Department of Computing

Multilinear Analysis of Face Image Ensembles

Santu Rana

This thesis is presented for the Degree of
Doctor of Philosophy
of
Curtin University of Technology

June 2010

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgement has been made. This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

Santu Rana

Contents

Abstract

Machine based face recognition is an important area of research that has attracted significant attention over the past few decades. Recently, multilinear models of face images have gained prominence as an alternative method for face recognition. Against linear techniques, multilinear models offer the advantage of having more complex models. Against kernel and manifold based non-linear techniques, the advantage lies in having more intuitive and computationally frugal modelling. In this thesis, we present an in-depth analysis and understanding of different properties associated with multilinear analysis and propose three different face recognition algorithms and a unified framework addressing open issues in face recognition.

We first propose a face recognition algorithm primarily to address the limitations of the existing multilinear based algorithms in the form of their inability to handle test images that are in unseen conditions. The algorithm is based on the construction of a new representational basis *multilinear eigenmodes*, enabling representation and classification of faces at unseen conditions. Subsequently, we propose a second algorithm to address the high computational complexity of the first algorithm. We define a set of *person-specific bases* to represent person-specific images under all variations, and based on this propose an efficient recognition algorithm. Next, we propose a framework of face recognition based on an interpretation of the multilinear analysis as a factor analysis paradigm. We the reformulate all the multilinear based algorithms to link them to a single optimization framework. A theoretical comparison of these algorithms is performed revealing the fundamental differences between them and their applicability in different face recognition scenarios. Experiments performed to compare multilinear analysis based methods to the leading linear and non-linear techniques reveal the superiority of our second algorithm on both measures of recognition accuracy and test speed.

Next, we address the issue of inadequate training samples that arise in many application scenarios. We introduce a novel "friendly-hostile" paradigm, in which we propose a mechanism to compensate for the low number of training samples of hostile people by learning the structure of face images from a large training set of the friendly people. The formulation is built on a novel synthesis paradigm that is based on the unique factorization properties of the multilinear analysis. Experimental results show significant performance gain in comparison to the conventional methods. We also discuss issues concerning unbalanced datasets, wherein some people may be under-represented than others in the training set. This results in different apriori bias per class, affecting conventional recognition algorithms. Based on theory and experiments we demonstrate that our algorithm does not get affected by any such imbalance in bias and produces consistent performance in all situations.

Acknowledgements

It is a pleasure to thank the many people who has made this thesis possible.

It is difficult to overstate my gratitude to my PhD supervisor, Prof. Svetha Venkatesh. She was always there when I needed any assistance, be it for discussions in technical matters or for matters that are not so technical. Throughout the PhD period, she provided constant encouragement, sound advice, good teaching, good company, and an unflinching show of confidence. I would have been lost without her.

I am indebted to my co-supervisor, Dr. Mihai Lazarescu, for all the insightful discussions that shaped the course of my PhD from time to time. His constant co-operation in every matter made the research a pleasurable experience.

I am grateful to Dr. Wanquan Liu for all the technical assistance. His valuable inputs enriched the thesis in great many ways.

I am thankful to all the members of our lab for making my stint a wonderful experience. Especially, to Patrick and Senjian for the technical discussions. To Jeff, for correcting English. To Mary, for the much needed support during my troubled trip to Florida and for assistance in several other ways.

I wish to thank my housemates, Pradeep, Paresh, Rahul, Prashant and Vishal for the camaraderie, entertainment, and care they provided.

I am grateful to all my friends for their caring and emotional support whenever I needed them. Special thanks to Prasanta, Gargidi, Sourangshu, Kunal, Anindya, Saikatda, Ali, Arif, Pal, Maity, Khan, Das, Samir, Sandy, Noor and Sreetama, for always being there within my reach.

I wish to thank my entire extended family for providing a loving environment for me. Among them, Mama, Mohankaku, Chotomama, Chotokaku, Mana and Didi deserve special mention.

Lastly, and most importantly, I wish to thank my parents, Biva R. Rana and Nikunja B. Rana. They raised me, supported me, taught me, and loved me. To them I dedicate this thesis.

List of Figures

List of Tables

Relevant Publications

Part of this thesis and some related work have been published or in communication. The list of these publications is provided below.

- **Chapter 4:** Rana, S., Liu, W., Lazarescu, M., and Venkatesh, S. (2008b). Recognising faces in unseen modes: A tensor based approach. Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, pages 1−8, ©IEEE. Reproduced with permission.

- **Chapter 5:** Rana, S., Liu, W., Lazarescu, M., and Venkatesh, S. (2008a). Efficient tensor based face recognition. Pattern Recognition, 2008. ICPR 2008. 19th International Conference on, pages 1−4, ©IEEE. Reproduced with permission.

- **Chapter 6:** Rana, S., Liu, W., Lazarescu, M., and Venkatesh, S. (2009b). A unified tensor framework for face recognition. Pattern Recognition, 42(11), 2850−2862, ©Elsevier. Reproduced in accordance with: `http://www.elsevier.com/wps/find/authorsview.authors/copyright#whatrights`.

- **Chapter 7:** Rana, S., Liu, W., Lazarescu, M., and Venkatesh, S. (2009a). Face recognition from few samples: A tensor based approach. (*In communication*).

Please note that the paper *"Efficient tensor based face recognition"* was nominated for the **IBM Best Students Paper Award** in **ICPR 2008**.

Notations

Symbol	Meaning
$\mathcal{A}, \mathcal{B}$	**Tensors**: tensors will always be denoted using calligraphic fonts.
$A_{(k)}$	**Tensor unfolding**: Unfolding of $\mathcal{A}$ in the k'th mode.
$\times_k$	**Mode-k product**: Matrix multiplication to a tensor at the k'th mode.
$\otimes$	**Kronecker product**
$\mathcal{T}$	**Image Tensor**
$\mathcal{S}$	**Core Tensor**

Chapter 1

Introduction

Machine based face recognition is an important research topic that has attracted significant attention over the last few decades. From authentication systems to law enforcement scenarios, the ability to recover the identity from a face image in a non-intrusive way can significantly change the way many things are done. For example, an ATM system with face recognition technology would enable the user to authenticate by looking at a camera instead of entering a PIN number, significantly reducing the chance of identity thefts. Similarly, a city with smart surveillance cameras equipped to identify criminals in real time will drastically simplify the job of law enforcement agencies. Furthermore, there are multitudes of possibilities that can be realised in the space of human-computer interaction through successful application of face recognition technologies.

Humans possess an unmatchable ability to identify faces in a great variety of conditions. The goal of the face recognition research is to equip machines with the same ability. The basic premise of face recognition is similar to pattern recognition problems, wherein a classifier is trained from a corpus of labelled training face images, and given a test face image, the classifier seeks the most appropriate label. The basic assumption is that adequate labelled training images are available and for camera based automated face recognition, machines can detect and crop the faces from the camera images or videos. Face detection in images is also a demanding task and a research problem in its own right, however, we do not address this issue in this thesis. Whilst the face recognition problem is similar to the basic pattern recognition problem, the uniqueness of face structures serve as the main research motivation to develop specialized face image modelling and to pursue domain induced feature selection to address the problem in a principled way. Mostly, face recognition involves using 2D grayscale or color face images in the visible light spectrum, however, with advancement of technologies many researchers have proposed using beyond visual range 2D images and 3D scanning data as possible improvements to the limitation of useful features available in the 2D visual range images. However, most of these technologies are still in development and require costly re-installation. Therefore, face recognition techniques with normal 2D images retains its importance and relevance in real world applications.

The first successful demonstration of face recognition algorithm based on using 2D images is considered to be the Eigenface algorithm of (Turk and Pentland, 1991b), (Turk and Pentland, 1991a). The algorithm proceeds by vectorizing the 2D images and then finding a representative subspace employing Principal Component Analysis. Several other linear models such as Independent Component Analysis, Fisher Discriminant Analysis etc. were adopted for face recognition with limited and inconsistent success. Later, non-linear models are introduced primarily based on kernel mechanisms to remove limitations imposed by the linearity in the models. However, definite success and consistent performance of those algorithms in widely different situations remained an issue. Face recognition based on manifold modelling has been studied in ISOMAP (Tenenbaum *et al.*, 2000), LLE (Roweis and Saul, 2000), (Saul and Roweis, 2003) and Laplacian Eigenmap (Belkin and Niyogi, 2003), (Belkin and Niyogi, 2001). The most prominent amongst them is LPP (Locality Preserving Projection) by (He and Niyogi, 2003). Whilst LPP performs consistently better than linear or kernel based methods and is considered as the state of art, the limitation lies in the fact that it is domain agnostic *i.e.* it does not offer any scope for domain specific modelling of faces. Moreover, the performance of LPP in different application scenarios leaves scope for further improvement. Lately, multilinear modelling has been proposed to provide a manifold modelling technique that explicitly takes into account the way face images undergo variations due to different influencing factors such as lighting, pose, expression etc. The basic premise of multilinear modelling is to model face images as multilinear functions of the various influencing factors over the pixel space. Multilinear modelling of face image ensembles was first proposed in (Vasilescu and Terzopoulos, 2002b) based on decomposing face image tensor using Higher Order SVD (De Lathauwer *et al.*, 2000). Face recognition based on such modelling is proposed in (Vasilescu and Terzopoulos, 2002a). Preliminary experiments demonstrates performance gain over the basic PCA based Eigenface (Turk and Pentland, 1991b) method.

1.1 Aims

This thesis aims to develop face recognition algorithms by extending our understanding of multilinear modelling of face image ensembles and investigate useful properties that are uniquely associated with such modelling. In particular, we seek to answer the following questions:

- Can we address the limitation of the existing multilinear modelling based face recognition algorithm of (Vasilescu and Terzopoulos, 2002a), which fails to handle unseen variations (such as unseen lighting or viewpoint) by proper understanding and utilization of the multilinear models and the analysis paradigm?
- Can we develop an efficient algorithm that can handle test faces in unseen conditions?
- Can we formulate a general framework for face recognition based on the multilinear modelling that offers a basis for comparison between different algorithms? To what extent do these algorithms compare with other face recognition approaches?
- Can we discover any unique attribute associated with multilinear modelling that can be effectively utilized to solve the real-world problem of face recognition from a few samples?

1.2 Approach and Contributions

Our approach is based on developing a deeper understanding of the different aspects of multilinear modelling of face image ensembles. The theoretical contribution of this thesis is the development and in-depth study of face recognition approaches based on multilinear models and analysis. The practical contribution is the development of an algorithm that is better than several leading face recognition algorithms, along with an algorithm that can address the “few samples” issue, a commonly faced problem in real-world scenarios. We detail our approaches and contributions below:

- The first question is addressed by understanding the multilinear decompositions and identifying the importance of the core tensor. Our argument is based on the fact that the core tensor contains the variation information for all factors and thus, if we need to address the issue of the recognition of faces with unseen variations then the core tensor must be exploited. Based on this understanding, we develop the notions of different multilinear identities and finally propose an algorithm based on a joint *multilinear eigenmode* space (MPCA-JS[1]) that is shown to offer consistently superior recognition performance to the existing method (Vasilescu and Terzopoulos, 2002a) on unseen test cases.

  The contribution of this part of the work is mostly theoretical entailing an understanding of the intricacies of multilinear models and analysis, determining the reasons as to why existing multilinear based algorithms fail to handle certain scenarios and then proposing a basic algorithm that convincingly addresses these limitations.

- Whilst we are able to achieve our first objective, we realized that the proposed algorithm lags when it comes to speed of testing. To retain the same recognition capability with a lower computational demand, we exploit the insights that whilst the first algorithm fully utilizes the benefits of multilinear decomposition, there is considerable redundancy in the solution. In the first solution we treat the person space as any other generating space, however, when we are interested in matching identities, we can restrict ourself only to the locations of the person identities in the person space. We remove the redundancy and propose a second algorithm (MPCA-PS[2]) that is able to maintain the level of recognition performance of the first algorithm in an efficient manner.

  This part of the work has both theoretical and practical significance. On the theoretical side, we exploit some intrinsic features of multilinear analysis and are able to present a low complexity algorithm, whilst retaining the recognition performance of the previously proposed algorithm. On the practical side, the high speed of this algorithm coupled with good recognition rates makes it an attractive choice for real-world applications.

- Whilst we solve the efficiency issue, we observe that the two algorithms yield different relative performance over different datasets. This brings us to the

[1]MPCA-JS: Multilinear Principal Component Analysis - Joint Space
[2]MPCA-PS: Multilinear Principal Component Analysis - Person Space

third question as to whether these discrepancies can be answered at a more theoretical level. This is addressed by developing an optimization framework that is fundamental to the way face recognition is performed from the multilinear models. All the algorithms are rederived in this framework, revealing their underlying strategies, as they solve the same optimization problem. A theoretical comparison among all these methods is then outlined in a concise manner. We also perform extensive experimentation with these algorithms vis-a-vis leading face recognition approaches over publicly available benchmark datasets.

This is the most significant part of the work in terms of both theoretical and practical contributions. The study of multilinear analysis based on a single unified framework completes the understanding of the whole multilinear modelling and analysis paradigm. Existing algorithms are analysed and compared based on their intrinsic qualities, providing useful insights about their applicability in different scenarios. Also, from the experimental results concerning evaluation of our multilinear based algorithms in comparison to the leading face recognition approaches, such as LDA and LPP, we show that our second algorithm consistently provides superior performance.

- The fourth question arises naturally from the consideration of real-world deployment of face recognition applications. Whilst most face recognition approaches implicitly assume the availability of a large number of training samples, in real-world situations it is hard to meet such requirements as the "people of interest" are mostly hostile in nature and thus, we may have access to only a few of their images. To solve this, we propose a novel friendly-hostile paradigm, where large numbers of training samples are collected from friendly people to compensate for the small training set from hostile people. We construct a novel synthesis paradigm based on the multilinear analysis framework and provide a way to model the face manifolds of each hostile people based on the average structures learnt from the friendly people. As the manifold describes the set of all possible images of a person, we are able to fully compensate for their lack of adequate training images. The recognition is carried out by seeking the manifold that best represents a given test image. We also apply this to the situation of unbalanced datasets, which may arise even in cooperative situations wherein the cost of image acquisition can be reduced by choosing to collect only a few training images for many of them. Extensive evaluation of the proposed approach is performed for both the “few samples” and ”unbalanced dataset” scenarios and the performance is compared with leading face recognition approaches.

This is a significant application of multilinear models, where the unique properties have been exploited to provide an elegant solution to a practical problem in face recognition. The theoretical work is mostly based on the framework developed before, with suitable extensions. The practical significance of this work is immense, especially in surveillance scenarios where the availability of training images for criminals is a significant issue.

1.3 Outline of the Thesis

The rest of the thesis is organized as follows: In Chapter 2 we provide related background and literature, starting with different face recognition techniques based on Local Feature Analysis, 3D modelling and beyond visual range images, linear and non-linear models using holistic face features, before focusing on multilinear analysis and its application for face image ensembles and face recognition. We also include a brief review on the work of face recognition when adequate training images are not available, a specific application scenario that is addressed in this thesis.

In Chapter 3 we provide the related mathematical background, starting with definitions related to tensors and their properties, tensor decomposition based on Higher Order Singular Value Decomposition (HOSVD), Multilinear Principal Component Analysis (Multilinear PCA), and finally, the framework of face image analysis with Multilinear PCA framework with some illustrative examples.

In Chapter 4 we propose our first algorithm (MPCA-JS) based on the multilinear analysis of face image ensembles. The primary motivation of this work is the limitation of the existing multilinear based algorithm (Vasilescu and Terzopoulos, 2002a), which fails to recognise faces when they are at very different lighting or viewpoints from the training faces. This limitation is addressed by constructing a joint *multilinear eigenmode* space as a representational basis. Experiments show its superior performance to the existing method.

In Chapter 5 we propose an improved algorithm (MPCA-PS) by removing the redundancy in the previous MPCA-JS algorithm. The algorithm is developed by constructing *person-specific eigenmode* subspaces for person specific face image representation. By doing this we are able to reduce the complexity order significantly. In experiments we find it to be significantly faster, whilst the performance is found to be almost similar, and in some situations significantly better than MPCA-JS.

In Chapter 6 we present a unified framework that unites all the above multilinear analysis based recognition algorithms into a common mathematical basis. We analyse the multilinear analysis from the factor analysis perspective and devise an optimization problem relating a face image to its associated factors. Next, we show that all the above algorithms solve this same optimization problem adopting different assumptions and employing different strategies. The algorithms are then compared theoretically to understand their applicability in different scenarios. We also compare all the multilinear based algorithms to some of the leading face recognition approaches, such as LDA and LPP on a number of benchmark datasets, and demonstrate that our algorithm MPCA-PS provides the best recognition performance in terms of both accuracy and speed.

In Chapter 7 we use the framework developed in the previous chapter to address the issue of face recognition when adequate training images are not available. We propose a solution based on a novel "friendly-hostile" paradigm, where we learn a generic structure of face images from the full dataset of the friendly people, and then formulate a way to transfer the structure to each hostile people so as to provide a formal description of their face manifolds, thereby, eliminating the need for large training samples. We discuss the "few samples" problem from two perspectives: one, where the hostile people are the actual persons of interest and are represented by a few samples, and second, where both friendly and hostile people are part of the recognition classes, but the hostile people have very low representation in the face dataset compared to the friendly people, leading to an unbalanced training dataset. Our algorithm show good performance for both scenarios, comprehensively surpassing the performance of other leading recognition algorithms.

Finally, Chapter 8 provides a summary of work in the thesis, its contributions, and discusses directions for future work.

Chapter 2

Related Background

In this chapter we review literature relevant to the work presented in this thesis. For humans, faces serve as an important modality for interaction, both as an identity and for conveying emotions. Perception of faces is so important that the human brain is understood to have a dedicated visual processing unit that exclusively deals with face images acquired visually (Logothetis and Gauthier, 2000). Therefore, with the arrival of computers and automation where increasing interaction between man and machines is seen as the way forward, there is a pressing need to impart machines with the ability of human like face perception. Initial research on machine based face recognition can be dated back to the 1960's (Bledsoe, 1964). Other important early work includes the work of (Kelly, 1971) and a seminal work by (Kanade, 1977), presenting basic insights in this area. Subsequently, correlating with the advancement of computation and wider availability of powerful desktop computers, face recognition research took off in the 1990's and ever since has established itself as a major research track across Artificial Intelligence, Computer Vision and Pattern Recognition. An excellent overview on the decade's effort of machine perception of faces is available in (Zhao *et al.*, 2003) along with reviews in (Gross *et al.*, 2001), (Bowyer *et al.*, 2006) and (Scheenstra *et al.*, 2005) covering different aspects of the problem and discussing various research challenges that exist in this area.

Our review is presented with the following structure: Section 2.1 provides a review on the recognition approaches based on local features and outlines potential problems of identification of local features. Section 2.2 provides a brief review on the emerging area of face recognition research that includes 3D imaging and beyond visual range imaging. Section 2.3 provides a detailed review of the main focus of our work: recognition based on holistic face features, with sub-sections discussing both linear and non-linear methods. Discussion on non-linear methods include both the Kernel based and the Manifold based approaches. Section 2.4 provides the review on the multilinear analysis framework and its application to face recognition problems. In Section 2.5, we review attempts to solve the recognition problem when adequate training images are unavailable, and finally the chapter ends with some closing remarks in Section 2.6. Additionally, the related mathematical background in presented in Chapter 3.

2.1 Face Recognition based on Local Features

Early face recognition approaches (Kelly, 1971) and (Kanade, 1977) were mostly based on configural matching of local face features *i.e.* eyes, nose, mouth etc. However, the seminal work of (Turk and Pentland, 1991b), (Turk and Pentland, 1991a) was based on considering the face image as a whole. Subsequently, the research in face recognition went into two different directions: one, in which face images are processed as holistic features similar to (Turk and Pentland, 1991b)(Turk and Pentland, 1991a), and the other, where recognition is carried out through local features analysis (LFA). Initial exploration in the latter direction includes the work of (Cox *et al.*, 1996), based on a mixture-distance approach and of (Samaria and Young, 1994), where Hidden Markov Models (HMM) were used over a strip of pixels covering salient features (eyes, forehead, nose, and mouth) and the transition statistics is learnt as the important feature. The HMM approach was further extended in (Nefian and Hayes, 1998) and (Nefian and Hayes, 1998) based on using 2D-DCT features in place of pixels. However, the most successful framework of LFA is considered to be the Elastic Bunch Graph based algorithm, as described in (Wiskott *et al.*, 1997). In their approach, the face image is represented by a bunch graph (a stack of Gabor features at each node), with nodes positioned in the salient locations on the face. The matching of face bunch graphs is performed following the Dynamic Link Architecture framework of (Buhmann *et al.*, 1990), (Lades *et al.*, 1993). This approach was found to be better at handling occlusions and different poses (Krger *et al.*, 1997). However, the fundamental problem of recognition approaches based on LFA is the identification/localization of the salient locations, which can be as difficult as the recognition problem itself. Moreover, recent results from neurology suggest that holistic face feature gets precedence over local features in face identification (Sinha *et al.*, 2006), which suggests that holistic features may be important for machines too. It is argued that fusing the holistic and LFA processing techniques may be beneficial, as in (Penev and Atick, 1996), however, one still has to resolve the feature localization problem before using this approach. Therefore, most of the research efforts in automatic face recognition, including the work presented in this thesis, considers only holistic features.

2.2 Non-traditional Methods: 3D Face, Infrared, Multi-spectral Imaging

Recently, there has been a growing interest in face recognition based on images taken at beyond visual range spectrum and recognition based on 3D face modelling. Face images taken at the visual range of electromagnetic spectrum are widely available and easily understood. However, images taken in dimly lit places may contain extreme shadows that affect the proper sensing of features, which in turn can affect the recognition performance. To address this issue, some researchers have proposed the use of images taken in infrared or near-infrared bands, which are not affected by external lighting as the human body is itself a source of infrared radiation (Socolinsky *et al.*, 2003), (Zou *et al.*, 2005), (Li *et al.*, 2006), and (Chu *et al.*, 2007). Fusing information from multiple spectrum bands has also been studied in (Pan *et al.*, 2003). Insightful surveys of such approaches can be found in the works of (Socolinsky and Selinger, 2002) and (Kong *et al.*, 2005). However, it should be noted that imaging beyond visual range spectrum is not widespread as yet.

Apart from the illumination issue, the other challenge in face recognition is to recognize faces at different poses. Since a human head has non-planar geometry, face images appear to be substantially different depending on head orientation or camera viewpoints. This complicates the task of recognition as most of the algorithms are primarily based on measuring similarity among the training and test samples. Approaches based on local feature analysis may overcome this issue to some extent by learning the correspondence among features at different poses, as discussed in (Ashraf *et al.*, 2008), however, a comprehensive framework has still not been realised. To address this issue, many researchers have proposed the use of 3D range information of the faces. Significant works based on this can be found in (Gordon, 1992), (Achermann *et al.*, 1997), (Achermann and Bunke, 2000) and (Hesher *et al.*, 2003). However, it is observed that using only range information does not yield desirable results, as many discriminating features are based on skin tones and textures. Studies in face recognition using both the range and texture images (3D+2D) have shown improved results compared to using 3D or 2D information separately (Chang *et al.*, 2003), (Tsalakanidou *et al.*, 2004), (Lu and Jain, 2005), and (Li *et al.*, 2005b). Surveys of such 3D based methods are available in (Bowyer *et al.*, 2006) and (Scheenstra *et al.*, 2005). However, similar to the infrared technology, 3D imaging technology is still not widely adopted and thus the benefit of using 3D information may elude us in years to come.

Some researchers have proposed using shape-from-shading (Zhao and Chellappa, 2000) or depth map recovery from stereo image pairs (Lao *et al.*, 2000), (Medioni and Waupotitsch, 2003) for 3D model construction when range data is unavailable. However, the underlying technology of both shape-from-shading and stereo depth estimation is still not dependable when used in unstructured scenes. Another type of work that involves the application of 3D models is the use of 3D morphable models. In this approach, a generic 3D model is modified to suit a subject's face based on single/few images taken from different viewpoints in a standard laboratory setting. This 3D model is then utilized to generate multitudes of face images covering different poses for the training of standard 2D image based classifiers. Descriptions of such models can be found in (Blanz *et al.*, 2002), (Blanz and Vetter, 2003), (Romdhani and Vetter, 2003), (Weyrauch *et al.*, 2004), and (Jiang *et al.*, 2005). Though this method can handle pose variations quite well, it has its own drawbacks: first, it does not allow the modelling of certain features such as hair, which may contain important discriminative information, and second, fitting the 3D model is extremely computationally intensive, making it non-viable in many real world scenarios. Therefore, the majority of face recognition research is still limited to the recognition problem based on 2D images taken with visual range cameras.

2.3 Face Recognition based on Holistic Features

Face recognition based on holistic features takes the whole face image, suitably cropped, for both training and testing. Predominately, the gray-level pixel values constitute the features and 2D images are first vectorized based on row-ordering to obtain the face feature vectors. In this section, first we present the linear classification methods that use these types of features, such as Eigenface (Turk and Pentland, 1991b), Fisherface (Belhumeur *et al.*, 1997) etc. Next, we briefly discuss other types of holistic features such as DCT coefficients, Edgemaps etc., that have been investigated in place of pixel values, and then conclude our discussion by reviewing work on nonlinear modelling techniques based on both the Kernel feature mapping and the manifold learning approaches.

2.3.1 Linear Methods

As noted previously, the first seminal work in face recognition is considered to be the Eigenface method, due to (Turk and Pentland, 1991b) and (Turk and Pentland, 1991a). Influenced by the work on image analysis by Karhunen-Loève transform (Kirby and Sirovich, 1990), they propose performing Principal Component Analysis (PCA) on training faces constituted by a collection of vectorized intensity images to obtain a reduced face subspace. Facial images are projected on that low dimensional subspace to obtain concise descriptions and a Nearest-Neighbour classifier (Cover and Hart, 1967) is employed for similarity matching. The combination of a simple and elegant solution and repeatability of experiments made the algorithm extremely popular and was instrumental in bringing face recognition research into the focus of scientific communities. Soon it was realised that although the algorithm handles lighting changes well, it fails to handle pose variations. To address the issue, (Pentland *et al.*, 1994) propose a modular approach where instead of creating a general subspace, a set of pose dependent subspaces are constructed and testing was preceded by identifying the correct pose of the test face and then using the right subspace for projection. Whilst the original classifier was based on an NN based approach, some research efforts emphasized using different classification techniques. Notably, the work in (Moghaddam and Pentland, 1997) propose modelling parametric probability distributions for each person in the eigenspace and then employ Bayesian decision theory for classification. Another notable distance measure used is the nearest-feature line distance, as discussed in the works of (Li and Lu, 1999), and (Zhou *et al.*, 2000), which can potentially improve the generalization capability of the classifier.

Another unsupervised data analysis approach that is similar in line to the PCA technique is the Independent Component Analysis (ICA). Whilst PCA obtains a subspace that is optimal for data compression, ICA is built to achieve statistical independence based on higher-order statistics of data. ICA was primarily invented in the signal processing community for blind deconvolution of mixed signals (Comon, 1989), and subsequently its scope for data analysis was also identified (Comon, 1994). The basic premise of ICA is of two types (Bell and Sejnowski, 1995),(A. Hyvrinen and Oja, 2001):

- **Type-I**: Generate components that are statistically independent to one another,
- **Type-II**: Generate statistically independent representations of training image vectors, similar to Factorial Coding (Schmidhuber, 1992).

ICA for face image representation is applied on the PCA-compressed data, then an eigenvalue type problem is solved based on the efficient algorithm by (Bell and Sejnowski, 1995). Initial studies on the performance of ICA in face recognition were reported in (Stewart Bartlett *et al.*, 1998), (Bartlett, 2001) and in the work of (Bartlett M.S. and T.J., 2002). Though initially it was promoted to be superior to the PCA based Eigenface type methods, the claim was later disputed in works of (Bruce *et al.*, 2002), and (Draper *et al.*, 2003). Based on extensive experiments, they conclude that whilst in some cases ICA clearly outperforms PCA, in some other cases the reverse can be true.

Both the Principal Component Analysis and the Independent Component Analysis provide optimal subspace for representation based on different sets of criteria. Whilst PCA is optimal from the data compression point of view, ICA is optimal if one wishes to have statistical independence among the components or the representations. However, none of the methods ask the important question, what subspace is best suited for the face classification task? The particular features that are important for distinguishing different face classes may well reside in low variance directions that PCA/ICA ignore to achieve optimal representations. Addressing this issue, Fisher Discriminant Analysis (FDA) (Fisher, 1936), or the Linear Discriminant Analysis (LDA) as commonly referred, looks for a projection subspace that maximizes class separability by maximizing the ratio of within class scatter to the between class scatter (Duda and Hart, 1973). LDA for face recognition was initially studied in the works of (Swets and Weng, 1996), (Etemad and Chellappa, 1997). Whilst (Etemad and Chellappa, 1997) suggests computing the subspace directly on the pixel space itself, the work of (Swets and Weng, 1996) suggests computing LDA on the eigenfeature space; e.g. LDA followed by the PCA of the image data matrix. Direct computation of subspace may be attractive but is limited by the size of training data. Often image vectors are far larger than the number of images per class, rendering the computed within-class scatter matrix to be singular. To avoid the singularity issue, PCA was seen to be a necessary step before LDA. Initial evaluation of this PCA+LDA method shows significant improvements of recognition accuracy over the traditional Eigenface method (Belhumeur *et al.*, 1997). Consequently, in (Zhao *et al.*, 1998) and (Zhao *et al.*, 1999) the authors propose a unified PCA+LDA framework to obtain a regularized LDA space that is shown to have superior generalization ability. There were several innovations in their work, which are:

- Construction of a universal PCA space, such that any new face images can be incorporated without retraining the PCA subspace.
- The PCA subspace is also chosen based on the eigenvectors which are face-like, instead of blindly choosing the largest eigenvectors.
- The LDA subspace is also shown to preserve discriminative power when additional face classes are incorporated without retraining the subspace.

These innovations and superior results of the LDA based classifiers against contemporary methods made it popular among the research communities. Whilst their method uses simple NN matching, a boosting based classifier on the LDA subspace (Lu *et al.*, 2006) or a neural network based classifier based on training an RBF network on the LDA feature space (Er *et al.*, 1999) were also considered. However, in subsequent efforts to rigorously evaluate LDA vis-a-vis PCA based Eigenface methods in (Martinez and Kak, 2001) shows that PCA may outperform the LDA based methods when the number of samples per class is small or, when the training data non-uniformly samples the underlying distribution. Subsequently, this is corroborated in (del Solar and Navarrete, 2005a) through their experiments on the FERET (Phillips *et al.*, 2000) data. Various improvements to standard LDA procedures have been proposed over time and notable among them was the Direct LDA (D-LDA) method. It is argued that the standard method of doing PCA prior to LDA to avoid singularity in the within-class scatter may adversely affect the class separability issue, as PCA may cull the important directions that contain discriminating information. The works in (Chen *et al.*, 2000) and (Yu and Yang, 2001) propose ways to address the singularity issue without performing PCA, *i.e.* performing LDA directly on the pixel space using some regularization schemes. Other important modifications can be found in (Lotlikar and Kothari, 2000), where a fractional step dimensionality reduction mechanism is proposed for more accurate weight computation of LDA with a weighted between class distance, and (Lu *et al.*, 2003b), where the same idea is explored in the D-LDA framework.

A different approach of subspace selection is discussed in the works of (Liu and Wechsler, 1998)(Liu and Wechsler, 2000), wherein the subspace selection is based on attaining an optimal combination of data compression (the objective of PCA) and classification accuracy (the objective of LDA). Based on the structural risk minimization framework proposed in (Vapnik, 1995), the authors propose a Genetic Algorithms (GA) based minimization procedure to reduce the empirical risk whilst maintaining desired level of class separability. The seed subspace is obtained by performing PCA, and GA minimization is performed iteratively by reorienting that subspace to achieve the above mentioned objective. Preliminary results on FERET suggest improved recognition accuracy over PCA and LDA, however, GA being computation intensive, is infeasible in many situations.

2.3.2 Holistic Features beyond Pixel Values

Until now we have considered only intensity images as the holistic feature, wherein images are represented as vectors in a pixel domain. However, there are several other holistic features that have been considered for face representation. Face recognition by coefficient vectors in the Fourier space has been studied in (Spies and Ricketts, 2000) and (El-Arief *et al.*, 2007) for scale and rotation invariant face recognition. Discrete Cosine Transform (Rao and Yip, 1990), which approximately follows the KL transform has also been studied as a data-independent subspace for face image representation and recognition (Hafed and Levine, 2001), (Pan *et al.*, 2000) and (Chen *et al.*, 2006). Contrary to the data-dependent subspace analysis of PCA, DCT provides a data-independent subspace, thereby saving computation cost. Also, in (Chen *et al.*, 2006) it is outlined that in the logarithmic DCT subspace images show lighting invariance property. However, in terms of the recognition performance, the DCT subspace fails to achieve consistent performance benefits over optimal data-dependent models of PCA. Extension of DCT based face features have been studied with limited success in (Zhu *et al.*, 2003), (Jing and Zhang, 2004), and (Chen *et al.*, 2005), wherein PCA and LDA are performed on the DCT space and in (Er *et al.*, 2005), where an RBF neural network is used for classification. Approaches based on Wavelet subband spaces have also been studied in (Feng *et al.*, 2000), and (Perlibakas, 2004), where PCA is performed on a specific subband of the Wavelet decomposition. In (Mandal and Wu, 2008), the Discrete Curvelet Transform was considered in place of wavelets, achieving limited success. All these data-independent subspace modelling techniques, whilst good for compression applications, lack the problem-specific knowledge that PCA or LDA achieve from analysis of training data, and in general are poor for face recognition. Another interesting holistic feature that has been studied is Edgemaps, where different types of edgemaps are used as features and the similarity of features is measured using set-theoretic distances (Takcs, 1998), (Gao and Leung, 2002). The use of Gabor features for face recognition has been discussed in (Aguilar-Torres *et al.*, 2007). However, none of the features have been thoroughly tested for their efficacy in face recognition. Intensity images as the holistic feature is widely used and well understood, and the same tradition is followed in our work.

2.3.3 Nonlinear Methods: Kernel Mapping

The linear subspace methods, as discussed above, are limited by the linear relationship between different features. Due to the complex nature of relationship among images, the linear subspaces may fail to generate well-separated face clusters, leading to poor classification accuracy. One way of addressing this lack of linear separability is by introducing non-linear classifiers such as Support Vector Machines (SVM) (Phillips, 1999), (Guo *et al.*, 2001). However, usually face recognition problems entail a large number of classes with small numbers of samples per class and in this setting, SVM can be expensive. Advances in kernel methods provide an alternative way of handling non-linearity (Schölkopf *et al.*, 1998). It introduces an efficient "kernel trick" to transform the primary feature space into a high-dimensional and often infinite-dimensional derived feature space, which is made up of non-linear combinations of the primary features. Assuming that with our choice of kernel function the classes are well-separated in this non-linear feature space, a linear classifier is sufficient. Initially the kernelized version of PCA was developed in (Schölkopf *et al.*, 1998) and its application to face recognition was studied in (Yang *et al.*, 2000). Essentially, PCA is performed on the derived non-linear feature space using different choices of kernels and a linear classifier is used for recognition. With the appropriate choice of kernel functions, the face images exhibit linear separability in the derived feature space, and therefore, a low complexity linear classifier will be sufficient to provide high speed testing. The experiments performed on several benchmarked datasets in (Yang *et al.*, 2000) show huge improvement over the classical PCA based Eigenface method. Following their success, nearly all the linear methods are "kernelized" to exploit the advantage of working in a nonlinear feature space. Face recognition based on Kernel-LDA (Mika *et al.*, 1999a), (Mika *et al.*, 1999b) was studied in (Liu *et al.*, 2002). Later on, face recognition using both Kernel-PCA and Kernel-LDA is thoroughly analysed in (Yang, 2001) and (Yang, 2002b). Variants of discriminant analysis rederived in the kernel framework are also proposed in (Lu *et al.*, 2002), (Lu *et al.*, 2003a). Further, in (Yang *et al.*, 2005) the authors showed that among the two possibilities of Kernel-LDA techniques, LDA applied on the Kernel-PCA space provides better performance than performing Kernel-LDA on the PCA subspace. Though kernel methods in general provide higher recognition rate than linear methods, they have their own shortcomings. According to (Gupta *et al.*, 2002), and (del Solar and Navarrete, 2005b) the shortcomings are:

- Kernel methods are considerably slower than their linear counterparts.
- Kernel parameter adjustments are difficult and data dependent.
- Kernel methods do not always provide performance benefits over linear methods as in situations where the classes are close to linearly separable, and in such cases, the introduction of kernels may actually degrade the performance.

Altogether, these drawbacks show that the kernel methods do not always provide viable options for real-world scenarios.

2.3.4 Nonlinear Methods: Manifold Learning

Recently, a number of studies have found that face images, specially when different poses are considered, lie in a non-linear manifold in the pixel space (Seung and Lee, 2000), (Roweis *et al.*, 2002), (Tenenbaum *et al.*, 2000), (Roweis and Saul, 2000). Moreover, recent neurological results in (Roweis and Saul, 2000) shows that the human brain possibly processes face images using manifolds. PCA and LDA by their formulation as linear operators only see the Euclidean structure of the face space, therefore, they are not equipped to understand the complex nonlinear structure of the face manifold. Kernel methods such as KPCA and Kernel-LDA, whilst providing non-linear feature mapping and consequently an ability to learn non-linear manifolds are lacking in two aspects: first, they do not consider the structure of face manifold explicitly, rather the structure is enforced by the choice of kernel functions, and second, they generally come with a high computational burden. Other types of nonlinear techniques have been proposed to learn the intrinsic non-linear manifold from the training data, notable among such methods are:

- ISOMAP (Tenenbaum *et al.*, 2000), which performs a global manifold learning through estimating the geodesic distance between all pairs of data points.
- LLE (Roweis and Saul, 2000) and (Saul and Roweis, 2003), which performs a locally coherent projection of a KNN (K Nearest Neighbour) graph of the data points by minimizing locally linear reconstruction errors, and

- Laplacian Eigenmap (Belkin and Niyogi, 2003), (Belkin and Niyogi, 2001), where a similar KNN graph between the input data points with edge distance weighted with a heat-kernel is embedded in a lower dimensional manifold by solving a generalized eigenvalue problem consisting of the Laplacian of the graph geometrically, which is similar to finding the eigenfunctions of the Laplace-Beltrami operator on the face manifold.

These methods are only helpful in finding the intrinsic manifold for the best representation, however, they cannot be directly used for face recognition as none of the methods have proposed a mechanism to evaluate new test points on the constructed manifold. Few approaches have been developed that use geodesic distances returned by the ISOMAP algorithm to compute the scatter matrix for further discriminant analysis of the data (Yang, 2002a), (Li *et al.*, 2005a). However, they are merely ad-hoc in nature since the geodesic distance is still undefined on the new test points.

Subsequently, the work in (He and Niyogi, 2003) extends the idea presented in the Laplacian Eigenmap algorithm by inserting a linear projection operator in the optimization step. The optimization now finds the optimal linear approximations of the eigenfunctions of the Laplace-Beltrami operator of the face manifold. Using this approach, along with the optimal projection of the data points (as in Laplacian Eigenmaps) a linear projection operator is obtained that solves the problem of mapping new data points to the manifold. Their algorithm is suitably called Locality Preserving Projection (LPP) because it preserves the local structure of the data in the new projection space. The efficacy of LPP for face recognition is also studied extensively in (He *et al.*, 2003) and (He *et al.*, 2005a), and is shown to outperform both the PCA and LDA by significant margins. Particularly, in (He *et al.*, 2005a) the authors provide further insights into the nature of the projection by providing a statistical interpretation of LPP and establish connections among PCA, LDA and LPP in a single graph embedding framework. Kernel extension of LPP is also studied in (Cheng *et al.*, 2005) and (Feng *et al.*, 2006). The authors of (Cheng *et al.*, 2005) employ LPP directly on the non-linear feature space, however, in (Feng *et al.*, 2006) the authors argue that performing LPP in the non-linear space of KPCA may be a better option. However, both these approaches provide only marginal performance benefit over plain LPP and their high cost of computation makes them undesirable in many situations. LPP like algorithms have also been studied in (Min *et al.*, 2004) and (Zheng and Yang, 2006), where the authors pose the problem in a variational setup. In (Pang *et al.*, 2005b) and (Pang *et al.*, 2005a), the authors follow the footsteps of LPP to insert a linear projection operator in the LLE framework. Whilst they provide alternative ways of modelling nonlinear manifolds with few functional advantages over LPP, their superiority to LPP for the recognition task has not been realised.

2.4 Multilinear Analysis and Its Application to Face Recognition

Historically, multifactor analysis to model multiway data took birth in the psychometric literature (Tucker, 1966) for the exploratory analysis of human psychological test results. The output of any psychological test is necessarily influenced by multiple independent factors and multifactor analysis was proposed to separate the influencing factors, assuming that they interact in a multilinear fashion *i.e.* the output is linearly related to each of the factors when others are fixed. Recently, in (Harshman and Lundy, 1994), the authors show the application of multifactor analysis in the context of phonetic analysis to successfully recover the latent relationship between different factors. Consequently, their approach is applied in many different types of domains as a tool for exploratory data analysis, notably, in DEDICOM (Bader *et al.*, 2006), where multifactor analysis is performed on the Enron email corpus to extract social relationship among the senders and recipients, in (Acar *et al.*, 2005) to analyse chatroom data, and in (Kolda *et al.*, 2005) to analyse web link data. An excellent survey of multiway data analysis and its application to different domains can be found in (Acar and Yener, 2009).

In computer vision applications, multiway data is generally termed as tensor data, consistent with the definition of tensor data in mathematics and similarly, the analysis is referred as the tensor analysis. Mathematically, tensors are higher-order abstractions of vectors and matrices and the elements of the tensor product of vector spaces. All the tools of tensor analysis are based on the separation of the underlying vector spaces, akin to the multiway data analysis framework. Early tensor analysis frameworks were based on factorizing tensors into the sum of rank-1 tensors (Harshman and Lundy, 1994). Though they are useful for exploratory analysis, they are not found suitable for vision related applications, as most vision problems deal with the representational issues rather than the actual relationship among the factors. Moreover, the nature of vision-related data requires managing different factors differently, which cannot be enforced in early frameworks. The NP-hard nature of the rank-1 decompositions was also a deterrent for such data intensive applications. Tensor analysis for vision problems took off after the seminal work of (De Lathauwer *et al.*, 2000) (Lathauwer *et al.*, 2000), who describe a different factorization method for tensors that addresses the issue of assigning different importance to factors by crafting the notion of multilinear-rank. In comparison to the rank-1 decomposition, where all the separated vector spaces have the same dimension, in the multilinear rank framework each vector space is allowed to have a different dimensionality. The decomposition is primarily based on the matrix SVD algorithm, making it easily understood as well as computationally inexpensive. The decomposition problem is formulated as a multilinear least squares problem, which is then solved using Alternating Least Squares (ALS) (Kroonenberg and Leeuw, 1980), and at each step Singular Value Decomposition (SVD) is used to estimate the least squares solution. The algorithm is suitably called the Higher Order Singular Value Decomposition (HOSVD) as the output of the decomposition exhibits similar properties as that of matrix SVD. HOSVD has been successfully applied in several vision related problems such as modelling of human gait (Haiping Lu; Plataniotis, 2006), facial expressions (Wang and Ahuja, 2003), image superresolution (Gong, 2005), face image hallucination (Liu *et al.*, 2007), face transfer (Vlasic *et al.*, 2005), and multi-dimensional textures (Vasilescu and Terzopoulos, 2004).

Face analysis based on multilinear factorization was first proposed in the seminal work on Tensorface by (Vasilescu and Terzopoulos, 2002b). They propose Multilinear Principal Component Analysis (MPCA) based on the Higher Order Singular Value Decomposition (HOSVD), which is then applied to a face dataset modelled as tensor data. Similar to Eigenface like techniques, they work on the vectorized intensity images and a tensor is created by stacking all the images in a hypercube with different variations modelled along separate dimensions of the hypercube. For example, when we only consider lighting variations among the training set, then a tensor of order 3 (*i.e.* a 3D hypercube) is created, where typically the first dimension is used to index people, the second for lighting and the third for pixels. PCA analysis is performed across all the dimensions (more commonly termed 'modes') and the interpretation of the result is obtained by introducing the concepts of *eigenpersons*, *eigenlighting*, *eigenviewpoints*, and the *eigenpixels*. Eigenface type methods rely on representing face images by linear combinations of *eigenfaces*, therefore, in the situation of multi-factor variation the *eigenfaces* fail to provide any intuitive interpretation. However, in the Tensorface (Vasilescu and Terzopoulos, 2002b) framework, the factors are separated and one can describe eigen-components in different contexts, such as *eigenpersons*, which captures the variations in the identity space, *eigenlighting* similarly captures the variations in the lighting space and so on. A face image of a particular person at particular variations can then be sufficiently described by computing the right linear combinations in the different eigenspaces, or by a multilinear combination between different factors. The work in (Vasilescu and Terzopoulos, 2002b) limited itself to the analysis part and described the potential benefit of tensor modelling as a flexible representational framework, where one can control the contribution of different eigen-components in the representation space, based on the importance of each mode. The first work on face recognition based on Multilinear PCA framework was proposed by (Vasilescu and Terzopoulos, 2002a), which showed improved performance over the Eigenface method. However, in their approach, only the *eigenpersons* are effectively used for recognition, whilst decompositions in other modes are used optionally to reduce the dimensionality of the associated vector-spaces (*e.g.* removing the dimensions with low variance). More precisely, if we want to identify persons, assuming we only model lighting and viewpoint variations, a set of *eigenmodes* are calculated for each combination of lighting and viewpoint. These *eigenmodes* are similar to *eigenfaces*, however, whilst *eigenfaces* capture variations over all the images, *eigenmodes* capture variations of images over person identities, at a particular combination of lighting and viewpoint. A test image is projected on each set of *eigenmodes* and a set of possible identity vectors is generated. It is known that in the *eigenperson* space, each person is represented by an unique identity vector, therefore, a test image can be classified by finding the closest pair among the generated coefficient vector set and the known identity vector of the persons in our database. Similar to Multilinear PCA, (Vasilescu, 2005) propose Multilinear Independent Component Analysis (MICA) with application to face recognition and in (Vasilescu, 2007), where the author performs MICA decomposition on 3D range images for face recognition. Another novel application of

HOSVD in face recognition can be found in the work of (Lee *et al.*, 2005a), where the recognition scheme is primarily based on the application of 3D morphable models with the representation vectors consisting of shape and reflectance parameters in place of pixel values. At this point, we note that there has been a recent surge of research papers on face recognition using tensor modelling. However, there are two distinctive ways tensor modelling is applied: first, by modelling face images as a multilinear functional form, and second, by considering a 2D image or a 3D color image or a multi-dimensional face features as tensor objects and constructing classifiers for such tensor valued objects. The works in (Wang and Ahuja, 2005), (He *et al.*, 2005b), (Yan *et al.*, 2007b) and (Yan *et al.*, 2007a) fall in this category. However, our work falls in the first category of using multilinear face model, similar to the concept of (Vasilescu and Terzopoulos, 2002b).

2.5 Face Recognition from Few Samples

The face recognition approaches discussed above rely heavily on the availability of large number of training samples. Since most of the methods fundamentally work compare training and testing samples in a suitable representational subspace, it is imperative to provide training images in a wide range of conditions to make the classifier robust. However, in real-world surveillance scenarios, where persons of interest are mainly non-cooperative criminals, collecting adequate training images is an impossible task. Whilst there is a wealth of literature for traditional face recognition scenarios where the availability of a large number of training images is often implicitly assumed, there are relatively fewer works that deal with the case when only few training images are available.

Traditional approaches addressing "few samples" issue can be classified into two distinct categories. The first category includes the work of (Wu and Zhou, 2002) (Chen *et al.*, 2004), (Jung *et al.*, 2004), who propose to extract as much information as possible from the few sample images. However, they fail to address the problem of information that is absent in the training samples such as images at widely different lighting or viewpoints, the type of test images one can expect in the real-world face recognition task. The second category addresses the issue by providing synthesis models to generate training samples at novel conditions, such as at novel lighting or novel viewpoints. Notable works amongst them are, (Beymer and Poggio, 1995), where faces corresponding to different poses are generated within a small angle of variation by applying affine transformation to the frontal face images, and (Georghiades *et al.*, 2001b), where the authors propose an illumination cone model based on the lambertian assumption to learn an approximate global lighting model and then define an affine model in the reconstructed 3D face space to model pose changes. The latter approach is notable in that it handles lighting issues quite well, however, it fails to adequately handle the pose issue. Moreover, this approach does not address situations when other sources of variation are present, such as expression, ageing etc.

Recently, the problem has been approached from a completely new direction, where a generic face dataset of people, who are different from the people of interest is used to learn a generalized space of face images, so that even a few samples projected on that subspace can achieve better localization and consequently improved recognition performance. In (Martínez, 2002) the authors propose to learn an expression invariant subspace from the generic dataset and in (Member-Kim and Member-Kittler, 2005) a similar approach is proposed to learn a view-invariant subspace to ultimately perform classification of persons of interest with few samples. In (Wang *et al.*, 2005) the authors pose the problem as a feature selection problem on the PCA subspace of the generic dataset, while the same authors propose a modification in (Wang *et al.*, 2006) by introducing a multi-learner framework to improve the recognition performance. In (Majumdar and Ward, 2008) the authors propose to learn a generic LDA subspace similar to that of (Wang *et al.*, 2005).

2.6 Closing Remarks

In this chapter we have reviewed the releated background in face recognition. Face recognition is an exciting research field having a large body of scientific effort, however, our main focus is on the holistic features based methods that use pixel values as the primary features, alongside providing a brief review on other types of works. The review shows how face recognition approaches based on simple PCA based methods changed over the years into more complex modelling such as Kernel feature mapping and Non-linear manifold modelling to achieve higher performance whilst being robust to extreme variations in different factors that influence a face image. We also discuss Multilinear analysis and the implication of this model as a tool for non-linear data modelling. Face recognition based on this modelling is also discussed in detail, while mentioning successful applications of this model in other vision related applications. We also discuss the real-world issues regarding the application of existing face recognition technologies when adequate training data is hard to obtain or costly, along with a brief review on different attempts to solve this problem. In the next chapter, we present the necessary mathematical background related to multilinear analysis and its application.

Chapter 3

Related Mathematical Background

In this chapter we present the mathematical background related to the multilinear analysis of face image ensembles. Historically, multilinear analysis was known as multi-factor decomposition, employed primarily to separate mixing factors assuming that the factors are independent and the mixing is a multilinear one *i.e.* each of the factors linearly influence the output when others are fixed (Harshman and Lundy, 1994). The decomposition is performed by expressing the data tensor as a sum of rank-1 tensors. A rank-1 tensor is by definition an outer product of vectors, where each vector corresponds to one factor. The rank of the decomposition is determined by the number of rank-1 tensors under the summation. However, this model is unsuitable for face recognition for two main reasons. First, the decomposition does not allow the factors to have different ranks or equivalently different dimensions of factor spaces, however, in most of the face recognition problems the number of pixels is many times larger than the number of people, and hence assigning same dimensions to both the pixel factor and the person factor is unsuitable. Second, this decomposition is a NP-hard problem, thus the computation can be expensive. Subsequently, (De Lathauwer *et al.*, 2000) propose a different decomposition framework based on the concept of multilinear rank, allowing one to assign different dimensions to different factor spaces with a reasonable complexity in terms of computation. This decomposition exhibits properties similar to the matrix SVD and is suitably termed Higher Order Singular Value Decomposition (HOSVD).

At this point we distinguish between the two terms commonly used in regards to the application of tensors:

- **Tensor Analysis:** This implies dealing with the tensor valued objects and understanding their nature, most commonly for tensor subspace computation and classification, akin to the way vector valued objects are treated.

- **Multilinear Analysis:** This implies that a data tensor is analysed with respect to its underlying factors. Since this implies multilinear relationships among the factor spaces, it is termed multilinear analysis.

The former is outside the scope of our discussion as the thesis is based on the multilinear analysis of a tensor created from an ensemble of face images.

The organization of the chapter is as follows: In Section 3.1, we present important concepts related to tensors. In Section 3.2, we present the tensor decomposition problem and its solution based on the Higher Order Singular Value Decomposition (HOSVD) and discuss its various properties. Following this, we present an iterative subspace method to improve the solution obtained from HOSVD. Next, in Section 3.3, we discuss the Multilinear PCA framework and present its application to face image ensembles, and finally, Section 3.4 concludes the chapter.

3.1 Tensor Concepts

A tensor $\mathcal{A}$ is defined as an element of the tensor space $\mathbb{R}^{I_1\times\ldots\times I_N}$ and is represented by a multidimensional array as:

$$\mathcal{A} = [a_{i_1,\ldots,i_N}]_{i,\ldots,i_N=1}^{I_1,\ldots,I_N} \tag{3.1.1}$$

The dimensions of a tensor are more commonly termed *modes*.

3.1.1 Tensor Properties

Definition 1. ***Addition and scalar multiplication:*** *For the tensor* $\mathcal{B} = [b_{i_1,\ldots,i_N}] \in \mathbb{R}^{I_1\times\ldots\times I_N}$ *and* $\alpha \in \mathbb{R}$,

$$\mathcal{A} + \mathcal{B} := [a_{i_1,\ldots,i_N} + b_{i_1,\ldots,i_N}] \tag{3.1.2}$$

and,

$$\alpha\mathcal{A} := [\alpha a_{i_1,\dots,i_N}] \tag{3.1.3}$$

Definition 2. ***Matricization of Tensors: Tensor Unfolding at k'th Mode*** *Matricization of a tensor refers to the process of creating a matrix from a tensor of order > 2. This operation is needed to create more complex tensor operations using the usual matrix operations as the building blocks. It is often referred to as tensor unfolding and defined with respect to the mode of unfolding. Following (Lathauwer* et al.*, 2000) the formal definition is given as:*

- *$A_{(k)}$ is denoted as tensor unfolding of $\mathcal{A}$ at k'th mode, and defined as $A_{(k)} \in \mathbb{R}^{I_k \times (I_{k+1} \dots I_N I_1 \dots I_{k-1})}$ and it contains the element $a_{i_1..i_N}$ at i_k'th row and at $[(i_{k+1}-1)I_{k+2}\dots I_N I_1 \dots I_{k-1} + (i_{k+2}-1)I_{k+3}\dots I_N I_1 \dots I_{k-1} + \dots + (i_N - 1)I_1 \dots I_{k-1} + (i_1 - 1)I_2 \dots I_{k-1} + (i_2 - 1)I_3 \dots I_{k-1} + \dots + i_{k-1}]$'th column.*

Intuitively, the operation indicates slicing the tensor along a particular direction depending on the mode of unfolding, and then putting the slices side-by-side in a matrix. The pictorial description is given in the Fig 3.1 for a third order tensor. Correspondingly, folding of an unfolded tensor refers to the reverse operation to get back the tensor from the unfolded representation.

Definition 3. ***Multilinear rank of a tensor:*** *The multilinear rank for the tensor $\mathcal{A}$ is a k-tuple $(r_1, \dots, r_N)$, where*

$$r_k = dim(R(A_{(k)})) = rank(A_{(k)}) \tag{3.1.4}$$

$R(A) = \{y|y = Ax\}$ is the range space of the matrix A, and $rank(A)$ is the conventional matrix rank.

Definition 4. ***Matrix times Tensor: Mode-k Multiplication*** *The mode-k multiplication of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ by a matrix $U \in \mathbb{R}^{I_k \times J_k}$ is denoted by $\mathcal{B} = \mathcal{A} \times_k U$. $\mathcal{B} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_{k-1} \times J_k \times I_{k+1} \dots \times I_N}$ and the entries of $\mathcal{B}$ are defined by,*

$$b_{i_1 \dots i_{k-1} j_k i_{k+1} \dots i_N} = \sum_{i_k=1}^{I_k} A_{i_1 \dots i_{k-1} i_k i_{k+1} \dots i_N} . U_{i_k, j_k} \tag{3.1.5}$$

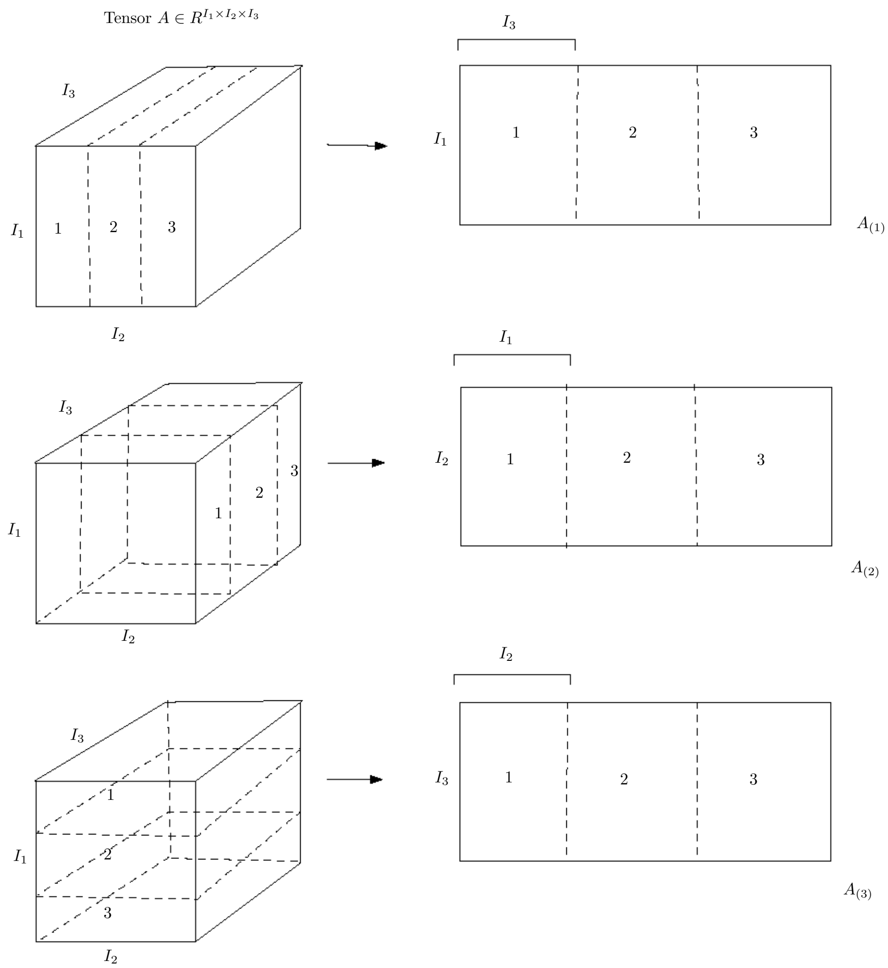

Figure 3.1: Unfolding of a third order tensor $\in \mathbb{R}^{3\times3\times3}$ in three different modes.

Equivalently, it can be shown that,

$$(\mathcal{A} \times_k U)_{(k)} = U^T A_{(k)} \tag{3.1.6}$$

The following identities hold for mode multiplications:

- *For* $U \in \mathbb{R}^{J_{k_1} \times I_{k_1}}$ *and* $V \in \mathbb{R}^{J_{k_2} \times I_{k_2}}$

$$(\mathcal{A} \times_{k_1} U) \times_{k_2} V = (\mathcal{A} \times_{k_2} V) \times_{k_1} U \tag{3.1.7}$$

- *For* $U, V \in \mathbb{R}^{I_k \times I_k}$

$$(\mathcal{A} \times_k U) \times_k V = \mathcal{A} \times_k (UV) \tag{3.1.8}$$

Definition 5. ***Inner product, tensor norm:*** *Inner product between the tensors* $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ *and* $\mathcal{B} \in \mathbb{R}^{I_1 \times \times I_2 \times \ldots \times I_N}$ *is defined as:*

$$\langle \mathcal{A}, \mathcal{B} \rangle := \sum_{i_1,\ldots,i_N} a_{i_1,\ldots,i_N} b_{i_1,\ldots,i_N} \tag{3.1.9}$$

The corresponding tensor norm is:

$$||\mathcal{A}|| = \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle} \tag{3.1.10}$$

This is the standard Frobenius norm for the tensor case. Also, we note that the norm is invariant under orthogonal transformation, i.e.

$$||\mathcal{A}|| = ||\mathcal{A} \times_1 V^{(1)} \times_2 V^{(2)} \ldots \times_N V^{(N)}|| \tag{3.1.11}$$

for orthogonal matrices $V^{(1)}, V^{(2)}, \ldots, V^{(N)}$.

3.2 Tensor Decomposition

In this section, we first present the tensor factorization problem similar in spirit to the matrix SVD. A matrix $D \in \mathbb{R}^{I_1 \times I_2}$ can be decomposed using SVD as a weighted sum of rank-1 matrices as:

$$D \simeq \sum_{i=1}^{min(I_1,I_2)} \alpha_i u_i \otimes v_i \tag{3.2.1}$$

where $\alpha_i \geq 0$ are the singular values, and u_i, v_i are the left-sided and right-sided singular vectors, respectively. Similarly, for tensors the decomposition can be formulated as expressing a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times ... \times I_N}$ as a weighted sum of rank-1 tensors:

$$\mathcal{A} \simeq \sum_{i=1}^{min(I_1,...,I_N)} \beta_i u_i^{(1)} \otimes u_i^{(2)} \otimes ... \otimes u_i^{(N)} \tag{3.2.2}$$

However, this canonical decomposition has one critical restriction in that it enforces the dimensions of all the separated vectors spaces, as defined by $[u_i^{(k)}]$ for $k = 1, ..., N$, to remain the same. This restriction is not suitable for many applications where the dimensions in one mode of a tensor can be arbitrarily larger than other modes. For a face image tensor, the dimension of a vectorized image can be excessively larger than the number of people, and hence, this factorization will generate a very poor approximation of the original tensor. To overcome this problem we formulate the problem based on the concept of *multilinear rank*. The formulation of tensor decomposition based on *multilinear rank* is similar to the formulation of SVD when deduced from the best-low rank approximation for matrices. The problem is stated below using formal notation.

Definition 6. ***Best rank-***$(r_1, ..., r_N)$ ***approximation:*** *Given* $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times ... \times I_N}$ *and the desired multilinear rank* $(r_1, ..., r_N)$*, find* $\mathcal{B} \in \mathbb{R}^{r_1 \times ... \times r_N}$ *and unitary matrices* $U^{(k)} \in \mathbb{R}^{I_k \times r_k}$ *for* $k = 1, ..., N$*, that solves:*

$$\min ||\mathcal{A} - \mathcal{S} \times_1 U^{(1)} \times_2 U^{(2)} ... \times_N U^{(N)}||^2 \tag{3.2.3}$$

Unfortunately, the above approximation problem does not have any closed-form solution and the existing algorithms are only able to find approximate solutions. Higher Order Singular Value Decomposition (HOSVD) offers the simplest solution with reasonably good approximation.

3.2.1 Higher Order Singular Value Decomposition

Following the work of (De Lathauwer *et al.*, 2000), we present the Higher Order Singular Value Decomposition (HOSVD) algorithm. The algorithm is primarily based on the application of SVD on tensors to compute a suboptimal solution to the *best rank* $(r_1, ..., r_N)$ approximation problem of 3.2.3. It is presented as the following theorem.

Theorem 1. ***Higher Order SVD:*** *Every* $(I_1 \times ... \times I_N)$*-tensor* $\mathcal{A} \in \mathbb{R}^{I_1 \times ... \times I_N}$ *can be written as the product:*

$$\mathcal{A} = \mathcal{S} \times_1 U^{(1)} \times_2 U^{(2)} ... \times_N U^{(N)} \tag{3.2.4}$$

in which,

1. $U^{(N)} = (U_1^{(N)} U_2^{(N)} ... U_N^{(N)})$ *is a unitary* $(I_N \times I_N)$ *matrix.*
2. $\mathcal{S}$ *is a* $(I_1 \times I_2 \times ... \times I_N)$*-tensor of which the subtensors* $\mathcal{S}_{i_n=\alpha}$*, obtained by fixing the k'th index to* α*, have the properties of:*
   (a) *all orthogonality: two subtensor* $\mathcal{S}_{i_n=\alpha}$ *and* $\mathcal{S}_{i_n=\beta}$ *are orthogonal for all possible values of* n, α *and* β *subject to* $\alpha \neq \beta$*:*

   $$\langle \mathcal{S}_{i_n=\alpha}, \mathcal{S}_{i_n=\beta} \rangle = 0, \text{ when } \alpha \neq \beta \tag{3.2.5}$$

   (b) *ordering:*

   $$||\mathcal{S}_{i_n=1}|| \geq ||\mathcal{S}_{i_n=2}|| \geq ... \geq ||\mathcal{S}_{i_n=I_N}|| \geq 0 \tag{3.2.6}$$

   for all possible values of n

   while the orthogonality of $U^{(k)}$*s and the all-orthogonality of* $\mathcal{S}$ *are the basic foundations, the ordering is more of a convention to enforce a particular ordering of the columns of* $U^{(k)}$*s.*

The decomposition shares many similarities to the conventional matrix SVD: first, the left and right singular vectors of SVD are generalized as the n-mode singular vectors and second, the role served by the singular values is maintained by the subtensors of the "core-tensor" $\mathcal{S}$. The dissimilarity here is that whilst for the matrix case the singular value matrix is pseudo-diagonal, in the case of tensors the core tensor $\mathcal{S}$ is not guaranteed to be pseudo-diagonal, and more often than not, it is a full tensor. The reason is that for the tensor case, the degrees of freedom offered by a $min(I_1, I_2, ..., I_n)$ non-zero diagonal core-tensor is not enough to model tensors with $I_1 I_2 ... I_N$ degrees of freedom.

3.2.1.1 HOSVD Computation

A matrix representation of the HOSVD can be obtained by unfolding $\mathcal{A}$ and $\mathcal{S}$ in the 3.2.4:

$$A_{(n)} = U^{(n)} S_{(n)} (U^{(n+1)} \otimes ... \otimes U^{(N)} \otimes U^{(1)} \otimes ... \otimes U^{(n-1)})^T \tag{3.2.7}$$

Now the properties of (3.2.5) and (3.2.6) imply that $S_{(n)}$ has mutually orthogonal rows, having Frobenius norms equal to $\sigma_1^{(n)}, \sigma_2^{(n)}, ..., \sigma_{I_n}^{(n)}$, with the ordering of $\sigma_1^{(n)} \geq \sigma_2^{(n)} \geq ... \geq \sigma_{I_n}^{(n)} \geq 0$. Let us define a diagonal matrix Σ as:

$$\Sigma := diag(\sigma_1^{(n)}, \sigma_2^{(n)}, ..., \sigma_{I_n}^{(n)}) \tag{3.2.8}$$

and a column-wise orthonormal matrix $V^{(n)} \in \mathbb{R}^{I_{n+1} ... I_N I_1 ... I_{n-1} \times I_n}$ as:

$$V^{(n)^T} := \tilde{S}_{(n)} (U^{(n+1)} \otimes ... \otimes U^{(N)} \otimes U^{(1)} \otimes ... \otimes U^{(n-1)}) \tag{3.2.9}$$

where $\tilde{S_{(n)}}$ is a normalized version of $S_{(n)}$ with the rows scaled to unit-norm, therefore,

$$S_{(n)} = \Sigma \tilde{S}_{(n)} \tag{3.2.10}$$

Putting it all back to the 3.2.7, we obtain:

$$A_{(n)} = U^{(n)} \Sigma V^{(n)^T} \tag{3.2.11}$$

$U^{(n)}, V^{(n)^T}$ are orthonormal matrices and Σ is a diagonal matrix. This is exactly the form of matrix SVD on the n-mode unfolded $A_{(n)}$. It is also easy to deduce that as the matrices $U^{(n)}$ for $n = 1, ..., N$ are orthonormal, the core tensor $\mathcal{S}$ can be obtained by switching the mode product from $\mathcal{S}$ to $\mathcal{A}$ as:

$$\mathcal{S} = \mathcal{A} \times_1 U^{(1)^T} \times_2 U^{(2)^T} ... \times_N U^{(N)^T} \tag{3.2.12}$$

Following this, we can state the HOSVD algorithm as follows:

1. For $i_n = 1, ..., i_N$ compute SVD of the n-mode unfolding of tensor $\mathcal{A}$ and set the left singular matrix as $U^{(n)}$.
2. Compute core tensor as:

$$\mathcal{S} = \mathcal{A} \times_1 U^{(1)^T} \times_2 U^{(2)^T} ... \times_N U^{(N)^T} \tag{3.2.13}$$

Next, we discuss the applicability of HOSVD for the *best rank* $(r_1, ..., r_N)$ approximation problem of 3.2.3 and the quality of approximation that is obtained by restricting the rank of individual vector space of $U^{(n)}$ for $n = 1, ..., N$.

3.2.1.2 Approximation Property

Assume that the original tensor $\mathcal{A}$ has a multilinear rank of $= (R-1, ..., R_N)$ and the reduced multilinear rank we want is $= (r_1, ..., r_N)$, where $1 \leq r_k \leq R_k$ (for $k = 1, .., N$). Let us say that this is done by removing the $(r_{k+1} - R_N)$ entries of $U^{(k)}$ corresponding to the lowest singular values of SVD of $A_{(k)}$ and setting the corresponding entries in $\mathcal{S}$ to zeros. Then, the approximated tensor $\tilde{\mathcal{A}} = \mathcal{S} \times_1 U^{(1)} \times_2 U^{(2)} ... \times_N U^{(N)}$ has the approximation error:

$$||\mathcal{A} - \tilde{\mathcal{A}}|| \;\; = \;\; \sum_{i_1=1}^{R_1} ... \sum_{i_N=1}^{R_N} s^2_{i_1 ... i_N} - \sum_{i_1=1}^{r_1} ... \sum_{i_N=1}^{r_N} s^2_{i_1 ... i_N} \tag{3.2.14}$$

where $s_{i_1 ... i_N}$ is the $i_1 ... i_N$'th entry to $\mathcal{S}$ and the error is only dependent on $\mathcal{S}$ as the Frobenius norm is not affected by orthonormal transformation *i.e.* multiplication by $U^{(k)}$. Now we further manipulate the 3.2.14 as:

$$
\begin{aligned}
||\mathcal{A} - \tilde{\mathcal{A}}|| &= \sum_{i_1=1}^{R_1} ... \sum_{i_N=1}^{R_N} s^2_{i_1 ... i_N} - \sum_{i_1=1}^{r_1} ... \sum_{i_N=1}^{r_N} s^2_{i_1 ... i_N} \\
&= \sum_{i_1=r_1+1}^{R_1} ... \sum_{i_N=r_N+1}^{R_N} s^2_{i_1 ... i_N} \\
&\leq \sum_{i_1=r_1+1}^{R_1} ... \sum_{i_N=+1}^{i_N=R_N} s^2_{i_1 ... i_N} + ... + \sum_{i_1=1}^{R_1} ... \sum_{i_N=r_N+1}^{R_N} s^2_{i_1 ... i_N} \\
&= \sum_{i_1=r_1+1}^{R_1} {\sigma^{(1)}_{i_1}}^2 + ... + \sum_{i_N=r_N+1}^{R_N} {\sigma^{(N)}_{i_N}}^2
\end{aligned}
\tag{3.2.15}
$$

where $\sigma^{(n)}_{i_k}$ is the n-mode singular values at i_k'th position.

As we see, the approximation error is bounded by the sum of squares of smaller singular values from all the modes. It is not an optimal solution, however, as the smaller singular values are assumed to be smaller in comparison with the larger singular values *i.e.* assuming ($\sigma^{(n)}_{r_n} \gg \sigma^{(n)}_{r_n+1}$). Then the solution provided by the HOSVD approximation $\tilde{\mathcal{A}}$ offers a close approximation of the tensor $\mathcal{A}$. Next, we present an algorithm based on orthogonal iteration to reduce the approximation error further.

3.2.2 Orthogonal Subspace Iteration

The approximation obtained by HOSVD for the *best rank* $(r_1, ..., r_N)$ approximation problem of 3.2.3 can be improved by an iterative method as proposed in the work of (Lathauwer *et al.*, 2000). Following the previous notation, the tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times ... \times I_N}$ is approximated by a tensor $\tilde{\mathcal{A}} = \mathcal{S} \times_1 U^{(1)} \times_2 ... \times_N U^{(N)} \in \mathbb{R}^{I_1 \times ... \times I_N}$ with the multilinear rank of $(r_1, ..., r_N)$. The error of approximation is given as:

$$||\mathcal{A} - \tilde{\mathcal{A}}||^2 \tag{3.2.16}$$

We want to minimize the error further by estimating better $\tilde{\mathcal{A}}$. In formal notation that is equivalent to,

$$\min f(\tilde{\mathcal{A}}) := ||\mathcal{A} - \tilde{\mathcal{A}}||^2 \tag{3.2.17}$$

We expand the 3.2.17 as:

$$\begin{aligned} f(\tilde{\mathcal{A}}) &= ||\mathcal{A} - \tilde{\mathcal{A}}||^2 \\ &= ||\mathcal{A}||^2 - 2\langle \mathcal{A}, \tilde{\mathcal{A}} \rangle + ||\tilde{\mathcal{A}}||^2 \end{aligned} \tag{3.2.18}$$

Now from the definition of the inner product, we can write as:

$$\begin{aligned} \langle \mathcal{A}, \tilde{\mathcal{A}} \rangle &= \langle \mathcal{A}, \mathcal{S} \times_1 U^{(1)} \times_2 \ldots \times_N U^{(N)} \rangle \\ &= \langle \mathcal{A} \times_1 {U^{(1)}}^T \times_2 \ldots \times_N {U^{(N)}}^T, \mathcal{S} \rangle \\ &= \langle \mathcal{S}, \mathcal{S} \rangle \\ &= ||\mathcal{S}||^2 \end{aligned} \tag{3.2.19}$$

Moreover, from the property of the Frobenius norm we know that $||\tilde{\mathcal{A}}|| = ||\mathcal{S}||$. Substituting this and 3.2.19 in 3.2.18 we obtain:

$$\begin{aligned} f(\tilde{\mathcal{A}}) &= ||\mathcal{A}||^2 - 2||\mathcal{S}||^2 + ||\mathcal{S}||^2 \\ &= ||\mathcal{A}||^2 - ||\mathcal{S}||^2 \end{aligned} \tag{3.2.20}$$

Hence, we can conclude that minimizing $f(\tilde{\mathcal{A}})$ is equivalent to maximizing $||\mathcal{S}||^2$. Let us define the optimization equation:

$$\max g := ||\mathcal{S}||^2 = ||\mathcal{A} \times_1 {U^{(1)}}^T \times_2 \ldots \times_N {U^{(N)}}^T||^2 \tag{3.2.21}$$

This optimization problem again does not have any closed-form solution. However, we can formulate an iterative solution based on the Alternating Variable Method (Kroonenberg and Leeuw, 1980). The solution proceeds in stages, fixing all the variables except one at each stage and then maximise 3.2.21 based on the free variables. Assuming that at the k'th stage we have fixed all the variables except $U^{(n)}$, the optimization problem at this stage becomes:

$$\begin{aligned} g(U^{(n)}) &= \|\mathcal{A} \times_1 U^{(1)^T} ... \times_{n-1} U^{(n-1)^T} \times_{n+1} U^{(n+1)^T} ... \times_N U^{(N)^T} \times_n U^{(n)}\|^2 \\ &= \|\mathcal{M}^{(n)} \times_n U^{(n)}\|^2 \\ &= \|U^{(n)}(M^{(n)})_{(n)}\|^2 \end{aligned} \tag{3.2.22}$$

where $\mathcal{M}^{(n)} = \mathcal{A} \times_1 U^{(1)^T} ... \times_{n-1} U^{(n-1)^T} \times_{n+1} U^{(n+1)^T} ... \times_N U^{(N)^T}$ and in the last step the mode-product is expressed as the matrix product. We also want to preserve the orthogonality of the matrix $U^{(n)}$. With this constraint the optimization problem becomes:

$$\begin{gathered} \max \|U^{(n)}(M^{(n)})_{(n)}\|^2 \\ s.t.\ U^{(n)^T} U^{(n)} = I \end{gathered} \tag{3.2.23}$$

This is a standard problem in linear algebra and the solution is provided by the SVD decomposition of $(M^{(n)})_{(n)}$ and $U^{(n)}$ is assigned the left singular matrix. We perform this at every stage for $n = 1, ..., N$, and at each stage use the newest value from the previous estimates. Once we finish computing all the $U^{(n)}$, we run the same process again until convergence is achieved. The combined procedure is termed as Tensor SVD and given in Algorithm 1.

3.2.2.1 Convergence

In the Tensor SVD algorithm, we used the Alternating Least Square (ALS) to optimize the multilinear maximization problem. The alternating variable method, whilst providing the simplest tool for solving multivariate optimization problems, suffers from poor convergence and cannot recover from local minima. However, we assume that the HOSVD provides a reasonably good estimate (Eldén and Savas, 2009) and as we improve the estimate at each iteration, at the end of the algorithm we can expect a solution that is sufficiently close to optimal. Alternative to ALS is the work of (Eldén and Savas, 2009), in which the authors propose a gradient based approach on a Grassmann manifold to achieve faster convergence. However, we found that the Tensor SVD algorithm is good enough for our purposes and leave such extensions for future consideration.

Algorithm 1: Tensor SVD

Input: $\mathcal{A} \in \mathbb{R}^{I_1 \times \ldots \times I_N}$ and desired multilinear rank $(r_1, ..., r_N)$ and desired approximation error level ϵ

Output: $\mathcal{S} \in \mathbb{R}^{r_1 \times \ldots \times r_N}$ and orthogonal matrices $U^{(n)} \in \mathbb{R}^{I_n \times r_n}$ for $n = 1, .., N$.

1. **Initial estimates (HOSVD):** Perform HOSVD on the tensor $\mathcal{A}$, as:

$$\mathcal{A} = \mathcal{S} \times_1 U^{(1)} ... \times_N U^{(N)}$$

   and set $\{U_0^{(n)}\}_{n=1}^N$ as the initial estimates for the following iterative procedure.

2. **Iterate until convergence:**

   - Compute $\mathcal{M}_{k+1}^{(1)} = \mathcal{A} \times_2 {U_k^{(2)}}^T \times_3 {U_k^{(3)}}^T ... \times_N {U_k^{(N)}}^T$, and set $U_{k+1}^{(1)}$ =left singular matrix of $SVD(M_{k+1}^{(1)})_{(1)}$
   - Compute $\mathcal{M}_{k+1}^{(2)} = \mathcal{A} \times_1 {U_{k+1}^{(1)}}^T \times_3 {U_k^{(3)}}^T ... \times_N {U_k^{(N)}}^T$, and set $U_{k+1}^{(2)}$ =left singular matrix of $SVD(M_{k+1}^{(2)})_{(2)}$
   - ...
   - Compute $\mathcal{M}_{k+1}^{(N)} = \mathcal{A} \times_1 {U_{k+1}^{(1)}}^T \times_2 U_{k+1}^{(2)} ... \times_{N-1} {U_{k+1}^{(N-1)}}^T$, and set $U_{k+1}^{(N)}$ =left singular matrix of $SVD(M_{k+1}^{(N)})_{(N)}$
   - Check convergence: $||\mathcal{A} - \mathcal{S} \times_1 U^{(1)} \times_2 U^{(2)} ... \times_N U^{(N)}||^2 \leq \epsilon$ or a certain numbers of iteration have been reached.

Converged values: $U^{(1)}, ..., U^{(N)}, \mathcal{S} = \mathcal{A} \times_1 {U^{(1)}}^T \times_2 ... \times_N {U^{(N)}}^T$

3.3 Multilinear Principal Component Analysis of Face Image Ensembles

In this section, we will present the multilinear analysis framework of face image ensembles. The analysis framework is primarily based on the Principal Component Analysis (PCA) of a data matrix, however, as tensors have higher numbers of modes than a matrix, PCA for the tensor data is built on the concept of multilinearity and termed as Multilinear PCA. We proceed by presenting the tensor model of face images and its decomposition with Multilinear PCA. In the end we also illustrate the various facets of Multilinear PCA decomposition, applying it on a small face dataset.

Face images taken from a person can differ based on the conditions of lighting, the viewpoints of cameras, expressions in the face, ageing etc. For a face image ensemble, being images from multiple people, identities of different persons also add another mode of variability. In the multilinear analysis framework all the factors that influence variations in a face image ensemble are assumed to be multilinear *i.e.* the influences are independently linear and the factors are mutually independent, and the analysis of the ensemble is performed through the analysis of those factor spaces.

Given a face image ensemble of N_p persons, each having images at exactly the same N_l number of lighting and N_v number of viewpoints (total number of images per person=$N_l N_v$), a face tensor is constructed as:

$$\mathcal{T}_{(i_p,i_l,i_v)} = I_{P_{i_p},L_{i_l},V_{i_v}} \tag{3.3.1}$$

where, $I_{P_{i_p},L_{i_l},V_{i_v}} \in \mathbb{R}^{N_x}$ is the image vector of i_p'th person at i_l'th lighting and i_v'th viewpoint. T is a tensor of order 4 and, $\mathcal{T} \in \mathbb{R}^{N_p \times N_l \times N_v \times N_x}$. A pictorial representation of the tensor is presented in Fig 3.2.

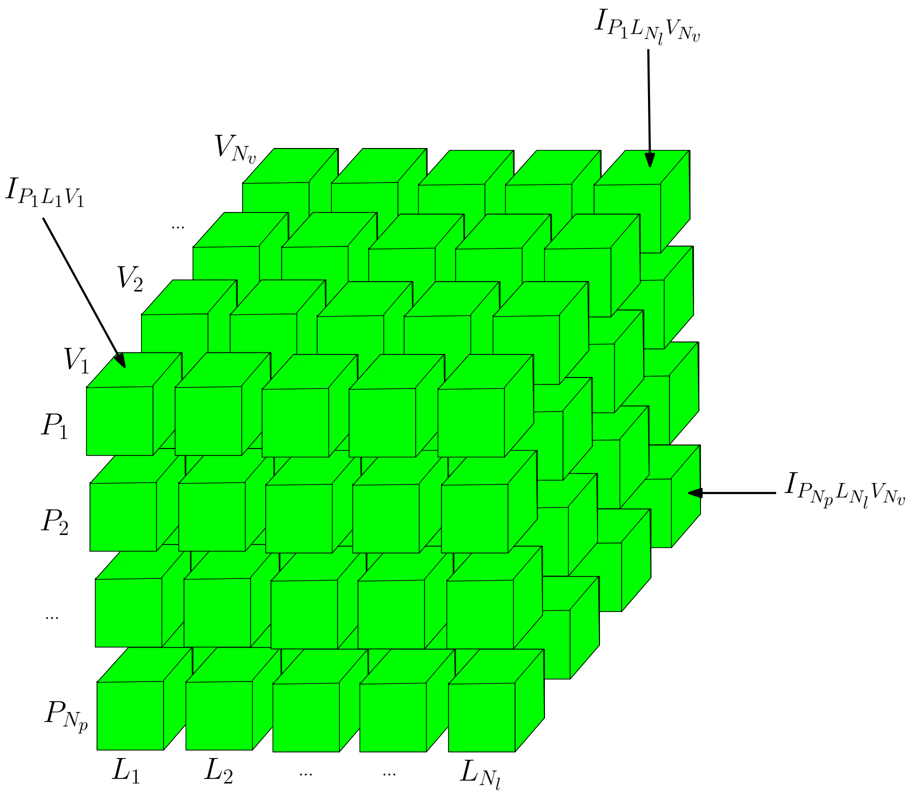

Figure 3.2: A face tensor $\mathcal{T} \in \mathbb{R}^{N_p \times N_l \times N_v \times N_x}$ constructed from the images of N_p persons each having images at exactly the same N_l number of lighting and N_v number of viewpoints, and the size of image vectors is N_x.

3.3.1 Multilinear PCA

Multilinear PCA of the face image tensor $\mathcal{T}$, as defined in 3.3.1, is performed by decomposing it using Tensor SVD (Algorithm 1), similar to the way PCA is performed by computing SVD of the image data matrix. The Tensor SVD of the face tensor yields four orthogonal subspaces and a core tensor as:

$$\mathcal{T} = \mathcal{S} \times_1 U^P \times_2 U^L \times_3 U^V \times_4 U^X \tag{3.3.2}$$

where the columns of U^P, U^L, U^V and U^X define the person, lighting, viewpoint and the pixel subspaces respectively. For the interpretation of the decomposition we will extend the concept of Principal Component Analysis into the multilinear analysis framework. The columns in U^X denote the traditional *eigenfaces* and the columns of U^P, U^L and U^V, denote the $N'_p(N'_p \leq N_p)$, $N'_l(N'_l \leq N_l)$ and $N'_v(N'_v \leq N_v)$ dominant eigenvectors of the person, lighting and viewpoint subspaces respectively. We refer to these eigenvectors as *eigen-person*, *eigen-lighting* and *eigen-viewpoint*, respectively.

Also, it is to be noted that each row of U^P corresponds to a particular person. We refer i_p'th row of U^P as the *person-space* representation of the person i_p. Interestingly, all the training images of i_p'th person, irrespective of any variation, are represented by a single point in the person space. Similarly, i_l'th row of U^L refers to the *lighting space* representation of the i_l'th lighting and i_v'th row of U^V refers to the *viewpoint space* representation of the i_v'th viewpoint. This unique association of vectors with each identity across all the modes is one of the useful characteristics of this framework and it will serve as a key element for solving various problems presented in subsequent chapters.

The core tensor, $\mathcal{S} \in \mathbb{R}^{N'_p \times N'_l \times N'_v \times N'_x}$ controls the mutual interaction between the person, lighting, viewpoint and pixel subspaces. It is an important element of the multilinear analysis as it can be viewed as a primitive structure that can be appropriately utilized to construct different representational images, which can be utilized to achieve different types of representation.

3.3.2 Example

In this section we illustrate the output of the multilinear decomposition on an example face image tensor. The tensor is created from a small dataset of 20 persons having images in 20 different lighting conditions (Gao *et al.*, 2008). Fig 3.3 shows the appearance of 20 persons and Fig 3.4 shows the lighting variation across the 20 different lighting conditions. The image vectors are of size 4200.

The face tensor $\mathcal{T}$ is created from the above dataset following Eq. 3.3.1. $\mathcal{T}$ is a tensor of order 3 and $\mathcal{T} \in \mathbb{R}^{20\times20\times4200}$. After factorizing $\mathcal{T}$ using Tensor SVD (Algorithm 1) we obtain:

$$\mathcal{T} = \mathcal{S} \times_1 U^P \times_2 U^L \times_3 U^X \tag{3.3.3}$$

Based on the concepts of *multilinear analysis* as discussed before, $\mathcal{S}$ is called the core tensor and the columns of $U^P \in \mathbb{R}^{20\times 20}, U^L \in \mathbb{R}^{20\times 20}$ and $U^X \in \mathbb{R}^{4200\times 400}$ defines the *eigenpersons*, *eigenlightings* and *eigenpixels* respectively. The *eigenpixel* is equivalent to the *eigenfaces* (Turk and Pentland, 1991b), however, the concepts of *eigenpersons* and *eigenlighting* is novel in the context of *multilinear analysis*. In Fig 3.5, Fig 3.6 and Fig 3.7 we illustrate the images of the 20 *eigenpersons* at actual neutral lighting, images of the first actual person at the 20 *eigenlightings*, and images of the first 5 *eigenpersons* at first 4 *eigenlightings*, respectively.

Figure 3.3: Multilinear analysis example: Images of 20 persons at the neutral lighting condition from PEAL dataset (Gao *et al.*, 2008).

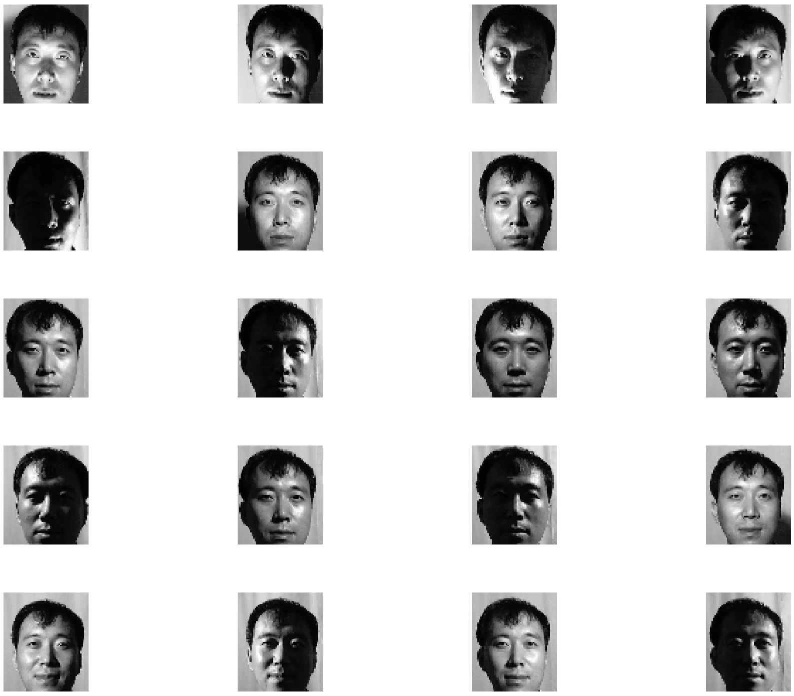

Figure 3.4: Multilinear analysis example: Images of the first person at 20 different lighting conditions.

Figure 3.5: Multilinear analysis example: Images of *eigenpersons* at neutral lighting. The images are obtained from the tensor $\mathcal{A}_P = \mathcal{S} \times_2 U^L \times_3 U^X$. Images are individually shifted and scaled to gray-scale for illustrative purposes. As the images show, the first *eigenperson* corresponding to the largest eigenvalue in the person mode decomposition captures the average face, followed by other variations.

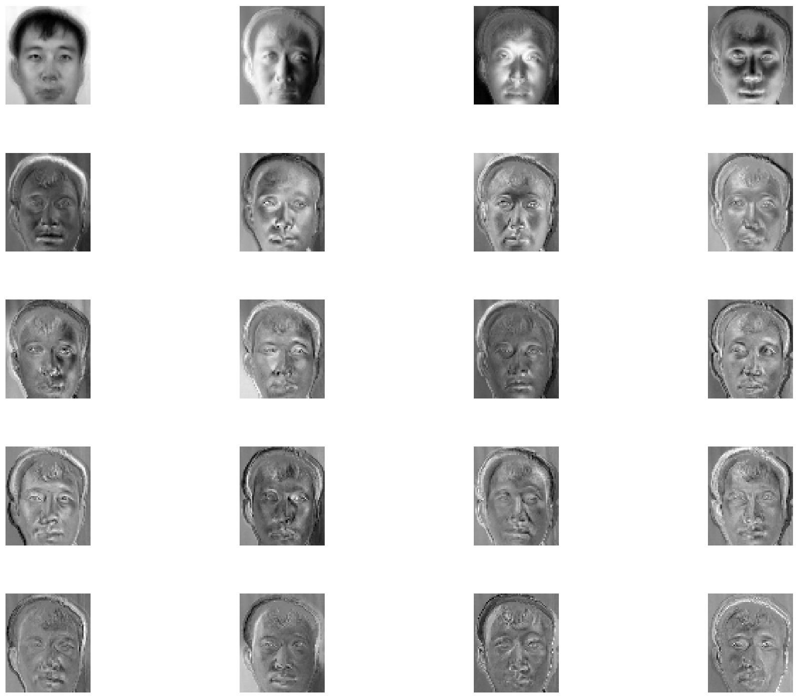

Figure 3.6: Multilinear analysis example: Images of the first person at all the *eigenlightings*. The images are obtained from the tensor $\mathcal{A}_L = \mathcal{S} \times_1 U^P \times_3 U^X$. Images are individually shifted and scaled to gray-scale for illustrative purposes. As the images show, the first *eigenlighting* corresponding to the largest eigenvalue in the lighting mode decomposition, captures the average lighting scenario, which is incidentally uniform lighting over whole of the face. Other *eigen*-variations are prominently recovered in subsequent images.

Figure 3.7: Multilinear analysis example: Images of the first 5 *eigenpersons* at the first 4 *eigenlightings*. The images are obtained from the tensor $\mathcal{A}_P = \mathcal{S} \times_3 U^X$. Images are individually shifted and scaled to grayscale for illustrative purposes.

Similar *eigen*-modes can be discerned from a face tensor with different and higher numbers of modes. It is also obvious that *multilinear analysis* adds many important facets over linear analysis (such as PCA) by defining *eigenpersons*, *eigenlightings*, etc. along with defining the *eigenpixels*, which are actually equivalent to the classical *eigenfaces*. This implies that *multilinear analysis* is potentially more powerful than the simpler linear analysis tools. Subsequent chapters will explore different eigenfactors and their inter-relations further to uniquely solve some of the problems associated with machine based face recognition.

3.4 Closing Remarks

In this chapter we presented the basic mathematical framework relating to tensor decomposition and Multilinear PCA that will be used extensively in this thesis. The basic premise of multilinear analysis is to model a face image ensemble as a face image tensor and to perform decomposition similar to SVD, the particular algorithms for the tensor case is termed as the Tensor SVD. The decomposition of tensor with Tensor SVD is then interpreted using PCA like analysis. This analysis framework is suitably termed Multilinear PCA as it deals with multiple modes in a multilinear fashion. We also present a basic interpretation of the Multilinear PCA of the face image tensor along with examples illustrating different aspects. In subsequent chapters we will suitably broaden the horizon of the multilinear analysis framework to propose unique approaches for face recognition, and subsequently provide solutions to some of the basic issues concerning face recognition application to real-world scenarios.

Chapter 4

Recognising faces in unseen modes: a tensor based approach

4.1 Introduction

As outlined in the introduction chapter, the implication of a practically useful face recognition technology is far reaching. In applications ranging from machine perception to security applications the effect will be tremendous. Traditional face recognition algorithms, whilst providing good recognition rates for face images at controlled circumstances, are vulnerable to face images subjected to uncontrolled variations in lighting and viewpoints. In order to be robust, face recognition algorithms should be able to identify faces irrespective of variations in facial expression, viewpoint or lighting condition. Popular face recognition algorithms such as the Eigenface (Turk and Pentland, 1991b) and the Fisherface (Belhumeur *et al.*, 1997) are only suitable for situations where the identity of the person is the only factor being considered, encountering difficulty when there are variations in lighting, viewpoint, expression etc. This is because these linear models are not intrinsically equipped to deal with variations in more than one factor *i.e.* person identity. Attempts have been made to overcome limitations in linear models by introducing non-linearity in the classification stage, however, the results are inconsistent and the algorithms are computationally expensive (Yang *et al.*, 2000).

In comparison, multilinear models (Vasilescu and Terzopoulos, 2002b) provide an alternative to accommodate variations across multiple factors in a simplistic manner, providing a structure in non-linearity without excessive complication in the model. Facial images are organised as a data tensor, with different factors of variation modelled as different modes of the data tensor. Subsequent application of Higher Order SVD, a generalisation of SVD for higher order matrices, generates subspaces related to every factor of variation. The power of such modelling is that it enables us to construct effective representations depending on the variations observed in each subspace and the importance given to the associated factor (Vasilescu and Terzopoulos, 2002b). The effectiveness of such a representation results in better face recognition performance than the linear models, as reported by Vasilescu *et.al.* in (Vasilescu and Terzopoulos, 2002a). However, in their approach, only the person-mode decomposition is used for recognition, whilst other mode decompositions are used optionally to reduce the dimensionality of associated vector spaces (*e.g.* removing the dimensions with low variance). More precisely, if we want to identify persons when the facial images are only subjected to varying lighting conditions and viewpoints, a set of *eigenmodes* are derived for each combination of lighting and viewpoint. These *eigenmodes* are similar to *eigenfaces*, however, whilst *eigenfaces* capture variations over all the images, *eigenmodes* capture variations over identities at particular combinations of lighting and viewpoint. These eigenmodes constitute the basis of vector space at each combination of lighting and viewpoint. The notion of multilinearity implies that a person in the training set has the same coefficient vector across all the bases. A test image is projected on every basis and a set of candidate coefficient vectors is generated. This set is then compared pair-wise to the set of person-specific coefficient vectors to find the best matching pair.

An analysis of these approaches reveals the following shortcomings:

1. Though multilinear decomposition is used, essentially a set of coupled bases is computed, based only on person-mode decomposition. This, we believe, is a severe under-utilisation of the multilinear decomposition, which provides a mechanism to unearth the hidden *multilinear relationship* among all the factors (*i.e.* person, lighting and viewpoint).

2. The recognition procedure in (Vasilescu and Terzopoulos, 2002a) (Wang and Ahuja, 2003) requires linear projection at each lighting-viewpoint combination, and that makes the algorithm expensive when a large number of lighting conditions and viewpoints are available in the training set.

3. Testing images at unseen lighting or viewpoints have not been investigated well as the projection bases do not contain any information regarding variations in these factors.

Based on these observations and motivated to fully use the information of *multilinear relationship* among the factors, we propose a recognition approach overcoming these shortcomings. In particular, we base our approach on the core tensor, which represents how the various factors interact with one another in a *multilinear way* to create an image. When the core tensor is multiplied by the pixel space eigenvectors (*i.e. eigenimages*), it transforms the *eigenimages* into *eigenmodes*. Contrary to the previous person-specific *eigenmodes*, these *multilinear eigenmodes* are the result of interaction among all the *eigenfactors*: *eigenpersons*, *eigenlightings* and *eigenviewpoints*. This *multilinear eigenmodes* is then used as the projection basis in the subsequent recognition algorithm. The proposed basis contains information on the variations of all the factors, thereby it is equipped to handle unseen variation in factors. A linear projection operator is then defined over this basis to obtain a joint person-lighting-viewpoint description for a face image. Further, we provide a mechanism to compute the description vectors for the training set in an efficient way, directly from the decomposition result. We store the description vectors for all the training images along with the projection operator. Given a test image, we find the description vector by employing the projection operator and then seek the closest training image.

The main contributions of this work are:

1. Novel insights are provided for the core tensor and its importance is established as the controlling entity of the face tensor model. This was not mentioned in previous works.
2. Scalability is achieved by exploiting the variance information of the factor spaces, providing higher control over the recognition performance. For example, if we know that in test scenarios there will only be small variation in the viewpoints, we can discard most of the dimensions in the viewpoint space. This can potentially increase both the speed and accuracy of testing.

3. Our proposed approach is also applicable to test images taken at unseen conditions as the variations in all the factor spaces are properly accounted in the proposed projection basis. In comparison, the existing approach (Vasilescu and Terzopoulos, 2002a) effectively uses only the person space, which renders it ineffective when test images are at unseen conditions.

4. The proposed algorithm follows the conventional testing framework of projection followed by comparison, providing easy deployment in place of other approaches such as, Eigenface, Fisherface etc.

The chapter is organized as follows: in section 4.2, we briefly review the tensor model of face image ensembles and the Multilinear PCA algorithm, followed by a brief overview on the existing recognition approach in section 4.3. The proposed recognition approach is developed in section 4.4. In section 4.5, we provide some experimental results, and finally section 4.6 concludes the chapter.

4.2 Review: Tensor Model of Face Image Ensembles

We follow the tensor model as described in the section 3.3, where the face image tensor $\mathcal{T}$ is given by,

$$\mathcal{T}_{(i_p,i_l,i_v)} = I_{P_{i_p},L_{i_l},V_{i_v}} \tag{4.2.1}$$

with the usual notation of $I_{P_{i_p},L_{i_l},V_{i_v}}$ being the image vector of i_p'th person at i_l'th lighting and i_v'th viewpoint. The Multilinear PCA decomposition of $\mathcal{T}$ is similarly followed and is represented as:

$$\mathcal{T} = \mathcal{S} \times_1 U^P \times_2 U^L \times_3 U^V \times_4 U^X \tag{4.2.2}$$

with the usual definition of $\mathcal{S}, U^P, U^L$, and U^V being followed. For a face image tensor $\mathcal{T}$ with additional modes, suitable extension of terminology can be achieved.

4.3 Existing Recognition Approach

We start with the Multilinear decomposition of the tensor $\mathcal{T}$ as defined in (4.2.2), and define $\mathcal{B}$ as:

$$\mathcal{B} = \mathcal{S} \times_2 U^L \times_3 U^V \times_4 U^X \tag{4.3.1}$$

Therefore, $\mathcal{B} \in \mathcal{R}^{N_p' \times N_l \times N_v \times N_x}$. If $B_{(person)}$ denotes the unfolding of the tensor $\mathcal{B}$ in the person mode, then

$$B_{(person)} = \begin{bmatrix} I_{P_1^e L_1 V_1} & I_{P_1^e L_2 V_1} & .. & I_{P_1^e L_{N_l} V_{N_v}} \\ I_{P_2^e L_1 V_1} & I_{P_2^e L_2 V_1} & .. & .. \\ .. & .. & .. & .. \\ .. & .. & .. & .. \\ I_{P_{N_p'}^e L_1 V_1} & .. & .. & I_{P_{N_p'}^e L_{N_l} V_{N_v}} \end{bmatrix} \tag{4.3.2}$$

where, $I_{P_{i_p}^e L_{i_l} V_{i_v}}$ is the image of i_p'th *eigenperson* for the i_l'th lighting and i_v'th viewpoint. This can be rewritten as:

$$B_{(person)} = \begin{bmatrix} b_{(1,1)} & b_{(2,1)} & .... & b_{(N_l,N_v)} \end{bmatrix} \tag{4.3.3}$$

where,

$$b_{(i_l,i_v)} = \begin{bmatrix} I_{P_1^e L_{i_l} V_{i_v}} \\ I_{P_2^e L_{i_l} V_{i_v}} \\ .... \\ .... \\ I_{P_{N_p'}^e L_{i_l} V_{i_v}} \end{bmatrix} \tag{4.3.4}$$

$b_{(i_l,i_v)}$ is used as the projection basis for the (i_l, i_v)'th combination. This results in $N_l \times N_v$ number of distinct projection bases. It is also evident from (4.3.1) and (4.2.2) that,

$$\begin{aligned} T &= \mathcal{B} \times_1 U^P \\ or, \quad T_{(i_l,i_v)} &= U^P b_{(i_l,i_v)} \\ or, \quad I_{P_k, L_{i_l}, V_{i_v}} &= c_k b_{(i_l,i_v)} \end{aligned} \tag{4.3.5}$$

where, c_k is the k'th row of the matrix U^P and it is specific to the k'th person. The recognition algorithm is based on (4.3.5). A test image is projected on the basis of $b_{(i_l,i_v)}$, for all i_l and i_v to generate a set of candidate coefficient vectors $\{c_{i_l,i_v}\}$. The best matching c_{p_m} (*i.e.* that minimizes $||c_p - c_{i_l,i_v}||$, for all i_l, i_v and p) identifies the test image as that of the person p_m. As this method uses lighting-viewpoint specific subspaces for projection, we term it MPCA-LV.

4.4 Proposed Recognition Approach

Motivated by our desire to exploit *multilinear relations* amongst the factors, we present a new recognition approach using the information in the core tensor. We derive the recognition scheme for the Multilinear PCA decomposition. Starting from (4.2.2) we define $\mathcal{A}$ as:

$$\mathcal{A} = \mathcal{S} \times_4 U^X \tag{4.4.1}$$

Let $A_{(person)}$ denote the unfolding of the tensor $\mathcal{A}$ in the person mode, then $A_{(person)}$ can be written as:

$$A_{(person)} = \begin{bmatrix} I_{P_1^e L_1^e V_1^e} & I_{P_1^e L_2^e V_1^e} & .. & I_{P_1^e L_{N_l'}^e V_{N_v'}^e} \\ I_{P_2^e L_1^e V_1^e} & I_{P_2^e L_2^e V_1^e} & .. & .. \\ .. & .. & .. & .. \\ .. & .. & .. & .. \\ I_{P_{N_p'}^e L_1^e V_1^e} & .. & .. & I_{P_{N_p'}^e L_{N_l'}^e V_{N_v'}^e} \end{bmatrix} \tag{4.4.2}$$

where, $I_{P_{i_p}^e L_{i_l}^e V_{i_v}^e}$ is the image of i_p'th *eigenperson* for i_l'th *eigen-lighting* and i_v'th *eigen-viewpoint*. Further, $A_{(person)}$ can be rewritten as:

$$A_{(person)} = \begin{bmatrix} a_{(1,1)} & a_{(2,1)} & .... & a_{(N_l',N_v')} \end{bmatrix} \tag{4.4.3}$$

where,

$$a_{(i_l,i_v)} = \begin{bmatrix} I_{P_1^e L_{i_l}^e V_{i_v}^e} \\ I_{P_2^e L_{i_l}^e V_{i_v}^e} \\ .... \\ I_{P_{N_p'}^e L_{i_l}^e V_{i_v}^e} \end{bmatrix} \tag{4.4.4}$$

We define a multilinear eigen-space as:

$$\tilde{A} = \begin{bmatrix} a_{(1,1)} \\ a_{(2,1)} \\ .... \\ .... \\ a_{(N'_l,N'_v)} \end{bmatrix} \tag{4.4.5}$$

$\tilde{A}$ constitutes a vector space spanned by the rows of $\tilde{A}$. The $N'_p \times N'_l \times N'_v$ rows combine elements of *eigen-person*, *eigenlighting* and *eigenviewpoint* and form the basis of this vector space. We call this basis *multilinear eigenmodes* and use it as the projection basis for the facial image.

The proposed basis has two distinct advantages. First, compared to the conventional tensor based method, the proposed basis involves *eigenmodes* across person as well as all the factors of variation, thereby truly exploiting the *multilinear relations* obtained from the multilinear decompositions. Second, like PCA and unlike the conventional tensor based method, the proposed method defines a unified basis for projection. However, in the case of multi-factor variations in the dataset, our proposed basis is physically more interpretable than that of the PCA.

The projection matrix, P for the basis $\tilde{A}$ is:

$$P = \tilde{A}^{+} \tag{4.4.6}$$

where, $\tilde{A}^{+}$ is the *Moore-Penrose pseudoinverse* of $\tilde{A}$. Next we will show that the coefficient vector of training images for the projection matrix in (4.4.6) can be directly calculated from the matrices U^P, U^L and U^V in an efficient way.

Definition 7. *Let us define M_P as:*

$$M_P = \begin{bmatrix} U^P & 0 & 0 & 0 \\ 0 & U^P & 0 & 0 \\ .... & .... & .... & .... \\ 0 & 0 & 0 & U^P \end{bmatrix} \tag{4.4.7}$$

where, U^P is repeated diagonally for $N_l \times N_v$ times.

Let us define M_L as:

$$M_L = \begin{bmatrix} UL & 0 & 0 & 0 \\ 0 & UL & 0 & 0 \\ .... & .... & .... & .... \\ 0 & 0 & 0 & UL \end{bmatrix} \tag{4.4.8}$$

here, UL is repeated diagonally for N_v times and is defined as:

$$UL = \begin{bmatrix} UL(1,1) \\ UL(2,1) \\ .... \\ .... \\ UL(N'_p, 2) \\ UL(1,2) \\ .... \\ .... \\ UL(N'_p, N_l) \end{bmatrix} \tag{4.4.9}$$

where $UL(i,j)$ is a row vector of size $N'_p \times N'_l$ defined as:

$$UL(i,j)_{(i+N'_p\times(k-1))} = \begin{cases} U^L(j,k) & \textit{for } k = 1, ..., N'_l \\ 0 & \textit{otherwise} \end{cases} \tag{4.4.10}$$

Let us define M_V as:

$$M_V = \begin{bmatrix} UV(1,1) \\ UV(2,1) \\ .... \\ .... \\ UV(N'_p \times N'_l, 1) \\ UV(1,2) \\ .... \\ .... \\ UV(N'_p \times N'_l, N_v) \end{bmatrix} \tag{4.4.11}$$

where $UV(i,j)$ is a row vector of size $N'_p \times N'_l \times N'_v$ defined as:

$$UV(i,j)_{(i+N'_p \times N'_l(k-1))} = \begin{cases} U^V(j,k) & \textit{for } k = 1,..,N'_v \\ 0 & \textit{otherwise} \end{cases} \tag{4.4.12}$$

Theorem 2. *Let $M = M_P \times M_L \times M_V$. If m_k is the k'th row of the matrix M then,*

$$m_k = I_{P_{i_p} L_{i_l} V_{i_v}} \times P \tag{4.4.13}$$

where, $i_p = ((k-1) \bmod N_p + 1)$, $i_l = ((\lceil \frac{k}{N_p} \rceil - 1) \bmod N_l + 1)$ and $i_v = (\lceil \frac{k}{N_p \times N_l} \rceil)$.

Proof. Let us refer to $\tilde{A}$ in (4.4.5), which is:

$$\tilde{A} = \begin{bmatrix} a_{(1,1)} \\ a_{(2,1)} \\ .... \\ .... \\ a_{(N'_l, N'_v)} \end{bmatrix} \tag{4.4.14}$$

From the definition of $a_{(i_l, i_v)}$ (4.4.4) we can observe that $\tilde{A}$ contains images of interaction between N'_p *eigen-persons,* N'_l *eigenlightings* and N'_v *eigenviewpoints*. The images in $\tilde{A}$ are organized in such a way that the index for the *eigen-persons* varies the fastest, followed by the index for the *eigenlightings* and finally, the index for the *eigenviewpoints* varies the slowest. This structure implies that the image of interaction between the i'_p'th *eigen-person*, i'_l'th *eigen-lighting* and i'_v'th *eigenviewpoint*, $I_{P^e_{i'_p} L^e_{i'_l} V^e_{i'_v}}$, can be found at the

$$(N'_p \times N'_l \times (i'_v - 1) + N'_p \times (i'_l - 1) + i'_p)\text{'th row of } \tilde{A}$$

We know that the image due to the interaction between the i'_p'th *eigen-person*, i'_l'th *eigen-lighting* and i_v'th actual viewpoint, $I_{P^e_{i'_p} L^e_{i'_l} V_{i_v}}$, can be calculated as:

$$I_{P^e_{i'_p} L^e_{i'_l} V_{i_v}} = \sum_{i'_v=1}^{N'_v} U^V(i_v, i'_v) \times I_{P^e_{i'_p} L^e_{i'_l} V^e_{i'_v}} \tag{4.4.15}$$

Continuing,

$$\begin{aligned} I_{P^e_{i'_p} L^e_{i'_l} V_{i_v}} &= \sum_{i'_v=1}^{N'_v} \{U^V(i_v, i'_v) \times (N'_p \times N'_l \times (i'_v - 1) \\ &\qquad + N'_p \times (i'_l - 1) + i'_p)\text{'th row of } \tilde{A}\} \\ &= [i_v\text{'th row of } U^V] \times Y \times \tilde{A} \end{aligned} \tag{4.4.16}$$

where, Y is a *selection* matrix of size $N'_v \times (N'_p \times N'_l \times N'_v)$ whose i'_v'th row is defined as:

$$Y_{i'_v,k} = \begin{cases} 1 & \text{if } k = \{(N'_p \times N'_l \times (i'_v - 1) + \\ & \qquad N'_p \times (i'_l - 1) + i'_p)\} \\ 0 & \text{otherwise} \end{cases} \tag{4.4.17}$$

Clearly, the first row of Y has entry 1 at the $(N'_p \times (i'_l - 1) + i'_p)$'th position and the rest are zero, the second row has entry 1 at $(N'_p \times N'_l + N'_p \times (i'_l - 1) + i'_p)$'th position and the rest are zero, the third row has an entry 1 at $(N'_p \times N'_l \times 2 + N'_p \times (i'_l - 1) + i'_p)$'th position and the rest are zero and so on. Hence if,

$$Z = [i_v\text{'th row of } U^V] \times Y \tag{4.4.18}$$

then,

$$Z_{(i + N'_p \times N'_l (k-1))} = \begin{cases} U^V(i_v, k) & \text{for } k = 1, .., N'_v \\ 0 & \text{otherwise} \end{cases} \tag{4.4.19}$$

where $i = N'_p \times (i'_l - 1) + i'_p$ and Z is same as the definition of $UV(i, i_v)$ in (4.4.12). Therefore, from (4.4.16) and (4.4.18), we can write:

$$\begin{aligned} I_{P^e_{i'_p} L^e_{i'_l} V_{i_v}} &= Z \times \tilde{A} \\ &= UV(N'_p \times N'_l (k-1), i_v) \times \tilde{A} \end{aligned} \tag{4.4.20}$$

Let us denote,

$$B = M_V \times \tilde{A} \tag{4.4.21}$$

where M_V is as defined in (4.4.11). From (4.4.11) and (4.4.20) it is easy to see that M_V transforms the images of $\tilde{A}$ into the images, which are due to the interaction of N'_p *eigenpersons* and N'_l *eigenlightings* at N_v actual viewpoints. The organization of the images in B is similar to $\tilde{A}$. Now we can formulate a similar argument to prove that pre-multiplication of B by M_L generates images of N'_p *eigenpersons* at

N_l actual lighting conditions and N_v actual viewpoints. Let $C = M_L \times B$ then consequently pre-multiplication of C by M_P generates images of N_p actual persons at N_l actual lighting conditions and N_v actual viewpoints. Let

$$\begin{aligned} D &= M_P \times C \\ &= M_P \times M_L \times B \\ &= M_P \times M_L \times M_V \times \tilde{A} \end{aligned} \tag{4.4.22}$$

The structure of D is again similar to $\tilde{A}$ and the image of i_p'th person at i_l'th lighting and at i_v'th viewpoint can be found at,

$$(N_p \times N_l \times (i_v - 1) + N_p \times (i_l - 1) + i_p)\text{'th row of } \mathcal{D} \tag{4.4.23}$$

Let us define $M = M_P \times M_L \times M_V$, then from (4.4.22) we obtain:

$$\begin{aligned} D &= M \times \tilde{A} \\ \Longrightarrow \quad M &= D \times \tilde{A}^{+} \\ &= \mathcal{D} \times P \end{aligned} \tag{4.4.24}$$

where $P = \tilde{A}^{+}$. From (4.4.23) and from (4.4.24) we observe that $(N_p \times N_l \times (i_v - 1) + N_p \times (i_l - 1) + i_p)$'th row of M contains the coefficient of projection of the i_p'th person at i_l'th lighting and at i_v'th viewpoint ($I_{P_{i_p} L_{i_l} V_{i_v}}$), for the projection matrix P. That implies that if m_k is the k'th row of the matrix M then,

$$m_k = I_{P_{i_p} L_{i_l} V_{i_v}} \times P \tag{4.4.25}$$

where,
$$k = (N_p \times N_l \times (i_v - 1) + N_p \times (i_l - 1) + i_p) \tag{4.4.26}$$

Solving (4.4.26) for i_p, i_l and i_v we obtain, $i_p = ((k-1) \bmod N_p + 1)$, $i_l = ((\lceil \frac{k}{N_p} \rceil - 1) \bmod N_l + 1)$ and $i_v = (\lceil \frac{k}{N_p \times N_l} \rceil)$. □

It follows from the above theorem that matrix $M = M_P \times M_L \times M_V$ contains the coefficients of projection of all the training images for the projection matrix P and thus provides an efficient way to compute the coefficient set. Each row m_k of the matrix M refers to a training image, whose person, lighting and viewpoint indices are provided by the above theorem.

4.4.1 Algorithm

The algorithm for testing is given in Algorithm 2. As our method uses joint person-lighting-viewpoint space description for recognition, we term it MPCA-JS.

Algorithm 2: Testing algorithm for MPCA-JS

Input: Test image I_T, Projection matrix P and the Coefficient matrix M.

Output: Person identity of the test image p_T

1. Find the description vector m_T for the test image as,

$$m_T = I_T \times P$$

2. Use a Nearest Neighbour classifier to find the best matching description vector m_b *i.e.* that minimizes,

$$\min_k ||m_T - m_k|| \text{ for } k = 1, ..., (N_p \times N_l \times N_v)$$

where, m_k is the k'th row of the matrix M. The distance measure that we use is the *cosine distance*. For two vectors $\mathbf{a}$ and $\mathbf{b}$ the *cosine distance* between them is defined as:

$$cosine\_dist(\mathbf{a}, \mathbf{b}) = \frac{\langle \mathbf{a} \, , \mathbf{b} \rangle}{||\mathbf{a}|| \, ||\mathbf{b}||}$$

3. The person identity, p_T for the test image is the person identity of the best matching vector m_b and it is given by,

$$p_T = ((b-1) \bmod N_p + 1)$$

4.4.1.1 Complexity Analysis

The MPCA-JS algorithm has a complexity of matrix multiplication of $O(N_pN_lN_vN_x)$ to obtain the description vector of the test image, and a search complexity of $O(N_pN_lN_v)$ to determine the nearest neighbour in the description vector space.

4.5 Experiments, Analysis and Evaluation

In this section we evaluate our proposed face recognition approach, MPCA-JS against the existing multilinear analysis based method MPCA-LV (Vasilescu and Terzopoulos, 2002a) and the PCA based Eigenface (Turk and Pentland, 1991b). The experiments are conducted on a number of publicly available benchmark datasets. Specifically, we use PEAL (Gao *et al.*, 2008), YaleB Frontal (Georghiades *et al.*, 2001a) and Extended YaleB (Lee *et al.*, 2005b) face database for experiments. The Extended YaleB database contains images of 38 persons at 64 different lighting conditions and at 9 different viewpoints. The YaleB frontal is a subset of the Extended YaleB database, which contains images of 38 persons at 64 different lighting conditions, at the frontal viewpoint only. PEAL is a face-database of Chinese nationals at different poses, lighting conditions and expressions (Gao *et al.*, 2008). For our experiment we have only chosen frontal images of 20 persons at 20 different illuminations. The image size for the datasets are as follow: for YaleB Frontal, $[48 \times 42]$, for Extended YaleB, $[32 \times 32]$, and for the PEAL dataset, $[70 \times 60]$. Prior to the experiments, all the images were cropped and their eye-points were manually aligned. Then all the image vectors were normalized to unity. For HOSVD and other tensor operations, we used the tensor toolbox developed by Bader and Kolda in MATLAB™ (Bader and Kolda, 2007).

For the PEAL dataset, four sets of experiments are performed with 5, 7, 9 & 11 lighting conditions as training, and the rest for testing. For experiments on YaleB Frontal dataset, a similar four sets of experiments are performed with 5, 10, 15 & 20 lighting conditions as training, whilst the rest are used for testing. For experiments on the Extended YaleB database, images at 16 representative lighting conditions and at 5 representative viewpoints are used for training and the rest for testing. Thus, in the experiments concerning PEAL and YaleB Frontal datasets, the test images are at unseen lighting conditions and for the Extended YaleB database, the test images are either at unseen lighting or viewpoints or at both unseen lighting conditions and viewpoints. Each set of experiments on YaleB Frontal and PEAL dataset is repeated 20 times on 20 random partitions of the datasets and the average of the results are reported. The performance of the proposed approach is compared with the PCA(Turk and Pentland, 1991b) and the existing approach, MPCA-LV (Vasilescu and Terzopoulos, 2002a).

Traditionally, when PCA is used for recognition, the last few eigenvectors are removed to improve performance. Here we use *energy thresholding* to retain the top-k eigenvectors such that

$$\min_k \frac{\sum_{j=1}^{k} \lambda_j}{\sum_{j=1}^{n} \lambda_j} \geq thresh \tag{4.5.1}$$

where, λ_j is the eigenvalue corresponding to the j'th eigenvector and $\lambda_{j-1} \geq \lambda_j$, n is the total number of eigenvectors and $thresh$ is the user specified threshold. In the case of Multilinear PCA, the same *energy thresholding* is used to select the "k" in each mode. In our experiment we set thresholds for PCA and for the pixel mode of MPCA as 0.96, and for all other modes of MPCA the threshold is set as 0.99.

Regarding distance measures: PCA based recognition method uses the Euclidian distance measure, and multilinear based methods use the Cosine distance measure, as suggested in the prior work of (Vasilescu and Terzopoulos, 2002a). All the experiments have been performed in the Matlab environment, running on a Intel Xeon 8×2.3GHz server with 16GB RAM and without code optimization.

Experimental results on the PEAL dataset are presented in Figures 4.1 and 4.2, showing recognition accuracies and test times respectively. Similarly, Figures 4.3 and 4.4 present recognition accuracies and test times for experiments on the YaleB Frontal dataset and finally, Figure 4.5 presents results of experiments on the Extended YaleB dataset.

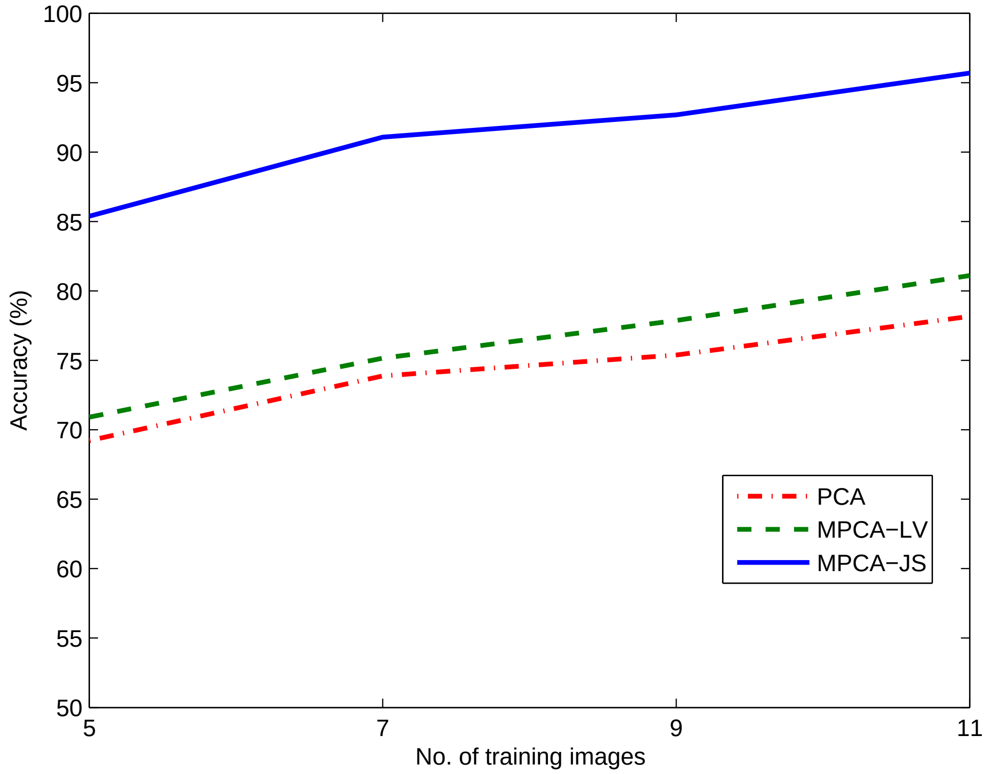

Figure 4.1: Recognition accuracy comparison on PEAL Lighting variation dataset. 20 persons with images at 20 different lighting conditions are divided into mutually exclusive training and testing parts. Four sets of experiments are performed with randomly chosen {5,7, 9, 11} lighting conditions for training and the rest for testing. Recognition accuracy is the average classification accuracy over 20 random partitions.

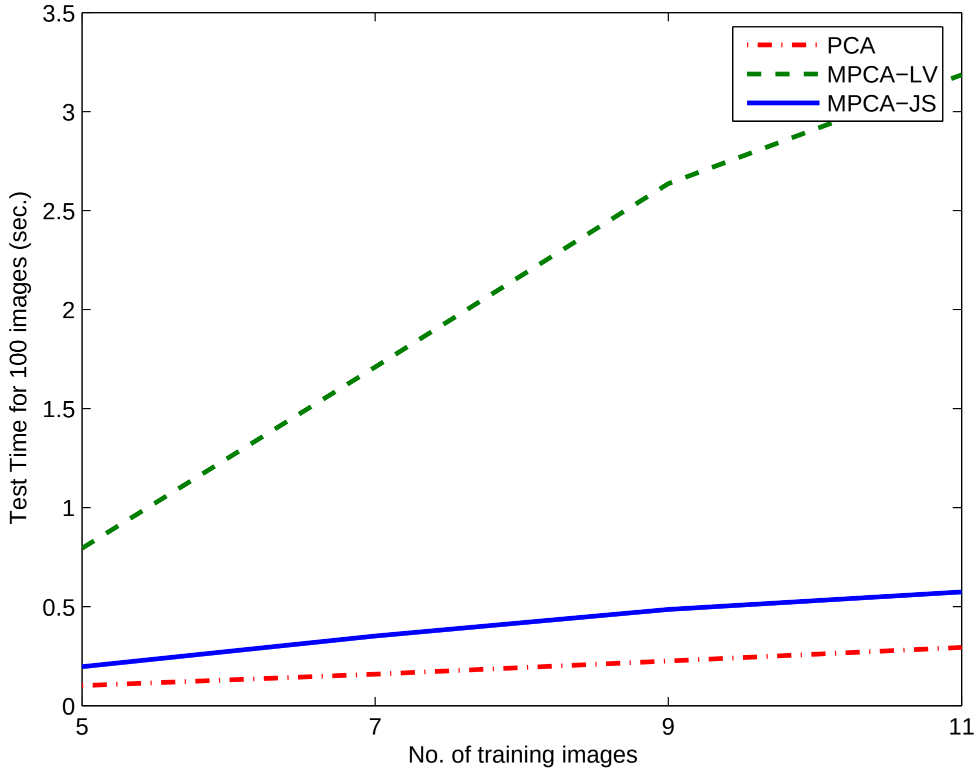

Figure 4.2: Testing time comparison on PEAL Lighting variation dataset. 20 persons with images at 20 different lighting conditions are divided into mutually exclusive training and testing parts. Four sets of experiments are performed with randomly chosen {5,10,15,20} lighting conditions for training and the rest for testing. Test time is for 100 test cases averaged over 20 random partitions.

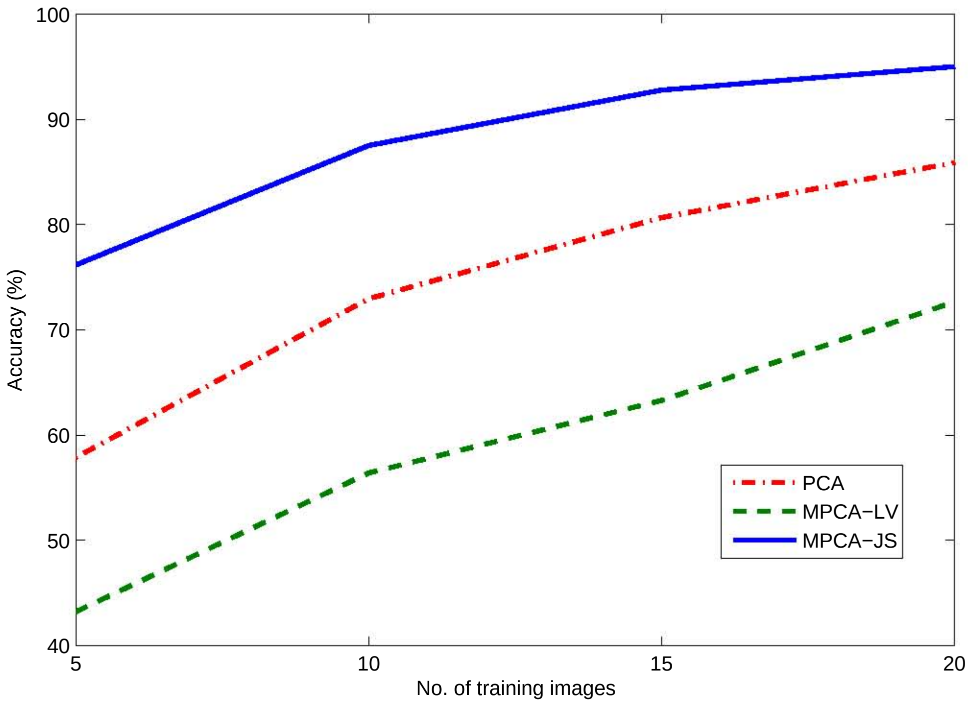

Figure 4.3: Recognition accuracy comparison on YaleB Frontal Lighting variation dataset. 38 persons with images at 64 different lighting conditions are divided into mutually exclusive training and testing parts. Four sets of experiments are performed with randomly chosen {5,10,15,20} lighting conditions for training and the rest for testing. Recognition accuracy is the average classification accuracy over 20 random partitions.

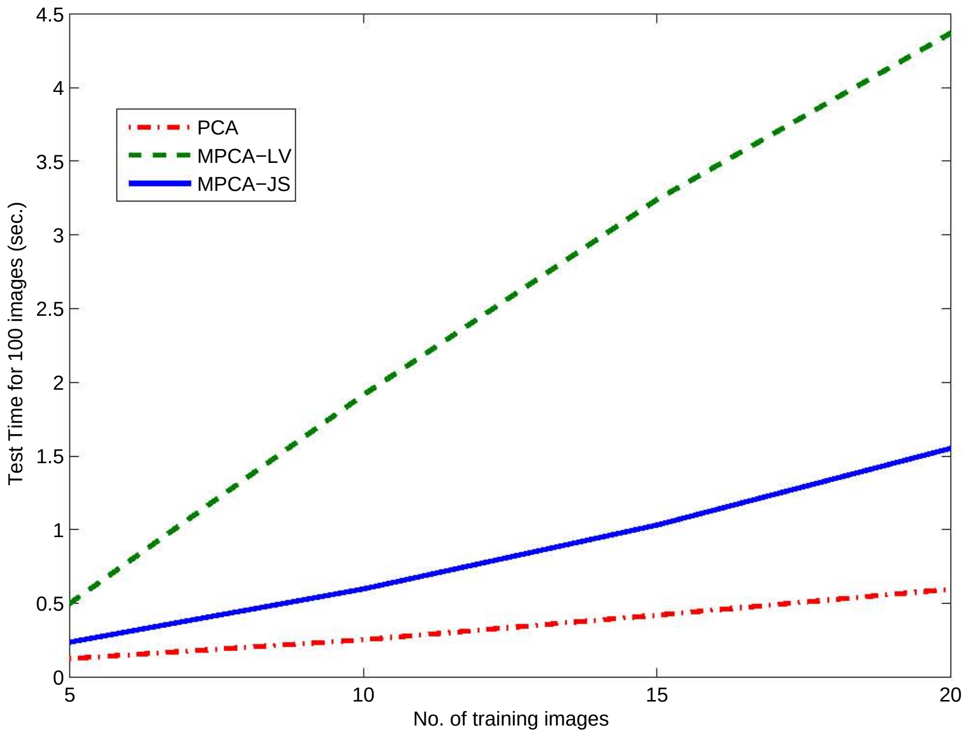

Figure 4.4: Testing time comparison on YaleB Frontal dataset with only Lighting variation. 38 persons with images at 64 different lighting conditions are divided into mutually exclusive training and testing parts. Four sets of experiments are performed with randomly chosen {5,10,15,20} lighting conditions for training and the rest for testing. Test time is for 100 test cases averaged over 20 random partitions.

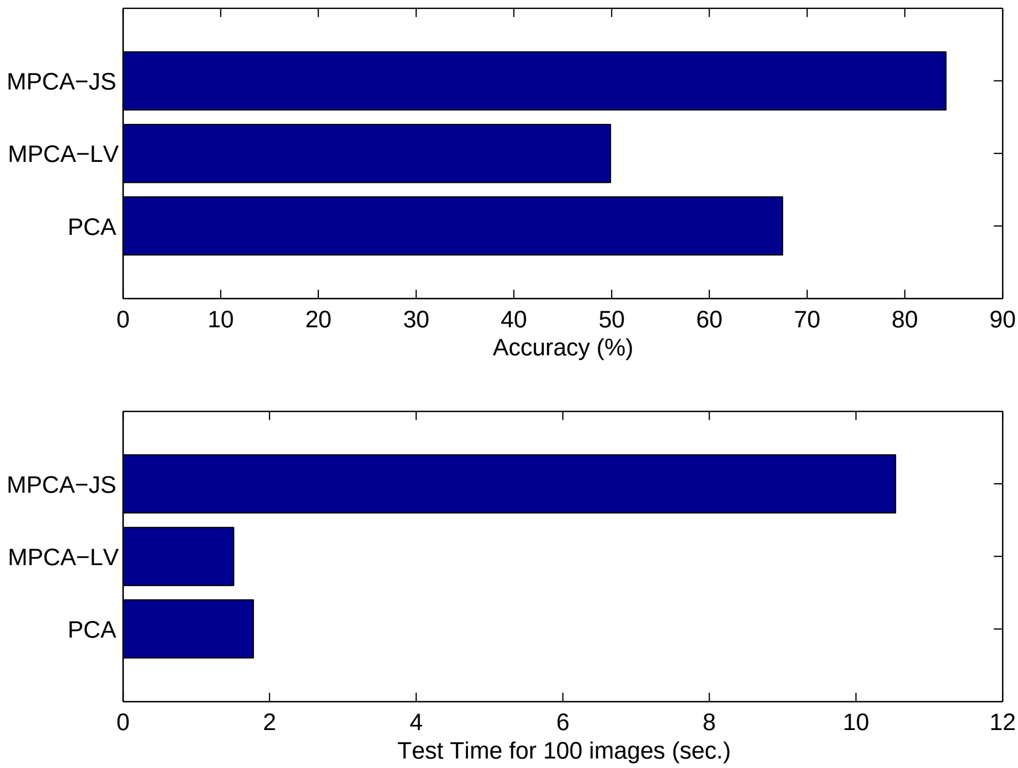

Figure 4.5: Comparative results on Extended YaleB Lighting+Viewpoint variation dataset. 38 persons with images at 64 different lighting and 9 different viewpoints is divided into mutually exclusive training and testing parts with training set containing images of all persons at 16 representative lighting conditions and at 5 representative viewpoints, and the rest constituting the testing set. The test time is the time taken to classify the first 100 test images.

4.5.1 Observations

As can be seen from the Figures 4.1–4.5, in terms of recognition accuracy, our proposed approach always outperforms the existing approach, MPCA-LV . Also the recognition accuracy for our approach is considerably higher than the PCA based recognition approach. MPCA-LV, whilst performing better than the PCA on the PEAL and the YaleB frontal datasets, performs worse in the experiment on the Extended YaleB dataset. This proves our claim that the full power of multilinear analysis is not thoroughly utilized in the MPCA-LV. However, as MPCA-JS utilizes all the factor spaces and their decompositions effectively, its performance is always better than the PCA based method.

In terms of test time, MPCA-JS is always slower than the PCA, and faster than MPCA-LV except when using the Extended YaleB dataset. The speed of MPCA-JS essentially depends on the number of comparisons needed to be performed on the multilinear eigenmode space, which in turn depends on the number of training images. As the number of training images in Extended YaleB is high, the speed is slower. However, for small datasets it is faster than the MPCA-LV algorithm and even if it is slower in the Extended YaleB dataset, in terms of recognition accuracy it provides the best result.

Overall, we can say that our approach, MPCA-JS, will in most cases be able to provide higher recognition accuracy compared to the PCA and the existing multilinear analysis based approach, MPCA-LV (Vasilescu and Terzopoulos, 2002a). The MPCA-LV shows inconsistent performance compared to PCA and is not guaranteed to provide benefits over PCA in all situations. However, this inconsistency is not the result of any drawback related to the multilinear analysis as we see that MPCA-JS, based on the same analysis framework, is able to achieve consistently higher performance than the PCA based method. In terms of testing speed, MPCA-JS is slower than the PCA, however, the difference of speed is not observed to be excessive.

4.6 Closing Remarks

In this chapter we propose a novel face recognition approach based on the multilinear analysis of face images that is capable of handling unseen variations in test images by exploiting the interaction of subspaces in an effective manner. Experimental results on the publicly available datasets show the superiority of our proposed recognition approach over the conventional multilinear recognition approach. Whilst in some situations the existing approach performs worse than the PCA, our proposed approach consistently outperforms it. Though the recognition performance provides encouraging results, we also appreciate the fact that there is scope for improvement once we take into account the other properties of core tensor and further exploit the factor spaces. In the next chapter, we propose an approach that improves upon testing time considerably by effectively exploiting the independence property of the factor spaces, whilst keeping the recognition performance intact. However, this chapter retains its fundamental importance by providing significant insights into the intricacies of the multilinear analysis of face images ensembles.

Chapter 5

Efficient Tensor Based Face Recognition

5.1 Introduction

In the previous chapter, we proposed a multilinear analysis based face recognition approach, MPCA-JS, observed to be better at handling images at unseen modes compared to the existing multilinear analysis based approach of (Vasilescu and Terzopoulos, 2002a) and the PCA based Eigenface approach. The central idea was to fully utilize the factorization of the face image tensor by exploiting the information in the core tensor. The core tensor serves as an important entity by playing the role of a connection tensor between different factor subspaces and, therefore, in the previous chapter we argued that the core tensor needs to be effectively utilized to fully realize the benefit of multilinear analysis. Multilinear analysis, in comparison to linear analysis, is able to unearth more complex structures (multilinear manifold vs. linear manifold) of the face space, therefore as a minimum, any multilinear analysis based recognition approach should perform better than linear analysis techniques such as PCA. However, that was not the case with the existing tensor based method. The MPCA-JS algorithm is developed on the concepts of *multilinear eigenmodes*, obtained from the core tensor. The *multilinear eigenmodes* are the result of multilinear analysis encompassing all the modes: persons, lighting, viewpoints etc., and the proposed recognition algorithm MPCA-JS is based on using the set of *multilinear eigenmodes* as the basis for representation and comparison. The functioning of the algorithm is similar to the standard Eigenface algorithm of (Turk and Pentland, 1991b), in which test images are compared to the training images on a suitable projection basis, which is the set of *multilinear eigenmodes* for MPCA-JS, and the closest matching training image is sought. Preliminary experiments on a few benchmark datasets demonstrated the superiority of the MPCA-JS approach against both the PCA and the existing approach (Vasilescu and Terzopoulos, 2002a), which is termed MPCA-LV. However, further investigation of the MPCA-JS approach reveals the following weaknesses,

- Though the space of *multilinear eigenmodes* provides a general framework for representation, it is inefficient for testing on a large database as the testing involves comparison with the coefficients vectors of all the training images. This results in a testing complexity of $O(N_T)$, where N_T is the total number of training images.
- When there are multiple modes of variation with very few training instances in each of them, the MPCA-JS algorithm encounters difficulties dealing with the sparseness of training data in the high-dimensional *multilinear eigenmode* space (the size of the space increases in a multiplicative way with the number of modes). This leads to poor generalisation by the classifier and may result in poor performance.

Based on these observations, we propose a new recognition approach to overcome the shortcomings outlined above. More precisely, we define *person-specific eigenmodes* that capture the variation across lightings and viewpoints for a particular person. These *person-specific eigenmodes* are used as components of a person-specific basis. A set of bases is defined with one particular basis corresponding to one particular person. Each of these bases spans a limited subspace of the whole image space, therefore, projections on these bases results in loss of information. However, projecting an image onto the same person-specific basis the image belongs to results in the smallest loss of information. We use the reconstruction error to quantify the loss of information. The basis, which gives the lowest reconstruction error, reveals the identity of the test image. The proposed method can be thought of as a compartmentalization of the general *multilinear eigenmode* space corresponding to different persons. Testing involves projection to N_p (number of persons in the database) number of bases and comparison of N_p number of reconstruction errors, resulting in a testing complexity of only $O(N_p)$. As $N_p << N_T$, we overcome the inefficiency of MPCA-JS. Moreover, as the individual *person-specific eigenmode* spaces are much smaller than the space of *multilinear eigenmodes*, we overcome the generalization problem as well.

The main contributions of this work are:

1. Developing a novel algorithm based on the same multilinear analysis based approach as the previous method MPCA-JS, however, with significantly lower computational complexity. The complexity of the new algorithm is $O(N_p)$ (where N_p is the number of persons), which is several times lower than the complexity of MPCA-JS which has a complexity of $O(N_T)$, where N_T is the total number of images in the dataset. Most importantly the reduction in complexity order is not observed to significantly affect the recognition rate.
2. The new algorithm is also potentially superior to MPCA-JS as the segregation of multilinear eigenmodes from the whole *multilinear eigenmodes* provides much smaller subspaces of *person-specific eigenmodes* compared to the higher dimensional subspace of *multilinear eigenmodes*. This can potentially alleviate the generalization problem as faced by MPCA-JS in certain scenarios.

The chapter is organized as follows: in section 5.2 we briefly recount the tensor model of face image ensembles and the Multilinear PCA algorithm, followed by a brief overview of the MPCA-JS in section 5.3. The proposed recognition approach is developed in section 5.4; in section 5.5, we evaluate our algorithm against the performance of MPCA-JS on a few benchmark datasets, and finally in section 5.6, we conclude the chapter.

5.2 Review: Tensor Model of Face Image Ensembles

We follow the tensor model as followed in the previous chapter and as described in the section 3.3, where the face image tensor $\mathcal{T}$ is given by,

$$\mathcal{T}_{(i_p,i_l,i_v)} = I_{P_{i_p},L_{i_l},V_{i_v}} \tag{5.2.1}$$

with the usual notation of $I_{P_{i_p},L_{i_l},V_{i_v}}$ being the image vector of i_p'th person at i_l'th lighting and i_v'th viewpoint. The Multilinear PCA decomposition of $\mathcal{T}$ is similarly followed from there and is represented as:

$$\mathcal{T} = \mathcal{S} \times_1 U^P \times_2 U^L \times_3 U^V \times_4 U^X \tag{5.2.2}$$

with the usual definition of $\mathcal{S}, U^P, U^L$, and U^V being followed. For a face image tensor $\mathcal{T}$ with additional modes, suitable extension of terminology can be achieved.

5.3 Review: MPCA-JS

MPCA-JS, as described in the previous chapter, uses *multilinear eigenmodes* as the basis for face image representation and recognition. From (5.2.2) let us define,

$$\mathcal{A} = \mathcal{S} \times_4 U^X \tag{5.3.1}$$

and if A_{person} denotes the unfolding of the tensor $\mathcal{A}$ in the person mode, then

$$A_{person} = \begin{bmatrix} I_{P_1^e L_1^e V_1^e} & .. & I_{P_1^e L_{N_l'}^e V_{N_v'}^e} \\ I_{P_2^e L_1^e V_1^e} & .. & .. \\ \ldots & \ldots & \ldots \\ I_{P_{N_p'}^e L_1^e V_1^e} & .. & I_{P_{N_p'}^e L_{N_l'}^e V_{N_v'}^e} \end{bmatrix} \tag{5.3.2}$$

where, $I_{P_i^e L_j^e V_k^e}$ is one specific *multilinear eigenmode*. These *multilinear eigenmodes* are used as the components for the projection basis. The coefficient vectors of projection for the training images are calculated separately from U^P, U^L, U^V, and stored for testing. Testing involves projecting the test image on the basis and comparing that with the stored coefficient vectors to find out the nearest training image, resulting in a testing complexity of $O(N_p \times N_l \times N_v)$.

5.4 Proposed Approach

Motivated by our desire to address the weaknesses of MPCA-JS, we present a new algorithm utilising *person-specific eigenmodes* as the basis for projection. We derive our approach starting from the multilinear PCA decomposition of face tensor $\mathcal{T}$, as in (5.2.2) and then perform unfolding and matricization on the tensor, $\mathcal{T}$ to define *person-specific eigenmodes*. The formulation is presented below.

Let us denote,

$$\mathcal{B} = \mathcal{S} \times_1 U^P \tag{5.4.1}$$

and if B_{person} denotes the unfolding of the tensor $\mathcal{B}$ in the person mode, then

$$B_{person} = \begin{bmatrix} I^e_{P_1 L^e_1 V^e_1} & \cdots & \cdots & I^e_{P_1 L^e_{N'_l} V^e_{N'_v}} \\ I^e_{P_2 L^e_1 V^e_1} & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ I^e_{P_{N_p} L^e_1 V^e_1} & .. & \cdots & I^e_{P^e_{N_p} L^e_{N'_l} V^e_{N'_v}} \end{bmatrix} \tag{5.4.2}$$

where $I^e_{P_{i_p} L^e_{i_l} V^e_{i_v}}$ is the *eigen-image* of the i_p'th person at i_l'th *eigen-lighting* and at i_v'th *eigen-viewpoint*. Now, let us define:

$$\tilde{B}_k = \begin{bmatrix} I^e_{P_k L^e_1 V^e_1} \\ I^e_{P_k L^e_2 V^e_1} \\ \cdots \\ I^e_{P_k L^e_{N'_l} V^e_{N'_v}} \end{bmatrix} \tag{5.4.3}$$

where, $\tilde{B}_k$ for $k = 1, ..., N_p$ are the person-specific bases. For a test image I_T, the projection on the basis $\tilde{B}_K$ will generate a description of how the images of a person with identity K at different *eigen-lighting* and *eigen-viewpoint* are combined to create the image. $\tilde{B}_K$ spans only a limited subspace, corresponding to the images of K'th person out of the whole image space. Therefore, the projections on $\tilde{B}_K$ can only preserve a certain fraction of information of the image I_T. However, if I_T belongs to person K, then a projection on $\tilde{B}_K$ will preserve the maximum information of the image, out of all $\{\tilde{B}_k\}_{k=1}^{N_p}$. We will use the minimum reconstruction error as the criterion to quantify the information preserved in a certain person-specific base. Given a test image, we will project it on all the *person-specific bases* and measure the reconstruction errors via back projection to the image space. The *person-specific basis* corresponding to the minimum reconstruction error denotes the identity of the test image. As the algorithm is based on *person-specific bases*, we denote it as MPCA-PS. The algorithm for testing is presented in Algorithm 3.

Algorithm 3: Testing algorithm for MPCA-PS

Input: Test image I_T, and $\{\tilde{B}_k\}_{k=1}^{N_p}$.

Output: Identity of the test image.

1. For every $k = 1, ..., N_p$ compute
   - $c_T^k = I_T \times \tilde{B}_k^+$
   - $e^k = ||I_T - c_T^k \times \tilde{B}_k||$

   The distance measure we use is Euclidean distance.

2. If e^{k_m} is the smallest of $\{e_k\}_{k=1}^{N_p}$ then the identity of I_T is k_m.

5.4.1 Complexity Analysis

In terms of computational complexity, we have N_p number of bases, resulting in N_p matrix multiplications, and N_p distance calculations. Hence, our proposed method is roughly of the order $O(N_p)$. Compared to that, MPCA-JS has only one matrix multiplication, and $(N_p \times N_l \times N_v)$ distance calculations, resulting in an algorithm of the order $O(N_p \times N_l \times N_v)$.

5.4.2 Remarks

In the previous chapter we proposed a general multilinear space encompassing all the factors *i.e.* person, lighting, viewpoint etc. Though the space of *multilinear eigenmodes* provides a general framework for representation, it is inefficient for testing on large databases due to the testing complexity being of the order $O(N_T)$, where $N_T = N_p \times N_l \times N_v$ is the number of training images. The proposed approach can be thought of as a compartmentalization of the general multilinear space corresponding to the person identities, resulting in testing complexity of $O(N_p)$. As $N_p << N_T$, we overcome the inefficiency of MPCA-JS, whilst experiments demonstrate that the recognition performance remains intact. Moreover, as each of the *person-specific bases* are N_p times smaller than the original *multilinear eigenmodes* space, the generalization problem as faced by MPCA-JS while dealing with the *multilinear eigenmodes* is reduced in this case.

5.5 Experiments, Analysis and Evaluation

In this section, we evaluate our proposed recognition approach MPCA-PS, primarily against the recognition approach developed in the previous chapter, MPCA-JS and also with the PCA based Eigenface algorithm. Experiments are conducted on a number of publicly available benchmark datasets. Specifically, we used PEAL (Gao *et al.*, 2008), YaleB Frontal (Georghiades *et al.*, 2001a) and Extended YaleB (Lee *et al.*, 2005b) face databases. The datasets are the same as described in Section 4.5. Additionally, we used the Weizmann face database, having face images of 28 persons at 5 different viewpoints $(0°, \pm 17°, \pm 34°)$, 3 different lighting conditions and 2 different expressions. The image size for the Weizmann dataset is $[64 \times 44]$. Prior to the experiments, all the images were cropped, their eye-points aligned manually, then all the image vectors were normalized to unity. For HOSVD and other tensor operations, we used the tensor toolbox developed by Bader and Kolda (Bader and Kolda, 2007).

For the PEAL dataset, four sets of experiments are performed with randomly chosen 5, 7, 9 and 11 lighting conditions for training. and the rest for testing. Similarly, for the YaleB Frontal dataset, four sets of experiments are performed with randomly chosen 5, 10, 15 and 20 lighting conditions for training with the rest being used for testing. For the experiments on the Extended YaleB database, 16 representative lighting conditions at 5 representative viewpoints are used for training and the rest for testing. For the experiments on Weizmann face database, 3 representative viewpoints (at $0°$ and $\pm 34°$) at 2 randomly selected lighting conditions and all the expressions are used for training and the rest for testing. Thus, in both the PEAL and YaleB Frontal database, the test images are at unseen lighting conditions. For both Extended YaleB database and Weizmann database, the test images are either at unseen lighting conditions, or at unseen viewpoints, or at both unseen lighting conditions and viewpoints. For the experiments on PEAL and YaleB Frontal, each set of experiments is repeated 20 times and the average results are reported. As mentioned before, the performance of the proposed recognition procedure is compared with PCA based Eigenface (Turk and Pentland, 1991b) and the previously proposed multilinear analysis based algorithm, MPCA-JS.

We use the same *energy thresholding*, as discussed in the previous chapter to select *eigenvectors* in different modes: 0.96 for pixel modes and 0.99 for all other modes for multilinear PCA, and 0.96 for PCA. The distance measures are also the same: PCA uses Euclidean and MPCA-JS uses Cosine. All the experiments were performed in the Matlab environment, running on a Intel Xeon 8×2.3GHz server with 16GB RAM and without code optimization.

Experimental results on YaleB Frontal dataset are presented in Figures 5.1 and 5.2, showing the recognition accuracy and test time respectively. Similarly, Figures 5.3 and 5.4 presents the recognition accuracy and test time for experiments on the PEAL lighting variation dataset. Figure 5.5 presents the results of experiments on the Extended YaleB dataset, and finally Figure 5.6 presents the recognition results for Weizmann dataset.

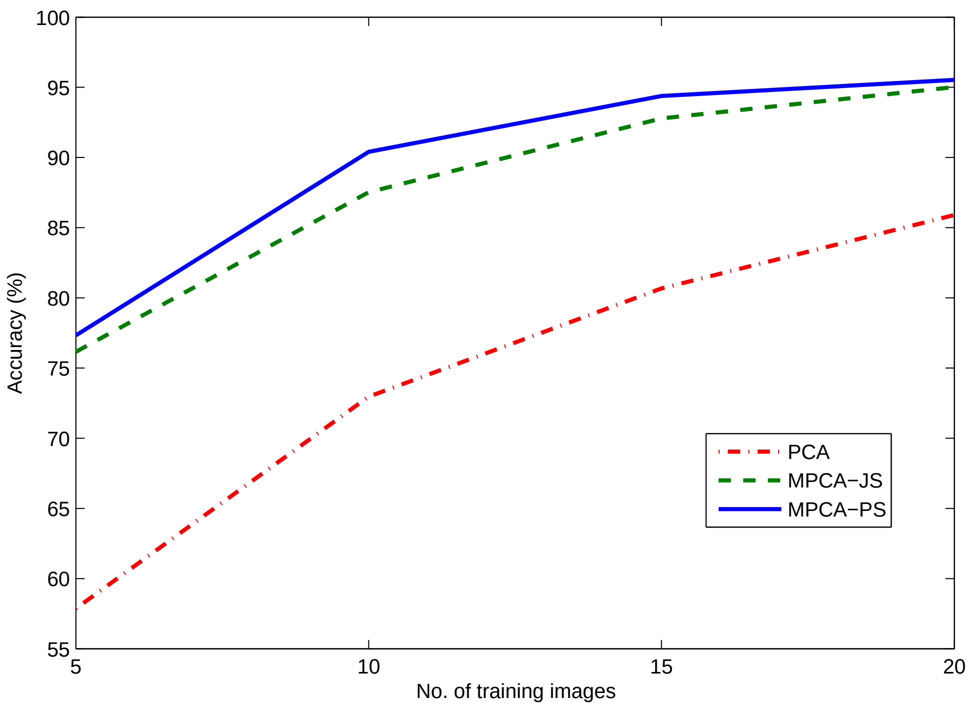

Figure 5.1: Recognition accuracy comparison on YaleB Frontal Lighting variation dataset. 38 persons with images at 64 different lighting conditions are divided into mutually exclusive training and testing parts. Four sets of experiments are performed with randomly chosen {5,10,15,20} lighting conditions for training and the rest for testing. Recognition accuracy is the average classification accuracy over 20 random partitions.

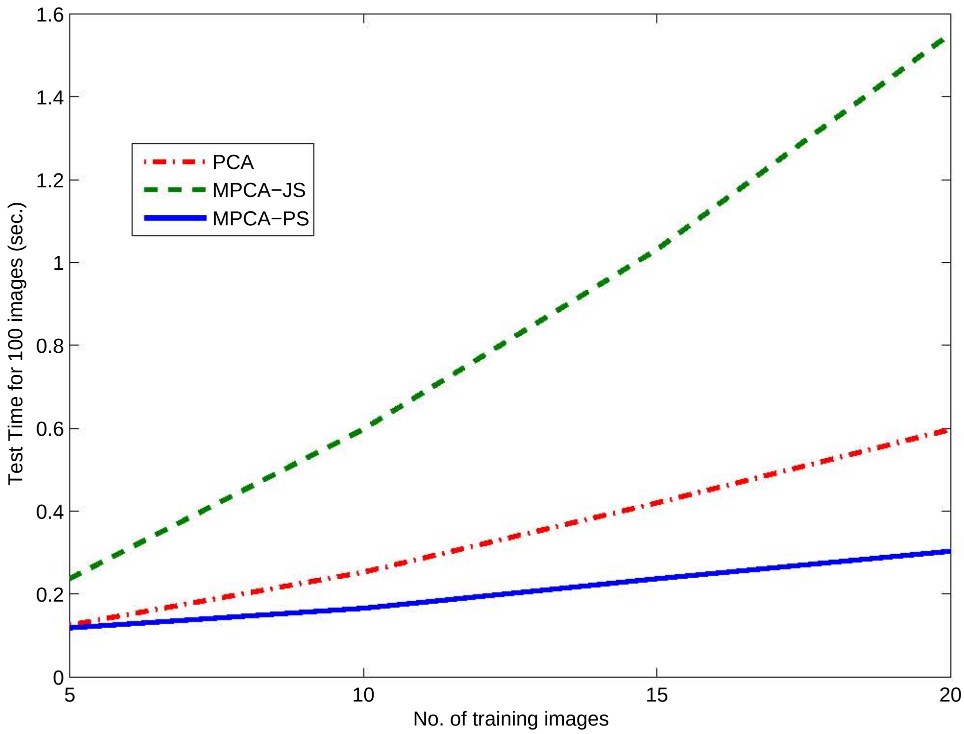

Figure 5.2: Testing time comparison on YaleB Frontal dataset with only Lighting variation. 38 persons with images at 64 different lighting conditions are divided into mutually exclusive training and testing parts. Four sets of experiments are performed with randomly chosen {5,10,15,20} lighting conditions for training and the rest for testing. Test time is for 100 test cases averaged over 20 random partitions.

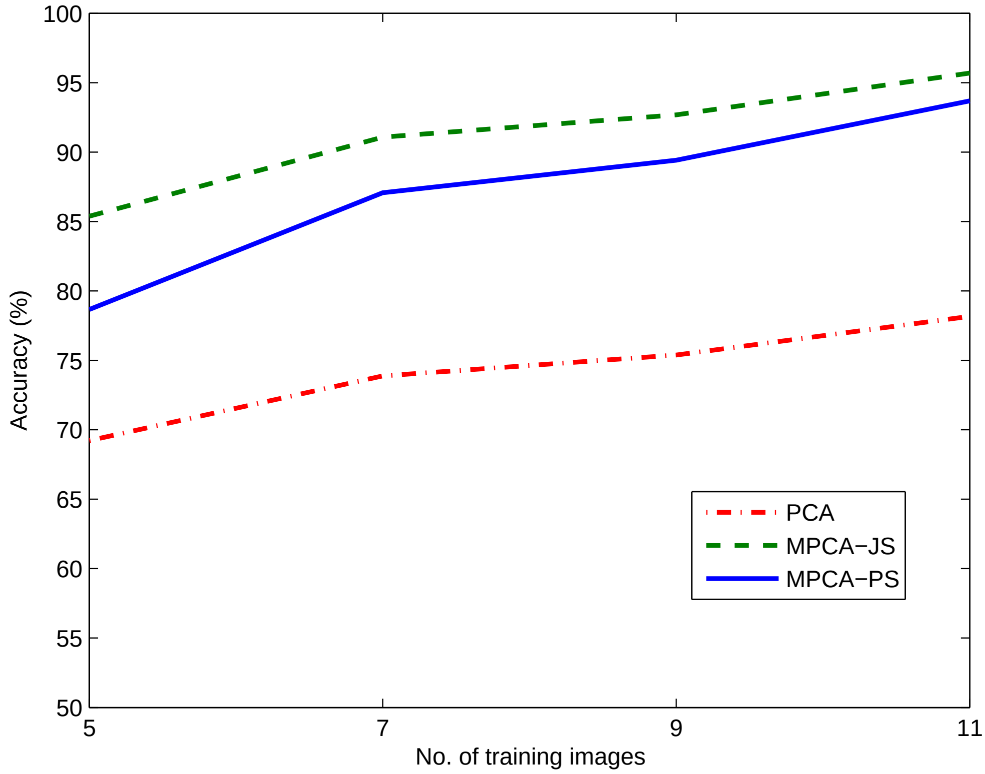

Figure 5.3: Recognition accuracy comparison on PEAL Lighting variation dataset. 20 persons with images at 20 different lighting conditions are divided into mutually exclusive training and testing parts. Four sets of experiments are performed with randomly chosen {5,7, 9, 11} lighting conditions for training and the rest for testing. Recognition accuracy is the average classification accuracy over 20 random partitions.

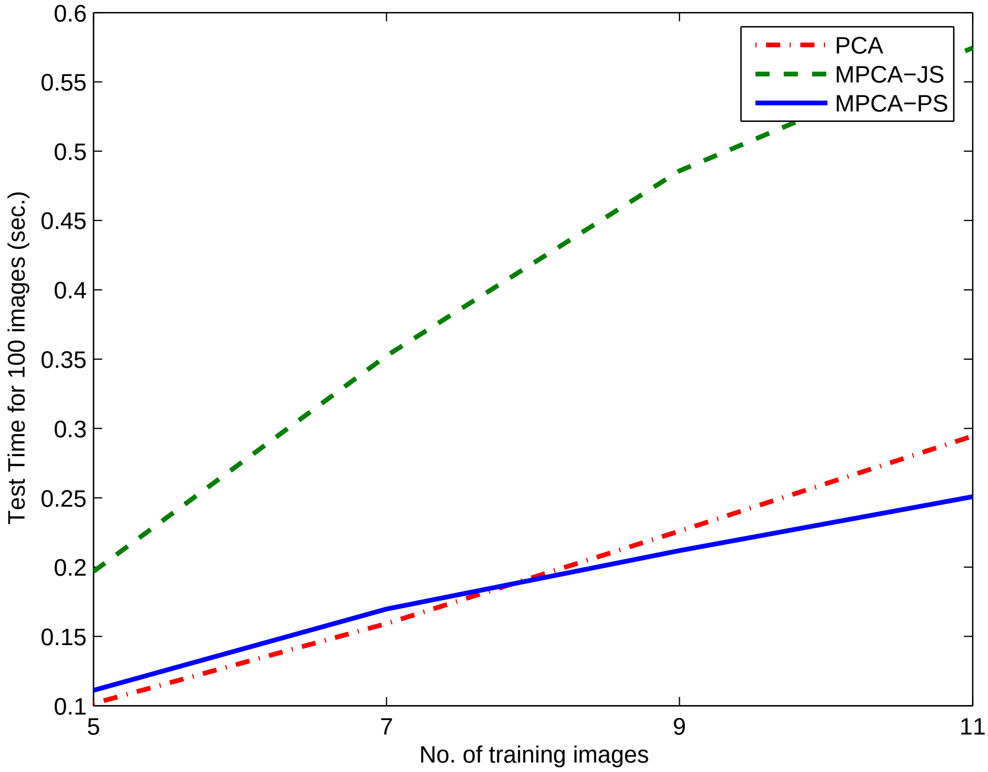

Figure 5.4: Testing time comparison on PEAL Lighting variation dataset. 20 persons with images at 20 different lighting conditions are divided into mutually exclusive training and testing parts. Four sets of experiments are performed with randomly chosen {5,10,15,20} lighting conditions for training and the rest for testing. Test time is for 100 test cases averaged over 20 random partitions.

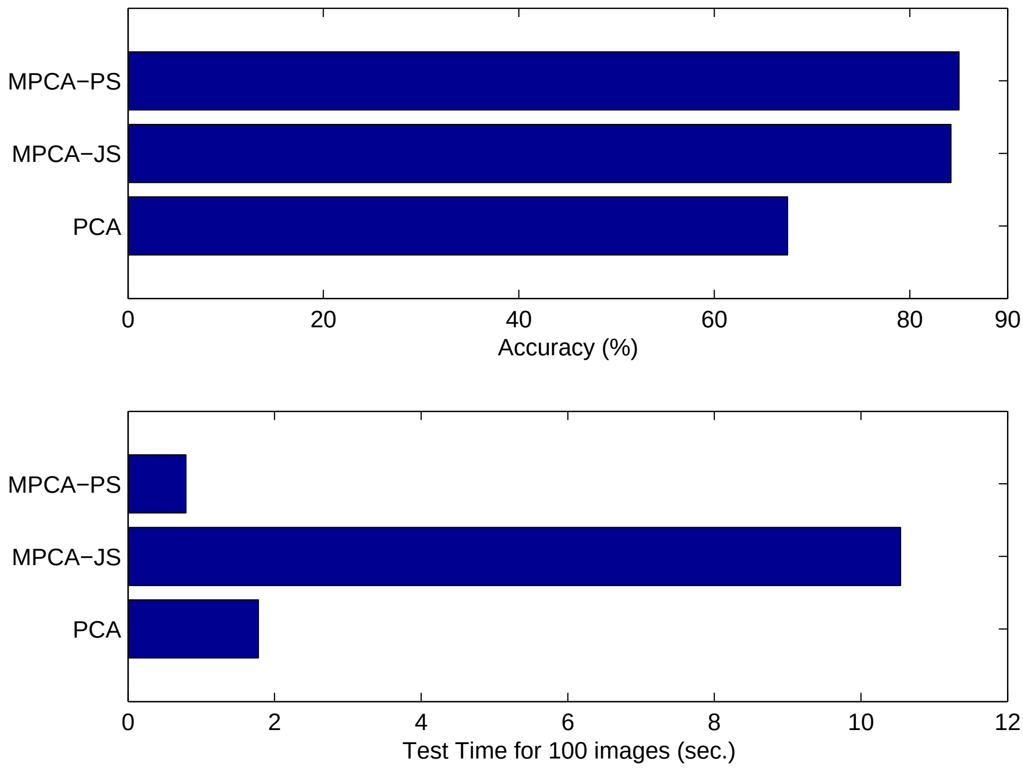

Figure 5.5: Comparative results on Extended YaleB Lighting+Viewpoint variation dataset. 38 persons with images at 64 different lighting and 9 different viewpoints is divided into mutually exclusive training and testing parts with training set containing images of all persons at 16 representative lighting conditions and at 5 representative viewpoints, and the rest constituting the testing set. The test time is the time taken to classify 100 test images.

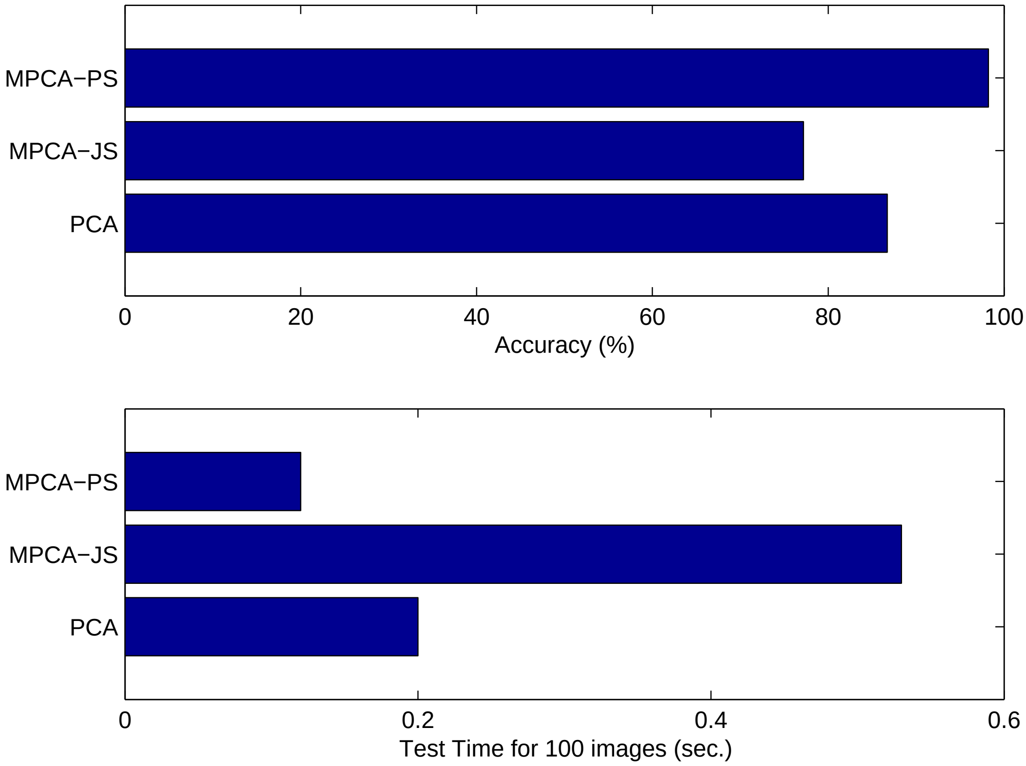

Figure 5.6: Comparative results on Weizmann Lighting+Viewpoint+Expression variation dataset. 28 persons with images at 3 different lighting, 5 different viewpoints and 2 different expressions is divided into mutually exclusive training and testing parts with training set containing images of all persons at 2 representative lighting conditions, 3 representing viewpoints and at all 2 expressions, and the rest constitute the testing set. The test time is the time taken to classify 100 test images.

5.5.1 Observations

From the experimental results as shown in the Figures 5.1-5.6, we can make the following observations:

- In the YaleB Frontal dataset (Figures 5.1) and 5.2, the proposed MPCA-PS approach provides a superior performance when compared with PCA and MPCA-JS both in terms of recognition accuracy and speed. Whilst the recognition accuracy is only slightly higher, the speed is around 8 times better than the MPCA-JS approach.

- Similarly, in the Extended YaleB dataset (Figure 5.5), the performance is superior to both the PCA and MPCA-JS in terms of both recognition accuracy and speed. MPCA-PS is more than 20 times faster than MPCA-JS.

- However, in the PEAL dataset (Figures 5.3 and 5.4) MPCA-PS shows a lower recognition rate than the MPCA-JS, but provides a better recognition rate than PCA. In terms of testing speed, MPCA-PS is still the fastest. The poor recognition accuracy of MPCA-PS in this dataset can be attributed to the fact that the database is of Chinese nationals, who share a great deal of similarity in appearance, thereby, there is much overlap among the *person-specific eigenmode* spaces introducing errors in the classification.

- The result observed on the Weizmann dataset (Figure 5.6) is an interesting one. In this experiment, MPCA-PS shows near perfect classification performance, whilst the recognition performance of MPCA-JS is considerably worse than the PCA. This is undesirable since a recognition method based on multilinear analysis should be able to perform better than the linear approach of PCA. The reason can be attributed to the fact that when the number of modes are high with only a few samples in each mode, the resultant *multilinear eigenmodes* becomes large (as the dimension of the space increases in the multiplicative way with the numbers of modes) and sparse (as only a few images are available) resulting in poor generalization of the NN classifier. However, the MPCA-PS classification scheme is not directly dependent on the number of training samples and thus is insensitive to such issues. Also, we note that in these experiments, the MPCA-PS is able to demonstrate its speed advantage by being faster than both the MPCA-JS and PCA.

Overall, we observe that the MPCA-PS algorithm is able to provide the fastest classification with recognition performance consistently better than PCA and similar to that of MPCA-JS.

5.6 Closing Remarks

In this chapter we propose a new multilinear analysis based face recognition approach, MPCA-PS, that demonstrates huge speed advantages in testing when compared to the algorithm developed in the last chapter, MPCA-JS. The increase in efficiency is obtained through effectively manoeuvring the decomposition products, whose physical interpretation is similar to compartmentalizing the *multilinear eigenmode* space corresponding to person identities. The testing requires projecting the test image onto each *person-specific basis* and then seeking the basis that is the most representative for the test image. In experiments, we found the speed advantage of MPCA-PS is significant over the MPCA-JS approach when compared on large datasets. In terms of recognition rate, MPCA-PS is always better than the PCA and similar to that of MPCA-JS. However, we note that the experimental results provide only the indicative performance profiles for these two algorithms, and such empirical comparisons may be of limited use when choosing an algorithm to be used in a database which is qualitatively different from the experimented databases. To address this concern, in the next chapter we present a comparison based on the intrinsic qualities of these algorithms in a unified theoretical framework.

Chapter 6

A Unified Tensor Framework for Face Recognition

6.1 Introduction

In the previous two chapters we have developed and discussed two different tensor based algorithms for face image recognition, namely MPCA-JS and MPCA-PS, and the existing work of (Vasilescu and Terzopoulos, 2002a) that we term MPCA-LV. In terms of performance, we observe that MPCA-LV is not suitable when test faces are in unseen modes, a limitation which was effectively addressed by MPCA-JS. Further, we observe that MPCA-JS is costly as it has a high complexity order and we propose a new approach, MPCA-PS, that addresses the efficiency issue whilst maintaining similar recognition performance. All of these approaches are based on the same Multilinear PCA decomposition, however, they differ widely in their development. We developed our previous approaches mostly through mechanical operations, including tensor unfolding, matrix reorganization etc., however, the development lacked a comprehensive mathematical formalism. The descriptions through *eigenmodes* and its variants are, at best, intuitive in nature and used mainly for justifying the approach rather than for the analysis of the solutions. For example, MPCA-PS algorithm offers comparable performance to MPCA-JS, but at a significantly low computational cost. However, it is not quite obvious as to why and how much impact it has on the quality of solution and overall recognition accuracy as this increase in efficiency is achieved. The only means of understanding is through a comparison that is totally empirical, and has the associated problem of dealing with inconsistent order of performance in different scenarios. Therefore, we aim to develop a framework that will enable us to perform explicit and non-empirical comparisons between different tensor based face recognition algorithms, solely based on their quality of solutions whilst enabling us to predict performance in general scenarios.

Tensor analysis through any form of multilinear decomposition is basically factor analysis with different assumptions on the relations of the factors. For face images, tensor analysis provides an ability to express a face image in terms of its constituent factors (*i.e.* person, lighting, viewpoint etc.), while the relation between these factors is governed by the choice of specific multilinear decompositions (*eg* Multilinear PCA/ Multilinear ICA etc.) used. Therefore, the basic principle of any tensor based approach can be best understood in terms of the way it utilizes different factors. In this chapter, we build a framework based on the factor analysis principles and re-formulate all tensor based face recognition methods in this framework to obtain a clearer and compact interpretation.

We start by developing an optimization problem relating a face image to its constituent factors (*eg.* person, lighting, viewpoint etc.). This stems directly from the way face images are modelled through a tensor and its decomposition. From factor analysis principles, we know that all factors are independent and thus invariant to the changes of other factors. The property of invariance between different factors implies that all the images of a person have the same value for its person-factor irrespective of the lighting or viewpoint conditions. Similarly, this notion carries over to all the other factors. This independence property of factors is useful in the sense that if we can obtain the person-factor value or person-space representation of an image, we can readily identify the image by matching it to the stored person-space representations of the known persons. The optimization problem can be solved directly using the Alternating Least Squares (ALS) method and the test image is factorized to obtain the person-factor value that can be used for recognition. However, in practice ALS is extremely slow in convergence and may fail to converge if the factor spaces are high-dimensional or highly correlated, as noted in (Lathauwer *et al.*, 2000). As there are no better ways to solve such optimization problems, we investigate alternative ways based on our understanding of the specific multilinear structure arising from the Multilinear PCA decomposition of the face image tensor and the domain knowledge of face image and recognition problem. In order to do so, we first develop a general framework that offers two key advantages:

1. It allows different assumptions about factors to be explicitly articulated and incorporated in precise mathematical form.

2. It enables the development of different strategies for optimization that exploit specific factor spaces and their relationships.

Within these parameters, we discover that the existing methods of MPCA-LV, MPCA-JS and MPCA-PS solve the same optimization problem by adopting different assumptions about the factors and employing different strategies on the factor spaces and their relations. We thoroughly describe the direct approach as well as alternative derivations of MPCA-LV, MPCA-JS, and MPCA-PS in a single framework. Comparative analysis of these assumptions and strategies reveals the quality of solution offered by these methods and provides the desired measure of comparison. Not only does this framework provide a theoretical mode of comparison, it also helps us to analyse failure modes for different recognition scenarios, and predicts their performance in any scenario, which is critical when considered from an application point of view.

The outline of this chapter is as follows: in section 6.2 we review the tensor model for multilinear face image analysis, while in section 6.3 we elaborate our framework and formulate four different face recognition methods, followed by a theoretical comparison and an analysis of the failure modes. Section 6.4 contains experimental results and detailed analysis of the results based on the proposed framework, and section 6.5 concludes the chapter.

6.2 Review: Tensor Model of Face Image Ensembles

We follow the tensor model as followed in the previous chapter and as described in Section 3.3, where the face image tensor $\mathcal{T}$ is given by,

$$\mathcal{T}_{(i_p,i_l,i_v)} = I_{P_{i_p},L_{i_l},V_{i_v}} \tag{6.2.1}$$

with the usual notation of $I_{P_{i_p},L_{i_l},V_{i_v}}$ being the image vector of i_p'th person at i_l'th lighting and i_v'th viewpoint. The Multilinear PCA decomposition of $\mathcal{T}$ is similarly followed from there and is represented as:

$$\mathcal{T} = \mathcal{S} \times_1 U^P \times_2 U^L \times_3 U^V \times_4 U^X \tag{6.2.2}$$

with the usual definition of $\mathcal{S}, U^P, U^L$, and U^V being followed. For a face image tensor $\mathcal{T}$ with additional modes, suitable extension of terminology can be achieved.

6.3 Proposed Recognition Framework

From (6.2.2), if $u_p(\in \mathbb{R}^{N'_p}), u_l(\in \mathbb{R}^{N'_l})$ and $u_v(\in \mathbb{R}^{N'_v})$ are respectively the person space, lighting space and viewpoint space projections for a test image I_T, we have,

$$I_T = \mathcal{S} \times_1 u_p \times_2 u_l \times_3 u_v \times_4 U^X \tag{6.3.1}$$

To represent the test image in the tensor framework, we need to compute u_p, u_l and u_v for I_T. For this purpose, we can formulate a multilinear least squares problem as:

$$\min_{u_p, u_l, u_v} ||I_T - \mathcal{S} \times_1 u_p \times_2 u_l \times_3 u_v \times_4 U^X||_2 \tag{6.3.2}$$

6.3.1 Approach 1: MPCA-ML

We have mentioned earlier that each person in the training database has a unique representation in the person space. Therefore, the most direct way to perform recognition is to compare the person space projection of the test image I_T with the person space representation of all the training persons. We can solve the multilinear optimization problem (6.3.2) directly using the Alternating Least Squares method (ALS) (Kroonenberg and Leeuw, 1980). In this method u_p, u_l and u_v are initialized to some random values and then the equation (6.3.2) is alternatively optimized over a single variable while keeping others fixed. After a sufficient number of iterations, we hope to achieve the convergence, indicated by minute difference in the consecutive estimates of the values of u_p, u_l and u_v. The algorithm is terminated when all the differences are below a certain threshold or a certain number of iterations have reached. The recognition algorithm is presented in Algorithm 4.

Algorithm 4: Testing algorithm for the MPCA-ML

Input: Test image I_T and $\mathcal{S}, U^X$

Output: Identity of the test image.

1. ALS method: initialize $u_p \in \mathbb{R}^{N'_p}, u_l \in \mathbb{R}^{N'_l}$ and $u_v \in \mathbb{R}^{N'_v}$ to some random values.

   - Step 1: Compute ${u_p}^{new}$ as,

   $$ {u_p}^{new} = I_T \times ((\mathcal{S} \times_2 u_l \times_3 u_v \times_4 U^X)_{(person)})^+ $$

   - Step 2: Compute ${u_l}^{new}$ as,

   $$ {u_l}^{new} = I_T \times ((\mathcal{S} \times_1 {u_p}^{new} \times_3 u_v \times_4 U^X)_{(lighting)})^+ $$

   - Step 3: Compute ${u_v}^{new}$ as,

   $$ {u_v}^{new} = I_T \times ((\mathcal{S} \times_1 {u_p}^{new} \times_2 {u_l}^{new} \times_4 U^X)_{(viewpoint)})^+ $$

   - Step 4: Check for convergence and set $u_p = {u_p}^{new}$, $u_l = {u_l}^{new}$ and $u_v = {u_v}^{new}$.

   - Step 5: Go to Step 1, if convergence of u_p, u_l and u_v is not achieved or a preset number of iteration has been reached.

2. If $\{c_{i_p}\}_{i_p=1}^{N_p}$ are the rows of U^P and if i_{p_m} minimizes

   $$ \min_{i_p} ||u_p - c_{i_p}|| $$

   then identity of I_T is i_{p_m}.

The major advantage of this solution is that it is easy to incorporate any prior knowledge of lighting or pose of the test image by initializing or fixing u_l and u_v to certain values. Also, we obtain all the factor-space projection of the test image, enabling us to estimate the lighting or pose of the face in the same framework. However, as discussed earlier, ALS suffers from poor convergence rates and may fail to converge in case of high-dimensional factor spaces. This results in an extremely expensive method coupled with poor performance for large databases.

6.3.2 Approach 2: MPCA-LV

As seen in the previous algorithm, even though a direct multilinear solution using ALS offers a simple approach to obtain a solution, the computation requirement can be high even for small databases. Also, we estimate both the lighting space and viewpoint space projection of the test image, though only the person space projection is sufficient for recognition. Considering this, the computational cost can be reduced if we avoid estimating u_l, u_v and instead solve (6.3.2) over a fixed set of $\{(u_l, u_v)\}$. Let (u_l^*, u_v^*) be a member from the set $\{(u_l, u_v)\}$, then equation (6.3.2) can be reformulated as:

$$\min_{u_p} ||I_T - \mathcal{S} \times_1 u_p \times_2 u_l^* \times_3 u_v^* \times_4 U^X||_2 \tag{6.3.3}$$

Following the definition of mode-k multiplication to a tensor, this can equivalently be written as:

$$\min_{u_p} ||I_T - u_p \times (\mathcal{S} \times_2 u_l^* \times_3 u_v^* \times_4 U^X)_{(person)}||_2 \tag{6.3.4}$$

Fortunately, this is a linear equation and the optimal u_p for (u_l^*, u_v^*) can be computed as:

$$u_p = I_T \times (\mathcal{S} \times_2 u_l^* \times_3 u_v^* \times_4 U^X)^+_{(person)} \tag{6.3.5}$$

where the superscript $^+$ implies *Moore-Penrose pseudo-inverse*. For each possible pairs of $\{u_l, u_v\}$, we obtain an optimal solution for u_p, resulting in a set of $\{u_p\}$. To create the set $\{u_l, u_v\}$, we assume that our training database is well-represented, implying that the lighting conditions and pose of the testing images are very close to the training lighting conditions and pose. Therefore, we fix u_l to the set of training lightings and u_v to the set of training viewpoints. Let $u_l^{k_1} = k_1$'th row of U^L and $u_v^{k_2} = k_2$'th row of U^V, then the set is created as, $\{(u_l^{k_1}, u_v^{k_2})$ for $k_1 = 1, ..., N_l$ $k_2 = 1, ..., N_v\}$. Clearly the cardinality of the set is $N_l N_v$. Using (6.3.5), we generate a set of person space representation for I_T, over the set $\{(u_l, u_v)\}$ as:

$$\begin{aligned} u_p^{k_1 k_2} &= I_T \times (\mathcal{S} \times_2 u_l^{k_1} \times_3 u_v^{k_2} \times_4 U^X)^+_{(person)} \quad \forall\, k_1, k_2 \\ &= I_T \times A_{k_1 k_2} \end{aligned} \tag{6.3.6}$$

The set of person space representations, $\{u_p^{k_1k_2}$ for $k_1 = 1, ..., N_l$ $k_2 = 1, ..., N_v\}$ is then compared pairwise to the person space representation of the training persons and the best matching person is found. The recognition algorithm is presented in Algorithm 5.

Algorithm 5: Testing algorithm for MPCA-LV

Input: Test image I_T, and $\mathcal{S}, U^X$

Output: Identity of the test image.

Precalculate: $A_{k_1k_2} = (\mathcal{S} \times_2 u_l^{k_1} \times_3 u_v^{k_2} \times_4 U^X)^+_{(person)}$

1. For $k_1 = 1, ..., N_l$ and $k_2 = 1, ..., N_v$ compute,

$$u_p^{k_1k_2} = I_T \times A_{k_1k_2}$$

2. If $\{c_{i_p}\}_{i_p=1}^{N_p}$ are the rows of U^P and if i_{p_m} minimizes

$$\min_{i_p,k_1,k_2} ||u_p^{k_1k_2} - c_{i_p}||$$

then identity of I_T is i_{p_m}.

This algorithm is an attempt to solve the optimization problem in deterministic time, though at the cost of obtaining a suboptimal solution. The suboptimal solution is near-optimal when our assumption that the test images do not have large variations compared to the training images is valid. However, if the variation is large, the estimation of u_p would be inaccurate and the recognition performance based on u_p would be poor. On the other hand, we still retain the advantage of incorporating prior knowledge in terms of manipulating u_l and u_v individually, either by fixing them to a certain set of values or excluding a certain subset of values from the existing set. It is also obvious that this approach is similar to the idea described in (Vasilescu and Terzopoulos, 2002a). However, by re-deriving this on our framework, we are able to gain more insights on how this algorithm is expected to behave in different scenarios. Specifically, we show that the inability of this approach to handle faces at unseen modes is because it is not designed to handle such conditions.

6.3.3 Approach 3: MPCA-JS

In this approach we exploit the structure of the core tensor to cast (6.3.2) as a linear least squares problem. From (6.3.2), we obtain the person space, lighting space and viewpoint space descriptions for a test image. The core tensor contains the multilinear relationship between the person space, lighting space and the viewpoint space. The person space, lighting space and viewpoint space projection vectors interact with the core tensor in the mode-specific linear way to approximate the image. Through manipulating the structure inside the core tensor, we will shift the multilinearity from the core tensor side to the description side. An image will now be described by a multilinear function of person space, lighting space and viewpoint space projection vectors and will be approximated by a linear multiplication of the description vector with a matrix derived from the core tensor. As a result, instead of multilinear optimization, we obtain a linear equation to compute the description vector for a test image, thereby saving computation cost and solving it optimally. Based on this idea, we provide the necessary mathematical derivation to develop a recognition algorithm.

Let us start by defining:

$$\mathcal{A} = \mathcal{S} \times_4 U^X \tag{6.3.7}$$

With analogy to the definition of T, we can state that,

$$\mathcal{A}_{(i_p,i_l,i_v)} = I_{P^e_{i_p} L^e_{i_l} V^e_{i_v}} \tag{6.3.8}$$

where $I_{P^e_{i_p} L^e_{i_l} V^e_{i_v}}$ is the image of i_p'th eigenperson at i_l'th eigenlighting and i_v'th eigenviewpoint. Following the usual definition of u_p, u_l and u_v for the test image I_T we can write that,

$$I_T \;=\; \mathcal{S} \times_1 u_p \times_2 u_l \times_3 u_v \times_4 U^X \tag{6.3.9}$$

And, following the definition of $\mathcal{A}$ from (6.3.7) we can rewrite the above as:

$$I_T \;=\; \mathcal{A} \times_1 u_p \times_2 u_l \times_3 u_v \tag{6.3.10}$$

Next we provide the formulation to shift the multilinearity from the core tensor side to the description vector side. Applying the rule of mode multiplication, (6.3.10) can be written as:

$$\begin{aligned}
I_T &= \sum_{i_v=1}^{N'_v} u_v(i_v).\sum_{i_l=1}^{N'_l} u_l(i_l).\sum_{i_p=1}^{N'_p} u_p(i_p).I_{P^e_{i_p}L^e_{i_l}V^e_{i_v}} \\
&= \sum_{i_v=1}^{N'_v} u_v(i_v).\sum_{i_l=1}^{N'_l} u_l(i_l).u_p.\begin{bmatrix} I_{P^e_1 L^e_{i_l} V^e_{i_v}} \\ ... \\ I_{P^e_{N'_p} L^e_{i_l} V^e_{i_v}} \end{bmatrix} \\
&= \sum_{i_v=1}^{N'_v} u_v(i_v).\Big\{u_l(1).u_p\begin{bmatrix} I_{P^e_1 L^e_1 V^e_{i_v}} \\ ... \\ I_{P^e_{N'_p} L^e_1 V^e_{i_v}} \end{bmatrix} + ... + u_l(N'_l).u_p\begin{bmatrix} I_{P^e_1 L^e_{N'_l} V^e_{i_v}} \\ ... \\ I_{P^e_{N'_p} L^e_{N'_l} V^e_{i_v}} \end{bmatrix}\Big\}
\end{aligned} \tag{6.3.11}$$

$$\begin{aligned}
\Rightarrow I_T &= \sum_{i_v=1}^{N'_v} u_v(i_v).u_l.\begin{bmatrix} u_p\begin{bmatrix} I_{P^e_1 L^e_1 V^e_{i_v}} \\ ... \\ I_{P^e_{N'_p} L^e_1 V^e_{i_v}} \end{bmatrix} & ... & 0 \\ ... & ... & ... \\ 0 & ... & u_p\begin{bmatrix} I_{P^e_1 L^e_{N'_l} V^e_{i_v}} \\ ... \\ I_{P^e_{N'_p} L^e_{N'_l} V^e_{i_v}} \end{bmatrix} \end{bmatrix} \\
&= \sum_{i_v=1}^{N'_v} u_v(i_v).u_l.\underset{(N'_l \times N'_p N'_l)}{\begin{bmatrix} u_p & ... & 0 \\ ... & ... & ... \\ 0 & ... & u_p \end{bmatrix}}.\begin{bmatrix} \begin{bmatrix} I_{P^e_1 L^e_1 V^e_{i_v}} \\ ... \\ I_{P^e_{N'_p} L^e_1 V^e_{i_v}} \end{bmatrix} \\ ... \\ ... \\ \begin{bmatrix} I_{P^e_1 L^e_{N'_l} V^e_{i_v}} \\ ... \\ I_{P^e_{N'_p} L^e_{N'_l} V^e_{i_v}} \end{bmatrix} \end{bmatrix}
\end{aligned} \tag{6.3.12}$$

Let us denote:

$$\begin{bmatrix} u_p & ... & 0 \\ ... & ... & ... \\ 0 & ... & u_p \end{bmatrix} . \begin{bmatrix} \begin{bmatrix} I_{P_1^e L_1^e V_{i_v}^e} \\ ... \\ I_{P_{N_p'}^e L_1^e V_{i_v}^e} \end{bmatrix} \\ ... \\ \begin{bmatrix} I_{P_1^e L_{N_l'}^e V_{i_v}^e} \\ ... \\ I_{P_{N_p'}^e L_{N_l'}^e V_{i_v}^e} \end{bmatrix} \end{bmatrix} = Y_{i_v} \tag{6.3.13}$$

then we can rewrite (6.3.12) as:

$$\begin{aligned} I_T &= \sum_{i_v=1}^{N_v'} u_v(i_v).u_l.Y_{i_v} \\ &= u_v. \begin{bmatrix} u_l.Y_{i_v=1} & ... & 0 \\ ... & ... & ... \\ 0 & ... & u_l.Y_{i_v=N_v'} \end{bmatrix} = u_v. \underset{(N_v' \times N_l' N_v')}{\begin{bmatrix} u_l & ... & 0 \\ ... & ... & ... \\ 0 & ... & u_l \end{bmatrix}} . \begin{bmatrix} Y_{i_v=1} \\ .. \\ .. \\ Y_{i_v=N_v'} \end{bmatrix} \end{aligned} \tag{6.3.14}$$

Substituting Y_{i_v} with its definition and after few more steps we obtain:

$$\begin{aligned} I_T &= u_v. \underset{(N_v' \times N_l' N_v')}{\begin{bmatrix} u_l & ... & 0 \\ ... & ... & ... \\ 0 & ... & u_l \end{bmatrix}} . \underset{(N_v' N_l' \times N_v' N_l' N_p')}{\begin{bmatrix} u_p & ... & ... & ... & 0 \\ ... & u_p & ... & ... & ... \\ ... & ... & ... & ... & ... \\ ... & ... & ... & ... & ... \\ 0 & ... & ... & ... & u_p \end{bmatrix}} . \underset{(N_v' N_l' N_p' \times N_x)}{\begin{bmatrix} I_{P_1^e L_1^e V_1^e} \\ ... \\ I_{P_{N_p'}^e L_1^e V_1^e} \\ I_{P_1^e L_2^e V_1^e} \\ ... \\ I_{P_{N_p'}^e L_{N_l'}^e V_1^e} \\ I_{P_1^e L_1^e V_2^e} \\ ... \\ I_{P_{N_p'}^e L_{N_l'}^e V_{N_v'}^e} \end{bmatrix}} \\ \Rightarrow I_T &= f(u_p, u_l, u_v) \,.\, \tilde{A} \end{aligned} \tag{6.3.15}$$

$f : \mathbb{R}^{N'_p} \times \mathbb{R}^{N'_l} \times \mathbb{R}^{N'_v} \rightarrow \mathbb{R}^{N'_p N'_l N'_v}$ is a multilinear function over u_p, u_l, u_v. $\tilde{\mathcal{A}}$ contains images of all the eigenpersons at all the eigenlightings and eigenviewpoints, and therefore, a derivative from the core tensor. Let $d_T = f(u_p, u_l, u_v)$ be the description vector for I_T, then we can reformulate (6.3.2) as:

$$\min_{d_T} ||I_T - d_T \,.\, \tilde{A}||_2 \tag{6.3.16}$$

Fortunately, this is a linear problem and the least squares solution for d_T is computed as:

$$d_T = I_T \times \tilde{A}^{+} \tag{6.3.17}$$

It is to be noted that, the optimal solution for d_T implies optimal solution for u_p, u_l and u_v as well, in contrast to the previous algorithm MPCA-LV, where we had a suboptimal solution of the optimization problem. However, on the flip side, given d_T, we cannot compute u_p uniquely, and therefore, we cannot use the usual recognition approach of comparing u_p in the person space as followed by both MPCA-ML and MPCA-LV. We need to compute description vectors for all the training images and store them for recognition. Recognition follows an Eigenface like algorithm, where d_T is compared to the stored description vectors and the best matching training image is found. Next, we will show that the description vectors of training images can be directly calculated from the matrices U^P, U^L and U^V in an efficient way.

Definition 8. *Let us define C_P as:*

$$C_P = \begin{bmatrix} C_P^{(1)} \\ C_P^{(2)} \\ \ldots \\ \ldots \\ C_P^{(N_p)} \end{bmatrix} \tag{6.3.18}$$

where, $C_P^{(k)}$ is defined as:

$$C_P^{(k)} = \begin{bmatrix} u_p^{(k)} & 0 & \ldots & 0 \\ 0 & u_p^{(k)} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & u_p^{(k)} \end{bmatrix} \tag{6.3.19}$$

where, $u_p^{(k)}$($=$ *k'th row of* U^P) *is repeated diagonally for* $N_l' N_v'$ *times.*

Similarly we can define C_L *as:*

$$C_L = \begin{bmatrix} C_L^{(1)} \\ C_L^{(2)} \\ ... \\ ... \\ C_L^{(N_l)} \end{bmatrix} \tag{6.3.20}$$

where, $C_L^{(k)}$ *is defined as:*

$$C_L^{(k)} = \begin{bmatrix} u_l^{(k)} & 0 & ... & 0 \\ 0 & u_l^{(k)} & ... & 0 \\ ... & ... & ... & ... \\ ... & ... & ... & ... \\ 0 & 0 & ... & u_l^{(k)} \end{bmatrix} \tag{6.3.21}$$

where, $u_l^{(k)}$($=$ *k'th row of* U^L) *is repeated diagonally for* $N_v' N_p$ *times.*

And C_V *as:*

$$C_V = \begin{bmatrix} C_V^{(1)} \\ C_V^{(2)} \\ ... \\ ... \\ C_V^{(N_v)} \end{bmatrix} \tag{6.3.22}$$

where, $C_V^{(k)}$ *is defined as:*

$$C_V^{(k)} = \begin{bmatrix} u_v^{(k)} & 0 & ... & 0 \\ 0 & u_v^{(k)} & ... & 0 \\ ... & ... & ... & ... \\ ... & ... & ... & ... \\ 0 & 0 & ... & u_v^{(k)} \end{bmatrix} \tag{6.3.23}$$

where, $u_v^{(k)}$($=$ *k'th row of* U^V) *is repeated diagonally for* $N_p N_l$ *times.*

Based on the above definitions, the following theorem provides the desired relationship between the image description vector and C_P, C_L & C_V, which were in turn calculated from the matrices U^P, U^L and U^V.

Theorem 3. *Let $M = C_V \times C_L \times C_P$. If m_k is the k'th row of the matrix $\mathcal{M}$ then,*

$$m_k = I_{P_{i_p} L_{i_l} V_{i_v}} \times \tilde{\mathcal{A}}^{+} \tag{6.3.24}$$

where the indices for person, lighting and viewpoint is gives as,

- $i_p = ((k-1) \bmod N_p + 1)$
- $i_l = ((\lceil \frac{k}{N_p} \rceil - 1) \bmod N_l + 1)$
- $i_v = (\lceil \frac{k}{N_p \times N_l} \rceil)$

Proof. Let us start with the definition of $\tilde{A}$:

$$\tilde{A} = \begin{bmatrix} I_{P_1^e L_1^e V_1^e} \\ \cdots \\ I_{P_{N_p'}^e L_1^e V_1^e} \\ I_{P_1^e L_2^e V_1^e} \\ \cdots \\ I_{P_{N_p'}^e L_{N_l'}^e V_1^e} \\ I_{P_1^e L_1^e V_2^e} \\ \cdots \\ \cdots \\ I_{P_{N_p'}^e L_{N_l'}^e V_{N_v'}^e} \end{bmatrix} \tag{6.3.25}$$

Let,

$$D_1^{(k)} = C_P^{(k)} \times \tilde{A} \tag{6.3.26}$$

then substituting the $C_P^{(k)}$ from (6.3.19) and $\tilde{\mathcal{A}}$ from (6.3.25),

$$D_1^{(k)} = \begin{bmatrix} u_p^k & 0 & ... & 0 \\ 0 & u_p^k & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & u_p^k \end{bmatrix} \cdot \begin{bmatrix} I_{P_1^e L_1^e V_1^e} \\ ... \\ I_{P_{N_p'}^e L_1^e V_1^e} \\ I_{P_1^e L_2^e V_1^e} \\ ... \\ I_{P_{N_p'}^e L_{N_l'}^e V_1^e} \\ I_{P_1^e L_1^e V_2^e} \\ ... \\ ... \\ I_{P_{N_p'}^e L_{N_l'}^e V_{N_v'}^e} \end{bmatrix}$$

$$\Rightarrow D_1^{(k)} = \begin{bmatrix} \sum_{j=1}^{N_p'} u_p^k(j).I_{P_j^e L_1^e V_1^e} \\ \sum_{j=1}^{N_p'} u_p^k(j).I_{P_j^e L_2^e V_1^e} \\ .... \\ \sum_{j=1}^{N_p'} u_p^k(j).I_{P_j^e L_{N_l'}^e V_1^e} \\ \sum_{j=1}^{N_p'} u_p^k(j).I_{P_j^e L_1^e V_2^e} \\ .... \\ .... \\ \sum_{j=1}^{N_p'} u_p^k(j).I_{P_j^e L_{N_l'}^e V_{N_v'}^e} \end{bmatrix} = \begin{bmatrix} I_{P_k L_1^e V_1^e} \\ I_{P_k L_2^e V_1^e} \\ .... \\ I_{P_k L_{N_l'}^e V_1^e} \\ I_{P_k L_1^e V_2^e} \\ .... \\ .... \\ I_{P_k L_{N_l'}^e V_{N_v'}^e} \end{bmatrix} \tag{6.3.27}$$

where, $I_{P_k L_{k_1}^e V_{k_2}^e} = \sum_{j=1}^{N_p'} u_p^k(j).I_{P_j^e L_{k_1}^e V_{k_2}^e}$ is the image of k'th actual person at k_1'th eigenlighting and k_2'th eigenviewpoint.

Let us define:

$$D_1 = C_P \times \tilde{\mathcal{A}} \tag{6.3.28}$$

then substituting C_P from (6.3.18) we obtain:

$$D_1 = \left[\begin{array}{c} C_P^{(1)} C_P^{(2)} ......C_P^{(N_p)} \end{array} \right]^T \tilde{\mathcal{A}} \ = \left[\begin{array}{c} D_1^{(1)} D_1^{(2)} ......D_1^{(N_p)} \end{array} \right]^T \ = \left[\begin{array}{c} I_{P_1 L_1^e V_1^e} \\ ... \\ I_{P_1 L^e_{N_l'} V_1^e} \\ I_{P_1 L_1^e V_2^e} \\ ... \\ I_{P_1 L^e_{N_l'} V^e_{N_v'}} \\ I_{P_2 L_1^e V_1^e} \\ ... \\ ... \\ I_{P_{N_p} L^e_{N_l'} V^e_{N_v'}} \end{array} \right] \tag{6.3.29}$$

Following a similar derivation:

$$D_2 = C_L \times D_1 \ = \left[\begin{array}{c} I_{P_1 L_1 V_1^e} \\ ... \\ I_{P_1 L_1 V^e_{N_v'}} \\ I_{P_2 L_1 V_1^e} \\ ... \\ I_{P_{N_p} L_1 V^e_{N_v'}} \\ I_{P_1 L_2 V_1^e} \\ ... \\ ... \\ I_{P_{N_p} L_{N_l} V^e_{N_v'}} \end{array} \right] \tag{6.3.30}$$

and finally:

$$D_3 = C_V \times D_2 \;=\; \begin{bmatrix} I_{P_1 L_1 V_1} \\ \cdots \\ I_{P_{N_p} L_1 V_1} \\ I_{P_1 L_2 V_1} \\ \cdots \\ I_{P_{N_p} L_{N_l} V_1} \\ I_{P_1 L_1 V_2} \\ \cdots \\ \cdots \\ I_{P_{N_p} L_{N_l} V_{N_v}} \end{bmatrix} \tag{6.3.31}$$

The structure of D_3 is such that the image $I_{P_{i_p} L_{i_l} V_{i_v}}$ can be found at

$$(N_p \times N_l \times (i_v - 1) + N_p \times (i_l - 1) + i_p)\text{'th row of } D_3 \tag{6.3.32}$$

Now,

$$\begin{aligned} D_3 &= C_V \times D_2 \\ &= C_V \times C_L \times D_1 \\ &= C_V \times C_L \times C_P \times \tilde{A} \\ &= \mathcal{M} \times \tilde{A} \\ => \quad M &= D_3 \times \tilde{A}^+ \quad (^+ \text{ implies pseudoinverse}) \end{aligned} \tag{6.3.33}$$

If m_k is the k'th row of M and if

$$k = N_p \times N_l \times (i_v - 1) + N_p \times (i_l - 1) + i_p \tag{6.3.34}$$

then,

$$\begin{aligned} m_k &= k\text{'th row of } D_3 \times \tilde{A}^+ \\ &= I_{P_{i_p} L_{i_l} V_{i_v}} \times \tilde{A}^+ \end{aligned} \tag{6.3.35}$$

Solving (6.3.34) for i_p, i_l and i_v we obtain, $i_p = ((k-1) \bmod N_p + 1)$, $i_l = ((\lceil \frac{k}{N_p} \rceil - 1) \bmod N_l + 1)$ and $i_v = (\lceil \frac{k}{N_p \times N_l} \rceil)$. □

It follows from the Theorem 3 that the matrix $M = C_V \times C_L \times C_P$ contains the description vectors, $f(u_p, u_l, u_v)$ of all training images. Each row m_k of the matrix

M refers to the description vector of a training image, whose person, lighting and viewpoint indices are provided by the above theorem. The algorithm for testing is presented in the Algorithm 6.

Algorithm 6: Testing algorithm for MPCA-JS

Input: Test image I_T, Projection matrix $P = \tilde{A}^+$ and the Coefficient matrix M.

Output: Person identity of the test image p_T

1. Find the description vector m_T for the test image as,

$$m_T = I_T \times P$$

2. Use a Nearest Neighbour classifier to find the best matching description vector m_b *i.e.* that minimizes,

$$\min_k ||m_T - m_k|| \text{ for } k = 1, ..., (N_p \times N_l \times N_v)$$

where, m_k is the k'th row of the matrix M. The distance measure we use is the *cosine distance*. For two vectors $\mathbf{a}$ and $\mathbf{b}$ the *cosine distance* between them is defined as,

$$cosine\_dist(\mathbf{a}, \mathbf{b}) = \frac{\langle \mathbf{a} \, , \mathbf{b} \rangle}{||\mathbf{a}|| \, ||\mathbf{b}||}$$

3. The person identity, p_T for the test image is the person identity of the best matching vector m_b and it is given by,

$$p_T = ((b-1) \bmod N_p + 1)$$

Clearly, this algorithm provides a deterministic approach to solve (6.3.2) without losing the optimality in solution, and hence, is superior to both the previous algorithms. Another major advantage is that it follows the conventional method of recognition *i.e.* projection (generation of description vector) followed by classification. Therefore, sophisticated classifiers can be used to improve the performance of the testing algorithm. However, a naive implementation using NN classifier can be enormously expensive, as the search complexity is $O(N_p N_l N_v N_p' N_l' N_v')$. In comparison, MPCA-ML has a search complexity of only $O(N_p N_p')$ and MPCA-LV has a search complexity of only $O(N_l N_v N_p N_p')$. Another point to be noted is that it uses a modified descriptor for recognition as opposed to the person factor used in the previous method. Also, it is clear from the algorithm that this approach is similar to the ideas presented in Chapter 4. However, through this alternative derivation we are able describe the methods in a cleaner way by explicitly stating the quality of solution it offers and then linking the performance to it. Significantly, this approach is distinct from the previous approaches in the sense that it uses a multilinear combination of all factors as the descriptor instead of the usual person space representation. The power of discrimination of this descriptor may be reduced considerably as more and more factors are added with the person-factor. Therefore, this method will provide good performance when the number of factors is low, and may fail when the number becomes higher.

6.3.4 Approach 4: MPCA-PS

In all the previous approaches, whilst solving (6.3.2) we allow u_p to be a free variable and solve the optimization problem. For recognition, we either directly use u_p for recognition as in MPCA-ML and MPCA-LV, or use a description that is a function of u_p, as in MPCA-JS. However, as we recognize persons that are present in our training database, we expect u_p to be the same as that of person space representation of the person the test image belongs to. Following this idea, we can simplify the optimization problem of (6.3.2), by finding its minima over the set of person space representations of the training persons. As each individual row of U^P refers to a specific person, we make a candidate set of u_p such that $\{u_p^k = k\text{'th row of } U^P,\ \text{for } k = 1, 2, ..., N_p\}$. For each u_p^k we generate a multilinear optimization subproblem that is multilinear in u_l, u_v. We follow similar techniques as in MPCA-JS to solve these multilinear subproblems and finally we compute the overall minima out of those subproblems. The specific u_p^k, which provides the overall minima, reveals the identity of the test image as that of person k. Specifically, we modify (6.3.2) as:

$$\min_{k,u_l,u_v} ||I_T - \mathcal{S} \times_1 u_p^k \times_2 u_l \times_3 u_v \times_4 U^X||_2 \tag{6.3.36}$$

Next, we develop a mechanism to solve this optimization problem. We start by defining:

$$\mathcal{B}_k = \mathcal{S} \times_1 u_p^k \times_4 U^X \tag{6.3.37}$$

Clearly, $\mathcal{B} \in \mathbb{R}^{1\times N_l' \times N_v' \times N_x}$. With analogy to the definition of T, we can state that,

$$\mathcal{B}_k(1, i_l, i_v) = I_{P_k L^e_{i_l} V^e_{i_v}} \tag{6.3.38}$$

where, $I_{P_k L^e_{i_l} V^e_{i_v}}$ is the image of k'th person at i_l'th eigenlighting and i_v'th eigenviewpoint. If the test image, I_T belongs to the K'th person and, u_l, u_v are its projection in lighting and viewpoint space respectively then,

$$\begin{aligned} I_T &= \mathcal{S} \times_1 u_p^K \times_2 u_l \times_3 u_v \times_4 U^X \\ &= \mathcal{B}_K \times_2 u_l \times_3 u_v \end{aligned} \tag{6.3.39}$$

Following the rule of mode multiplication, (6.3.39) can be written as:

$$\begin{aligned}
I_T &= \textstyle\sum_{i_v=1}^{N_v'} u_v(i_v). \sum_{i_l=1}^{N_l'} u_l(i_l). I_{P_K L^e_{i_l} V^e_{i_v}} \\
&= \textstyle\sum_{i_v=1}^{N_v'} u_v(i_v). u_l. \begin{bmatrix} I_{P_K L^e_1 V^e_{i_v}} \\ \cdots \\ I_{P_K L^e_{N_l'} V^e_{i_v}} \end{bmatrix} \\
&= u_v(1).u_l \begin{bmatrix} I_{P_K L^e_1 V^e_1} \\ \cdots \\ I_{P_K L^e_{N_l'} V^e_1} \end{bmatrix} + ... + u_v(N_v').u_l \begin{bmatrix} I_{P_K L^e_1 V^e_{N_v'}} \\ \cdots \\ I_{P_K L^e_{N_l'} V^e_{N_v'}} \end{bmatrix} \\
&= u_v. \begin{bmatrix} u_l \begin{bmatrix} I_{P_K L^e_1 V^e_1} \\ \cdots \\ I_{P_K L^e_{N_l'} V^e_1} \end{bmatrix} & \cdots & 0 \\ 0 & \cdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \cdots & u_l \begin{bmatrix} I_{P_K L^e_1 V^e_{N_v'}} \\ \cdots \\ I_{P_K L^e_{N_l'} V^e_{N_v'}} \end{bmatrix} \end{bmatrix}
\end{aligned} \tag{6.3.40}$$

$$\Rightarrow I_T = u_v \underbrace{\begin{bmatrix} u_l & ... & 0 \\ ... & ... & ... \\ 0 & ... & u_l \end{bmatrix}}_{N'_v \times N'_l N'_v} . \begin{bmatrix} \begin{bmatrix} I_{P_K L^e_1 V^e_1} \\ ... \\ I_{P_K L^e_{N'_l} V^e_1} \end{bmatrix} \\ ... \\ ... \\ \begin{bmatrix} I_{P_K L^e_1 V^e_{N'_v}} \\ ... \\ I_{P_K L^e_{N'_l} V^e_{N'_v}} \end{bmatrix} \end{bmatrix} \tag{6.3.41}$$

$$\Rightarrow I_T = g(u_v, u_l).\tilde{B}_K$$

where, $g : \mathbb{R}^{N'_l} \times \mathbb{R}^{N'_v} \rightarrow \mathbb{R}^{N'_l N'_v}$ is a multilinear function over u_l, u_v. $\tilde{B}_k$ contains images of k'th person at all the eigenlightings and eigenviewpoints. Let, $c_T = g(u_v, u_l)$ is the description vector of lighting and pose for I_T then we can reformulate (6.3.36) corresponding to the person K as:

$$\min_{c_T} ||I_T - c_T \times \tilde{B}_K||_2 \tag{6.3.42}$$

This is a linear optimization problem and the least squares solution for c_T, specific to $k = K$ is computed as:

$$c_T^K = I_T \times \tilde{B}_K^+ \tag{6.3.43}$$

In order to find the optimal k, we can write (6.3.36) as:

$$\min_{k} ||I_T - c_T^k \times \tilde{B}_k||_2 \tag{6.3.44}$$

The algorithm for testing is presented in Algorithm 7.

In this approach we enforce suboptimality to the solution by optimizing (6.3.2) over a set of u_p. However, for a fixed u_p we solve for optimal values of u_l and u_v. Though both MPCA-LV and this approach generate suboptimal solutions, the fundamental difference between them is that, in MPCA-LV we fix u_l, u_v to the training lighting and viewpoint cases and optimally solve for u_p, whereas in this case we fix u_p to the training persons and optimally solve for u_l, u_v. Fixing u_l, u_v implies that we expect the lighting condition and viewpoint of a test image to be almost similar to any of the training lighting and viewpoint cases. However, fixing u_p implies

Algorithm 7: Testing algorithm for MPCA-PS

Input: Test image I_T, and $\{\tilde{B}_k\}_{k=1}^{N_p}$.

Output: Identity of the test image.

1. For every $k = 1, ..., N_p$ compute
   - $c_T^k = I_T \times \tilde{B}_k^+$
   - $e^k = ||I_T - c_T^k \times \tilde{B}_k||$
     The distance measure we use is Euclidean distance.
2. If e^{k_m} is the smallest of $\{e_k\}_{k=1}^{N_p}$ then the identity of I_T is k_m.

that we expect the test image to have the same person space representation as the person the test image belongs to. Clearly, the later assumption is more logical and the suboptimal solution for this case should almost be near-optimal, implying we should expect similar recognition performance as that of MPCA-JS. Moreover, as this algorithm has search complexity of only $O(N_pN_x)$, it is much more efficient than MPCA-JS. On the down side, due to the way recognition is performed, we cannot further reduce the time. In comparison, MPCA-JS can be made faster by using sophisticated classifiers. It is also obvious that this approach is similar to the ideas presented in Chapter 5. However, with this derivation we are able to express it in a clearer fashion by explicitly stating the quality of solution it offers. Firstly, it solves the optimization problem almost-optimally, thereby overcoming the drawback of MPCA-LV. Secondly, it utilizes person space indirectly for recognition thereby overcoming the limitation of MPCA-JS.

6.3.5 Theoretical Analysis of the Four approaches

As shown in Table 6.1, we are able to describe all four methods in a coherent way in terms of the assumption they make and the strategies they adopt to solve the multilinear optimization problem of Eq. (6.3.2). Critically, we are also able to outline their failure modes based on the analysis of solution mechanisms they follow and get an understanding of their expected behaviours and consequent employability

Method	**Assumptions**	**Solution Strategy (A)**	**Quality of Solution**	**Recognition Method (B)**	**Recognition Cost (A+B)**
MPCA -ML	None	Brute force - using Alternating Least squares (ALS)	Optimal if global optima is reached *(however ALS fails to converge if factor spaces are high-dimensional or factors are highly correlated)*	Compare u_p with stored person space projection of known persons	Indefinite + $O(N_p N'_p)$
MPCA -LV	Lighting and Viewpoint for the test image is almost similar to training lightings and viewpoints.	Fix (u_l, u_v) to training lighting-viewpoint combination and simplify (6.3.2) to obtain $N_l N_v$ no. of linear optimization problems (6.3.4). Solve all the linear optimization problems to generate a candidate set of u_p.	Sub-optimal *(extremely poor if test lighting/ viewpoint conditions are quite different from that of the training set)*	Compare the candidate set of u_p with stored person space projection of known persons in a pair-wise manner.	$N_l N_v$ projections + $O(N_l N_v N_p N'_p)$
MPCA -JS	None	Manipulate the structure of core tensor to cast the multilinear optimization problem of (6.3.2) as a linear optimization problem in (6.3.16). The solution is obtained as a descriptor, that is a multilinear function of all the factors.	Optimal *(however, the factors are inseparable and the multilinear descriptor may not have the same discriminative power as that of u_p)*	Compare the description vector (of length $N'_v N'_l N'_p$) with stored description vectors of all training images.	Only one projection + $O(N_v N_l N_p N'_v N'_l N'_p)$
MPCA-PS	person space projection of the test image should be close to the person space representation of the training person the test image belongs to.	Fix u_p to the person space representation of training persons and divide (6.3.2) into N_p no. of multilinear optimization subproblems, each specific to one person (6.3.36). Manipulate the structure of core tensor to cast each of them as a linear optimization problem (6.3.42).	Sub-optimal *(however, highly optimal if testing involves only known persons and it indirectly uses u_p to discriminate)*	Among the N_p optimization problems, find the one which generate the least error in approximating the test image.	N_p projections + $O(N_p N_x)$

Table 6.1: Summary of the four approaches.

in different face recognition scenarios. Firstly, the direct approach MPCA-ML is not preferable for two reasons: (1) the failure of the optimization routine when the factor spaces are high-dimensional as in the cases of large databases and (2) indefinite cost of testing. MPCA-LV, though a linear time algorithm is heavily dependent on the assumption that the training dataset contains images that cover all possible test lighting/viewpoint conditions, and therefore, is brittle when these conditions are not met. MPCA-JS overcomes all the previous problems and offers the best solution for the optimization problem. However, the solution obtained is in terms of a descriptor from which different factors are impossible to separate. The discriminating power of this descriptor is hard to predict and could be worse than using unique person factors, particularly when the number of factors goes higher because of the reduced influence of the unique person factor in the descriptor. For MPCA-PS the overall solution is suboptimal as we do not compute person-factor value, however, for each fixed value of the person-factor (respective to each known persons) we solve for all other factors optimally. Thereby, the solution is nearly optimal if the test images are only from known persons. With this strategy, MPCA-PS can on one hand overcome the shortcomings of the MPCA-LV approach while on the other hand it improves upon MPCA-JS by keeping the person factor separate and using it indirectly for recognition. By being dependent on the person factor for discrimination, it is able to provide low complexity testing whilst at the same time providing a safeguard from any performance variation due to varying number of factors as in the case of MPCA-JS. Hence, we can expect that MPCA-PS will provide a consistent performance over any kind of dataset. MPCA-ML will fail for larger databases. MPCA-LV will perform poorly when test conditions are unseen while MPCA-JS will typically provide good performance with the condition that an increase in the number of factors will negatively impact the overall performance. In the next section we will validate these theoretical insights through experiments on different databases.

6.4 Experiments, Analysis and Evaluation

In this section, we provide experimental results of the different tensor based approaches we derived on a number of publicly available benchmark datasets. Specifically, we use PEAL (Gao *et al.*, 2008), YaleB Frontal (Georghiades *et al.*, 2001a), Extended YaleB (Lee *et al.*, 2005b) and Weizmann face database for experiments. The datasets are the same as described in (Section 4.5 and Section 5.5) except for PEAL, for which a larger version of the dataset is used in this chapter.

PEAL is a face-database of Chinese nationals at different poses, lightings and expressions (Gao *et al.*, 2008). From this database we created a PEAL Lighting variation dataset by choosing images of 150 persons at different lighting conditions. The lighting conditions have three different modes: light focussing on the middle of the faces (M), light focussing from up (U) and light focussing from down (D). Each mode has light focussing from five different directions ($0°, \pm 45°, \pm 90°$), resulting in 15 different lighting conditions, though images at all 15 lighting conditions are not available for all persons. Primarily the light used is of incandescent type, however, occasionally fluorescent lamps were used, resulting in a maximum of 5 extra lighting conditions. In the dataset the minimum number of images a person has is 6 whilst the maximum number is 20. However, only three images per person are available at exactly the same lighting conditions, *i.e.* at M-0°, U-0° and D-0°.

For the YaleB Frontal dataset, four sets of experiments are performed with randomly chosen 5, 10, 15 and 20 lighting conditions for training with the rest being used for testing. For the experiments on the Extended YaleB database, 16 representative lighting conditions at 5 representative viewpoints are used for training and the rest for testing. For the experiments on Weizmann face database 3 representative viewpoints (at 0° and $\pm$ 34°), at 2 randomly selected lighting conditions and at all the expressions are used for training and the rest for testing. For the experiments on the PEAL lighting variation dataset, 3 lighting conditions at M-0°, U-0° and D-0° are used for training and the rest 1040 images for testing. Thus, in both the PEAL and YaleB Frontal database, the test images are at unseen lighting. For both Extended YaleB database and Weizmann database, the test images are either at unseen lighting or unseen viewpoints or both. For the experiments on YaleB Frontal, each set of experiments is repeated 20 times and the average and the standard deviations of the results are reported. Experiments are performed with MPCA-ML, MPCA-LV (Vasilescu and Terzopoulos, 2002a), MPCA-JS, MPCA-PS along with PCA (Turk and Pentland, 1991b) for baselining and LDA(Belhumeur *et al.*, 1997) and LPP(He and Niyogi, 2003) to benchmark results produced by the tensor based methods against the leading ones. All the timings reported here are for Matlab code running on a Intel Xeon 8×2.3GHz server with 16GB RAM and the programs were written without any extra effort for code optimization.

We have used *energy thresholding* to select the number of *eigenvectors* to be retained (*i.e.* for PCA and for Pixel mode of MPCA the energy threshold = 0.96, at all other modes of MPCA the threshold = 0.99). LDA and LPP is performed in the reduced dimension after performing PCA on the data with the same threshold to maintain uniformity. Face recognition using PCA, LDA and LPP uses Euclidean distance measure for comparison, while all multilinear analysis based approaches use the cosine distance measure. For MPCA-ML, convergence is achieved when norm of the difference of two consecutive estimates of u_p, u_l, u_v is less than 0.000001 or the number of iterations have reached 2000.

Table 6.2 and 6.3 shows the recognition performance and test time for experiments on YaleB dataset. Experimental results for Extended YaleB database, Weizmann database and PEAL dataset are tabulated in Table 6.4, 6.5 and 6.6, respectively.

Recognition method	***Avg. Accuracy (%) (std. deviation)***			
	5 train	**10 train**	**15 train**	**20 train**
PCA	53.71 (9.6)	69.20 (5.3)	77.69 (4.8)	88.29 (3.9)
MPCA-ML	71.67 (11.5)	85.92 (4.3)	91.37 (2.3)	93.36 (2.0)
MPCA-LV	43.22 (14.3)	56.41 (9.6)	63.30 (9.3)	72.66 (8.6)
MPCA-JS	76.16 (10.0)	87.52 (5.8)	92.78 (2.5)	95.02 (1.7)
MPCA-PS	**77.33** (11.1)	**90.40** (4.1)	**94.38** (1.5)	**95.52** (1.5)
PCA+LDA	72.53 (11.7)	85.07 (6.1)	90.63 (3.6)	93.44 (3.0)
PCA+LPP	72.98 (10.8)	86.97 (4.3)	91.84 (2.4)	94.38 (2.0)

Table 6.2: Recognition performance on YaleB Frontal lighting variation dataset.

Recognition method	***Avg. Time (sec.)*** 5 train	**10 train**	**15 train**	**20 train**
PCA	0.11	0.21	0.34	0.45
MPCA-ML	14.75	18.83	25.76	32.70
MPCA-LV	0.83	2.86	4.37	5.85
MPCA-JS	0.27	0.85	1.41	2.20
MPCA-PS	0.10	**0.15**	**0.19**	**0.24**
PCA+LDA	**0.09**	0.16	0.25	0.31
PCA+LPP	0.10	0.21	0.32	0.43

Table 6.3: Testing time for YaleB Frontal lighting variation dataset. Test time is for 100 test images.

	Recognition method PCA	**MPCA-ML**	**MPCA-LV**	**MPCA-JS**	**MPCA-PS**	**PCA +LDA**	**PCA +LPP**
Accuracy (%)	67.48	75.95 (79.77)	49.89	84.20	**85.03**	**85.03**	84.30
Test time (sec.)	1.78	283.54	1.51	10.54	**0.79**	1.23	1.64

Table 6.4: Experimental results for Extended YaleB dataset with lighting+viewpoint variation. Test time is for 100 test images.

	Recognition method						
	PCA	**MPCA-ML**	**MPCA-LV**	**MPCA-JS**	**MPCA-PS**	**PCA +LDA**	**PCA +LPP**
Accuracy (%)	86.71	84.52 (86.31)	69.64	77.18	**98.21**	97.82	91.67
Test time (sec.)	0.20	40.38	0.20	0.53	**0.12**	0.12	0.20

Table 6.5: Experimental results for Weizmann database with lighting+viewpoint+expression variation. Test time is for 100 test images.

	Recognition method						
	PCA	**MPCA-ML**	**MPCA-LV**	**MPCA-JS**	**MPCA-PS**	**PCA +LDA**	**PCA +LPP**
Accuracy (%)	74.79	74.04 (83.08)	75.10	**92.79**	87.79	89.89	88.94
Test time (sec.)	2.45	445.01	1.31	2.14	**1.23**	1.68	2.44

Table 6.6: Experimental results for PEAL lighting variation dataset. Test time is for 100 test images.

From the Tables 6.2–6.6 we list the following observations:

- In most of the experiments, MPCA-PS outperforms all other approaches including LDA and LPP. Particularly, in the Weizmann dataset, it showed excellent recognition accuracy. We attribute this to the nature of the dataset, where the appearance of persons are very different from each other, resulting in lesser overlap among *person-specific eigenmode* spaces. The opposite reasoning applies to the PEAL dataset, where persons share some similarities in appearance, resulting in a higher overlap of the *person-specific eigenmode* spaces. Therefore, we observe that MPCA-PS performed somewhat worse than MPCA-JS in the PEAL dataset, though comparable to LDA and LPP. Nevertheless, efficiency wise, MPCA-PS always outperforms other approaches significantly. Also in Table 6.2, note that the MPCA-PS has the least variation in performance due to different partitioning of YaleB Lighting variation dataset when larger number of training images is used, which shows the robustness of the algorithm. In Extended YaleB dataset, which is the largest among all the datasets we have, MPCA-PS is almost two times faster than PCA and MPCA-LV, almost 350 times faster than MPCA-ML, and around 14 times faster than MPCA-JS.

- In most cases, MPCA-JS performs similarly to MPCA-PS, except in the experiments on the Weizmann dataset. There, it performs worse than PCA. The reason as pointed out correctly in Chapter 5, is the poor generalization of the NN classifier in the high-dimensional *multilinear eigenmode* space. Also, it is to be noted that in the PEAL dataset, it demonstrated higher performance than both LDA and LPP.

- In all the experiments, MPCA-LV performs very poorly in terms of recognition accuracy, which is reflective of our analysis discussed before that this method is not suitable for application for recognition under unseen lightings and viewpoints.

- Finally, we observe that MPCA-ML produces poor performance except in the case of YaleB Lighting variation dataset, which concurs with our discussion of this method in the corresponding section. The better performance in YaleB is the result of the small number of variables as well as small dimensions. Compared to YaleB lighting variation dataset, YaleB Extended dataset has 3 variables, Weizmann has 4 variables and although PEAL dataset has only 2 variables, the person space is considerably higher (129 compared to ≤ 38 for YaleB) which results in poor performance of MPCA-ML in these datasets. We also note the high testing time required for this method. To further study the behaviour of ALS we experimented on YaleB Extended, Weizmann and PEAL dataset, by initializing person space variable u_p with the expected solution, *i.e.* the known person space representation of the test person. The intent is to find the performance of MPCA-ML when the ALS gets closer to the global minima. These accuracies are shown in brackets in the MPCA-ML columns of Tables 6.4, 6.5 and 6.6. We observe the recognition for this case gets closer to the best recognition rate (as obtained from MPCA-JS and MPCA-PS). This supports our viewpoint that the poor performance of MPCA-ML is the result of the poor behaviour of ALS and not because of the way the problem is formulated.

From the observations above we can conclude that the approaches MPCA-LV and MPCA-ML are not useful for practical purposes. Whilst MPCA-LV is limited by its design that makes it unsuitable for unseen test scenarios, MPCA-ML can potentially be improved by introducing an efficient and more accurate optimization routine than the one based on ALS. Between the MPCA-JS and MPCA-PS, whilst both provide some of the best results compared to LDA and LPP based approaches, the latter is able to achieve consistent performance at lower computational costs. Hence, we can state that the MPCA-PS algorithm is the best candidate among all the multilinear analysis based approaches considered.

6.5 Closing Remarks

In this chapter, we have unified different tensor based approaches for face recognition in a single optimization framework. The major contribution has been the development of a formal way for an explicit, non-empirical comparison among the tensor based face recognition approaches. In turn this enables us to gain a deeper understanding of the applicability of these algorithm in different face recognition scenarios. From the analysis, we determine that the MPCS-PS is likely to produce the best overall performance for all types of datasets despite the fact that it provides a sub-optimal solution. MPCA-JS provides an optimal solution and is likely to produce good recognition performance when the number of factors is small. MPCA-ML is impractical when dealing with large databases due to its large computational requirement and also may not be suitable when factor spaces are high dimensional as the underlying optimization process ALS encounters difficulty in such scenarios. Finally, the MPCA-LV algorithm is destined to produce poor performance when test conditions are different from the training conditions.

In addition to the theoretical analysis, we present an evaluation of all the algorithms against the leading face recognition approaches such as, LDA and LPP. From the experimental results we observe that:

- The algorithms behaved the way as predicted from the theoretical analysis.
- The algorithm MPCA-PS demonstrates superior performance in most of the experiments to both the LDA and LPP based approaches, generally accepted as the leading approaches in face recognition, and at a lower computational expense. Hence, it is the best candidate for use in real world face recognition applications.

However, we note that the training for MPCA-PS requires a large number of training images to construct a full face tensor. This can be a limiting factor in some real-world scenarios where adequate data is not available. In the next chapter, we address the data-inadequacy issue and present a solution based primarily on the framework presented in this chapter.

Chapter 7

Face Recognition From Few Samples: A Tensor Based Approach

7.1 Introduction

In the preceding chapters, we have discussed different tensor based face recognition algorithms and among them our proposed algorithms namely, MPCA-JS and MPCA-PS are found to produce recognition performance that is similar or better than the leading face recognition approaches. However, the basic assumption behind this evaluation was that the training set is well represented with a large number of training samples. Most of the current face recognition algorithms, such as Eigenface (Turk and Pentland, 1991b), Fisherface (Belhumeur *et al.*, 1997), LPP (He and Niyogi, 2003) etc. depend heavily on the availability of a large number of training exemplars for sound performance, as the underlying principle of those methods is based on finding the closest matching training image to a given test image. Among the tensor based approaches, testing in MPCA-JS is based on computing the closest training image, and hence, is affected by the number of training samples. On the other hand, MPCA-PS, being based on individually modelling subspaces for each person, is not directly affected by the number of training samples, though it initially requires large training samples to perform multilinear analysis and obtain individual subspaces. These approaches offer suitable alternatives when designing a face recognition system in a collaborative environment where collection of enough training samples are easier with user assistance. However, in scenarios such as video surveillance, this condition is hard to meet as the persons of interest are almost always criminals. As they are hostile, there may be only a few training images available. This insufficiency of training data leads to poor generalization of tensor based approaches and the current approaches. As discussed in Chapter 2, traditional approaches to address such shortcomings rely on synthesizing virtual training samples using different types of face models (Jung *et al.*, 2004) (Georghiades *et al.*, 2001b) (Beymer and Poggio, 1995). However, a detailed physical modelling of various factors that affect a face image is not available and synthesis of face images with a simplified model or ad-hoc processes do not help in recognition. Thus face recognition in such hostile scenarios remains an open problem and serves as the motivation for this work.

An alternative approach to generic face modelling is to learn the relationship between different factors directly from data. We propose to model this scenario by considering a "friendly-hostile" paradigm. We collect as many images as required from the "friendly" set of people, and assume that only a few (1-5 images) are available for the "hostile" or the persons of interest. From the friendly database we learn a generic structure of face images under the influence of different factors such as lighting, viewpoint expression etc., and then we formulate a way to transfer the structure to each hostile person so as to provide a formal description of their face manifolds, eliminating the need for large training samples. We note that in the synthesis based on face models a generic external model is required, however in our case, we unearth the model from the face dataset of friendly people, thereby, providing a realistic and useful data-dependent model. The advantage of this data-dependent model is that it can be effectively used to learn the details of a particular area of the global face manifold in scenarios when we know that people under surveillance are from a particular gender or race, by collecting data from similar subjects only.

We employ the concepts of multilinear analysis to introduce a face recognition algorithm based on the above "friendly-hostile" paradigm that is capable of classifying the hostile persons, by learning effectively from the friendly persons dataset. The application of multilinear models is based on the fact that a complete and independent separation is achieved. The independence of factors implies that a face image is independently controlled by the different factors, which in turn implies that only the knowledge of person-identity vector of a person is sufficient to describe its whole face subspace. We first construct a tensor from the face dataset of friendly people and obtain person-identity, lighting, viewpoint etc. factor spaces. Then, for every hostile person we provide a least square based optimization technique to obtain person-identity vectors from the few training images available on the basis of the factor spaces obtained from the factorization of the friendly dataset. We describe the whole approach in a synthesis framework providing interpretation of the face subspaces in an intuitive manner. Next, we re-derive the MPCA-PS algorithm to assert that for recognition also it is sufficient to know only the person-identity vectors of the hostile persons.

Until now, the situation we considered is the small dataset problem, prevalent in almost all surveillance scenarios. However, the proposed "friendly-hostile" paradigm can also be effectively utilized in scenarios where persons of interest are totally co-operative as in the situation of face recognition based access control systems. Using current techniques, if we were to construct a face recognition system that can handle extreme variation of lighting or viewpoints, then we must supply example face images at a number of conditions to reasonably cover the whole face space, thereby, making collection of a training set costly, time-consuming and tedious. However, we can apply our proposed "friendly-hostile" paradigm in such scenarios and save both time and cost. This is a typical unbalanced dataset scenario. The unbalance in the training dataset can also occur in many other scenarios, such as, when persons of interest are actually made up of both friendly and hostile people. Algorithms such as, Eigenface, Fisherface, LPP, and MPCA-JS are unable to handle such imbalance as the underlying NN-type classifiers are affected by the apriori-bias based on the number of training samples; persons with a higher number of training samples have higher apriori-bias than the persons with a lower number of training samples. However, the recognition algorithm for our approach is MPCA-PS, based on modelling subspaces corresponding to each class and thus is not affected by such imbalance and offers solutions to practical aspects of real world face recognition.

The main contributions of our work are:

1. Construction of a novel synthesis perspective on tensor factorisation that facilitates the development of the algorithm for training and testing. This approach

does not rely on filling in the “missing images” in the training set, rather it finds the more general description in terms of the face subspace for each person from the available training images, and recognition is performed by effectively using these subspaces.

2. Construction of a framework to solve the unbalanced dataset problem, an effective tool to address many real-world issues. We formulate the friendly-hostile paradigm where friendly and hostile group are defined based on the number of training images. This imbalance in the dataset creates problems for algorithms such as Eigenface(Turk and Pentland, 1991b), Fisherface (Belhumeur *et al.*, 1997), LPP (He and Niyogi, 2003) etc. due to the high apriori bias towards the friendly people with large number of training images. However, our recognition approach is not affected by such bias as each person, no matter how many training images they have, are represented by only one identity vector.

The outline of the chapter is as follows: in Section 7.2 we propose the “Friendly-Hostile” paradigm on which our solution is based. In Section 7.3, we present the outline of the proposed approach. Section 7.4 reviews the tensor model of face image ensembles and Section 7.5 presents a novel synthesis framework based on the multilinear analysis of face image ensembles. In Section 7.6, we present the recognition framework of the “few samples” problem, followed by experimental evaluation in Section 7.7, and finally, Section 7.8 concludes the chapter.

7.2 Problem Formulation: Friendly-Hostile paradigm

As mentioned previously, our training database is made up of friendly and hostile persons. Friendly people were asked to cooperate to generate a complete dataset meeting all the requirements of training data for tensor based recognition approaches (*i.e.* face images taken at all the representative lighting-viewpoint combinations and the lighting-viewpoints are exactly matched across all persons). However, the hostile persons are not cooperative and we have access to only a few images per person at varied lighting-viewpoints across persons. We denote the full dataset $D = D_F \cup D_H$, where D_F denotes the dataset containing the images of the friendly persons and D_H denotes the images of the hostile persons. The number of friendly persons is N_p^f and that of hostile persons is N_p^h and our aim is to devise a tensor based face recognition scheme that can classify images of all the $N_p^f + N_p^h$ persons (unbalanced dataset) or just the N_h hostile persons (small dataset) as the application requires.

7.3 Outline of the Proposed Approach

We frame our solution based on a novel synthesis perspective of tensor factorization. From the factorization of a full tensor, we obtain a core tensor and four factor spaces of *eigenpersons, eigenlightings, eigenviewpoints* and *eigenpixels*. The core tensor is equivalent to the *eigenvalues* as in the case of matrix decomposition and is similarly ordered with respect to the different *eigenfactors*. Once we multiply the core tensor with the pixel subspace we obtain an ordered set of *eigenimages* in the core tensor, termed as core-image tensor, which are the result of interaction between different *eigenpersons, eigenlightings* and *eigenviewpoints*. The images in the core-image tensor serve as the multilinear basis for face images. To obtain a face image of a person at a particular lighting and viewpoint from this multilinear basis, we need a set of three identity vectors: one that describes the person's identity and one each describing that particular lighting condition and viewpoint. These identity vectors are then multiplied in the respective modes of the core-image tensor to obtain the desired image, formulation for which stems directly from the way multilinear decomposition is performed. It is to be noted that these three identity vectors are completely independent of each other and they are uniquely associated with their corresponding entity. Assuming the *eigenimages* in the core-image tensor are general enough to represent the universe (or the subset we are interested in) of face images, we can build a synthesis framework which enables us to synthesize face images of any person at any lighting/viewpoint condition through a defined combination of these *eigenimages*. Accordingly, the unique identity vector associated with a persons identity is sufficient to completely define all the images at all possible conditions. We use the images of friendly people to construct a full tensor and then subsequently assume that the core-image tensor thus obtained is general enough to represent the images of the hostile people. The person identity vectors for friendly people are obtained directly from the factorization result. To obtain the person-identity vectors for the hostile persons, we formulate an optimization problem following the same synthesis framework. The optimization problem is solved using an ALS (Alternating Least Square) based algorithm and upon convergence we obtain the required identity vector of that person. With the core-image tensor and all the person-identity vectors of both the friendly and hostile persons, we use the MPCA-PS algorithm for recognition. The MPCA-PS algorithm is derived from the synthesis paradigm to suitably extend it for our approach.

7.4 Review: Tensor Model of Face Image Ensembles

We follow the tensor model as followed in the previous chapter and as described in Section 3.3, where the face image tensor $\mathcal{T}$ is given by,

$$\mathcal{T}_{(i_p,i_l,i_v)} = I_{P_{i_p},L_{i_l},V_{i_v}} \tag{7.4.1}$$

with the usual notation of $I_{P_{i_p},L_{i_l},V_{i_v}}$ being the image vector of i_p'th person at i_l'th lighting and i_v'th viewpoint. The Multilinear PCA decomposition of $\mathcal{T}$ is similarly followed from there and is represented as:

$$\mathcal{T} = \mathcal{S} \times_1 U^P \times_2 U^L \times_3 U^V \times_4 U^X \tag{7.4.2}$$

with the usual definition of $\mathcal{S}, U^P, U^L$, and U^V being followed. The core tensor $\mathcal{S}$ is not guaranteed to be diagonal, though the subtensors of $\mathcal{S}$ are ordered and mutually orthogonal. Subtensors are analogues to the concept of *eigenvalue* for matrix decomposition. The subtensor $\mathcal{S}_{(i_p=k_1,i_l=k_2,i_v=k_3,i_x=k_4)}$ is a scalar and corresponds to the k_1'th *eigenperson*, k_2'th *eigenlighting*, k_3'th *eigenviewpoint* and k_4'th *eigenpixel*. Let,

$$\mathcal{A} = \mathcal{S} \times_4 U^X \tag{7.4.3}$$

then in the tensor $\mathcal{A}(\in \mathbb{R}^{N'_p \times N'_l \times N'_v \times N_x})$ the subtensors remain ordered in the first, second and the third mode whilst the ordering vanishes in the pixel mode. A particular subtensor $\mathcal{A}_{(i_p=k_1,i_l=k_2,i_v=k_3)}$ is a vector ($\in \mathbb{R}^{N_x}$) and represents the image which corresponds to the k_1'th *eigenperson*, k_2'th *eigenlighting*, k_3'th *eigenviewpoint*. Tensor A is termed as the core-image tensor and serves as the foundation for the synthesis framework described in the next section.

7.5 Multilinear model for face image synthesis

From the Multilinear PCA decomposition of tensor $\mathcal{T}$ in (7.4.2), and following the rule of mode-multiplication we note that a particular element of $\mathcal{T}$ is obtained by,

$$\begin{aligned} \mathcal{T}_{i_p,i_l,i_v,i_x} &= \mathcal{S} \times_1 u_p^{i_p} \times_2 u_l^{i_l} \times_3 u_v^{i_v} \times_4 u_x^{i_x} \\ \Rightarrow \mathcal{T}_{i_p,i_l,i_v} &= \mathcal{S} \times_1 u_p^{i_p} \times_2 u_l^{i_l} \times_3 u_v^{i_v} \times_4 U^X \end{aligned} \tag{7.5.1}$$

where, the subtensor $\mathcal{T}_{i_p,i_l,i_v} \in \mathbb{R}^{N_x}$ is a vector and $u_p^{i_p}, u_l^{i_l}, u_v^{i_v}$ are i_p'th, i_l'th, and i_v'th rows of U^P, U^L, and U^V, respectively. From (7.4.1), we know that $\mathcal{T}_{i_p,i_l,i_v}$ represents the image of $I_{P_{i_p},L_{i_l},V_{i_v}}$, therefore, from (7.5.1):

$$I_{P_{i_p},L_{i_l},V_{i_v}} = \mathcal{S} \times_1 u_p^{i_p} \times_2 u_l^{i_l} \times_3 u_v^{i_v} \times_4 U^X \tag{7.5.2}$$

Substituting $\mathcal{A} = \mathcal{S} \times_4 U^X$, as defined in (7.4.3), we obtain,

$$I_{P_{i_p},L_{i_l},V_{i_v}} = \mathcal{A} \times_1 u_p^{i_p} \times_2 u_l^{i_l} \times_3 u_v^{i_v} \tag{7.5.3}$$

The above equation states that the eigenimages of $\mathcal{A}$ can be combined in a multilinear way to generate face images, where person identity ($u_p^{i_p}$), lighting conditions ($u_l^{i_l}$) and the viewpoint direction ($u_v^{i_v}$) can be varied independently to obtain images of different persons at desired lighting and viewpoints. To obtain a generalized description of (7.5.3), let us assume that the eigenimages of A are general enough to capture all the variations of the universe of face images under lighting and viewpoint variations only. Then, we can extend the span of u_p to $\mathbb{R}^{N'_p}$ and we would be able to generate an infinite number of distinct persons. Similarly we can assume $u_l \in \mathbb{R}^{N'_l}$ and $u_v \in \mathbb{R}^{N'_v}$ and would be able to generate face images at infinite number of distinct lighting and viewpoint combination. We denote u_p as person identity, u_l as lighting identity and u_v as viewpoint identity vectors. Now the generalized face image synthesis model can be written as:

$$I = \mathcal{A} \times_1 u_p \times_2 u_l \times_3 u_v \tag{7.5.4}$$

where I is a face image whose identity, lighting conditions and viewpoint directions are governed by the choice of person, lighting and viewpoint identity vectors. The rows of U^P, U^L and U^V serve as the reference for persons, lighting conditions and viewpoints respectively.

7.6 Recognition framework

We will use MPCA-PS as the base recognition algorithm. However, as described before, the algorithm in its current form can only be applied to complete datasets with large numbers of training samples and thus can only deal with the training data for friendly persons. To extend this to cover hostile persons with few training images, we need to develop a deeper understanding of its working principles. A similar effort was made in the previous chapter to formulate MPCA-PS as an optimization problem, however, the resultant formulation is still limited to the complete database scenario. Here we will analyse the algorithm based on our particular need of building a classifier for hostile persons with only a few training samples.

We start by factoring the full dataset D_F by Multilinear PCA and the decomposition is given as:

$$D_F = \mathcal{S}_F \times_1 U^P \times U^L \times_3 U^V \times_4 U^X \tag{7.6.1}$$

where $\mathcal{S}_F \in \mathbb{R}^{N_p'^f \times N_l' \times N_v' \times N_x'}$, $U^P \in \mathbb{R}^{N_p{}^f \times N_p'^f}, U^L \in \mathbb{R}^{N_l \times N_l'}, U^V \in \mathbb{R}^{N_v \times N_v'}$ and $U^X \in \mathbb{R}^{N_x \times N_x'}$. We can compute the core-image tensor as:

$$\mathcal{A}_F = \mathcal{S}_F \times_4 U^X \tag{7.6.2}$$

For a test image I_T belonging to the k'th friendly persons, the synthesis equation (7.5.4) tells us that I_T has been synthesized as:

$$I_T = \mathcal{A}_F \times_1 u_p^k \times_2 u_l^t \times_3 u_v^t \tag{7.6.3}$$

where, the u_p^k is the person-identity vector of the k'th friendly person (k'th row of U^P), u_l^t corresponding to the lighting conditions and u_v^t to the viewpoint of the test image. In reality, we do not know the identity of the test image, which necessitates computation of the k for I_T along with other synthesis parameters, u_l^t and u_v^t. To compute them we formulate an optimization problem as:

$$\min_{k,u_l^t,u_v^t} ||I_T - \mathcal{A}_F \times_1 u_p^k \times_2 u_l^t \times_3 u_v^t||^2 \tag{7.6.4}$$

where, $k = 1, ..., N_p^f$. The solution $k = k_{min}$ is the identity of the test image. Generally, test images are not always at referenced lightings or viewpoints, therefore, we cannot iterate over the sets of u_l^t (rows of U^L) and u_v^t (rows of U^V) to find the minima. We need to solve (7.6.4) over all possible values of u_l^t and u_v^t and to proceed let us denote:

$$\mathcal{B}^k = \mathcal{A}_F \times_1 u_p^k \, (\in \mathbb{R}^{1\times N_l'\times N_v'\times N_x}) \tag{7.6.5}$$

and now, (7.6.4) can be written as:

$$\min_k(\min_{u_l^t,u_v^t} ||I_T - \mathcal{B}^k \times_2 u_l^t \times_3 u_v^t||^2) \tag{7.6.6}$$

The inner optimization problem is a multilinear problem, which can be solved via ALS method. However, ALS is a slow method and there is no guarantee for convergence. Hence, we will first convert the inner multilinear optimization problem into a linear least square problem and then use standard matrix operations to solve that. Let us denote:

$$e_k = \min_{u_l^t,u_v^t} ||I_T - \mathcal{B}^k \times_2 u_l^t \times_3 u_v^t||^2 \tag{7.6.7}$$

By the definition of mode-multiplication we can write:

$$\begin{aligned} \mathcal{B}^k \times_2 u_l^t \times_3 u_v^t &= \textstyle\sum_{i_l} u_l^t(i_l) \sum_{i_v} u_v^t(i_v)\mathcal{B}^k_{1,i_l,i_v} \\ &= \textstyle\sum_{i_l} u_l^t(i_l)(u_v^t.B^k_{1,i_l}) \\ &= u_l \begin{bmatrix} u_v & ... & 0 \\ ... & ... & ... \\ 0 & ... & u_v \end{bmatrix} \begin{bmatrix} \mathcal{B}^k_{1,1} \\ ... \\ ... \\ \mathcal{B}^k_{1,N_l'} \end{bmatrix} \end{aligned} \tag{7.6.8}$$

$$\Rightarrow \mathcal{B}^k \times_2 u_l^t \times_3 u_v^t \quad = \quad g(u_l^t, u_v^t).B^k \tag{7.6.9}$$

where, $\mathcal{B}^k_{1,i_l}$ is the $(1, i_l)$'th subtensor of $\mathcal{B}^k$, $g : \mathbb{R}^{N'_l} \times \mathbb{R}^{N'_v} \to \mathbb{R}^{N'_l N'_v}$ is a multilinear function over u_l^t, u_v^t and $B^k (\in \mathbb{R}^{N'_l N'_v \times N_x})$ is a matrix constructed from the subtensors of $\mathcal{B}$. From the deductions of (7.6.9), (7.6.7) can be expressed as:

$$e_k = \min_{g(u_l^t, u_v^t)} ||I_T - g(u_l^t, u_v^t) B^k||^2 \tag{7.6.10}$$

The least square solution of e_k is given by,

$$e_k = ||I_T - I_T (B^k)^+ B^k||^2 \tag{7.6.11}$$

where, $^+$ implies *pseudoinverse*. Hence, the original optimization problem of (7.6.6) can be written as:

$$\min_k ||I_T - I_T (B^k)^+ B^k||^2 \tag{7.6.12}$$

The solution $k = k_{min}$ is the desired person index of the test image. Essentially, the matrix B^k is same as the matrix related to the *person-specific eigenmode space* described in chapter 5, and the formulation of the recognition problem is same as the MPCA-PS algorithm proposed there. However, this derivation helps us to clearly identify the key requirements for this algorithm: 1) a core-image tensor $\mathcal{A}_F$ and 2) person-identity vectors u_p^k for all persons in our database, to calculate the B^k using (7.6.5) and (7.6.9). The recognition method described in (7.6.12) is not applicable to the hostile persons as we have neither their core-image tensor defined, nor have their person-identity vectors.

The core-image tensor is composed of the images of different *eigenpersons* at various *eigenlightings* and *eigenviewpoints*. For friendly persons, we assumed that necessary images were taken at different lightings and viewpoints to compute sufficiently general subspaces of lightings and viewpoints. Therefore, the subspaces of eigenlightings and eigenviewpoints should remain the same for hostile persons too. To make the core-image tensor $\mathcal{A}_F$ applicable to hostile persons, the only assumption we need to make is that the space of eigenpersons are general enough to represent even the hostile persons, which is true if we have sufficient variation among the friendly persons. Assuming sufficient variation in our friendly datasets, $\mathcal{A}_F$ can be considered as the core-image tensor for both the hostile persons and friendly persons. With this, the only missing part to employ the recognition scheme of MPCA-PS is the person-identity vectors for the hostile persons. Next, we discuss how to compute the person-identity vector of a hostile person from their set of training images.

7.6.1 Computing person-identity vectors from a set of training images

Let I is an image of the hostile person, then the synthesis equation (7.5.4) implies that,

$$I = \mathcal{A}_F \times_1 u_p \times_2 u_l \times_3 u_v \tag{7.6.13}$$

where u_p is the desired person-identity vector. u_p can be obtained by solving the following optimization equation:

$$\min_{u_p, u_l, u_v} ||I - \mathcal{A}_F \times_1 u_p \times_2 u_l \times_3 u_v||^2 \tag{7.6.14}$$

For a set of images $\{I_m\}_{m=1}^{M}$, the integrated optimization problem can be expressed as:

$$\min_{u_p, u_l^m, u_v^m} \sum_m ||I_m - \mathcal{A}_F \times_1 u_p \times_2 u_l^m \times_3 u_v^m||^2 \tag{7.6.15}$$

In the above problem, we kept the person-identity vectors (u_p) the same for all images, however, we allow each image to have different lighting (u_l^m) and viewpoint (u_v^m). To solve this problem, we use the Alternating Least Squares method. The number of free variables in this problem is $2M + 1$, therefore, we have $2M + 1$ numbers of single-variable least square problems. These are:

$$\begin{gathered}
\min_{u_p} \sum_m ||I_m - \mathcal{A}_F \times_1 u_p \times_2 u_l^m \times_3 u_v^m||^2 \\
\min_{u_l^m} ||I_m - \mathcal{A}_F \times_1 u_p \times_2 u_l^m \times_3 u_v^m||^2 \text{ for } m = 1, ..., M \\
\min_{u_v^m} ||I_m - \mathcal{A}_F \times_1 u_p \times_2 u_l^m \times_3 u_v^m||^2 \text{ for } m = 1, ..., M
\end{gathered} \tag{7.6.16}$$

To solve for u_p let us denote:

$$J_P = \sum_m ||I_m - \mathcal{A}_F \times_1 u_p \times_2 u_l^m \times_3 u_v^m||^2 \tag{7.6.17}$$

The only variable unknown here is u_p. Let $\mathcal{A}_F \times_2 u_l^m \times_3 u_v^m = W_P^m$. $W_P^m \in \mathbb{R}^{N_p' \times 1 \times 1 \times N_x}$. It is easy to note that $W_P^m \times_1 u_p = u_p.W_P^m$. With this (7.6.17) can be written as:

$$J_P = \sum_m ||I_m - u_p.W_P^m||^2 \tag{7.6.18}$$

Equating $\frac{dJ_P}{du_p}$ to zero we obtain:

$$\begin{aligned} \frac{dJ_P}{du_p} &= \frac{d}{du_p} \textstyle\sum_m ||I_m - u_p.W_P^m||^2 & (7.6.19) \\ \Rightarrow \ 0 &= \frac{d}{du_p} \textstyle\sum_m (I_m - u_p.W_P^m)(I_m - u_p.W_P^m)^T \\ &= -2\sum_m I_k(W_P^m)^T + 2u_p \sum_m W_P^m (W_P^m)^T & (7.6.20) \\ \Rightarrow \ u_p &= (\sum_m I_m(W_P^m)^T).(\sum_m W_P^m(W_P^m)^T)^+ \end{aligned}$$

To solve the u_l^m, let us denote the objective function as:

$$J_L^m = ||I_m - \mathcal{A}_F \times_1 u_p \times_2 u_l^m \times_3 u_v^m||^2 \tag{7.6.21}$$

The variables u_p and u_v^k are fixed. We denote a constant $W_L^m = \mathcal{A}_F \times_1 u_p \times_3 u_v^m$, where $W_L^m \in \mathbb{R}^{1 \times N_l' \times 1 \times N_x}$. It can be easily shown that $W_L^m \times_2 u_l^m = u_l^k.W_L^m$, by suppressing the scaler dimensions. Hence (7.6.21) can be re-written as:

$$J_L^m = ||I_m - u_l^m.W_L^m||^2 \tag{7.6.22}$$

The minima for this least square solution is well-known and is:

$$u_l^m = I_m.(W_L^m)^+ \tag{7.6.23}$$

In a similar fashion, we can derive the optimal values for u_v^m as:

$$u_v^m = I^m.(W_V^m)^+ \tag{7.6.24}$$

where, $W_V^m = \mathcal{A}_F \times_1 u_p \times_2 u_l^m$.

We start from some random values for u_p, u_l^m and u_v^m and solve the $2M + 1$ optimization problems of (7.6.16) sequentially. This is iterated until the values converge or a pre-set number of iterations is reached. A concise description of the solution is presented in Algorithm 8.

Algorithm 8: Algorithm to compute person-identity vector from a set of images

Input: core-image tensor $\mathcal{A}_F$ and training set for a hostile person $\{I_m\}_{m=1}^{M}$.

Output: person-identity vector of this person u_p.

Set u_l^m and u_v^m to some random values. Iterate the following till convergence:

- $u_p = (\sum_m I_m (W_P^m)^T).(\sum_m W_P^m (W_P^m)^T)^+$, where

  $W_P^m = A_F \times_2 u_l \times_3 u_v$

- For $m = 1, ..., M$
  - $u_l^m = I^m.(W_L^m)^+$, where $W_L^m = \mathcal{A}_F \times_1 u_p \times_3 u_v^m$
  - $u_v^m = I^m.(W_V^m)^+$, where $W_V^m = \mathcal{A}_F \times_1 u_p \times_2 u_l^m$

7.6.2 Recognition

First, we calculate the person-identity vectors of all the hostile persons using Algorithm 1. Then, the core-image tensor along with the person-identity vectors of the friendly and hostile persons are used to compute $\{B^k\}_{k=1}^{N_p^f+N_p^h}$ following (7.6.5) and (7.6.9). We follow the recognition Eq. (7.6.12) to build the classifier for the hostile persons only as in the case of small dataset problem. The algorithm is presented in Algorithm 9.

Similarly, for the unbalanced dataset scenario, we build the classifier for all the friendly and hostile persons and the algorithm is presented in Algorithm 10.

Algorithm 9: Recognition algorithm for small training dataset

Input: Test image I^T, core-image tensor $\mathcal{A}_F$ and the set of person-identity vectors $\{u_p\}_{k=1}^K$ for $K = N_p^h$ persons.

Output: Identity of the test image.

Precalculate: $\mathcal{B}^k = \mathcal{A}_F \times_1 u_p^k$ and then $B^k = [\mathcal{B}_{1,1}^k ... \mathcal{B}_{1,N_l'}^k]^T$

- For every $k = 1, ..., K$
  - Compute, $e_k = ||I_T(E - (B^k)^+ B^k)||^2$
- Find $k = k_{min}$ for which e_k is the smallest. Output k_{min}.

Algorithm 10: Recognition algorithm for unbalanced dataset

Input: Test image I^T, core-image tensor $\mathcal{A}_F$ and the set of person-identity vectors $\{u_p\}_{k=1}^K$ for $K = N_p^f + N_p^h$ persons.

Output: Identity of the test image.

Precalculate: $\mathcal{B}^k = \mathcal{A}_F \times_1 u_p^k$ and then $B^k = [\mathcal{B}_{1,1}^k ... \mathcal{B}_{1,N_l'}^k]^T$

- For every $k = 1, ..., K$
  - Compute, $e_k = ||I_T(E - (B^k)^+ B^k)||^2$
- Find $k = k_{min}$ for which e_k is the smallest. Output k_{min}.

From the analysis of the proposed approach we can list the following useful properties of our proposed solution:

- Once we have a general enough core-image tensor learnt from the friendly persons, the addition of hostile persons can be performed incrementally without changing the existing classifier, which is desirable when retention of past training data for future updates is not feasible due to the large size of the data.
- All persons, irrespective of the number of training images, are represented by only one person-identity vector. Therefore, in our approach, recognition for hostile persons is not affected by any bias that results from large training samples of the friendly persons.

7.7 Experiments and Analysis

We use YaleB (Georghiades *et al.*, 2001a), Extended YaleB (Lee *et al.*, 2005b), Weizmann and PEAL (Gao *et al.*, 2008) face databases for experiments. As described in the previous chapters, the Extended YaleB database contains images of 38 persons at 64 different lightings and at 9 different viewing directions, whilst the YaleB Frontal dataset has images at the frontal viewpoint only. In the Weizmann face database, we have 28 persons at 5 different viewpoints ($0°, \pm 17°, \pm 34°$), 3 different lightings and 2 different expressions. In the PEAL Lighting variation dataset we have images of 100 persons at different lighting conditions. The number of available images per person lies between 6-15. Prior to the experiments, all images were cropped, their eye-points aligned manually and the image vectors normalized to unity. For HOSVD and other tensor operations, we used the tensor toolbox developed by Kolda *et. al.* in MATLAB^{TM} (Bader and Kolda, 2007). In all experiments we compare the performance of our approach with PCA (Turk and Pentland, 1991b), Fisherface (PCA+LDA) (Belhumeur *et al.*, 1997) and LPP (He and Niyogi, 2003). Experiments were performed using Matlab^{TM}, running on a Intel Xeon 8$\times$2.3 GHz server with 16GB of RAM.

7.7.1 ALS Convergence Study:

In this experiment we study the rate of ALS convergence for the estimation of person-index vectors from the hostile persons. It is of paramount importance to estimate the person-index vector correctly from the few training images available to achieve good recognition performance. However, previous studies show that whilst ALS is the simplest iterative method for multi-variate optimization, it can be extremely slow to converge. We study the rate of convergence for both YaleB Frontal (two variables: person-index and lighting-index vectors) and Extended YaleB (three variables: person-index, lighting-index and viewpoint-index vectors) to obtain insights on the lower bound of iteration numbers to estimate the various index vectors within a satisfactory error tolerance, and more importantly to achieve good recognition performance on the hostile persons test images.

For experiments on YaleB frontal dataset, the whole dataset of 38 persons is randomly partitioned into 25 friendly and 13 hostile persons. The friendly persons are represented by all the 64 images at different lighting conditions and the core tensor is computed. The hostile person training dataset is constructed with only 3 images at selected lighting conditions. The test-set for hostile persons is created with all the 64 images of the 13 hostile persons. Figure 7.1 shows the average of consecutive estimation difference for the person-index, lighting-index and viewpoint-index vectors, averaged over all the persons and number of training images with respect to iteration numbers. As the plot shows the consecutive estimation difference reduces drastically at the beginning and slows down later, however, the difference goes well below 10^{-3} after iteration 35, which is satisfactory considering that the recognition performance has already achieved saturation after around 20 iterations. A similar experiment is performed on the Extended YaleB dataset and we see a similar trend (Figure 7.2), where the consecutive estimation error for the index vectors falls below 10^{-3} after 40 iterations and the recognition performance flattens out after 25 iterations. For the Extended YaleB dataset we needed few more iterations to achieve convergence similar to YaleB frontal dataset as the number of vectors to be estimated is one more. Therefore, from this experiment we conclude that if we set the iterations number at a suitably high number, we can assume satisfactory performance from the ALS algorithm for the estimation of the person-index vectors of the hostile persons. For all the following experiments, we will set the iterations number at 200 for all the datasets. We assume that training can take as much time as needed, and our primary aim is to obtain convergence of the ALS algorithm for the satisfactory estimation of the person-index vectors.

7.7.2 Evaluating Small Dataset Performance:

In this experiment we evaluate the performance of our algorithm for application to face recognition from few samples on YaleB frontal lighting variation dataset, Extended YaleB dataset with lighting and viewpoint variation, Weizmann dataset with

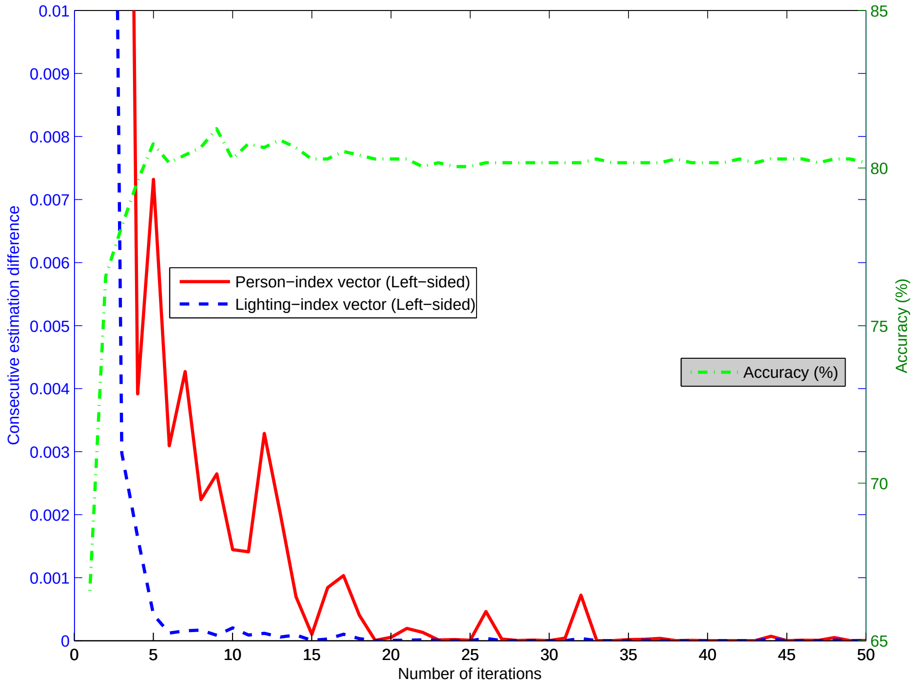

Figure 7.1: Experimental results on YaleB Frontal database for ALS convergence study. The whole database was randomly partitioned into 25 friendly and 13 hostile persons. 25 Friendly persons were represented by all 64 images at different lighting conditions, and 13 hostile persons were represented with only 3 images per person at selected lighting conditions. The plot shows the average of consecutive estimation differences for the person-index, lighting-index and viewpoint-index vectors. Also the accuracy on the hostile-persons test set, when the classifier was built for all the 38 persons (Friendly+Hostile).

lighting, viewpoint and expression variation and PEAL lighting variation dataset. The recognition performance is compared with that of PCA, LDA and LPP based face recognition algorithms.

For the YaleB frontal dataset having 38 persons, we perform the experiment with 25 people as friendly and the other 13 people as hostile persons. The friendly dataset is created with all the images of the friendly persons (64 images for each person), and is used to learn the core-image tensor. For hostile persons, only a few images at selected lighting conditions are used for training and the whole dataset is used

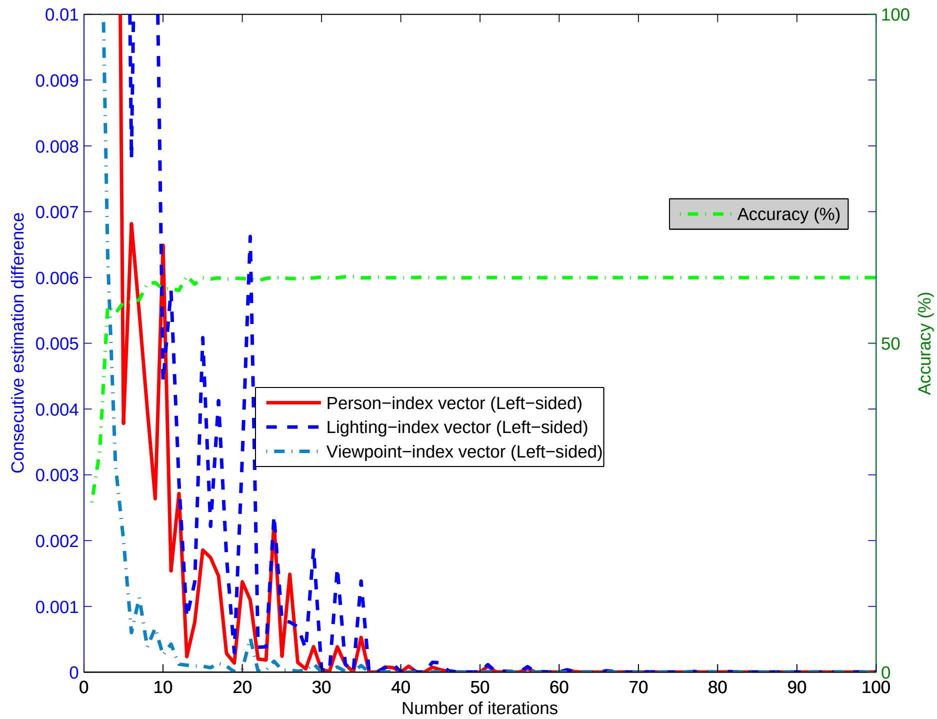

Figure 7.2: Experimental results on Extended YaleB database for ALS convergence study. The whole database was randomly partitioned into 25 friendly and 13 hostile persons. 25 Friendly persons were represented by all 576 images, and 13 hostile persons were represented with only 9 images per person at ambient lighting and selected viewpoints. The plot shows the average of consecutive estimation differences for the person-index, lighting-index and viewpoint-index vectors. Also the accuracy on the hostile-persons test set, when the classifier was built for all the 38 persons (Friendly+Hostile).

for testing (64 images for each person). The experiment is repeated 50 times with different combinations of friendly and hostile people and the average recognition rate for only the hostile people is shown in Table 7.1.

No. of training images	***Recognition method***			
	PCA	**PCA +LDA**	**LPP**	**Proposed**
1	40.77	-	-	**72.68**
3	44.42	61.03	61.78	**80.57**
5	52.65	65.18	70.25	**83.90**
7	66.83	69.88	77.46	**86.31**

Table 7.1: Experimental results on YaleB frontal lighting variation dataset. Testing dataset contains 13 persons, whilst separate 25 persons are used for learning core-image tensor.

Similar experiments are performed on Extended YaleB dataset with 25 Friendly-13 Hostile persons. The friendly dataset is created with all the images of the friendly persons (576 images for each person), and is used to learn the core-image tensor. For hostile persons, only a few images at selected viewpoints and lighting conditions are used for training and the whole dataset is used for testing (576 images for each person). The experiment is repeated 50 times with different combinations of friendly and hostile people and the average recognition rate for the hostile persons are presented in Table 7.2.

For the Weizmann dataset, we divided the dataset into 13 friendly and 15 hostile persons. The friendly dataset is created with all the images of the friendly persons (30 images for each person), and is used to learn the core-image tensor. For hostile persons, only a few images at selected viewpoints and lighting conditions are used for training and the whole dataset is used for testing (30 images for each person). The experiment is repeated 50 times with different combinations of friendly and hostile people and the average recognition rate for the hostile persons are presented in Table 7.3.

No. of training images	***Recognition method***			
	PCA	**PCA +LDA**	**LPP**	**Proposed**
1	31.42	-	-	**49.29**
3	46.58	52.65	54.22	**66.29**
5	49.73	64.39	61.50	**69.60**
7	54.66	70.61	67.49	**73.10**

Table 7.2: Experimental results on Extended YaleB dataset. Testing dataset contains 13 persons, whilst separate 25 persons are used for learning core-image tensor.

For the PEAL dataset, we divided the dataset of 100 persons into 28 friendly, who have 15 images each at matching lighting conditions and 72 hostile persons, who have 9 images each at different lighting conditions. The friendly dataset is only used to learn the core-image tensor. For hostile persons, only a few images at selected lighting conditions are used for training and the whole dataset is used for testing (9 images for each person). The recognition rate for the hostile persons are presented in Table 7.4.

7.7.2.1 Observations:

As can be seen from the results our proposed approach performs well across all the datasets and is better than the LDA or LPP based methods, specifically, when there is only lighting variation as in the YaleB frontal lighting variation dataset in Table 7.1 and PEAL lighting variation dataset, in Table 7.4. For the PEAL dataset, we also note that we have 72 persons in the small dataset, whilst only 28 persons were available for learning the generic manifold, which shows that our approach has the potential to handle large datasets. For experiments on Extended YaleB dataset with lighting and viewpoint variation, as in Table 7.2, though we have superior results than other methods, we needed more training images to achieve an

No. of training images	***Recognition method***			
	PCA	**PCA +LDA**	**LPP**	**Proposed**
1	70.41	-	-	**83.14**
2	74.27	73.83	75.80	**88.93**

Table 7.3: Experimental results on Weizmann dataset. Testing dataset contains 15 persons, whilst separate 13 persons are used for learning core-image tensor.

No. of training images	***Recognition method***			
	PCA	**PCA +LDA**	**LPP**	**Proposed**
1	51.70	-	-	**73.76**
2	57.72	77.62	71.75	**81.17**

Table 7.4: Experimental results on PEAL dataset. Testing dataset contains 72 persons, whilst separate 28 persons are used for learning core-image tensor.

acceptable level of recognition accuracy compared to the YaleB lighting variation dataset. It is to be noted that for this dataset the lighting variation is extreme and coupled with viewpoint variation, this represents a challenging dataset. To illustrate, in the experiments of Weizmann dataset (Table 7.3), though we have lighting, viewpoint and expression variation, we were able to achieve much better levels of accuracy even with smaller numbers of training images, which is evidently due to the absence of extreme variations in any of the modes.

7.7.3 Evaluating unbalanced dataset performance

In this experiment we evaluated the performance of our algorithm in unbalanced dataset scenarios. For the YaleB frontal dataset we assumed that among the total

38 persons, we have access to all the images for 25 persons, however, only 5 images are available for the other 13 persons. We performed two experiments: in Scenario 1 - we showed the recognition performance when only the 5 persons are needed to be classified and in Scenario 2 - we showed the recognition performance when the classifier needs to classify all the 38 persons. The testing dataset is the whole dataset. For the 25 persons, the testing set is a proper subset of the training set, therefore, they will be accurately recognised. However, this is not the case for the remaining 13 persons, who are represented in training data with only 5 samples each and we tabulate recognition results for this group only. Therefore, for Scenario 1, we will have only 13 people, each having 5 training samples and the classifier built for only them, which represents a balanced dataset. However, in Scenario 2, 25 persons having large training images (64 each) and 13 persons having only 5 images each, resulting in an unbalanced dataset scenario. The recognition performance is presented in Table 7.5. We perform a similar experiment on the Weizmann dataset

	Recognition method			
	PCA	**PCA +LDA**	**LPP**	**Proposed**
Scenario 1	58.55	73.46	78.35	**85.44**
Scenario 2	32.08	48.71	50.03	**81.50**

Table 7.5: Experimental results on YaleB frontal lighting variation dataset. 25 Persons are represented with full training set and 13 people are represented with only 5 images at selected lighting conditions. Testing is performed on all the available images, and the result is shown only for the 5 added persons in two scenarios: Scenario 1 - standalone classifier for the added 13 people only and Scenario 2: classifier for all the 38(25+13) persons.

with 13 persons having a full training set and the remaining 15 persons with only 2 images available per person. The result is tabulated in a similar way with the recognition accuracy reported for only the 15 persons with small training set in two scenarios: Scenario 1 - the classifier is built only for these 15 persons, and Scenario 2 - the classifier is built for all the 28 persons, making it an unbalanced dataset scenario (Table 7.6).

	Recognition method			
	PCA	**PCA +LDA**	**LPP**	**Proposed**
Scenario 1	72.23	70.56	74.18	**87.39**
Scenario 2	34.66	44.82	47.13	**85.32**

Table 7.6: Experimental results on Weizmann dataset. 13 Persons are represented with full dataset and 15 people are represented with only 2 images at selected lighting, viewpoint and expression conditions. Testing is performed on all the available images, and the result is shown only for the 5 added persons in two scenarios: Scenario 1 - standalone classifier for the added 15 people only and Scenario 2: classifier for all the 28(13+15) persons.

7.7.3.1 Observations:

As can be seen from the results in Tables 7.5 – 7.6, our approach achieves superior performance, and is least affected by unbalance in the dataset. The performance drop due to unbalance, as can be seen by comparing figures within the columns, is only 2-3% for our method, whilst being >20% for other methods. This shows that even if some people are omitted during the training image collection exercise, our approach achieves good recognition performance.

7.8 Closing Remarks

In this chapter, we address the inapplicability of the previously developed tensor based face recognition methods such as MPCA-JS or MPCA-PS to real-world scenarios where collecting training images is challenging. Those methods inherently assume a co-operative world wherein a full training database containing a number of images per persons are available at all specified lighting and viewpoints. Whilst this is true in some situations, collecting images is a problem in situations such as law-enforcement, where only a few images are available for the persons of interest.

To address this issue we provided a new paradigm of "friendly-hostile", in which we collect as many images as needed from the friendly, co-operative persons, wherein only a few images are available for hostile persons. The whole database of friendly-hostile is an incomplete database and a full tensor cannot be created, and therefore, previous methods cannot be applied. We solve this problem by learning a generalized face images structure from the images of friendly database using a novel multilinear synthesis perspective and then utilize it to compensate for the low number of images for the hostile persons. Experimental results on YaleB frontal, Extended YaleB, Weizmann and PEAL datasets show the superiority of the proposed approach, both in terms of accuracy and speed against the leading methods such as Fisherface and LPP.

The solution is also significant when considered from the perspective of unbalanced datasets. As the ratio of number of images between friendly and hostile people are highly skewed favouring the friendly people, strong apriori bias is created towards them. Standard approaches such as Eigenface, Fisherface, LPP find it difficult to correctly classify the test images belonging to the hostile persons due to this strong bias favouring the friendly people. However, our method is not affected by such bias, and thus does not face any difficulty in classification as demonstrated in our experiments.

Chapter 8

Conclusion

8.1 Summary

This thesis has presented an investigation into multilinear modelling of face image ensembles and its application in face recognition. We present two face recognition algorithms for conventional applications, a framework of face recognition in multilinear analysis paradigm and a specialized algorithm for application in "few samples" scenarios. The applications are built through proper interpretation and analysis of multilinear models for face images.

We start by presenting a novel face recognition algorithm in Chapter 4, primarily to address the limitations of the existing multilinear based face recognition algorithms that fail to handle test faces at unseen conditions. We propose our algorithm (MPCA-JS) based on the properties of core tensor and noting its importance as an entity that contains information on variations in all the factors (person identities, lighting conditions, viewpoints etc.). From the core tensor we construct a new basis of *multilinear eigenmodes* and use that in the recognition algorithm for face image representation and comparison.

In chapter 5, we propose an efficient algorithm (MPCA-PS) that is able to provide similar or better performance than MPCA-JS at a fraction of the computational cost. The improvement in efficiency is obtained by removing the redundancies present in the previous algorithm in the way it treats the person identity space. Instead of treating identity space as a generating space like lighting or viewpoint, we limit our focus to only the identity vectors of the known persons and reduce the order of representational basis by significant order. The procedure for recognition is altered suitably to obtain an efficient algorithm that is found to be significantly faster, especially on large datasets.

In chapter 6, we present a unified framework of face recognition based on multilinear analysis. From the factor analysis principle we develop an optimization equation that defines the way a face image is related to all the influencing factors. We then show that all the above algorithms are built on implicitly solving this same optimization problem by different means. We discuss the way they differ in their approach and their relative strengths and weaknesses in different face recognition scenarios. With this our understanding about the multilinear models and analysis is completed. In this chapter we also test our proposed algorithms against the leading face recognition algorithms, and demonstrate the superiority of our algorithm MPCA-PS, in terms of both recognition accuracy and test time.

Next in chapter 7, we apply our understanding to solve the “few samples” issue, as encountered in many real-world face recognition applications. Whilst most of the traditional face recognition algorithms implicitly assume the availability of a large number of training samples to be robust, in many real-world situations, especially in law enforcement application, gathering large training sets for criminals is impossible and only a few images per person may be available. Traditional algorithms fail comprehensively in such scenarios. We propose a new algorithm based on the same multilinear framework that is shown to provide good performance even when only 1-3 images per person are used for training. We propose a novel “friendly-hostile” framework, where a large training set of “friendly” people are used to compensate fully the lack of training samples of the “hostile” people. We also discuss “few samples” issue from the perspective of unbalanced datasets, wherein some people may be under-represented than others in the training set. Experiments on unbalanced datasets demonstrate that our algorithm does not get affected by any such imbalance in bias and produces consistent performance in all situations.

8.2 Future work

Opportunities for future works lie in several directions. From the theoretical stand point, we envisage introducing non-linearity into the multilinear framework. Essentially what we took advantage of the multilinear analysis, is the ability to decompose face images into its constituent factors. However, the factor spaces are only linear

Euclidean vector spaces and for many factors, such as viewpoint or expression, the linearity assumption is stretched when large variations are present. It may be possible to model the individual factor spaces as some form of non-linear space based on kernel extension or metric modification via manifold learning. That way we can retain advantage of the multi-factor framework as developed in this thesis, and at the same time taking advantage of the non-linearity in the factor spaces.

Another area that needs some attention is the scalability issue of factorization. For application in big datasets having thousands of persons, it may be computationally infeasible to factorize a large tensor using the conventional Tensor SVD algorithm. Specialized factorization algorithms having the ability to handle large tensors need to be formulated for such scenarios.

In this whole thesis we have emphasized the use of pixel values as the preferred features, however, many recent developments suggest other types of features can be equally powerful for recognition purposes. It will be interesting to investigate different features and interpret the factorization result based on the new features and discover their relative advantages over the pixel values.

Bibliography

A. Hyvrinen, J. K. and Oja, E. (2001). *Independent Component Analysis*. John Wiley & Sons.

Acar, E. and Yener, B. (2009). Unsupervised multiway data analysis: A literature survey. *IEEE Trans. on Knowl. and Data Eng.*, **21**(1), 6–20.

Acar, E., Camtepe, S. A., Krishnamoorthy, M. S., and Yener, B. (2005). Modeling and multiway analysis of chatroom tensors. In *IEEE International Conference on Intelligence and Security Informatics*, pages 256–268. Springer, Germany.

Achermann, B. and Bunke, H. (2000). Classifying range images of human faces with hausdorff distance. In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, volume 2, pages 809–813 vol.2.

Achermann, B., Jiang, X., and Bunke, H. (1997). Face recognition using range images. In *VSMM '97: Proceedings of the 1997 International Conference on Virtual Systems and MultiMedia*, page 129, Washington, DC, USA. IEEE Computer Society.

Aguilar-Torres, G., Sanchez-Perez, G., Nakano-Miyatake, M., and Perez-Meana, H. (2007). Face recognition algorithm using the discrete gabor transform. In *Electronics, Communications and Computers, 2007. CONIELECOMP '07. 17th International Conference on*, pages 35–35.

Ashraf, A., Lucey, S., and Chen, T. (2008). Learning patch correspondences for improved viewpoint invariant face recognition. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8.

Bader, B., Harshman, R., and Kolda, T. (2006). Temporal analysis of social networks using three-way dedicom. Technical Report SAND2006-2161, Sandia National Laboratories.

Bader, B. W. and Kolda, T. G. (2007). Tensor toolbox version 2.2 copyright 2007, sandia national laboratories, available at *http://csmr.ca.sandia.gov/∼tgkolda/TensorToolbox/*.

Bartlett, M. S. (2001). *Face Image Analysis by Unsupervised Learning.* Kluwer Academic Publishers, Norwell, MA, USA.

Bartlett M.S., M. J. and T.J., S. (Nov 2002). Face recognition by independent component analysis. *Neural Networks, IEEE Transactions on*, **13**(6), 1450–1464.

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

Belkin, M. and Niyogi, P. (2001). Locality preserving projections. In *Advances in Neural Information Processing Systems 15.* MIT Press.

Belkin, M. and Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, **15**(6), 1373–1396.

Bell, A. J. and Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, **7**(6), 1129–1159.

Beymer, D. and Poggio, T. (1995). Face recognition from one example view. *Computer Vision, IEEE International Conference on*, **0**, 500.

Blanz, V. and Vetter, T. (2003). Face recognition based on fitting a 3d morphable model. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **25**(9), 1063–1074.

Blanz, V., Romdhani, S., and Vetter, T. (2002). Face identification across different poses and illuminations with a 3d morphable model. In *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, pages 192–197.

Bledsoe, W. W. (1964). The model method in facial recognition. Technical report, Tech. Report PRI:115, Panoramic Research Inc., Palo Alto, CA.

Bowyer, K. W., Chang, K., and Flynn, P. (2006). A survey of approaches and challenges in 3d and multi-modal 3d + 2d face recognition. *Computer Vision and Image Understanding*, **101**(1), 1 – 15.

Bruce, K. B., Draper, B. A., Beveridge, J. R., and She, K. (2002). Pca vs. ica: A comparison on the feret data set. In *Joint Conference on Information Sciences*, pages 824–827.

Buhmann, J., Lades, M., and von der Malsburg, C. (1990). Size and distortion invariant object recognition by hierarchical graph matching. In *Neural Networks, 1990., 1990 IJCNN International Joint Conference on*, pages 411–416 vol.2.

Chang, K. I., Bowyer, K. W., and Flynn, P. J. (2003). Face recognition using 2d and 3d facial data. In *ACM Workshop on Multimodal User Authentication*, pages 25–32.

Chen, L.-F., Liao, H.-Y. M., Ko, M.-T., Lin, J.-C., and Yu, G.-J. (2000). A new lda-based face recognition system which can solve the small sample size problem. *Pattern Recognition*, **33**(10), 1713 – 1726.

Chen, S., Zhang, D., and Zhou, Z.-H. (2004). Enhanced (pc)2 a for face recognition with one training image per person. *Pattern Recogn. Lett.*, **25**(10), 1173–1181.

Chen, W., Er, M. J., and Wu, S. (2005). Pca and lda in dct domain. *Pattern Recognition Letters*, **26**(15), 2474 – 2482.

Chen, W., Er, M. J., and Wu, S. (2006). Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, **36**(2), 458–466.

Cheng, J., Liu, Q., Lu, H., and Chen, Y.-W. (2005). Supervised kernel locality preserving projections for face recognition. *Neurocomputing*, **67**, 443 – 449. Geometrical Methods in Neural Networks and Learning.

Chu, R., Liao, S., and Zhang, L. (2007). Illumination invariant face recognition using near-infrared images. *IEEE Trans. Pattern Anal. Mach. Intell.*, **29**(4), 627–639. Senior Member-Li, Stan Z.

Comon, P. (1989). Separation of stochastic processes. In *Higher-Order Spectral Analysis, 1989. Workshop on*, pages 174–179.

Comon, P. (1994). Independent component analysis, a new concept? *Signal Process.*, **36**(3), 287–314.

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

Cox, I. J., Ghosn, J., Yianilos, P. N., and com, p. n. n. (1996). Feature-based face recognition using mixture-distance. In *CVPR '96: Proceedings of the 1996 Conference on Computer Vision and Pattern Recognition (CVPR '96)*, page 209, Washington, DC, USA. IEEE Computer Society.

De Lathauwer, L., Moor, B. D., and Vandewalle, J. (2000). A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications*, **21**(4), 1253–1278.

del Solar, J. R. and Navarrete, P. (2005a). Eigenspace-based face recognition: a comparative study of different approaches. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, **35**(3), 315–325.

del Solar, J. R. and Navarrete, P. (2005b). Eigenspace-based face recognition: a comparative study of different approaches. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, **35**(3), 315–325.

Draper, B. A., Baek, K., Bartlett, M. S., and Beveridge, J. R. (2003). Recognizing faces with pca and ica. *Computer Vision and Image Understanding*, **91**(1-2), 115 – 137. Special Issue on Face Recognition.

Duda, R. O. and Hart, P. E. (1973). *Pattern classification and scene analysis.* A Wiley-Interscience Publication, New York: Wiley, 1973.

El-Arief, T. I., Nagaty, K. A., and El-Sayed, A. S. (2007). Eigenface vs. spectroface: a comparison on the face recognition problems. In *SPPR'07: Proceedings of the Fourth conference on IASTED International Conference*, pages 321–327, Anaheim, CA, USA. ACTA Press.

Eldén, L. and Savas, B. (2009). A newtin-grassmann method for computing the best multilinear rank -(r_1, r_2, r_3) approximation of a tensor. *In communication.*

Er, M. J., Wu, S., and Lu, J. (1999). Face recognition using radial basis function (rbf) neural networks. In *Decision and Control, 1999. Proceedings of the 38th IEEE Conference on*, volume 3, pages 2162–2167 vol.3.

Er, M. J., Chen, W., and Wu, S. (2005). High-speed face recognition based on discrete cosine transform and rbf neural networks. *Neural Networks, IEEE Transactions on*, **16**(3), 679–691.

Etemad, K. and Chellappa, R. (1997). Discriminant analysis for recognition of human face images. *Journal of the Optical Society of America A*, **14**, 1724–1733.

Feng, G., Hu, D., Zhang, D., and Zhou, Z. (2006). An alternative formulation of kernel lpp with application to image recognition. *Neurocomputing*, **69**(13-15), 1733 – 1738. Blind Source Separation and Independent Component Analysis - Selected papers from the ICA 2004 meeting, Granada, Spain, Blind Source Separation and Independent Component Analysis.

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

Fisher, R. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, (7), 179–188.

Gao, W., Cao, B., Shan, S., Chen, X., Zhou, D., Zhang, X., and Zhao, D. (Jan. 2008). The cas-peal large-scale chinese face database and baseline evaluations. *Systems, Man and Cybernetics, Part A, IEEE Transactions on*, **38**(1), 149–161.

Gao, Y. and Leung, M. (2002). Face recognition using line edge map. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **24**(6), 764–779.

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

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

Gong, K. J. S. (17-21 Oct. 2005). Multi-modal tensor face for simultaneous super-resolution and recognition. *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, **2**, 1683–1690 Vol. 2.

Gordon, G. (1992). Face recognition based on depth and curvature features. In *Computer Vision and Pattern Recognition, 1992. Proceedings CVPR '92., 1992 IEEE Computer Society Conference on*, pages 808–810.

Gross, R., Shi, J., and Cohn, J. (2001). Quo vadis face recognition? In *Third Workshop on Empirical Evaluation Methods in Computer Vision*.

Guo, G., Li, S. Z., and Chan, K. L. (2001). Support vector machines for face recognition. *Image and Vision Computing*, **19**(9-10), 631 – 638.

Gupta, H., Agrawal, A., Pruthi, T., Shekhar, C., and Chellappa, R. (2002). An experimental evaluation of linear and kernel-based methods for face recognition. In *Applications of Computer Vision, 2002. (WACV 2002). Proceedings. Sixth IEEE Workshop on*, pages 13–18.

Hafed, Z. M. and Levine, M. D. (2001). Face recognition using the discrete cosine transform. *Int. J. Comput. Vision*, **43**(3), 167–188.

Haiping Lu; Plataniotis, K.N.; Venetsanopoulos, A. (2006). Multilinear principal component analysis of tensor objects for recognition. *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, **2**, 776–779.

Harshman, R. A. and Lundy, M. E. (1994). Parafac: parallel factor analysis. *Comput. Stat. Data Anal.*, **18**(1), 39–72.

He, X. and Niyogi, P. (2003). Locality preserving projections. In *Advances in Neural Information Processing Systems 16*. MIT Press.

He, X., Yan, S., Hu, Y., and Zhang, H.-J. (2003). Learning a locality preserving subspace for visual recognition. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 385–392 vol.1.

He, X., Yan, S., Hu, Y., and Niyogi, P. (2005a). Face recognition using laplacianfaces. *IEEE Trans. Pattern Anal. Mach. Intell.*, **27**(3), 328–340. Fellow-Zhang, Hong-Jiang.

He, X., Cai, D., and Niyogi, P. (2005b). Tensor subspace analysis. In *In Advances in Neural Information Processing Systems 18 (NIPS*. MIT Press.

Hesher, C., Srivastava, A., and Erlebacher, G. (2003). A novel technique for face recognition using range imaging. In *Signal Processing and Its Applications, 2003. Proceedings. Seventh International Symposium on*, volume 2, pages 201–204 vol.2.

Jiang, D., Hu, Y., Yan, S., Zhang, L., Zhang, H., and Gao, W. (2005). Efficient 3d reconstruction for face recognition. *Pattern Recognition*, **38**(6), 787 – 798. Image Understanding for Photographs.

Jing, X.-Y. and Zhang, D. (2004). A face and palmprint recognition approach based on discriminant dct feature extraction. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, **34**(6), 2405–2415.

Jung, H.-C., Hwang, B.-W., and Lee, S.-W. (2004). Authenticating corrupted face image based on noise model. *Automatic Face and Gesture Recognition, IEEE International Conference on*, **0**, 272.

Kanade, T. (1977). Computer recognition of human faces. *Interdisciplinary Systems Research*, **47**.

Kelly, M. D. (1971). *Visual identification of people by computer*. Ph.D. thesis, Stanford University, Stanford, CA, USA.

Kirby, M. and Sirovich, L. (1990). Application of the karhunen-loeve procedure for the characterization of human faces. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **12**(1), 103–108.

Kolda, T., Bader, B., and Kenny, J. (2005). Higher-order web link analysis using multilinear algebra. In *Data Mining, Fifth IEEE International Conference on*, pages 8 pp.–.

Kong, S. G., Heo, J., Abidi, B. R., Paik, J., and Abidi, M. A. (2005). Recent advances in visual and infrared face recognition–a review. *Computer Vision and Image Understanding*, **97**(1), 103 – 135.

Krger, N., Ptzsch, M., and von der Malsburg, C. (1997). Determination of face position and pose with a learned representation based on labelled graphs. *Image and Vision Computing*, **15**(8), 665 – 673. British Machine Vision Conference.

Kroonenberg, P. and Leeuw, J. D. (1980). Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika*, **45**, 69–97.

Lades, M., Vorbruggen, J. C., Buhmann, J., Lange, J., von der Malsburg, C., Wurtz, R. P., and Konen, W. (1993). Distortion invariant object recognition in the dynamic link architecture. *IEEE Trans. Comput.*, **42**(3), 300–311.

Lao, S., Sumi, Y., Kawade, M., and Tomita, F. (2000). 3d template matching for pose invariant face recognition using 3d facial model built with isoluminance line based stereo vision. In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, volume 2, pages 911–916 vol.2.

Lathauwer, L. D., Moor, B. D., and Vandewalle, J. (2000). On the best rank-1 and rank-(r1,r2,. . .,rn) approximation of higher-order tensors. *SIAM J. Matrix Anal. Appl.*, **21**(4), 1324–1342.

Lee, J., Moghaddam, B., Pfister, H., and Machiraju, R. (17-21 Oct. 2005a). A bilinear illumination model for robust face recognition. *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, **2**, 1177–1184 Vol. 2.

Lee, K.-C., Ho, J., and Kriegman, D. (May 2005b). Acquiring linear subspaces for face recognition under variable lighting. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **27**(5), 684–698.

Li, R.-F., Hao, H.-W., Tu, X.-Y., and Wang, C. (2005a). Face recognition using kfd-isomap. In *Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on*, volume 7, pages 4544–4548 Vol. 7.

Li, S. and Lu, J. (1999). Face recognition using the nearest feature line method. *Neural Networks, IEEE Transactions on*, **10**(2), 439–443.

Li, S. Z., Zhao, C., Ao, M., and Lei, Z. (2005b). Learning to fuse 3d+2d based face recognition at both feature and decision levels.

Li, S. Z., Zhang, L., Liao, S., Zhu, X., Chu, R., Ao, M., and He, R. (2006). A near-infrared image based face recognition system. In *FGR '06: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*, pages 455–460, Washington, DC, USA. IEEE Computer Society.

Liu, C. and Wechsler, H. (1998). Face recognition using evolutionary pursuit. In *ECCV '98: Proceedings of the 5th European Conference on Computer Vision-Volume II*, pages 596–612, London, UK. Springer-Verlag.

Liu, C. and Wechsler, H. (2000). Evolutionary pursuit and its application to face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, **22**(6), 570–582.

Liu, Q., Huang, R., Lu, H., and Ma, S. (2002). Face recognition using kernel-based fisher discriminant analysis. In *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, pages 197–201.

Liu, W., Tang, X., and Liu, J. (2007). Bayesian tensor inference for sketch-based facial photo hallucination. In *Artificial Intelligence, 2007. IJCAI 2007. 20th International Joint Conference on*, page 21412146.

Logothetis, N. K. and Gauthier, I. (1 February 2000). Is face recognition not so unique after all? *Cognitive Neuropsychology*, **17**, 125–142(18).

Lotlikar, R. and Kothari, R. (2000). Fractional-step dimensionality reduction. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **22**(6), 623–627.

Lu, J., Plataniotis, K., and Venetsanopoulos, A. (2002). A kernel machine based approach for multi-view face recognition. In *Image Processing. 2002. Proceedings. 2002 International Conference on*, volume 1, pages I–265–I–268 vol.1.

Lu, J., Plataniotis, K., and Venetsanopoulos, A. (2003a). Face recognition using kernel direct discriminant analysis algorithms. *Neural Networks, IEEE Transactions on*, **14**(1), 117–126.

Lu, J., Plataniotis, K., and Venetsanopoulos, A. (2003b). Face recognition using lda-based algorithms. *Neural Networks, IEEE Transactions on*, **14**(1), 195–200.

Lu, J., Plataniotis, K., Venetsanopoulos, A., and Li, S. (2006). Ensemble-based discriminant learning with boosting for face recognition. *Neural Networks, IEEE Transactions on*, **17**(1), 166–178.

Lu, X. and Jain, A. K. (2005). Integrating range and texture information for 3d face recognition. In *WACV-MOTION '05: Proceedings of the Seventh IEEE Workshops on Application of Computer Vision (WACV/MOTION'05) - Volume 1*, pages 156–163, Washington, DC, USA. IEEE Computer Society.

Majumdar, A. and Ward, R. (2008). Pseudo-fisherface method for single image per person face recognition. In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, pages 989–992.

Mandal, T. and Wu, Q. (2008). Face recognition using curvelet based pca. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4.

Martinez, A. and Kak, A. (2001). Pca versus lda. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **23**(2), 228–233.

Martínez, A. M. (2002). Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. *IEEE Trans. Pattern Anal. Mach. Intell.*, **24**(6), 748–763.

Medioni, G. and Waupotitsch, R. (2003). Face recognition and modeling in 3d. In *AMFG 2003; IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, pages 232–2003.

Member-Kim, T.-K. and Member-Kittler, J. (2005). Locally linear discriminant analysis for multimodally distributed classes for face recognition with a single model image. *IEEE Trans. Pattern Anal. Mach. Intell.*, **27**(3), 318–327.

Mika, S., Ratsch, G., Weston, J., Scholkopf, B., and Mullers, K. (1999a). Fisher discriminant analysis with kernels. In *Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop*, pages 41–48.

Mika, S., Rtsch, G., Weston, J., Schlkopf, B., Smola, A., and Mller, K.-R. (1999b). Invariant feature extraction and classification in kernel spaces.

Min, W., Lu, K., and He, X. (2004). Locality pursuit embedding. *Pattern Recognition*, **37**(4), 781 – 788. Agent Based Computer Vision.

Moghaddam, B. and Pentland, A. (1997). Probabilistic visual learning for object representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **19**(7), 696–710.

Nefian, A. and Hayes, M.H., I. (1998). Hidden markov models for face recognition. In *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, volume 5, pages 2721–2724 vol.5.

Pan, Z., Rust, A., and Bolouri, H. (2000). Image redundancy reduction for neural network classification using discrete cosine transforms. In *Neural Networks, 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on*, volume 3, pages 149–154 vol.3.

Pan, Z., Healey, G., Prascad, M., and Tromberg, B. (2003). Face recognition in hyperspectral images. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 1, pages I–334–I–339 vol.1.

Pang, Y., Yu, N., Li, H., Zhang, R., and Liu, Z. (2005a). Face recognition using neighborhood preserving projections. *Lecture Notes in Computer Science*, **3768/2005**, 854–864.

Pang, Y., Zhang, L., and Liu, Z. (2005b). Neighborhood preserving projections (npp): A novel linear dimension reduction method. *Lecture Notes in Computer Science*, **3644/2005**, 117–125.

Penev, P. and Atick, J. (August 1996). Local feature analysis: a general statistical theory for object representation. *Network: Computation in Neural Systems*, **7**, 477–500(24).

Pentland, A., Moghaddam, B., and Starner, T. (1994). View-based and modular eigenspaces for face recognition. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on*, pages 84–91.

Perlibakas, V. (2004). Face recognition using principal component analysis and wavelet packet decomposition. *Informatica*, **15**(2), 243–250.

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

Phillips, P. J. (1999). Support vector machines applied to face recognition. In *Proceedings of the 1998 conference on Advances in neural information processing systems II*, pages 803–809, Cambridge, MA, USA. MIT Press.

Rao, K. R. and Yip, P. (1990). *Discrete cosine transform: algorithms, advantages, applications*. Academic Press Professional, Inc., San Diego, CA, USA.

Romdhani, S. and Vetter, T. (2003). Efficient, robust and accurate fitting of a 3d morphable model. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 59–66 vol.1.

Roweis, S., Saul, L. K., and Hinton, G. E. (2002). Global coordination of local linear models. In *Advances in Neural Information Processing Systems 14*, pages 889–896. MIT Press.

Roweis, S. T. and Saul, L. K. (2000). Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science*, **290**(5500), 2323–2326.

Samaria, F. and Young, S. (1994). Hmm-based architecture for face identification. *Image and Vision Computing*, **12**(8), 537 – 543.

Saul, L. K. and Roweis, S. T. (2003). Think globally, fit locally: unsupervised learning of low dimensional manifolds. *J. Mach. Learn. Res.*, **4**, 119–155.

Scheenstra, A., Ruifrok, A., and Veltkamp, R. C. (2005). A survey of 3d face recognition methods. In *Lecture Notes in Computer Science*, pages 891–899. SpringerVerlag.

Schmidhuber, J. (1992). Learning factorial codes by predictability minimization. *Neural Computation*, **4**(6), 863–879.

Schölkopf, B., Smola, A., and Müller, K.-R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput.*, **10**(5), 1299–1319.

Seung, H. S. and Lee, D. D. (2000). COGNITION: The Manifold Ways of Perception. *Science*, **290**(5500), 2268–2269.

Sinha, P., Balas, B., Ostrovsky, Y., and Russell, R. (2006). Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE*, **94**(11), 1948–1962.

Socolinsky, D. and Selinger, A. (2002). A comparative analysis of face recognition performance with visible and thermal infrared imagery. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, volume 4, pages 217–222 vol.4.

Socolinsky, D. A., Selinger, A., and Neuheisel, J. D. (2003). Face recognition with visible and thermal infrared imagery. *Computer Vision and Image Understanding*, **91**(1-2), 72 – 114. Special Issue on Face Recognition.

Spies, H. and Ricketts, I. (2000). Face recognition in fourier space. In *Proc. Vision Interface Conf.*, Montreal, Canada.

Stewart Bartlett, M., Lades, H. M., and Sejnowski, T. J. (1998). Independent component representations for face recognition. In B. E. Rogowitz and T. N. Pappas, editors, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 3299 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pages 528–539.

Swets, D. and Weng, J. (1996). Using discriminant eigenfeatures for image retrieval. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **18**(8), 831–836.

Takcs, B. (1998). Comparing face images using the modified hausdorff distance. *Pattern Recognition*, **31**(12), 1873 – 1881.

Tenenbaum, J. B., Silva, V. d., and Langford, J. C. (2000). A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science*, **290**(5500), 2319–2323.

Tsalakanidou, F., Malassiotis, S., and Strintzis, M. (2004). Integration of 2d and 3d images for enhanced face authentication. In *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, pages 266–271.

Tucker, L. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, **31**, 279–311.

Turk, M. and Pentland, A. (1991a). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, **3**(1), 71–86.

Turk, M. and Pentland, A. (3-6 Jun 1991b). Face recognition using eigenfaces. *Computer Vision and Pattern Recognition, CVPR 1991, IEEE Computer Society Conference on*, pages 586–591.

Vapnik, V. N. (1995). *The nature of statistical learning theory*. Springer-Verlag New York, Inc., New York, NY, USA.

Vasilescu, M. Alex O.; Terzopoulos, D. (21-23 Aug. 2007). A tensor algebraic approach to image synthesis, analysis and recognition. *3-D Digital Imaging and Modeling, 2007. 3DIM '07. Sixth International Conference on*, pages 3–12.

Vasilescu, M.A.O.; Terzopoulos, D. (20-25 June 2005). Multilinear independent components analysis. *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, **1**, 547–553 vol. 1.

Vasilescu, M. and Terzopoulos, D. (2002a). Multilinear image analysis for facial recognition. *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, **2**, 511–514 vol.2.

Vasilescu, M. A. O. and Terzopoulos, D. (2002b). Multilinear analysis of image ensembles: Tensorfaces. In *ECCV '02: Proceedings of the 7th European Conference on Computer Vision-Part I*, pages 447–460, London, UK. Springer-Verlag.

Vasilescu, M. A. O. and Terzopoulos, D. (2004). Tensortextures: multilinear image-based rendering. In *SIGGRAPH '04: ACM SIGGRAPH 2004 Papers*, pages 336–342, New York, NY, USA. ACM.

Vlasic, D., Brand, M., Pfister, H., and Popović, J. (2005). Face transfer with multilinear models. *ACM Trans. Graph.*, **24**(3), 426–433.

Wang, H. and Ahuja, N. (16-16 Oct. 2003). Facial expression decomposition. *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 958–965 vol.2.

Wang, H. and Ahuja, N. (2005). Rank-r approximation of tensors using image-as-matrix representation. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 346–353 vol. 2.

Wang, J., Plataniotis, K., and Venetsanopoulos, A. (2005). Selecting discriminant eigenfaces for face recognition. *Pattern Recognition Letters*, **26**(10), 1470 – 1482.

Wang, J., Plataniotis, K. N., Lu, J., and Venetsanopoulos, A. N. (2006). On solving the face recognition problem with one training sample per subject. *Pattern Recogn.*, **39**(9), 1746–1762.

Weyrauch, B., Heisele, B., Huang, J., and Blanz, V. (2004). Component-based face recognition with 3d morphable models. In *Computer Vision and Pattern Recognition Workshop, 2004. CVPRW '04. Conference on*, pages 85–85.

Wiskott, L., Fellous, J.-M., Kruger, N., and von der Malsburg, C. (1997). Face recognition by elastic bunch graph matching. In *Image Processing, 1997. Proceedings., International Conference on*, volume 1, pages 129–132 vol.1.

Wu, J. and Zhou, Z.-H. (2002). Face recognition with one training image per person. *Pattern Recogn. Lett.*, **23**(14), 1711–1719.

Yan, S., Xu, D., Lin, S., Huang, T., and Chang, S.-F. (2007a). Element rearrangement for tensor-based subspace learning. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1–8.

Yan, S., Xu, D., Yang, Q., Zhang, L., Tang, X., and Zhang, H.-J. (2007b). Multilinear discriminant analysis for face recognition. *Image Processing, IEEE Transactions on*, **16**(1), 212–220.

Yang, J., Frangi, A. F., Yang, J.-y., and Jin, Z. (2005). Kpca plus lda: A complete kernel fisher discriminant framework for feature extraction and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, **27**(2), 230–244. Senior Member-Zhang, David.

Yang, M. H. (2001). Face recognition using kernel methods. In *Advances in Neural Information Processing System*. MIT Press.

Yang, M.-H. (2002a). Face recognition using extended isomap. In *Image Processing. 2002. Proceedings. 2002 International Conference on*, volume 2, pages II–117–II–120 vol.2.

Yang, M. H. (2002b). Kernel eigenfaces vs. kernel fisherfaces: Face recognition using kernel methods. In *FGR '02: Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, page 215, Washington, DC, USA. IEEE Computer Society.

Yang, M.-H., Ahuja, N., and Kriegman, D. (2000). Face recognition using kernel eigenfaces. In *Image Processing, 2000. Proceedings. 2000 International Conference on*, volume 1, pages 37–40 vol.1.

Yu, H. and Yang, J. (2001). A direct lda algorithm for high-dimensional data – with application to face recognition. *Pattern Recognition*, **34**(10), 2067 – 2070.

Zhao, W., Chellappa, R., and Krishnaswamy, A. (1998). Discriminant analysis of principal components for face recognition. In *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, pages 336–341.

Zhao, W., Chellappa, R., and Phillips, P. (1999). Subspace linear discriminant analysis for face recognition. Technical report, Center for Automation Research, University of Maryland, College Park, MD.

Zhao, W., Chellappa, R., Phillips, P. J., and Rosenfeld, A. (2003). Face recognition: A literature survey. *ACM Comput. Surv.*, **35**(4), 399–458.

Zhao, W. Y. and Chellappa, R. (2000). Sfs based view synthesis for robust face recognition. In *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, pages 285–292.

Zheng, Z. and Yang, J. (2006). Supervised locality pursuit embedding for pattern classification. *Image and Vision Computing*, **24**(8), 819 – 826.

Zhou, Z., Li, S., and Chan, K. L. (2000). A theoretical justification of nearest feature line method. In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, volume 2, pages 759–762 vol.2.

Zhu, J., Vai, M., and Mak, P. (2003). Face recognition using 2d dct with pca. In *Biometric Recognition (Sinobiometrics03), The 4th Chinese Conference on*, Beijing, PR China.

Zou, X., Kittler, J., and Messer, K. (2005). Face recognition using active near-ir illumination. In *BMVC '05: Proceedings of the British Machine Vision Conference*.