Face Recognition
Semisupervised Classification, Subspace Projection and Evaluation Methods

Edited by S. Ramakrishnan

FACE RECOGNITION - SEMISUPERVISED CLASSIFICATION, SUBSPACE PROJECTION AND EVALUATION METHODS

Edited by **S. Ramakrishnan**

Face Recognition - Semisupervised Classification, Subspace Projection and Evaluation Methods
http://dx.doi.org/10.5772/61471
Edited by S. Ramakrishnan

Contributors

Konrad Gromaszek, Jarferr Yang, Shih-Ming Huang, Yang-Ting Chou, Mala Sundaram, Ambika Mani, Mingbo Zhao, Yuan Gao, Zhao Zhang, Bing Li, Srinivasan Ramakrishnan

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.

We are IntechOpen,
the first native scientific
publisher of Open Access books

3,350+
Open access books available

108,000+
International authors and editors

114M+
Downloads

151
Countries delivered to

Our authors are among the
Top 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

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

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

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

Meet the editor

Dr. S. Ramakrishnan is a Professor and the Head of the Department of Information Technology, Dr. Mahalingam College of Engineering and Technology, Pollachi, India. Dr. Ramakrishnan is a reviewer of 25 international journals such as *IEEE Transactions on Image Processing*, *IET Journals* (Formally IEE), *ACM Computing Reviews*, Elsevier Science Journals, Springer Journals, and Wiley Journals. He is a guest editor of special issues in three international journals, including *Telecommunication Systems Journal of Springer*. He has published 116 papers in international, national journals and conference proceedings. Dr. S. Ramakrishnan has published a book on wireless sensor networks for CRC Press, USA, two books on speech processing for InTech Publisher, Croatia, and a book on computational techniques for Lambert Academic Publishing, Germany.

Contents

Preface

Pattern recognition has gained significant attention due to the rapid explosion of internet- and mobile-based applications. Among the various pattern recognition applications, face recognition is always being the center of attraction. With so much of unlabeled face images being captured and made available on internet (particularly on social media), conventional supervised means of classifying face images become challenging. This clearly warrants for semi-supervised classification and subspace projection. Another important concern in face recognition system is the proper and stringent evaluation of its capability. This book is edited keeping all these factors in mind. This book is composed of five chapters covering introduction, overview, semi-supervised classification, subspace projection, and evaluation techniques.

Chapter 1 provides a brief introduction to ensure maximum coherence and relatedness of the remaining four chapters of this book, and also explains the nature and purpose of the book.

Chapter 2 offers a brief overview of the face recognition systems and its components. This chapter highlights the important applications of face recognition. The authors of Chapter 2 presented a comprehensive overview of various classical face recognition methods. This chapter helps the new readers in understanding the various inquisitives of face recognition.

Chapter 3 discuses semi-supervised classification method. The authors have beautifully narrated the need and background for the semi-supervised face recognition. Existing methods of semi-supervised classification is deeply studied and authors of this chapter have identified the research gaps. Also, they have proposed a new algorithm as a solution to the gaps by conducting extensive experiments.

Chapter 4 focuses on latest technique named linear regression and its variants, over and above the classical subspace projection techniques. Important and critical issues in face recognition namely partial occlusion, illumination variance, different expression, pose variance, and low resolution are all addressed and presented.

Finally, Chapter 5 presents various important and critical metrics that should be used to evaluate the performances of the face recognition system. All the metrics are presented from the basics, and the authors have also provided case studies to demystify the myths in the performance evaluation of face recognition system.

Overall, this book is brief and comprehensive and will be a useful resource for the graduate students, researchers, and practicing engineers in the field of machine vision and computer science and engineering.

Dr. S. Ramakrishnan
Professor and Head, Department of Information Technology,
Dr. Mahalingam College of Engineering and Technology,
Pollachi, India

Introductory Chapter: Face Recognition - Overview, Dimensionality Reduction, and Evaluation Methods

S. Ramakrishnan

Additional information is available at the end of the chapter

Face recognition is one of most popular and powerful applications in modern computing industries [1–4]. It has found applications ranging from person identification (surveillance) [5] to emotion identification (human–machine interaction) [6]. Over the past few decades, researchers in field of computers and electrical and electronics engineering have worked continuously to improve the performances of the face recognition systems. In-spite of these continuous efforts, there are still a plenty of scope for the new and additional research in the field of face recognition. This is due to the popularization of light-weight computing devices, increased customer expectations, and business competitions.

Now the world best cameras in terms of resolution are available in smart phones at affordable price, and CCD cameras are found even in houses and almost in all commercial, business, and office environments including small-sized enterprises. Amount of face images being captured keep on increasing, and recognition of faces among these huge databases makes the task further challenging. One of most the important subtopics in face recognition is dimensionality reduction [1], because storing and processing of these high-resolution face images from huge database using light-weight devices require dimensionality reduction.

Several different face recognition systems, including hardware (cameras, memory disk, and processors) and software, are available in the market. These face recognition systems provide better performance in one aspect and lack in other aspect. Comprehensive evaluation the performances of face recognition systems is the need of the hour.

Keeping these factors in mind, this book on "face recognition" is focusing on dimensionality reduction and evaluation methods. This book is brief but comprehensive. Other than this introductory chapter, this book has four more chapters, two chapters for dimensionality reduction and one for an overview of the face recognition systems and evaluation methods.

Rest of this introductory chapter is spared for providing a brief outlines, linkages, importance, and significances of the four chapters of this book.

Chapter 1: This chapter provides an overview of face recognition, various issues in face recognition, and different methods of face recognition and applications of face recognition. We strongly encourage the young readers to thoroughly study this chapter to get the bird's eye view of face recognition. Advanced readers can proceed to Chapter 2 directly.

Typical complex engineering applications requires various submodules and proper fine tuning of all those modules to make the application perfect. Face recognition, one of the toughest complex engineering applications, certainly requires number of submodules. A few important submodules are pre-processing, face detection and normalization, feature database and classifier. These building blocks are presented in Chapter 1 in a simple way. Various challenges in face recognition include [7–10]: scale invariance, rotation invariance, translation invariance, illumination invariance, and emotion invariance. All these make the task difficult for the face recognition system. These challenges are discussed in Chapter 1.

Authors of Chapter 1 presented a comprehensive overview of various classical face recognition methods. Classification of 18 different classical face recognition algorithms based on local and holistic features is also presented in this chapter nicely. Over and above, the classical face recognition methods and modern face recognition methods are briefly introduced in Chapter 1. Modern techniques include artificial neural networks, wavelets-based methods, descriptor-based method, 3D methods, and video-based techniques. Advantages and disadvantages of both classical and modern methods are narrated in Chapter 1. This will help the students to choose an appropriate technique for doing their projects. Eight different potential applications of face recognition systems are highlighted in Chapter 1. Ideally, through reading of Chapter 1 will be of immense help for the young readers.

Chapter 2: Traditional pattern recognition methods can either be a supervised learning or unsupervised learning. Face recognition methods comes under supervised learning methods. Supervised learning requires proper and complete labeling of all patterns and objects. Due to social media and in general internet, the amount of face images being generated is steeply increasing. Most of these face images are not labeled by required for the face recognition system to provide satisfactory performance. Hence, a new type of learning method, which is a subtype of supervised leaning called, semi-supervised learning method is being applied to modern face recognition methods [11–16]. This chapter is dealing with this new learning method and also addressing dimensionality reduction concept in semi-supervised learning.

Semi-supervised learning methods can be grouped under transductive learning or induction learning. Authors of this chapter have systematically presented the state-of-the-art methods and nicely introduced their contribution in this chapter. Authors of Chapter 2 have proposed a new and effective algorithm for semi-supervised dimensionality using local and global regression. The algorithm proposed in this chapter is capable of reducing dimensions of both transductive learning and induction learning. Proposed algorithm is explained from the first principles so that the readers with pattern recognition or image processing background can easily understand and apply this in their projects. Presentation of the proposed algorithm is

excellent as it has proper mix of analytical and descriptive treatments. Theorems employed by the authors are also provided and over and above the proposed concepts are illustrated with intermediate results. This is a must read subsection for the young learners.

In addition to the smooth and neat presentation of the proposal and related works, the authors of Chapter 2 have conducted extensive experiments and beautifully presented the results along with appropriate discussions. Experiments are conducted not only using synthetic dataset but also using three of the real-world bench mark datasets, namely UMNIST, extended Yale B, and MIT-CBCL. Experimental results are also compared with existing methods. This chapter is well written and much useful for the both young and senior researchers working in pattern recognition.

Chapter 3: Among the various challenges of a typical pattern recognition system, dimensionality reduction is one of important tasks. Image processing applications such as face recognition should focus on dimensionality reduction for better performance. Subspace projection techniques are highly useful and classical option in face recognition is useful for reducing the dimension. Principle component analysis (PCA) and linear discriminant analysis (LDA) are both popular and powerful subspace projection techniques over the past few decades [17] and applied in almost all pattern recognition systems [18–26].

In face recognition, input–output pairs are known as it is mostly supervised. Here, linear regression that used to fit a linear function to a set of input–output pairs is latest technique and also comes under subspace projection. Chapter 3 is focusing on latest technique named linear regression and its variants, over and above the classical subspace projection techniques. Important and critical issues in face recognition, namely partial occlusion, illumination variance, different expression, pose variance, and low resolution are all addressed and presented.

This chapter is self-contained and comprehensive. Authors of this chapter have provided a brief overview of how face images are represented and recognized. Two of the classical subspace projection methods, namely PCA and LDA are briefly presented in this chapter. This quick introduction will help even advanced readers to recall the basics. In addition to this similarity metrics used in the classifier stage of face recognition systems are also presented.

Various latest subspace optimization techniques, namely linear regression classification, robust linear regression classification, improved principal component regression, unitary regression classification, linear discriminant regression classification, generalized linear regression classification, and trimmed linear regression are all presented. These eight methods are discussed in this chapter with correct blend of mathematical equations and theoretical descriptions.

Authors have conducted extensive experiments are presented the results. Performance analysis is carried out on the benchmark datasets, namely Yale B, AR, FERT, and FEI. Comparative analysis of the various subspace projection methods and linear regression and its variants are also provided precisely in this chapter. This chapter is self-reliant and will be useful to both young and advanced readers.

Chapter 4: Performance evaluation is one of most important aspects in face recognition applications [2, 3, 27–29]. Recognition rate (or classification accuracy) is the commonly used metrics to analyze the performance of the face recognition methods. But there are several other important and critical metrics available for performance evaluation of the system. In this chapter, those metrics are presented and demonstrated. A brief outline of face recognition techniques and methods are provided in this chapter. Four important component of a confusion matrix, namely true positive, true negative, false positive, and false negative are presented. Based on these four parameters, seven significant evaluation metrics, namely precision, recall, sensitivity, specificity, fallout, error rate, and accuracy are presented in this chapter. Receiver operating characteristics (ROC) curve analysis is presented sensitivity and specificity. Salient points in ROC analysis are illustrated clearly for all possible performances of face recognition methods.

Like ROC combines sensitivity and specificity, F-score combines precision and recall, and this metric is better explained in this chapter. In addition to these metrics, the following metrics are also briefed: false match rate, false non-match rate, equal error rate, failure to enroll rate, and failure to capture rate.

Authors of this chapter have conducted experiments to analyze the performances of the face recognition using these metrics. Three different case studies are presented using face images from the benchmark datasets. Whoever developing face recognition system finds this chapter useful.

Final word: This book has five chapters including this introductory. This book can be a brief material and will be highly useful for students, researchers, and practicing engineering working in pattern recognition, image processing, and machine vision.

Author details

S. Ramakrishnan[*]

Address all correspondence to: ram_f77@yahoo.com

Department of Information Technology, Dr. Mahalingam College of Engineering and Technology, Pollachi, Coimbatore, India

References

[1] Ramakrishnan, S., and Ibrahiem M.M.EI Emary. Computational techniques and algorithms for image processing (ISBN: 978-3-8433-5802-6) & (ISBN: 3843358028). Lambert Academic Publishing, Germany, 2012.

[2] Gonzalez, Rafael C., and Richard E. Woods. Digital image processing (3rd Edition) (ISBN: 013168728X). Prentice-Hall, Inc., Upper Saddle River, 2006.

[3] Jain, Anil K., Fundamentals of digital image processing. Prentice Hall Inc., Upper Saddle River, 1989.

[4] Duda, Richard O., and Peter E. Hart, Pattern classification and scene analysis, vol. 3. Wiley, New York, 1973.

[5] Szeliski, Richard, Computer vision: algorithms and applications. Springer, New York, 2010.

[6] Ming, Yue. Robust regional bounding spherical descriptor for 3D face recognition and emotion analysis. *Image and Vision Computing*, 35:14–22, 2015.

[7] Ding, Changxing, Chang Xu, and Dacheng Tao. Multi-task pose-invariant face recognition. *IEEE Transactions on Image Processing*, 24(3):980–993, 2015.

[8] Yang, Allen Y., et al. Fast-minimization algorithms for robust face recognition. *IEEE Transactions on Image Processing*, 22(8):3234–3246, 2013.

[9] Liao, Shengcai, Anubhav K. Jain, and Stan Z. Li. Partial face recognition: alignment-free approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(5):1193–1205, 2013.

[10] Jun, Bongjin, and Daijin Kim. Robust face detection using local gradient patterns and evidence accumulation. *Pattern Recognition*, 45(9):3304–3316, 2012.

[11] Yu, Jun, Meng Wang, and Dacheng Tao. Semisupervised multiview distance metric learning for cartoon synthesis. *IEEE Transactions on Image Processing*, 21(11):4636–4648, 2012.

[12] Criminisi, Antonio, Jamie Shotton, and Ender Konukoglu. Decision forests: a unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. *Foundations and Trends® in Computer Graphics and Vision*, 7(2–3):81–227, 2012.

[13] Gao, Quanxue, et al. A novel semi-supervised learning for face recognition. *Neurocomputing*, 152:69–76, 2015.

[14] Yi, Yugen, et al. Semi-supervised local ridge regression for local matching based face recognition. *Neurocomputing*, 167:132–146, 2015.

[15] Yang, Shuyuan, et al. Semisupervised dual-geometric subspace projection for dimensionality reduction of hyperspectral image data. *IEEE Transactions on Geoscience and Remote Sensing*, 52(6):3587–3593, 2014.

[16] Zhao, Mingbo, et al. Learning from normalized local and global discriminative information for semi-supervised regression and dimensionality reduction. *Information Sciences*, 324:286–309, 2015.

[17] Imran Naseem, Roberto Togneri and Mohammed Bennamoun, Linear regression for face recognition, *IEEE Transactions on PAMI*, 32(11):2106–2112, 2010.

[18] Zhang, Yong-Qin, et al. Guided image filtering using signal subspace projection. *Image Processing, IET*, 7(3):270–279, 2013.

[19] Pyatykh, Stanislav, Jurgen Hesser, and Lei Zheng. Image noise level estimation by principal component analysis. *IEEE Transactions on Image Processing*, 22(2):687–699, 2013.

[20] Wan, Tao, Chenchen Zhu, and Zengchang Qin. Multifocus image fusion based on robust principal component analysis. *Pattern Recognition Letters*, 34(9):1001–1008, 2013.

[21] Moulin, Christophe, et al. Fisher linear discriminant analysis for text-image combination in multimedia information retrieval. *Pattern Recognition*, 47(1):260–269, 2014.

[22] Shu, Xin, Yao Gao, and Hongtao Lu. Efficient linear discriminant analysis with locality preserving for face recognition. *Pattern Recognition*, 45(5):892–1898, 2012.

[23] Huang, Shih-Ming, and Jar-Ferr Yang. Linear discriminant regression classification for face recognition. *IEEE Transactions on Signal Processing Letters*, 20(1):91–94, 2013.

[24] Sharma, Alok, and Kuldip K. Paliwal. A two-stage linear discriminant analysis for face-recognition. *Pattern Recognition Letters*, 33(9):1157–1162, 2012.

[25] Huang, Shih-Ming, and Jar-Ferr Yang. Improved principal component regression for face recognition under illumination variations. *IEEE Transactions on Signal Processing Letters*, 19(4):179–182, 2012.

[26] Naseem, Imran, Roberto Togneri, and Mohammed Bennamoun. Robust regression for face recognition. *Pattern Recognition*, 45(1):104–118, 2012.

[27] Yang, Meng, et al. Gabor feature based robust representation and classification for face recognition with Gabor occlusion dictionary. *Pattern Recognition*, 46(7):1865–1878, 2013.

[28] Huang, Shih-Ming, and Jar-Ferr Yang. Unitary regression classification with total minimum projection error for face recognition. *IEEE Transactions on Signal Processing Letters*, 20(5):443–446, 2013.

[29] Galbally, Javier, Sébastien Marcel, and Julian Fierrez. Image quality assessment for fake biometric detection: application to iris, fingerprint, and face recognition. *IEEE Transactions on Image Processing*, 23(2):710–724, 2014.

Face Recognition: Issues, Methods and Alternative Applications

Waldemar Wójcik, Konrad Gromaszek and
Muhtar Junisbekov

Additional information is available at the end of the chapter

Abstract

Face recognition, as one of the most successful applications of image analysis, has recently gained significant attention. It is due to availability of feasible technologies, including mobile solutions. Research in automatic face recognition has been conducted since the 1960s, but the problem is still largely unsolved. Last decade has provided significant progress in this area owing to advances in face modelling and analysis techniques. Although systems have been developed for face detection and tracking, reliable face recognition still offers a great challenge to computer vision and pattern recognition researchers. There are several reasons for recent increased interest in face recognition, including rising public concern for security, the need for identity verification in the digital world, face analysis and modelling techniques in multimedia data management and computer entertainment. In this chapter, we have discussed face recognition processing, including major components such as face detection, tracking, alignment and feature extraction, and it points out the technical challenges of building a face recognition system. We focus on the importance of the most successful solutions available so far. The final part of the chapter describes chosen face recognition methods and applications and their potential use in areas not related to face recognition.

Keywords: face recognition, biometric identification, methods, applications, image processing

1. Introduction

Recent advances in automated face analysis, pattern recognition and machine learning have made it possible to develop automatic face recognition systems to address these applications. On the one hand, recognising face is natural process, because people usually do it effortlessly without much conscious. On the other hand, application of this process in area of computer vision remains a difficult problem. Being part of a biometric technology, automated face recognition has a plenty of desirable properties. They are based on the important advantage— non-invasiveness. The various biometric methods can be distinguished into physiological (fingerprint, DNA, face) and behavioural (keystroke, voice print) categories. The physiological approaches are more stable and non-alterable, except by severe injury. Behavioural patterns are more sensitive to human overall condition, such as stress, illness or fatigue.

The brief analysis of the face detection techniques using effective statistical learning methods seems to be crucial as practical and robust solutions.

Figure 1 points out the basic elements of the typical face recognition system.
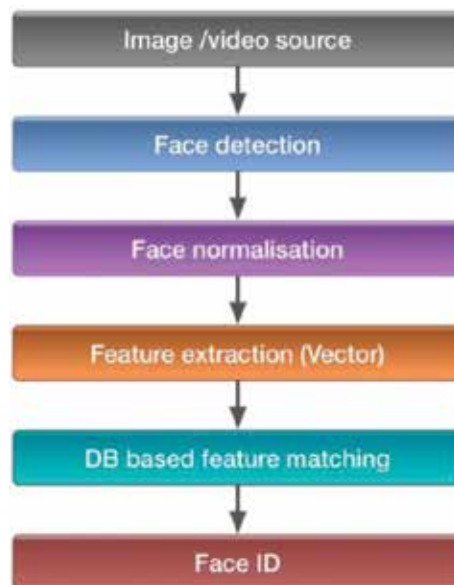
Figure 1. Crucial elements of the typical face recognition system.

Face detection performance is a key issue, so techniques for dealing with non-frontal face detection are discussed. Subspace modelling and learning-based dimension reduction methods are fundamental to many current face recognition techniques. Discovering such subspaces so as to extract effective features and construct robust classifiers stands another challenge in this area. Face recognition has merits of both high accuracy and low intrusive, so

it has drawn the attention of the researches in various fields from psychology, image processing to computer vision.

The first stage is face detection in the acquired image that is regardless of scale and location. It often uses an advanced filtering procedure to distinguish locations that represent faces and filters them with accurate classifiers. It is notable that all translations, scaling and rotational variations have to be dealt in the face detection phase. For example, regarding to [1,2], facial expressions and hairstyle changes or smiling and frowning face still stands important variations during pattern recognition stage.

In the next step, anthropometric data set-based system predicts the approximate location of the principal features such as eyes, nose and mouth. Of course, whole procedure is repeated to predict the subfeatures, relative to principal features, and verified with collocation statistic to reject any mislocated features.

Dedicated anchor points are generated as the result of geometric combinations in the face image and then it starts the actual process of recognition. It is carried out by finding local representation of the facial appearance at each of the anchor points. The representation scheme depends on approach. In order to deal with such complication and find out the true invariant for recognition, researchers have developed various recognition algorithms.

There are several boundaries for current face recognition technology (FERET). In [3,4] was provided early benchmark of face recognition technologies. While under ideal conditions, performance is excellent, under conditions of changing illumination, expression, resolution, distance or aging, performance decreases significantly. It is the fact that face recognition systems are still not very robust regarding to deviations from ideal face image. Another problem is an effective way of storing and access granting to facial code (or facial template) stored as a set of features and extracted from image or video.

Considering roughly presented elements above of the complex process of face recognition, a number of limitations and imperfections can be seen. They require clarification or replacing by new algorithms, methods or even technologies.

In this chapter, we have discussed face recognition processing, including major components such as face detection, tracking, alignment and feature extraction, and it points out the technical challenges of building a face recognition system. We focus on the importance of the most successful solutions available so far.

The final part of the chapter describes chosen face recognition methods and applications and their potential use in areas not related to face recognition.

The need for this study is justified by an invitation to participate in the further development of a very interesting technology, which is face recognition.

Despite the fact, there is continual performance improvement regarding several face recognition technology areas, and it is worth to note that current applications also impose new requirements for its further development.

2. Previous methods

2.1. Classical face recognition algorithms

There has been a rapid development of the reliable face recognition algorithms in the last decade. The traditional face recognition algorithms can be categorised into two categories: holistic features and local feature approaches. The holistic group can be additionally divided into linear and nonlinear projection methods.

Many applications have shown good results of the linear projection appearance-based methods such as principal component analysis (PCA) [5], independent component analysis (ICA) [6], linear discriminate analysis (LDA) [7,8], 2DPCA [9] and linear regression classifier (LRC) [10].

However, due to large variations in illumination conditions, facial expression and other factors, these methods may fail to adequately represent the faces. The main reason is that the face patterns lie on a complex nonlinear and non-convex manifold in the high-dimensional space.

In order to deal with such cases, nonlinear extensions have been proposed like kernel PCA (KPCA), kernel LDA (KLDA) [11] or locally linear embedding (LLE) [12]. The most nonlinear methods using the kernel techniques, where the general idea consists of mapping the input face images into a higher-dimensional space in which the manifold of the faces is linear and simplified. So the traditional linear methods can be applied.

Although PCA, LDA and LRC are considered as linear subspace learning algorithms, it is notable that PCA and LDA methods focus on the global structure of the Euclidean space, whereas LRC approach focuses on local structure of the manifold.

These methods project face onto a linear subspace spanned by the eigenface images. The distance from face space is the orthogonal distance to the plane, whereas the distance in face space is the distance along the plane from the mean image. These both distances can be turned into Mahalanobis distances and given probabilistic interpretations [13].

Following these, there have been developed: KPCA [14], kernel ICA [15] and generalised linear discriminant analysis [16].

Despite strong theoretical foundation of kernel-based methods, the practical application of these methods in face recognition problems, however, does not produce a significant improvement compared with linear methods.

Another family of nonlinear projection methods has been introduced. They inherited the simplicity from the linear methods and the ability to deal with complex data from the nonlinear ones. Among these methods, it is worth to underline: LLE [17] and locality preserving projection (LPP) [18]. They produce a projection scheme for training data only, but their capability to project new data items is questionable.

In the second category, local appearance features have certain advantages over holistic features. These methods are more stable to local changes such as expression, occlusion and

misalignment. The common representative method names local binary patterns (LBPs) [19,20]. The neighbouring changes around the central pixel in a simple but effective way are described by LBP. It is invariant monotonic intensity transformation and supports small illumination variations. Many LBP variants are proposed to improve the original LBP such as histogram of Gabor phase patterns [21] and local Gabor binary pattern histogram sequence [22,23]. Generally, the LBP is utilised to model the neighbouring relationship jointly in spatial, frequency and orientation domains [22].

It allows to explore efficiently discriminant and robust information in the pattern. Further development of the mentioned subspace approaches represents discriminant common vectors (DCVs) approach [24].

The DCV method collects the similarities among the elements in the same class and drops their dissimilarities. Thus, each class can be represented by a common vector computed from the within scatter matrix.

In case of testing an unknown face, the corresponding feature vector is computed and associated to the class with the nearest common vector. Sometimes, kernel discriminative common vectors [25] or improved discriminative common vectors and support vector machine (SVM) [26] are introduced for the face recognition task.

Similarly to the LLE method, neighbourhood preserving projection (NPP) and orthogonal NPP (ONPP) are introduced in [27,28]. These approaches preserve the local structure between samples. To reflect the intrinsic geometry of the local neighbourhoods, they use data-driven weights by solving a least-squares problem. ONPP forces the mapping to be orthogonal and then solves an ordinary eigenvalue problem. NPP requires solving a generalised eigenvalue problem, regarding to imposing a condition of orthogonality on the projected data.

Block diagram of the traditional face recognition approaches is presented in **Figure 2**.
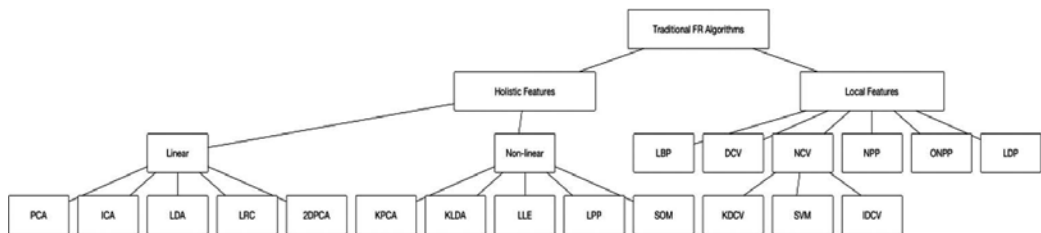
Figure 2. Traditional face recognition algorithms.

However, it is still unclear how to select the neighbourhood size and how to assign optimal values for other hyper-parameters; for them, sparsity preserving projections [29,30] and LPPs [31] are also applied for face recognition.

In [32], a multi-linear extension of the LDA method called discriminant analysis with tensor representation is proposed. It is different from preserving projection methods and implements discriminant analysis directly on the natural tensorial data to preserve the neighbourhood

structure of tensor feature space. Another method of supervised and unsupervised multi-linear NPP (MNPP) for face recognition is presented in [33]. A survey of multi-linear methods can be found in [11]. They operate directly on tensorial data rather than vectors or matrices and solve problems of tensorial representation for multidimensional feature extraction and recognition. Multiple interrelated subspaces are obtained in the MNPP method by unfolding the tensor over different tensorial directions. The order of the tensor space determines the number of subspaces derived by MNPP [34,35].

2.2. Artificial Neural Networks in face recognition

In [11,36,37], artificial neural networks are used to solve nonlinear problem. To recognise human faces, a non-convergent chaotic neural network is suggested in [38].

A radial basis function neural network integrated with a non-negative matrix factorisation to recognise faces is presented in [39]. Moreover, for face and speech verifications, [40] utilise a momentum back propagation neural network. Non-negative sparse coding method to learning facial features using different distance metrics and normalised cross-correlation for face recognition is applied in [41].

A posterior union decision-based artificial neural network approach is proposed in [33,34]. It has elements of both neural networks and statistical approaches and replenishes methods for recognising face images with partial distortion and occlusion.

Unfortunately, this approach, like other statistical-based methods, is inaccurate to model classes given only a single or a small number of training samples [42,43].

2.3. Gabor wavelet-based solutions

Gabor wavelets have been widely used for face representation by face recognition researchers [44,45,46], and Gabor features are recognised as better representation for face recognition in terms of (rank-1) recognition rate [47]. Moreover, it is demonstrated to be discriminative and robust to illumination and expression variations [48]. When only one sample image per enrolled subject is available, [49] propose adaptively weighted sub-Gabor array for face representation and recognition.

Moreover, two kinds of strategies to capture Gabor texture information: Gabor magnitude-based texture representation (GMTR) and Gabor phase-based texture representation (GPTR), are proposed in [50].

Gamma density to model the Gabor magnitude distribution characterises GMTR approach. The GPTR is characterised by the generalised Gaussian density for modelling the Gabor phase distribution. It allows the estimated model parameters to be served as texture representation of the face.

The Gabor wavelet applied at fixed positions, in correspondence of the nodes of a square-meshed grid superimposed to the face image, is presented in [51]. Each subpattern of the partitioned face image is defined as the extracted Gabor features that belong to the same row of the square-meshed grid which are then projected to lower dimension space by Karhunen–

Loeve transform. The obtained features of each subpattern, which are weighted using genetic algorithm (GA), are used to train a Parzen Window Classifier. Finally, matching process is done by combining the classifiers using a weighted sum rule.

The learning approach based on Gabor features and kernel supervised Laplacian faces for face recognition under the classifier fusion framework is introduced in [52]. The Gabor features obtained from each channel as a new sample of the same class are used to adopt the classifier fusion strategy. Such approach is useful for improving the performance of the recognition results.

Histogram of Gabor phase feature is proposed in [53]. In [54,55,56,57,58], the patch-based histograms of local patterns are concatenated together to form the representation of the face image via learned local Gabor patterns. The feature representation problem by providing a learning method instead of simple concatenation or histogram feature is presented in [59]. In [60], the Gabor features were adopted for the sparse representation (SR)-based classification and a Gabor occlusion dictionary was learned under the well-known SR framework.

The main drawback of Gabor-based methods is that the dimensionality of the Gabor feature space is significantly high since the face images are convolved with a bank of Gabor filters.

To overcome this problem, Adaboost algorithm [61] and entropy and genetic algorithms (GA) [62] are used to select the most discriminative Gabor features.

However, selecting the most useful method from so many Gabor features is very time-consuming [61]. Furthermore, extracting the Gabor features is computationally intensive, so the features are currently useless for real-time applications [63]. A simplified version of Gabor wavelets is introduced in [64]. Unfortunately, the simplified Gabor features are more sensitive to lighting variations in reference to the original Gabor features.

2.4. Face descriptor-based methods

Local feature-based face image description provides a global description. So local features of the image are evaluated in the neighbouring pixels and then aggregated to form the final global description [65,66]. This is unlike global methods in which the entire image is utilised to produce each feature, where the first steps start with the description of the face realised at a pixel level by making use of the local neighbourhood of each pixel. Then, the image is divided into a number of subregions, and from each subregion, a local description is formed as a histogram of the pixel level descriptions calculated in the previous step. Next, the information of the regions is combined into the final descriptor by concatenating the partial histograms [67,68].

To determine image descriptors that are able to improve classification performance of multi-option recognition as well as pair matching of face images seems to be a complex problem [65,69,70].

Learning the most discriminant local features that can minimise the difference of the features between images of a same individual and maximise that between images from other people

depending on the nature of these descriptors, which compute an image representation from local patch statistics stands the main idea of the approach.

The face verification accuracy ranked on the LFW benchmark after face verification using multiple local descriptors designed to capture statistics of local patch similarities is proposed in [34]. Enhancing the face recognition performance by introducing the discriminative learning into three steps of LBP-like feature extraction is presented in [71].

The discriminant image filters, the optimal soft sampling matrix and the dominant patterns are all learned from images. The general advantage of these methods is compact, highly discriminative and easy to extract learning-based descriptor. These methods are discriminative and robust to illumination and expression changes.

2.5. 3D-based face recognition

As 3D capturing process is becoming cheaper and faster [72], it is commonly thought that the use of 3D sensing has the potential for greater recognition accuracy than 2D. The advantage behind using 3D data is that depth information does not depend on pose and illumination, and therefore, the representation of the object does not change with these parameters, making the whole system more robust. 3D-based techniques can achieve better robustness to pose variation problem than 2D-based ones. A comprehensive survey of the 3D face recognition approaches is presented in [73].

A method for face recognition across variations in pose, which combines deformable 3D models with a computer graphics simulation of projection and illumination, can be found in [74]. In this method, faces are represented by model parameters for 3D shape and texture. Their 3D morphable models are combined with spherical harmonics illumination representation [75] to recognise faces under arbitrary unknown lighting.

Using facial symmetry to handle pose variation in 3D face recognition is presented in [76], where an automatic landmark detector is used. It helps to estimate pose and detects occluded areas for each facial scan. Subsequently, an annotated face model is registered and fitted to the scan. During fitting, facial symmetry is used to overcome the challenges of missing data [77].

There is a generic 3D elastic model for pose invariant face recognition proposed in [29]. It is constructed for each subject in the database using only a single 2D image by applying the 3D generic elastic model (3DGEM) approach. Each 3D model is subsequently rendered at different poses within a limited search space about the estimated pose, and the resulting images are matched against the test query. Finally, the distances between the synthesised images and test query are computed by using a simple normalised correlation matcher to show the effectiveness of the pose synthesis method to real-world data.

In [78], a geometric framework for analysing 3D faces, with the specific goals of comparing, matching and averaging their shapes, is proposed to represent facial surfaces by radial curves emanating from the nose tips.

3D face recognition approach based on local geometrical signatures called facial angular radial signature (ARS) that can approximate the semi-rigid region of the 3D face is proposed in [79].

The authors employed KPCA to map the raw ARS facial features to mid-level features to improve the discriminating power. Finally, the resulting mid-level features are combined into one single feature vector and fed into the SVM to perform face recognition [80, 81, 82, 83, 84, 85, 86].

The drawback of using 3D data in face recognition is that these face recognition approaches need all the elements of the system to be well calibrated and synchronised to acquire accurate 3D data (texture and depth maps). The existing 3D face recognition approaches rely on a surface registration or on complex feature (surface descriptor) extraction and matching techniques. They are, therefore, computationally expensive and not suitable for practical applications. Moreover, they require the cooperation of the subject making them not useful for uncontrolled or semi-controlled scenarios where the only input of the algorithms will be a 2D intensity image acquired from a single camera.

2.6. Video-based face recognition

The analysis of video streams of face images has received increasing attention in biometrics [87]. An immediate advantage in using video information is the possibility of employing redundancy present in the video sequence to improve still image systems. Although significant amount of research has been done in matching still face images, the use of videos for face recognition is relatively less explored [88]. The first stage of video-based face recognition (VFR) is to perform re-identification, where a collection of videos is cross-matched to locate all occurrences of the person of interest [89].

Generally, VFR approaches can be classified into two categories based on how they leverage the multitude of information available in a video sequence: (i) sequence based and (ii) set based, where at a high level, what most distinguishes these two approaches is whether or not they utilise temporal information [90, 91].

The formulation of a probabilistic appearance-based face recognition approach is extended in [92]. Originally, it was defined to do recognition from a single still image as previously explained, to work with multiple images and video sequences. In [93], there is the constrained subspace spanned from face images of a clip into a convex hull and then calculate the nearest distance of two convex hulls as the between-set similarity. Thus, each test and training example is a set of images of a subject's face, not just a single image, so recognition decisions need to be based on comparisons of image sets.

In [94], VFR task is converted into the problem of measuring the similarity of two image sets, where the examples from a video clip construct one image set. The authors consider face images from each clip as an ensemble and formulate VFR into the joint sparse representation (JSR) problem. In JSR, to adaptively learn the sparse representation of a probe clip, they simultaneously consider the class-level and atom-level sparsity, where the former structures the enrolled clips using the structured sparse regulariser and the latter seeks for a few related examples using the sparse regulariser.

In order to identify the most important advantages and imperfections, discussed above methods are summarised in **Table 1**.

No.	Method	Advantages	Disadvantages
1.	Classical face recognition algorithms	Focuses on local structure of the manifold. These methods project face onto linear subspace spanned by the eigenface images. The distance from face space is orthogonal to the plane of mean image, so may be easily turned to Mahalanobis distances with probabilistic interpretation	These methods may fail to adequately represent faces when large variations in illumination facial expressions and other factors occur. Regarding to [34], applying kernel-based nonlinear methods do not produce a significant improvement comparing to linear methods. LLE, LLP and LBP brought simple and effective way to describe neighbouring changes in face description. Subspace approaches were applied in DCV- and SVM-based methods. Preserving the local structure between samples is the domain of NPP and ONPP methods. The problem is that it is still unclear how to select the neighbourhood size or assign optimal values for them
2.	Artificial neural networks	Radial basis function artificial neural network is naturally integrated with non-negative matrix factorisation. Also other approaches for process simplification regarding to ANNs native linearisation feature and computation speed up. Ideal solution, especially for recognising face images with partial distortion and occlusion	The main disadvantage of this approach is requirement of greater number of training samples (instead one or limited number). It is inaccurate in the same way like other statistically based methods
3.	Gabor wavelets	The Gabor wavelets exhibit desirable characteristics of capturing salient visual properties like spatial localisation orientation selectivity and spatial frequency. Different biometrics applications favour this approach	The drawback of the Gabor-based methods is significantly high dimensionality of the Gabor feature space since face image is convolved with a bank of Gabor filters. Approach is computationally intensive and impractical for real-time applications. Additionally, simplified Gabor features are sensitive to lightning variations

No.	Method	Advantages	Disadvantages
4.	Face descriptor-based methods	The main idea behind developing image descriptors is to learn the most discriminant local features that minimise difference between images of the same individual and maximise that between images from the other people. These methods are discriminative and robust to illumination and expression changes. They offer compact, easy to extract and highly discriminative descriptor	Approach is computationally intensive during descriptor extraction stage, but encouraging simplicity and performance in reference to online applications
5.	3D-based face recognition	Extend traditional 2D capturing process and has greater potential for accuracy. The depth information does not depend on the pose and illumination making solution more robust	Require all the elements of the 3D face recognition system to be well calibrated and synchronised to existing 3D data. Computationally expensive and not suitable for practical applications
6.	Video-based recognition	The main advantage of the approach is possibility of employing redundancy present in video to improve still image systems	Relatively poorly investigated. Multiply problems with measuring similarity of two (or more) images

Table 1. Face recognition methods overview.

Methods indicated in the **Table 1** illustrate the evolution of face recognition technology. The huge potential of face descriptor-based methods ought to be emphasised, regarding to the fact the local descriptor idea has been recently recognised as the most crucial design framework for face identification and verification tasks [34].

3. Face recognition applications

Many published works mention numerous applications in which face recognition technology is already utilised including entry to secured high-risk spaces such as border crossings as well as access to restricted resources [95, 96, 97]. On the other hand, there are other application areas in which face recognition has not yet been used. The potential application areas of face recognition technology can be outlined as follows [34]:

- Automated surveillance, where the objective is to recognise and track people [98].

- Monitoring closed circuit television (CCTV), the facial recognition capability can be embedded into existing CCTV networks, to look for lost children or other missing persons or tracking known or suspected criminals.

- Image database investigations, searching image databases of licensed drivers, benefit recipients and finding people in large news photograph and video collections [99, 100], as well as searching in the Facebook social networking web site [101].

- Multimedia environments with adaptive human computer interfaces (part of ubiquitous or context aware systems, behaviour monitoring at childcare or centres for old people, recognising customers and assessing their needs) [102].

- Airplane-boarding gate, the face recognition may be used in places of random checks merely to screen passengers for further investigation. Similarly, in casinos, where strategic design of betting floors that incorporates cameras at face height with good lighting could be used not only to scan faces for identification purposes, but possibly to afford the capture of images to build a comprehensive gallery for future watch-list, identification and authentication tasks [103].

- Sketch-based face reconstruction, where law enforcement agencies in the world rely on practical methods to help crime witnesses reconstruct likenesses of faces [104]. These methods range from sketch artistry to proprietary computerised composite systems [105, 106, 107].

- Forensic applications, where a forensic artist is often used to work with the eyewitness in order to draw a sketch that depicts the facial appearance of the culprit according to his/her verbal description. This forensic sketch is used later for matching large facial image databases to identify the criminals [108, 109]. Yet, there is no existing face recognition system that can be used for identification or verification in crime investigation such as comparison of images taken by CCTV with available database of mugshots. Thus, utilising face recognition technology in the forensic applications is a must as discussed in [110, 111].

- Face spoofing and anti-spoofing, where a photograph or video of an authorised person's face could be used to gain access to facilities or services. Hence, the spoofing attack consists in the use of forged biometric traits to gain illegitimate access to secured resources protected by a biometric authentication system [112, 113]. It is a direct attack to the sensory input of a biometric system, and the attacker does not need previous knowledge about the recognition algorithm. Research on face spoof detection has recently attracted an increasing attention [114], introducing few number of face spoof detection techniques [115, 116, 117]. Thus, developing a mature anti-spoofing algorithm is still in its infancy and further research is needed for face spoofing applications [118, 119].

There have been envisaged many applications for face recognition, but most of commercial ones exploit only superficially the great potential of this technology. Most of the applications are notable limited in their ability to handle pose, lighting changes or aging.

In reference to **access control**, face verification during face-based PC logon has become feasible, but seems to be very limited. Naturally, such PC verification system can be extended in the future for authentic single sign-on to multiple networked services or transaction authorisation or even for access to encrypted files. For example, banking sector is rather conservative in deploying such a biometrics. They estimated high risk in loosing customers disaffected by being falsely rejected than they might gain in fraud prevention. It is the reason for robust passive acquisition systems development with low false rejection.

The most of physical access control systems uses face recognition combination with other biometrics, for example speaker identification and lip motion [120].

One of the most interest in face recognition in application domain is associated with surveillance. Regarding to the generous type of information it contains, video is the medium of choice for surveillance. For applications that require identification, face recognition is the best biometric for video data. The biggest advantage of this approach is passive participation of subject (human). The whole process of recognition and identification can be carried out without the person's knowledge.

Although the development of face recognition surveillance systems has already begun, the technology seems to not accurate enough. It also brings additional problems concerning highly extensive perception in the data gathering and computing side of such complex solutions.

Another future domain, where face recognition is expected to become important, is area of pervasive or ubiquitous computing. Computing devices equipped with sensors become more widespread in reference to together networking. Such approach will allow envisage a future where the most of everyday objects are going to have some computational power, allowing to precisely adapt their behaviour to various factors including time, user, user control or host.

This vision assumes easy information exchange, also including images between devices of different types.

Currently, the most of devices have simple user interface, controlled only by active commands on the part of the user. Some of the devices are able to sense environment and acquire information about the physical word and the people within their region of interest. One of the crucial part of smart devices of human awareness is knowing the identity of the users close to a device, even currently implemented in several smartphones with different results. It is important when contributed with other biometrics regarding to passive nature of face recognition.

4. Conclusion

Face recognition is still a challenging problem in the field of computer vision. It has received a great deal of attention over the past years because of its several applications in various domains. Although there is strong research effort in this area, face recognition systems are far from ideal to perform adequately in all situations form real world. Paper presented a brief

survey of issues methods and applications in area of face recognition. There is much work to be done in order to realise methods that reflect how humans recognise faces and optimally make use of the temporal evolution of the appearance of the face for recognition.

Author details

Waldemar Wójcik[1], Konrad Gromaszek[1*] and Muhtar Junisbekov[2]

*Address all correspondence to: k.gromaszek@pollub.pl

1 Institute of Electronics and Information Technology, Lublin University of Technology, Lublin, Poland

2 Taraz State University, Taraz, Jambyl, Kazakhstan

References

[1]   Lin S.: 'An introduction to Face Recognition Technology', Informing Science, 2000, 3, pp.1-6.

[2]   An, L., Kafai, M., Bhanu, B.: 'Dynamic Bayesian network for unconstrained face recognition in surveillance camera networks', IEEE J. Emerg. Sel. Top. Circuits Syst., 2013, 3, (2), pp. 155–164.

[3]   Philips, P. J., Moon H., Rauss P., Rizivi S.: 'The FERET September 1996 Database and Evaluation Procedure', Audio- and Video-based Biometric Person Authentication, Lecture Notes in Computer Science, vol. 1206, 395-402, Springer 1997.

[4]   Liao, S., Lei, Z., Yi, D., Li, S.: 'A benchmark study of large-scale unconstrained face recognition'. Int. Joint Conf. on Biometrics (IJCB 2014), Florida, USA, 2014, pp. 1–8.

[5]   Turk, M., Pentland, A.: 'Eigenfaces for recognition', J. Cogn. Neurosci., 1991, 3, (1), pp. 71–86.

[6]   Bartlett, M.S., Movellan, J.R., Sejnowski, T.J.: 'Face recognition by independent component analysis', IEEE Trans. Neural Netw., 2002, 13, (6), pp. 1450–1464.

[7]   Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: 'Eigenfaces vs. _sherfaces: recognition using class speci_c linear projection', IEEE Trans. Pattern Anal. Mach. Intell., 1997, 19, (7), pp. 711–720.

[8]   Hu, H., Zhang, P., De la Torre, F.: 'Face recognition using enhanced linear discriminant analysis', IET Comput. Vis., 2010, 4, (3), pp. 195–208.

[9] Yang, J., Zhang, D., Frangi, A.F., Yang, J.-Y.: 'Two-dimensional PCA: a new approach to appearance-based face representation and recognition', IEEE Trans. Pattern Anal. Mach. Intell., 2004, 26, (1), pp. 131–137.

[10] Naseem, I., Togneri, R., Bennamoun, M.: 'Linear regression for face recognition', IEEE Trans. Pattern Anal. Mach. Intell., 2010, 32, (11), pp. 2106–2112.

[11] Lu, J., Plataniotis, K.N., Venetsanopoulos, A.N.: 'Face recognition using kernel direct discriminant analysis algorithms', IEEE Trans. Neural Netw., 2003, 14, (1), pp. 117–126.

[12] He, X., Yan, S., Hu, Y., Niyogi, P., Zhang, H.-J.: 'Face recognition using Laplacian faces', IEEE Trans. Pattern Anal. Mach. Intell., 2005, 27, (3), pp. 328–340.

[13] Perlibakas, V.: 'Distance measures for PCA-based face recognition', Pattern Recognit. Lett., 2004, 25, (6), pp. 711–724.

[14] Kim, K.I., Jung, K., Kim, H.J.: 'Face recognition using kernel principal component analysis', IEEE Signal Process. Lett., 2002, 9, (2), pp. 40–42.

[15] Bach, F., Jordan, M.: 'Kernel independent component analysis', J. Mach. Learn. Res., 2003, 3, pp. 1–48.

[16] Ji, S., Ye, J.: 'Generalized linear discriminant analysis: a uni_ed framework and ef_cient model selection', IEEE Trans. Neural Netw., 2008, 9, (10), pp. 1768–1782.

[17] Roweis, S., Saul, L.: 'Nonlinear dimensionality reduction by locally linear embedding', Science, 2000, 290, (5500), pp. 2323–2326.

[18] Xiaofei, H., Partha, N.: 'Locality preserving projections'. Int. Conf. on Advances in Neural Information Processing Systems (NIPS'03), 2003, pp. 153–161.

[19] Ahonen, T., Hadid, A., Pietikäinen, M.: 'Face description with local binary patterns: application to face recognition', IEEE Trans. Pattern Anal. Mach. Intell., 2006, 28, (12), pp. 2037–2041.

[20] Suruliandi, A., Meena, K., Reena, R.R.: 'Local binary pattern and its derivatives for face recognition', IET Comput. Vis., 2012, 6, (5), pp. 480–488.

[21] Zhang, B., Shan, S., Chen, X., Gao, W.: 'Histogram of Gabor phase patterns (HGPP): a novel object representation approach for face recognition', IEEE Trans. Image Process., 2007, 16, (1), pp. 57–68.

[22] Yang, B., Chen, S.: 'A comparative study on local binary pattern (LBP) based face recognition: LBP histogram versus LBP image', Neurocomputing, 2013, 22, pp. 620–627.

[23] Zhang, B., Gao, Y., Zhao, S., Liu, J.: 'Local derivative pattern versus local binary pattern: face recognition with high-order local pattern descriptor', IEEE Trans. Image Process., 2010, 19, (2), pp. 533–544.

[24] Cevikalp, H., Neamtu, M., Wilkes, M., Barkana, A.: 'Discriminative common vectors for face recognition', IEEE Trans. Pattern Anal. Mach. Intell., 2005, 27, (1), pp. 4–13.

[25] Jing, X.-Y., Yao, Y.-F., Yang, J.-Y., Zhang, D.: 'A novel face recognition approach based on kernel discriminative common vectors (KDCV) feature extraction and RBF neural network', Neurocomputing, 2008, 71, pp. 3044–3048.

[26] Wen, Y.: 'An improved discriminative common vectors and support vector machine based face recognition approach', Expert Syst. Appl., 2012, 39, (4), pp. 4628–4632.

[27] Kokiopoulou, E., Saad, Y.: 'Orthogonal neighborhood preserving projections: a projection based dimensionality reduction technique', IEEE Trans. Pattern Anal. Mach. Intell., 2007, 29, (12), pp. 2143–2156.

[28] [Yanwei, P., Lei, Z., Zhengkai, L., Nenghai, Y., Houqiang, L.: 'Neighborhood preserving projections (NPP): a novel linear dimension reduction method', Lect. Notes Comput. Sci., 2005, 3644, pp. 117–125.

[29] Prabhu, U., Jingu, H., Marios, S.: 'Unconstrained pose-invariant face recognition using 3D generic elastic models', IEEE Trans. Pattern Anal. Mach. Intell., 2011, 33, (10), pp. 1952–1961.

[30] Qiao, L., Chena, S., Tan, X.: 'Sparsity preserving projections with applications to face recognition', Pattern Recognit., 2010, 43, (1), pp. 331–341.

[31] Jiwen, L., Yap-Peng, T.: 'Regularized locality preserving projections and its extensions for face recognition', IEEE Trans. Syst. Man Cybern. B, Cybern., 2009, 40, (3), pp. 1083–4419.

[32] Lu, J., Plataniotis, K.N., Venetsanopoulos, A.N., Stan, Z.L.: 'Ensemble-based discriminant learning with boosting for face recognition', IEEE Trans. Neural Netw., 2006, 17, (1), pp. 166–178.

[33] Abeer, A.M., Woo, W.L., Dlay, S.S.: 'Multi-linear neighborhood preserving projection for face recognition', Pattern Recognit., 2014, 47, (2), pp. 544–555.

[34] Hassaballah M., Aly S.: 'Face recognition: challenges, achievements and future directions', IET Computer Vision, 2015, Vol. 9, Iss. 4, pp. 614–626.

[35] Lu, H., Plataniotis, K.N., Venetsanopoulos, A.N.: 'A survey of multilinear subspace learning for tensor data', Pattern Recognit., 2011, 44, (7), pp. 1540–1551.

[36] Pang, S., Kim, D., Bang, S.Y.: 'Face membership authentication using SVM classi_cation tree generated by membership-based LLE data partition', IEEE Trans. Neural Netw., 2005, 16, (2), pp. 436–446.

[37] Zhang, B., Zhang, H., Ge, S.: 'Face recognition by applying wavelet subband representation and kernel associative memory', IEEE Trans. Neural Netw., 2004, 15, pp. 166–177.

[38] Li, G., Zhang, J., Wang, Y., Freeman, W.J.: 'Face recognition using a neural network simulating olfactory systems', Lect. Notes Comput. Sci., 2006, 3972, pp. 93–97.

[39] Zhou, W., Pu, X., Zheng, Z.: 'Parts-based holistic face recognition with RBF neural networks', Lect. Notes Comput. Sci., 2006, 3972, pp. 110–115.

[40] Park, C., Ki, M., Namkung, J., Paik, J.K.: 'Multimodal priority veri_cation of face and speech using momentum back-propagation neural network', Lect. Notes Comput. Sci., 2006, 3972, pp. 140–149.

[41] Bhavin, J.S., Martin, D.L.: 'Face recognition using localized features based on nonnegative sparse coding', Mach. Vis. Appl., 2007, 18, (2), pp. 107–122.

[42] Jiwen, L., Yap-Peng, T., Gang, W.: 'Discriminative multimanifold analysis for face recognition from a single training sample per person', IEEE Trans. Pattern Anal. Mach. Intell., 2013, 35, (1), pp. 39–51.

[43] Singh, R., Vatsa, M., Noore, A.: 'Face recognition with disguise and single gallery images', Image Vis. Comput., 2009, 72, (3), pp. 245–257.

[44] Gu, W., Xiang, C., Venkatesh, Y., Huang, D., Lin, H.: 'Facial expression recognition using radial encoding of local Gabor features and classi_er synthesis', Pattern Recognit., 2012, 45, (1), pp. 80–91.

[45] Serrano, S., Diego, I., Conde, C., Cabello, E.: 'Recent advances in face biometrics with Gabor wavelets: a review', Pattern Recognit. Lett., 2010, 31, (5), pp. 372–381.

[46] Shen, L., Bai, L.: 'A review on Gabor wavelets for face recognition', Pattern Anal. Appl., 2006, 9, (2–3), pp. 273–292.

[47] Zhang, W.C., Shan, S.G., Chen, X.L., Gao, W.: 'Are Gabor phases really useless for face recognition?', Int. J. Pattern Anal. Appl., 2009, 12, (3), pp. 301–307.

[48] Zhao, S., Gao, Y., Zhang, B.: 'Gabor feature constrained statistical model for ef_cient landmark localization and face recognition', Pattern Recognit. Lett., 2009, 30, (10), pp. 922–930.

[49] Kanan, H., Faez, K.: 'Recognizing faces using adaptively weighted sub-Gabor array from a single sample image per enrolled subject', Image Vis. Comput., 2010, 28, (3), pp. 438–448.

[50] Yu, L., He, Z., Cao, Q.: 'Gabor texture representation method for face recognition using the Gamma and generalized Gaussian models', Image Vis. Comput., 2010, 28, (1), pp. 177–187.

[51] Nanni, L., Maio, D.: 'Weighted sub-Gabor for face recognition', Pattern Recognit. Lett., 2007, 28, (4), pp. 487–492.

[52] Zhao, Z.-S., Zhang, L., Zhao, M., Hou, Z.-G., Zhang, C.-S.: 'Gabor face recognition by multi-channel classi_er fusion of supervised kernel manifold learning', Neurocomputing, 2012, 97, pp. 398–404.

[53] Zhang, B., Shan, S., Chen, X., Gao, W.: 'Histogram of Gabor phase patterns: a novel object representation approach for face recognition', IEEE Trans. Image Process., 2007, 16, (1), pp. 57–68.

[54] Xie, S., Shan, S., Chen, X., Chen, J.: 'Fusing local patterns of Gabor magnitude and phase for face recognition', IEEE Trans. Image Process., 2010, 19, (5), pp. 1349–1361.

[55] Xu, Y., Li, Z., Pan, J.-S., Yang, J.-Y.: 'Face recognition based on fusion of multi-resolution Gabor features', Neural Comput. Appl., 2013, 23, (5), pp. 1251–1256.

[56] Chai, Z., Sun, Z., Mndez-Vzquez, H., He, R., Tan, T.: 'Gabor ordinal measures for face recognition', IEEE Trans. Inf. Forensics Sec., 2014, 9, (1), pp. 14–26.

[57] Liu, C., Wechsler, H.: 'Gabor feature based classi_cation using the enhanced _sher linear discriminant model for face recognition', IEEE Trans. Image Process., 2002, 11, (4), pp. 467–476.

[58] Liu, C., Wechsler, H.: 'Independent component analysis of Gabor features for face recognition', IEEE Trans. Neural Netw., 2003, 14, (4), pp. 919–928.

[59] Ren, C.-X., Dai, D.-Q., Li, X., Lai, Z.-R.: 'Band-reweighed Gabor kernel embedding for face image representation and recognition', IEEE Trans. Image Process., 2014, 32, (2), pp. 725–740.

[60] Yang, M., Zhang, L., Shiu, S., Zhang, D.: 'Gabor feature based robust representation and classi_cation for face recognition with Gabor occlusion dictionary', Pattern Recognit., 2013, 46, (7), pp. 1865–1878.

[61] Serrano, A., de Diego, I., Conde, C., Cabello, E.: 'Analysis of variance of Gabor _lter banks parameters for optimal face recognition', Pattern Recognit. Lett., 2011, 32, (15), pp. 1998–2008.

[62] Perez, C., Cament, L., Castillo, L.E.: 'Methodological improvement on local Gabor face recognition based on feature selection and enhanced Borda count', Pattern Recognit., 2011, 44, (4), pp. 951–963.

[63] Oh, J., Choi, S., Kimc, C., Cho, J., Choi, C.: 'Selective generation of Gabor features for fast face recognition on mobile devices', Pattern Recognit. Lett., 2013, 34, (13), pp. 1540–1547.

[64] Choi, W.-P., Tse, S.-H., Wong, K.-W., Lam, K.-M.: 'Simpli_ed Gabor wavelets for human face recognition', Pattern Recognit., 2008, 41, (3), pp. 1186–1199.

[65] Chen, J., Shan, S., He, C., et al.: 'WLD: a robust local image descriptor', IEEE Trans. Pattern Anal. Mach. Intell., 2010, 32, (9), pp. 1705–1720.

[66] Jabid, T., Kabir, M., Chae, O.: 'Facial expression recognition using local directional pattern (LDP)'. IEEE Int. Conf. on Image Processing (ICIP), Hong Kong, China, 2010, pp. 1605–1608.

[67] Bereta, M., Karczmarek, P., Pedrycz, W., Reformat, M.: 'Local descriptors in application to the aging problem in face recognition', Pattern Recognit., 2013, 46, (10), pp. 2634–2646.

[68] Bereta, M., Pedrycz, W., Reformat, M.: 'Local descriptors and similarity measures for frontal face recognition: a comparative analysis', J. Vis. Commun. Image Represent., 2013, 24, (8), pp. 1213–1231.

[69] Cao, Z., Yin, Q., Tang, X., Sun, J.: 'Face recognition with learning-based descriptor'. IEEE Conf. on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 2010, pp. 2707–2714.

[70] Farajzadeh, N., Faez, K., Pan, G.: 'Study on the performance of moments as invariant descriptors for practical face recognition systems', IET Comput. Vis., 2011, 4, (4), pp. 272–285.

[71] Lei, Z., Pietikäinen, M., Stan, Z.L.: 'Learning discriminant face descriptor', IEEE Trans. Pattern Anal. Mach. Intell., 2014, 36, (2), pp. 289–302.

[72] Ira, K.-S., Ronen, B.: '3D face reconstruction from a single image using a single reference face shape', IEEE Trans. Pattern Anal. Mach. Intell., 2011, 33, (2), pp. 394–405.

[73] Bowyer, K., Chang, K.P., Flynn, P.: 'A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition', Comput. Vis. Image Underst., 2006, 101, (1), pp. 1–15.

[74] Blanz, V., Vetter, T.: 'Face recognition based on _tting a 3D morphable model', IEEE Trans. Pattern Anal. Mach. Intell., 2003, 25, (9), pp. 1063–1074.

[75] Zhang, L., Samaras, D.: 'Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics', IEEE Trans. Pattern Anal. Mach. Intell., 2006, 28, (3), pp. 351–363.

[76] Passalis, G., Panagiotis, P., Theoharis, T., Kakadiaris, I.A.: 'Using facial symmetry to handle pose variations in real-world 3D face recognition', IEEE Trans. Pattern Anal. Mach. Intell., 2011, 33, (10), pp. 1938–1951.

[77] Lei, Y., Bennamoun, M., El-Sallam, A.: 'An ef_cient 3D face recognition approach based on the fusion of novel local low-level features', Pattern Recognit., 2013, 46, (1), pp. 24–37.

[78] Drira, H., Ben Amor, B., Srivastava, A., Daoudi, M., Slama, R.: '3D face recognition under expressions, occlusions and pose variations', IEEE Trans. Pattern Anal. Mach. Intell., 2013, 35, (9), pp. 2270–2283.

[79] Lei, Y., Bennamoun, M., Hayat, M., Guo, Y.: 'An ef_cient 3D face recognition approach using local geometrical signatures', Pattern Recognit., 2014, 47, (2), pp. 509–524.

[80] Andrea, F.A., Michele, N., Daniel, R., Gabriele, S.: '2D and 3D face recognition: a survey', Pattern Recognit. Lett., 2007, 28, (14), pp. 1885–1906.

[81] Cai, L., Da, F.: 'Estimating inter-personal deformation with multi-scale modelling between expression for three-dimensional face recognition', IET Comput. Vis., 2012, 6, (5), pp. 468–479.

[82] Chen, Q., Yao, J., Cham, W.K.: '3D model-based pose invariant face recognition from multiple views', IET Comput. Vis., 2007, 1, (1), pp. 25–34.

[83] Lu, X., Jain, A.K.: 'Matching 2.5D face scans to 3D models', IEEE Trans. Pattern Anal. Mach. Intell., 2006, 28, (1), pp. 31–43.

[84] Al-Osaimi, F., Bennamoun, M., Mian, A.: 'An expression deformation approach to non-rigid 3D face recognition', Int. J. Comput. Vis., 2009, 81, (3), pp. 302–316.

[85] Bronstein, A.M., Bronstein, M.M., Kimmel, R.: 'Three-dimensional face recognition', Int. J. Comput. Vis., 2005, 64, (1), pp. 5–30.

[86] Chang, K., Bowyer, K., Flynn, P.: 'An evaluation of multimodal 2D + 3D face biometrics', IEEE Trans. Pattern Anal. Mach. Intell., 2005, 27, (4), pp. 619–624.

[87] Marin-Jimenez, M., Zisserman, A., Eichner, M., Ferrari, V.: 'Detecting people looking at each other in videos', Int. J. Comput. Vis., 2014, 106, (3), pp. 282–296.

[88] O'Toole, A., Harms, J., Snow, S., Hurst, D., Pappas, M., Abdi, H.: 'A video database of moving faces and people', IEEE Trans. Pattern Anal. Mach. Intell., 2005, 27, (5), pp. 812–816.

[89] Poh, N., Chan, C.H., Kittler, J., et al.: 'An evaluation of video-to-video face veri_cation', IEEE Trans. Inf. Forensics Sec., 2010, 24, (8), pp. 781–801.

[90] Best-Rowden, L., Klare, B., Klontz, J., Jain, A.: 'Video-to-video face matching: establishing a baseline for unconstrained face recognition'. Biometrics: Theory, Applications and Systems (BTAS), Washington DC, USA, 2013.

[91] Barr, J., Boyer, K., Flynn, P., Biswas, S.: 'Face recognition from video: a review', Int. J. Pattern Recognit. Artif. Intell., 2012, 26, (5), pp. 53–74.

[92] Zhang, Y., Martinez, A.: 'A weighted probabilistic approach to face recognition from multiple images and video sequences', Image Vis. Comput., 2006, 24, (6), pp. 626–638.

[93] Cevikalp, H., Triggs, B.: 'Face recognition based on image sets'. IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR'10), San Francisco, CA, USA, 2010, pp. 2567–2573.

[94] Cui, Z., Chang, H., Shan, S., Ma, B., Chen, X.: 'Joint sparse representation for video-based face recognition', Neurocomputing, 2014, 135, (5), pp. 306–312.

[95] Anil, K., Arun, A., Karthik, N.: 'Introduction to biometrics' (Springer, New York, USA, 2011.

[96] Phillips, P.J., Flynn, P.J., Scruggs, T., et al.: 'Overview of the face recognition grand challenge'. IEEE Conf. Computer Vision and Pattern Recognition, San Diego, CA, USA, 2005, pp. 947–954.

[97] Stan, Z.L., Jain, A.: 'Handbook of face recognition' (Springer, New York, USA, 2005).

[98] Kamgar-Parsi, B., Lawson, W., Kamgar-Parsi, B.: 'Toward development of a face recognition system for watchlist surveillance', IEEE Trans. Pattern Anal. Mach. Intell., 2011, 33, (10), pp. 1925–1937.

[99] Ortiz, E., Becker, B.: 'Face recognition for web-scale datasets', Comput. Vis. Image Underst., 2013, 108, pp. 153–170.

[100] Ozkana, D., Duygulu, P.: 'Interesting faces: a graph-based approach for finding people in news', Pattern Recognit., 2010, 43, (5), pp. 1717–1735.

[101] Pinto, N., Stone, Z., Zickler, T., Cox, D.: 'Scaling up biologically-inspired computer vision: a case study in unconstrained face recognition on Facebook'. IEEE Computer Vision and Pattern Recognition, Workshop on Biologically Consistent Vision, Colorado Springs, USA, 2011, pp. 35–42.

[102] Best-Rowden, L., Han, H., Otto, C., Klare, B., Jain, A.: 'Unconstrained face recognition: identifying a person of interest from a media collection'. Technical Report, Technical Report MSU-CSE-14-1, Michigan State University, 2014.

[103] Introna, L., Wood, D.: 'Picturing algorithmic surveillance: the politics of facial recognition systems', Surveillance Soc., 2004, 2, (2/3), pp. 177–198.

[104] Han, H., Klare, B., Bonnen, K., Jain, A.: 'Matching composite sketches to face photos: a component-based approach', IEEE Trans. Inf. Forensics Sec., 2013, 8, (1), pp. 191–204.

[105] Gao, X., Zhong, J., Tian, C.: 'Sketch synthesis algorithm based on E-Hmm and selective ensemble', IEEE Trans. Circuits Syst. Video Technol., 2008, 18, (4), pp. 487–496.

[106] Tang, X., Wang, X.: 'Face photo-sketch synthesis and recognition', IEEE Trans. Pattern Anal. Mach. Intell., 2009, 31, (11), pp. 1955–1967.

[107] Tang, X., Wang, X.: 'Face sketch recognition', IEEE Trans. Circuits Syst. Video Technol., 2004, 14, (1), pp. 50–57.

[108] Brendan, F.K., Zhifeng, L., Anil, K.J.: 'Matching forensic sketches to mug shot photos', IEEE Trans. Pattern Anal. Mach. Intell., 2011, 33, (3), pp. 639–646.

[109] Klum, S., Han, H., Jain, A., Klare, B.: 'Sketch based face recognition: forensic vs. composite sketches'. Sixth IAPR Int. Conf. Biometrics (ICB'13), Madrid, Spain, 2013, pp. 1–8.

[110] Jain, A., Klare, B., Park, U.: 'Face matching and retrieval in forensics applications', IEEE Multimedia, 2012, 19, (1), pp. 20–28.

[111] Jain, A.K., Klare, B., Park, U.: 'Face recognition: some challenges in forensics'. IEEE Int. Conf. on Automatic Face Gesture Recognition and Workshops (FG 2011), Santa Barbara, CA, USA, 2011, pp. 726–733.

[112] Erdogmus, N., Marcel, S.: 'Spoo_ng in 2D face recognition with 3D masks and anti-spoo_ng with kinect'. The IEEE Sixth Int. Conf. Biometrics: Theory, Applications and Systems (BTAS 2013), Washington, DC, USA, 2013, pp. 1–6.

[113] Marcel, S., Nixon, M., Li, S.: 'Handbook of biometric anti-spoo_ng: trusted biometrics under spoo_ng attacks' (Springer, New York, USA, 2014).

[114] Määttä, J., Hadid, A., Pietikäinen, M.: 'Face spoofing detection from single images using texture and local shape analysis', IET Biometrics, 2012, 1, (1), pp. 3–10.

[115] Chingovska, I., Anjos, A., Marcel, S.: 'On the effectiveness of local binary patterns in face anti-spoo_ng'. IEEE Int. Conf. Biometrics Special Interest Group (BIOSIG), Darmstadt, Germany, 2012, pp. 1–7.

[116] de Freitas Pereira, T., Anjos, A., De Martino, J., Marcel, S.: 'LBP-TOP based counter-measure against face spoo_ng attacks'. Int. Workshop on Computer Vision with Local Binary Pattern Variants (ACCV), Daejeon, Korea, 2012, pp. 121–132.

[117] Zhang, Z., Yan, J., Liu, S., Lei, Z., Yi, D., Li, S.Z.: 'A face antispoo_ng database with diverse attacks'. Fifth IAPR Int. Conf. on Biometrics (ICB), New Delhi, India, 2012, pp. 26–31.

[118] Chingovska, I., Rabello dos Anjos, A., Marcel, S.: 'Biometrics evaluation under spoo_ng attacks', IEEE Trans. Inf. Forensics Sec., 2014, 9, (12), pp. 2264–2276.

[119] Hadid, A.: 'Face biometrics under spoo_ng attacks: vulnerabilities, countermeasures, open issues, and research directions'. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 2014, pp. 113–118.

[120] Orubeondo A. : 'A New Face for Security', InfoWorld.com.

A Generally Semisupervised Dimensionality Reduction Method with Local and Global Regression Regularizations for Recognition

Mingbo Zhao, Yuan Gao, Zhao Zhang and Bing Li

Additional information is available at the end of the chapter

Abstract

The insufficiency of labeled data is an important problem in image classification such as face recognition. However, unlabeled data are abundant in the real-world application. Therefore, semisupervised learning methods, which corporate a few labeled data and a large number of unlabeled data into learning, have received more and more attention in the field of face recognition. During the past years, graph-based semisupervised learning has been becoming a popular topic in the area of semisupervised learning. In this chapter, we newly present graph-based semisupervised learning method for face recognition. The presented method is based on local and global regression regularization. The local regression regularization has adopted a set of local classification functions to preserve both local discriminative and geometrical information, as well as to reduce the bias of outliers and handle imbalanced data; while the global regression regularization is to preserve the global discriminative information and to calculate the projection matrix for out-of-sample extrapolation. Extensive simulations based on synthetic and real-world datasets verify the effectiveness of the proposed method.

Keywords: Semi-supervised Learning, Dimensionality Reduction, Local and Global Regressions, Face Recognition, Transductive and Inductive Learning

1. Introduction

In the real world, there are ever-increasing vision face data generated from Internet surfing and daily social communication. These metadata can be labeled or unlabeled, and accordingly be

utilized for image retrieval, summarization, and indexing. To handle these datasets for realizing the above tasks, automatic annotation is an elementary step, which can be formulated as a pattern classification problem and accomplished by learning-based techniques. Traditionally, the supervised-learning-based methods, such as Linear discriminant analysis (LDA) and Support Vector Machine (SVM), can deliver satisfactory recognition accuracy given that the number of labeled data is adequate. But labeling a huge amount of data is expensive and time consuming. On the other hand, the unlabeled data are sufficient and can be easily obtained from real-world application. Therefore, semisupervised learning-based methods that utilize a few of labeled data and a huge amount of unlabeled data are becoming more and more popular than only relying on the supervised learning methods [27–33].

Recently, since two pioneer semisupervised methods, i.e., Gaussian Fields and Harmonic Functions (GFHF) and Learning with Local and Global Consistency (LLGC), have been proposed in 2003 and 2004, respectively, graph-based semisupervised learning methods have received considerable research interest in the area of semisupervised learning. These methods usually represent both labeled and unlabeled sets by a graph, and then utilize their graph Laplacian matrix to characterize the manifold structure. Finally, different learning tasks such as image classification, clustering, and dimensionality reduction are performed on the graph Laplacian matrix. For example, GFHF and LLGC work in a transductive way by directly propagating the class label information from the labeled set to the unlabeled set along the graph, where the labels of unlabeled data can be estimated. Other similar works include Random Walk [5] and Special Label Propagation (SLP) [8]. However, the transductive learning methods cannot predict the class labels of new-coming samples, hence suffering the out-of-sample problem.

To solve the out-of-sample problem, inductive learning methods are proposed during the past decades. Typical methods for inductive learning are Manifold Regularization (MR) [1] and Semisupervised Discriminant Analysis (SDA) [2]. The MR tries to learn a projection matrix by adding the graph Laplaican regularized term to the cost function of original supervised methods. Therefore, both unlabeled and new-coming data can be cast into a low-dimensional subspace, hence the out-of-sample problem can be naturally solved [7, 9, 10, 16]. For example, MR has extended the regularized least square and SVM to their semisupervised learning extensions, i.e., Laplacian regularized least squares (Lap-RLS) and Laplacian SVM by adding a manifold regularized term. Similarly, Cai et al. [2] have extended LDA to SDA for semisupervised dimensionality reduction.

It should be noted that the success of semisupervised learning is based on how to utilizing the unlabeled data for characterizing the distribution of labels in data space. Several methods including Locally Linear Reconstruction [11, 12, 20], Local Regression and Global Alignment [13, 14], and Local Spline Regression [18, 19] have been developed to discover the intrinsic manifold structure of data. However, when we do semisupervised classification, the data points lying far away the data manifold are noisy for learning the correct classifier and can deteriorate the classification performance. On the other hand, sampling in real-world applications is usually not uniform. As a result, the sampled data may be imbalanced or with multi-density distribution. None of the aforementioned methods focus on solving the two problems.

In this chapter, we develop an effective semisupervised dimensionality reduction method, i.e., Local and Global Regression (LGR), for face recognition with outliers and imbalanced face data. In order to both handle transductive and inductive learning problems, LGR aims to sufficiently learn the classification function by using all data. In detail, the presented method first extends the original supervised regression term to a supervised loss term and a global regression regularized term, where the loss term is to fix the inconsistency between the predicted labels and initial labels, while the global regression term is to sufficiently learn the classification function using all training data and to obtain the projection matrix for handling out-of-sample problem. Furthermore, to capture the local discriminative information, a set of weighted local classification functions are adopted for each dataset to estimate the labels of its nearby data, where the weight is to reduce the outliers bias and to deal with imbalanced data. Thus, both local and global discriminative information of dataset can be preserved by the proposed LGR method.

The main contributions of this work are as follows: (1) we propose a new effective method for semisupervised dimensionality reduction, which can handle both transductive and inductive learning problems; (2) we develop a graph Laplacian matrix, which can characterize both local geometrical and discriminative information, as well as reduce the bias of outliers and handle imbalanced data; (3) we have also established the connection between the proposed method and other state-of-the-art methods. Theoretical analysis has shown that many popular semi-supervised methods such as LRGA can be viewed as the special cases of the proposed method. Extensive simulations based on synthetic and real-world datasets verify the effectiveness of the proposed method.

This chapter is organized as follows. In Section 2, the notations and motivations are first given. We then propose our LGR method for both handling transductive and inductive learning problems. Finally, we also establish the connection between the proposed method and other state-of-the-art methods. Section 3 demonstrates the extensive simulations and the final conclusions are drawn in Section 4.

2. The proposed method

2.1. Notation and motivation

In semi-supervised learning, we define $X = \{X_l, X_u\} = \{x_1, x_2, \ldots, x_{l+u}\} \in R^{D \times (l+u)}$ be the data matrix where the first l and the remaining u columns are the labeled and unlabeled samples, respectively; $Y_l = \{y_1, y_2, \ldots, y_j\} \in R^{c \times l}$ be the binary label matrix with each column y_j repre-senting the class assignment of x_j, i.e. $y_{ij} = 1$, as the class matrix, where $y_{ij} = 1$, if x_j belongs to the ith class; $y_{ij} = 0$, otherwise, D and c are the numbers of features and classes, respectively. We also let $L = D - W$ be the graph Laplacian matrix associated with both labeled and unlabeled sets [17],where W is the weight matrix defined as $w_{ij} = \exp(- \| x_i - x_j \|^2 / 2\sigma^2)$, if x_i is within the k

nearest neighbor of x_j or if x_j is within the k nearest neighbor of x_i; w_{ij}=0, otherwise, D is a diagonal matrix satisfying $D_{ii} = \sum_{j=1}^{l+u} w_{ij}$.

Most semi-supervised learning methods utilize the Gaussian function based affinity matrix. As point out in references [11, 12], the Gaussian function based affinity matrix is found to be oversensitive to the Gaussian variance; only a slight variation on the variance may affect the results dramatically. Thus, Gaussian function based affinity matrix is not a good method for handling image classification. The method developed should be robust to the parameters.

Second, when carrying out semisupervised classification, the samples lying far away from the data manifold are outliers which may lead to learn an incorrect classifier and deteriorate the classification performance. Considering **Figure 1(a and b)** as examples, we generalize a two-cycle and two-moon datasets with outliers. Considering the distribution of two data, the ideal decision boundary should lie in the gap between two data sub-manifolds. However, since there are many outliers around the data manifold, these outliers will blur the clear distribution of the whole data and are noisy to learn a correct classifier. Therefore, it is very important to develop a method that can adaptively reduce the effects of outliers.

Third, in real-world applications, sampling is usually not uniform. Consequently, the sampled data can be imbalanced or follows multi-density distribution. **Figure 1(c)** shows a two-plate dataset with two classes: each class follows a Gaussian distribution but with different cores and density. Obviously, the data points (left data points) in the high-density area will take more important part than those (right data points) in the low-density area when to learn a classifier, which may cause incorrect classification results. The method developed should handle such imbalanced data with multi-density distribution.

The method developed should also solve the out-of-sample problem. To address the above problems, we, in this paper, propose a new semisupervised learning method, which is based on local and global regression.
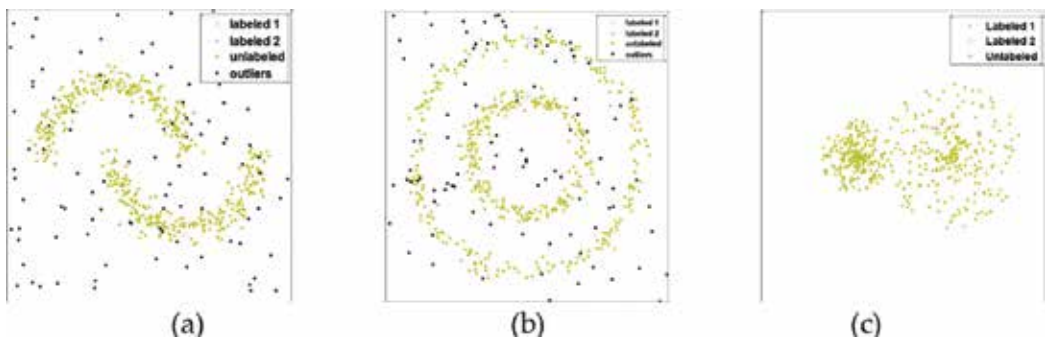

Figure 1. (a) Two-cycle dataset; (b) two-moon dataset; (c) two-plate dataset.

2.2. Local and global regression

We start from the supervised least-squares regression. The least-square regression is to fix a linear model $y_j = V^T x_j + b^T$ by regressing X on Y:

$$\min \sum_{j=1}^{l} \left\| V^T x_j + b^T - y_j \right\|_F^2 + \alpha_t \|V\|_F^2,$$ (1)

where V is the projection matrix that is to project the new-coming samples and b is the bias term. Although the label y_j of x_j ($j \leq l$) has already been known, since l is usually very small, the classification function $z_j = V^T x_j + b$ may not be sufficiently trained due to the small sample size. To solve this problem, we introduce $Z = \{Z_l, Z_u\} = \{z_1, z_2, \ldots, z_{l+u}\} \in R^{c \times (l+u)}$ as a set of estimated labels to play the same roll by replacing $V^T x_j + b$ with z_j and add a regression term to Eq. (**1**) as follows:

$$\min \sum_{i=1}^{l} \|z_i - y_i\|_F^2 + \alpha_r \left(\sum_{j=1}^{l+u} \left\| V^T x_j + b^T - z_j \right\|_F^2 + \eta \|V\|_F^2 \right).$$ (2)

According to Eq. (**2**), the classification function $z_j = V^T x_j + b$ can be sufficiently learned by using all the predicted labels and to fix to their original labels. In other meaning, the global discriminative information can be preserved by the regression term of Eq. (**2**). Furthermore, to grasp the local discriminative information, we induce a local regression function for each data sample x_j. We denote $N_k(x_j)$ as the k neighborhood set of x_j with itself, $X_j = \{x_{j_0}, x_{j_1}, \ldots, x_{j_{k-1}}\} \in R^{D \times k}$ as the local data matrix formed by all samples in $N_k(x_j)$, where $\{j_1, j_1, \ldots, j_k\}$ is the index set of $N_k(x_j)$ and $j_1 = j$, $x_{j_1} = x_j$. We also denote $Z_j = \{z_{j_0}, z_{j_2}, \ldots, z_{j_{k-1}}\} \in R^{c \times k}$ as the local low-dimensional label matrix in $N_k(x_j)$. Then, the local regression function for all data samples can be given as follows:

$$\min_{Z_j, V_j, b_j} \sum_{j=1}^{l+u} \left(\sum_{i=0}^{k-1} \left\| V_j^T x_{j_i} + b_j^T - z_{j_i} \right\|_F^2 + \eta \|V_j\|_F^2 \right).$$ (3)

However, minimizing the above total errors over all data samples tends to force each local error $\alpha_{j_i} = \| V_j^T x_{j_i} + b_j^T - z_{j_i} \|_F$ similar to each other. Given some cases that the dataset includes some outliers, assuming all the local regression errors equally may emphasize the effects from outliers and weaken the effects from normal data. In this section, to weaken the effects from outliers, we add a weight vector $\Gamma_j = \{\tau_{j_1}, \tau_{j_2}, \ldots, \tau_{j_k}\} \in R^{1 \times k}$ for each local data patch x_j in order to penalize each regression error, which can be shown as

$$\min_{Z_j, V_j, b_j} \sum_{j=1}^{l+u} \left(\sum_{i=0}^{k-1} \tau_{j_i} \left\| V_j^T x_{j_i} + b_j^T - z_{j_i} \right\|_F^2 + \eta \left\| V_j \right\|_F^2 \right). \tag{4}$$

In the following section, we will discuss how to select the weight τ_{j_i}. Our motivation is to let the weight of local error α_{j_i} be large given x_{j_i} are the normal data and in the contrast to let the weight be small given x_{j_i} is outlier. In detail, to obtain local projection matrix V_j and bias b_j, we perform derivatives to Eq. (**4**) w.r.t. V_j and b_j to zeros. Then, Eq. (**4**) will be reduced to

$$\min_Z \sum_{j=1}^{l+u} Tr\left(Z S_j L_j S_j^T Z^T \right) = \min_Z Tr\left(Z L_d Z^T \right), \tag{5}$$

where $L_j = H_j - H_j X_j^T (X_j H_j X_j^T + \eta I)^{-1} X_j H_j$; $S_j \in R^{(l+u) \times k}$ is the selected matrix satisfying $(S_j)_{pq} = 1$, if x_p is the qth neighbors to x_p; $(S_j)_{pq} = 0$, otherwise, $L_d = \sum_{j=1}^{l+u} \left(S_j L_j S_j^T \right)$ is the local graph Laplacian matrix. Similarly, by setting the derivatives of Eq. (**2**) w.r.t. V and b to zero, we have

$$\begin{cases} b = \left(e Z^T - e X^T V \right) / e e^T \\ V = \left(X L_c X^T + \eta I \right)^{-1} X L_c Z^T \end{cases}, \tag{6}$$

where $e \in R^{1 \times (l+u)}$ is a unit vector and $L_c = I - e^T e / e e^T$ is used for centering the samples by subtracting the mean of all samples. With b and V in Eq. (**6**), the global regression term in Eq. (**2**) can be written as

$$\left\| V^T X + b^T e - Z \right\|_F^2 + \eta \left\| V \right\|_F^2 = Tr\left(Z L_g Z^T \right), \tag{7}$$

where $L_g = L_c - L_c X^T \left(X L_c X^T + \eta I \right)^{-1} X L_c$ is the global graph Laplacian matrix. By integrating Eq. (**7**) with Eq. (**2**), we formulate our method as follows:

$$J(Z) = \min_Z Tr\left((Z - Y) U (Z - Y)^T \right) + \alpha_m Tr\left(Z L_d Z^T \right) + \alpha_r Tr\left(Z L_g Z^T \right), \tag{8}$$

where $U \in R^{(l+u) \times (l+u)}$ is the diagonal matrix with the first l and the remaining u diagonal elements as 1 and 0, respectively; the second term describes the local discriminative structure of data; the third term describes the global discriminative structure; and α_m and α_r are the two balancing parameters. Since both local and global regressions are regularized in our method,

we refer our method as **LGR**. Finally, by performing derivatives of $J(Z)$ w.r.t. z to zero, we can calculate the solution of z as

$$Z = YU\left(U + \alpha_m L_d + \alpha_r L_g\right)^{-1}. \tag{9}$$

Then, we can obtain the optimal projection matrix and bias term by replacing z in Eq. (**6**).

2.3. Weight selection for bias reduction

In this section, we consider how to select the weights in the proposed method suggested in Section 2.2. Note, our goal of using the weights is to weaken the effects of outliers and the weight τ_{j_i} should be set to a small value if x_{j_i} is an outlier. Then we can make the weight τ_{j_i} inversely proportional to the distance between x_{j_i} and a center μ_j, i.e., $\tau_{j_i} = 1/\parallel x_{j_i} - \mu_j \parallel$. Such a center is expected to represent the idea center of data in the neighborhoods of x_j and should be far away from outliers. Hence, the weight τ_{j_1} is usually small if x_{j_i} is an outlier. But this center μ_j is unknown. We next present an iterative approach to calculate μ_j and the weight τ_{j_1} simultaneously. The approach is converged and proved afterward.

---

1. Initialize μ_j^0 as the average center of all data points in the local patch of x_j.

2. Update $\tau_{j_i}^t$ for each x_{j_i} as $\tau_{j_i} = 1/\parallel x_{j_i} - \mu_j^{t-1} \parallel$ and form the weight matrix Γ_j^t.

3. Update $\mu_j^t = \sum_{i=0}^{k-1} \tau_{j_i}^t x_{j_i} / \sum_{i=0}^{k-1} \tau_{j_i}^t = X_j \Delta_j^t e / e^T \Delta_j^t e$.

4. Iterate steps 2 and 3 until $\sum_{i=0}^{k-1} \parallel x_{j_i} - \mu_j^t \parallel$ no changes. Output $\tau_{j_i}^t$.

Table 1. Iterative approach for calculating the weight.

Table 1 shows the basic steps of the iterative approach. Following **Table 1**, the weight $\tau_{j_i}^t$ at each iteration is updated from the last μ_j^{t-1} and the newly updated center μ_j^t is calculated from current $\tau_{j_i}^t$. The whole iterations are continued until convergence, so that the weight $\tau_{j_i}^t$ can be adaptively and iteratively re-weighted to minimize $\sum_{i=0}^{k-1} \parallel x_{j_i} - \mu_j^t \parallel$. In addition, as can be seen in simulation of **Figure 2**, the updated μ_j^t will be adaptively re-weighted to be close to the main center of most data points, while the updated $\tau_{j_i}^t$ will be weaken if x_{j_i} is outliers or be strengthened if x_{j_i} is close to the ideal center. We next discuss a theorem to guarantee the convergence of the approach of **Table 1**.

*Theorem 1. The approach in **Table 1** will monotonically decrease the objective function $\sum_{i=0}^{k-1} \parallel x_{j_i} - \mu_j^t \parallel$ until convergence.*

Proof. According to step 3 in **Table 1**, we know that

$$\mu_j^t = \arg\min_{\mu_j^t} \sum_{i=0}^{k-1} \tau_{j_i}^t \left\| x_{j_i} - \mu_j^t \right\|_F^2, \tag{10}$$

where $\tau_{j_i} = 1 \big/ \| x_{j_i} - \mu_j^{t-1} \|$ as in step 2 of **Table 1**. Following Eq. (**10**), we have

$$\sum_{i=0}^{k-1} \left\{ \left\| x_{j_i} - \mu_j^t \right\|^2 \Big/ \left\| x_{j_i} - \mu_j^{t-1} \right\| \right\} \leq \sum_{i=0}^{k-1} \left\{ \left\| x_{j_i} - \mu_j^{t-1} \right\|^2 \Big/ \left\| x_{j_i} - \mu_j^{t-1} \right\| \right\}. \tag{11}$$

Based on the lemma in reference [6] that $2\sqrt{a} - a\big/\sqrt{b} \leq 2\sqrt{b} - b\big/\sqrt{b}$ holds for any two nonzero value, we have

$$\sum_{i=0}^{k-1} \left\{ 2\left\| x_{j_i} - \mu_j^t \right\| - \frac{\left\| x_{j_i} - \mu_j^t \right\|^2}{\left\| x_{j_i} - \mu_j^{t-1} \right\|} \right\} \leq \sum_{i=0}^{k-1} \left\{ 2\left\| x_{j_i} - \mu_j^{t-1} \right\| - \frac{\left\| x_{j_i} - \mu_j^{t-1} \right\|^2}{\left\| x_{j_i} - \mu_j^{t-1} \right\|} \right\}. \tag{12}$$

By summing Eqs. (**11**) and (**12**) in two sides, we have

$$\sum_{i=0}^{k-1} \left\| x_{j_i} - \mu_j^t \right\| \leq \sum_{i=0}^{k-1} \left\| x_{j_i} - \mu_j^{t-1} \right\|. \tag{13}$$

Eq. (**14**) indicates that the objective function $\sum_{i=0}^{k-1} \| x_{j_i} - \mu_j^t \|$ is monotonically decreased in each iteration. Since there is a lower bound in the objective function ($\sum_{i=0}^{k-1} \| x_{j_i} - \mu_j^t \| \geq 0$), the iterative approach will certainly converge. We thus prove Theorem 1. Finally, by incorporating the weight for reducing the bias for each local regression error into Eq. (**4**), we can reduce the bias of outliers of data samples.

Here, in order to show the convergence of the approach, we simply show an example in **Figure 2(a)**, where we generalize eight normal data points and two outliers in R^2. **Figure 2(b)** shows the converged route of μ, where we start μ^0 as the average mean of all data points and mark μ^t in each iteration with t. From **Figure 2(b)**, we can observe that the optimal solution μ^t will iterative close to the main center of normal data while be far away from the outliers. **Figure 2(c)** shows the converged curve of approach as discussed in **Table 1**. From **Figure 2(c)**, we can observe that the objective $\sum_{i=0}^{k} \| x_i - \mu^t \|$ will monotonically decrease until convergence. **Figure 2(d)** shows the converged weight of data points. From **Figure 2(d)**, we can observe the weights of normal data points are strengthen while those of outliers can be reduced.
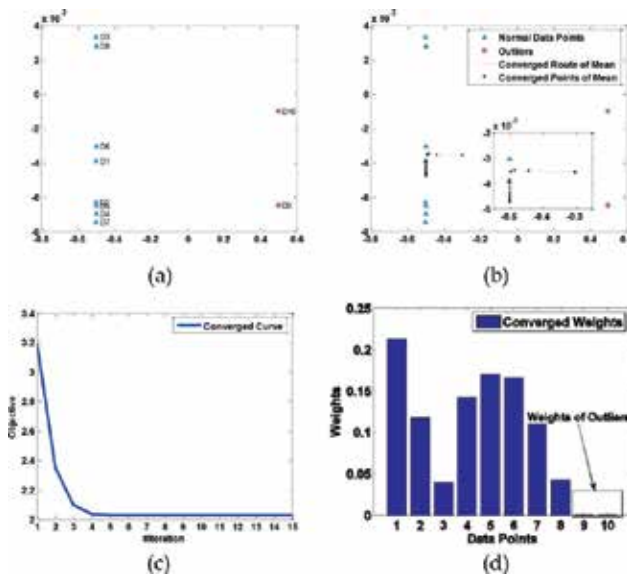
Figure 2. The convergence of the approach in **Table 1**: (a) original data, (b) the converged route of mean, (c) the converged curve of objective, (d) the converged weight.

2.4. Normalizing graph Laplacian matrix

It can be easily proved that L_d is a graph Laplacian matrix (see the Appendix). But L_d may not be a normalized graph Laplacian matrix. As pointed in references [8, 23], the normalization can strengthen the local regressions in the low-density region and weaken those in the high-density region. Since the data sampling is usually uniform in practice, normalization is useful for handling the case when the density of dataset varies dramatically. In this section, we show that by choosing a special weight vector Γ_j for each X_j, L_d can be a normalized graph Laplacian matrix.

Specifically, let us consider a data sample x_j and let K_l be the index set of those neighborhoods; set $N_k(x_j)$ contains x_j as a neighbor of x_l, i.e., if $j \in K_l$, then $x_l \in N_k(x_j)$, where x_l can be denoted as x_{j_i} in the neighborhood set $N_k(x_j)$, and $i=i(l, j)$ is the local index depending on l and j. Obviously, if x_l is in the low-density area, it has sparse neighbors and K_l is relatively small. As a result, its connections to other samples will be weaker than that which has large K_l. Here, to strengthen the connections of samples in the low-density area, we need to normalize the weights corresponding to each K_l. Let τ_j^l be the weight of x_{j_i} and l be the global index of x_{j_i}. We then define $\tau_{j_i} = \tau_j^l$ as follows:

$$\tau_{j_i} = \tau_j^l \leftarrow \frac{\tau_j^l}{\sum_{i \in K_l} \tau_i^l}. \tag{14}$$

Hence, based on this definition, we have the following theorem:

Theorem 2. *With the normalization for each w_{j_i} as in Eq. (**14**), $L_d$is both graph Laplaican matrix and normalized graph Laplacian matrix*.

Proof. The proof that L_d is a graph Laplacian matrix can be seen in the Appendix. In order to prove L_d is a normalized graph Laplacian matrix, we need prove L_d can be reformulated in the form of $L_d = I - W_d$ and the sum of each row or column of the affinity matrix W_d is equal to 1. Note $L_d = \sum_{j=1}^{l+u} \left(S_j L_j S_j^T \right)$ and $L_j = H_j - H_j X_j^T (X_j H_j X_j^T + \eta I)^{-1} X_j H_j$, where $H_j = \Delta_j - \left(\Delta_j e_k^T e_k \Delta_j \right) / \left(e_k \Delta_j e_k^T \right)$, we first define the affinity matrix W_d as follows:

$$W_d = \sum_{j=1}^{l+u} \left(S_j W_j^d S_j^T \right), \tag{15}$$

where each W_j^d satisfies

$$W_j^d = \left(\Delta_j e_k^T e_k \Delta_j \right) / \left(e_k \Delta_j e_k^T \right) - H_j X_j^T (X_j H_j X_j^T + \eta I)^{-1} X_j H_j. \tag{16}$$

Then, L_d can be reformulated as

$$L_d = \sum_{j=1}^{l+u} \left(S_j \Delta_j S_j^T \right) - \sum_{j=1}^{l+u} \left(S_j W_j^d S_j^T \right). \tag{17}$$

Here, for each $S_j \Delta_j S_j^T$, we have $S_j^T e^T = e_k^T \Rightarrow S_j \Delta_j S_j^T e^T = S_j \Gamma_j^T$, where $S_j \Gamma_j^T \in R^{(l+u) \times 1}$ is a column vector by putting each τ_j^l to its global index l corresponding to x_{j_i}. We thus have

$$\left\{ \sum_{j=1}^{l+u} \left(S_j \Delta_j S_j^T \right) \right\} e^T = \sum_{j=1}^{l+u} \left(S_j \Gamma_j^T \right) = e^T. \tag{18}$$

The second equation holds as $\sum_{i \in K_l} \tau_i^l = 1$; hence, the sum of all $S_j \Gamma_j^T$ in each element is equal to 1. Then, following Eq. (**18**), it indicates $\sum_{j=1}^{l+u} \left(S_j \Delta_j S_j^T \right)$ is an identity matrix, i.e., $\sum_{j=1}^{l+u} \left(S_j \Delta_j S_j^T \right) = I$. Then based on the above analysis, we can reformulate L_d in the form of $L_d = I - W_d$. In addition, since L_d is a graph Laplaican matrix (as proved in the Appendix), it satisfies $L_d e^T = 0$, then we have

$$
\begin{aligned}
L_d e^T = 0 \Rightarrow & \left\{ \sum_{j=1}^{l+u} \left(S_j \Delta_j S_j^T \right) - \sum_{j=1}^{l+u} \left(S_j W_j^d S_j^T \right) \right\} e^T = 0 \\
\Rightarrow & \left\{ \sum_{j=1}^{l+u} \left(S_j \Delta_j S_j^T \right) \right\} e^T = \left\{ \sum_{j=1}^{l+u} \left(S_j W_j^d S_j^T \right) \right\} e^T, \\
\Rightarrow & e^T = \left\{ \sum_{j=1}^{l+u} \left(S_j W_j^d S_j^T \right) \right\} e^T \\
\Rightarrow & W_d e^T = e^T \text{ or } e W_d = e
\end{aligned}
\tag{19}
$$

which indicates that the sum of each column or row of W_d is equal to 1. We thus prove the theorem. Theorem 2 indicates that by choosing a special weight vector τ_{j_i} for each x_{j_i}, L_d can be both graph Laplacian matrix and normalized graph Laplacian matrix.

Here, it should be noted that if x_l is an outlier, its local weights can be significantly decreased, whether taking x_l as a neighbor of itself or of other data points. Otherwise, the normalization does not change the magnitude of its original local weights. For some data points in the low-density area, normalizing the weights can increase the information convection through those points. Finally, the basic steps of the proposed LGR are given in **Table 2** and the flowchart by utilizing the proposed LGR method for face recognition is given in **Figure 3**.

Input: Data matrix $X \in R^{D \times (l+u)}$, the initial label matrix $Y \in R^{c \times (l+u)}$, and other related parameters.

Output: The projection matrix $V^* \in R^{D \times d}$ and estimated label matrix $Z^* \in R^{c \times (l+u)}$.

Algorithm:

1. Determine the weight for each local patch based on Table 1.

2. Normalize the weight as in Eq. (**14**).

3. Form local regression regularized term L_d as in Eq. (**5**) with special local weight vector.

4. Form global regression regularized term L_g as in Eq. (**7**).

5. Solve the regression problem as in Eq. (**8**):

$$
J(Z) = \min_Z Tr \left((Z - Y) U (Z - Y)^T \right) + \alpha_m Tr \left(Z L_d Z^T \right) + \alpha_r Tr \left(Z L_g Z^T \right),
$$

and calculate estimated label matrix $Z^* = YU \left(U + \alpha_m L_d + \alpha_r L_g \right)^{-1}$ as in Eq. (**9**). Output $V^* = \left(X L_c X^T + \eta I \right)^{-1} X L_c Z^{*T}$.

6. Calculate the projection matrix V^* by replacing z^* to Eq. (**6**) as $V^* = \left(X L_c X^T + \eta I \right)^{-1} X L_c Z^{*T}$. Output V^*.
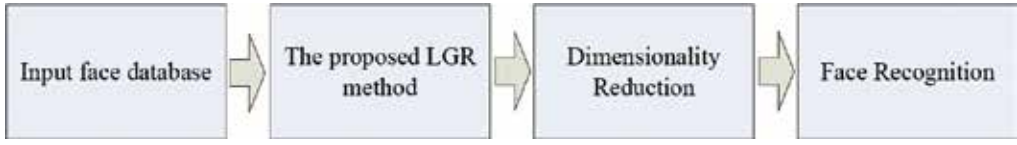

Table 2. The proposed LGR.

Figure 3. Flowchart by utilizing the proposed LGR for face recognition.

2.5. Discussion and relative work

In this section, we discuss the relationship of Learning from Local and Global Information (LLGDI) with other state-of-the-art methods including MR, Flexible Manifold Embedding (FME), and Local Regression and Global Alignment (LRGA).

2.5.1. Relationship to manifold regularization (Lap-RLS/L) [1]

The goal of MR [1] is to develop a semisupervised learning strategy by extending the original supervised methods, such as RLS and SVM to their semisupervised learning versions, i.e., Laplacian RLS and Laplacian SVM. For example, Lap-RLS/L is to fix a linear model $y_j = V^T x_j + b^T$ by regressing X on Y and simultaneously to preserve the manifold smoothness in the embeddings of both the labeled and the unlabeled set. The objective function of Lap-RLS/L can be given as

$$J(V,b) = \min \sum_{j=1}^{l} \left\| V^T x_j + b^T - y_j \right\|_F^2 + \alpha_t \|V\|_F^2 + \alpha_m Tr\left(V^T XLX^T V\right). \tag{20}$$

However, it can be observed that Lap-RLS/L cannot sufficiently train the classification function due to the utilization of labeled samples, though it uses manifold term as complementary. Hence, the proposed LGR is superior to Lap-RLS/L.

2.5.2. Relationship to FME [7, 10]

Nie et al. has proposed another unified framework, i.e., FME [7, 10], for semisupervised dimensionality reduction, in which they verify that LLGC, GFHF, and Lap-RLS/L are only special cases in the framework. The basic objective function of FME can be given as

$$J(V,Z,b) = \min \sum_{i=1}^{l} \left\| z_i - y_i \right\|_F^2 + \alpha_m Tr\left(ZLZ^T\right) + \alpha_r \left(\left\| V^T X + b^T e - Z \right\|_F^2 + \eta \|V\|_F^2 \right). \tag{21}$$

It can be observed that Eq. (**22**) is almost the same as the objective function of LGR in Eq. (**10**), when we consider $L_d \rightarrow L$. However, LGR has utilized a weighted and normalized local discriminative Laplacian matrix to preserve manifold and discriminative structure in a dataset. This is a better way than only relying on neighborhood graph.

2.5.3. Relationship to LRGA [13, 14]

Recently, Yang et al. has proposed semisupervised transductive learning method, namely, LRGA [13, 14], for multimedia retrieval. They share the similar concept with the proposed method. The basic objective function of LRGA can be given as

$$J(Z) = \min_{Z, V_j, b_j} \sum_{i=1}^{l} \|z_i - y_i\|_F^2 + \alpha_m \sum_{j=1}^{l+u} \left(\sum_{i=1}^{k} \|V_j^T x_{j_i} + b_j^T - z_{j_i}\|_F^2 + \eta \|V_j\|_F^2 \right). \qquad (22)$$

It can be noted that LRGA is a special case of LGR when $\alpha_r = 0$. Therefore, LRGA is only a transductive learning method and cannot handle the out-of-sample problem, while LGR is a transductive and inductive learning method. Another superiority of LGR over LRGA is that LGR has adopted a weighted normalized each local regression term. Thus, as shown in the simulation results, LLGDI can handle outliers and multi-density dataset remarkably.

3. Simulation results

In this section, we will evaluate the proposed LGR based on three synthetic datasets and two real-world datasets.

3.1. Synthetic datasets

In this section, we evaluate the performance of the proposed LGR and SLP for transductive learning. The SLP is an extensive method to GFHF, LLGC, and Random Walk (RW) hence, it is representative. Here, we utilize two-moon and two-cycle datasets in **Figure 1(a and b)** for
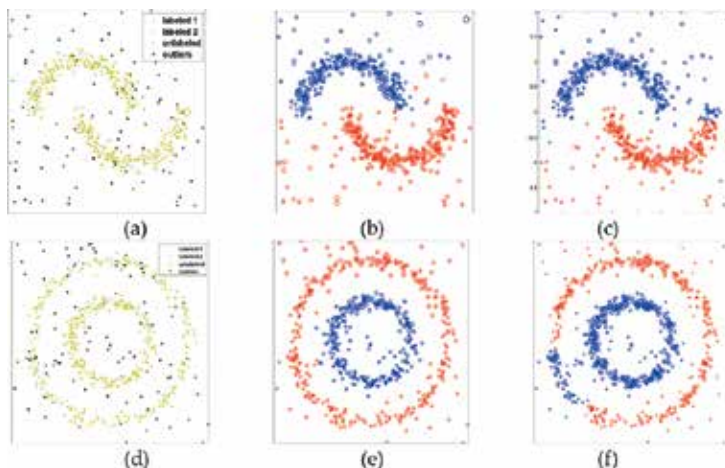

Figure 4. Toy examples for transductive learning: (a) and (d) the original data of two-moon and two-cycle datasets; (b) and (e) the results of LGR; (c) and (f) the results of SLP.

evaluation. **Figure 4** shows the results of LGR and SLP for transductive learning. From **Figure 4**, we can see that LGR can achieve better simulation result than SLP, in a way that less data are misclassified in LGR than SLP. This indicates the proposed LGR is robust to the outliers.

We also evaluate the inductive performance of the proposed LGR for handling the out-of-sample problem. **Figure 5** shows the gray images of decision surfaces and boundaries learned by LGR, which are formed as follows: for each pixel, we form the its gray value as the difference from each pixel to its nearest labeled data of different classes in the reduced subspace. Here, we set the reduced dimensionality as 1. Then, we form the decision boundaries by the pixels with the value 0. Following **Figure 5**, we can observe that the proposed LGR can learn clear decision boundary that can well separate two classes, which verifies the effectiveness of LGR for handling the out-of-sample problem.

To show the merit of normalization, we utilize two-plate dataset in **Figure 1(c)** for evaluation. Our goal is to show LGR can handle multi-density dataset. **Figure 6** shows the gray images of decision surfaces and boundaries learned by LGR without normalization and LGR with normalization. From **Figure 6**, we can observe that LGR without normalization cannot find proper boundary. However, LGR with normalization can achieve better performance, as there are less missing-classified data points separated by the decision boundary, which becomes more distinctive and accurate. The improved results are believed to be due to the fact that normalization can strengthen the local regressions in the low-density region and weaken those in the high-density region. This is proved to be advantageous to be used for multi-density dataset.
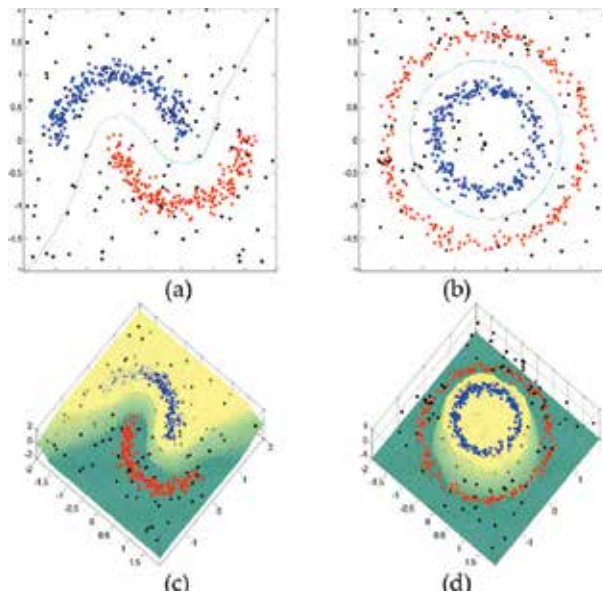

Figure 5. Toy examples for inductive learning: decision surfaces and boundaries learned by LGR. (a) and (c) Two-moon dataset; (b) and (d) two-cycle dataset.
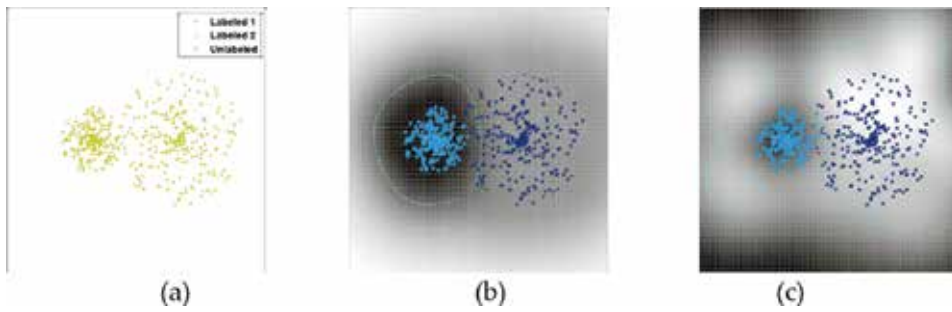
Figure 6. Gray image of reduced space learned by LGR without normalization and LGR with normalization: two-plate dataset. (a) Original dataset; (b) LGR without normalization; and (c) LGR with normalization.

3.2. Semisupervised face recognition based on real-world benchmark datasets

For handling the face recognition problem, we use three real-world face datasets to evaluate the performance of methods, which include UMNIST: cannot find the full name [24], Extended Yale-B [25], and Massachusetts Institute of Technology Center for Biological and Computational Learning (MIT-CBCL) [26] datasets. The UMIST dataset is a multi-view face dataset, consisting of 1012 images of 20 peoples, each covering a wide range of poses from profile to frontal views. Therefore, the UMIST has widely been used for general purpose face recognition under different face poses. The size of each image is 112×92 with 256 gray levels per pixel. In our simulation, we down-sample the size of each image to 28×23 and no other preprocessing is performed. The Extended Yale-B dataset contains 16,123 images of 38 human subjects under 9 poses and 64 illumination conditions. Because of the illumination variability, the same object can appear dramatically different even when viewed in fixed pose. Hence, this is another challenge for face recognition, and Extended Yale-B dataset are extensively used for testing appearance-based face recognition methods. Similar to the UMIST dataset, the images are also cropped and resized to 32×32 pixels. This dataset now has around 64 near frontal images under different illuminations per individual. The MIT-CBCL dataset provides 3240 synthetic images rendered from 3D head models of 10 peoples. The head models are generated by fitting a morphable model to the high-resolution training images. Different from UMNIST dataset, the MIT-CBCL dataset is based on the 3D morphable model, which is rendered under varying pose and illumination conditions making the face recognition task more challengeable. The size of each image is originally 200×200 with 256 gray levels per pixel. In our simulation, we down-sample the size of each image to 32×32 and no other preprocessing is performed. The detailed information of dataset and some sampled images of real-world datasets can be shown in **Table 3** and **Figure 7**. For each dataset, we randomly select 10, 50 and 30 samples from each class as training samples for UMNIST, Extended Yale-B, and MIT-CBCL datasets. The test set is then formed by the selected or all remaining samples. The data partitioning for each dataset is also given in **Table 3**.

Next, we compare our method with other supervised and semisupervised dimension reduction methods. These methods include Regularized Linear discriminant analysis (RLDA), SDA

[2], Lap-RLS/L [1], least-square solution for solving SDA in Eq. (**16**) (in **Table 1**, we refer to it as LS-SDA) [28], FME [7, 10], and the proposed LGR. Note that Principal Component Analysis (PCA) is an unsupervised method while RLDA is supervised methods, and the remaining methods LGR are all semisupervised methods. The simulation settings are as follows: for SDA, Lap-RLS/L, two parameters, i.e., α_t and α_m, need to be determined for balancing the trade-off between the manifold and Tikhonov terms. We use fivefold cross validation to determine the best values and the candidate set is $\{10^{-9}, 10^{-6}, 10^{-3}, 10^{0}, 10^{3}, 10^{6}, 10^{9}\}$. The above candidate set is also used for determining the best value for the Tikhonov term parameter α_t in RLDA and the addition regularized parameter α_r in FME and LGR. In order to eliminate the null space before performing dimension reduction, the training sets in all datasets are preliminarily processed with PCA operator. Since most of methods, such as RLDA, SDA, Lap-RLS/L and FME, and the proposed LGR have a limited rank of c–1, we simply reduce the dimensionality of all methods to c–1. All methods used labeled set in the output reduced subspace to train a nearest neighborhood classifier in order to evaluate the classification accuracy of test set. We also compare the performance of nearest neighborhood classifier with other state-of-the-art methods as a baseline.

Dataset	Database Type	#Samples	#Dim	#Class	#Training per Class	#Test per Class
UMNIST	Face	1012	1024	20	20	Remains
Extended Yale-B	Face	16123	1024	38	50	Remains
MIT-CBCL	Face	3240	1024	10	30	30

Table 3. Dataset information and data partition for each dataset.

(a)   (b)   (c)
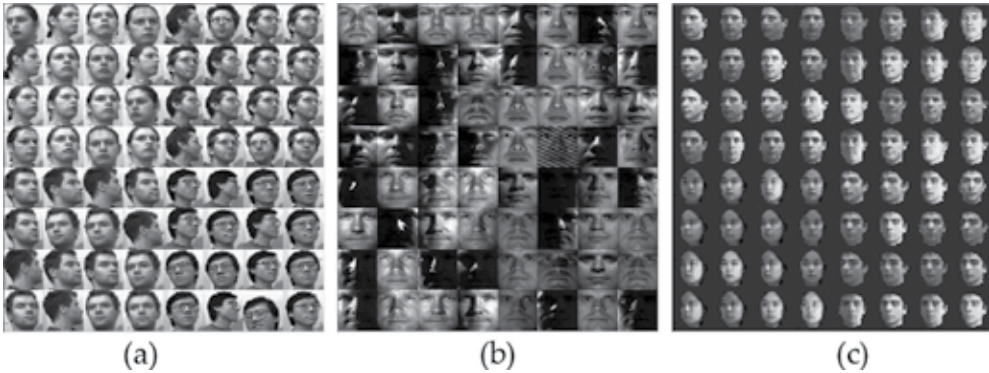
Figure 7. Sample images of real-world datasets: (a) UMNIST dataset, (b) Extended Yale-B dataset, (c) MIT-CBCL dataset.

The average accuracies over 20 random splits with the above parameters for each dataset are shown in **Table 4**. From the simulation results, we can obtain the following observation: (1) given sufficient labeled samples, all the supervised and semisupervised dimension reduction methods outperform nearest neighborhood classifier due to the utilization of label information

and feature extraction; (2) the semisupervised dimension reduction methods are better than the corresponding supervised methods. For example, SDA outperforms RLDA by about 5–6% in COIL100 dataset with two labeled samples per class. For other datasets, it can outperform by 2–3%. This indicates that by incorporating the unlabeled set into the training procedure, the classification performance can be markedly improved, as the manifold structure embedded in the dataset is preserved; (3) we also observe that both SDA and the least-square solution in **Table 1** can achieve the same classification results due to the reason as analyzed in Section 3; (4) the proposed LGR can deliver better accuracies than those delivered by other semisupervised dimension reduction methods such as SDA and Lap-RLS/L by about 3–4% in most datasets. The improvement can even achieve almost 8% in ETH80 dataset with two labeled samples per class. The improvement is believed to be true that LGR aims to characterize both local and global discriminative information embedded in dataset, which is better to handle classification problem; (5) we observe that LGR outperform FME by about 2% in most cases. The main reason is that LGR has utilized a weighted normalized local discriminative Laplacian matrix to preserve both manifold and discriminative structures in dataset, which is better than only relying on neighborhood graph.

Dataset	Method	4 labeled samples per class		7 labeled samples per class		10 labeled samples per class	
		Unlabeled	Test	Unlabeled	Test	Unlabeled	Test
		Mean±std	Mean±std	Mean±std	Mean±std	Mean±std	Mean±std
UMNIST	Baseline	81.1±0.9	80.2±1.0	88.6±0.7	88.3±0.7	93.1±0.6	93.0±0.7
	RLDA	85.2±0.6	85.0±0.7	90.7±0.5	90.4±0.6	95.3±0.4	94.4±0.5
	SDA	86.4±0.7	86.3±0.7	92.1±0.6	91.7±0.7	96.2±0.4	95.4±0.5
	LS-SDA	86.4±0.7	86.3±0.7	92.1±0.6	91.7±0.7	96.2±0.4	95.4±0.5
	Lap-RLS/L	86.6±0.7	86.0±0.8	91.9±0.3	91.9±0.4	95.7±0.5	95.3±0.6
	FME	88.2±0.6	87.7±0.6	93.1±0.3	92.9±0.4	96.7±0.5	96.1±0.5
	LGR	89.2±0.4	88.9±0.5	94.2±0.2	93.8±0.4	97.9±0.6	97.2±0.4

Table 4. Average classification accuracy over 20 random splits on unlabeled set and test set of different datasets (means±standard derivations).

4. Conclusion

In this chapter, we propose a semisupervised method, namely LGR, for face recognition. With the above analysis, the following conclusions can be drawn: (1) the proposed LGR can achieve better results in face recognition than those delivered by other state-of-the-art methods as more discriminative information are captured based on local and global regressions, (2) the proposed LGR is robust to outliers and can handle the imbalanced data, and (3) the proposed LGR can deal with out-of-sample extrapolation to estimate the labels of new-coming face data by casting it to the global projection matrix.

Appendix

In order to prove that L_d is graph Laplacian matrix, we need to prove L_d is positive semidefinite matrix and the sum of each row or column of L_d is equal to zero. We first have the following Lemmas:

Lemma 1. *For each local patch X_j, L_j can be reformulated as follows:*

$$L_j = \eta G_j \left(G_j^T X_j^T X_j G_j + \eta I \right)^{-1} G_j^T, \tag{23}$$

where $G_j = \left(I - \Delta_j e_k^T e_k / \left(e_k \Delta_j e_k^T \right) \right) \Delta_j^{-1/2} \in R^{k \times k}$.

Proof. First, it can be easily noted that $G_j G_j^T = H_j$, which is verified as follows:

$$
\begin{aligned}
G_j G_j^T &= \left(I - \Delta_j e_k^T e_k / \left(e_k \Delta_j e_k^T \right) \right) \Delta_j \left(I - \Delta_j e_k^T e_k / \left(e_k \Delta_j e_k^T \right) \right)^T \\
&= \left(\Delta_j - \Delta_j e_k^T e_k \Delta_j / \left(e_k \Delta_j e_k^T \right) \right) \left(I - \Delta_j e_k^T e_k / \left(e_k \Delta_j e_k^T \right) \right)^T \\
&= \Delta_j - 2 \Delta_j e_k^T e_k \Delta_j / \left(e_k \Delta_j e_k^T \right) + \Delta_j e_k^T \left(e_k \Delta_j e_k^T \right) e_k \Delta_j / \left(e_k \Delta_j e_k^T \right)^2 \\
&= \Delta_j - \Delta_j e_k^T e_k \Delta_j / \left(e_k \Delta_j e_k^T \right) = H_j
\end{aligned}
\tag{24}
$$

Then, we have

$$
\begin{aligned}
L_j &= H_j - H_j X_j^T (X_j H_j X_j^T + \eta I)^{-1} X_j H_j \\
&= G_j G_j^T - G_j G_j^T X_j^T (X_j G_j G_j^T X_j^T + \eta I)^{-1} X_j G_j G_j^T \\
&= G_j G_j^T - G_j G_j^T X_j^T X_j G_j \left(G_j^T X_j^T X_j G_j + \eta I \right)^{-1} G_j^T. \\
&= G_j G_j^T - G_j \left(G_j^T X_j^T X_j G_j + \eta I - \eta I \right) \left(G_j^T X_j^T X_j G_j + \eta I \right)^{-1} G_j^T \\
&= \eta G_j \left(G_j^T X_j^T X_j G_j + \eta I \right)^{-1} G_j^T
\end{aligned}
\tag{25}
$$

The second equation holds as $A(A^T A + \lambda I)^{-1} = (AA^T + \lambda I)^{-1} A$, for any matrix A. Thus, Lemma 1 is proved.

Lemma 2. *Given a positive semidefinite matrix C, DCD^T is a positive semidefinite matrix for any matrix D.*

Lemma 3. *Given a set of positive semidefinite matrixes* $\{C_1, C_2 \dots, C_n\}$ *then* $\sum_{j=1}^{n} C_j$ *is a positive semidefinite matrix.*

We neglect the proofs of Lemmas 2 and 3 as they can be seen in reference [15]. Then with Lemmas 1–3, we can easily prove Theorem 2 as follows:

Proof of Theorem 2. Note that following Lemma 1, we reformulate each L_j as $L_j = \eta G_j \left(G_j^T X_j^T X_j G_j + \eta I \right)^{-1} G_j^T$. It can be noted $\left(G_j^T X_j^T X_j G_j + \eta I \right)^{-1}$ is a positive semidefinite matrix, then, following Lemmas 2 and 3, we have each $\eta S_j G_j \left(G_j^T X_j^T X_j G_j + \eta I \right)^{-1} G_j^T S_j^T$ is a positive semidefinite matrix and L_d, i.e.,

$$L_d = \sum_{j=1}^{l+u} \left(S_j L_j S_j^T \right) = \sum_{j=1}^{l+u} \left(\eta S_j G_j \left(G_j^T X_j^T X_j G_j + \eta I \right)^{-1} G_j^T S_j^T \right). \tag{26}$$

is also a positive semidefinite matrix. In addition, for each $\eta S_j G_j \left(G_j^T X_j^T X_j G_j + \eta I \right)^{-1} G_j^T S_j^T$, we have $S_j^T e^T = e_k^T$ and

$$G_j^T e_k^T = \left(e_k^T - \left(e_k^T \Delta_j e_k \right) e_k^T \big/ \left(e_k \Delta_j e_k^T \right) \right) \Delta_j^{-1/2} = \left(e_k^T - e_k^T \right) \Delta_j^{-1/2} = 0$$

$$\Rightarrow \eta S_j G_j \left(G_j^T X_j^T X_j G_j + \eta I \right)^{-1} G_j^T S_j^T e^T = 0 \tag{27}$$

$$\Rightarrow L_d e^T = \sum_{j=1}^{l+u} \left(\eta S_j G_j \left(G_j^T X_j^T X_j G_j + \eta I \right)^{-1} G_j^T S_j^T \right) e^T = 0$$

which indicates that the sum of each row or column of L_d is equal to zero. We thus prove L_d is graph Laplacian matrix.

Author details

Mingbo Zhao[1*], Yuan Gao[1*], Zhao Zhang[2] and Bing Li[3]

*Address all correspondence to: mbzhao4@gmail.com

*Address all correspondence to: ethan.y.gao@my.cityu.edu.hk

1 Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong S. A. R.

2 School of Computer Science and Technology, Soochow University, Suzhou, P. R. China

3 School of Economics, Wuhan University of Technology, Wuhan, P. R. China

References

[1]   M. Belkin, P. Niyogi, V. Sindhwani. Manifold regularization: a geometric framework for learning from labeled and unlabeled samples. *Journal of Machine Learning Research*, 7:2399–2434, 2006.

[2]   D. Cai, X. He, J. Han. Semi-supervised discriminant analysis. *IEEE International Conference on Computer Vision*, Rio de Janeiro, Brazil, IEEE, 1–7, 2007.

[3]   X. Zhu, Z. Ghahramani, J. D. Lafferty. Semi-supervised learning using Gaussian fields and harmonic functions. In *Proceedings of ICML*, Washington DC, USA, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2003.

[4]   D. Zhou, O. Bousquet, T. N. Lal, J. Weston, B. Scholkopf. Learning with local and global consistency. In *Proceedings of NIPS*, Vancouver, Canada, Massachusetts Institute of Technology Press, Cambridge, MA, USA, 2004.

[5]   M. Szummer, T. Jaakkola. Patially labeled classification with Markov random walks. In *Proceedings of NIPS*, Vancouver, Canada, Massachusetts Institute of Technology Press, Cambridge, MA, USA, 2002.

[6]   F. Nie, H. Huang, X. Cai, C. Ding. Efficient and robust feature selection via joint L21-norms minimization. In *Proceedings of NIPS*, Vancouver, Canada, Massachusetts Institute of Technology Press, Cambridge, MA, USA 2010.

[7]   F. Nie, D. Xu, I. W. H. Tsang, C. Zhang. Flexible Manifold Embedding: a framework for semi-supervised and unsupervised dimension reduction. *IEEE Transactions on Image Processing*, 19(7):1921–1932, 2010.

[8]   F. Nie, S. Xiang, Y. Liu, C. Zhang. A general graph based semi-supervised learning with novel class discovery. *Neural Computing and Application*, 19(4):549–555, 2010.

[9]   F. Nie, D. Xu, X. Li, S. Xiang. Semi-supervised dimensionality reduction and classification through virtual label regression. *IEEE Transactions on Systems, Man and Cybernetics, Part B*, 41(3):675–685, 2011.

[10]   F. Nie, D. Xu, I. W. H. Tsang, C. Zhang. A flexible and effective linearization method for subspace learning. *Graph Embedding for Pattern Analysis*, 177–203, Yun Fu, Yunqian Ma, Eds. Springer, New York, 2013.

[11]   F. Wang, C. Zhang. Label propagation through linear neighborhoods. *IEEE Transactions on Knowledge and Data Engineering*, 20(1):55–67, 2008.

[12]   J. Wang, F. Wang, C. Zhang, H. C. Shen, L. Quan. Linear neighborhood propagation and its applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(9): 1600–1615, 2009.

[13] Y. Yang, F. Nie, D. Xu, J. Luo, Y. Zhuang, Y. Pan. A multimedia retrieval framework based on semi-supervised ranking and relevance feedback. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(5):723–742, 2012.

[14] Y. Yang, D. Xu, F. Nie, J. Luo, Y. Zhuang. Ranking with local regression and global alignment for cross medial retrieval. In *Proceedings of MM*, Beijing, China, ACM New York, NY, USA, 2009.

[15] Y. Yang, F. Nie, S. Xiang, Y. Zhuang, W. Wang. Image clustering using local discriminant models and global integration. *IEEE Transactions on Image Processing*, 19(10):2761–2773, 2010.

[16] D. Wang, F. Nie, H. Huang. Large-scale adaptive semi-supervised learning via unified inductive and transductive model. In *Proceeding of KDD*, New York, NY, USA, ACM New York, NY, USA, 2014.

[17] X. He, S. Yan, Y. Hu, P. Niyogi, H. Zhang. Face recognition using Laplacian faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3):328–340, 2005.

[18] S. Xiang, F. Nie, C. Zhang. Semi-supervised classification via local spline regression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(11):2039–2053, 2010.

[19] S. Xiang, F. Nie, C. Zhang. Nonlinear dimensionality reduction with local spline embedding. *IEEE Transactions on Knowledge and Data Engineering*, 21(9):1285–1298, 2009.

[20] S. T. Roweis, L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 209: 2323–2326, 2000.

[21] S. A. Nene, S. K. Nayar, H. Murase. Columbia object image library (COIL-100). *Technical Report CUCS-005-96*, Columbia University, New York, NY, 1996.

[22] B. Leibe, B. Schiele. Analyzing appearance and contour based methods for object categorization. In *Proceedings of CVPR*, Madison, Wisconsin, USA, IEEE, 2003.

[23] R. Johnson, T. Zhang. On the effectiveness of Laplacian normalization for graph semi-supervised learning. *Journal of Machine Learning Research*, 8:1489–1517, 2007.

[24] D. B. Graham, N. M. Allinson. Characterizing virtual eigensignatures for general purpose face recognition in face recognition: from theory to application. *NATO ASI Series F, Computer and Systems Sciences*, 163:446–456, 1998.

[25] K. C. Lee, J. Ho, D. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5): 947–963, 2005.

[26] B. Weyrauch, J. Huang, B. Heisele, V. Blanz. Component-based Face Recognition with 3D Morphable Models. *First IEEE Workshop on Face Processing in Video, IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington DC, USA, IEEE, 2004.

[27] M. Zhao, Z. Zhang, T. W. S. Chow, Trace ratio criterion based generalized discriminative information for semi-supervised dimensionality reduction. *Pattern Recognition*, 45(4):1482–1499, 2012.

[28] M. Zhao, Z. Zhang, H. Zhang. Learning from local and global discriminative information for semi-supervised dimensionality reduction. *The International Joint Conference on Neural Networks (IJCNN)*, 1–8, Dallas, TX, USA, IEEE, 2013.

[29] M. Zhao, Z. Zhang, T. W. S. Chow, B. Li. Soft label based linear discriminant analysis for image recognition and retrieval. *Computer Vision and Image Understanding*, 121:86–99, 2014.

[30] M. Zhao, Z. Zhang, T. W. S. Chow, B. Li. A general soft label based linear discriminant analysis for semi-supervised dimension reduction. *Neural Networks*, 55:83–97, 2014.

[31] M. Zhao, T. W. S. Chow, Z. Zhang, B. Li. Automatic image annotation via compact graph based semi-supervised learning. *Knowledge Based Systems*, 76:148–165, 2015.

[32] M. Zhao, C. Zhan, Z. Wu, P. Tang. Semi-supervised image classification based on local and global regression. *IEEE Signal Processing Letters*, 22(10):1666–1670, 2015.

[33] M. Zhao, T. W. S. Chow, Z. Wu, Z. Zhang, B. Li. Learning from normalized local and global discriminative information for semi-supervised regression and dimensionality reduction. *Information Sciences*, 324(10):286–309, 2015.

Advances of Robust Subspace Face Recognition

Yang-Ting Chou, Jar-Ferr Yang and
Shih-Ming Huang

Additional information is available at the end of the chapter

Abstract

Face recognition has been widely applied in fast video surveillance and security systems and smart home services in our daily lives. Over past years, subspace projection methods, such as principal component analysis (PCA), linear discriminant analysis (LDA), are the well-known algorithms for face recognition. Recently, linear regression classification (LRC) is one of the most popular approaches through subspace projection optimizations. However, there are still many problems unsolved in severe conditions with different environments and various applications. In this chapter, the practical problems including partial occlusion, illumination variation, different expression, pose variation, and low resolution are addressed and solved by several improved subspace projection methods including robust linear regression classification (RLRC), ridge regression (RR), improved principal component regression (IPCR), unitary regression classification (URC), linear discriminant regression classification (LDRC), generalized linear regression classification (GLRC) and trimmed linear regression (TLR). Experimental results show that these methods can perform well and possess high robustness against problems of partial occlusion, illumination variation, different expression, pose variation and low resolution.

Keywords: subspace projection, principal component analysis, linear discriminant analysis, linear regression classification, robust linear regression classification, ridge regression, improved principal component regression, unitary regression classification, linear discriminant regression classification, generalized linear regression classification, trimmed linear regression

1. Introduction

From the tragic 911 incident in 2011, more and more researches focus on the security issues with computational intelligence. How to avoid the tragic event from happening again and how to quickly identify the terrorists and suspects before or after the tragic event happens are very important. Therefore, the effectiveness of security is being examined almost everywhere. The discoveries of current security vulnerabilities along with the exploration of new methods should be constantly investigated to improve the security systems. Security measures with computational intelligence are used to improve the safety of our everyday lives.

The security issue is that we should recognize the wanted criminals and invaders from their biometric characteristics such as face, fingerprint, iris, palm and so on. Among these biometrics signals, the face image is easier and more direct to be captured by distanced cameras than others. For instance, the face images of any suspect who walks through the hotel lobby will be recorded by cameras. Hence, computer vision technologies with cameras can be applied to realize intelligent video surveillance systems. Since many face images of criminals and terrorists are available in the police department, they can be used to identify if the unknown face images are them from the distributed cameras. Thus, an efficient face recognition system could help to improve security. The face recognition systems would not only be helpful in identifying the criminals and terrorists, but also be used to search missing persons or identify the incident of weak person. Thus, face recognition systems with surveillance cameras have been already installed in many locations such as department store, airports and supermarkets. Besides, if the face recognition systems installed at home can timely detect the user's facial expression, the smart service for the user can be properly introduced accordingly.

The goal of face recognition is to distinguish a specific identity and its outlook from face images. However, in realistic situations, such as video surveillance and access control, face recognition task might encounter great challenges such as different facial expressions, illumination variations, partial occlusions and even low resolution problems, which will degrade the face recognition performance and result in severe security complications. For example, the image captured by a CCTV camera at a distance would have a very low resolution which degrades the recognition performance significantly. Besides, in the testing phase, the face image is a factor which is out of control. In other words, the person may not be on a frontal pose and may not be a pure image, that is the person may be wearing glasses, hat, or mask, or even with some lighting influence and expressions. Over past years, subspace projection optimizations have been widely proposed to solve this problem with linear [1] and non-linear [2] approaches. The principal component analysis (PCA) [3, 4] and linear discriminant analysis (LDA) [5] are the two typical examples of linear transform approaches which attempt to seek a low-dimensional subspace for dimensionality reduction. The nonlinear projection approaches also have been used in many literatures like the kernel PCA (KPCA) [6] and kernel LDA (KLDA) [7] which can uncover the underlying structure when the samples lie on a nonlinear manifold structure in the image space.

Recently, the linear regression classification (LRC) proposed in 2010 by Naseem *et al*. [8] has been treated as an effective subspace projection method, which performs well on face recog-

nition. Moreover, the robust linear regression classification (RLRC) [9] estimating regression parameters by using the robust Huber estimation was introduced to achieve robust face recognition under illumination variation and random pixel corruption. Ridge regression (RR) [10] estimated the regression parameters by using a regularized least square method to model the linear dependency in the spatial domain. Huang *et al*. and Chou *et al*. presented several improved approaches of LRC, including improved-PCA-LRC [11], LDA-LRC [12], unitary-LRC [13], and generalized-LRC [14, 15] for dealing with different situations like facial expressions, lighting changes, and pose variations. Lai *et al*. [16] utilized the least trimmed square (LTS) as a robust estimator to detect the contaminated pixels from query for boosting the performance under the partial occlusion situation.

The rest of this chapter is organized as follows. With the overview of fundamentals and facial representation, several famous face recognition algorithms are first presented in Section 2. Section 3 is dedicated to present several advances of subspace projection optimizations for robust face recognition technologies including RLRC, RR, improved principal component regression (IPCR), unitary regression classification (URC), linear discriminant regression classification (LDRC), generalized linear regression classification (GLRC) and trimmed linear regression (TLR). The performances of the aforementioned projection methods will be shown in Section 4. Finally, conclusions are drawn in Section 5.

2. Fundamentals of face recognition and representation

As shown in **Figure 1**, the typical face recognition system contains two major parts: face detection and face recognition. In this section, the face detection methods are first briefly introduced. Then, the well-known subspace project methods are reviewed. Finally, the similarity measures of image feature vectors are overviewed. Generally, the unknown data vector will be projected into a certain subspace, a similarity measure will be used to classify it. To narrow down the computation and increase the recognition accuracy, the first step of the recognition system, called face detection, is to detect and crop the face region from the image or video.
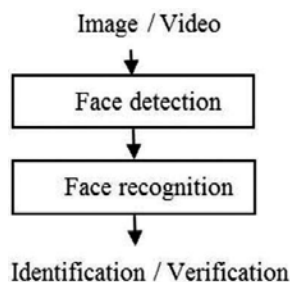

Figure 1. The simplified flow chart of face recognition system.

2.1. Face detection

The methods of face detection [17–20], can be separated into neural network, feature-based, and color-based approaches. Neural network approach [21] trains the facial class and non-facial class while a new image or video can be detected based on the prior training data. The well-known method is AdaBoost learning algorithm [22, 23]. Feature-based approach is to utilize the facial feature for detecting facial region. For example, the corresponding positions of the eyes, nose, and mouth are useful features; moreover, the shape of face, which is almost like an ellipse, can be included. Rule-based algorithm [24] and elliptical edge [25] are two popular feature-based methods. Color-based approach as [26] adopts the variance of skin color to detect if the region is face or not. For example, the face region in grayscale should not change immense while the eyes, mouth and hair should be darker than the other part of face.

Once the face areas are detected by a selected face detection method, their face images in size of $a{\times}b$ pixels could be projected into another subspace such as principal space, kernel space, frequency space and so on, in order to find a proper set of features for boosting the recognition performance. Assume there are C subjects. Each class is with N training color images. For the ith class, $i = 1, 2, …, C$, the jth training color image in size of $a{\times}b$ pixels with K components is formed a data matrix as $\boldsymbol{v}_{i,j,k} \in \boldsymbol{R}^{a{\times}b{\times}K}$ for $j = 1, 2, …, N$ and $k = 1, 2, …, K$. For example, $K = 3$ color components, $k = 1, 2,$ and 3 denote the red, green, blue channels, respectively. For some recognition algorithms, $\boldsymbol{v}_{i,j,k} \in \boldsymbol{R}^{a{\times}b{\times}K}$ is transformed to grayscale as $\boldsymbol{g}_{i,j}=c_1\boldsymbol{v}_{i,j,1}+c_2\boldsymbol{v}_{i,j,2}+c_3\boldsymbol{v}_{i,j,3}$, where are c_1, c_2, and c_3 are fixed in visualization. The gray image $\boldsymbol{g}_{i,j}$ is reshaped into one column vector as $\boldsymbol{x}_{i,j} \in \boldsymbol{R}^{M{\times}1}$ where $M = a{\times}b$. In the testing phase, an unknown color image, $\boldsymbol{z} \in \boldsymbol{R}^{M{\times}K}$, is given. In order to predict unknown $\boldsymbol{z}$ by training data, it should be transformed to grayscale, be normalized and be reshaped into a column vector as $\boldsymbol{y} \in \boldsymbol{R}^{M{\times}1}$.

2.2. Subspace projection methods

The famous subspace projection methods, such as PCA and LDA are reviewed in the following sections.

2.2.1. Principal component analysis (PCA)

The PCA method is widely used for dimensionality reduction in the computer vision field, especially for face recognition technology. In the PCA, the data is represented as a linear combination of an orthonormal set of vectors that maximize the data scatter across all images. The first principal component represents the most variability of the image as possible while the second one represents the second most, and so on. The flow chart to find PCA transformation bases is shown in **Figure 2**. The main objective of the PCA is to reduce the dimension of the feature image $\boldsymbol{x}_{i,j}$ to retain a few principal components. This means that most of the useless information would be reduced, and the remaining data could be well represented in a lower dimension space by the PCA.

As shown in **Figure 2**, the derivations of the PCA transformation bases are stated in the following equations. First, the feature face image should remove the global mean to become:

$$\widetilde{x}_{i,j} = x_{i,j} - \overline{x}_{\text{global}} \tag{1}$$

where $\overline{x}_{global} = \frac{1}{C \cdot N} \sum_C \sum_N x_{i,j}$ is a global mean vector of all facial image vectors.
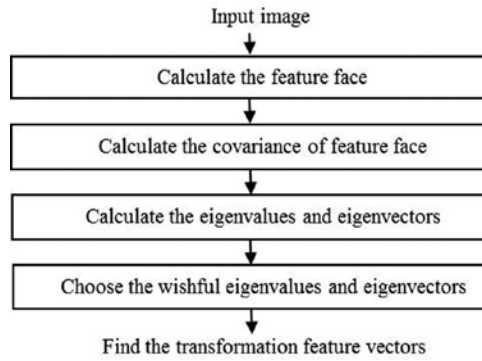

Figure 2. The flow chart for finding PCA transformation.

After the computation of the feature face images, we can obtain $M{\times}M$ covariance matrix of all feature face images as:

$$Qu = ru \tag{2}$$

Based on the covariance matrix, the eigenvectors and eigenvalues can be retrieved by singular value decomposition (SVD) or eigen-decomposition as:

$$P_{PCA} = \{u_1, u_2, \ldots, u_P\}, P \leq M \tag{3}$$

where $r = \{r_1, r_2, \ldots, r_M\}$ is a set of total M descending-ordered eigenvalues and their corresponding eigenvectors $u = \{u_1, u_2, \ldots, u_M\}$ According to the expected dimension, we can choose P principal components. Thus, the PCA transformation with the P largest eigenvectors, the PCA transformation P_{PCA} with $P{\times}M$ size can be formed by the corresponding P eigenvalues as:

$$w_{PCA,i,j} = P_{PCA}^T \widetilde{x}_{i,j} \tag{4}$$

Finally, we can achieve the PCA features, $w_{PCA,i,j} \in R^{P{\times}1}$, by multiplying PCA transformation and the feature image vector as:

$$P_{LDA} = \underset{P}{\text{argmax}} \frac{|P^T S_B P|}{|P^T S_W P|} \tag{5}$$

On the other hand, the testing image vector y can be projected onto PCA subspace by P_{PCA}. The PCA subspace $\hat{y}_{PCA}$ can be written as:

$$\hat{y}_{LDA} = P_{LDA}^T y \tag{6}$$

And the similarity measure based on this feature data vector is calculated to determine the final result.

2.2.2. Linear discriminant analysis

Fisher proposed the LDA for recognition which is a kind of statistical analysis method like the PCA. But the difference is that the LDA can discriminate the different subjects even though the maximum variance subspaces among them are overlapped as shown in **Figure 3**. The goal of the LDA is that these projections onto a line will be well separated by disparate classes and be well concentrated by the same class.
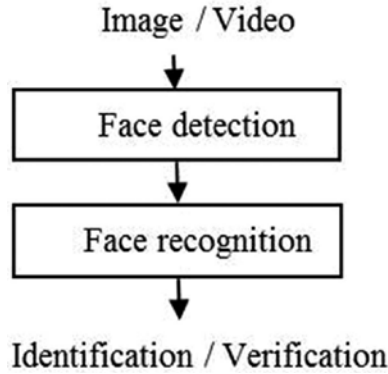
Image / Video

Face detection

Face recognition

Identification / Verification

Figure 3. Comparison of LDA and PCA in projection space.

Thus, the concept of LDA is to seek the optimal projection by maximizing the ratio of between-class and within-class scatter. Fisher utilizes a criterion to optimize this problem as:

$$L_{1,i,j} = \left\| w_{i,j} - \hat{y} \right\|_1 = \sum_{m=1}^{M} \left| w_{i,j}^{(m)} - \hat{y}^{(m)} \right| \tag{7}$$

where $S_B = \sum_{i=1}^{C} \sum_{q=1,q\neq i}^{C} (\bar{x}_{local,i} - \bar{x}_{local,q})(\bar{x}_{local,i} - \bar{x}_{local,q})^T$ is the between-class matrix where $\bar{x}_{local,i} = \frac{1}{N} \sum_{j=1}^{N} x_{i,j}$ is a local mean vector of the ith class. And $S_W = \frac{1}{C} \sum_{i=1}^{C} (X_i - \bar{x}_{local,i})(X_i - \bar{x}_{local,i})^T$ is the within-class matrix where X_i is concatenated by the ith data set of N training gray images. Then, the optimal projection matrix, W_{LDA}, can be solved by computing generalized SVD or eigen-decomposition as:

$$\cos\theta_{i,j} = \frac{\boldsymbol{w}_{i,j}^T\hat{\boldsymbol{y}}}{\|\boldsymbol{w}_{i,j}\|_2\|\hat{\boldsymbol{y}}\|_2} = \frac{\left|\Sigma_{h=1}^M w_{i,j}^{(h)}\hat{y}^{(h)}\right|^2}{\Sigma_{h=1}^M\left|w_{i,j}^{(h)}\right|^2 \Sigma_{h=1}^M\left|\hat{y}^{(h)}\right|^2} \tag{8}$$

where Λ is the diagonal eigenvalue matrix. We apply the optimal projection matrix to convert the face feature vector $\boldsymbol{x}_{i,j}$ into a new discriminant vector, $\boldsymbol{w}_{Fisher,i,j}$ as:

$$\boldsymbol{w}_{Fisher,i,j} = \boldsymbol{P}_{LDA}^T\boldsymbol{x}_{i,j} \tag{9}$$

In the same way, the testing image vector is projected onto LDA subspace by $\boldsymbol{P}_{LDA}$ and can be represented as:

$$\mathbf{y} = \boldsymbol{X}_i\boldsymbol{\beta}_i, i = 1,2,\dots,C \tag{10}$$

And the final result can be determined by using similarity measure based on this feature vector.

2.3. Similarity measures

There exist three distance measures [27–29] such as the city block distance (Taxicab geometry, L_1), Euclidean distance (L_2) and L_∞ norm distance. These distance measures are defined from two column vectors $\boldsymbol{w}_{i,j}$ and $\hat{\boldsymbol{y}}$ which can be obtained from the subspace projection like PCA subspace $\{\boldsymbol{w}_{PCA,i,j}, \hat{\boldsymbol{y}}_{PCA}\}$, LDA subspace $\{\boldsymbol{w}_{LDA,i,j}, \hat{\boldsymbol{y}}_{LDA}\}$, and the other projections with dimensionality of M or P. The distance measures, L_1, L_2, and L_∞ can be respectively written as:

$$\hat{\boldsymbol{y}}_{LRC,i} = \boldsymbol{X}_i\left(\boldsymbol{X}_i^T\boldsymbol{X}_i\right)^{-1}\boldsymbol{X}_i^T\boldsymbol{y}, i = 1,2,\dots,C \tag{11}$$

$$\hat{\boldsymbol{y}}_{LRC,i} = \boldsymbol{H}_i\boldsymbol{y}, i = 1,2,\dots,C \tag{12}$$

and

$$i^* = \underset{i}{\operatorname{argmin}}\left\|\hat{\boldsymbol{y}}_{LRC,i} - \boldsymbol{y}\right\|_2 = \underset{i}{\operatorname{argmin}}\left\|\boldsymbol{H}_i\boldsymbol{y} - \boldsymbol{y}\right\|_2, i = 1,2,\dots,C \tag{13}$$

where $\boldsymbol{x}_{i,j}^{(m)}$ and $\boldsymbol{y}^{(m)}$ are the mth component of $\boldsymbol{x}_{i,j}$ and $\boldsymbol{y}$ column vectors, respectively.

However, these vectors satisfy the Cauchy-Schwarz inequality as:

$$\hat{\boldsymbol{\beta}}_{RR,i} = \underset{\boldsymbol{\beta}_i}{\operatorname{argmin}}\{\|\mathbf{y} - \boldsymbol{X}_i\boldsymbol{\beta}_i\|_2^2 + \lambda\|\boldsymbol{\beta}_i\|_2^2\}, i = 1,2,\dots,C \tag{14}$$

To ignore the amplitudes of two feature data vectors, the similarity measure can be also defined by a cosine criterion as:

$$\cos\theta_{i,j} = \frac{\boldsymbol{w}_{i,j}^T\widehat{\boldsymbol{y}}}{\|\boldsymbol{w}_{i,j}\|_2\|\widehat{\boldsymbol{y}}\|_2} = \frac{\left|\sum_{h=1}^M w_{i,j}^{(h)}\widehat{y}^{(h)}\right|^2}{\sum_{h=1}^M\left|w_{i,j}^{(h)}\right|^2\sum_{h=1}^M\left|\widehat{y}^{(h)}\right|^2} \tag{15}$$

3. Advances of subspace projection optimization

In this section, the advances of subspace projection optimization are presented for robust face recognition system. Then, the well-known subspace projection methods including LRC, RLRC, RR, IPCR, URC, LDRC, GLRC and TLR are introduced.

3.1. Linear regression classification (LRC)

For applying the linear regression to estimate the class specific model, all N training gray images from the same class are concatenated as:

$$\boldsymbol{P}_{PCAZ} = \{\boldsymbol{u}_{n+1}, \boldsymbol{u}_{n+2}, \dots, \boldsymbol{u}_P\}, P \leq M \tag{16}$$

where $\boldsymbol{X}_i$ is in the size of $M{\times}N$ and is called class-specific model. In other words, the ith class is represented by a vector space $\boldsymbol{X}_i$, which is called the regressor for each subject, in the training phase.

In the testing phase, if an unknown column vector $\boldsymbol{y}$ belongs to the ith class, its linear combination can be rewritten in terms of the training data from the ith class and can be formulated as:

$$\boldsymbol{w}_{PCAZ,i,j} = \boldsymbol{P}_{PCAZ}^T\widetilde{\boldsymbol{x}}_{i,j} \tag{17}$$

where $\boldsymbol{\beta}_i \in \boldsymbol{R}^{N\times1}$ is the vector of regression parameters. The goal of the linear regression is to find the regression parameters by minimizing the residual errors as:

$$\widehat{\boldsymbol{\beta}}_{PCAZ,i} = \underset{\boldsymbol{\beta}_i}{\text{argmin}}\left\{\left\|\boldsymbol{y}_{PCAZ} - \boldsymbol{w}_{PCAZ,i}\boldsymbol{\beta}_i\right\|_2^2\right\}, i = 1,2,\dots,C \tag{18}$$

The regression coefficients, $\boldsymbol{\beta}_i$, can be solved through the least-square estimation method and can be represented as:

$$\widehat{\boldsymbol{\beta}}_{PCAZ,i} = \left(\boldsymbol{X}_{PCAZ,i}^T\boldsymbol{X}_{PCAZ,i}\right)^{-1}\boldsymbol{X}_{PCAZ,i}^T\boldsymbol{y}_{PCAZ}, i = 1,2,\dots,C \tag{19}$$

For each class i, the regressed vector $\widehat{\boldsymbol{y}}_{LRC,i}$ can be predicted through the regression parameters $\widehat{\boldsymbol{\beta}}_{LRC,i}$ and predictors $\boldsymbol{X}_i$ as

$$\underset{\boldsymbol{P}_{URC}}{\text{argmin}} \sum_{i=1}^{C} \sum_{j=1}^{N} tr\left[\boldsymbol{P}_{URC}^{T}(\boldsymbol{x}_{i,j} - \tilde{\boldsymbol{x}}_i)(\boldsymbol{x}_{i,j} - \tilde{\boldsymbol{x}}_i)^T \boldsymbol{P}_{URC}\right] = \underset{\boldsymbol{P}_{URC}}{\text{argmin}} \, tr\left[\boldsymbol{P}_{URC}^{T}\boldsymbol{E}_{URC}\boldsymbol{P}_{URC}\right] \tag{20}$$

By substituting Equation (19) into Equation (20), the predicted response vector $\hat{\boldsymbol{y}}_i$ can be rewritten as:

$$\underset{\boldsymbol{P}_{LDRC}}{\text{argmin}} \, \frac{E_{BC}}{E_{WC}} \tag{21}$$

Theoretically, we can treat Equation (21) as a class-specific projection as:

$$\frac{E_{BC}}{E_{WC}} = = \frac{\frac{1}{NC(C-1)}\sum_{i=1}^{C}\sum_{j=1}^{N}\sum_{q=1,q\neq i}^{C}\left\|\tilde{x}_{i,j} - \hat{x}_{i,j,q}^{inter}\right\|^2}{\frac{1}{NC}\sum_{i=1}^{C}\sum_{j=1}^{N}\left\|\tilde{x}_{i,j} - \hat{x}_{i,j}^{intra}\right\|^2} \tag{22}$$

where $\hat{\boldsymbol{y}}_{LRC,i}$ is the projection of $\boldsymbol{y}$ onto the subspace of the ith class by the projection matrix $\boldsymbol{H}_i = \boldsymbol{X}_i(\boldsymbol{X}_i^T\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i^T$.

In the LRC approach, the minimum reconstruction error is adopted for determining the final result. In other words, the distance between predicted response vector $\hat{\boldsymbol{y}}_{LRC,i}$ and unknown column vector $\boldsymbol{y}$ will be smallest when the unknown column vector belongs to the training vector space of class i. Therefore, the identity i^* can be determined by minimizing the Euclidean distance between the predicted response vector and unknown vector as:

$$\boldsymbol{E}_w = \frac{1}{NC}\sum_{i=1}^{C}\sum_{j=1}^{N}(\boldsymbol{x}_{i,j} - \boldsymbol{x}_{i,j}^{intra})(\boldsymbol{x}_{i,j} - \boldsymbol{x}_{i,j}^{intra})^T \tag{23}$$

3.2. Robust linear regression classification (RLRC)

The LRC has been claimed that classical statistical methods are robust, but they are only robust in the fact of true cases. Once the data distribution is in fact of false cases, the regression parameter under original least square estimation could be inaccurate. In other words, the original least square estimation is inefficient and can be biased in the presence of outliers. There exist several approaches for robust estimation like R-estimator [30, 31] and L-estimator [30, 32]. However, M-estimator is now shown superiority due to their generality, efficiency and high breakdown point [30, 33]. Based on the M-estimator, the optimal function becomes:

$$\underset{\boldsymbol{P}_{LDRC}}{\text{argmin}} \, \frac{E_{BC}}{E_{WC}} = \underset{\boldsymbol{P}_{LDRC}}{\text{argmin}} \, \frac{\boldsymbol{P}_{LDRC}^{T}\boldsymbol{E}_b\boldsymbol{P}_{LDRC}}{\boldsymbol{P}_{LDRC}^{T}\boldsymbol{E}_w\boldsymbol{P}_{LDRC}} \tag{24}$$

where

$$\rho(\|\mathbf{y} - \mathbf{X}_i\boldsymbol{\beta}_i\|) = \begin{cases} \frac{1}{2\gamma}\|\mathbf{y} - \mathbf{X}_i\boldsymbol{\beta}_i\|^2 & , \text{ for } \|\mathbf{y} - \mathbf{X}_i\boldsymbol{\beta}_i\| \leq \gamma \\ \|\mathbf{y} - \mathbf{X}_i\boldsymbol{\beta}_i\| - \frac{1}{2}\gamma, & \text{ for } \|\mathbf{y} - \mathbf{X}_i\boldsymbol{\beta}_i\| > \gamma \end{cases} \tag{25}$$

and $\rho(\bullet)$ is a symmetric function and γ being a tuning constant, also called the Huber threshold.

3.3. Ridge regression (RR)

The goal of the RR is to find and minimize the residual errors and their penalty as:

$$\boldsymbol{E}_b\boldsymbol{u}_l = \lambda_l\boldsymbol{E}_w\boldsymbol{u}_l, l = 1, 2, \dots, \varphi \tag{26}$$

where λ is the regularization parameter. Comparing with linear regression, the RR adds a penalty, $\lambda\|\boldsymbol{\beta}_i\|_2^2$, to the regression model to reduce the variance of the model. The regression parameter vectors can be computed by:

$$\boldsymbol{y}_k = \boldsymbol{X}_{i,k}\boldsymbol{\beta}_{GLRC,i}, i = 1, 2, \dots, C; \; k = 1, 2, \dots, K \tag{27}$$

3.4. Improved principal component regression (IPCR)

Multicollinearity denotes the interrelations among the independent variables. In the linear regression, the regression estimation could be imprecise because the multicollinearity phenomenon would inflate the variance and covariance. To overcome the problem of multicollinearity, various approaches have been proposed. IPCR is one of the powerful approaches.

The IPCR is a two-step classification method. In the first step, the PCAZ is adopted to transform the observed variables into the new decorrelated components. Then, the first n components are dropped because these components are very sensitive to the lighting changes. Mathematically, the PCA process is used in all training samples including covariance matrix evaluation as Equation (2), and eigen-decomposition estimation as Equation (3). Then, we can obtain a set of eigenvectors, $\boldsymbol{u}=\{u_1, u_2, \dots, u_M\}$, and a set of eigenvalues, $r=\{r_1, r_2, \dots, r_M\}$ with $r_1 \geq r_2 \geq \dots \geq r_M$. As above mentioned, we drop first n components and the projection matrix can be express as:

$$\widehat{\boldsymbol{\beta}}_{GLRC,i} = \underset{\boldsymbol{\beta}_{GLRC,i}}{\operatorname{argmin}} \left\{ \sum_{k=1}^{K} \left(\boldsymbol{y}_k - \boldsymbol{X}_{i,k}\boldsymbol{\beta}_{GLRC,i}\right)^T \left(\boldsymbol{y}_k - \boldsymbol{X}_{i,k}\boldsymbol{\beta}_{GLRC,i}\right) \right\} \tag{28}$$

The PCAZ features, $w_{PCAZ,i,j} \in \boldsymbol{R}^{P\times1}$, can be obtained by multiplying the projection matrix and the average image vector as:

$$\boldsymbol{r}_k = \sum_{i=1}^{C} |\widehat{\boldsymbol{y}}_{i,k} - \boldsymbol{y}_k| \tag{29}$$

In order to apply LRC to estimate class specific model, feature vectors should be grouped according to the class-membership. Hence, for the ith class, we have

$w_{PCAZ,i}$=$[w_{PCAZ,i,1}, w_{PCAZ,i,2}, \ldots, w_{PCAZ,i,N}]$. In the testing phase, an unknown column vector, y, is transformed to PCAZ subspace as $y^{(PCAZ)}$. In the second step, the new subspace of PCAZ projection is used in LRC such that we can seek more reliable regression coefficients for each subject for face recognition. The goal of regression becomes to minimize the residual errors as:

$$y_k = X_{i,k}\widetilde{\boldsymbol{\beta}}_{GLRC,i} \tag{30}$$

The regression parameter vectors can be rewritten as a matrix form as:

$$\widetilde{\boldsymbol{\beta}}_{GLRC,i} = \underset{\boldsymbol{\beta}_{GLRC,i}}{\arg\min} \sum_{k=1}^{K} \alpha_k \left(y_k - X_{i,k}\widetilde{\boldsymbol{\beta}}_{GLRC,i}\right)^2 \tag{31}$$

3.5. Unitary regression classification (URC)

The total within-class projection error from all classes cannot be taken in previous mentioned methods for classification that would degrade the recognition accuracy. The URC is proposed to minimize the total within-class projection error from all classes for LRC to improve the robustness for pattern recognition.

Instead of original space, we hope to find a global unitary rotation P_{URC}=$[s_1,\ldots,s_\Psi]$ with $\Psi \leq M$, which can rotate the original data space to a new compact w_{URC} data space as:

$$w_{URC,i,j} = P_{URC}^T x_{i,j} \tag{32}$$

to achieve the total minimum projection error of all training data stated as:

$$\underset{P_{URC}}{\arg\min} \sum_{i=1}^{C} \sum_{j=1}^{N} \left\| w_{URC,i,j} - \widetilde{w}_i \right\|^2 \tag{33}$$

where $\widetilde{w}_i$=$\widetilde{H}_{URC,i}w_{URC,i,j}$ is the within-class projection to make the objective function be well-posed. In w_{URC} data space, the ith class projection matrix can be obtained by following $\widetilde{H}_{URC,i}$=$W_{URC,i}\left(W_{URC,i}^T W_{URC,i}\right)^{-1}W_{URC,i}^T$ where $W_{URC,i}$=$[W_{URC,i,1}, W_{URC,i,2}, \ldots, W_{URC,i,N}]$. The unitary rotation matrix, P_{URC}, is used to achieve the total minimum within-class projection error for LRC. From minimum reconstruction error, the objective function in T data space can be represented as:

$$\underset{P_{URC}}{\arg\min} \sum_{i=1}^{C} \sum_{j=1}^{N} \left\| w_{URC,i,j} - \widetilde{H}_{URC,i}w_{URC,i,j} \right\|^2$$
$$= \underset{P_{URC}}{\arg\min} \sum_{i=1}^{C} \sum_{j=1}^{N} \left\| P_{URC}^T x_{i,j} - \widetilde{H}_{URC,i}P_{URC}^T x_{i,j} \right\|^2 \tag{34}$$

By substituting $W_{URC,i}$=$P_{URC}^T X_i$ into $\widehat{H}_{URC,i}$=$W_{URC,i}\left(W_{URC,i}^T W_{URC,i}\right)^{-1}W_{URC,i}^T$, the objective function becomes:

$$\underset{\boldsymbol{P}_{URC}}{\text{argmin}} \sum_{i=1}^{C} \sum_{j=1}^{N} tr\left[\boldsymbol{P}_{URC}^{T}(\boldsymbol{x}_{i,j} - \widetilde{\boldsymbol{x}}_i)(\boldsymbol{x}_{i,j} - \widetilde{\boldsymbol{x}}_i)^{T}\boldsymbol{P}_{URC}\right] = \underset{\boldsymbol{P}_{URC}}{\text{argmin}} \, tr[\boldsymbol{P}_{URC}^{T}\boldsymbol{E}_{URC}\boldsymbol{P}_{URC}] \qquad (35)$$

where $E_{URC} = \sum_{i=1}^{C} \sum_{j=1}^{N} (\boldsymbol{x}_{i,j} - \widetilde{\boldsymbol{x}}_i)(\boldsymbol{x}_{i,j} - \widetilde{\boldsymbol{x}}_i)^{T}$, also called within-class projection error matrix. The projection matrix, $\boldsymbol{P}_{URC} = [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_\Psi]$, can be solved by evaluating eigen-decomposition as:

$$\boldsymbol{E}_{URC}\boldsymbol{s}_l = \lambda_l \boldsymbol{s}_l, l = 1, 2, \ldots, \Psi \qquad (36)$$

where $\lambda_\Psi \geqq \ldots \geqq \lambda_l \geqq \ldots \geqq \lambda_1 \geqq 0$.

3.6. Linear discriminant regression classification (LDRC)

Although the previous methods including LRC, RLRC, and IPCRC can perform well on face recognition, we cannot guarantee that the projection subspace in LRC or IPCRC is most discriminatory. When the projection subspaces among the different subjects overlap, the recognition result would be incorrect. To obtain an effective discriminant subspace for LRC, the LRC with discriminant analysis is presented by maximizing the ratio of the between-class reconstruction error (BCRE) to the within-class reconstruction error (WCRE) by the LRC.

Mathematically, all images are collected from C classes as $X = [\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_C] = [\boldsymbol{x}_{1,1}, \ldots, \boldsymbol{x}_{i,j}, \ldots, \boldsymbol{x}_{C,N}]$. LDRC is to find an optimal projection by maximizing the BCRE over the WCRE for the LRC such that the LRC on the optimal subspace has better discrimination for classification. The goal of LDRC is to maximize the objective function as:

$$\underset{\boldsymbol{P}_{LDRC}}{\text{argmin}} \frac{E_{BC}}{E_{WC}} \qquad (37)$$

where $\boldsymbol{P}_{LDRC} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_\varphi]$ is the optimal projection matrix, and E_{BC} and E_{WC} denote the BCRE and WCRE, respectively. The original space, $\boldsymbol{x}_{i,j}$, can be mapped into the subspace, $\widetilde{\boldsymbol{x}}_{i,j} = \boldsymbol{P}_{LDRC}^{T}\boldsymbol{x}_{i,j}$. Hence, the objective function can be rewritten as:

$$\frac{E_{BC}}{E_{WC}} == \frac{\frac{1}{NC(C-1)}\sum_{i=1}^{C}\sum_{j=1}^{N}\sum_{q=1,q\neq i}^{C}\left\|\widetilde{\boldsymbol{x}}_{i,j} - \widehat{\boldsymbol{x}}_{i,j,q}^{inter}\right\|^{2}}{\frac{1}{NC}\sum_{i=1}^{C}\sum_{j=1}^{N}\left\|\widetilde{\boldsymbol{x}}_{i,j} - \widehat{\boldsymbol{x}}_{i,j}^{intra}\right\|^{2}} \qquad (38)$$

where $\widetilde{\boldsymbol{x}}_{i,j,q}^{inter} = \boldsymbol{H}_q^{\widetilde{\boldsymbol{x}}}\widetilde{\boldsymbol{x}}_{i,j}$ denotes the inter-class projection of $\widetilde{\boldsymbol{x}}_{i,j}$ by the LRC from the different q^{th} class and $\boldsymbol{x}_{i,j}^{intra} = H_{i,j}^{\widetilde{\boldsymbol{x}}}\widetilde{\boldsymbol{x}}_{i,j}$ denotes the intra-class projection of $\widetilde{\boldsymbol{x}}_{i,j}$ by the LRC in the same class. The $\boldsymbol{x}_{i,j}$ is used to instead of $\widetilde{\boldsymbol{x}}_{i,j}$ as:

$$\frac{E_{BC}}{E_{WC}} = \frac{\frac{1}{NC(C-1)}\sum_{i=1}^{C}\sum_{j=1}^{N}\sum_{q=1,q\neq i}^{C}\left\|\widetilde{\boldsymbol{x}}_{i,j} - \widehat{\boldsymbol{x}}_{i,j,q}^{inter}\right\|^{2}}{\frac{1}{NC}\sum_{i=1}^{C}\sum_{j=1}^{N}\left\|\widetilde{\boldsymbol{x}}_{i,j} - \widehat{\boldsymbol{x}}_{i,j}^{intra}\right\|^{2}} = \frac{\frac{1}{NC(C-1)}\sum_{i=1}^{C}\sum_{j=1}^{N}\sum_{q=1,q\neq i}^{C}\left\|\boldsymbol{P}_{LDRC}^{T}\boldsymbol{x}_{i,j} - \boldsymbol{H}_q^{\widetilde{\boldsymbol{x}}}\boldsymbol{P}_{LDRC}^{T}\boldsymbol{x}_{i,j}\right\|^{2}}{\frac{1}{NC}\sum_{i=1}^{C}\sum_{j=1}^{N}\left\|\boldsymbol{P}_{LDRC}^{T}\boldsymbol{x}_{i,j} - \boldsymbol{H}_{i,j}^{\widetilde{\boldsymbol{x}}}\boldsymbol{P}_{LDRC}^{T}\boldsymbol{x}_{i,j}\right\|^{2}} \qquad (39)$$

With some algebraic deduction, the form becomes:

$$\frac{E_{BC}}{E_{WC}} = \frac{\frac{1}{NC(C-1)}\sum_{i=1}^{C}\sum_{j=1}^{N}\sum_{q=1,q\neq i}^{C} tr\left[P_{LDRC}^{T}(x_{i,j}-x_{i,j,q}^{inter})(x_{i,j}-x_{i,j,q}^{inter})^{T}P_{LDRC}\right]}{\frac{1}{NC}\sum_{i=1}^{C}\sum_{j=1}^{N} tr\left[P_{LDRC}^{T}(x_{i,j}-x_{i,j}^{intra})(x_{i,j}-x_{i,j}^{intra})^{T}P_{LDRC}\right]} = \frac{tr(P_{LDRC}^{T}E_{b}P_{LDRC})}{tr(P_{LDRC}^{T}E_{w}P_{LDRC})} \tag{40}$$

where

$$E_{b} = \frac{1}{NC(C-1)}\sum_{i=1}^{C}\sum_{j=1}^{N}\sum_{q=1,q\neq i}^{C}(x_{i,j}-x_{i,j,q}^{inter})(x_{i,j}-x_{i,j,q}^{inter})^{T} \tag{41}$$

and

$$E_{w} = \frac{1}{NC}\sum_{i=1}^{C}\sum_{j=1}^{N}(x_{i,j}-x_{i,j}^{intra})(x_{i,j}-x_{i,j}^{intra})^{T} \tag{42}$$

is inter-class and intra-class reconstruction error, respectively. In other words, the objective function can be represented as:

$$\underset{P_{LDRC}}{\text{argmin}}\ \frac{E_{BC}}{E_{WC}} = \underset{P_{LDRC}}{\text{argmin}}\ \frac{P_{LDRC}^{T}E_{b}P_{LDRC}}{P_{LDRC}^{T}E_{w}P_{LDRC}} \tag{43}$$

For solving the optimization problem, Equation (43) can be reformulated as the following:

$$\underset{P_{LDRC}}{\text{argmin}}\ P_{LDRC}^{T}E_{b}P_{LDRC}, s.t. P_{LDRC}^{T}E_{w}P_{LDRC} = \vartheta \tag{44}$$

where ϑ is a constant. The projection matrix, $P_{LDRC}=[u_{1}, u_{2},\ldots, u_{\varphi}]$, can be solved by evaluating eigen-decomposition as:

$$E_{b}u_{l} = \lambda_{l}E_{w}u_{l}, l = 1,2,\ldots,\varphi \tag{45}$$

where $\lambda_{1}\geqq\ldots\geqq\lambda_{l}\geqq\ldots\geqq\lambda_{\varphi}$.

3.7. Generalized linear regression classification (GLRC)

In real-world recognition applications, the input images generally have multiple components which can overcome the unexpected effects such as pose variations, limited image information and so on. For color face recognition, the GLRC with membership grade (MG) criteria is proposed to defend the unexpected effects.

Mathematically, each channel component is separately normalized and transformed to one column vector such that $v_{i,j,k}\in R_{p\times q\times K} \rightarrow x_{i,j,k}\in R_{d\times K}$, where $d = p\cdot q$. In the ith class, the kth component of N training images is collected as:

$$X_{i,k} = [x_{i,1,k}, x_{i,2,k}, \dots, x_{i,N,k}] \in R^{d \times N} \tag{46}$$

for $i = 1, 2, \dots, C$ and $k = 1, 2, \dots, K$, where $X_{i,k}$ is treated as the kth-channel collected training data of the ith class in the training phase.

For the test image, the kth-channel testing image, z_k, is normalized and reshaped into a column vector as $y_k \in R^{d \times 1}$. For the kth component, the linear combination of $X_{i,k}$ from the ith class for the test vector y_k becomes:

$$y_k = X_{i,k} \boldsymbol{\beta}_{GLRC,i}, i = 1,2,\dots,C; \; k = 1,2,\dots,K \tag{47}$$

where $\boldsymbol{\beta}_{GLRC,i} \in R^{N \times 1}$ is an ideal projection vector of the ith-class regression parameter for all channels. In order to estimate the projection vector, the objective function becomes:

$$\widehat{\boldsymbol{\beta}}_{GLRC,i} = \underset{\boldsymbol{\beta}_{GLRC,i}}{\operatorname{argmin}} \left\{ \sum_{k=1}^{K} (y_k - X_{i,k} \boldsymbol{\beta}_{GLRC,i})^T (y_k - X_{i,k} \boldsymbol{\beta}_{GLRC,i}) \right\} \tag{48}$$

After solving the optimization problem, the regression vector can be expressed as:

$$\widehat{\boldsymbol{\beta}}_{GLRC,i} = \left(\sum_{k=1}^{K} X_{i,k}^T X_{i,k} \right)^{-} \left(\sum_{k=1}^{K} X_{i,k}^T y_k \right) \tag{49}$$

In order to achieve optimal performance, the different components should be treated as unequally important. Thus, the absolute sum of prediction residual of the kth component after the direct least square optimization is given as:

$$r_k = \sum_{i=1}^{C} |\widehat{y}_{i,k} - y_k| \tag{50}$$

where $\widehat{y}_{i,k} = X_{i,k} \boldsymbol{\beta}_{GLRC,i}$, $i = 1, 2, \dots, C$. Based on the statistical opinion, we define the importance of the kth component to be inverse of the normalized absolute sum of prediction residual, which is expressed by:

$$\alpha_k = \frac{1}{r_k + \varepsilon} \sum_{k=1}^{K} r_k \tag{51}$$

where ε is a tiny value which is used to avoid $r_k = 0$. The larger the residual, r_k is, the less important the kth component will be. For the GRLC optimization, we propose the linear combination of $X_{i,k}$ of the kth component in the ith class for the test vector y_k becomes:

$$y_k = X_{i,k} \widetilde{\boldsymbol{\beta}}_{GLRC,i} \tag{52}$$

where $\tilde{\beta}_{GLRC,i} \in R^{N \times 1}$ is the vector of the ith-class total regression parameters to achieve the GRLC optimization as:

$$\tilde{\beta}_{GLRC,i} = \underset{\beta_{GLRC,i}}{\operatorname{argmin}} \sum_{k=1}^{K} \alpha_k \left(y_k - X_{i,k} \tilde{\beta}_{GLRC,i} \right)^2 \tag{53}$$

The optimal total regression parameter vector, $\tilde{\beta}_{GLRC,i}$ can be given by:

$$\tilde{\beta}_{GLRC,i} = \left(\sum_{k=1}^{K} \alpha_k X_{i,k}^T X_{i,k} \right)^{-} \left(\sum_{k=1}^{K} \alpha_k X_{i,k}^T y_k \right) \tag{54}$$

The prediction, $\hat{y}'_{i,k}$ is then expressed as $\hat{y}'_{i,k} = X_{i,k} \tilde{\beta}_{GLRC,i}$.

For identity recognition, the minimum prediction error of the GRLC should be further designed to compute the similarity between the prediction vector $\hat{y}'_{i,k}$ and the query vector y. The similarity in terms of minimization of prediction errors of total K components can be designed by the following MG criteria as:

$$i^* = \underset{i}{\operatorname{argmin}} \left\{ \sum_{k=1}^{K} \left(1 + \left(\frac{d_{i,k}}{\bar{d}_k + \varepsilon} \right)^t \right)^{-1} \right\} \tag{55}$$

where $d_{i,k} = \alpha_k \| \hat{y}'_{i,k} - y_k \|$, $\bar{d}_k = \frac{1}{N} \sum_{i=1}^{N} \alpha_k \| \hat{y}'_{i,k} - y_k \|$ and t is the pre-selected fuzzy factor.

3.8. Trimmed linear regression (TLR)

For the occlusion situations, the previous methods including LRC, RLRC, IPCR, URC, LDRC, and GLRC are not suitable because the existing methods treat all pixels as equally import. Conversely, if the outliers can be detected and trimmed from the testing image and the corresponding training samples, the mechanism still can work. Hampel identifier [34, 35] for outlier detection is highly thought of by the researchers because it can make out the extreme values easily. An advantage of Hampel identifier is that it adopts median absolution deviation (MAD), which is a powerful measure in statistics, for removing the masking data. Mathematically, the Hampel identifier can be expressed as:

$$\frac{|\Delta - media(\Delta)|}{MAD/0.6745} > 2.24 \tag{56}$$

where Δ is a data set, media(Δ) denotes the media value of Δ data set. The number of 0.6745 is a probable error of standard deviation. When the ratio is larger than 2.24, the data will be abandoned. For example, there is a data set, [2, 3, 3, 4, 4, 250]. The sample mean is 44.33, sample variance is 100.76, sample median is 3.5, MAD equals to 0.5, and the detection rule by mean and median is:

$$\frac{|250-44.33|}{100.76} = 2.04 \tag{57}$$

and

$$\frac{|250-3.5|}{0.5 \cdot 0.6745} = 332.52 \tag{58}$$

respectively. We can observe that the Hampel identifier excludes the outlier easier than the other one.

For the face recognition, the error of estimation can be presented as:

$$\boldsymbol{e}_i = \boldsymbol{y} - \boldsymbol{X}_i \boldsymbol{\beta}_i \tag{59}$$

where error is a zero mean distribution. In order to detect the occlusion part, each pixel should suffice the Hampel identifier estimation as:

$$\varepsilon' = \left\{ \zeta \mid \frac{|e_i[\zeta]-0|}{\frac{median(|e_i[\zeta]-0|)}{0.6745}} < 2.24 \right\} \tag{60}$$

where ε' is the indices of all pixels, that is $\varepsilon' = \{1, 2, \ldots, M\}$. The real median of noise is zero. From the Equation (60), the pure pixels, ε', are found out. In other words, the pure pixels are taken for regression estimation. The training data can be rewritten as $\boldsymbol{X}_{TLR,i} = [\boldsymbol{x}_{TLR,i,1}, \boldsymbol{x}_{TLR,i,2}, \ldots, \boldsymbol{x}_{TLR,i,N}] \in R^{\tau \times N}$ and testing sample becomes $\boldsymbol{y}_{TLR} \in R^{\tau \times 1}$ where τ is the number of elements in ε' and $\tau < M$. The objective function becomes:

$$\widetilde{\boldsymbol{\beta}}_{TLR,i} = \underset{\boldsymbol{\beta}_{TLR,i}}{\operatorname{argmin}} \left\| \boldsymbol{y}_{TLR} - \boldsymbol{X}_{TLR}\widetilde{\boldsymbol{\beta}}_{TLR,i} \right\|_2^2 \tag{61}$$

The regression parameter vectors can be represented as:

$$\widehat{\boldsymbol{\beta}}_{TLR,i} = \left(\boldsymbol{X}_{TLR,i}^T \boldsymbol{X}_{TLR,i} \right)^{-1} \boldsymbol{X}_{TLR,i}^T \boldsymbol{y}_{TLR} \tag{62}$$

4. Experimental results

In order to verify the recognition accuracy, the well-known databases including Yale B, AR, FERET, and FEI are utilized. In the experiments, we evaluate the mentioned method against low resolution problem coupled with facial expressions, illumination changes, pose variations, and partial occlusions.

4.1. Yale B database

The Yale B database contains 10 subjects [36, 37]. Each subject has 64 illumination images with 9 different poses. The Yale B can be divided into five subset based on angle of the light source directions as shown in **Figure 4**. In the experiments, the first subset with normal pose is used for training and the remaining subsets (Subset 2 to 5) with normal pose are utilized for testing. All images are cropped and resized to 30×25 pixels. **Table 1** reveals that IPCRC performs better than the traditional subspace projection like PCA and LDA. Moreover, the IPCRC can also outperform the LRC, RLRC and RR. The reason is that the original subspace cannot represent the data distribution very well. Besides, PCA subspace is very sensitive to illuminant variations. However, IPCRC not only can transform to PCA subspace, but also can defend the illumination variations by removing the top n components. Thus, IPCRC possesses higher robustness to illuminations than the other methods.

Methods	Subset 2	Subset 3	Subset 4	Subset 5
PCA	89.81	47.04	21.90	15.26
LDA	95.14	75.14	34.76	10.00
LRC	100.00	100.00	91.86	52.11
RLRC	100.00	100.00	92.86	60.00
RR	100.00	100.00	92.57	53.68
IPCRC	100.00	100.00	95.00	64.21
LDRC	100.00	100.00	97.14	56.84
URC	100.00	100.00	90.71	53.68

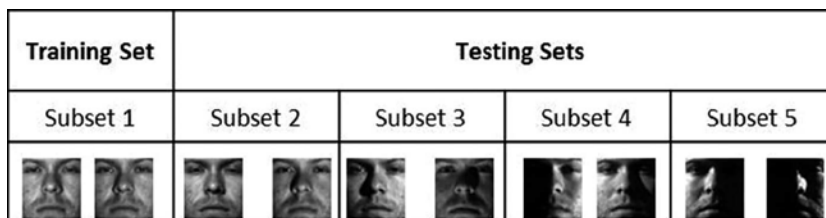**Table 1.** Accuracy (%) comparisons on Yale B.

Figure 4. The experimental design and some samples of cropped and aligned illustration from Yale B face database.

4.2. FERET database

Furthermore, we experiment on the FERET face database [38, 39] for the purpose of verifying the performance among the different subspace projections. In the experiments, we select four facial images including fa, fb, ql, and qr from 300 subjects as **Figure 5**. All images are converted, cropped, and downsampled to 30×25 pixels with grayscale. As the **Figure 5** shown, the fa and

fb samples are small pose and rotation changes; conversely, the ql and qr samples are major pose variations. In order to obtain a reliable result, cross-validation experimental procedure is adopted. In other words, three images per person are used for training while the fourth image is used for testing. **Table 2** shows that the average recognition accuracy (ARA) in URC performs outstandingly. We can observe that the RLRC and IPCRC are highly sensitive to pose variations but in spite of these, methods perform well in noisy and illuminated face images, respectively.

	PCA	LDA	LRC	RLRC	RR	IPCRC	URC
fa	80.67	87.33	94.00	91.67	85.67	92.33	96.00
fb	81.00	84.33	92.33	90.00	83.67	91.00	95.33
ql	65.67	63.00	71.00	69.33	63.33	66.00	73.00
qr	68.33	72.00	75.00	74.33	70.00	68.67	84.33
ARA	73.92	76.67	83.08	81.33	75.67	79.50	87.17

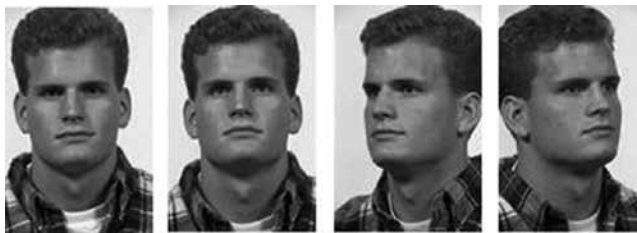**Table 2.** Accuracy (%) comparisons on FERET.

Figure 5. Samples (fa, fb, ql, qr) of one subject from FERET face database.

4.3. AR database

AR face database [40, 41] was conducted by Martinez and Benavente in 1998. This database contains 4000 mug shots of 126 subjects (70 males and 56 females) with different variations such as facial expressions, lighting changes and partial occlusions. For normal case, each subject contains 26 images in two sessions. The first session (AR1 ~ AR13), containing 13 photos, includes facial expression, different lighting changes, and partial occlusions (sunglasses and scarf) with lighting changes. The second session (AR14 ~ AR26) duplicates the same way of first session two weeks later as shown in **Figure 6**. In the experiments, 100 subjects are selected and all images are cropped and resized into 30×25 pixels with grayscale. We classify the images into four different expressions including neutral (AR4, AR14), happy (AR2, AR3), angry (AR1, AR17), and screaming (AR15, AR16) expressions. The single-one-expression training strategy is adopted to present the performance. For example, if neutral expression images are used for training, the happy, angry, and screaming expressions are used as query images. **Table 3** reveals that the LDRC achieves the best performance in all cases. Moreover,

we can observe that the happy expression images for training obtain higher performance than the others; conversely, the screaming expression images for testing can obtain lowest performance. On the other hand, the partial occlusion situations are used to discussion. In this experiments, the expression variation images (AR1~AR4, AR14~AR17) are utilized as training set, and testing sets are separated in two cases including sunglasses (AR8, AR21) and scarf (AR11, AR24). All images are cropped and resized into 42×30 pixels with grayscale. In the **Table 4**, we can observe two points. First, the TLRC can perform better than the other methods under sunglasses occlusion or scarf occlusion. Second, the upper bound occlusion seems to obtain higher performance than the lower bound occlusion. In other words, the mouth features are more useful than the eye features.

Training Set	Testing Set	PCA	LDA	LRC	RLRC	RR	IPCRC	LDRC	GLRC
N	H,A,S	88.75	86.53	89.03	87.50	84.17	88.19	91.39	89.00
H	N,A,S	90.97	89.17	92.08	91.67	82.33	92.08	93.61	54.33
A	N,H,S	90.00	85.69	90.00	89.17	83.33	89.44	90.42	89.83
S	N,H,A	88.19	84.58	87.78	86.53	77.88	87.58	90.42	55.50

Table 3. Accuracy (%) comparisons on AR.

Figure 6. Samples of one subject from AR face database.

Training Set	Testing Set	PCA	LDA	SRC	LRC	RLRC	RR	IPCRC	GLRC	TLRC
AR1~ AR4; AR14~ AR17	Sunglasses (AR8, AR21)	42.5	20.5	87.0	65.5	47.5	59.0	44.5	90.5	100.0
	Scarf (AR11, AR24)	7.0	33.5	59.5	12.5	9.5	10.5	6.0	35.5	94.5

Table 4. Accuracy (%) comparisons under partial occlusion problem on AR.

4.4. FEI database

The FEI face database [42, 43] contains 200 subjects (100 males and 100 females). Each subject has 14 images with different pose variations (image1~image10), facial expressions (image11~image12), and illumination variations (image13~image14) as shown in **Figure 7**. In the experiments, all images are resized to 24×20 pixels with grayscale and the "leave-one-out" strategy" is adopted. From **Table 5**, it can be seen that the IPCRC is more robust to severe lighting variation (image 14) and URC is good at facial profiles (image 1, image 10). All in all, the ARA of URC performs the best.

		PCA	LDA	LRC	RLRC	RR	IPCRC	LDRC	URC	GLRC
Test Image	1	91.0	91.5	92.0	89.0	89.0	86.0	95.0	95.0	97.0
	2	99.5	99.0	100.0	100.0	100.0	99.0	99.5	100.0	100.0
	3	97.0	99.0	99.5	99.5	99.5	99.5	99.0	100.0	100.0
	4	98.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	5	97.5	99.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	6	96.5	99.5	99.0	99.0	99.0	99.0	99.5	100.0	99.5
	7	99.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	8	99.0	99.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	9	97.0	99.5	99.5	99.5	99.5	99.5	100.0	100.0	100.0
	10	79.5	83.0	83.5	78.0	78.0	72.0	86.0	91.5	91.5
	11	98.0	100.0	99.5	99.5	99.5	99.5	100.0	99.5	100.0
	12	97.0	99.5	99.0	99.0	99.0	97.5	99.0	98.5	99.5
	13	47.0	87.5	97.0	97.5	97.5	99.0	98.5	99.5	94.5
	14	23.5	39.5	79.5	91.0	91.0	92.5	83.0	88.5	77.5
ARA		92.04	96.73	97.62	97.00	97.00	96.23	97.11	98.77	97.11

Table 5. Accuracy (%) comparisons on FEI.

Figure 7. Samples of one subject from FEI database.

4.5. Discussions

From the experimental results, we can observe that IPCRC has a good performance under illumination situation. The reason is that the first n components in IPCRC are removed. The first n components are very sensitive to the lighting changes. However, although IPCRC has better performance under the lighting changes, it cannot handle the pose variations and occlusion problems very well. For the pose variations, the URC performs better than the other subspace methods because URC attempts to minimize the total intra-class reconstruction error to find an optimal projection which can decrease the pose influence. LDRC embeds discriminant analysis into the LRC for seeking an optimal projection matrix such that the LRC on that subspace has high discriminatory ability for classification. In other words, LDRC can perform better than LRC and IPCRC in most cases. In the occlusion situation, the TLRC can effectively remove the masking data and project onto a more reliable subspace.

5. Conclusions

In this chapter, we presented several subspace projection methods for robust face recognition to deal with different practical situations such as pose variations, lighting changes, facial expressions, and partial occlusions.

For illumination variation task in face recognition, an improved principal component classification can be used to solve the multicollinearity problem and can perform better recognition accuracy than the original linear regression and RR. For the pose variations, a URC has been presented to minimize the total within-class projection error from all classes for LRC to improve the robustness for pattern recognition. Moreover, a LDRC has been proposed to overcome facial expressions by maximizing the ratio of the BCRE to the WCRE by the LRC. For the partial occlusions, a trimmed regression classification is used to remove unreliable pixels by the Hampel identifier. Finally, experimental results have revealed the comparisons with different subspace projection optimizations.

Author details

Yang-Ting Chou, Jar-Ferr Yang[*] and Shih-Ming Huang

*Address all correspondence to: jarferryang@gmail.com

Institute of Computer and Communication Engineering, Department of Electrical Engineering, National Cheng Kung University, Tainan, Taiwan

References

[1] Yang, J.; Zhang, D.; Frangi, A.F.; Yang, J-Y. Two-dimensional PCA: a new approach to appearance-based face representation and recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2004;26(1):131–137.

[2] Shawe-Taylor, J.; Cristianini, N., editors. Kernel methods for pattern analysis. Cambridge University Press Inc., 2004. ISBN: 0521813972.

[3] Turk, M.; Pentland, A. Eigenfaces for recognition. Journal of Cognitive Neuroscience. 1991;3(1):71–86.

[4] Belhumeur, P.N.; Hespanha, J.P.; Kriegman, D. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1997;19(7):711–720.

[5] Martínez, A.M.; Kak, A.C. Pca versus lda. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2001;23(2):228–233.

[6] Schölkopf, B., Smola, A.; Müller, K.R. Nonlinear component analysis as a kernel eigenvalue problem. Neural Computation. 1998;10(5):1299–1319.

[7] Yang, M.H. Kernel eigenfaces vs. kernel fisherfaces: face recognition using kernel methods. In: 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition; 2002. p. 0215.

[8] Naseem, I.; Togneri, R.; Bennamoun, M. Linear regression for face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2010;32(11):2106–2112.

[9] Naseem, I.; Togneri, R.; Bennamoun, M. Robust regression for face recognition. Pattern Recognition. 2012;45(1):104–118.

[10] Xue, H.; Zhu, Y.; Chen, S. Local ridge regression for face recognition. Neurocomputing. 2009;72(4):1342–1346.

[11] Huang, S.M.; Yang, J.F. Improved principal component regression for face recognition under illumination variations. IEEE Signal Processing Letters. 2012;19(4):179–182.

[12] Huang, S.M.; Yang, J.F. Linear discriminant regression classification for face recognition. IEEE Signal Processing Letters. 2013;20(1):91–94.

[13] Huang, S.M.; Yang, J.F. Unitary regression classification with total minimum projection error for face recognition. IEEE Signal Processing Letters. 2013;20(5):443–446.

[14] Chou, Y.T.; Yang, J.F. Identity recognition based on generalised linear regression classification for multi-component images. IET Computer Vision. 2016;10(1):18–27.

[15] Chou, Y.T.; Yang, J.F. Object recognition based on generalized linear regression classification in use of color information. In: 2014 IEEE Asia Pacific Conference on In-Circuits and Systems; 2014. 272–275.

[16] Lai, J.; Jiang, X. Robust face recognition using trimmed linear regression. In: ICASSP; 2013. 2979–2983.

[17] Yang, G.; Huang, T.S. Human face detection in a complex back-ground. Pattern Recognition. 1994;27(1):53–63.

[18] Hotta, K.; Kurita, T.; Mishima, T. Scale invariant face detection method using higher-order local autocorrelation features extracted from log-polar image. In: Third IEEE International Conference on Automatic Face and Gesture Recognition; 1998. 70–75.

[19] Han, C.C.; Liao, H.Y.M.; Yu, G.J.; Chen, L.H. Fast face detection via morphology-based pre-processing. Pattern Recognition. 2000;33(10):1701–1712.

[20] Fasel, B. Fast multi-scale face detection. IDIAP-Com-04-1998. 1998.

[21] Rowley, H.A.; Baluja, S.; Kanade, T. Neural network-based face detection. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1998;20(1):23–38.

[22] Viola, P.; Jones, M.J. Robust real-time face detection. International Journal of Computer Vision. 2004;57(2):137–154.

[23] Lienhart, R.; Maydt, J. An extended set of haar-like features for rapid object detection. In: 2002 International Conference on Image Processing; 2002. 1–900.

[24] Kotropoulos, C.; Pitas, I. Rule-based face detection in frontal views. In: 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing; 1997. 2537–2540.

[25] Birchfield, S. An elliptical head tracker. In: IEEE Conference Record of the Thirty-First Asilomar Conference on Signals, Systems & Amp; 1997. 1710–1714.

[26] Wu, H.; Chen, Q.; Yachida, M. Face detection from color images using a fuzzy pattern matching method. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1999;21(6):557–563.

[27] Orfanidis, S. SVD, PCA, KLT, CCA, and All That. Rutgers University Electrical and Computer Engineering Department, Optimum Signal Pro-cessing. 2007;1–77.

[28] Golub, G.H.; Van Loan, C.F., editors. Matrix Computations. JHU Press. 2012.

[29] Watkins, D.S., editors. Fundamentals of Matrix Computations. John Wiley & Sons Inc.. 2004. ISBN: 978-0-470-52833-4.

[30] Huber, P.J., editors. Robust Statistics. Berlin Heidelberg: Springer. 2011.

[31] Heyde, C.C., editors. Quasi-Likelihood and Its Application: A General Approach to Optimal Parameter Estimation. Springer-Verlag New York Berlin Heidelerg Inc.. 1997. ISBN: 0-387-98225-6.

[32] Fraiman, R.; Meloche, J.; García-Escudero, L.A.; Gordaliza, A.; He, X.; Maronna, R.; Yohai, V.J.; Sheather, S.J.; McKean, J.W.; Small, C.G.; Wood, A.; Fraiman, R.; Meloche, J., editors. Multivariate L-Estimation. Test. 1999; 8(2): 255-317.

[33] Hampel, F.R.; Ronchetti, E.M.; Rousseeuw, P.J.; Stahel, W.A., editors. Robust Statistics: The Approach Based on Influence Functions. John Wiley & Sons Inc.. 2011. ISBN: 9781118186435.

[34] Wilcox, R.R., editors. Applying Contemporary Statistical Techniques. Elsevier Inc.. Gulf Professional Publishing. 2003. ISBN: 978-0-12-751541-0.

[35] Wilcox, R.R., editors. Introduction to Robust Estimation and Hypothesis Testing. Academic Press. 2012.

[36] Georghiades, A.S.; Belhumeur, P.N.; Kriegman, D. From few to many: illumination cone models for face recognition under variable lighting and pose. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2001;23(6):643–660.

[37] Georghiades, A.S.; Belhumeur, P.N.; Kriegman, D. Extended Yale Face Database B. Available from: http://vision.ucsd.edu/~leekc/ExtYaleDatabase/

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

[39] Phillips, P.J.; Moon, H.; Rizvi, S.A.; Rauss, P.J. The FERET Database. Available from: http://www.itl.nist.gov/iad/humanid/feret/feret_master.html

[40] Martinez, A.M. The AR face database. In: CVC Technical Report #24; 1998.

[41] Martinez, A.M. The AR Face Database. Available from: http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html

[42] OLIVEIRA; JR, L. L.; Thomaz, C. E. Captura e alinhamento de imagens: Um banco de faces brasileiro. Relatório de iniciação científica, Depto. Eng. Elétrica da FEI, São Bernardo do Campo, SP, 2006. 10: 1-10.

[43] OLIVEIRA; JR, L. L.; Thomaz, C. E. FEI Face Database. Available from: http://www.fei.edu.br/~cet/facedatabase.html

Face Recognition: Demystification of Multifarious Aspect in Evaluation Metrics

Mala Sundaram and Ambika Mani

Abstract

Face recognition has become an interesting research area in the recent era, and blends knowledge from various disciplines such as neuroscience, psychology, statistics, data mining, computer vision, pattern recognition, image processing, and machine learning. A new opportunity is obtained using the application of statistical methods for evaluating the performance of the system. Evaluation methods are the yardstick to examine the efficiency and performance of any face recognition system. Methods for performance evaluation seek to distinguish, compare, and interpret the various factors such as characteristics of subjects, location, illumination, and images. In this chapter, we show how to adapt popular performance measures commonly used in face recognition research, including—precision, recall, *F*-measure, fallout, accuracy, efficiency, sensitivity, specificity, error rate, receiver operating characteristics (ROC). This work serves as an introduction to performance measures, and as a practical guide for using them in research.

Keywords: face recognition, feature extraction, face detection, evaluation metrics, biometric

1. Introduction

The human face plays an interesting role in conveying people's identity in social interaction, biometric systems, law enforcement, security, and surveillance systems [1]. Variety of applications including biometric face recognition technology showed significant attention using the human face as a key to security [2]. As compared with other biometrics systems using fingerprint, iris, and palm print, face recognition has trenchant advantages because of its noncontact process. Face images can be captured from a distance without concerning the person, and the identification process does not require interacting with the person.

Face recognition is one of the major and rapidly thriving fields over the past two decades. This research area straddles researchers from multiple disciplines including data mining, image processing, pattern recognition, neuroscience, psychology, computer vision, and machine learning, etc. The face recognition system can identify one or more individuals from the still images or video by using a stored database of faces [3, 4]. This is a classification problem focusing on automatic face recognition. The main aspect of the face recognition systems is training the system with images from the known persons and classifying the newly coming test images into one of the classes.

Performance evaluation method is the yardstick to analyse the efficiency of any face recognition system. The assessment is essential for understanding the quality of the model or the technique, for refining parameters in the iterative process of learning and for selecting the most adequate model or strategy from a given set of models or techniques [5]. Several criteria are used to evaluate models for different tasks. This chapter goes through general ideas and the techniques used for evaluating the face recognition systems.

The chapter is structured as follows: Section 2 gives the intricate discussion on the face recognition techniques and methods, Section 3 throws light on the various aspects of the evaluation metrics, Section 4 discuss about the ways of assessing the system, Section 5 details the experimental analysis with case studies, and finally Section 6 concludes the chapter with future direction.

2. Face recognition techniques and methods

The human brain is highly adapted for face recognition, by remembering faces better than other patterns, and prefers to look at them over other patterns. Now a days computes also compensates in this research field. Facial recognition systems are applications of computers that examine the digital images of individuals for the purpose of identifying them [6]. The process of face recognition is influenced by many factors such as shape, size, pose, occlusion, and illumination. A human face is an extremely complex object with features that can vary over time. It is covered with nonuniformly textured material skin, which makes face object difficult to model. Skin of the face is influenced by perspiration level. The skin colour changes when the individual is embarrassed or becomes warm.

Facial recognition, have two different applications: basic and advanced. Basic facial recognition identifies faces or nonfaces such as cookies and animals. If it is a face, then the system looks for eyes, a nose, and a mouth. Advanced facial recognition deals with the question on a particular face. This includes unique features: the width of nose, wideness of the eyes, the depth and angle of the jaw, the height of cheekbones, and the distance between the eyes, and creates a unique numerical code. Using these numerical codes, the system then matches that image with another image and identifies how similar the images are to each other. The image sources for facial recognition include pre-existing photos from various databases and video camera signals.

Generally, a face recognition system consists of the following steps: Face detection, feature extraction, and face recognition as in **Figure 1**.
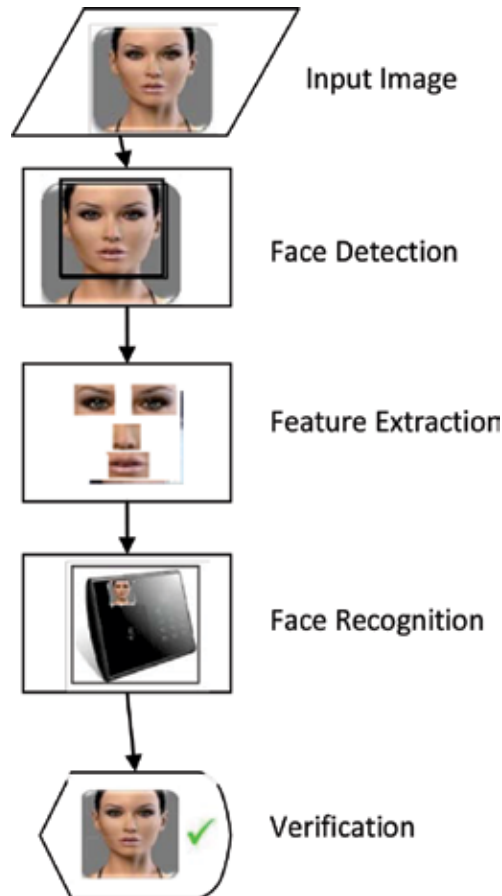

Figure 1. General structure of the face recognition system.

2.1. Face detection

The main function of this step is to determine the human faces and its location in a given image. The expected outputs are patches within each face or features of the face in the input image. It can also be regarded as object detection to find location and size of all objects in a given image. Face detection could be used for region-of-interest detection, object detection, video and image classification, etc., as in Ref [7–9] (**Figure 2**).

2.2. Feature extraction

In this phase, human-face patches are extracted from images to improve the accuracy of face recognition. To recognize human faces, extracting the prominent characteristics on the face
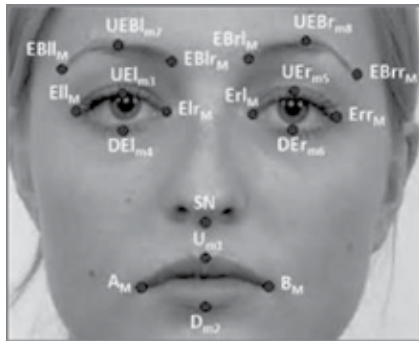
Figure 2. Feature detection.

features such as eyes, nose, and mouth together with their geometry distribution is applied. There are differences in face shape, size, and structure of these organs, so the faces are differing in thousands of ways so as to recognize them. One familiar technique is to extract the shape of the nose, eyes, chin, and mouth, and then distinguish the face by distance and size of those organs. The next method is to use a flexible model to illustrate the shape of the organs on face cleverly. A face patch is next transformed into a feature vector with rigid dimension (**Figure 3**).
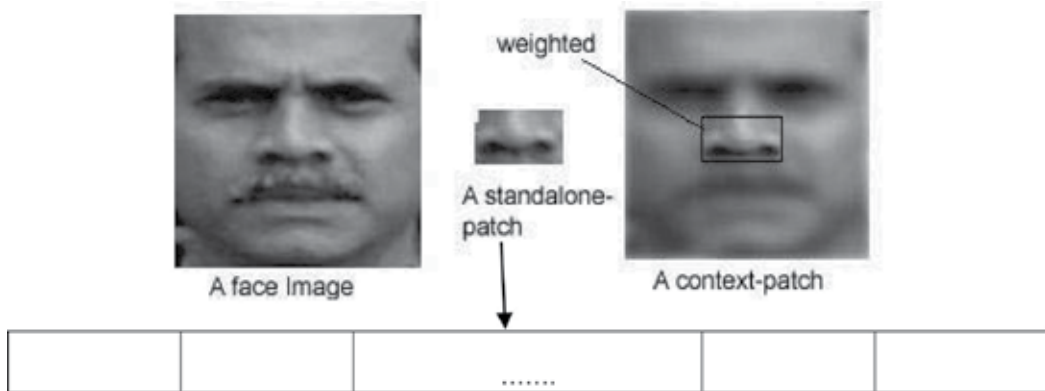
Figure 3. Feature extraction and feature vector representation.

2.3. Face recognition

Recognition of face from feature extraction and feature vector representation is the final step. A face data base is needed to achieve an automatic recognition. In the face database, for each person, several images are taken and their characters are stored. When an input face image comes in, the face detection and feature extraction are performed first. Then compare the characteristic features to each face of class stored in the database. The common approach of face recognition is identification and verification [10]. In face identification, the system probes

for the given face image to tell who he/she is, while in face verification, given a face image, the system validates true or false about the identification (**Figure 4**).
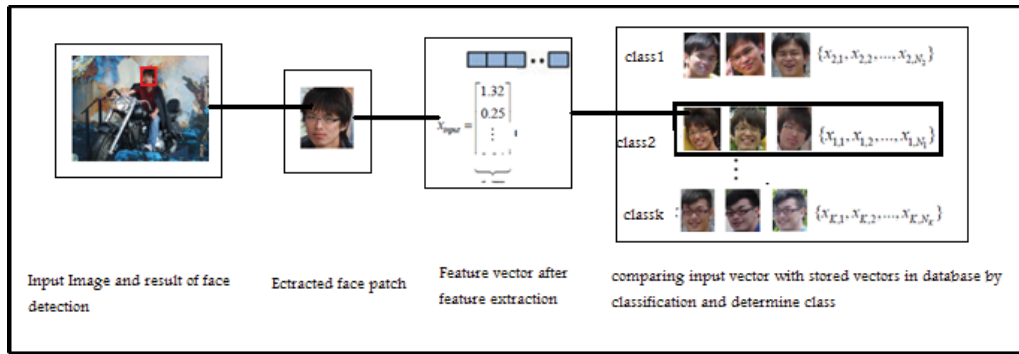

Figure 4. Steps in face recognition system.

3. Multifarious aspect in evaluation metrics

Now a days various measures are utilized for evaluating the performance of the face recognition system. This section elaborates some of them. The standard approach to deal with face recognition system evaluation revolves round the ground truth notion of positive and negative detection. Table 1, shows the confusion matrix. The terms positive and negative reveal the asymmetric condition on detection tasks where one class is the relevant pattern class and another class is the nonrelevant class.

Ground truth\detection	Detected (Positive)	Rejected (Negative)
Relevant	True positive (TP)	False negative (FN)
Nonrelevant	False positive (TP)	True negative (FN)

Table 1. Confusion matrix.

In the case of binary recognition or two class recognition, the system has to differentiate between face and nonface criteria. The true positive means the portion of face images to be detected by the system, while the false positive means the portion of nonface images to de detected as faces. The term true positive here has the same meaning as the detection rate and recall. False positives implies wrongly matching the individuals with photos in the database, and false negatives means not catching people even when their photo is in the database. There are two main evaluation plots: the receiver operating characteristics (ROC) curve and the precision and recall (PR) curve. The ROC curve examines the relation between the true positive rate and the false positive rate, while the PR curve extracts the relation between detection rate (recall) and the detection precision.

3.1. Precision

Precision is the fraction of the detected images that square measure relevant to the user's wants. It is additionally referred to as reliability or repeatability and is that the degree to that recurrent measurements beneath unchanged conditions show an equivalent results. Equation (1) represents them.

$$Precison = \frac{No\,of\,true\,positive}{No\,of\,all\,detected\,patterns} \tag{1}$$

In binary classification, precision is additionally known as positive predictive value. It is represented in Equation (2).

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

3.2. Recall

Recall is the proportion of positive cases that were properly identified. It is the fraction of relevant images that are successfully detected. It is additionally referred to as true positive rate. Recall is calculated using Equation (3).

$$Recall = \frac{No\,of\,true\,positive}{No\,of\,relevant\,patterns} \tag{3}$$

In binary classification, recall is commonly referred to as sensitivity. It is denoted in Equation (4).

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

3.3. Fall out

Fall out is the proportion of nonrelevant images that are detected as positive, out of all nonrelevant images (Equation 5).

$$Fallout = \frac{|\,\{non-relevant\} \cap \{detected\}\,|}{|\,\{non-relevant\}\,|} \tag{5}$$

In case of binary category, fallout is closely associated with specificity and is capable (1 – specificity). It is often checked out as the chance that nonrelevant images are detected as positive (Equation 6).

$$Fallout = \frac{TN}{TN + FP} \tag{6}$$

3.4. *F*-measure

F-measure is additionally referred to as *F*-Score or *F*1-measure. It combines the exactness and recall. It computes the average of the precision and recall. A conventional *F*-measure is the harmonic mean of precision and recall. This score is used to give a summary of the PR curve. It will be denoted as in Equation 7:

$$F\text{-}mesure = \frac{2 \times precision \times recall}{precision + recall} \tag{7}$$

In binary classification it is denoted as in Equation 8:

$$F\text{-}measure = \frac{2 \times TP}{(2 \times TP + FP + FN)} \tag{8}$$

The harmonic mean is an additional intuitive then the arithmetic mean, once computing the quantitative relation. Therefore, the complete definition of *F*-measure is given by Equation 9.

$$F - measure = \frac{(\beta^2 + 1)\,PR}{\beta^2 P + R}$$

$$\text{where } \beta^2 = \frac{1 - \alpha}{\alpha} \quad \text{and} \quad \alpha \; \varepsilon \; [0,1] \text{ and } \beta \; \varepsilon \; [0,\infty] \tag{9}$$

$$\alpha = \frac{1}{2} \text{ or } \beta = 1 \text{ is commonly written as } F_1 \text{ or } F_{\beta=1} = \frac{2PR}{P + R}$$

β is the parameter that controls a balance between P and R. When $\beta = 1$, F_1 involves be similar to the harmonic mean of P and R. This is often also referred to as *F*-measure or balanced *F*-score since precision and recall are equally weighted. When $\beta > 1$ emphasize recall. When $\beta < 1$ emphasize precision.

3.5. Accuarcy

Accuracy is the proportion of classifications, over all the N examples that were correctly detected. Accuracy is defined as "the fraction of quantity of correct classification over the entire number of samples." The amount of predictions in classification techniques relies upon the counts of the test records properly or incorrectly predicted by the model [11]. These counts are tabulated into a confusion matrix (also referred as contingency) Table 1. The confusion matrix shows how the classifier is behaving for individual categories.

$$Accuracy = \frac{No \ of \ correctly \ detected \ pattern}{Total \ number \ of \ validation \ set} \tag{10}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{11}$$

3.6. Error rate

The fraction is the range quantity of misclassification over the overall number of validation samples. However, the system response to wrong answers is the motive behind the introduction of error rate. It is an acceptable performance measure for the comparison of classification techniques given the balanced datasets. Precision, recall, and *F*-measure are acceptable performance measures for unbalanced datasets (Equations 13 and 14).

$$Error \ rate = \frac{No \ of \ misclassification}{No \ of \ samples \ in \ the \ validation \ set} \tag{12}$$

$$e = \frac{FP+FN}{TP+TN+FP+FN} \tag{13}$$

3.7. Effectiveness

The effectiveness measure is based on F_β-measure. F_β "Measures the effectiveness of detection with respect to a user who attaches β times as much importance to recall as precision (Equation 14)."

$$E \ (effectiveness) \ is \ E(P,R) = 1 - \frac{(\beta^2+1)PR}{\beta^2 P + R} \tag{14}$$

where β determines the relative importance of precision (P) and recall (R)

3.8. Sensitivity

True positive rate (TPR) is named sensitivity, hit rate, and recall. An applied mathematical measure of how well a binary classification test properly identifies a condition probability of properly labelling members of the target class (Equation 15).

$$Senesitivity = \frac{TP}{TP+FN} \tag{15}$$

3.9. Specificity

True negative rate (TNR) is named specificity. It is an applied mathematics measure of how well a binary classification test properly identifies the negative cases (Eq. 16).

$$TNR = \frac{TN}{TN+FP} \qquad (16)$$

False positive rate (FPR) also called as alarm rate is denoted as in Eqs. 17 and 18:

$$FPR = \frac{FP}{TN+FP} \qquad (17)$$

$$Specificity = \frac{TN}{TN+FP} = 1 - False\,alarm \qquad (18)$$

3.10. Receiver operating characteristics

Receiver operating characteristics (ROC) is a graph used for organizing and visualizing the performance of a system. It is a distinct option for precision–recall curves [12]. ROC graphs are normally utilized in medical decision-making, and in recent years are used more and more in machine learning and data processing research. It is a graphical representation for displaying the transition between TPR and FPR. TPR indicates correctly classified or total positive values and plotted on the *y*-axis, whereas FPR indicates incorrectly classified or total negative values plotted on the *x*-axis.

The points on the top left of ROC have high TP Rate and low FP Rate, thus represents smart classifiers. ROC graphs are far more helpful for domains with skew category distribution and unequal classification error costs. For this ability, ROC graphs are far more popular than accuracy and error rate. ROC plot can also visualize characterization change between the False match rate (FMR) and False nonmatch rate (FNMR).

Generally, the matching technique performs a decision based on a threshold that determines how close the image is to a template. If the threshold is reduced, there will be fewer false nonmatches, but more false accepts. Similarly, a higher threshold will reduce the FMR, but increase the FNMR. This more linear graph illuminates the differences for higher performances (rarer errors).

In **Figure 5**, the value A depicts Conservative performance which makes positive performance only with a strong evidence, so few false positive errors. The value B indicates the Liberal performance and value C indicates the perfect performance.

Figure 5. Regions of ROC graphs.

Some of the additional measures to evaluate the performance of Face identification systems are the following: Recognition Rate, Verification Rate, Half Total Error Rate in Ref. [13], Genuine Acceptance Rate (GAR), False Acceptance Rate (FAR), and False Rejection Rate (FRR)

The Recognition Rate is the simplest measure. It relies on a list of gallery images (usually one per identity) and a list of probe images of the same identities. The Recognition Rate is the total number of correctly identified probe images divided by the total number of probe images.

Another evaluation measure is the Verification Rate as in Ref. [14]. It relies on a list of image pairs, where pair with the same and pairs with different identities are compared. Given the lists of similarities of types, the ROC graph can be computed, and finally the Verification Rate. There are some more measures, such as the Half Total Error Rate and similar, which rely on independent development and evaluation sets. Validation test is a kind of test used to identify faces. The verification system uses some measures (i.e., Equal Error Rate), while some other are usually adopted for recognition systems (i.e., Recognition Rate).

3.11. False match rate

It is also denoted as FMR or False Accept Rate (FAR). FMR is the probability that the system incorrectly matches the input pattern to a nonmatching template in the database. It gauges the percent of invalid inputs that are incorrectly accepted. Similarly, if the person is an imposter in reality, but the matching score is higher than the threshold, then he is treated as genuine. This increases the FMR also depends upon the threshold value.

3.12. False nonmatch rate

It is also denoted as FNMR or false reject rate (FAR). It is the probability that the system fails to detect a match between the input pattern and a matching template from the database. It measures the percent of valid inputs that are incorrectly rejected.

3.13. Equal error rate

It is denoted as crossover error rate (EER or CER) or the rate at which both acceptance and rejection error are equal. The value of the EER can be obtained from the ROC curve. The EER is a quick way to compare the accuracy of devices with different ROC curves. Normally, the device with the lowest EER is the most accurate.

3.14. Failure to enroll rate

Also represented as FTE or FER is the rate at which endeavors to create a template from an input is unsuccessful. This case is usually caused by low-quality inputs.

3.15. Failure to capture rate

FTC is the probability that the system fails to detect an input even when the input is presented correctly.

4. Evaluation of face recognition system

Recognition of faces relies on how flexible the system is for pose variations. If the aim of the system is to recognize only frontal faces, then just use few classifier and function. The number of images of each face relies on the training on an image and testing on the rest. The recognition from different angles depends on the type of images and training set accordingly, with at least one image for each pose per person. The number of images in the training set and test on the remainder depends on the application of the system.

There are three methods to measure accuracy in a face recognition task. The one that was most suitable might depend to an extent on what the end purpose was.

1. How accurate is the algorithm at detecting a person from a data set containing *many images of a person and various images of different people*.

2. How accurate is the algorithm at gaining knowledge with a *set of faces from training and testing datasets of same peoples' images*.

3. How accurate is the algorithm at identifying *more than one person* from a dataset containing *images of these people mixed with the other people*.

For Case 1, train the algorithm with a set of images of an individual person's face and test on a set of images that contain different images of the goal person as well as equal number of other people. This task would be a binary classification task and accuracy can be efficiently measured with the help of precision and recall then. For more generalized results, this test could be repeated using various people.

For Case 2, train on multiple images of several people and then test on different images of the same people (If the dataset contains limited persons, then leave-one-out methodology might

be useful). This type of multiclass classification problem can be evaluated with the help of confusion matrixes which would be helpful in evaluating this sort of test.

For Case 3, train the algorithm on a categorized training set of images of several people and then test on a set of images containing different images of the same people mixed with other images of faces (To recognize people from a crowd, then large number of different peoples' images can be mixed in the test dataset). This could be created as a binary classification (person of interest/not), or as a multiclass problem (each person is a separate class with others). If the test set contains unbalanced images, then various measures of accuracy with true negatives can be used.

5. Experimental analysis

Face recognition has various result challenges as in reference [15]. In this section we have employed the theoretical model for computing the various performance measures to evaluate the efficiency of the face recognition system in different aspects.

5.1. Case 1

This case study used publicly available AT&T database in reference [16] for recognition experiments. In the database, 10 different images of each of 40 persons (total 400 images) with deviations in angles, expressions, and facial details are conceived. A preview image of the Database of Faces is shown in **Figure 6**.

The comparison is performed using Support Vector Machine technique and the computational efficiency is tabulated in the Table 2 and depicted in **Figure 7**.

Figure 8 shows the accuracy measure of the various datasets obtained using various technique.
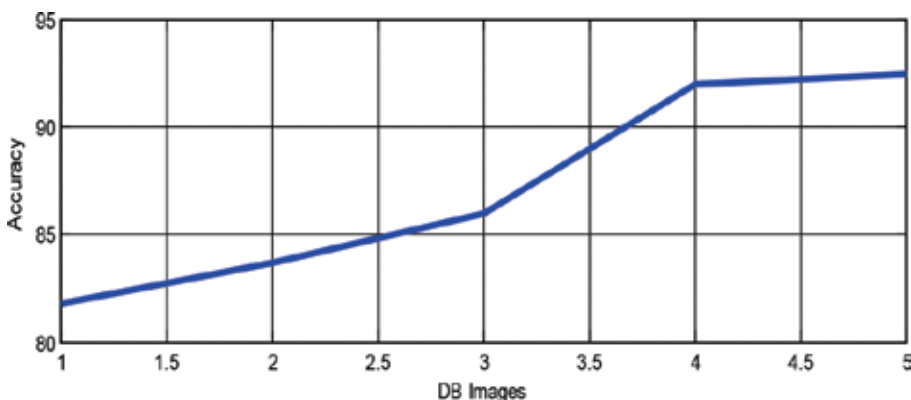

Figure 6. Sample collection of images in the dataset.

S.No	DB Images	Accuracy (Recognition Rate)
1	3	81.80%
2	4	83.70%
3	5	86.00%
4	6	92.00%
5	7	92.40%

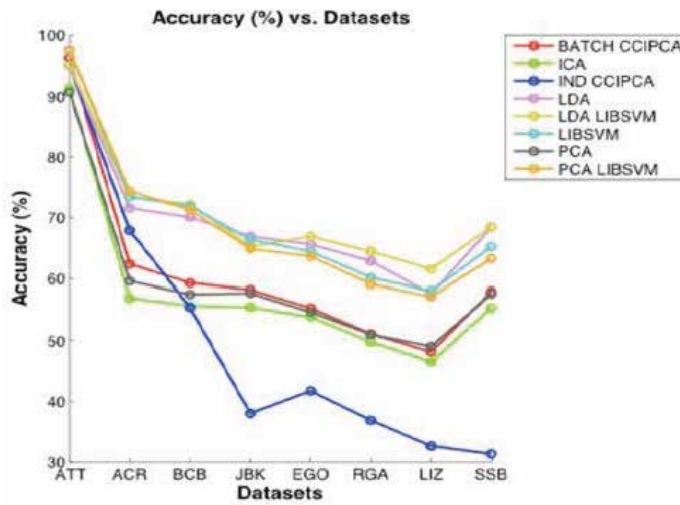**Table 2.** Accuracy of the recognition system.

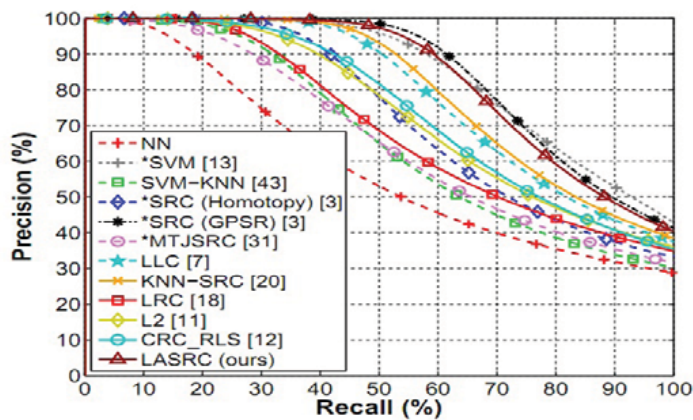

Figure 7. Accuracy of the recognition system.

Figure 8. Accuracy of the recognition system using various datasets.

5.2. Case 2

Brian C. Becker, gathered 800,000 face dataset from the Facebook social network as in reference [17] that models real-world situations where specific faces must be recognized and unknown identities must be rejected. Finally, the results are depicted using precision–recall curve as in **Figure 9**. The graph shows that as the precision increases recall decreases.
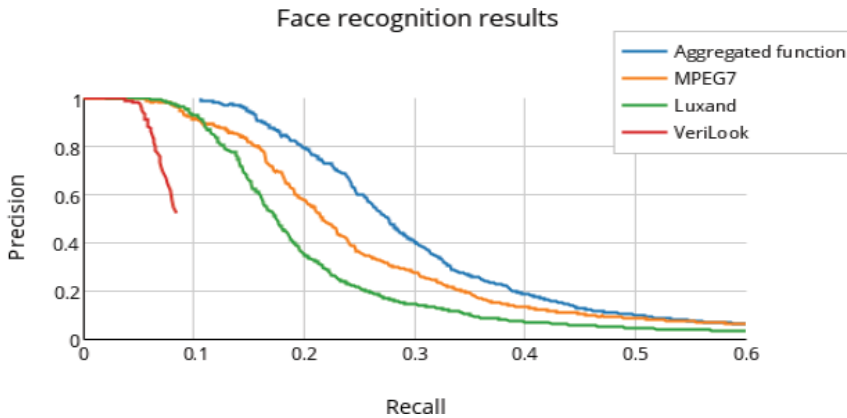

Figure 9. Precision and recall curve on our 800,000 Facebook dataset.

Nonreal time algorithms are marked with an asterisk (*). LASRC approach performs very similarly to nonreal time algorithms such SRC or SVMs but has the advantage of being real time. In fact, LASRC trains 100× faster than SVMs and classify 250× faster than SRC. Compared to other real-time methods, LASRC outperforms state-of-the- art least squares, sparse, and max-margin classifiers.

Face recognition is a technology for automatic detection and recognition of human faces on static images as stated in reference [18]. The main advantage of this technology is its ability to aggregate multiple face recognition and detection functions. Here we listed some of the commercial software for face recognition such as FaceSDK, VeriLook SDK, MPEG-7 descriptors + OpenCV. The following Table 3 and **Figure 10** show the values of precision and recall obtained using the listed software.

Name	Recall	Precision
OpenCV	55%	89%
FaceSDK	63%	83%
VeriLook SDK	73%	84%
Aggregation approach	62%	98%

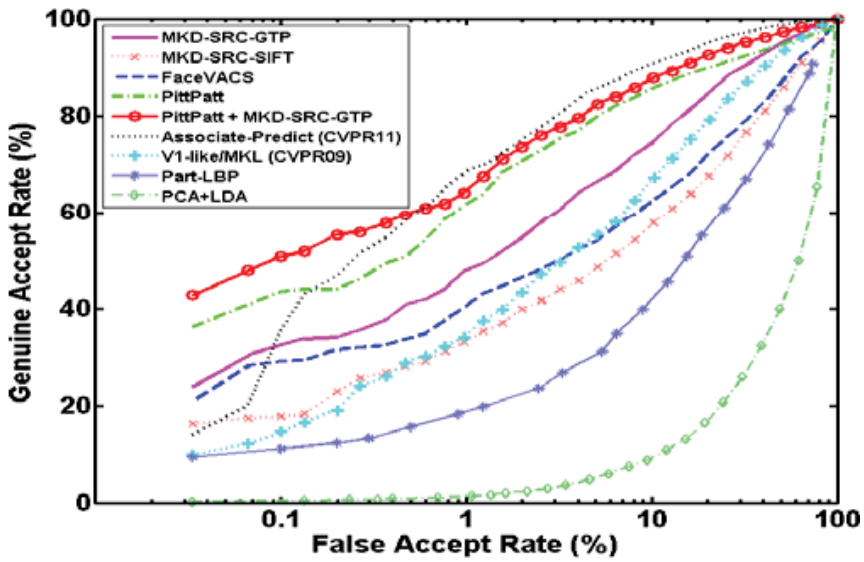**Table 3.** Compariosn results of precision and recall.

Figure 10. Comparison results of precision and recall.

5.3. Case 3

This case study used the LFW benchmark dataset, where the dataset is divided into 10 subsets for cross validation, with each subset containing 300 pairs of genuine matches and 300 pairs of impostor matches for verification. The mean values of FAR and Genuine Accept Rate (GAR) with fixed thresholds over all the 10 subsets are plotted in an ROC curve for **performance evaluation** as in reference [19] and **Figure 11**.
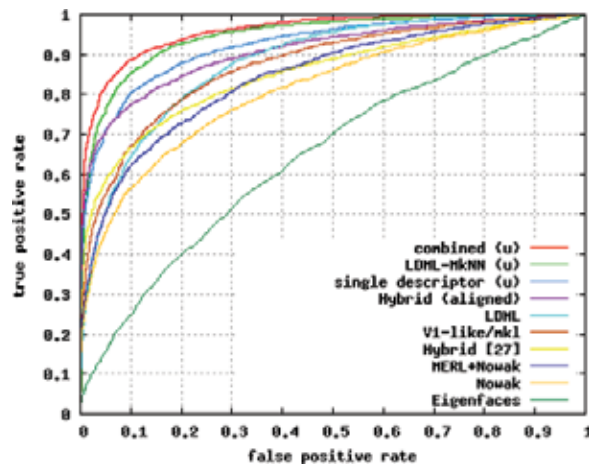

Figure 11. The ROC curves of the various face recognition algorithms.

The following ROC curves (**Figure 12**) are the average over ten-folds (FPR and TPR) of the LFW data set. The (u), indicates ROC curve is for the unrestricted setting.
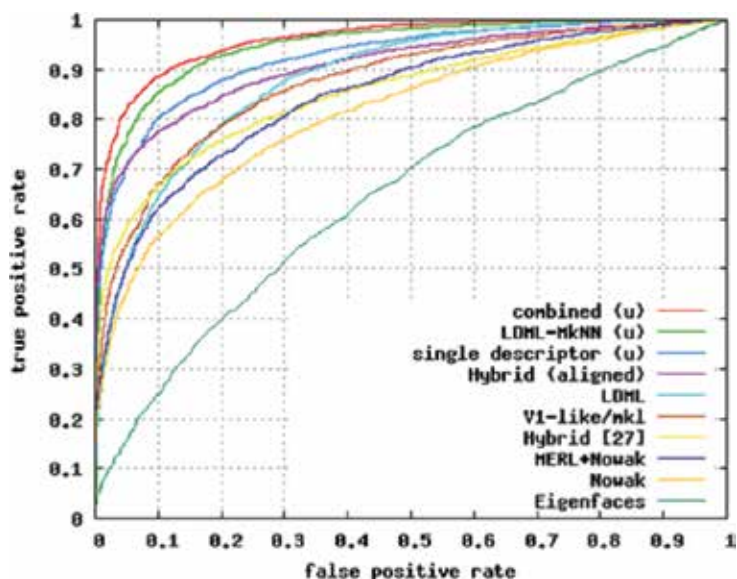

Figure 12. The ROC curves using TPR and FPR.

6. Conclusion and future work

This chapter presents a viewpoint about face recognition and the various ways to evaluate the face recognition system. The faces are highly complex patterns that often differ in only subtle ways, like changes in angle and lighting. Hence, the face recognition system should consider various factors such as facial expression change, aging, pose change, illumination change, scaling factor, frontal vs. profile presence and absence of spectacles, occlusion due to scarf, mask in front, beard, and moustache. Generally, when the training set contains faces of one person, then precision and recall could be used to evaluate accuracy. When the training set contains multiple faces of several people and test set contains the different faces of same people, then confusion matrixes would be helpful in evaluating the test face. When the training contains faces of interest with other faces, and the test set is an unbalanced one, then various measures of accuracy dominated by true negatives can be used to evaluate the face recognition. A complete face recognition system contains several subproblems where each one is an independent research problem. The line of future work includes the assessment of various machine learning algorithms used in face recognition with feature mining. However, next era face recognition are going to have tremendous application in smart environs, real time, and in much less-controlled situations.

Author details

Mala Sundaram[*] and Ambika Mani

*Address all correspondence to: ursmala@gmail.com

Department of Computer Science and Engineering, Anna University (BIT Campus), Tamil Nadu, India

References

[1] Kumar A., Kumar A. Adaptive security for human surveillance using multimodal open Set biometric recognition. In: Proceedings of the 22nd International Conference on Pattern Recognition (ICPR); 2014. pp. 405–410. DOI: 10.1109/ICPR.2014.78

[2] Hitoshi I. The world's best face recognition system to achieve safety and security in our society. Special Issue on Solutions for Society—Creating a Safer and More Secure Society NEC Technical Journal 2015. Vol.9, No.1, 48–51.

[3] Liao S., Jain A., Li S. Partial face recognition: alignment-free approach. IEEE Transactions on Pattern Analysis and Machine Intelligence 2013.Vol. 35, Issue 5, 1193–1205. DOI: 10.1109/TPAMI.2012.191

[4] Mohanty P., Sarkar S., Kasturi R., Phillips P. Subspace approximation of face recognition algorithms: an empirical study. IEEE Transactions on Information Forensics and Security 2008. Vol. 3, Issue 4, 734–748. DOI: 10.1109/TIFS.2008.2007242

[5] W-L. Chao. Integrated machine learning algorithms for human age estimation [thesis]. National Taiwan University; 2011.

[6] Zhao, R. Chellappa, A. Rosenfeld, P.J. Phillips, Face recognition: a literature survey. ACM Computing Surveys (CSUR) 2003. Vol. 35, Issue 4; 399–458. DOI: 10.1145/954339.954342

[7] Low E.H. Face detection: a survey. Computer Vision and Image Understanding 2001. Vol. 83, 236–274. DOI: 10.1006/cviu.2001.0921

[8] Ming-Hsuan Y., D.J. Kriegman, N. Ahuja. Detecting faces in images: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 2002. Vol. 24, Issue 1, 34–58. DOI: 10.1109/34.982883

[9] S. Zafeiriou, C. Zhang, Zhengyou Z..A survey on face detection in the wild: past, present and future. Computer Vision and Image Understanding 2015. Vol.138, Issue C, 1–24. DOI: 10.1016/j.cviu.2015.03.015

[10] Kumar N., Berg A., Belhumeur P., Nayar S. Describable Visual Attributes for Face Verification and Image Search, IEEE Transactions on Pattern Analysis and Machine Intelligence 2011. Vol. 33, Issue 10, 1962–1977. DOI: 10.1109/TPAMI.2011.48

[11] Aggarwal G., Biswas S., Flynn P.J., Bowyer K.W. Predicting performance of face recognition systems: An image characterization approach. In: Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW); Workshops, IEEE Press; 2011. pp. 52–59. DOI:10.1109/CVPRW.2011.5981784

[12] Fawcett, T. An introduction to ROC analysis. Pattern Recognition Letters 2006. Vol. 861–874. DOI: 10.1016/j.patrec.2005.10.010

[13] White D., Dunn J.D., Schmid A.C., Kemp R.I.. Error rates in users of automatic face recognition software. PLoS ONE 2015. Vol. 10, Issue 10, e0139827. doi:10.1371/journal.pone.0139827

[14] P. Ahlgren, L. Gr¨onqvist. Retrieval evaluation with incomplete relevance data: a comparative study of three measures. In: Proceedings of the 15th ACM international conference on Information and knowledge management (CIKM '06); New York, 2006. pp. 872–873 DOI: 10.1145/1183614.1183773

[15] Phillips P.J. Preliminary face recognition grand challenge results. In: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR'06), Washington. USA IEEE 2006. pp. 15–24. 10.1109/FGR.2006.87

[16] Database of Faces by AT & T [Internet]. 2016. Available from: http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html [Accessed: 2016-11-12]

[17] Brian C. Becker [Internet]. 2016. Available from: http://www.briancbecker.com/blog/research/web-scale-face-recognition/[Accessed: 2016-01-27]

[18] Laboratory of Data Intensive Systems and Applications (DISA)[Internet]. Available from http://disa.fi.muni.cz/[Accessed: 2016-01-11]

[19] Ho, W.H., Watters, P. A new performance evaluation method for face identification—regression analysis of misidentification risk. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2007); 2007; pp. 1–6. DOI: 10.1109/CVPR.2007.383276

Edited by S. Ramakrishnan

Pattern recognition has gained significant attention due to the rapid explosion of internet- and mobile-based applications. Among the various pattern recognition applications, face recognition is always being the center of attraction. With so much of unlabeled face images being captured and made available on internet (particularly on social media), conventional supervised means of classifying face images become challenging. This clearly warrants for semi-supervised classification and subspace projection. Another important concern in face recognition system is the proper and stringent evaluation of its capability. This book is edited keeping all these factors in mind. This book is composed of five chapters covering introduction, overview, semi-supervised classification, subspace projection, and evaluation techniques.

IntechOpen