it continues to move ahead on the points of the band; it freezes the values of existing points and, at last, incorporates some of those in front and calculate the new band around front (Chopp, 1993), (Malladi et al., 1993), (Adalsteinsson & Sethian, 1995). Ideas may be clarified by looking the main loop of method:

1) Points with the smallest value of T are tagged with Tial;
2) The Trial points becomes Alive and takes off from Close;
3) All Trial neighbors that are not Alive are tagged as Close and, if they are in Far, are removed and added to Close group;
4) The values of T for all the neighbors from the Eq. 8 are recalculated by solving the quadratic equation with only the values of the points that are Alive;
5) The cycle begins again from 1).

The algorithm works because the process of recalculation of the T values of upwind neighboring points produces a value less than those of accepted points (those in Alive). This feature, called monotone properties, allows to move forward by choosing the narrow strip of Trial points with the minimum value of T. In this way is possible to change neighbors without having to go back and to adjust the accepted values as is shown in Figure 4.

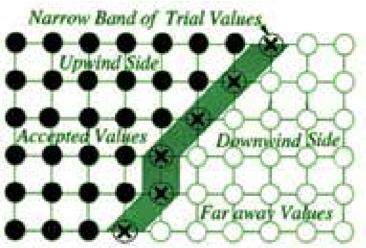


Fig. 4. Upwind construction of accepted values.

The Fast Marching methods are used to solve the Eq.10 by means of an update procedure, where it can imagine an uniform square grid and the goal is to update the value of T at the center point i,j. The Fast Marching methods, also, can be extended on a particular Triangulated Planar Domain (Kimmel & Sethian, 1998). To do it is necessary built a monotone update procedure on the Triangulated Mesh or, generally on an Arbitrary Triangulation (acute triangulation and obtuse triangulation) (Kimmel & Sethian, 1998).

5. GEODESIC DISTANCE

The Fast Marching Method algorithm can be used to compute distances on triangulated manifolds, and, hence, to construct minimal geodesics. Before proceeding, is useful to say a few words about a geodesic distances.

The geodesic distance between two pixels p and q is the length of the shortest path from p to q. Suppose $P = \{p_1, p_2, ..., p_n\}$ is a path between pixels p_1 and p_n, i.e. p_i and p_{i+1} are connected neighbors for $i \in \{1, 2, ..., n-1\}$ and p_i belong to the domain for all i. The path length $l(P)$ is defined as:

$$l(P) = \sum_{i=1}^{n-1} d_N(p_i, p_{i+1}) \tag{9}$$

the sum of the neighbors distances d_N between adjacent points in the path.

If the idea is to use the Fast Marching on triangulated manifolds first is necessary to solve the Eikonal equation on the triangulated surface with speed $|\nabla T| = F = 1$ to calculate the distance from a source point; then you can then go back (backtrack) along the gradient by solving the ordinary differential equation

$$\frac{dX(s)}{ds} = -\nabla T \tag{10}$$

where at the source points the distance is zero and $X(s)$ is the geodesic path. Succesively, using the Heun's method of integration of the second order on Triangulated surface is possible to switch to a first order. Applying the Fast Marching method to triangulated domains requires a careful analysis of the update of one vertex in a triangle, while the T values at the other vertices are given: when the inner part of the triangle is integrated, three neighboring triangles are used to interpolate T with a polynomial second-order, whose six coefficients calculated from data values associated with T vertices. The fast marching on triangulated domains method can compute the geodesic distance between one vertex and the rest of the n surface vertices in O(n) operations. Repeating this computation for each vertex, we compute a geodesic distance matrix Δ in O(n^2) operations. So we compute a geodesic distance matrix Δ, where

$$\delta\left(p_i, p_j\right) = \delta_{ij} \tag{11}$$

Each ij entry of Δ represents the square geodesic distance between the vertex i and the vertex j, that is δ_{ij}=GeodesicDistance(Vertex$_i$;Vertx$_j$), where

$$[\Delta] = \delta^2_{ij} \tag{12}$$

Thereby, given a triangulated surface, we apply the fast marching procedure for each vertex in a subset of the vertices as a source point and obtain the geodesic distance matrix, **Δ**.

6. Multidimensional Scaling

The main problem of human face is that it can not be considered a rigid object because it undergoes deformations resulting from facial expressions. So classical surface matching methods, aimed to find a Euclidean transformation of two surfaces which maximizes some shape similarity criterion, are more suitable for rigid objects than moving objects. The second problem is that facial surface has class of transformations that are not arbitrary, and in literature facial expressions can be modeled as isometric (or length-preserving) transformations that do not stretch and do not tear the surface, or more strictly, preserve the surface metric. When is studied a human face, the problem is implemented a deformable surface matching algorithm is to find a representation, which is the same for all isometric surfaces.

One of the algorithms most widely used for this purpose is the Multi-Dimensional Scaling (MDS). The idea is, preserving the relative distances between pairs of points, to transform a set of points in a high-dimensional space to a lower-dimensional one (Rohde, 2002). In most cases, in fact, is very complex to carry out results from studies handling high-dimensional vector spaces. So, if it is possible to obtain a set of vectors in a much lower-dimensional space, while preserving their similarity structure, the operations could be performed more efficiently. In other words multi-dimensional scaling (MDS) techniques are applied to extract a finite dimensional flat space in which geodesic distances are represented as Euclidean ones.

In the MDS method, the dissimilarity between pairs of objects of a collection are given as coefficients and then approximated as distances between corresponding pairs of entities in the visual representation. The quality of this approximation is expressed as a loss function, which produces the best value to its minimum. Is possible say that Richardson (Richardson, 1938) and Young and Householder (Young & Householder, 1938) may have officially initiated the multidimensional scaling literature, but frequent applications did not begin to appear until the papers by Torgerson (Torgerson, 1952) and, successively, by Shepard and Kruskal. Torgerson used a one-dimensional scaling technique to convert dissimilarity ratings to target distances and then attempted to find a set of points whose pairwise Euclidean distances best matched the target distances, according to mean-squared error. Though very effective, this technique is too complicated for many applications and a serious incident is that the correct scaling method is difficult to determine and can vary from problem to another. An improvement occurred with Shepard (Shepard, 1962) that supposed that the goal of MDS should be to obtain a monotone relationship between the actual point distances. In literature Torgerson's method has spread as MDS metric. On the contrary, the technique introduced by Shepard is known as non-metric MDS. Kruskal (Kruskal, 1964) further developed the procedure and defined the function known as stress function, relating the pairwise distances and the ranking of dissimilarities. The basic technique developed by Shepard and Kruskal remained the standard for most applications of MDS.

Schwartz (Schwartz et al, 1989), were the first to use multidimensional scaling (MDS) as a tool for studying curved surfaces by planar models. In his work, he applied an MDS technique to flatten convoluted cortical surfaces of the brain, onto a plane, in order to study their functional architecture. Zigelman (Zigelman et al., 2002) and Grossman (Grossman et al., 2002) extended some of these ideas to the problem of texture mapping and voxel-based cortex flattening. A generalization of this approach was introduced in the recent work of Elad e Kimmel (Elad & Kimmel, 2001), that proposed an efficient algorithm to construct a signature for isometric surfaces.

The problem can be so defined: suppose you have a collection of n objects and a way to determine the differences between each pair

$$\Delta = \lfloor \delta_{ij} : i, j = 1,\ldots,n \rfloor \tag{13}$$

The metric Multidimensional Scaling is a procedure to find a configuration of n points in a space of *p* size, usually Euclidean, so the point

$$x_i^* = (x_{i1}^*,\ldots,x_{ip}^*)^T \tag{14}$$

represents uniquely the object *i*, and the Euclidean distance between points x_i^* and x_j^* is:

$$d_{ij}(X^*) = \left\| x_i^* - x_j^* \right\| = \sqrt{\sum_{a=1}^{p} \left(x_{ia}^* - x_{ja}^* \right)^2} \tag{15}$$

The corresponding dissimilarity δ_{ij} for all pairs of objects (i, j) is so approssimated:

$$\underset{i<j}{\forall} d_{ij}(X^*) \approx \delta_{ij} \tag{16}$$

Is usually sufficient to consider each pair of objects (i, j) only once with i<j, since differences are symmetrical. Asymmetric matrix elements $\hat{\Delta}$ must be mediated:

$$\delta_{ij} = \delta_{ji} = \left(\hat{\delta}_{ij} + \hat{\delta}_{ji}\right)/2 \tag{17}$$

For a given configuration *X*, the approximation error that is obtained to represent the dissimilarity between objects i and j can be defined as follows:

$$e_{ij} \overset{def}{=} \left| d_{ij}(X) - \delta_{ij} \right| \tag{18}$$

The MDS - least squares technique (least-squares) defines the loss function as a weighted sum of normalized errors and possibly on all pairs of objects (i, j), and penalizes the overall approximation error. The minimum of this function on X is found through numerical optimization to obtain the desired configuration X *.

Each of the MDS algorithms is an instance of a heuristic minimization function well known. This heuristic provides a theoretical basis and is usually the result of a large number of options and variations derived from operations research professionals and other fields. Some of the choices are determined simply by the application domain, others may be based on theoretical or empirical evidence found in the literature.

7. Hidden Markov Models

Hidden Markov models (HMMs) are a well-developed technology for classification of multivariate data that have been used extensively in speech recognition, handwriting recognition and even sign recognition. They consist of states, possible transitions between states (and the probability of those transitions being taken) and a probability that in a particular state, a particular observation is made (see Fig. 5). The "Hidden" mean that the state of the HMM can not, in general, be known by looking at the observations. Then, they are also "Markov" because the probability of observing an output depends only on the current state and not on previous states. By looking at the observations, using an algorithm known as the Viterbi Algorithm, an estimate of the probability that a particular instance or stream observed was generated by that HMM can be computed (Rabiner, 1989).

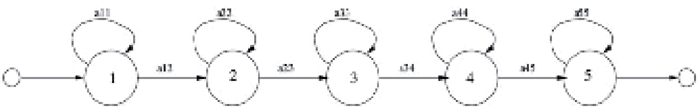


Fig. 5. Schema of Hidden Markov Models

For every class of interest, is possible apply HMMs, to calculate the probability returned by each of them and choose the most probable one.

7.1 One - Dimensional Hidden Markov Models

The main idea of this structure is to design a multi-state system which outputs this sequence while being in a state q_t at a given time t (Samaria & Young, 1994). At each state, the system is designed to output a certain observation from the vocabulary with a likelihood given by the output probabilities. At each time step the system will switch from its current state to the next (possibly stay in the same state) with a transition probability. The given of the transition probabilities will therefore implicitly design the (hidden) structure of the system. The state of the system at time $t = 1$ is given by its initial state probabilities.

The elements that characterised a HMMs are:

- $N=|S|$ is the number of states of the model. If S is the set of states, then $S=\{s_1,s_2,\dots,s_n\}$. s_i is one of the states that can be employed by the model. To observe the system are used T observation sequences, where T is the number of observations. The state of the model at time t is given by q_t *in* S, $1 < t < T$;
- $M=|V|$ is the number of different observation symbols. If V is the set of all possible observation symbols (also called the codebook of the model), then $V=\{v_1,v_2,\dots,v_M\}$;
- $A = \{a_{ij}\}$ is the state transition probability matrix, where a_{ij} is the probability that the state i became the state j:

$$a_{ij} = p(q_t = s_j \mid q_{t-1} = s_i) \tag{19}$$

  where $1 \le i\,;\, j \le N$, with constraint $0 \le a_{i,j} \le 1$, *and* $\sum_{j=1}^{N} a_{ij} = 1$, $1 \le i \le N$

- $B=\{b_j(k)\}$ the observation symbol probability matrix, $b_j(k)$ is the probability to have the observation k when the state is j:

$$b_j(k) = p\ (O_t = v_k \mid q_t = S_j) \tag{20}$$

  where $1 \le j \le N;\ 1 \le k \le M$; and O_t is the observation symbol at time t.

- $\Pi = \{\pi_1,\pi_2,\dots\pi_N\}$ is the initial state distribution:

$$\pi_i = p\ (q_j = S_i) \tag{21}$$

  where $1 \le j \le N$.

Using a shorthand notation, a HMM is defined by the following expression:

$$\lambda = (\mathrm{A},\mathrm{B},\pi). \tag{22}$$

The training of the model, given a set of sequences $\{O_i\}$, is usually performed using the standard Baum-Welch re-estimation, which determines the parameters (A,B,π) that maximize the probability $P(\{O_i\} \mid \lambda)$.

7.2 Two - Dimensional Hidden Markov Models

Pseudo Two-Dimensional Hidden Markov Models (P2D-HMMs) can be considered how a generalization of the One-Dimensional HMMs, helpful to represent two-Dimensional data. A word "Pseudo" is used because of the state alignments of consecutive columns are calculated independently of each other (Werner & Rigoll, 2001).

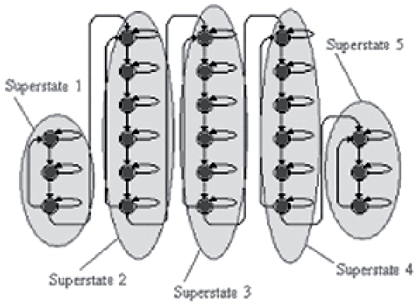


Fig. 6. Schema of Pseudo 2Dimensional Hidden Markov Model

In Fig. 6, P2D-HMMs are organized in a two-dimensional way: the horizontal states are named superstates; each of them consists of a One-dimensional HMMs in vertical direction. For recognition process a P2D-HMM can be transformed into an equivalent one-dimensional HMM.

The HMMs can be trained by the standard Baum-Welch algorithm and the recognition step can be carried out using the standard Viterbi algorithm.

The elements of an P2D-HMM are [Nefian & Hayes III, 1999] :

- N_0 are the number of superstates, and $S_0 = \{S_{0,i}\}$ $1 \le \boldsymbol{i} \le \boldsymbol{N_0}$ the set of superstates.
- The initial superstate distribution, $\Pi_0 = \{\pi_{0,i}\}$, were $\pi_{0,i}$ are the probabilities of being in super state i at time zero.
- The super state transition probability matrix,

$$\mathbf{A}_0 = \{a_{0,ij}\} \tag{23}$$

were $a_{0,ij}$ is the probability of transitioning from super state $\boldsymbol{i}$ to superstate $\boldsymbol{j}$.

- The parameters of the P2D-HMMs, which include:
  - ✓ The number of embedded states in the *kth* superstate, $\boldsymbol{N_1^{(k)}}$, and the set of embedded states, $S_1^{(k)}$ $\{S_1^{(k)}{}_{,i}\}$.
  - ✓ The initial state distribution, $\Pi_1^{(k)} = \{\pi_1^{(k)}{}_{,i}\}$, where $\pi_1^{(k)}{}_{,i}$ are the probabilities of being in state $\boldsymbol{i}$ of super state $\boldsymbol{k}$ at time zero.
  - ✓ The state transition probability of transitioning from state $\boldsymbol{k}$ to state $\boldsymbol{j}$.
- Finally, there is the state probability matrix,

$$B^{(k)} = \{b_i^{(k)}(O_{t0,\,t1})\} \tag{24}$$

for the set of observations where $\mathbf{O}_{t0,\ t1}$ represent the observation vector at row $\boldsymbol{t_0}$ and column $\boldsymbol{t_1}$. In a *continuous density* HMM, the states are characterized by continuous observation density functions. The probability density function that is typically used, is a finite mixture of the form

$$b_i^{(k)}(O_{t_0,t_1}) = \sum_{m=1}^{M} c_{im}^{(k)} N(O_{t_0,t_1}, \mu_{im}^{(k)}, U_{im}^{(k)}) \tag{25}$$

where $1 \le i \le N_1^{(k)}$, $c_{im}^{(k)}$ is the mixture coefficient for the *m*th mixture state ***i*** of super state k. $N(O_{t_0,t_1}, \mu_{im}^{(k)}, U_{im}^{(k)})$ is a Gaussian with mean vector $\mu_{im}^{(k)}$ and covariance matrix $U_{im}^{(k)}$. Let

$$\Lambda^{(k)} = \left\{\Pi_1^{(k)}, A_1^{(k)}, B_1^{(k)}\right\} \tag{26}$$

be the set of parameters that define the *kth* super state. Using a shorthand notation, an embedded HMM is defined as the triplet:

$$\lambda = (\Pi_0, A_0, \Lambda) \tag{27}$$

where:

$$\Lambda = \left\{\Lambda^{(1)}, \Lambda^{(2)}, \Lambda^{(N_0)}\right\} \tag{28}$$

Although more complex than a one-dimensional HMM, a P2D-HMM is more appropriate for data that are two-dimensional, and has a complexity proportional to the sum of the squares of the number of states:

$$\sum_{k=1}^{N_0} (N_1^{(k)})^2 \tag{29}$$

In Fig. 3 it's shown an example of a P2D-HMM with a structure 3-6-6-6-3: the superstates 1 and 5 are constituted by a left to right 1D-HMM with 3 states; instead the superstates 2, 3 and 4 are constituted by a left to right 1D-HMM with 6 states. This is a particular structure used in the face recognition system realized by the authors and proposed in the followed sections.

8. The Mesh and GAVADB 3D face database

The Mesh, three-dimensional surfaces of face of subject, is in a ASE format, chosen for its excellent readability. A file ASE, in fact, is perfectly compatible with a text file, it opens up with a common notepad to show the content. The 3D image is a graphical model with a surface constructed from polygons. The polygons are described by the graphics system as solid faces, rather than as hollow polygons, as is the case with wireframe models. Separate portions of mesh that make up the model are called polygon mesh and quadrilateral mesh. The mesh is stored as three-coordinates points x, y and z of each point of the mesh and the triangles defined by the points themselves. The number of a three-coordinates points are variable, in this system a Mesh have 3000 points.

To test the proposed system the GAVADB database (Moreno & Sanchez, 2004) is used. It is a 3D face database. It contains 549 three-dimensional images of facial surfaces. This database has 9 images for each of the 61 different individuals with 45 male and 16 female. The total of the individuals are Caucasian and their age is between 18 and 40 years old. The database provides systematic variations with respect to the pose and the facial expression. In particular there are: 2 neutral expressions with frontal views; 2 neutral expressions with views x-rotated with ±30°, looking up and looking down respectively; 2 neutral expressions with views y-rotated ±90°, left and right profiles respectively; 3 frontal gesture images laugh, smile and a random gesture chosen by the user, respectively.

8.1 The Implementation

Given a polyhedral approximation (Mesh) of the facial surface, S, the fast marching on triangulated domains method computes the geodesic distance between one vertex and the rest of the *n* surface vertices. Repeating this computation for each vertex, we compute a geodesic distance matrix Δ, where for each ij entry of Δ represents the square geodesic distance between the vertex i and the vertex j, that is δ_{ij}=GeodesicDistance(Vertex$_i$;Vertx$_j$), with [Δ]= δ^2_{ij} (see Eq 11, Eq. 12). One of the crucial steps in the construction of the canonical form of a given surface, is an efficient algorithm for the computation of δ_{ij}, that is the geodesic distances.

The resulting matrix Δ is invariant under isometric surface deformations, but is not a unique representation of isometric surfaces since it depends on arbitrary ordering and the selection of the surface points. Treating the squared mutual distances as a particular case of *dissimilarities*, it is possible to apply a dimensionality-reduction technique, that is *multidimensional scaling* (MDS), in order to embed the surface into a low-dimensional Euclidean space $\Re^m$. This is equivalent to finding a mapping between two metric spaces,

$$\begin{aligned} &\varphi : (S,\delta) \to (\Re^m, d) \\ &\varphi(p_i) = x_i \end{aligned} \tag{30}$$

that minimizes the embedding error,

$$\begin{aligned} &\varepsilon = f(\|\delta_{ij} - d_{ij}\|) \\ &d_{ij} = \|x_i - x_j\|_2 \end{aligned} \tag{31}$$

for some monotone function *f* that sums over all *ij*. The obtained *m*-dimensional representation is a set of points $x_i \in \Re^m (i = 1,\ldots,n)$, corresponding to the surface points p_i. Different MDS methods can be derived using different embedding error criteria (Borg et al. 1997). A particular case is the *classical scaling*, introduced by Young and Householder (Young et al. 1938), and that is a method here applied. The embedding in $\Re^m$ is performed by double-centering the matrix Δ

$$B = -\frac{1}{2} J \Delta J \tag{32}$$

(here $J = I - \frac{1}{n} U$; I is a *n*×*n* identity matrix, and *U* is a matrix consisting entirely of ones). The first *m* eigenvectors e_i, corresponding to the *m* largest eigenvalues of B, are used as the embedding coordinates

$$\begin{aligned} &x_i^j = e_i^j \\ &i = 1,\ldots,n;\ j = 1,\ldots,n \end{aligned} \tag{33}$$

where x_i^j denotes the *j*-th coordinate of the vector x_i. We refer to the set of points x_i obtained by the MDS as the *canonical form* of the surface. In the described work it is obtained a surface representation with *m*=20.

The autovectors values obtained by MDS that give the best characterization of a face surface, calculated Canonical Form, are saved in the binary format HTK (Hidden Markov Model ToolKit) in order to use the Hidden Markov Model to perform the face recognition. The binary file has an extension .bin, and contains the data saved in the format Big Endian, which is commonly used by Motorola processors (family 680x0), IBM and Sun as opposed to Intel and AMD's x86 family.

The Hidden Markov Model Toolkit (HTK) (Young and Young, 1994) is a toolkit for building and manipulating Hidden Markov Models. HTK is primarily used for speech recognition research although it has been used for numerous other applications such as the face recognition. HTK consists of a set of library modules and tools to train and to test the HMM and to run a recognition. In the proposed system is used the Pseudo 2-D HMM with topology 3-6-6-6-3 shown in Fig. 7.

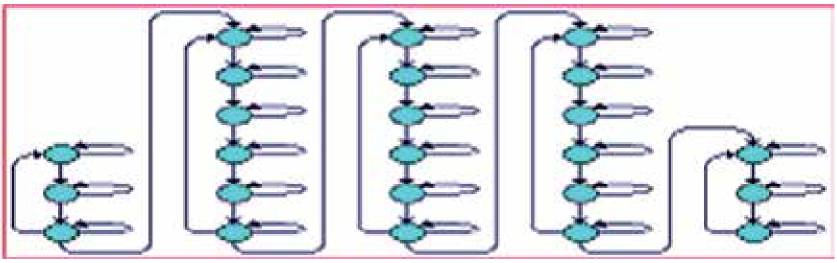


Fig. 7. Schema of used Pseudo 2D HMM

9. Experimental Results

After training of P2D-HMM on canonical forms constructed from GAVADB database, the authors performed several experiments and the results were encouraging, indeed the system achieved a rate of recognition equal to 98%. The robustness of the system was put to the test by the presence of various noise situations: incompleteness of mesh in the form of holes and occlusions. Moreover, comparing these outcomes with others different from other methods proposed in the literature and applied on the same 3D database, the proposed method results outperform all of them (see Table 2).

System	Authors	Results
Fishersurface (LDA)	(Heseltine et al., 2004a)	88.7%
Eigensurface (PCA)	(Heseltine et al., 2004b)	87.3%
Fishersurface (LDA)	(Heseltine et al., 2004c)	88.4%
3D matching + gray level	(Beumier & Acheroy, 2000)	98%
Morphable model [19]	(Blanz et al., 2002)	89%
3D eigenfaces [20]	(Xu et al, 2004)	71.1%
The presented system		**98%**

Table 2. Comparative results

Observing the Table 2, it is possible to notice that the proposed 3D facial recognition system obtained training a 3-6-6-6-3 P2D-HMM by expression-invariant representation (*Canonical Form*) of the databases faces, has reached reasonably good results, encouraging the authors to continue in this area for future developments in several directions such as: increasing the numbers of mesh points, that now are 3000; improving the Multi-Dimensional Scaling algorithm; reducing implementation time necessary to calculate the Canonical Form.

10. Reference

Adalsteinsson, D. & Sethian, J. A. (1995). A fast level set method for propagating interfaces. *Journal of Computational Physics*, Vol. 118, No. 2, 269–277

Ashbourn, J. (2000). *Biometrics: Advanced Identity Verification, The Complete Guide*, Springer, London

Berretti, S.; Del Bimbo, A. & Pala, P. (2008). SHREC'08 Entry: 3D Face Recognition using Integral Shape Information, *Proceedings of IEEE International Conference on Shape Modeling and Applications*, Stony Brook, New York, USA, pp. 255-256, June 2008

Beumier, C. & Acheroy, M. (2000a) Automatic Face Verification from 3D and Grey Level Clues, *Proceedings of RECPAD2000, 11th Portuguese conference on pattern recognition,* pp. 95-101, Porto, Portugal, May 11-12, 2000

Beumier, C. & Acheroy, M. (2000b). Automatic 3D Face Authentication. *Image and Vision Computing*, Vol. 18, 315-321

Blanz, V.; Romdhani, S. & Vetter, T. (2002). Face Identification across Different Poses and Illumination with a 3D Morphable Model, *Proceedings of IEEE International Automatic Face and Gesture Recognition (FGR'02)*, pp. 202-207, 2002

Borg, I. & Groenen, P. (1997). *Modern multidimensional scaling - theory and applications*, Springer-Verlag, ISBN 0-3879-4845-7, New York

Buciu, I. (2008). Overview of Face Recognition Techniques. *JEEE 2008*, 173-176

Cascetta, F. & De Luccia, F. (2004). Personal Identification Systems. *Mondo Digitale, No. 9*

Chopp, D. L. (1993). Computing Minimal Surfaces via Level Set Curvature Flow. *Journal of Computational Physics*, Vol. 106, 77-91

Chua, C.S.; Han, F. & Ho, Y.K. (2000). 3D human face recognition using point signature, *Proceedings of IEEE International Automatic Face and Gesture Recognition (FRG'00)*, pp. 233-238, 2000

Elad, A. & Kimmel, R. (2001). Bending Invariant Representations for Surfaces, *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'01)*, Vol. 1, pp. 168-174

Gordon, G.G. (1991). Face Recognition based on Depth Maps and Surface Curvature. *Geometric Methods in Computer Vision, SPIE Proceedings*, Vol. 1570, 234-247

Grossman, R.; Kiryati, N. & Kimmel, R. (2002). Computational surface flattening: a voxel-based approach, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 4, 433-441, ISSN 0162-8828

Heseltine, T.; Pears, N. & Austin, J. (2004a). Three-Dimensional Face Recognition: A Fishersurface Approach, *Proceedings of International Conference on Image Analysis and Recognition (ICIAR 2004*), pp. 684-691

Heseltine, T.; Pears, N. & Austin, J. (2004b). Three-Dimensional Face Recognition: An Eigensurface Approach, *Proceedings of International Conference on Image Processing (ICIP 2004),* pp. 1421-1424, ISBN 0-7803-8554-3, IEEE , New York

Heseltine, T.; Pears, N. & Austin, J. (2004c). Three-Dimensional Face Recognition Using Surface Space Combinations, *Proceedings of British Machine Vision Conference*, 2004

Hu, Y.; Jiang, D.; Yan, S.; Zhang, L. & Zhang, H. (2004). Automatic 3D Reconstruction for Face Recognition, *Proceedings of IEEE International Automatic Face and Gesture Recognition (FGR'04)*, pp. 17-19, Seoul, Korea, 2004

Jain, A. ; Bolle, R. & Pankanti, S. (1999). *Biometrics: Personal Identification in Networked Society*, Kluwer Academic Press, Boston

Jain, L.C.; Halici, U.; Hayashi, I.; Lee, S.B. & Tsutsui, S. (1999). *Intelligent biometric techniques in fingerprint and face recognition*, CRC Press, Boca Raton, Florida, USA

Kimmel, R. & Sethian, J. A. (1998). Computing geodesic paths on manifolds, *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 95, No. 15, pp. 8431-8435

Kruskal, J. B. (1964). Multidimensional scaling by optimising goodness of fit to a nonmetric hypothesis. *Psychometrika*, Vol. 29, 1–27

Lee, J.C. & Milios, E. (1990). Matching Range Images of Human Faces, *Proceedings of IEEE 3rd Interntional Conference on Computer Vision (ICCV 1990)*, pp. 722-726

Lee, M.W. & Ranganath, S. (2003). Pose-invariant face recognition using a 3D deformable model. *Pattern Recognition*, Vol. 36, No. 8, 1835-1846

Lu, X.; Colbry, D. & Jain, A.K. (2004). Three-Dimensional Model Based Face Recognition, *Proceedings of 17th International Conference on Pattern Recognition*, pp. 362-366, 2004

Malladi, R.; Sethian, J. A. & Vemuri, B. C. (1995). Shape modeling with front propagation: a level set approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 17, No. 2, 158 - 175

Moreno, A.B. & Sanchez, A. (2004). GavabDB: A 3D Face Database, *Proceedings of 2nd COST Workshop on Biometrics on the Internet: Fundamentals, Advances and Applications*, pp. 75-80, Vigo, Spain, March 2004

Nefian, A.V. & Hayes, M.H. III. (1999). An embedded HMM-based approach for face detection and recognition. *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP Proceedings., pp. 3553-3556 vol.6, 1999*

Pan, G.; Wu, Z. & Pan, Y. (2003). Automatic 3D Face Verification from Range Data. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'03)*, Vol. 3 (2003), 193-196

Pan, G.; Han, S.; Wu, Z. & Wang, Y. (2005). 3D Face Recognition using Mapped Depth Images, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05),* pp. 175-181

Pan, G. & Wu, Z. (2005). 3D Face Recognition from Range Data. *International Journal of Image and Graphics*, Vol. 5, No. 3, 573-593

Prabhakar, S.; Pankanti, S. & Jain A. K. (2003). Biometric Recognition: Security and Privacy Concerns. *IEEE Security & Privacy*, pp. 33-42, March-April 2003

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, pp. 257 – 286 Vol. 77 Issue 2, Feb 1989

Richardson, M. W. (1938). Multidimensional psychophysics, In: *Psychological Bulletin*, Vol. 35, 659-660

Rohde, D.L.T. (2002). Methods for binary multidimensional scaling. *Neural Computation*, Vol. 14, No. 5, 1195-1232 , ISSN 0899-7667

Samaria, F. & Young, S. (1994). HMM based architecture for face identification. *Image and computer Vision*, pp.537-543 vol. 12, October, 1994

Schwartz, E.L.; Shaw, A. & Wolfson, E. (1989). A numerical solution to the generalized mapmaker's problem: flattening nonconvex polyhedral surfaces. *IEEE Transactions on Pattern Analysis and machine Intelligence*, Vol. 2, No. 9, Sept. 1989, 1005-1008

Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. *Psychometrika*, Vol. 27, 125–139, 219–246

Song, Y.; Vergeest, J.S.M.; Wigers, T. & Langerak, T.R. (2006). Estimating principal deformable freeform features by curvature analysis, *Proceedings of 7th International Conference on Computer-Aided Industrial Design and Conceptual Design (CAIDCD '06)*, pp. 1-6, November 2006

Torgerson, W. S. (1952). Multidimensional scaling: I. Theory and method. *Psychometrika*, Vol. 17, 401–419

Trucco, E. & Verri, A. (1998). *Introductory Techniques for 3-D Computer Vision*, Prent. Hall Inc.

Wang, Y.; Pan, G. & Wu, Z. (2004). 3D Face Recognition using Local Shape Map, *Proceedings of IEEE International Conference on Image Processing (ICIP'04)*, 2004

Werner, S. & Rigoll, G. (2001). Pseudo 2-Dimensional Hidden Markov Models in Speech Recognition. *IEEE Workshop on Automatic Speech Recognition and Understanding,* pp. 441-444

Wu, Y.; Pan, G. & Wu, Z. (2003). Face Authentication based on Multiple Profiles Extracted from Range Data. *AVBPA'03, LNCS*, pp. 515-522 vol. 2688

Xu, C.; Wang, Y.; Tan, T. & Quan, L. (2004). A new attempt to face recognition using eigenfaces, *Proceedings of the 6th Asian Conf. on Computer Vision*, Vol. 2, pp. 884-889

Yacoob, Y. & Davis, L.S. (1994). Labeling of Human Face Components from Range Data. *CVGIP: Image Understanding*, Vol. 60, No.2, 168-178

Young, G. & Householder, G. S. (1938). Discussion of a set of points in terms of their mutual distances. *Psychometrika* Vol. 3, No. 1, Roger E. Millsap, (Ed.), 19-22

Young, S.J. (1994). The HTK Hidden Markov Model Toolkit: Design and Philosophy. *Entropic Cambridge Research Laboratory, Ltd*, Vol., 2, 2-44

Zhang, D. (2000). *Automated Biometrics Technologies and Systems*, Kluwer Academic Publishers, Boston

Zigelman, G.; Kimmel, R. & Kiryati, N. (2002). Texture mapping using surface flattening via multi-dimensional scaling. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 8, No. 2, 198-207

Understanding Correlation Techniques for Face Recognition: From Basics to Applications

A. Alfalou[1] and C. Brosseau[2]
1- ISEN Brest, L@bISEN, 20 rue Cuirassé Bretagne
CS 42807, 29228 Brest Cedex 2, France
2- Université Européenne de Bretagne, Université de Brest,
Lab-STICC and Département de Physique,
CS 93837, 6 avenue Le Gorgeu, 29238 Brest Cedex 3, France

1. Introduction

This chapter covers some of the latest developments in optical correlation techniques for face recognition using the concept of spectral fusion and placing a special emphasis on its practical aspects and applications for face recognition. Optical correlation is reviewed as a method to carry out instantly a decision on the target form (form to be recognized). A range of relevant practical topics is discussed, such as JTC (Joint Transform Correlator) and the Vander-Lugt architectures. Both of them are based on the "4f" optical setup and on a comparison between the target image and a reference image. The similarity between the two images is achieved by the detection of a correlation peak. The development of suitable algorithms for optical correlation and new technologies in electro-optical interfaces has significantly improved the processing capacity of optical correlators and reduce their sizes. To overcome the limitations of a decision taken by a simple detection of a correlation peak, more complex decision algorithms are required. These algorithms necessitate the integration of new correlation filters and the realization of multiple correlations and reconfigurable multi-channel architectures.

This chapter surveys also the main correlation filters and the main multi-decision correlator architectures. It will describe the development of a multi-reconfigurable architecture based of a new version of correlation filter, i.e. the segmented filter. More specifically, we will describe an all-optical and compact correlator and a new phase optimized filter which based on four ingredients: POF, sector, composite and segmented filters. To extend the application of the algorithm and improve this system, a correlation filter is adapted in order to recognize a 3D face using multiple cameras arranged in a pickup grid. To minimize the effect of overlapping between the different spectra, the filter optimization is realized by shifting different reference spectral images. This could increase the number of references included in the same filter.

2. Background and Notations of optical correlation

The use of optical correlation methods has evolved over the past thirty years. Thus, their scope has extended to security applications, as for example tracking and/or identifying for military applications (aircraft recognition, boats recognition, etc) or civilian applications (road signs recognition, face identification: bank, metro, airport). This interest for optical correlation techniques, based on the use of two Fourier transform (FT), is essentially due to the fact that it is possible to achieve an optical Fourier Transform instantly [1] (using a simple converging lens).

To perform optical correlation techniques, two major approaches of correlators are proposed and validated in the literature: JTC (Joint Transform Correlator) [2] and Vander Lugt correlator [3]. Both approaches are based:

1) On an all-optical set-up, called "4f" set-up [1],
2) On a comparison between the target image and reference image coming from a learning data-base,
3) On a simple detection of a correlation peak. The latter measures the similarity degree between the target image and the reference one.

We begin this section by recalling the principles of these two approaches:

2.1 Principle of the all optical filtering 4f setup

The 4f set-up (Fig. 1) is an all optical system composed of two convergent lenses. The 2D object **O** (displayed in the input plane) is illuminated by a monochromatic wave [1]. A first lens performs the FT of the input object **O** in the image focal plane (Fourier plane), $\mathbf{S_O}$. In this focal plane, a specific filter **H** is positioned (using optoelectronic interfaces). Next, a second convergent lens performs the inverse Fourier transform (FT^{-1}) in the output plane of the system to get the filtered image $\mathbf{O}'$ captured using a CCD (charge-coupled device) camera.

2.2 JTC : Joint transform correlator

Weaver and Goodman introduced the foundations of a new type of correlator for pattern recognition called JTC [2]. In their pioneering article [2], Weaver and Goodman demonstrated the possibility and the necessary conditions to achieve optically the convolution between two images displayed in the input plane of the JTC correlator: a target image (image to be recognized) and a reference image (image coming from a given data-base). Using this property and the classical "4f" set-up, Javidi and co-workers proposed [4] an optimized version of this JTC correlator. This optimized version allowing us to obtain very sharp and very intense peak correlations can increase the capacity of this type of correlators. This optimization is made possible by binarizing the joint spectrum[1] according to a well-defined threshold and using a SLM modulator. Another optimized version of the

[1] Joint spectrum: spectrum obtained after performing a FT of the input plane of the JTC correlator, containing the target image and the reference image placed at a distance *d* from the target [5-6].

JTC correlator was proposed by Javidi [5] by introducing a non-linearity in its Fourier plane (the JTC Fourier plane). This non-linearity in the Fourier plane increases the performances of this type of correlators.

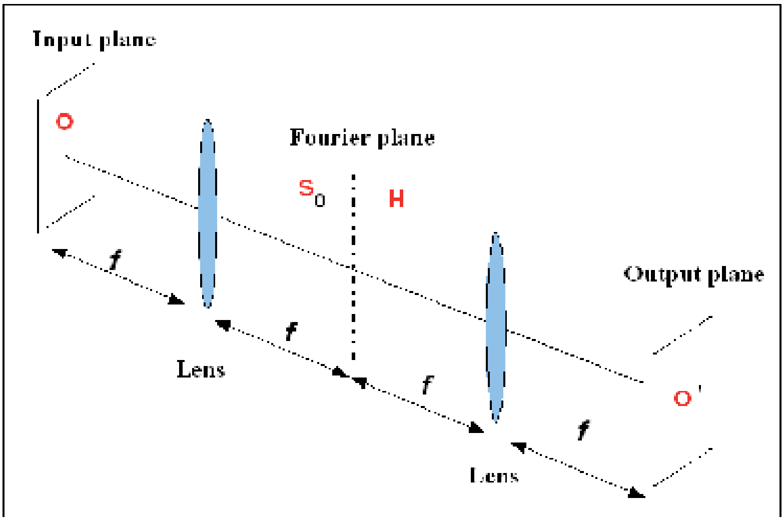


Fig. 1. All optical filtering 4f setup

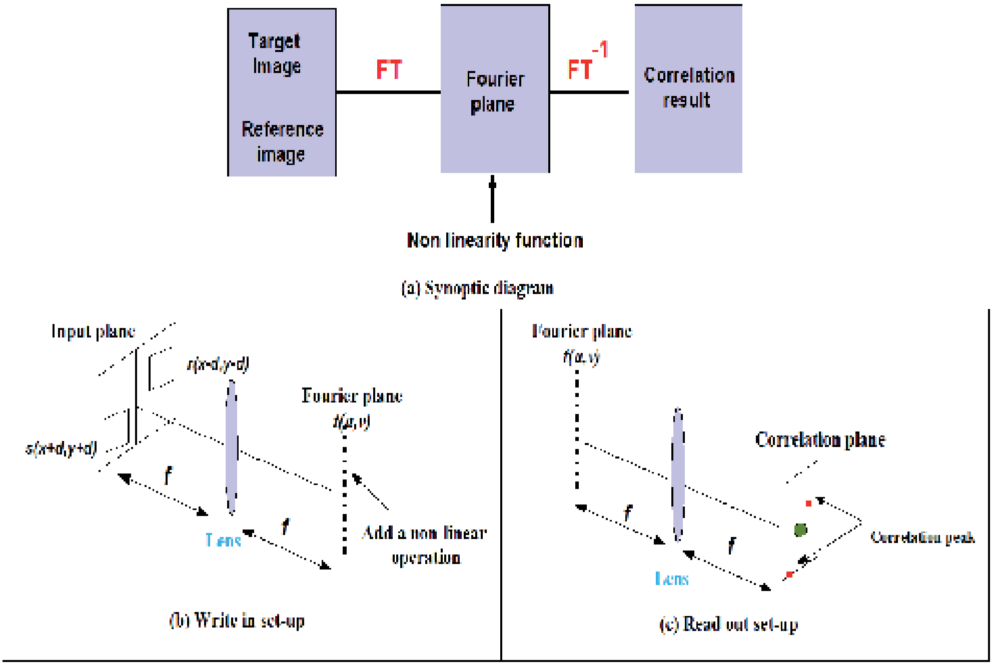
The architecture of a non-linear JTC correlator (JTC-NL) [6] is now described in some detail. Other types of architecture will be listed in subsection (2. 3). A synoptic diagram of the JTC set-up is presented in Fig. 2-a. Basically, it is an arrangement based upon the 4f set-up with a non-linear operation in its Fourier plane [5-6]. Its principle is to introduce, in the input plane, both the target and reference images separated by a given distance and to separate the writing and reading phases. In Ref.[6], this non-linear operation was achieved by using an Optically Addressed Spatial Light Modulator (OASLM) in the Fourier plane. A first beam coming from a laser illuminates the input plane, *I(x, y)* Eq. (1), which contains the scene, *s*(*x*, *y*), i.e. the target image to be recognized, and the image reference, *r*(*x-d*, *y-d*), where *d* represents the distance between the target and the reference images.

$$I(x,y)=s(x,y)+r(x-d,y-d). \tag{1}$$

The joint spectrum, obtained with Fourier transformed of *I(x,y)*, is recorded by the OASLM modulator and yields *t*(*u*, *v*) Eq. 2 (writing stage shown in Fig. 2-b).

$$\begin{aligned} t(u,v)=&|S(u,v)|\times \exp[\varphi_s(u,v)]+ \\ &|R(u,v)|\times \exp[\varphi_r(u,v)]\times \exp[-j(ud+vd)] \end{aligned}. \tag{2}$$

After a FT^{-1} of the joint spectrum, lighted with a second beam (reading stage shown in Fig. 2-c), the correlation between the target and the reference image on a CCD camera is obtained. The correlation plane has several peaks which correspond to the autocorrelation of the whole scene in the center (zero order), and two correlation peaks (corresponding to the reference and target images). The location of the different correlation peaks depends on the positions of the reference and the target images within the scene.



(a) Synoptic diagram

(b) Write in set-up

(c) Read out set-up

Fig. 2. Principle of the JTC correlator.

The performance of the JTC-NL correlators [5-6] proved to be larger than the linear JTC correlators [2-3]. Furthermore, by modifying the non-linearity in the Fourier plane of the JTC-NL correlator, it is possible to increase or decrease the discrimination power of the JTC-NL correlator. The interested reader may want to consult [7]. The optical implementation of the JTC correlator using an OASLAM modulator is simple to perform [6-7] and allows a high cadence (1 kHz) of processing (comparison between the target and reference images). Further adjustment in the non-linearity parameters can be easily made by changing, e.g. the voltage control of the modulator, the pulse duration, or the illumination of the modulator.

However, this correlator suffers from physical and algorithmic drawbacks. Indeed, the use of two SLM modulators in the JTC (a first one is used in the input plane to display jointly the target image and the reference one, and the other one is used in the Fourier plane to write/read the joint spectrum of JTC correlator) imposes a limited resolution depending on the technology used to manufacture these modulators. Besides the loss of information caused by these resolutions, set by the manufacturers, they also impose constraints on the choice of lens to be used for performing the FT of the input plane. The focal must verify equation (3).

$$f = \frac{N \operatorname{Re} s_{input-plane} \operatorname{Re} s_{Fourier-plane}}{\lambda}, \tag{3}$$

where $\operatorname{Re} s_{input-plane}$ denotes the resolution of the modulator used in the input plane, $\operatorname{Re} s_{Fourier-plane}$ is the resolution of the Fourier plane modulator, *N* is the number of pixels and λ is the wavelength. Similarly, the focal length of the second lens (to perform the second Fourier transform) is calculated by respecting the resolution of the modulator used in the Fourier plane and the resolution of the camera used to record the output plane: correlation plane.

Figure (3) shows an example of a correlation plane using a JTC-NL correlator [8]. In this plane, three peaks are visible: a central peak characterizing the zero-th order of the JTC correlator; the other two peaks are the two correlation peaks which show similar correlations between the target and reference images. The position of the two correlation peaks depends of the position of the target image relatively to the reference image placed both in the input plane.

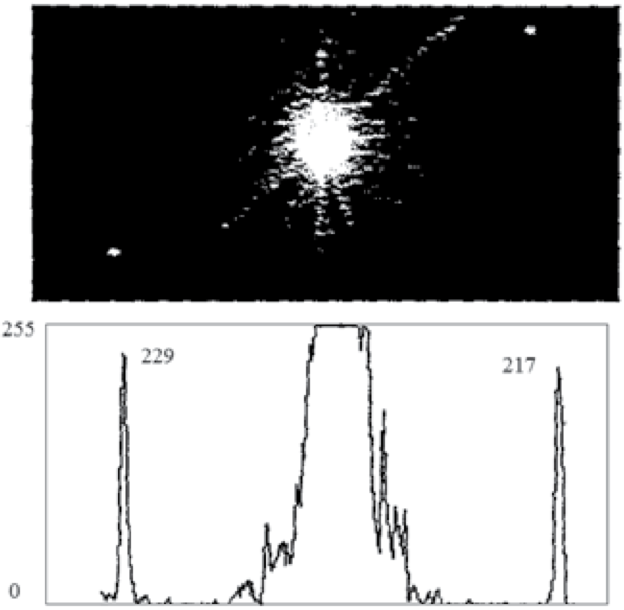


Fig. 3. Correlation plane obtained with an all Optical NL-JTC correlator [8].

2.3 The correlator JTC applied to facial recognition

To have a general overview on this type of correlator, we test by numerical simulations, some JTC architectures proposed in the literature. We examine the performance of this type of correlator vis-à-vis the face recognition application that interests us in this chapter. For this purpose, we begin by recalling the principles of several architectures implementing the

JTC correlator. All these architectures use the 4f set-up and differ from the treatment applied on the joint spectrum of the JTC correlator:

- *Classical JTC [2]: CL-JTC*

This classical architecture consists to record the intensity of the joint spectrum given by equation (2). The result is as follows

$$\begin{aligned} |t(u,v)|^2 \qquad & |S(u,v)|^2 + |R(u,v)|^2 \\ = & + \left(|S(u,v)| \times \exp[\varphi_s(u,v)]\right) \times \left(|R(u,v)| \times \exp[-\varphi_r(u,v) + j(ud+vd)]\right). \\ & + \left(|S(u,v)| \times \exp[-\varphi_s(u,v)]\right) \times \left(|R(u,v)| \times \exp[\varphi_r(u,v) - j(ud+vd)]\right) \end{aligned} \tag{4}$$

To obtain the correlation plane using the CL-JTC, we need to perform a FT^{-1} of the recorded intensity (Eq. 4) to get the three peaks in the correlation plane: one representing the zero-th order resulting from the FT^{-1} of the first line of Eq. 4, and two correlation peaks corresponding to the FT^{-1} of the second and third lines in Eq. 4. Studying these two correlation peaks determine the degree of resemblance between the target and the reference faces. However, this classical correlator has two major drawbacks rendering the decision of this type of correlator not always reliable:

1. It presents a very large zero-th order and very intense peak compared to the correlation peaks,
2. It is characterized by two low-intensity and large correlation peaks.

- *Binary JTC [4]: B-JTC*

Once the intensity of the joint spectrum (Eq. 4) is recorded, the binary JTC consists to binarize this intensity (Eq. 4) using the method detailed in equation (5). Then a FT^{-1} is performed to obtain the correlation plane.

$$SJ\_Bin(u,v) = \begin{cases} +1 & if \quad |t(u,v)|^2 \succ S \\ -1 & if \quad |t(u,v)|^2 \prec S \end{cases}, \tag{5}$$

where *SJ_Bin* is the binarized joint spectrum and *S* is a threshold that corresponds to the medium of the intensity joint spectrum values [4]. This correlator provides better performance than the classical JTC. It is more discriminating, has very sharp and intense correlation peaks and a very sharp zero-th order [4]. However, it is too sensitive especially to deformations of the target image relatively to the reference image (rotation, scale), which appear often in face recognition applications.

- *Nonlinear JTC* [5]: *NL-JTC*

To control the sensitivity of the JTC correlator, the authors of Ref [5] presented a non-linear version of this correlator. It consists in introducing a non-linearity function "$g(E)$" in the joint spectrum intensity (Eq. 4). The non-linearity function $g(E)$ is given by the following equation:

$$g(E(u,v)) = E^{K}(u,v) \quad with \quad \begin{cases} E(u,v) = |t(u,v)|^{2} \\ K \quad is\ a\ cons\tan t\ value \end{cases} , \tag{6}$$

where K represents the degree of freedom applied to our correlator. If K=0, we obtain the binary JTC correlate. Here we set $K = 0.5$ since this value leads to a good compromise between the power discrimination and the sensitivity [7-8].

- *JTC order without zeros [9]: NZ-JTC*

As was mentioned previously, the JTC correlator has a high zero-th order which is detrimental to obtain a good and reliable decision, especially for a small distance length between the object to recognize and the reference image. Reducing this distance is required for decreasing the size of the input plane that has a limited space-bandwidth product (*SBWP*). Indeed, the size of the input plane depends on the size of the scene (which contains the object to be recognized), the reference image and the distance between them. To solve this problem a first technique is proposed consisting in hiding the zero-th order by multiplying the correlation plane with a binary mask given by equation (7)

$$M(x,y) = \begin{cases} +0 & if \quad x\ and\ y \in [-c,+c] \\ 1 & if \quad otherwise \end{cases} , \tag{7}$$

where c is the width value of the desired mask. However, this mask cannot entirely solve the problem. Depending on the value of c, the correlation peaks can be filtered out. To overcome correctly the zero-th order problem, another technique was proposed and validated in the literature [9]. It consists in eliminating the first two terms of the joint spectrum (Eq. 4), i.e. $\left(|S(u,v)|^{2} + |R(u,v)|^{2} \right)$, which introduce this zero-th order.

This approach was chosen by us. For that purpose, we first record separately the spectra intensities of the two images presented in the input plane: the face to be recognized and the reference. Then, we subtract the two intensity values of the equation (4) to get

$$\begin{aligned} |t(u,v)|^{2} = {} & \left(|S(u,v)| \times \exp[\varphi_{s}(u,v)]\right) \times \left(|R(u,v)| \times \exp[-\varphi_{r}(u,v) + j(ud+vd)]\right) , \\ & + \left(|S(u,v)| \times \exp[-\varphi_{s}(u,v)]\right) \times \left(|R(u,v)| \times \exp[\varphi_{r}(u,v) - j(ud+vd)]\right) \end{aligned} \tag{8}$$

This equation contains only terms corresponding to the two correlation peaks. Then, by performing a FT^{-1} of this quantity only two correlation peaks are visible.

- *Fringe-Adjusted Joint Transform Correlation: FA-JTC [10]*

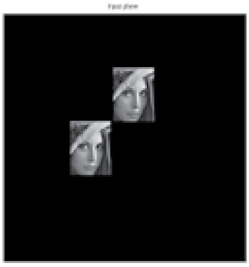
To optimize this JTC correlator and to render it more robust to noise, the authors of [10] suggested and validated a new approach of JTC, called Fringe-Adjusted Joint Transform Correlation FA-JTC. Basically, it consists in multiplying the intensity of the JTC joint spectrum (Eq 4) with the fringe-adjusted filter (Eq. 9)

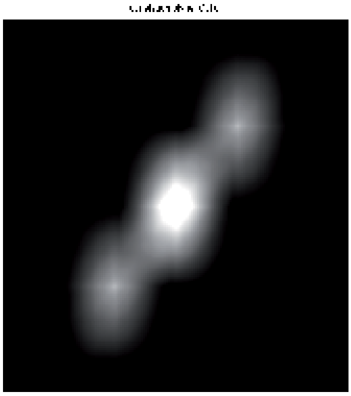
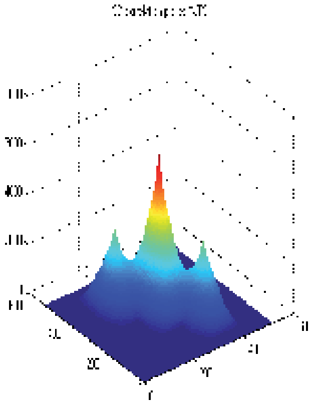
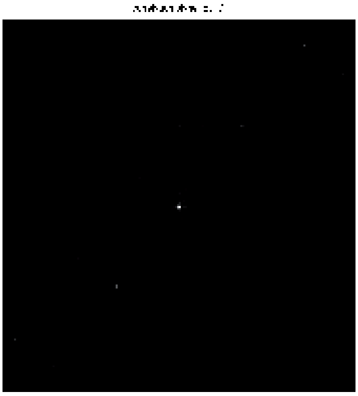
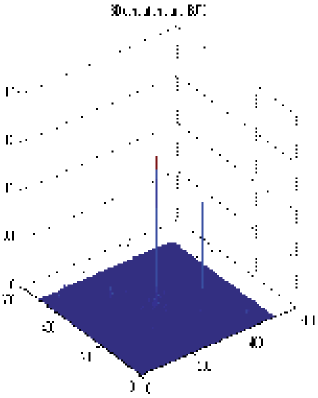
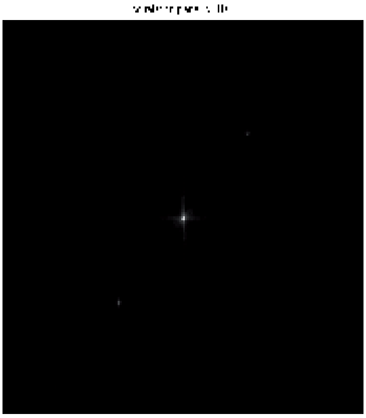
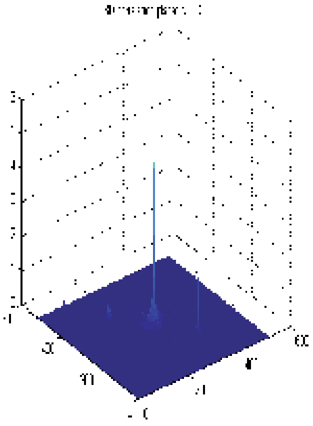
$$H(u,v) = \frac{G(u,v)}{N(u,v) + R(u,v)} \ , \tag{9}$$

where *G(u,v)* denotes a function to obtain an optical gain larger than 1 and *N(u,v)* is a function used to reduce the noise effect in the correlation peak and/or to delete the band limit the signal. A comparison of the FA-JTC method with the classical JTC and binary JTC shows that they are successful to increase the decision performance of the JTC correlator [10].

- *Comparison and conclusion*

In this section, we give an overview of the performance of several JTC correlators vis-à-vis the face recognition issue. The example chosen, Table (l-a), is to recognize the face of Lena with a reference face-image Lena. This study was conducted in a controlled environment: without noise.

Face to recognize / face reference (a)	 **Input plane**

Classical-JTC (b)	 Correlation plane: Classical JTC  3D correlation plane: Classical JTC
B-JTC (c)	 Correlation plane: B-JTC  3D correlation plane: B-JTC
NL-JTC (d)	 Correlation plane: NL-JTC  3D correlation plane: NL-JTC

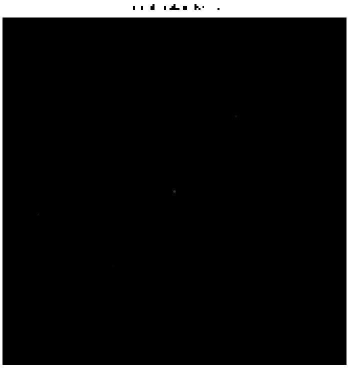
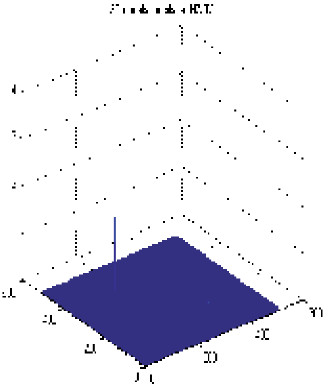
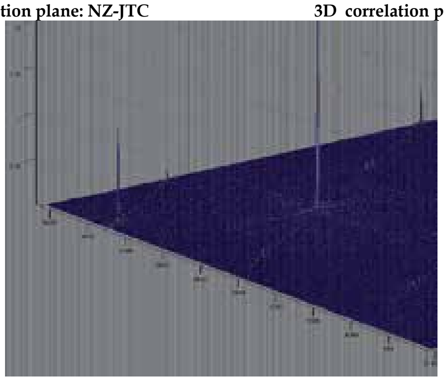
NZ-JTC (e)	 Correlation plane: NZ-JTC  3D correlation plane: NZ-JTC  Zoom of the 3D correlation plane

Table 1. 3D representation of the various correlation planes obtained with several approaches of the JTC correlator.

Table (1) presents information of the JTC correlator performances. The binary JTC correlator has very intense and sharp correlation peaks (Table (1-c)) compared to the values obtained with the classical JTC (Table 1-b). This results in a reliable and discriminating decision. The B-JTC correlators have good discriminating features, however they are too sensitive with respect to noise like for e.g. the deformation of the target face relative to the reference one, i.e. a small change between the reference face and the face to be recognized yields a significant decrease of the intensity of the correlation peaks. Thus, it is necessary to increase the number of reference images in the database to cover all possibilities that the target image can have in a given scene.

One solution to overcome this problem is realized by introducing a non-linearity in the Fourier plane. This allows us to control the performance of the JTC correlator (Table 1-d) and to increase or decrease its discriminating characteristic. By changing the non-linearity degree of the JTC correlator, different applications, with or without noise, can be considered. Suppression of the zero-th order can make easier the decision and increase the performance (Table 1-e).

Despite improvements found in the literature, the JTC architecture has still a serious drawback: it requires a large input *SBWP* [11]**.** The very nature of the JTC requires that the target and the reference (or references) images are placed side by side in the input plane (Figure 4). Assuming that N_s is the pixel size of the scene (that contains the target image with size=N_o) and N_r is the pixel size of the reference image, the pixel size of the input plane (Figure 4) N_e can be written as

$$N_e = N_s + 2 \times N_r = P \times N_r, \tag{10}$$

where *P* denotes the number of reference images to include in the input entry [11]. The use of many references in the input plane is necessary to overcome the problem of making a decision based on a single comparison between the target image with a single reference image. Thus, a decrease of the $SBWP\_input$ appears because of non-active zones to avoid overlapping of the correlation planes in the output plane of the JTC correlator [11]; the target and the reference images must be placed in the input plane at a distance *d*. Consequently, one can write $SBWP\_input = N_e^2$, indicating that $SBWP\_input$ increases when the size of the object is increased. For example if we want to recognize a face that has a size equal to (64×64) pixels, it is necessary to use an input plane with a size equal to (320×320) pixels. This plane has an unused area equals to $8 \times (64 \times 64)$ (with reference to Figure (4)) [11].

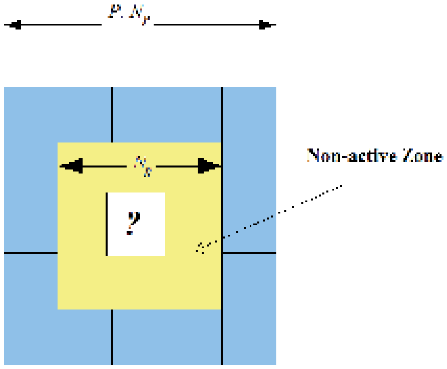


Fig. 4. Input plane of a JTC correlator using multiple references.

The implementation of an all-optical NL_JTC correlator is quite simple. It has a good compromise between discrimination and robustness; this compromise is mainly due to the use of non-linearity in the Fourier plane introduced via an optically addressed modulator (OASLM). However, the number of references that can be treated in a parallel manner is limited by the structure of the input plane ($SBWP\_input$), i.e. the target and references must be positioned in the input plane. In addition, the use of several references in the input plane can lead to saturation in the joint spectrum, when increasing the number of references [7,11].

To overcome this problem the Vander Lugt Correlator (**VLC**)-based technique can be used. This technique does not require a large input plane because only the scene is presented.

Thus, all the correlation features are introduced in the Fourier plane. This technique is now described.

2.4 Vander Lugt Correlator (VLC) and mono-channel approach

The synoptic diagram of the VLC is presented in Figure 5. Basically, this technique is based on the multiplication of the spectrum, S_0, of the target image O by a correlation filter, H, made from a reference image. The input plane is illuminated with a linearly polarized incident parallel beam. Located in the rear focal plane of a Fourier lens, a SLM is used in order to display the chosen correlation filter in the Fourier plane of the optical set-up. Many approaches for designing this filter can be found in the literature according to the specific object that needs to be recognized [1,3][13-19]. A second FT is then performed with a lens in a CCD Camera (correlation plane). This results in a more or less intense central correlation peak depending on the degree of similarity between the target object and the image reference.

Since the publication of the first optical correlator VLC [3], much research has been done to render it more discriminating and robust. The 1980s and 1990s have seen a large number of correlation filters to improve the Classical Matched filter (CMF) [3]. These improvements were intended to include physical constraints of the SLM modulators which are required to display filters in the optical set-up. A few words about the notation: $I(x,y)$ and $R_i(x,y)$ are the (2D) target image (to recognize) and the 2D reference image number #i, respectively. In like fashion, $S_I(u,v)$ and $S_R(u,v)$ denote the Fourier transform of the target image and the reference image, respectively. Subsequently, we will focus on presenting several filters which have been designed for 3D face recognition.

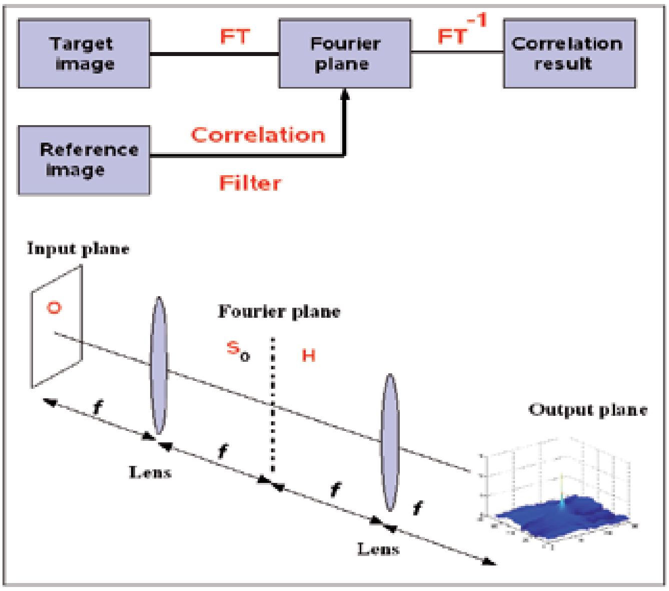


Fig. 5. Synoptic diagram of VLC.

- *Classical Matched filter: H_{CMF}*

This filler is defined as following:

$$H_{CMF}(u,v)=\frac{\alpha S^{*}_{R_i}(u,v)}{N(u,v)}, \tag{11}$$

where $S^{*}_{R_i}$ denotes the conjugate of the reference spectrum image, N is the spectral density of the background noise and α is a constant. This filter is very robust, but presents a low discriminating power [1,3,13-14]. Without noise this filter can be written as: $H_{CMF}=S^{*}_{R_i}(u,v)$.

- *Phase Only Filter: H_{POF}*

The POF filter is an optimized version of the classical matched filter (H_{CMF}) and is defined by equation (12). This filter has a very sharp peak, is very discriminating, but is too sensitive to specify noise arising e.g. in object deformation [13,20].

$$H_{POF} = \frac{S^{*}_{R_i}(u,v)}{\left|S^{*}_{R_i}(u,v)\right|}, \tag{12}$$

where $S^{*}_{R_i}(u,v)$ denotes the conjugate of the reference spectrum image and $\left|S^{*}_{R_i}(u,v)\right|$ is the module of the reference spectrum.

- *Binary Phase Only Filter: H_{BPOF}*

This filter is a binary version of the H_{POF} filter to take into account the physical constraints imposed by using fast and binary SLM [8,21]. Several techniques have been proposed to achieve this binarization, e.g. [21]. Here, this filter is binarized according to the following equation:

$$H_{BPOF}=\begin{cases}+1 & if \quad real\left(H_{POF}\right)\geq 0\\ -1 & if \quad real\left(H_{POF}\right)\prec 0\end{cases}. \tag{13}$$

Different types of binarization were considered, see e.g. Ref [22] for a review.

- *Phase Only Filter with region of support: H_{ROS_POF}*

The principle of this filter is very important for the definition of our multi-correlation segmented filter. It is defined as a multiplication of the H_{POF} filter with a pass-band function P [22-23] and can be written as following:

$$H_{ROS-POF} = H_{POF} P \quad , \tag{14}$$

where *P* is a pass-band function used to select the desired information in the H_{POF} filter. This H_{ROS_POF} filter permits to increase the robustness of the classical $H_{POF,}$ and has a good discriminating power.

The following set of filters was also proposed and discussed in the literature:

- $\mathbf{H_{AMPOF}}$: Amplitude Matched Phase only filter [24]
- $\mathbf{H_{CTMF}}$: Complex Ternary Matched Filter [25]
- $\mathbf{H_{FPF}}$: Fractional power filter [14]
- $\mathbf{H_{IF}}$: Inverse filter [14]
- $\mathbf{H_{PCMF}}$: Phase With Constrained Magnitude filter [26]
- $\mathbf{H_{PMF}}$: Phase mostly filter [27]
- $\mathbf{H_{QPF}}$: Quad-Phase filter [28-29]

All these filters have been proposed to optimize the performance of the decision taken with the classical correlation filter H_{CMF}. However, in every case, each filter uses a single reference. For a reliable decision, we must compare the target image with a large number of filters. These results in a huge increase of the time required to make a decision. In addition, a small part of the *SBWP* of the output plane is used (decision based on one point in the output plane: mono-correlation). To overcome these problems and improve the reliability of the decision, multi-correlation approaches constitute a useful option [8,30].

2.5 VLC and multi-correlation approaches

Multi-correlation consists in considering a larger part of the correlation plane [8,30] to cover all possible situations and have a discriminating power. Other words stated, we explore decision-making structures based on several correlations to overcome the drawbacks existing with a decision taken with a single reference. From a practical point of view, the use of several references can be realized with a temporal multiplexing of these references (Figure 6). For that purpose, we consider successively several (number of used references) correlations to make the decision. This obviously increases the computational time.

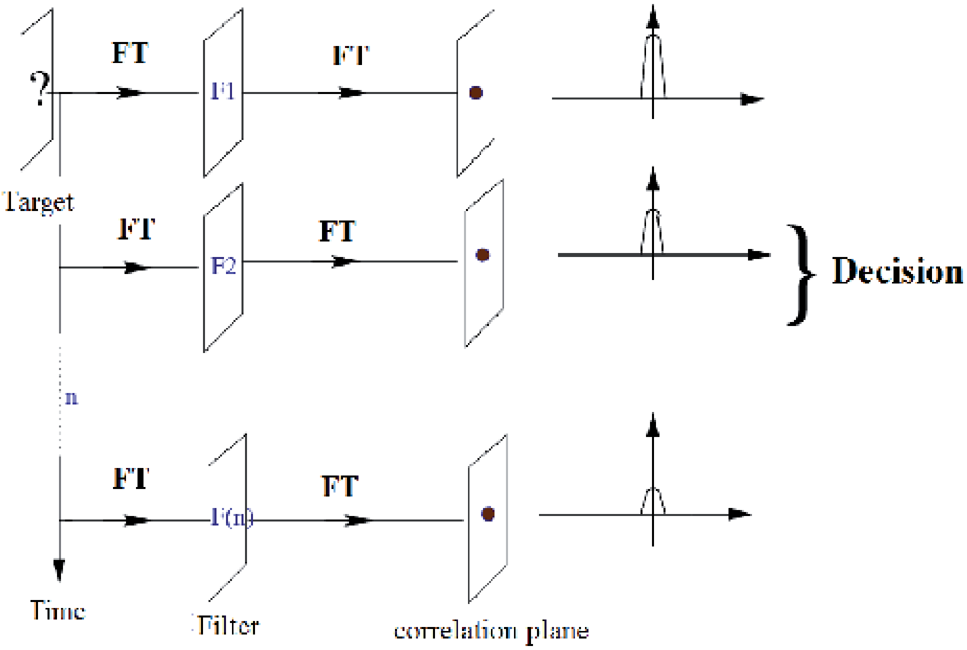


Fig. 6. Architecture using the temporal multiplexing of references.

Alfalou *et al*. [18-19] proposed and validated another solution that consists of spatially multiplexing these references together by combining their respective filters as shown in Figure (7). This solution performs simultaneously and independently many correlations. Then, the decision is taken by comparing all these correlations.

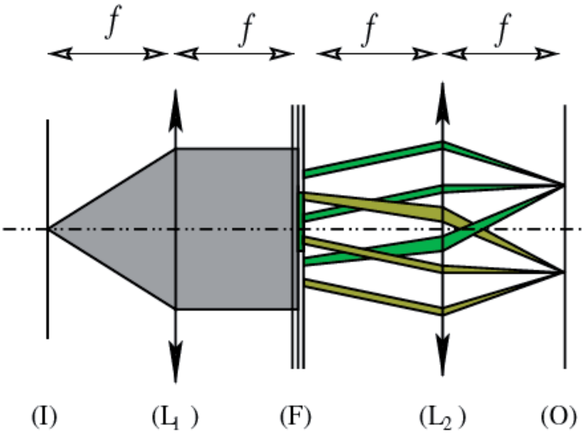


Fig. 7. Architecture using spatial multiplexing: principle of a correlator using a segmented composite filter for the multi-correlation.

Here, the filter that we seek to build is a filter allowing us to recognize one or several objects in a scene. To build such a filter, we have chosen to work with the composite filter approach for multi-correlation [18-19].

- *Composite Filter: H_{COMP}*

This filter allows one to merge multiple versions of a reference image. This type of filter is interesting to solve the correlation filter sensitivity vis-à-vis the rotation and the scaling operations. This filter can be defined as a linear combination of different reference versions:

$$H_{COM} = \sum_{i} a_i \, R_i \ , \tag{15}$$

where the a_is are coefficients to optimize a cost function defined for a desired application [16-17]. In the composite filter approach, each point of the Fourier plane is involved in the correlation of a given class according to the intensity of its spectrum at this point. However, this may be inconvenient if several classes use the same point in the Fourier plane.

- *Segmented Composite Filter: H_{seg}*

To get a reliable decision, it turned out to be necessary to perform a multi-decision and reconfigurable correlator, in order to display the different references rapidly. A first solution consists in adding a grating in the correlator in order to obtain several channels at the same time [8]. However this kind of architecture lacks programmability: this is due to the fact that the various channels (necessary to separate the several decisions) are implemented in the correlator by a fixed grating. After investigating programmability opportunities of the various correlators architectures, we opted for a phase only filter based correlator [3,20] for optical implementation simplicity. But, the programmability of such architectures turned out to be a serious issue. Consequently, the study was directed towards composite filters based architectures, where the programmability introduced into the Fourier plane is easier to implement [16-17]. The various references are merged together in the same filter and can be placed in the correlator using reconfigurable interfaces. A specific carrier is assigned to each reference for separating the correlation results.

However, this needs to cope with the issue of information encoding in the Fourier plane. To perform such encoding, various schemes of the Fourier plane were proposed. One of them is adapted to the composite filters for correlation and can be carried out using the concept of segmented composite filter [18-19]. This architecture is illustrated in Figure 7. It deals with a standard optical correlator using a spatial light modulator (SLM) with binary phase-like filters in the Fourier plane.

To improve the robustness of the H_{POF} filter, it was shown that the multiplication of the latest filter with a binary pass-band function ($P(i)=0$ $i<N$ *or* $P(i)=1$ *if* $i>N$, where N is chosen according to the specific application) increases the robustness of the POF filter against noise. To improve the pass-band function, Ding and co-workers [23] proposed an iterative method. This technique requires a large computing time and depends on the desired application. Moreover, it is based on a binary function (0 or 1) leading to a loss of Horner efficiency which can be detrimental for optical implementation. A similar technique was proposed in Refs. [18-19] to improve the pass-band function based on a non-iterative method and optimized by using the *SBWP* in the Fourier plane. This technique was shown to give a good Horner efficiency identical to a conventional H_{POF}.

This property was used to design our multi-correlation filter optimizing the use of the *SBWP* in the Fourier plane of the VLC correlator. We begin first by multiplying each filter (manufactured from one reference image) with a given pass-band function. Note that each pass-band function is chosen to eliminate all non-important areas in each filter. In these non-used areas information coming from another filter to obtain is included. Thus, a single filter containing information on several references images is used.

The significant increase in the processing capacity is due to the parallelism offered by the segmented composite filters, ensuring that an optimal use of the *SBWP* is available in the correlator. The parallelism provided by these filters offers two important advantages:

- First, it enables to increase the rate of correlations independently from the technology used for SLMs.
- Second, it offers a more efficient decision-making processes (the decision is made by the simultaneous consideration of several peaks).

- *Advantages of the segmented composite filter:*

In addition to parallelism, this filter allows one to obtain a better optimization of the *SBWP* available than conventional filters. With the composite filter, the phenomenon of local saturation in the Fourier plane is much more awkward than that met with the segmented filter, due to the fact that the manufacture of the composite filter is based on the local addition of spectral information coming from different references. On the other hand, the nature of the segmented composite filter, whose synoptic diagram is presented Figure (8), allows us to reduce the phenomenon of local saturation in Fourier plane. This result is obtained by segmenting the Fourier plane in several zones and by allotting a class to each zone. Separation at the output plane is obtained by adding a spatial carrier to each class.

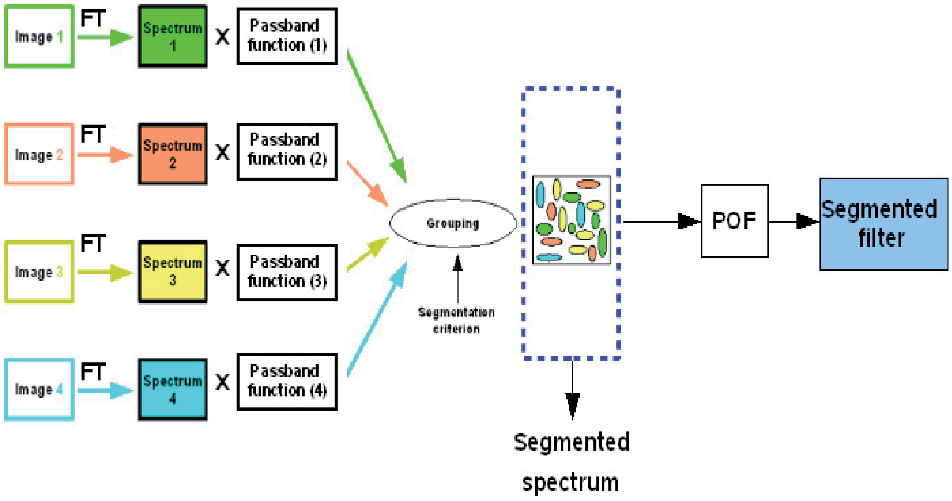


Fig. 8. Synoptic diagram of a correlator using a segmented composite filter for multi-correlation

The used criterion which was worked out in [18] is purely energetic and does not take into account the phase information. Here, the validation of such optimization of this criterion is illustrated. The decision to assign this pixel to a reference is based on the overall comparison:

$$\frac{E_{ij}^{k}\cos(\phi_{ij}^{k})}{\sum_{i=0}^{N}\sum_{j=0}^{N}E_{ij}^{k}} \geq \begin{cases} \dfrac{E_{ij}^{0}\cos(\phi_{ij}^{0})}{\sum_{i=0}^{N}\sum_{j=0}^{N}E_{ij}^{0}} \\ \dfrac{E_{ij}^{1}\cos(\phi_{ij}^{1})}{\sum_{i=0}^{N}\sum_{j=0}^{N}E_{ij}^{1}} \\ . \\ \dfrac{E_{ij}^{L-1}\cos(\phi_{ij}^{L-1})}{\sum_{i=0}^{N}\sum_{j=0}^{N}E_{ij}^{L-1}} \end{cases} , \qquad (16)$$

where L is the number of references, N is the size of the filter plane, and $E_{ij}(k,l)$ is the spectral intensity of the i-th class at the pixel (k,l). Figure (9) presents a segmented filter with a reference base made up of 4 classes (with three references per class, i.e. 9 references at all).

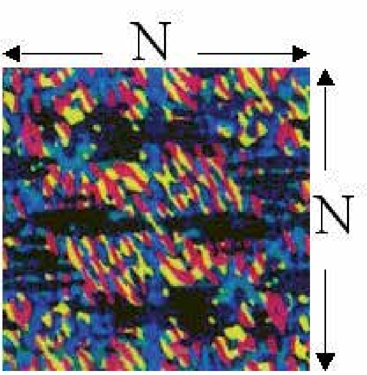


Fig. 9. Segmentation of each class in the Fourier plane

- *Criteria for assessing pattern-recognition*

Several criteria have been proposed to evaluate the correlation performance, i.e. recognition, discrimination [14, 31]. Here, two of them will be chosen for illustrative purpose. The first is the Peak-to-Correlation Energy (*PCE*) defined as the energy of the peak correlation normalized to the total energy of the correlation plane:

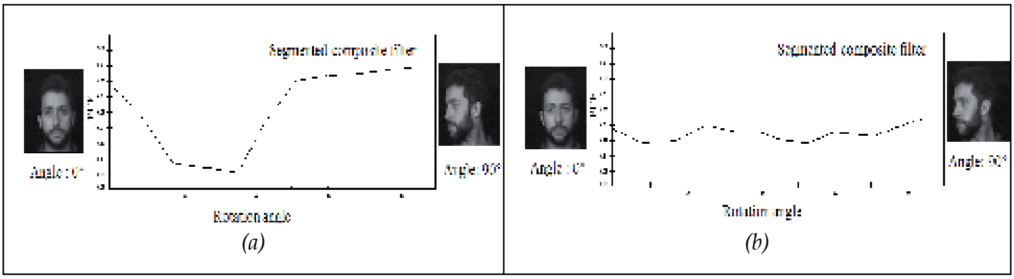
$$PCE = \frac{\sum_{i,j}^{N} E_{peak}(i,j)}{\sum_{i,j}^{M} E_{correlation\_plane}(i,j)}, \qquad (17)$$

where N denotes the size of the peak correlation spot, and M is the size of correlation plane. A second criterion is the SNR_{out}, defined as ratio of the energy of the peak correlation spot to the noise in the correlation plane:

$$SNR_{out} = \frac{\sum_{i,j}^{N} E_{peak}(i,j)}{\sum_{i,j}^{M} E_{correlation\_plane}(i,j) - \sum_{i,j}^{N} E_{peak}(i,j)} . \tag{18}$$

- *Performance of the segmented composite filter vis-à-vis the rotational invariance*

From a practical point of view, the Phase Only Segmented Composite Filter is a filter having a very good capacity of discrimination, but it is not robust. To illustrate such low robustness, we take a face onto which a rotation between 0° and 90° (0°, 10°, 20°, … 90°) is performed. The obtained faces are then compared with a single phase segmented composite filter made from two positions of faces: 0° and with a rotation equal to 90°. Figure (10-a) confirmed the low robustness of this kind of filter, i.e. the PCE is lowered by a factor of 4. To overcome this issue, we add a reference, a face rotated by 45° in the filter. The results, incorporating the same conditions as before, are presented in figure (10-b). These results demonstrate the increase of the robustness of the segmented filter by using only three references in each class (at 0 °, the second at 45° and a third to 90°).



(a) *(b)*

Fig. 10. Robustness of the segmented composite filter: (a) results obtained with a segmented filter used two references at 0° and 90°. (b) Results obtained with a segmented filter with three references at 0°, 45° and 90° [8]

- *Comparison between the classical composite filter and the segmented composite filter : application to face recognition*

Taking into account the limited space band-width product (SBWP), the quantity of information which can be introduced into the filter is rather limited. Consequently, the number of references which can be incorporated in the composite filter is low. Table (2) summarizes the PCE obtained numerically by introducing into the Fourier plane a single phase filter for a face recognition application. The evolution of the PCE versus the number of correlation is displayed in figure (11) (base consisted of 4 classes, with 4 references per class).

Correlation number	PCE : Composite filter	PCE : Segmented composite filter
1	0.256	0.323
2	0.163	0.309
3	0.088	0.181
4	0.066	0.141
5	0.062	0.155
6	0.033	0.138
7	0.033	0.121
8	0.029	0.121
9	0.028	0.113
10	0.029	0.099
11	0.016	0.090
12	0.012	0.090

Table 2. PCE values according various methods of calculation from filter considered

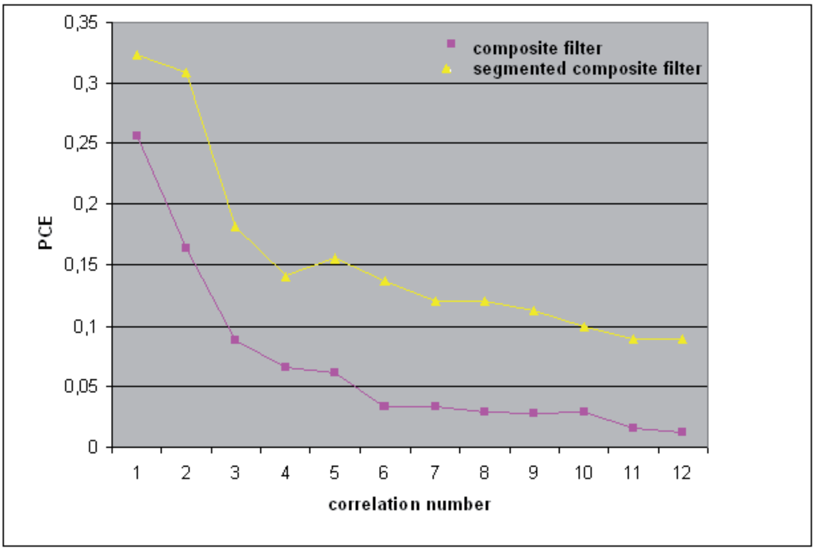


Fig. 11. PCE values obtained versus correlation number for two composite filters

In figure (11) the PCE decreases as a function of the number of correlations for different filters. It is also important to observe in figure (11), the quasi-impossibility of recognizing a face (with a classical composite filter) when more than 5 references are used. While the segmented composite filter makes it possible to incorporate 12 references the PCE loses only 1/3 of its initial value.

3. Multi-view segmented filter for multi-correlation: Application to 3D face recognition

In many practical applications, e.g. face recognition, 3D objects need to be recognized. In order to identify them, they should be first converted in 2D images. In this case, SLMs, used to display the input plane and the filters into a correlator are 2D. The 3D-2D conversion is generally achieved with a CCD camera resulting in a loss of information. To deal with this issue, we propose to use a basic correlator (VLC) and to merge information of both object shape and object depth in its Fourier plane. Two main methods for 3D object reconstruction can be found in the literature. The first is based on holography, i.e. the PSI technique [32-33]. Even though it can provide a reliable 3D object reconstruction, it is discarded here because it needs achieving complex transmittances using several specific reference waves. Furthermore, it requires the registration of these transmittances. The second method, called integral imaging technique (II) [34] is easy to achieve since it does not require complex amplitudes and acquisition of a large amount of information. Indeed, this technique is based on the registration of multi-views of a 3D target object. These different views are called elemental 2D images. They are taken using a matrix of fixed cameras (pickup grid). Each of these cameras captures a view of the 3D object (Figure (12)). Then, the elemental 2D images are used to rebuild the 3D object by a specific process [34].

3.1. Our approach to recognize 3D object

The first step in our approach consists in decomposing the 3D object into many elemental 2D images. For this purpose, we use a pickup grid, i.e. symmetric arrangement with three rows of cameras: A first row is associated to an angle of -15° view of the object, a second one to 0°, and the third one to +15°. In our arrangement, the cameras are placed at a distance equal to 3 cm each in length and width. To validate our approach, we used objects without noise: the subject is positioned in front of the camera placed at the center. To recognize the 3D object, we can correlate each 2D elemental image with its appropriate 2D correlation filter, thus giving an elemental decision. The final decision on the 3D object is taken by combining all these elemental decisions. However, it is a very long process that requires the *a priori* knowledge of the information on the object (especially on its orientation). To deal with this problem, we propose to merge the different elemental images together to produce a single image that contains all the necessary information on the various views of the 3D object. Next, we compare this image with only one filter. Two fusion techniques were used: the first one is to carry out the composite filter and the second one is to carry out the segmented filter. Next, we will detail the process to merge the input elemental images and the fabrication of the filter.

For the dual purpose of merging the information dealing with a 3D object and reducing it in a 2D image, we start by applying the technique used in the fabrication of composite filters. The 3D/2D image obtained is a combination of different elemental images. However, a major drawback of this technique lies in its saturation problem, especially when the number of elemental images is large, which eventually leads to a decrease of the performance of the correlator. This saturation is due to the fact that the technique we want to propose here should be optically implementable. SLMs are used to display the input images and filters that require 8-bit coding of information for the SEIKO modulator (amplitude modulator, with 256 gray levels) and 2-bit coding for the Displaytech modulator. The former modulator is used to display the target image in the input plane of our correlator. The latter modulator is used to display the various filters. To overcome the saturation, the number of merged elemental images is limited to three. A fusion technique based on the segmentation of the Fourier plane is employed [18-35].

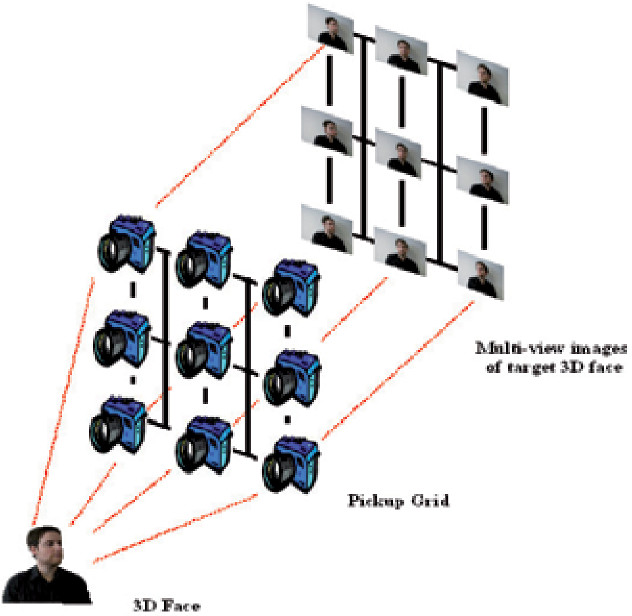


Fig. 12. Decomposition of the 3D object into several multi-view images.

3.2. Optimized multi-view segmented filter

After this short presentation of the principle of the segmented filter technique, an extension of this technique to 3D recognition is now proposed. For that purpose, we first consider a linear combination of different elemental 2D images (Figure 13-a) taken by one camera row in the pickup-grid (with reference to Figure 13-b). Next, each spectrum of merged elemental images is shifted (Figure 13-c). It was also necessary to shift the spectra to optimize the segmentation in the Fourier plane. This is required because the elemental 2D images are too close to each other. In practice, the shifting is realized optically by multiplying each merged elemental image by a specific phase term [18]. Afterwards, a segmentation to obtain the SMVSF is performed.

Particular attention was paid to find the appropriate value of the shifting parameter Δ that minimizes the overlap between the different spectra. For that purpose, we rely on previous work by Papoulis [36], who determined the necessary minimal size of a given spectrum by calculating the quantity called RMS duration

$$\Delta \;=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left|\nabla H(u,v)\right|^2 dudv=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x^2+y^2)\left|S_I(x,y)\right|^2 dxdy\,, \tag{19}$$

where $\nabla H(u,v)$ is the gradient of the spectrum, $S_I(u,v)$ denotes the spectrum of an image I, and $1/2\pi$ is a normalization factor.

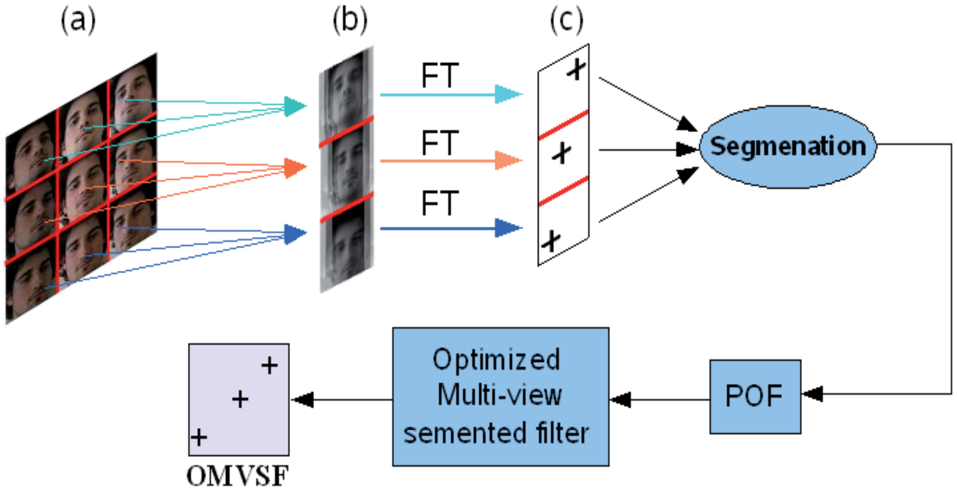


Fig. 13. Principle of the optimized multi-view segmented filter.

3.3. Proposed 3D correlator set-up

The principle of our approach is presented in Figure 14. The first operation is to add the different elemental target images. To realize this operation, we perform the same merging steps applied to fabricate our filter presented above. A multiplication with the SMVSF is followed by a FT^{-1} giving us the correlation plane. In the output plane, we get the result of simultaneous correlation of all elemental images obtained from the 3D face.

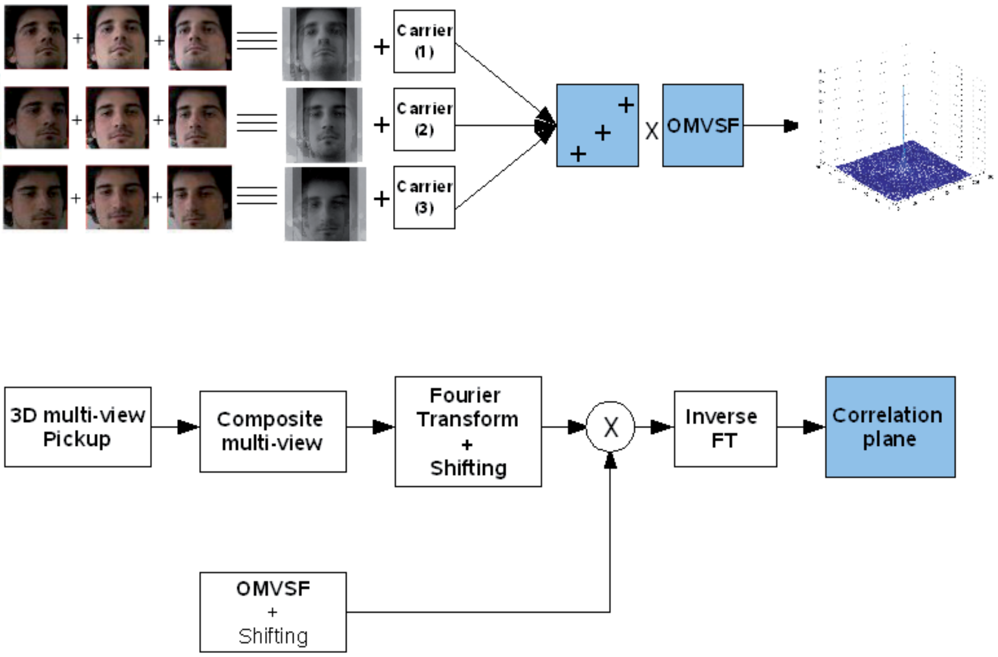


Fig. 14. Principle of the 3D Correlator set-up.

3.4. Results and Discussion

The purpose of this section is to validate the principle of our approach by numerical simulations. These simulations were conducted without noise, but taking into account the limitations imposed by using SLMs and their specific coding, i.e. 8 bit-coding of the input image and 2 bit-coding of the filter. We recall that the specific application we have in mind for these simulations is the recognition of a 3D face (with reference to Figure 15).

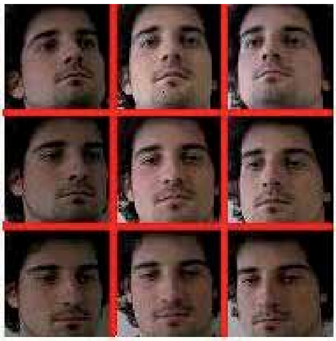


Fig. 15. Example of decomposition of a 3D face into multi-view 2D elemental images

If we apply our protocol for correlating the target image with a non-shifted multi-view segmented filter (Non-SMVSF), the correlation plane is presented in Figure 16(a). This figure shows the presence of a single correlation peak resulting from the autocorrelation between elemental input images with different corresponding elemental filters. It must be noted that the large width and the high level noise presented in the correlation plane (-38 dB) affects the performance of this filter. The segmentation of the different spectra is made from very similar spectra leading to strong spectral segmentation and consequently to the emergence of the isolated pixel phenomenon.

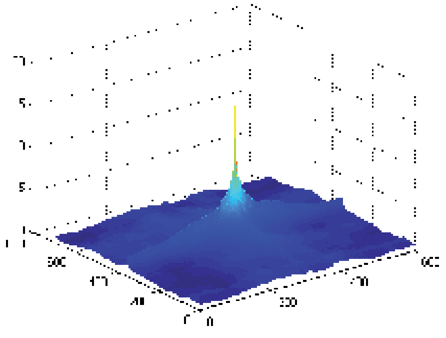
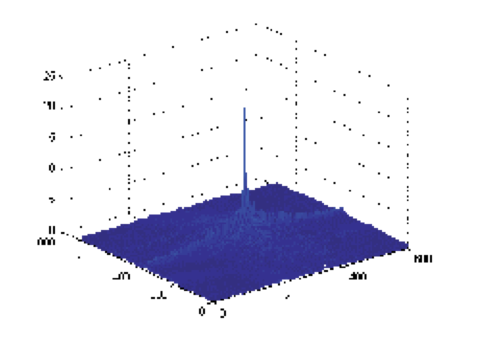
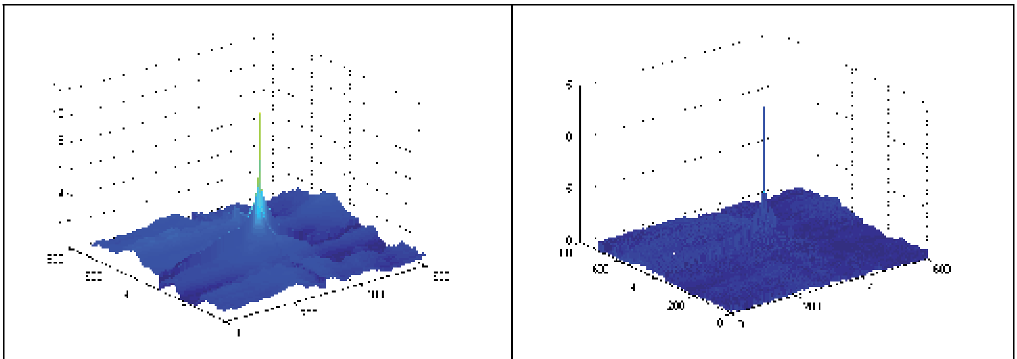
	
PCE = 0.00127 *SNR1*= -37.8 dB	*PCE* = 0.00221 *SNR1*= -32 dB
(a)	(b)

Fig. 16. Correlation plane (a) without spectra shifting, i.e. Non-SMVSF, and (b) with spectra shifting, i.e. SMVSF.

To overcome the problem of strong segmentation, we propose to optimize the use of the filter space *SBWP*. This is done by shifting the spectra to eliminate the high overlapping areas. The shifting value was calculated using the RMS duration criterion. Figure 16-b shows the correlation plane obtained with this optimization. It is seen that the correlation peak is higher and that the noise has been significantly reduced in the correlation plane (-32 dB). Form the implementation standpoint, it is necessary to binarize the filter. Figures (17-a) and (17-b) show the correlation result obtained by the binarization of two filters used in (Figure 16). The most compelling result is the significantly higher correlation peak observed in these figures.



PCE = 0.0051 *SNR1*= - 45.8 dB	*PCE*= 0.00221 *SNR1*= - 41 dB
(a)	(b)

Fig. 17. Correlation plane obtained by binarizing the (a) Non-Shifting MVSF filter, and (b) the SMVSF filter, respectively.

A new approach for the purpose of recognizing 3D objects has been presented and validated. This approach is based on the decomposition of the 3D target object and the 3D reference images by using the II method. Once the elemental images are obtained, they are merged together using composite and segmented techniques. To deal with the problem of isolated pixel the different spectra were shifted. This shifting has been calculated based on the RMS duration criterion. The work presented here allows one to correlate a 3D object with a single filter, to obtain a higher and sharper correlation peak, and to reduce the noise level in the correlation plane.

4. Conclusion

In this chapter, we introduced the principle of two optical correlation approaches which can be used to recognize faces: the JTC correlator and VLC correlator. Both are based on the use of the standard optical set-up called 4f. Although the optical implementation of the JTC correlator is easier to realize than the VLC correlator, the JTC requires the use of a very large input plane in order to introduce both the face to recognize and references. Consequently, we have chosen the VLC correlator to perform our face recognition application. This choice seems appropriate, especially when a new concept of correlation filter called segmented composite filter is employed. Having detailed this filter, we presented results showing good performances of this filter applied to face recognition. Moreover, to take into account the fact that a face is a 3D object, we proposed and validated an optimization of this segmented filter suitable for 3D face recognition. For future work, it would be interesting to test the robustness of this filter against noise and with respect to the modification of a 3D target object compared to 3D references images (rotation, scaling).

To have a broader view of face recognition problem and not to be limited only to the optical approaches, the reader may want to consult [37]. This reference presents a pure numerical approach for face recognition based on ICA (Independent Component Analysis). A comparison between these numerical and optical approaches is also provided in [37].

Finally, we point out the need of new research topics which can maximize the speed and improve the decision of the optical correlator for face recognition. Moreover, in some applications the correlator is not located in the same physical place as the target image to be recognized, thus requiring transmission and storage of the images before being processed. Therefore, it is necessary to develop appropriate techniques of image compression and encryption for recognizing a 3D face [38-39].

5. References

[1] Goodman, J. W. (1968). *Introduction to Fourier Optics*. McGraw-Hill, New York.

[2] Weaver C. S. & Goodman J. W. (1966). A Technique for Optically Convolving Two Functions. *Applied Optics,* 5, 1248-1249**.**

[3] Vander Lugt, A. (1964). Signal detection by complex spatial filtering. *IEEE Trans. Info. Theory,* IT-10, 139-145.

[4] Javidi, B. & Chung-Jung, K. (1988). Joint transform image correlation using a binary spatial light modulator at the Fourier plane. *Applied Optics*, 27, 4, 663-665.

[5] Javidi, B. (1989). Nonlinear joint power spectrum based optical correlation. *Applied Optics,* 28, 2358–2367.

[6] Guibert, L.; Keryer, G.; Servel, A.; Attia, M.; Mackenzie, H.; Pellat-Finet, P. & and de Bougrenet de la Tocnaye, J. L. (1995). On-board optical joint transform correlator for real-time road sign recognition. *Optical Engineering*, 34, 101–109.

[7] Keryer, G. (1996). *Etudes de corrélateurs optiques à corrélation jointe mono ou multicanaux : application à la reconnaissance des formes*. PhD Thesis, Université de Paris XI, Paris-France.

[8] Alfalou, A. (1999). *Implementation of Optical Multichannel Correlators: Application to Pattern Recognition*. PhD Thesis, Université de Rennes 1–ENST Bretagne, Rennes-France.

[9] Li, C. T.; Yin, S. & and Yu, F. T. S. (1998). Nonzero-order joint transform correlator. *Optical Engineering,* 37, 58–65.

[10] Alam, M. S. & Karim, M. A. (1993). Fringe-adjusted joint transform correlation. *Applied Optics,* 32, 4344-4350.

[11] Keryer, G; de Bougrenet de la Tocnaye, J. L. & Al Falou, A. (1997). Performance comparison of ferroelectric liquid-crystal-technology-based coherent optical multichannel correlators. *Applied Optics,* 36, 3043-3055.

[12] Horner, J. L. & Gianino, P.D. (1984). Phase-only matched filtering. *Applied Optics*, 23, 812-816.

[13] Javidi, B.; Odeh, S.F. & Chen, Y. F. (1988). Rotation and scale sensitivities of the binary phase-only filter. *Applied Optics,* 65, 233-238.

[14] Vijaya Kumar, B. V. K. & Hassebrook, L. (1990). Performance measures for correlation filters. *Applied Optics*, 29, 2997-3006.

[15] Horner, J. L. (1992). Metrics for assessing pattern-recognition performance. *Applied Optics,* 31, 165-166.

[16] de Bougrenet de la Tocnaye, J. L.; Quémener, E. & Pétillot, Y. (1997). Composite versus multichannel binary phase-only filtering. *Applied Optics*, 36, 6646-6653.

[17] Kumar, B. V. K. V. (1992). Tutorial survey of composite filter designs for optical correlators. *Applied Optics*, 31, 4773-4801.

[18] Alfalou, A.; Keryer, G. & de Bougrenet de la Tocnaye, J. L. (1999). Optical implementation of segmented composite filtering. *Applied Optics,* 38, 6129-6136.

[19] Alfalou, A.; Elbouz, M. & Hamam, H. (2005). Segmented phase-only filter binarized with a new error diffusion approach. *J. Opt. A: Pure Applied Optics*, 7, 183-191.

[20] Horner J. L. & Gianino, P. D. (1984). Phase-only matched filtering. *Applied Optics*, 23, 812-816.

[21] Horner, J. L.; Javidi, B. & Wang, J. (1992). Analysis of the binary phase-only filter. *Optics Communications,* 91, 189-192.

[22] Kumar B. V. K. V. (1994). Partial information Filters. *Digital Signal Processing-New York*, 4, 147-153.

[23] Ding, J.; Itoh, J. & Yatagai, T. (1995). Design of optimal phase-only filters by direct iterative search. *Optics Communications,* 118, 90-101.

[24] Awwal, A. A. S.; Karim, M.A. & Jahan, S.R. (1990). Improved correlation discrimination using an amplitude-modulated phase-only filter. *Applied Optics,* 29, 233-236.

[25] Dickey, F. M.; Kumar B. V. K. V.; Romero L.A. & Connelly J.M. (1990). Complex ternary matched filters yielding high signal-to-noise ratios. *Optical engineering*, 29, 9, 994-1001. ISSN 0091-3286.

[26] Kaura, M. A. & Rhodes, W. T. (1990). Optical correlator performance using a phase-with-constrained-magnitude complex spatial filter. *Applied Optics*, 29, 2587-2593.

[27] Richard, D. Juday, R. D. (1989). Correlation with a spatial light modulator having phase and amplitude cross coupling. *Applied Optics,* 28, 4865-4869.

[28] Dickey, F.M. & Hansche, B. D. (1989). Quad-phase correlation filters for pattern recognition. *Applied Optics,* 28, 1611-1613.

[29] Hansche, B. D.; Mason, J. J. & Dickey, F. M. (1989). Quad-phase-only filter implementation. *Applied Optics,* 28, 4840-4844.

[30] Mendlovic D. & Kirysche V. (1995). Two-channel computer-generated hologram and its application for optical correlation. *Optics Communications*, 116, 322-325.

[31] Horner, J.L. (1992). Metrics for assessing pattern-recognition performance. *Applied Optics,* 31, 165-166.

[32] Darakis, E. & Soraghan, J. J. (2007). Reconstruction domain compression of phase-shifting digital holograms. *Applied Optics*, 46, 351-356.

[33] Naughton, T. J.; Frauel, Y.; Javidi, B. & Tajahuerce, E. (2002). Compression of digital holograms for three-dimensional object reconstruction and recognition. *Applied Optics,* 41, 4124- 4132.

[34] Tavakoli, B.; Daneshpanah, M.; Javidi, B. & Watson,E. (2007). Performance of 3D integral imaging with position uncertainty. *Optics Express,* 15, 11889-11902.

[35] Farhat, M.; Alfalou, A.; Hamam, H. & Brosseau, C. (2009). Double fusion filtering based multi-view face recognition. *Optics Communications*, 282, 2136-2142.

[36] Papoulis, A. (1962). *The Fourier Integral and its Applications.* McGrawHill, New York.

[37] Alfalou, A.; Farhat, M. & Mansour, A. (2008). Independent Component Analysis Based Approach to Biometric Recognition. *Information and Communication Technologies: From Theory to Applications, 2008.* 7-11 April 2008 Page(s): 1-6. Digital Object Identifier 10.1109/ICTTA.2008.4530111.

[38] Alfalou, A. & Mansour, A. (2009). A new double random phase encryption scheme to multiplex & simultaneous encode multiple images. *Applied optics*, 48, 31, 5933-5946.

[39] Alfalou, A & Brosseau, C. (2009). Image Optical Compression and Encryption Methods. *OSA: Advances in Optics and Photonics*, 1, 589-636.

20

Parallel Face Recognition Processing using Neocognitron Neural Network and GPU with CUDA High Performance Architecture

Gustavo Poli and José Hiroki Saito
Libertas Integrated Schools
Federal University of São Carlos

1. Introduction

This chapter presents an implementation of the Neocognitron Neural Network, using a high performance computing architecture based on GPU (Graphics Processing Unit). Neocognitron is an artificial neural network, proposed by Fukushima and collaborators, constituted of several hierarchical stages of neuron layers, organized in two-dimensional matrices called cellular plains. For the high performance computation of Face Recognition application using Neocognitron it was used CUDA (Compute Unified Device Architecture) as API (Application Programming Interface) between the CPU and the GPU, from GeForce 8800 GTX of NVIDIA Company, with 128 ALU's. As face image databases it was used a face database created at UFSCar (Federal University of São Carlos), and the CMU-PIE (Carnegie Melon University - Pose, Illumination, and Expression) database. The load balancing through the parallel processing architecture was obtained by means of the distributed processing of the cellular connections as threads organized in blocks, following the CUDA philosophy of development. The results showed the viability of this type of device as a massively parallel data processing tool, and that smaller the granularity of the parallel processing, and the independence of the processing, better is its performance.

2. Motivation

The face recognition using machines is an active and actual research area. This is composed by multiple disciplines as image processing, pattern recognition, computer vision, artificial neural networks, and computer architectures. There are many commercial applications that implement Face Recognition Techniques, as in access control, and security using video camera.

Many countries use the face recognition techniques for several purposes. On China, for example, it was developed the immigrant recognition system to the cities of Shenzhen and Zhunhai (Terra, 2006). The system gives good results, especially when the fingerprint can´t be used to the recognition purpose, due to several problems as age, damage with chemical reactions, and so on.

Owing to the user-friendly (non-intrusive) property, the face recognition is attractive, despite of the extremely reliable methods of personal biometric identification such as fingerprint and iris scanning analysis.

As it can be seen there are major challenges on the issues of facial recognition, where you can highlight a relationship between two basics variables of the process: the degree of reliability/robustness of the technique being used and computational cost of this technique.

The goal of this chapter is the presentation of a computer architecture for face recognition, aiming its performance increasing through the use of a massively parallel data processing, achieved by the implementation of a Neocognitron neural network architecture, based on GPU (Graphic Processing Unit). To access the GPU as a device for scheduling purposes, it is used in majority the CUDA (Compute Unified Device Architecture), a library that extends the functions of language C, FORTRAN and Python in order to provide the GPU as a device for data processing.

3. Neocognitron Neural Network

Neocognitron is a massively parallel neural network, composed by serveral layers of neuron cells, proposed by Fukushima (Fukushima end Miyake, 1982)(Fukushima and Wake 1992)(Saito and Fukushima, 1998). In a brainway computer it corresponds to part of the human visual recognition system.

The Neocognitron neural network has the basic principle of operation extracting features in a hierarchical manner, i.e., performs the extraction of features in various stages. In the first stage, the extracted features are the simplest; and at the following stages, summing up the lines in different senses of rotation, the features will be presenting with more complexity. The characteristic of this network is that the features extracted by a stage have the informations only sent by the previous stage, as a feedforward neural network.

3.1 The Neocognitron Structure

The stages of a Neocognitron network are arranged in tiers, each of these layers has its own type/complexity of data being processed, and these consist of simple cells (Cell-S), complex cells (Cell-C) and activity cells (Cell-V).

The stages are compared by the Layer-S, of Cell-Ss. The Layer-C, of Cell-Cs; and Layer-V, of Cell-Vs. Within each layer there is a number of cell-plans, which are organized as two-dimensional array of cells, each cell with the ability of extracting the same features of the adjacent cells in the same cell-plan.

The stages function as a tool for organizing the process of extracting the characteristics or factors with a degree of complexity of the extracted pattern characteristics. The first stage, called the zero stage (Stage 0) is not used within the hierarchical scheme of feature extraction and it is used as the retina of the eye, capturing the pattern to be processed by the network. Figure 1 shows the stages of a Neocognitron with five stages.

The number of stages of a Neocognitron network depends on the size of the input pattern being processed by the network. The larger the size of the input pattern, greater is the number of stages required by the network. For example, an input pattern of 20 x 20 pixels, typically results in a network of three hierarchical stages.

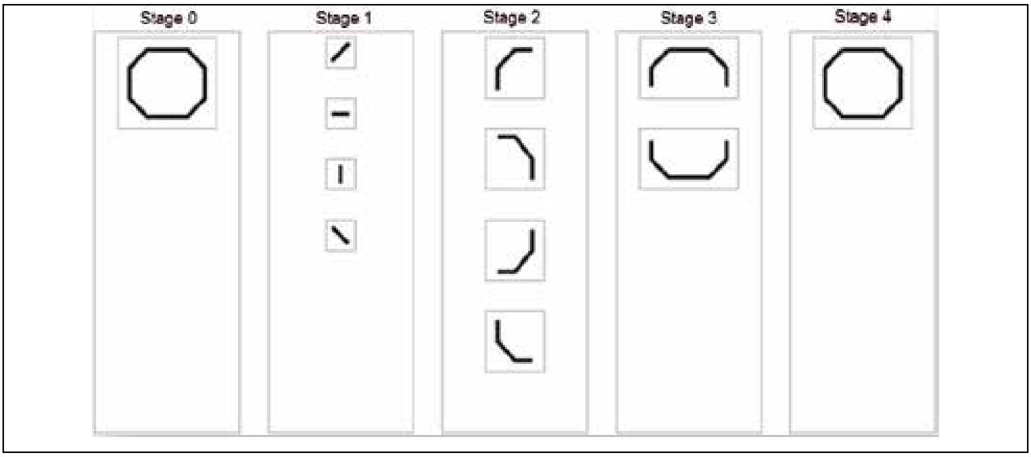


Fig. 1. Neocognitron representation with five stages.

Each stage of a Neocognitron network is divided into three layers: a simple layer (Layer-S), a complex layer (Layer-C) and a layer of activity (Layer-V). Assuming the Neocognitron with five stages, shown earlier (Figure 1), its representation in layers can be seen in Figure 2.

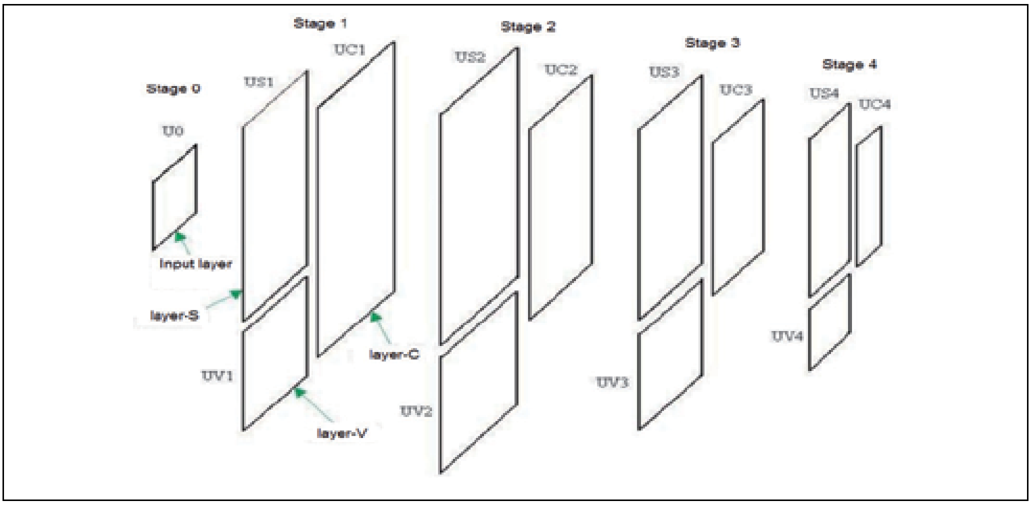


Fig. 2. Neocognitron representation with five stages with its layers.

In stage 0 there is only one layer, which is the input layer or input pattern. All other stages have three types of layers, one Layer-S, a Layer-V and a Layer-C.

Each layer is formed by a number of cellplans. The number of plans in each Layer-S and Layer-C is related to the number of features extracted by the stage of the network. A Layer-V is a single cell-plan layer. The size of the plans is equal to the same layer and it decreases as you climb the hierarchy of stages. Figure 3 shows the plans distributed in the cell layers of the network.

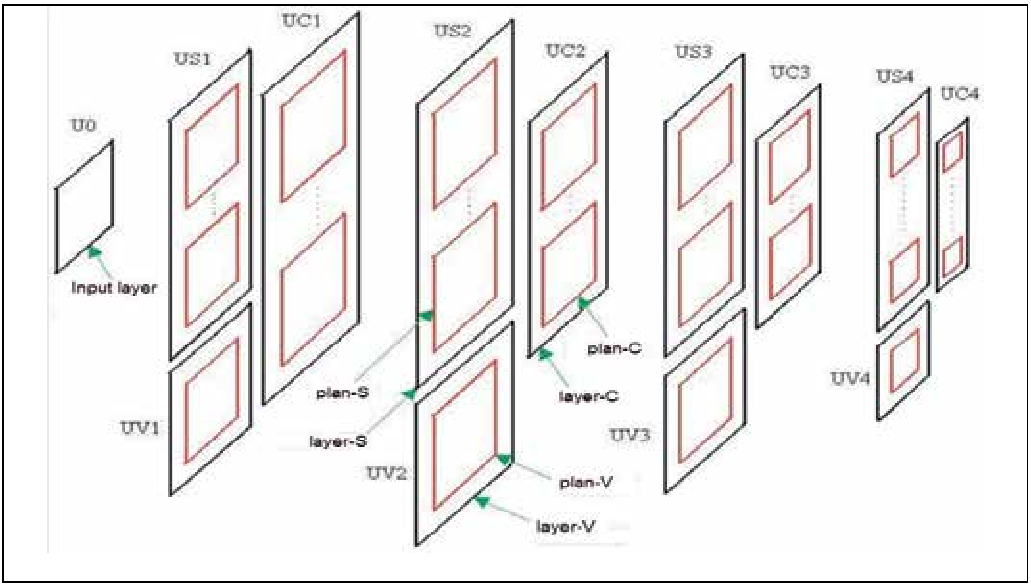


Fig. 3. Five stages Neocognitron representation with its layers and plans.

Each Plan-S, Plan-V, Plan-C and Input layer is formed by a set (array) of specialized cells. Figure 4 shows the cells distributed along the plans of the network. A Plan-C of the layer U_{C4}, the last stage of the network, contains only a single cell, whose activity indicates the recognition of the input pattern.

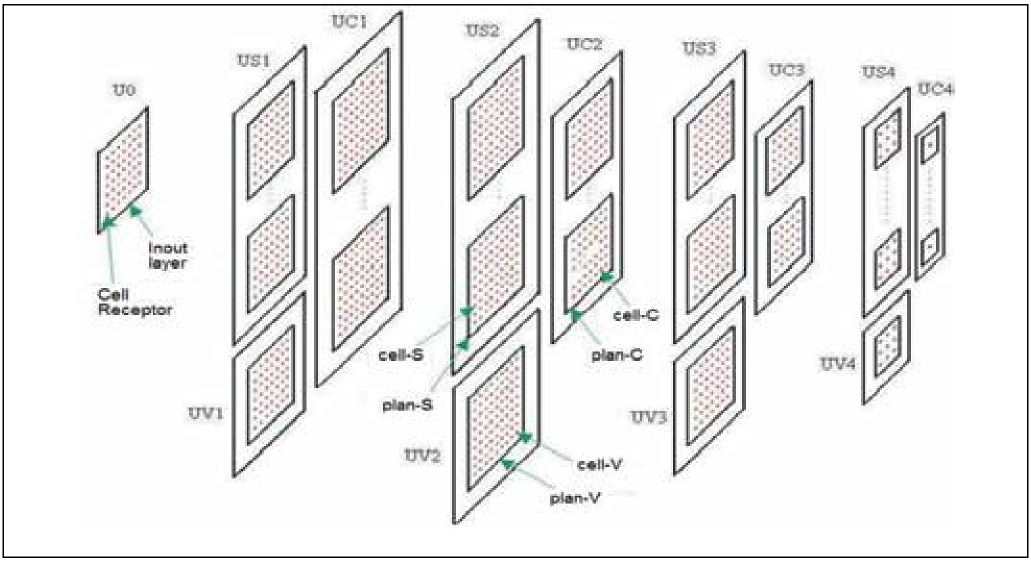


Fig. 4. Five stages Neocognitron representation with its layers, plans, and cells.

3.2 Weights and Connections

A characteristic of the Neocognitron is to have a large number of cells but a reduced number of connections. The cells are connected to a reduced connection area, of the previous layer. This characteristic of connectivity is different from the Multilayer Perceptron, in which a neuron of a layer is connected to all neurons of the previous layer.

For each connection there is a weight that is used to influence the amount of information that is transferred. Neocognitron has four types of weights: weight-a, weight-b, weight-c, and weight-d, whose uses are summarized as shown in Figure 5.

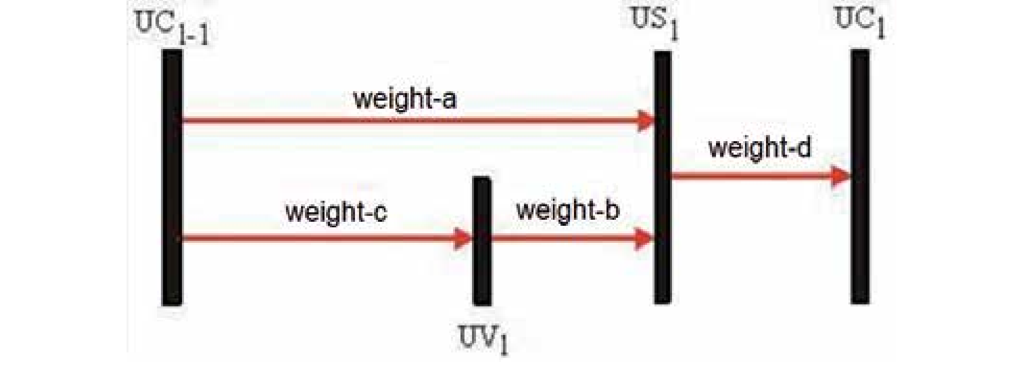


Fig. 5. Neocognitron weights (weight-a, weight-b, weight-c and weight-d) and its connections.

Within a cell-plan level, all cells share the same weight. This causes all cells in the same plan to observe the same feature, thus specializing the plan for the same feature in different positions.

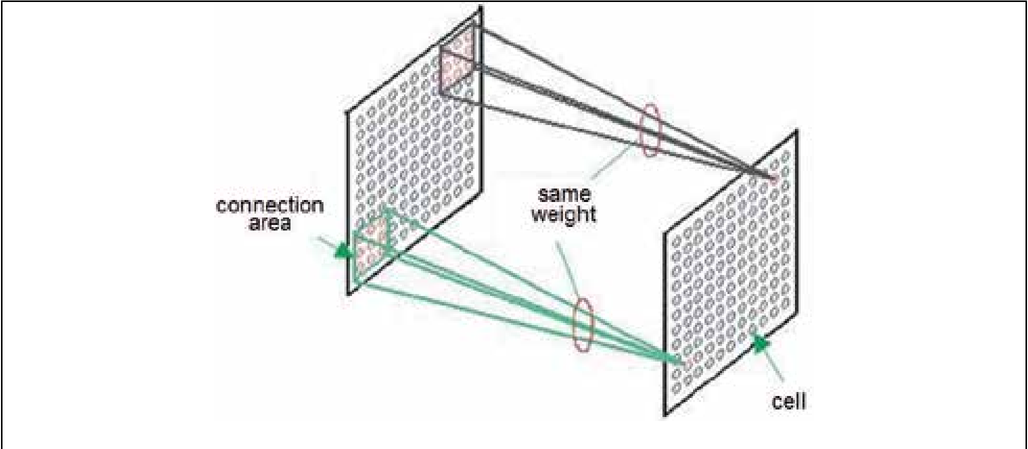


Fig. 6. Within a cellular level, all cells share the same weights.

You can even arrange the weights in two categories, which are modified by training (weight-a weight-b); and which are not modified, i.e., the values attributed to them, remain unchanged through the implementation of the network (weight-c and weight-d).

3.3 Processing Neocognitron Network

Each Cell-V calculates the input values of Cell-Cs from a small region connection area of all cell-plans of the previous Layer-C. The size of the connection area is the same for cells-V and cells-S in a stage of the network and it is determined at the time of construction of the network. An example of the connection area can be seen in Figure 7.

The value of a Cell-V represents the average activity of cells belonging to its area of connection and is used to inhibit the corresponding Cell-S. The exact specification of the function of Cell-V, $u_{Vl}(n)$, is given by Equation 1:

$$u_{vl}(n) = \sqrt{\sum_{k_{l-1}=1}^{K_l} \sum_{i \in S_l} C_l(i), u^2_{C_{l-1}}(n+i, k_{l-1})} \qquad (1)$$

where the weight should be $c_l \geq 0$; and u_{cl}-1(n+i, $k_{l\text{-}1}$) represents the input value, from the previous cell-plan $k_{l\text{-}1}$ at the position n+i. Here, i represents a position in a region S_l in a cell-plan.

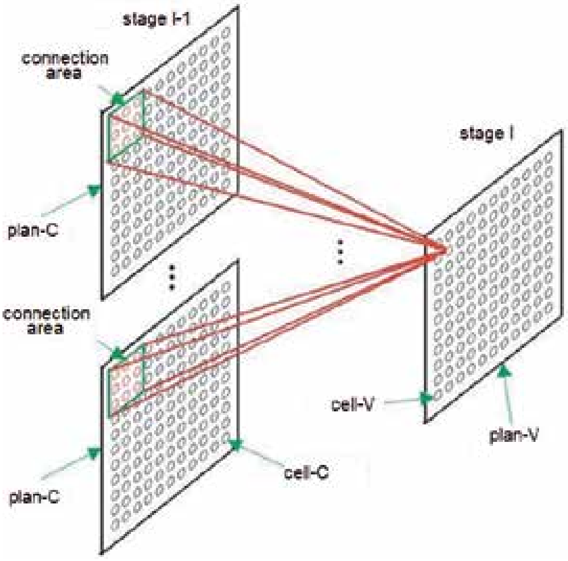


Fig. 7. Example of the connection area of a Cell-V.

The Cell-S evaluates the output values of cell-Cs in a connection area of all cell-plan of the Layer-C of the previous stage, or the input layer. As seen in the previous section, the size of the connection area is the same for cell-Ss and cell-Vs on the same stage (Figure 8).

The role of Cell-S is to recognize a feature in the connected area. To recognize a feature, a Cell-S uses the information in the connection area and information about activities in this area, informed by Cell-V. The feature extracted by a Cell-S is determined by weights on their input connections.

The feature extraction by a plan-S and the significance of the weights is easier to be observed in the cell layer U_o (first layer) of the network. In each cell of layer-S, U_{S1}, following the first layer, there is only one connection area and this area is the receptive field or area of connection of input pattern. Because all cells are equal, any cell in the same cell-plan can recognize the same feature. In the example, the feature is a vertical line that can be in different positions. So, the Cell-S, that is positioned in the connection area containing the feature (vertical line), responds, as outlined in the Plan-S in Figure 9.

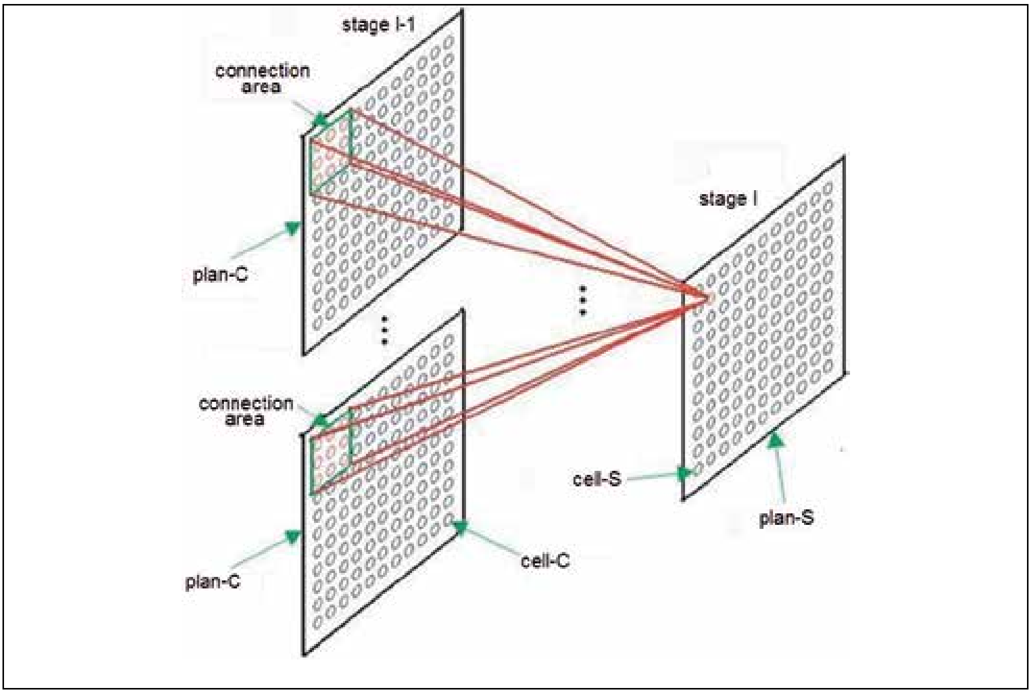


Fig. 8. Example of the connection area of a Cell-S.

The output value of a Cell-S is determined by Equation 2:

$$u_{Sl}(n,k_l) = \frac{\theta}{1-\theta} \cdot \varphi \left[\frac{1+\sum_{k_{l-1}=1}^{K_{l-1}} \sum_{i \in S_l} a_l(k_{l-1},i,k_l), u_{cl-1}(n+i,k_{l-1})}{1+\theta, b_l(k_l), u_{vl}(n)} - 1 \right] \tag{2}$$

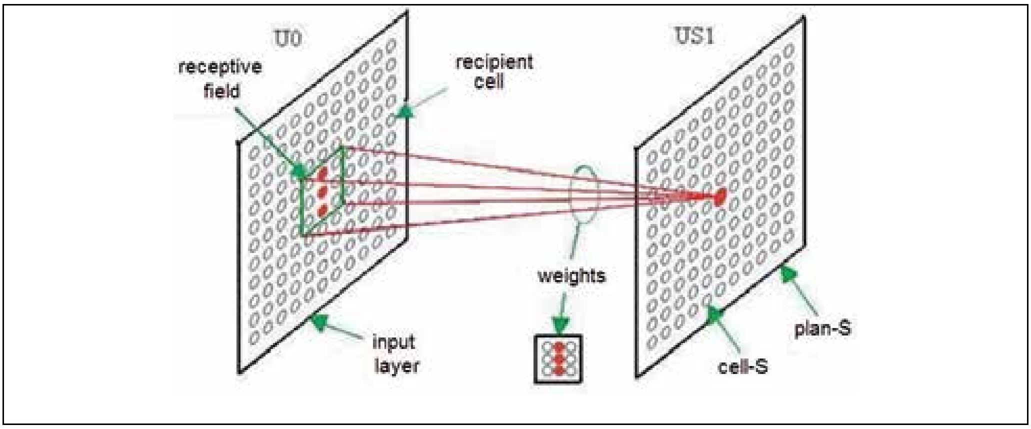


Fig. 9. Example of the connection area of a Cell-S.

The element θ is the threshold parameter with which you can modify the ability for Cell-S extract a particular feature. The weight-a, $a_l(k_{l-1},i,k_l)$, should be greater than or equal to zero, as well as weight-b, $b_l(k_l)$, and the activation function $\varphi[x] = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases}$.

The Cell-Ss have the ability to extract features not only trained but also distorted, or generalized. This capacity is influenced by the choice of parameter θ, called threshold. It is easy to understand, because the threshold θ multiplies the weighted value coming from the Cell-V, the denominator of the argument. Thus, the lower the value of θ, greater the ability of generalization of trained features.

The cell-C evaluates the output values of plan-S of earlier layer-S (Figure 10). The value of Cell-C depends on the activity of Cell-Ss in its area of connection. The greater the number of active Cell-Ss, greater is the activity of Cell-C. The equation of the Cell-C is described by Equation 3.

$$u_{Cl}(n,k_l) = \psi\left[\sum_{i\in S_l} d_l(i), u_{Sl}(n+i,k_l)\right] \tag{3}$$

As the weight-d, $d_l(i) \geq 0$ and $\psi[x] = \begin{cases} \frac{x}{1+x} & x \geq 0 \\ 0 & x < 0 \end{cases}$.

If a cell-C is active for a single cell-S, all the adjacent cells will be active, so that plan-C contains a blurred representation of the Plan-S. Moreover, as the blurring results in the cell-plan's adjacent values are very close, a small number of Cell-Cs is necessary for the next stage. This results in reducing the size of the Plan-C, in relation to the Plan-S, Figure 11.

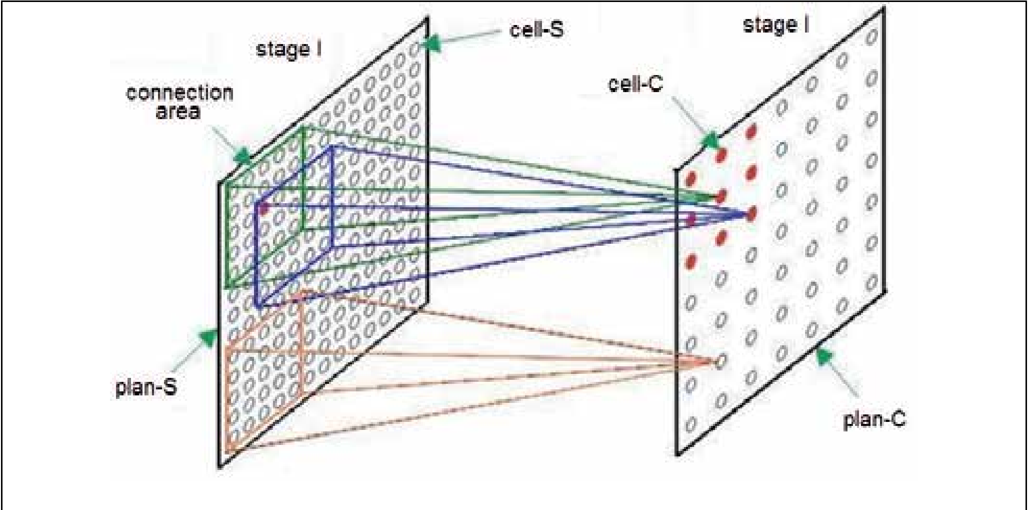


Fig. 10. Example of the connection area of a Cell-C.

3.4 Network Training

Although there are two main training methods for the Neocognitron network, it is described here the method originally designed, which is learning without supervision.

At first, the training follows as the majority of neural networks, i.e., it is showed a sample pattern, and data are propagated through the network, allowing the weights of the connections to fit progressively according to a given algorithm. After the weights are updated, the network receives a second pattern in the input layer, and the process repeats with all the training samples until the network classifies the patterns correctly.

Neocognitron network has the characteristic that all the cells in the same cell-plan share the same set of weights. Therefore, only a single cell of each plan must participate in training, and after that, distribute the whole weight to the other cells.

To better understand the operation, one can imagine all plans of a Layer-S stacked on each other, aligned so that the cells corresponding to a given location is directly above each other. Thus, it is possible to imagine several columns, cutting perpendicularly the planes. These columns create groups of cells-S, where all group members have receptive fields in the same location in the input layer.

With this model in mind, we can now apply a standard input and examine the response of Cell-Ss in each column. To ensure that each Cell-S provides a distinct response, one may start a_l weights with random small positive value and the weights b_l inhibitors with zero. First, note the plane and the position of the Cell-S whose response is the strongest in each column. Then it examines the plans individually so that if a plan has two or more of these Cell-Ss, it chooses only the Cell-S with the stronger response, subject to the condition that each cell is in a different column-S.

These Cell-Ss become the prototypes or representatives of all the cells in the respective plan. Once chosen the representatives, the updates of the weights are made in accordance with the Equation 4 and Equation 5, and all the cells of the same plan will be updated to be with the same weights:

$$\Delta a_l(k_{l-1}, v, k_l) = q_l c_l(u) u_{cl-1}(k_{l-1}, n+v) \tag{4}$$

$$\Delta b_l(k_l) = q_l v_{cl}(n) \tag{5}$$

Once the cells are updated to respond to a particular characteristic, they begin to emit responses smaller in relation to other features.

4. GPU as a Device to Generic Processing

Over the past 10 years, hitherto, it has seen the evolution of the GPU's as specialized hardware to process graphics and video output, and massive parallel processing of data for general computing. The power of data processing of GPU's has grown much faster than the CPU, and the main reason for this rapid growth of GPU's with respect to the CPU is due to the fact that the GPU's were born with the focus of intensive computing, with respect to data processing and massive parallel computing, as just the minimum requirements necessary to meet the needs of the scenario of computer graphics, like rendering, shadows in 3D scenes and others.

Thus the design of the GPU takes into account the existence of more transistors dedicated to a better process control and data flow, as illustrated schematically in Figure 12, which depicts the main elements: ALU, cache, and DRAM control for a CPU (Figure 12a) and a GPU (Figure 12b).

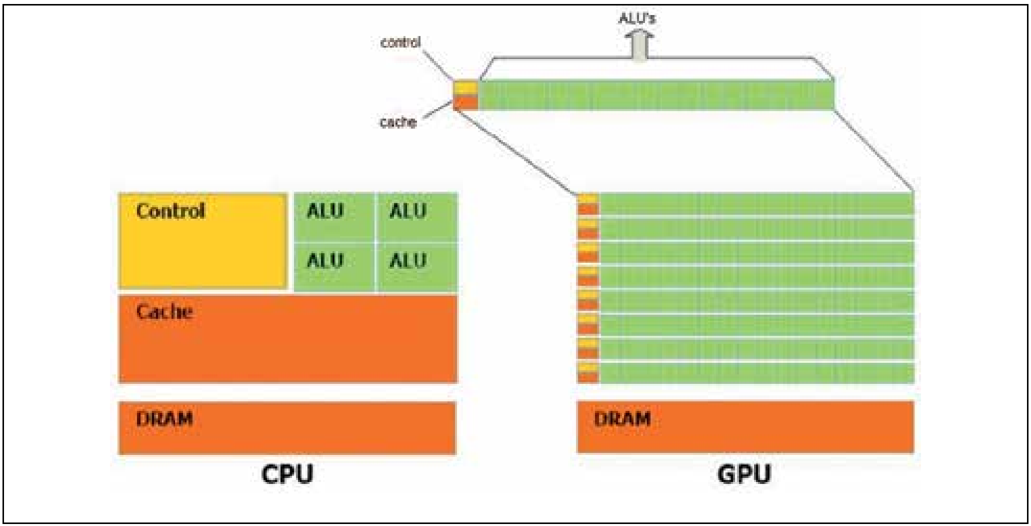


Fig. 11. GPU intended to use more transistors for Data Processing.

Many applications that process large data sets organized in a matrix/vector can use a model of parallel computing. In 3D rendering processes large arrays of pixels and vertices are organized so that they can be processed in parallel using threads. Similarly, applications of image processing, encoding and decoding, video scaling, stereo vision, artificial neural networks and pattern recognition can be processed in data blocks and pixels by parallel threads. In fact, many algorithms, even outside the area of image processing, can be accelerated through parallelization of data processing, specially signal processing, simulation of physical effects, computer models of financial or biological applications.

4.1 CUDA - Compute Unified Device Architecture

The development of applications that use the GPU as a device for "unconventional” parallel data processing, i.e., not specifically the graphics processing like rendering, is increasing. However the use of a GPU as a device that requires an adjustment of the traditional graphics card pipeline's, forcing the developer to take responsibility for certain control points in these processes, through graphics libraries that have an API for GPU's to become programmable, is annoying.

CUDA is a new architecture of hardware and software that was developed with the main objective of managing the parallel processing of data within the GPU device without the need to make the mapping of the routines and take responsibility for the execution of the pipeline system, through API chart.

In Figure 12 we have the software stack environment of CUDA, not necessarily for 4 layers of software and these are: (a) application, which is implemented by the browser software that makes use of GPU as a device data processing; (b) CUDA Library is a set of mathematical libraries, such as CUBLAS, an extension of a BLAS library functions algebra implemented in FORTRAN and CUFFT's a fast Fourier transform of 1, 2 and 3 dimensions; (c) where the CUDA runtime routines of other graphics libraries like OpenGL and DirectX are accessed to be processed on the GPU; and (d) CUDA Driver API that is the direct communication with the GPU.

In order to facilitate the development of computing solutions for general purpose, not just graphic, CUDA provides the GPU direct memory access to both writing (Figure 13) and for reading (Figure 14), just as a conventional CPU works.

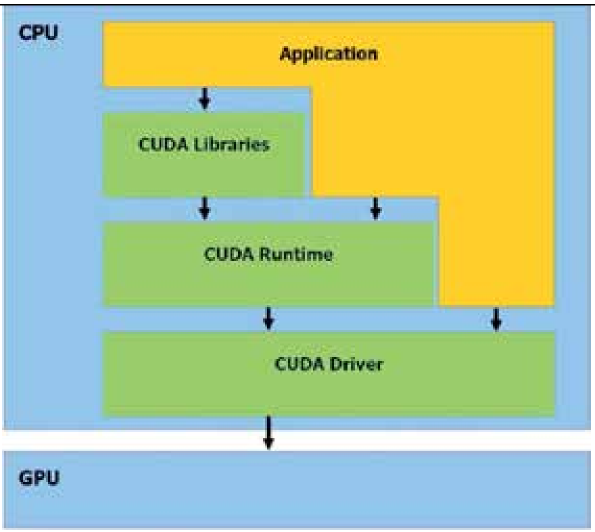


Fig. 12. CUDA Software Stack (NVIDIA, 2007).

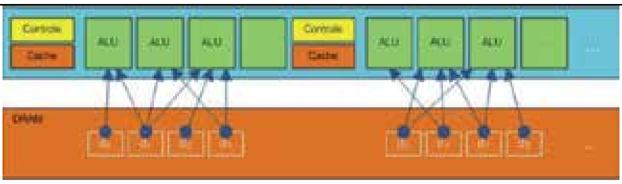


Fig. 13. GPU accessing memory to read (NVIDIA, 2007).

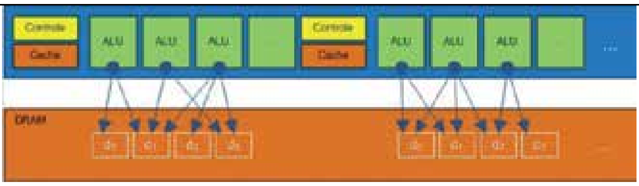


Fig. 14. GPU accessing memory to write (NVIDIA, 2007).

In these Figures 14 and 15, the data is read from or written to memory by the ALUs. In this architecture there is a parallel data cache and a shared memory, which has a high-speed access for both, writing and reading. The applications benefit from this structure by minimizing overfetch and round-trips of DRAM and reduce the need/dependence on the bandwidth of DRAM access.

4.2 CUDA Programming Model

In developing a parallel application via CUDA, GPU is viewed as a computer device capable of running a large number of threads in parallel. The GPU operates as a coprocessor of the CPU, which in the context of CUDA is called the host.

The part of the application, most suitable to be processed in the device, is a function performed several times with different data. These functions should be isolated and implemented within the scope of CUDA and are called the kernel that are executed within the device.

Both host and device (GPU) have one call to a DRAM memory device and a host memory. The call is made from a kernel due to the transfer of data between two memories. CUDA provides a set of functions for this feature (moving data between the two types of memory).

When a host application makes a call to a kernel, it is executed by a set of threads arranged in blocks of execution. These blocks in turn are grouped into grid blocks.

A block of threads is a lot of threads that work together cooperatively to get a better efficiency of data usage and shared memories, and their processing is synchronized. Each thread within a block is identified by a threadID, which is a combination of the number of thread with the block in which it is inserted.

The formation of a value of one threadID is complex, and to assist in this process, it can be specified a block to have two or three dimensions of arbitrary size, and identify each thread using a composite index of two or three instances, as shown in Table 1, where Dx, Dy, Dz are dimensions of the blocks, x, y, z are the coordinates, and threadID is obtained by calculating the expressions presented.

Block Size	Coordinate d Thread	threadID
D_x, D_y	x,y	$x+yD_x$
D_x, D_y, D_z	x,y,z	$x+yD_x+zD_xD_y$

Table 1. Formation of the address of the thread within the block

The number of threads that a block can contain is limited. As previously mentioned, blocks with the same dimensionality working in the execution of a single kernel can be grouped into a grid of blocks of threads. The call of this kernel is performed using a specific syntax which is reported beyond the normal parameters of the function to be processed on the device: data on grid (Dg), block (Db) and memory to be allocated (Ns).

As the threads, blocks also have an identification number within a grid, following a rule similar to the formation of the address of the threads, as shown in Table 2, where Dx, Dy indicate the dimensions of the grid, x, y the coordinates of blockID blocks and the identification number of the block calculated by the showed expression.

Grid Size	Coordinate d Block	blockID
D_x, D_y	x,y	$x+yD_x$

Table 2. Formation of the address of the block within the grid

In Figure 15 presents an overview of the structure of the threads running inside the device. This can be seen separating the two: host hardware (CPU) and the device (GPU), where the kernels called for implementation on the host are sent to the device, where the processing of threads arranged in blocks are divided into grids of processing.

It is worth calling attention here to the fact that kernels have distinct grid settings, and different blocks, as its dimensionality, as shown in Figure 15, where the size of the blocks and the grid used in kernel 2 is different from that used by the kernel 1.

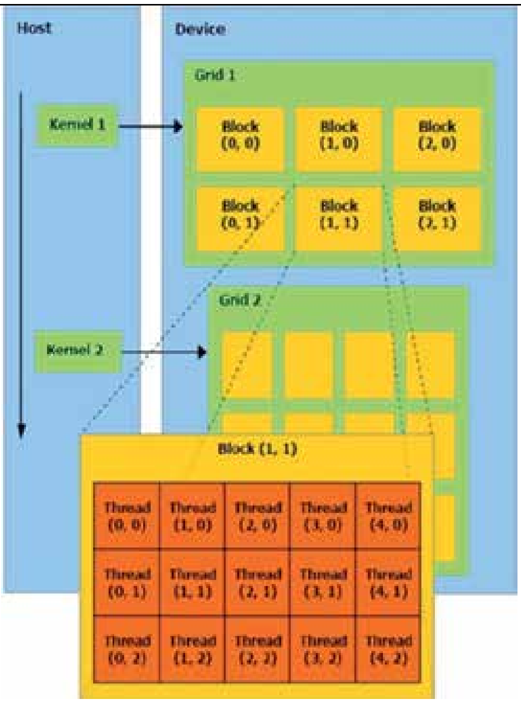


Fig. 15. Address of the threads within the blocks to the grid (NVIDIA, 2007).

4.3 CUDA Memory Model

The thread runs inside the device and only has (DRAM) memory access inside this, according to a set of access rules shown in Figure 16 and detailed in Table 3.

The threads can access registers (register) and memory space for reading and writing. The shared memory (shared) is accessed by blocks for writing and reading. The global memory is accessed by grid for reading and writing. Memories of constant and textures are accessed by the grid in read-only.

The global spaces, constant, and texture can be read or written by the host and are persistent across kernel calls during the same application.

Memory Space	When accessed	Rule
Register	by thread	Read/Write
Local	by thread	Read/Write
Shared	by block	Read/Write
Global	by grid	Read/Write
Constant	by grid	Read Only
Texture	by grid	Read Only

Table 3. Memory Model Access Rules

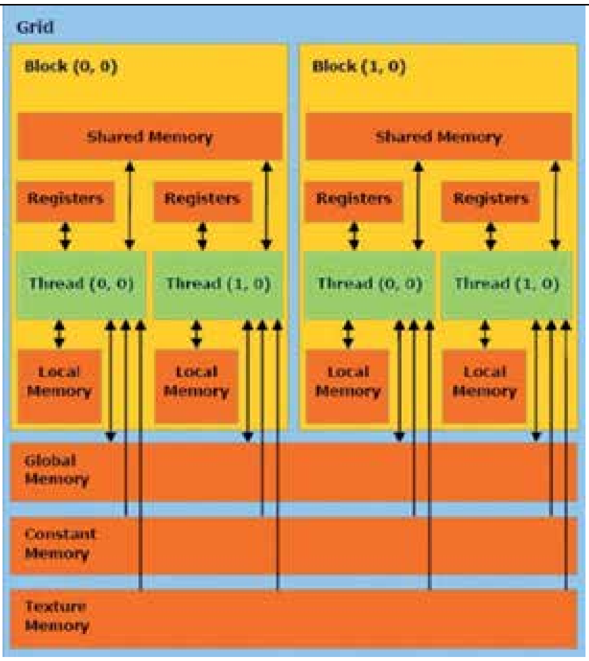


Fig. 16. CUDA Memory Model (NVIDIA, 2007).

5. Processing Neocognitron Network with CUDA

A face recognition system using Neocognitron neural network can be processed in two phases: the learning phase, and the recognition phase. This work have focused on recognition, since, the learning phase is ready. The training is being carried out by an application developed in Delphi (Saito and Abib, 2005) and it has as a product of its execution, the generation of a repository of data. This comprises a set of three types of binary files: one to store the number of plans for each stage, another type to store the weight-a e weight-b trained by the network and a third type with the results used by the layer of Cell-Cs of the stage.

At the recognition phase, one face image (input pattern) is shown to the system, and it executes and tries to identify the face. Figure 17 shows the block diagram of the parallel processing management algorithm of the face recognition phase using CUDA, in two parts, host and device.

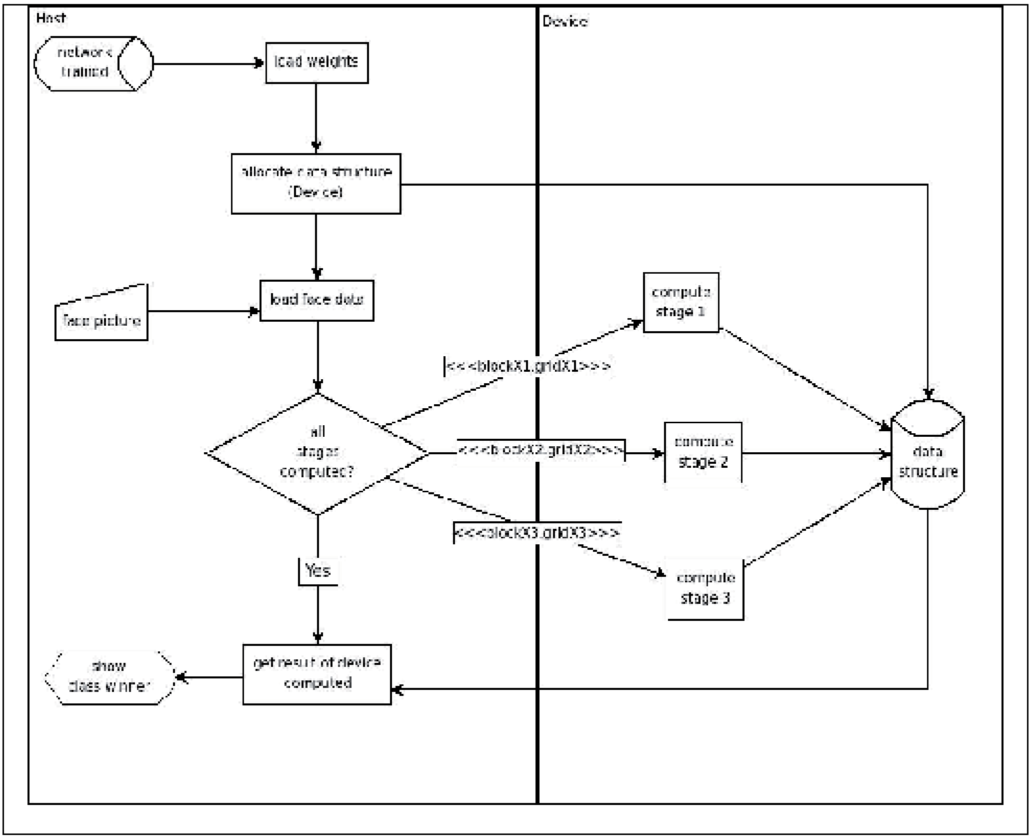


Fig. 17. Block Diagram Organization of Processing Neocognitron on CUDA device.

As can be seen in the block diagram (Figure 17), there are two repositories of data, one with the weights produced during the training phase of the network, and the second repository of data with the images of the faces to be recognized. We used two banks of the faces, one developed at UFSCar (Figure 19), consisting of fifty frontal images of six persons, in 57x57 pixels resolution; and the CMU PIE database (Figure 20) which consists of a large number of images of people in different poses, lighting and facial expressions. It used 13 cameras, 9 in the same horizontal line, each separated from 22.5^0. Other 4 cameras include 2 above and below the central camera, and 2 in the corners of the room. On Figure 19, the different positions of cameras are identified by ratings c02 ... c34. To obtain the change in lighting, it used a system with 21 flashes. Capturing images with and without the backlight, are obtained 43 different lighting conditions. For a variety of facial expressions were asked for people to neutral expressions, smile, blinking, and speaking. The database consists of 41368 images of 68 people.

Fig. 18. UFSCar Face Data Base.

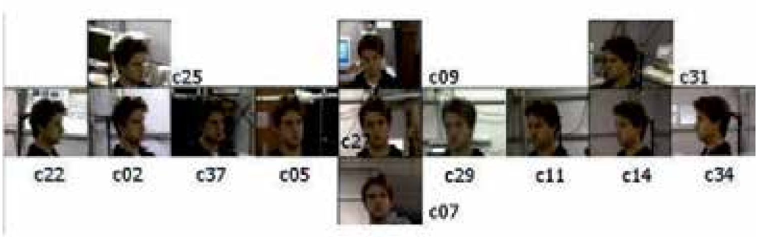


Fig. 19. CMU PIE database.

Since this work corresponds to the recognition phase, they were selected randomly 10 people of the database CMU PIE (4002, 4014, 4036, 4047, 4048, 4052, 4057, 4062, 4063, e 4067). They were selected the images of people speaking, because the existence of 60 images per pose, per person. Thus, during the experiments, they were used frontal images of people with size 640x486, and after the capture of the face region, the reduced image of size 57x57, as shown in Fig. 21.

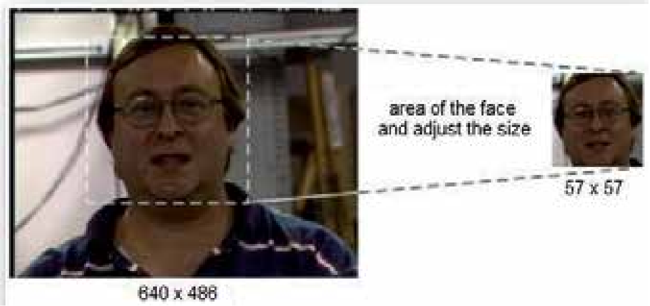


Fig. 20. Face picture used to recognition process.

At Figure 21, it can be seen the Neocognitron network processing, to the process of recognition. It may be noted the three stages of the network, represented by: stage 1, formed by all the layers U_{S1}, U_{C1} and U_{V1}; stage-2, formed by all the layers U_{S2}, U_{C2} and U_{V2} and stage-3, formed by all the layers U_{S3}, U_{C3} and U_{V3}. Also, it is presented the input layer U_0.

Table 4 shows the dimensionality of the weights used in the network, according to the stage where it is applied. The weight-a and weight-b are obtained by the training process of the network, which in the scope of this project is already done.

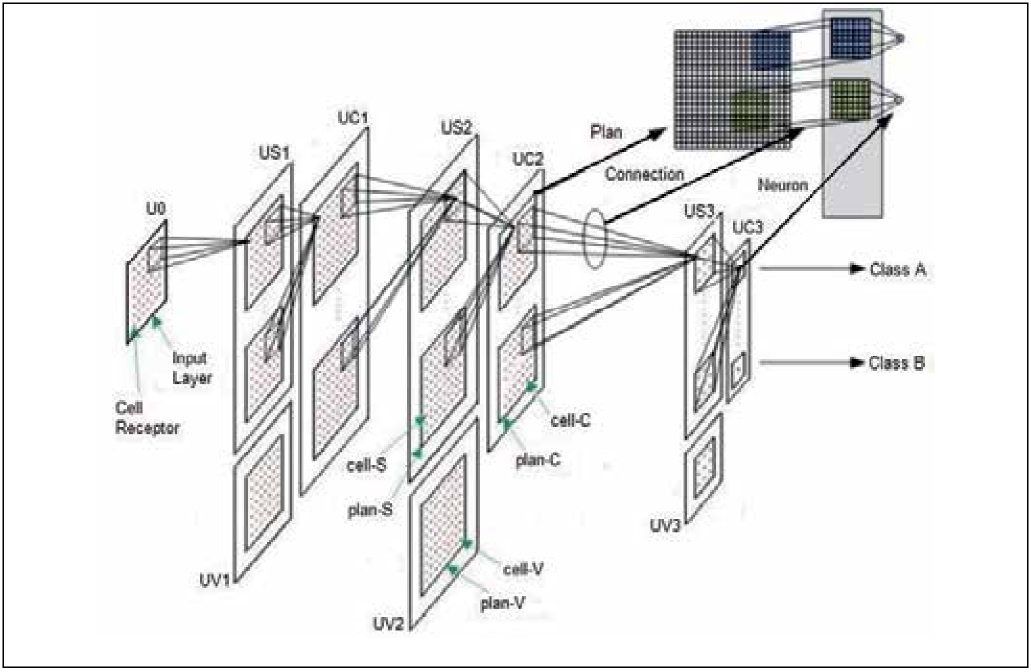


Fig. 21. Neocognitron network processing using GPU.

The weight-c and weight-d are fixed and defined at the time of implementing the network. On the Table 5, 6 and 7 are presented as matrices weight-c used in stages 1, 2 and 3 respectively.

Stage	Weight-a	Weight-b	Weight-c	Weight-d
1	7x7	1x1	7x7	5x5
2	7x7	1x1	7x7	5x5
3	5x5	1x1	5x5	3x3

Table 4. Dimensionality of the Weights used in the network.

0.017361	0.019231	0.019231	0.019231	0.019231	0.019231	0.017361
0.019231	0.019231	0.021635	0.021635	0.021635	0.019231	0.019231
0.019231	0.021635	0.021635	0.024038	0.021635	0.021635	0.019231
0.019231	0.021635	0.024038	0.026709	0.024038	0.021635	0.019231
0.019231	0.021635	0.021635	0.024038	0.021635	0.021635	0.019231
0.019231	0.019231	0.021635	0.021635	0.021635	0.019231	0.019231
0.017361	0.019231	0.019231	0.019231	0.019231	0.019231	0.017361

Table 5. Matrix of weight-c of stage 1.

0.017361	0.019231	0.019231	0.019231	0.019231	0.019231	0.017361
0.019231	0.019231	0.021635	0.021635	0.021635	0.019231	0.019231
0.019231	0.021635	0.021635	0.024038	0.021635	0.021635	0.019231
0.019231	0.021635	0.024038	0.026709	0.024038	0.021635	0.019231
0.019231	0.021635	0.021635	0.024038	0.021635	0.021635	0.019231
0.019231	0.019231	0.021635	0.021635	0.021635	0.019231	0.019231
0.017361	0.019231	0.019231	0.019231	0.019231	0.019231	0.017361

Table 6. Matrix of weight-c of stage 1.

0.035225	0.039628	0.039628	0.039628	0.035225
0.039628	0.039628	0.044031	0.039628	0.039628
0.039628	0.044031	0.048924	0.044031	0.039628
0.039628	0.039628	0.044031	0.039628	0.039628
0.035225	0.039628	0.039628	0.039628	0.035225

Table 7. Matrix of weight-c of stage 1.

Tables 8, 9 and 10 corresponds to the matrices of weight-d used in the processing of stages 1, 2, and 3, respectively.

0.72	0.72	0.72	0.72	0.72
0.72	0.81	0.9	0.81	0.72
0.72	0.9	1	0.9	0.72
0.72	0.81	0.9	0.81	0.72
0.72	0.72	0.72	0.72	0.72

Table 8. Matrix of weight-d of stage 1.

0.72	0.81	0.81	0.81	0.72
0.81	0.81	0.9	0.81	0.81
0.81	0.9	1	0.9	0.81
0.81	0.81	0.9	0.81	0.81
0.72	0.81	0.81	0.81	0.72

Table 9. Matrix of weight-d of stage 2.

0.81	0.9	0.81
0.9	1	0.9
0.81	0.9	0.81

Table 10. Matrix of weight-d of stage 3.

To carry out processing of any type of data within the CUDA, it must be able to be implemented within a hierarchical organization such as:

Grid >> Block >> Thread

meaning that the grid are composed by blocks, and the blocks by threads. The Neocognitron network also has an organization in a hierarchical structure, such as:

Stage >> Cell Plan >> Neuron

meaning that the stages are composed by several call-plans, and the cell-plans of a collection of cells, or neurons.

Analyzing the organization of the two architectures, it is possible to verify some points of correspondence, which can be seen in Figure 21 and listed in Table 11, which sees itself as a Grid equivalent to a Stage the block equivalent to all the neuron processing, and the thread equivalent to one connection processing. The correspondence between the two architectures facilitates the network modelling to be used at the CUDA/GPU environment.

CUDA	Neocognitron
Grid	Stage
Block	Cell Plan
Thread	Neuron

Table 11. Points of correspondence: CUDA x Neocognitron.

Another important factor of the validity of the correspondence between the architectures lies on the fact that there is an independence of the values of a huge amount of neurons at their processing. That is, the value of a neuron, in a cell-plan does not depend on the value of the neighbour neuron at the same plane, but on the data of the preceding stage. That validates the use of architecture as the GPU/CUDA.

The implementation of a project using the GPU/CUDA, determines that there are two processing environments, the host and device. It was developed a set of functions (processes) that run in the host and only a function that is performed on the device, as can be seen in Figure 17.

This kernel has the responsibility to process a single stage of the Neocognitron network, and is called by the host application in each stage within a specific order. It should be noted that the network model in this project has three stages.

Despite the processing of the network be similar for all stages, the Neocognitron network shows a reduction of dimensionality during its processing. The plan size of each stage is reduced from stage to stage, until the last stage, which has a single neuron in a plan, and a number of plans coincident to the number of classes to be recognized.

This is why the kernel is required at the time of processing a certain stage and can thus tell the GPU setting specific with respect to size of blocks of processing to be implemented. The models of cell-plans and connection area organizations, invoked by the kernels, by stage, are shown in Table 12. The goal is to process a plan with the greatest number of areas of possible connections.

Stage	Plan-S	Con. Area	Plan-C	Block	Grid
1	21x21	7x7	21x21	5x49	1
2	14x14	7x7	14x14	4x49	1
3	7x7	5x5	1x1	10x25	1

Table 12. Organization of cell-plans and connection area and its implementation in GPU/CUDA.

As a block processed in a single cycle of GPU/CUDA, the data of 16 multiprocessors has been that for the Stage-1. Each plan has a size of 21x21 (441) neurons and the connection area of this plan has a dimension of 7x7, resulting in 49 connections. The size of the block used for the processing of this stage was 49 threads, the same size of the connection area, 5 blocks in the grid processing simultaneously, or 245 connections computed simultaneously, as can be seen in Figure 22.

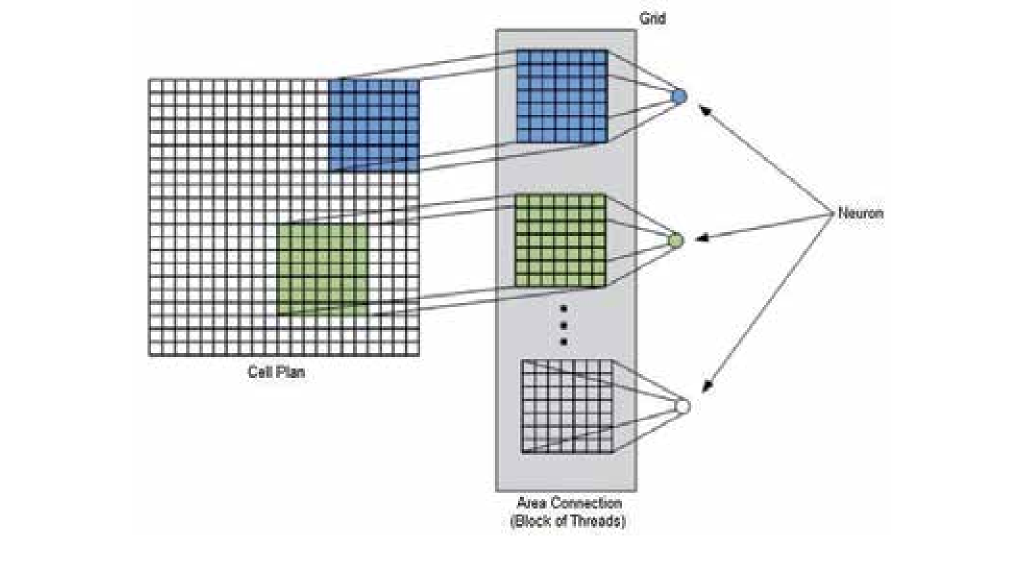


Fig. 22. Neuron connection processing diagram.

The total number of neurons processed simultaneously from 245, 245 and 250, to the stages 1, 2 and 3 respectively.

6. Results

Using the UFSCar and CMU-PIE human face databases, the recognition rate obtained is high and probably may be increased using more training images. The results of the recognition rate obtained by the two databases can be seen in Table 13.

Database Face	Rate Recognition
CMU-PIE	98%
UFSCar	97%

Table 13. Degree of accuracy of recognition of faces.

The total run time of the network was 0.118 seconds. The measure of time was obtained through the use of the control functions of processing time made available by the API CUDA, and used the cutCreateTimer functions, cutStartTimer, cutStopTimer, cutGetTimerValue and cutDeleteTimer.

Stage	Plan-S	time (sec)
1	95 plans	0.092
2	51 plans	0.022
3	47 plans	0.004

Table 14. Time, in seconds, spent during cell-plan processings.

Table 15 presents a comparison between the processing time in the GPU / CUDA, with the same network being processed in a single (mono) environment and processed in a cluster with 8 processors, values obtained by Ribeiro (Ribeiro 2002). It was made an adjustment in time goals for the work of Ribeiro, depending on the speed of processors used in their work and the existing today. The table is organized into three columns, where the first column is the computer architecture, the second column is the number of parallel processors, and the third column, processing time in seconds.

Architecture	Number of Processors	Time (sec)
Mono	1	48
Cluster	8	15
GPU/CUDA	128	0.118

Table 15. Comparing the processing time in different architectures.

By this table it is possible to calculate the speed-up and efficiency of processing between the architectures. These values are presented in Table 16, organized into three columns: computer architecture, Speed-up, and efficiency.

Architecture	Speed-up	Efficiency
Cluster	79.787	0.311
GPU/CUDA	255.319	0.999

Table 16. Comparation Speed-up and Efficiency in different architectures.

The total amount of memory used at the device was 439 MB, which represents an allocation of 57% of total memory. Since this is their distribution and consumption detailed at Table 17, in two columns, the first "Reserve Area" indicates where it allocated the amount of memory in Mega Bytes presented in the second column "Located Area".

Reserved Area	Located Area (Mb)
Stage 1	336
Stage 2	80
Stage 3	10
Other Variables	13

Table 17. Number of dedicated memory, of GPU, allocated for the implementation.

7. Conclusion

With the Neocognitron network processing within the GPU/CUDA presented in this work, we can conclude that there was a significant increase on the processing performance of the Neocognitron face recognition, showing the feasibility of using this method.

However the size of images used in the operation were small, 57 x 57, which allowed the full load of the structure of the network into the memory of the device where access is protected and high-speed, factors that may have influenced the results presented.

In an attempt to draw a line between the comparative Neocognitron network processed in the GPU/CUDA and traditional architecture it was verified through the calculation of speed-up, a gap, since they won a super-linear speed-up $S_p > p$. This occurred by differences in architecture. Moreover a high performance was observed when compared the time of processing.

Another conclusion on the implementation of this project is that this minimizes some existing common problems, when used other parallel computing environment, cluster that for example, you can quote:

- Synchronization: since the granularity of development within this device is not competing for shared memories (each thread has its point / area of memory) there is a need for loss of time for achieving a synchronization of processors;
- Network: if the whole process takes place within the same device there is the issue connection type and the speed of the entire GPU architecture; and
- Contention: there is no competition for resources by processors.

Another issue, where the GPU has advantages over the traditional architecture of high performance computing (such as cluster), is related to load balancing. Since the GPU architecture is focused on SIMD data processing type, the Neocognitron network implementation project, focused on block processing, is privileged, since an entire block is processed in a single cycle of processing.

However, the development of projects in GPU/CUDA environment presents as the main difficulty the modelling of streaming data processing. As seen in other studies by Poli (Poli et al. 2007) (Poli et al. 2008), it is not every applications that benefit with this architecture. It can be submitted three categories of possibilities for implementation of applications in the GPU: full potential for development, partial potential for development and the unfeasible development.

The applications that benefit with the processing in GPU/CUDA, are those that have large volumes of data on a matrix, and have their processing independent of the adjacent processing. The computer cost for decision making is significant. It is concluded that the shorter the granularity inside a model of data organized and structured to a vision processing in blocks will have a better gain in processing performance.

8. Acknowledgents

The authors would like to thank Prof. Simon Backer from Carnegie Mellon University for the kindness of sending the PIE database, used in this work.

9. References

Fukushima K. and Miyake S. (1982). *Neocognitron: A New Algorithm for Pattern Recognition Tolerant of Deformations and Shift in Position*, Pattern Recognition, Vol. 15, pages 455-469

Fukushima K. and Wake N. (1992). *Improved Neocognitron with Bend-Detection Cells*, IEEE - International Joint Conference on Neural Networks, Baltimore, Maryland, 1992

NVIDIA (2007). *NVIDIA CUDA Compute Unified Device Architecture - Programming Guide,* Publisher NVIDIA

Poli G., Levada A. M. L., Mari J. F., Saito J. H. (2007). *Voice Command Recognition with Dynamic Time Warping (DTW) using Graphics Processing Units (GPU) with Compute Unified Device Architecture (CUDA)*, SBAC-PAD International Symposium on Computer Architecture and High Performance Computing, 2007, pages 19-27

Poli G., Levada A. M.L., Saito J. H. , Mari J. F. , Zorzan M. R. (2008). *Processing Neocognitron of Face Recognition on High Performance Environment Based on GPU with CUDA Architecture (CUDA)*, SBAC-PAD International Symposium on Computer Architecture and High Performance Computing, 2008, pages 81-88

Saito J. H. and Fukushima K. (1998). *Modular Structure of Neocognitron to Pattern Recognition*, ICONIP'98, Fifth Int. Conf. on Neural Information Processing, Kitakyshu – Japan 1998

Hirakuri M. H. (2003). *Aplicação de Rede Neural Neocognitron para reconhecimenro de atributos faciais*, Dissertação de Mestrado - Universidade Federal de São Carlos, 2003

Ribeiro, L. J. (2002). *Paralelização da Rede Neural Neocognitron em Cluster SMPs.*, Dissertação de Mestrado da Universidade Federal de São Carlos (UFSCar), 2002

Sim T. , Bsat M. (2003). *The CMU Pose, Illumination, and Expression database*, IEEE Transaction on Pattern Analysis and and Machine Intelligence, 2003, pages 1615-1618

Saito J. H. and Abib S. (2005). *Using CMU PIE Human face database to a Convolutional Neural Network - Neocognitron*, ESANN2005 - European Symposium on Artificial Neural Network, 2005, pages 491-496

Terra Notícias (2006) *China implanta sistemas de reconhecimento facial biométrico,* http://noticias.terra.com.br/ciencia/interna/0,,OI956783-EI238,00.html

Edited by Milos Oravec

This book aims to bring together selected recent advances, applications and original results in the area of biometric face recognition. They can be useful for researchers, engineers, graduate and postgraduate students, experts in this area and hopefully also for people interested generally in computer science, security, machine learning and artificial intelligence.

Various methods, approaches and algorithms for recognition of human faces are used by authors of the chapters of this book, e.g. PCA, LDA, artificial neural networks, wavelets, curvelets, kernel methods, Gabor filters, active appearance models, 2D and 3D representations, optical correlation, hidden Markov models and others. Also a broad range of problems is covered: feature extraction and dimensionality reduction (chapters 1-4), 2D face recognition from the point of view of full system proposal (chapters 5-10), illumination and pose problems (chapters 11-13), eye movement (chapter 14), 3D face recognition (chapters 15-19) and hardware issues (chapters 19-20).

Photo by the-lightwriter / iStock

IntechOpen

ISBN 978-953-51-4550-9