Department of Computing

Robust Face Recognition via Multi-channel Models and Feature Selection

Mustafa M M Alrjebi

This thesis is presented for the Degree of

Doctor of Philosophy
at
Curtin University

October 2017

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.

Mustafa Alrjebi

Date

04/10/2017

Abstract

Automatically identifying people is essential in daily life such as finding missing children, identifying criminals and searching for people on internet, and so face recognition is a popular research topic in computer vision community. One obvious advantage of face recognition is without user cooperation. As face images are usually captured in uncontrolled environments, therefore, some conditions such as facial expressions, pose variations, lighting variations and occlusions can negatively affect the accuracy of face recognition. Current face recognition approaches against occlusions and poses variations are mainly still based on grey images or three color channels ignoring the role of full-color information from multi color channels. With regard to face recognition against illumination variations, research in this area is mainly based on normalising grey images utilising a single channel. Also, some important factors in face recognition are still not investigated properly. For example, some extreme facial expressions can lead to significant misalignment of face shapes, which can't be solved by the existing alignment methods, and result in very low recognition accuracy.

Regarding color face recognition, the use of color information is based essentially on three color channels with attempts to gain valuable discriminant information using different transformation methods from RGB images to other robust color spaces and models such as Normalised RGB and Discriminant Color Space (DCS). On the other hand, the Multiple Color Fusion (MCF) is the only attempt of effective use of multiple color channels resulting in a better performance against all other face image representation methods. Even though the MCF model shows high performance in comparison with other three channels based color spaces and models, it is only used in a holistic face manner ignoring the valuable local variations in different face areas. Moreover, the MCF and all other color spaces and models utilise different color channels with an assumption that all color channels have a similar/equal contribution to face recognition performance and in fact, this is not true. In this thesis, these problems will be systematically investigated and the main contributions are stated as follows in each chapter summary.

In Chapter 3, the extreme facial expressions problem will be addressed. This is motivated by the fact that mouth shape changes in extreme expressions will affect the face recognition performance significantly. One face recognition approach against mouth shape variations is proposed, in which a mouth shape correction algorithm is proposed. Technically, the mouth shape change is determined by detecting the lips and teeth connectivity, and then an algorithm is proposed to correct an open mouth to

a closed-mouth case morphologically. The effectiveness of the proposed method is evaluated via experiments showing high improvement in recognition accuracy.

In Chapter 4, the effective use of color information in different face areas is investigated. This thesis introduces a new multichannel color fusion in block wise named as the Two Directional Multiple Color Fusion (2D-MCF). In this approach, one represents a face image using multiple colors locally in block wise rather than in holistic face image. Each block in this model can have multiple color channels across different color spaces. In comparison with the existing representation methods, this method can achieve very high recognition rate for frontal face images. Moreover, this new model is integrated with the Partitioned Sparse Representation based Classifier (P-SRC) to tackle the problem of face recognition against occlusions in Chapter 5. The 2D-MCF representation is used for image representation, and the P-SRC with majority voting is used as a classifier. The proposed model has been compared with some of the state of the art methods based on RGB color space and greyscale images, and it shows very high improvement in recognition accuracy.

In chapter 6, we investigate the effective use of multiple color information in holistic face images with different resolutions. To the best of our knowledge, this is the first time to use multi-resolution color information in face recognition. This thesis proposes a weighted multiple color channel model based on holistic face image named as the Multi-resolution Multiple Color Fusion (MMCF). The idea behind this model is to assign a specific value to each color channel according to its impact on face recognition. In this method, the face image is represented by multiple color channels where different color channels can have different sizes (resolutions) based on the role of each channel in face recognition. As expected, the proposed method outperforms the existing methods including the MCF and 2D-MCF models. Moreover, the proposed method is integrated with one of state of the art deep learning network for pose correction, and such integration can achieve very high performance.

Finally, a new multi-channel representation is proposed to tackle the problem of illumination variations named as the Multilevel Threshold Local Binary Pattern Fusion MTLBP-F. This method utilises the grey image representation. After enhancing the image quality using the Discrete Cosine Transform, multi-channels are produced using multi level of Threshold Local Binary Pattern, and then the image representation is optimised for face recognition using the 2D-MCF criteria. In comparison with one of the best algorithms, the proposed method can achieve much better recognition accuracy.

In summary, this thesis proposes some novel methods based on image representation in 2D color and grey images to tackle the most challenging problems in face recognition. The mouth shape correction is a straightforward algorithm which can recover the face image from misalignment situation resulting in a high improvement in recognition accuracy. The main focus of this thesis is the effective use of multiple channels to represent face image either in color images or grey images including local representation in the 2D-MCF, weighted multiple channels in the MMCF model and multiple channels from grey images in the MTLBP-F. Our experiments show that these multi-channel models can compensate the insufficiency of discriminant information caused by occlusions, pose variations and illumination variations.

Acknowledgments

All praises are due to Almighty God “Allah”, Who provided me with the strength and willingness to undertake this work and the opportunity to contribute a drop in the sea of knowledge. I am most grateful to my supervisors, Wanquan Liu and Ling Li. Their open doors and persistent encouragement was invaluable for the completion of this research. Finally, these acknowledgements would not complete without appreciating the unwavering support of my family including my mother, my wife and children, and the memory of my father.

Published Work

This thesis includes the following works that have been published over the course of my PhD study, and they are listed in chapter order:

- Alrjebi, M. M., W. Liu and L. Li (2015). Face Recognition against Mouth Shape Variations. 2015 International Conference on Digital Image Computing: Techniques and Applications. (Chapter 3).
- Alrjebi M. M., W. L., Ling Li (2015). Two Directional Multiple Color Fusion for Face Recognition. International Conference on Computers, Communications, and Systems (ICCCS). Kanyakumari, India, IEEE**:** pp. 171-177. (Chapter 4).
- Alrjebi, M. M., N. Pathirage, W. Liu and L. Li (2017). "Face recognition against occlusions via color fusion using 2D-MCF model and SRC." Pattern Recognition Letters 95: 14-21. (Chapter 5).
- Mustafa M. Alrjebi, W. L., Ling Li (2017).Face Recognition against Illuminations using Two Directional Multi-Level Threshold-LBP and DCT. Accepted by Multimedia Tools and Applications. (Chapter 6).
- Mustafa M. Alrjebi, W. L., Ling Li (2017). Face Recognition against Pose Variations Using Multi-resolution Multiple Color Fusion. International Journal of Machine Intelligence and Sensory Signal Processing 1.4 (2016): 304-320. (Chapter 7).

Contents

List of Figures

List of tables

Chapter One

Introduction

Face Recognition is an attractive way to identify people automatically via machine learning. Unlike other biometrics techniques, face recognition is a natural way to recognise people that can be implemented automatically without person cooperation. Even though other biometrics such as fingerprint and iris scan are more accurate than face recognition, they need person cooperation with specified devices. Such cooperation is unmanageable in many situations such as identifying people from the internet, searching for missing people, surveillance systems and all other automatic human recognition systems. Any face recognition system contains three main components, which are input data capture (captured face images), feature extraction, and classifier design. The quality of each of these components has a direct impact on the recognition accuracy.

As it is well-known, all human faces are similar; they all contain the same organs that are distributed in different facial areas in the same manner which makes facial recognition very challenge. Moreover, some other factors can increase the complexity of this problem such as facial expressions, occlusions, pose variations and lighting conditions. These factors can deteriorate the required discriminant information and can be considered as noise. Consequently, we can divide the information contained in a facial image into three categories: the similarity information, discriminant Information, and noise. From these three categories, the important information we need to identify a face is the discriminant information, while similarity information and noise have a negative effect on face recognition performance.

All the above factors have the effect of reducing the discriminative information, which can negatively affect face recognition performance. For example, extreme facial expressions can deform face geometry leading to a mismatch between the query image and gallery images. Most importantly, the occlusions, pose variations, and lighting variations will cause a severe damage of discriminant information.

Many researches in face recognition have been conducted to deal with different facial expressions, but they all ignore the case of extreme expressions that can lead to deformation of face geometry. The mouth shape variation can be considered as a special case of extreme facial

expressions because it leads the face image to a misalignment situation. This problem cannot be solved by existing alignment methods because it deforms the face geometry seriously due to changing the ratio between the height and width of a face as well we changing the locations and shapes of some facial organs such as nose, mouth, cheeks and chin. We will address this face deformation problem in this thesis.

With regard to color face recognition, research in face recognition has been redirected to color images as the rapid growth of hardware technology can provide a high processing speed and large memory space. The use of color information shows a significant improvement in face recognition accuracy (Torres, Reutter et al. 1999). The RGB images were the first color images used for face recognition. It has been shown that the RGB channels can provide invaluable discriminant information which achieves much better performance than previous greyscale images. In order to enhance the efficiency of color information for color face recognition, various algorithms have been proposed to obtain color spaces specifically for face recognition task (Jian and Chengjun 2008, Yang and Liu 2008, Chengjun and Jian 2009, Jing, Liu et al. 2010, Yang, Liu et al. 2010). The aim is to create a new color space with invaluable discriminant information. In other words, these algorithms attempt to enrich the discriminant information, to minimise the similarity information, or to do both tasks simultaneously. Researchers attempt to find the best representation for face recognition by transforming the RGB representation to other color spaces which are expected to have better discriminant information. The common characteristic between these color spaces is that they all contain exactly three color channels. However, there is only one model known as the Multiple Color Fusion MCF (Li, An et al. 2011) which uses multiple color channels and shows much better recognition accuracy in comparison with other three channels based color spaces. The MCF model holistically utilises multiple color channels while ignoring the local variances in different facial areas. Moreover, all current color spaces and models assume that different color channels have the same contribution in face recognition which is not true in practice. We will address these issues and develop new approaches to improve the performance in this thesis.

With regard to the problem of face recognition against occlusions, research on this issue has been focused mainly on detecting and removing the occlusion and then performing face recognition. Some other methods like the Partitioned SRC（P-SRC） (Wright, Yang et al. 2009) are developed to deal with the problem without detecting the occlusion area. The input face data used in P-SRC is the greyscale image. As it is known that the greyscale images cannot provide sufficient discriminant information for face recognition, we will address the occlusion issue using color information in Chapter five.

With regard to face recognition against pose variations, one main idea of research in this problem is the pose correction, which implies that correcting the face image with pose angle to a frontal pose. This correction can significantly improve the recognition accuracy because it matches the input face image with the frontal gallery images. However, such correction can only solve the problem of the mismatch between the query image and the gallery images. One negative impact of pose variations is that the face area is only partially presented in the face images with pose angles, for example, a face image with pose of 90^0 shows only 50% of the face area which indicates that 50% of face information is missed. This missed information can significantly affect the recognition accuracy. Inappropriately, current research only pays attention to correcting pose angles to a frontal pose while ignoring the deficiency of necessary discriminant information. In this thesis, color information will be used effectively to compensate the shortage of discriminant information caused by pose variations.

Regarding face recognition against lighting variations, research in this area attempts to enhance image quality in different ways. The illumination adds unnecessary information to face image, and it deforms some local variances in the face image. Consequently, some methods attempt to enhance the image quality by normalising the image and removing/reducing the illumination information or by improving the local variances. All existing methods utilise the greyscale images with a single channel to represent the face image. This use of single channel seems unsatisfactory to provide the discriminant information required for face recognition. Color information will be used to address this issue in this thesis.

The aim of this thesis is to overcome the shortcomings mentioned above as illustrated in the next subsections

1.1. Aims and Approaches

As many face recognition systems still use 2D representation either for color or grey images, the aim of this thesis is to achieve robust face recognition by using 2-D information focusing on the image representation. Technically, we aim to demonstrate how image representation can enhance recognition accuracy in different conditions including the cases of extreme facial expressions, uncontrolled environments, presence of occlusions, lighting variations and pose variations. For the extreme expressions, our aim is to reconstruct an aligned face image from the deformation caused by the opened-mouth case. This study is based on landmark detections and some metrics on normal face shapes. As to occlusions, pose variations and lighting variations, these factors will lead to loss of important discriminant information, we aim to

compensate the discriminate information by enriching the features from multiple color channels. The objectives of this research can be summarised as follows:

1. Robust face recognition against mouth shape variations.
2. Robust frontal face recognition via multi-color information.
3. Robust face recognition against occlusions by using multi-color representation.
4. Robust face recognition against small pose variations using multi-color representation.
5. Robust face recognition against lighting variations using multi-channel filter representation based on greyscale images.

The first objective is to tackle the problem of mouth shape variations in face recognition. The recovery from an open mouth to closed mouth is our essential idea for this problem. Firstly, the proposed method can recognise the mouth shape and classify it into one of four categories which *are Open-mouth, Closed mouth, Open mouth with teeth and Smiling*. Secondly, it corrects these different shapes to the closed-mouth shape. For the second objective, we proposed a new hybrid color model named as the Two Directional Multiple Color Fusion (2D-MCF) based on more than three color channels in local patches. The aim is to achieve the best color representation locally in a block-wise way rather than in a holistic manner for a face image. For the third objective, we combine the 2D-MCF representation in conjunction with the P-SRC classifier (Wright, Yang et al. 2009) that widely used for face recognition against occlusions. For the fourth objective, we proposed to use the color information by revising an existing pose correction method based on deep learning technique. As the 2D-MCF is designed for frontal face recognition, a new *multicolor multi-resolution model* is proposed the first time in this thesis. This new model utilises multiple color channels with different resolutions to represent the face image holistically. Finally, for the fifth objective, we explore the multiple filter channels based on the Threshold LBP (Meng, Gao et al. 2010) to enrich a face image with invaluable discriminant information. The fusion decision is taken in a block-wise manner similar to the criteria used in 2D-MCF.

1.2. Contributions of this Thesis

All objectives mentioned in the previous subsection have been achieved in this thesis. The detailed contributions of this thesis can be interpreted as follows:

1.2.1. Innovative Techniques

All methods proposed in this thesis are innovative. They provide a significant improvement in recognition accuracy over other existing state of the art methods. Furthermore, we propose different new methods for different problems, including the mouth shape variations, occlusions, pose variations and lighting variations.

For mouth shape variations, a mouth shape correction method is proposed utilising the mouth landmarks based on the AR detection model (Liang, Liu et al. 2014). First, we detect the lips landmarks and teeth, and then the mouth shape is classified to one of four categories which are, *opened mouth, closed mouth, opened mouth with teeth and smiling*. Finally, the opened mouth cases are corrected to closed mouth using a proposed straightforward algorithm. Such correction can recover the face geometry to a normal form and also remove the unwanted mouth cavity area. This algorithm can be considered as a pre-processing step in any face recognition system, which can be performed to avoid misalignment caused by the Opened-mouth.

The second novel method is named as the Two Directional Multiple Color Fusion 2D-MCF. This method utilises multiple color channels locally in a block-wise manner to achieve the best color representation of each block in a face image via machine learning technique. This method takes the differences between various facial areas into account by considering shape and color information based on the assumption that different facial areas should have different color representations. This model is trained by using the MCF criteria locally, and then a holistic face image is used for face recognition after dividing into different small blocks.

The third method is the integration of the 2D-MCF model with the P-SRC approach (Wright, Yang et al. 2009). The 2D-MCF can enhance the image representation while the P-SRC (Wright, Yang et al. 2009) can cope with occlusions well. Their combination can improve recognition accuracy significantly in dealing with occlusions. In addition to that, we also create two different face datasets from the AR database and the FRGC database by adding occlusions to face images. These new datasets include a high percentage of occlusions, occlusions in an uncontrolled environment as well as different shapes/colors of mask occlusions.

The fourth method is to propose a new model employing multiple color channels with multi-resolutions. This new model, named as the Multi-resolution Multiple Color Fusion (MMCF), utilises various color channels for holistic face images. In this model, different color channels can have different resolutions. The essential idea is to adopt different resolutions to different

color channels according to the contribution of each channel in face recognition. The fusion decision is made in a holistic manner rather than a block-wise way, making this model very robust to small pose variations. For face recognition against large pose variations, this new model is integrated with the Stacked Progressive Auto Encoder (SPAE) deep network (Kan, Shan et al. 2014). The SPAE is used to correct different poses of different color channels of the MMCF model instead of correcting only greyscale images. The SPAE can match the query image with pose angle to the frontal gallery images, and the MMCF model compensates the deficiency of necessary discriminant information caused by pose variations. Such integration leads to a high performance of face recognition.

Finally, the fifth method named as the Two Directional Multilevel Threshold-LBP Fusion (2D-MTLBP-F) is proposed to tackle the problem of lighting variations. This method uses multiple filter channels from Threshold Local Binary Pattern TLBP based on greyscale images. The fusion decision is made based on the 2D-MCF criteria. The key idea of this approach is to compensate the insufficiency of discriminant information caused by illumination variations by adding more filter channels.

1.2.2. Theoretical Contribution

Theoretical ideas are important contributions because they motivate and assist research in the future. This thesis provides the following theoretical insights:
For mouth shape variations, we analyse the misalignment caused by opened-mouth shape and its negative impact on recognition accuracy. We show that the recovery can recover the normal face geometry leading to the recognition accuracy improvement significantly.

Secondly, for the 2D-MCF representation, we show the importance of local optimal representation. As different facial areas have different shapes and colors, the best representation of each facial area may vary from one local facial area to another. Consequently, we propose the 2D-MCF model which assigns a specific multiple channel representation to each block in a face image according to its performance in face recognition. Also, we demonstrate the significance of using this model to enhance the patch based face recognition against occlusions. Theoretically, a global learning after obtaining each optimal block representation is also significantly important as demonstrated in the second step of 2D-MCF.

Fourthly, the importance of using the weighted multiple channels is analysed in the MMCF model, where different weights are assigned to different channels in the proposed model. Also,

the shortcomings of using the deep learning technique based on greyscale images are analysed, and then we demonstrate the significance of using the MMCF model instead of greyscale images with the Stacked Progressive Auto Encoder SPAE (Kan, Shan et al. 2014) for pose correction and face recognition. I believe this is the first time to use multi-resolutions multi-colours facial images for face recognition.

Finally, I propose to use multiple filter channels to improve face recognition performance against illuminations based on greyscale images. We analyse the different information given by different TLBP levels and how they can provide valuable information for face recognition task. Also, we demonstrate the efficiency of using the 2D-MCF criteria via optimising a large number of TLBP filter channels for face recognition.

1.2.3. Experimental Setup

In this thesis, we have conducted intensive experimental demonstrations for the effectiveness of our different proposed methods. We have used different face databases that commonly used in existing literature including facial datasets in an uncontrolled environment, facial expression variations, occlusions, pose variations and lighting variations. We also create some special datasets for occlusions. Furthermore, some new types of experiments have been proposed, such as the cross database testing using deep learning technique and a high percentage of occlusions.

1.3. Thesis Structure

The rest of this thesis is organised as follows. In chapter 2, some related knowledge is presented. We review some of the most popular face recognition methods like Principle Component Analysis PCA (Turk and Pentland 1991), Linear Discriminant Analysis LDA (Etemad and Chellappa 1997) and SRC (Wright, Yang et al. 2009). Furthermore, we present the most related work to this thesis. Regarding grey images, the Local Binary Pattern LBP (Yang and Chen 2013) is presented. For color images, different three channel based color spaces are reviewed. Also, the multichannel color space named as the Multiple Color Fusion (MCF) (Li, An et al. 2011) is analysed. Finally, the deep learning networks are briefly illustrated.

In chapter 3, face recognition against mouth shape variations is presented. First, an analysis of the problem is highlighted, and then its solution is proposed. A recovery from the opened mouth is proposed by detecting and utilising mouth landmarks. Finally, the proposed method is evaluated experimentally with a specified dataset.

In chapter 4, a new multi-channel color model named as the Two Directional Multiple Color Fusion (2D-MCF) is developed. This method extends the original MCF model locally in a block-wise manner rather than on a holistic face image. The idea and proposed algorithm are justified. Finally, this new method is evaluated experimentally while outperforming different existing color spaces and models.

In chapter 5, a new method based on the 2D-MCF model and the P-SRC (Wright, Yang et al. 2009) is proposed. This new method is designed to tackle the problem of face recognition against occlusions. As occlusions appear locally in some face areas, the local optimal face representation based on 2D-MCF is used to solve this problem. The proposed method shows very high accuracy even with high occlusion percentages.

In chapter 6, a new Multi-Channel Multi-resolution representation is proposed. Each image can be represented with multi channels with different resolutions. The idea behind using different resolutions is that different color channels may have different importance for face recognition. The proposed method can achieve very high accuracy in comparison to other methods including the 2D-MCF. Furthermore, the proposed method is integrated with deep learning technique for pose corrections, and it can achieve very high accuracy even in extreme cases like large pose variations, in single gallery image and cross database testing.

In chapter 7, a new multi-channel filer representation named as the two Directional Multi-level Threshold Local Binary Pattern Fusion (2D-MTLBP-F) is proposed for face recognition against lighting variations. The proposed representation is obtained for grey-scale images. Firstly, Discrete Cosine Transform (DCT) (Rao 1990) is used to enhance image quality, and then the enhanced image is transformed into multi levels of LBP with different thresholds. The idea of 2D-MTLBP-F is used for the training to obtain the best representation. The proposed representation is evaluated with different datasets showing high accuracy compared with the existing state of the art methods.

Finally, the thesis is concluded in chapter 8 from the viewpoint of image representation and its role in face recognition. Also, some future research directions are suggested based on this work.

Chapter Two

Background

In this chapter, we present some background research on face recognition methods based on 2D images. Firstly, some face recognition techniques are presented in some details including Principle Component Analysis (PCA) (Turk and Pentland 1991), Linear Discriminant Analysis (LDA) (Etemad and Chellappa 1997), Sparse Representation Classifier (SRC) (Wright, Yang et al. 2009) and Deep learning technique (Deng 2014). Secondly, background knowledge on face image representation in 2D is presented in some details including grey images and some different color image representations. With regard to grey images, the Local Binary Pattern (LBP) (Ali, Hussain et al. 2012) is presented in some detail as it is one of the powerful methods used in face recognition. With regard to color image representation, we present some three channels based color spaces, for example but not limited to RGB, XYZ, YIQ, and HSV (Gonzalez, Woods et al. 2009). Thirdly, we will present some three channels based color models (based on training samples) such as the Discriminant Color Space (DCS) (Jian and Chengjun 2008) and the Holistic Orthogonal Analysis (HOA) (Jing, Liu et al. 2010). Further, we will clearly distinguish between color spaces and color models. Finally, the Multiple Color Fusion (MCF) model (Li, An et al. 2011) is presented in some details in this chapter.

2.1. Face Recognition based on Subspace Approaches

The direct use of face image matching for face recognition is inappropriate and can result in very low accuracy. Usually, the size of face image is large, and its pixel values contain irrelevant information. To avoid this problem, many dimensionality reduction methods have been proposed to attain small sized relevant information. Explicit examples of these methods are Principle Component analysis (PCA) (Turk and Pentland 1991), Linear Discriminant Analysis (LDA) (Etemad and Chellappa 1997), and Independent Component Analysis (ICA) (Chengjun and Jian 2009). The discriminant features obtained by any of these dimensionality reduction methods can be utilised for face recognition via using one of matching methods such

as the Nearest Neighbour NN (Cormen and ebrary 2009). In this section, we will describe the PCA, LDA as they are the most common techniques used for face recognition.

2.1.1. Principle Component Analysis (PCA)

As the discriminant information of face image is very sparse, direct use of pixel value matching for face recognition is not effective due to its high dimensions. A reduced subspace is required to improve the recognition accuracy. Principle Component Analysis proposed in (Turk and Pentland 1991) is widely used in many face recognition systems mainly for dimensionality reduction. This method can reduce the dimensionality of a given input image, and seek for the efficient directions to represent the original data. Consequently, it can reduce the computation time and improve the recognition accuracy. This method can be detailed as follows:

Given a training face dataset containing M images: m_1 , $m_2, \ldots\ldots, m_M$, where $m_i \epsilon \mathbb{R}^{N^2}$. The average face image can be computed using these training face images as follows:

$$\bar{m} = \frac{1}{m} \sum_{i=1}^{M} m_i \tag{2.1}$$

where m_i is N^2 vector representing the i^{th} image in the dataset, and M is the number of images in the data-set.

1. Let $A = [(\bar{m}_1), (\bar{m}_2), \ldots\ldots, (\bar{m}_M)]$ which is $N^2 \times M$ containing all images in the training data-set after subtracting the average image $\overline{m}$ from each image, i.e., $\bar{m}_i = m_i - \bar{m}$ with $i = 1, 2, 3, \ldots\ldots, M$.
2. The covariance matrix C which is $N^2 \times N^2$ is obtained as follows:

$$C = A.A^T \tag{2.2}$$

3. A matrix $L = A^T.A$ which is M by M is used instead of the matrix C. This can reduce the computation because $M \ll N^2$.
4. The eigenvalues λ and their eigenvectors $U \epsilon \mathbb{R}^{N^2 \times M}$ of L are obtained, and then the feature space is defined as $\Omega = [\Omega_1, \Omega_2, \ldots\ldots, \Omega_M] = U^T.A$.

5. To reduce the dimensionality, only $\grave{M}$ eigenvectors associated with the *largest ei*genvalues are used, where $\grave{M} \leq M$. This number of eigenvalues $\grave{M}$ is determined heuristically, and the dimension of U reduced to $N^2 \times \grave{M}$.
6. To recognise given face image X, firstly, its PCA feature vector Ω_X is obtained with $\grave{M}$ feature elements as follows: $\Omega_X = U^T.X$. To recognise the test image X, the obtained feature vector Ω_X is matched to the feature vectors of the training images using the Nearest Neighbour measure.

As it is seen from above, PCA doesn't use class labels and this implies that the discriminant information is not effectively utilized. This is unwise from a classification point of view and can result in poor discrimination for different classes. Linear Discriminant Analysis LDA solves this problem by using the class labels in its definition. In the next subsection, LDA is presented in details.

2.1.2. Linear Discriminant Analysis (LDA)

This section presents Linear Discriminant Analysis LDA for face recognition. LDA proposed in (Etemad and Chellappa 1997) can provide powerful discriminative information for face recognition. It requires images labels in the training. The idea is to maximise inter-class scatter and minimise intra-class scatter at same time. The steps of extracting discriminant features for face recognition can be outlined as follows.

1. The within class scatter matrix S_W and the between class scatter matrix S_B are obtained as follows.

$$S_W = \sum_{i=1}^{C} \sum_{m_{ij} \epsilon P_i} (m_{ij} - \bar{m}_i)(m_{ij} - \bar{m}_i)^T \tag{2.3}$$

$$S_B = \sum_{i=1}^{C} (\bar{m}_i - \overline{m})(\bar{m}_i - \overline{m})^T \tag{2.4}$$

   where P_i is the i^{th} class in the training dataset, C is the number of classes, m_{ij} is the j^{th} image in the i^{th} class, $\bar{m}_i$ is the mean of the i^{th} class, and $\overline{m}$ is the grand mean of all images in the dataset.
2. LDA aims to find a projection matrix Φ that maximises the following ratio

$$\frac{\mathrm{tr}(\Phi^T S_B \Phi)}{tr(\Phi^T S_W \Phi)} \tag{2.5}$$

The solution can be obtained by solving the following eigenvalue problem:

$$S_B \Phi = \lambda S_W \Phi \tag{2.6}$$

To reduce the dimensionality, only eigenvectors associated with the largest eigenvalues are used. Consequently, the dimension of the matrix Φ is reduced to $N^2 \times \grave{M}$, where $\grave{M}$ is the number of largest selected eigenvalues.

3. The dimensionality of the matrix A can be reduced by projecting it to the matrix Φ as: $A_{lda} = \Phi^T . A$.
4. To recognize face image X, its LDA features vector is obtained as follows: X_{lda}=$\Phi^T . X$

As in PCA, A_{lda} features can be used with a classifier such as NN to match any input face image to gallery images. However, if S_W is singular, its inverse does not exist. This problem can be solved by using the PCA subspace instead of a training image matrix including all images). Unfortunately, the use of PCA subspace may lead to loss of relevant discriminant information which can negatively affect the efficacy of LDA in face recognition. The Regularized LDA R-LDA (Lu et al., 2005) solves this problem without need to use the PCA subspace. The original objective function for R-LDA is to maximize the following ratio:

$$\frac{tr(\Phi^T S_b \Phi)}{tr(\Phi^T (S_W + \lambda \mathrm{I}) \Phi)} \tag{2.7}$$

where λ a regularisation parameter, and I is the identity matrix. In this thesis, we will use the R-LDA as it outperforms the original LDA.

2.2. Sparse Representation Classifier (SRC)

Face Recognition using Sparse Representation Classifier (SRC) was proposed in 2009 by (Wright, Yang et al. 2009). SRC assumes that any face image can be represented by a linear combination of training samples of the same person, where the training set for each person is over complete and any face image lies in a spanned subspace.

Let the i[th] class $A_i = [v_{i,1}, v_{i,2}, \ldots, v_{i,n_i}] \epsilon \mathbb{R}^{m \times n}$ represent all training samples of this class. A test face image $y \epsilon \mathbb{R}^m$ from the class i,

$$y = \alpha_{i,1} v_{i,1}, + \alpha_{i,2} v_{i,2} + \ldots\ldots + \alpha_{i,n_i} v_{i,n_i} \quad (2.8)$$

where $\alpha_{i,j} \epsilon \mathbb{R}$ and $j = 1, 2, \ldots\ldots, n_i$. If the whole training samples with K classes can be represented as follows:

$$A = [A_1, A_2, \ldots\ldots, A_k] = [v_{1,1}, v_{1,2}, \ldots, \quad v_{k,n_k}] \quad (2.9)$$

Ideally, the test sample y can be represented within the whole training set as follows:

$$y = A x_0 \, \epsilon \mathbb{R}^m \quad (2.10)$$

where $x_0 = [0, \ldots, 0, \ \alpha_{i,1}, \alpha_{i,2}, \ldots, \alpha_{i,n_i}, 0, \ldots\ldots, 0]^T \epsilon \mathbb{R}^m$, which supposed to have a zero values except those associated with same class of y. In practice, this problem can be formulated as a minimization problem.

$$\hat{x}_0 = min \|x\|_0 \ \ st \ Ax = y \quad (2.11)$$

If some mild conditions are satisfied, the above l^0 minimisation is equivalent to the following l^1 minimisation problem (Candes, Romberg et al. 2006, Candes and Tao 2006, Donoho 2006),

$$\hat{x}_1 = min \|x\|_1 \quad st \ Ax = y \quad (2.12)$$

To classify a given query image y, one first solves the sparse representation $\widehat{\boldsymbol{x}}$ in eq. 2.12. Ideally, the nonzero values of $\widehat{\boldsymbol{x}}$ will be associated with the column $\boldsymbol{A_i}$ of the matrix A, which represent the possible class for the query. Consequently, the test image is classified to the $\boldsymbol{i^{th}}$ class in the training set. However, the presence of noise may lead to some nonzero values outside the $\boldsymbol{i^{th}}$ class. To solve this problem, the input query image y can be classified to the class $\boldsymbol{A_i}$ if the coefficients associated with this class can well reproduce y. For such purpose, the input image $\boldsymbol{y}$ can be approximated as follows:

$$\hat{y}_i = A \delta_i(\hat{x}_1)$$

where δ_i is the function that selects the coefficients of the i^{th} class. Finally, the input image y is classified to the i^{th} class which minimizes the residuals between the approximation $\hat{y}_i$ and the input image y:

$$\min_i \quad r_i(y) = \ \|y - A\delta_i(\hat{x}_1)\|_2$$

In this thesis, we will use the SRC as it is one of the robust face recognition methods.

2.3. Deep Learning for Face Recognition

One of significant difficulties in face recognition is the selection of appropriate features for a particular classification problem. In most of feature selection schemes, the selection of such features is the user's responsibility which is usually time and effort consuming. Further, extracted features are often designed for specific, well-defined problems and may lack generalisation. In practice, the raw input data typically contains irrelevant information and comes with inappropriate form. Deep neural networks are one of the appropriate alternatives for feature selection, which are very popular in computer vision community recently. Using a multilayer architecture and nonlinear functions, deep neural networks are attractive in solving complex problems with high amount dimensions of data. In the next subsections, we first present the Artificial Neural Network as it is the basis of the deep networks, and then we briefly introduce the concept of deep learning techniques used in this thesis. As there are various models of deep networks, we only present the Auto Encoder which is closely related to our work.

2.3.1. Artificial Neural Network (ANN)

Neural Network is a collection of large number of connected atomic processes (neurons) working harmoniously to perform a particular task. ANN has been configured to solve some specific problems such as pattern recognition and classification.

Artificial Neuron is the atomic element of ANN. It contains input function, activation function, and output (Micheli-Tzanakou 2011). Figure.2 illustrates the architecture of an artificial neuron.

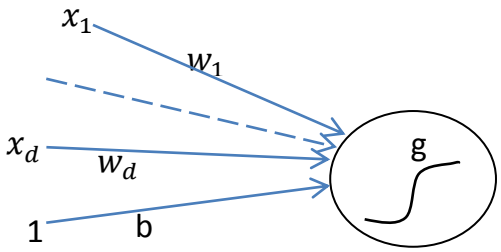


Figure 2.1 Artificial Neuron. W is the connection weighting, b is the neuron bias, and g is the activation function

The neuron output activation function is

$$h(x) = g\big(a(x)\big) = g(b + \sum_i w_i x_i).$$

where a(x) is the input activation function given below

$$a(x) = \mathrm{b} + \sum_{\mathrm{i}} \mathrm{w_i x_i}.$$

where b is the neuron bias, and w is the connection weighting. Usually, a simple Neural Network consists of three layers: The input layer, the hidden layer, and the output layer. The input layer contains raw input data fed into the network. The activity of each hidden unit depends on the activity of input units and the connection between the input and hidden units. The characteristic of the output unit is a direct result of the hidden layer. Figure.2 illustrates this architecture.

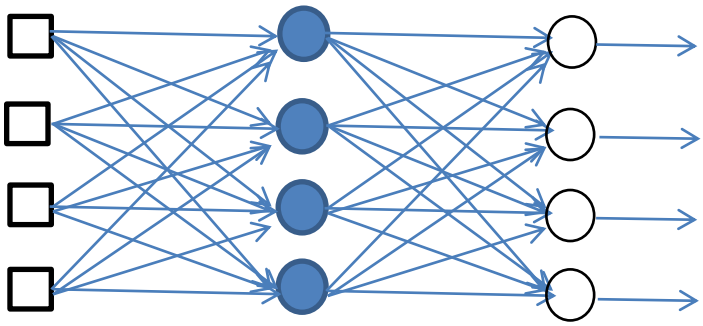


Figure 2.2. Simple Neural Network

The hidden units can be either active or inactive based on the values of weights. A hidden unit can choose what it represents by modifying the weights. The number of layers and neurons depends on the applications, and the neural network is trained to determine its parameters (w and b). The weights are usually initialised to be random values between -1 and 1, or between -0.5 and 0.5, and then these values are modified in each round based on different criteria.

Next, we will describe two types of neural networks.

1. Feed forward Neural Network

In this type of ANN, the signal flows in one direction from the input layer to the output layer, no feedback provided by nodes.

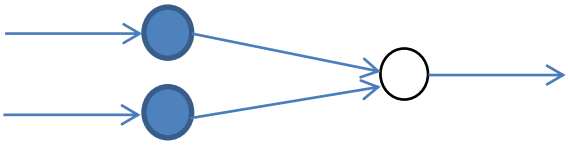


Figure 2.3. Feed forward Neural Network

2. Backpropagation Neural Network

This type of ANN attempts to minimise the mean square error by modifying the weights based on the output of nodes in a backward direction (Hecht-Nielsen 1988).

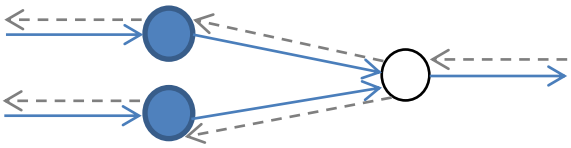


Figure 2.4. Backpropagation Neural Network

Problems with traditional neural networks: The traditional ANNs were once an attractive way for classification problems including face recognition (Ming and Fulcher 1996, Lawrence, Giles et al. 1997, Aitkenhead and McDonald 2003, Haddadnia and Ahmadi 2004) as it doesn't

require a user involvement in designing particular algorithms. However, there are some shortcomings when using this type of neural networks. First, training the entire neural network (i.e, all its parameters related to the interconnected neurons) is difficult as well as time consuming since all parameters are optimized centrally in the designed algorithms. In this case, it is hard to have too many layers for the neural networks; furthermore, random initialisation of weights may be far from the global optimal solution. Recently, deep learning techniques have spiked the popularity of neural networks since some new concepts are proposed to deal with these two critical issues. Next, we will briefly introduce deep leaning neural networks, especially the auto-encoder neural network.

2.3.2. Deep Learning Neural Networks

In deep learning neural networks, models are containing multiple layers in which one can learn data representation on multiple levels. Deep learning methods have resulted in significant improvement in many research areas such as speech recognition and visual object recognition. The key point of deep learning is that the features in different levels are not designed by humans. Instead they are learned from data which make deep learning a general approach. The optimization in obtaining the parameters is done layer by layer replacing the central optimization in traditional neural networks (Bengio 2009, Deng and Yu 2014, Schmidhuber 2015). In the deep neural networks, the random initialisation of the weights may lead to the trap of local minima/ maxima. Consequently, the network is pre-trained to initialise its weights before starting the training process. Next, we will consider the Auto Encoder as one of the popular Deep Neural Network models.

2.3.2.1. Auto Encoder

Auto Encoders attempt to reconstruct the input data during the encoding and decoding processes, and the data will be represented in multi-level of abstractions during the encoding and decoding. The output aims to reconstruct the ideal/target data as shown in Figure 2.5.

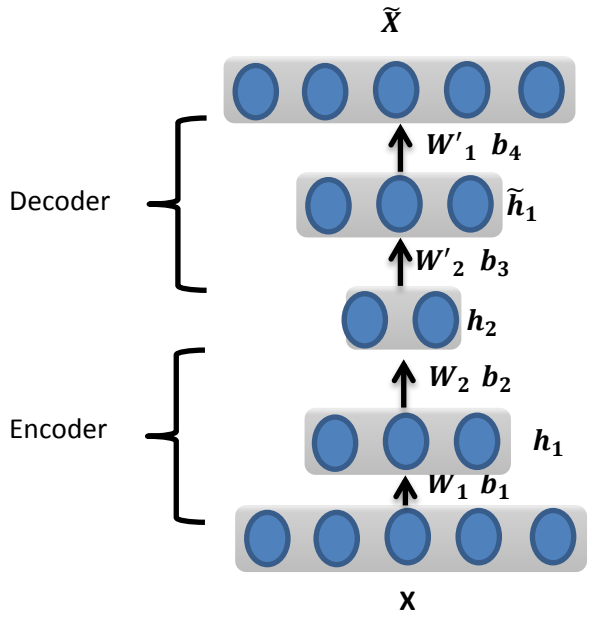


Figure 2.5. Architecture of Auto Encoder

An activation function g is used for encoding and decoding.

For encoding:

$$h_1 = g(W_1 X + b_1)$$

$$h_2 = g(W_2 h_1 + b_2)$$

For decoding:

$$\tilde{h}_1 = g(W'_2 h_2 + b_3)$$

$$\tilde{X} = g(W'_1 \tilde{h}_1 + b_4)$$

As shown in Figure 2.5, the output layer has the same number of nodes of the input layer. The part that computes the hidden layers is known as Encoder where the part that computes the output from the hidden representation is known as Decoder. The weights W_i are initialised to pre-trained values. For face recognition task, one of the hidden layer features is used with an appropriate face recognition technique (such as LDA). For other tasks like face pose correction (Lai, Dai et al. 2015) and de-noising (Pathirage, Li et al. 2015), the targeted face image is used in loss function instead of the original input image.

2.4. Image Representation for Face Recognition

Different image representations are used in face recognition and they are extracted from raw pixel values in an image (Yip and Sinha 2002, Jian and Chengjun 2008, Chengjun and Jian 2009, Jing, Liu et al. 2010, Yang, Liu et al. 2010, Li, An et al. 2011, Ali, Hussain et al. 2012). This data can be utilised either directly or indirectly for face recognition task. In this section, we will detail some image representations based on transforms. Grey and color models for images will be illustrated in some details hereafter.

2.4.1. Face Recognition for Grey Images

In early development stage of face recognition systems, the grey-scale images are the basis of image representation. Here, we will describe grey images and the related transform: Local Binary Pattern (LBP).

2.4.1.1. Grey Images

In this representation, any image is represented by one layer in 2D space called grey scale (Gonzalez, Woods et al. 2009). Each pixel in this greyscale is represented by a value indicating the intensity of this pixel. The grey scale can represent 256 grey-scales from 0 to 255, where 0 represents black, and 255 represents white. The early face recognition methods usually use the grey images rather than color images. Even though most of the recent techniques of face recognition utilise color information, grey images are still widely used in many facial recognition methods as we will see later in this thesis. One of the most popular image representation derived from the grey image is the Local Binary Pattern (LBP) (Ali, Hussain et al. 2012).

2.4.1.2. Local Binary Patterns

Local Binary Patterns (LBP) is widely used for face recognition due to the discriminative power and simplicity (Meng, Gao et al. 2010, PietikÃ¤Inen 2010, Yang and Chen 2013, Zhong and Zhang 2013, Zhou, Liu et al. 2014). The LBP describes the image texture based on local

variations. In the simple model of LBP, each pixel value is replaced by its LBP value which is computed based on its neighbouring pixels as illustrated below.

The binary values B of the neighbouring pixels Z are computed as follows.

$$B_j = \begin{cases} 0 & Z_j - P(x,y) < 0 \\ 1 & Z_j - P(x,y) \geq 0 \end{cases} j = 0,1,\dots,7 \tag{2.13}$$

where P(x,y) is the grey value of the pixel (x,y). The LBP value is computed as:

$$LBP(x,y) = \sum_{j=0}^{7} (2 \times B_j)^j \tag{2.14}$$

Figure 2.6 illustrates the method of obtaining the LBP value for a given pixel.

This representation can be used for face recognition in two different ways (Yang and Chen 2013).

1. LBP image. In this scenario, the LBP representation is used with one of face recognition methods such as PCA or LDA features (Ahonen, Hadid et al. 2004, Hadid, Pietikainen et al. 2004).
2. LBP histogram. The Histogram features are extracted in a block-wise manner, and then these features are reduced and used for face recognition (Ahonen, Hadid et al. 2006, Huang, Shan et al. 2011).

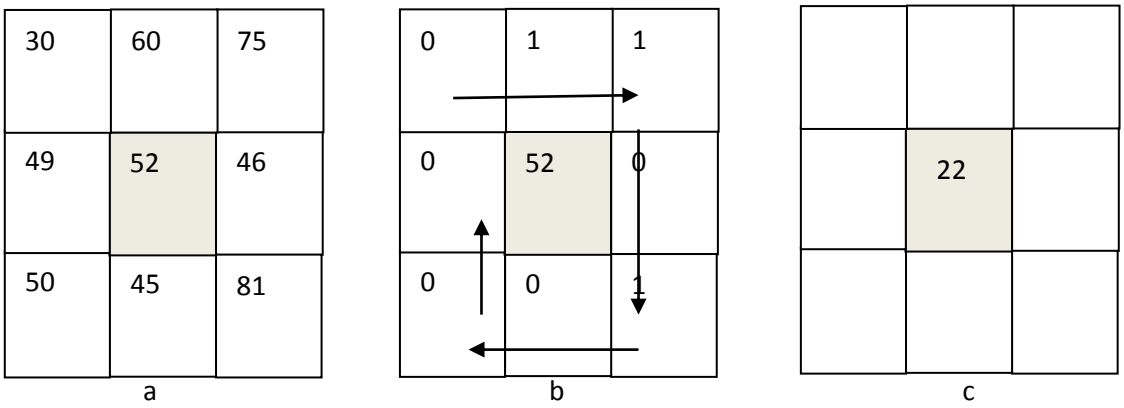


Figure 2.6 Pixel-wise LBP. a. 3X3 block. b. Binary values of neighboring pixels according to the LBP criteria. c. LBP value of centered pixel.

2.4.2. Color Images

As we mentioned earlier, the color information provides valuable discriminant information which is very useful for face recognition. The basis of all color spaces is the RGB (Red, Green and Blue) as all other color spaces can be obtained from the RGB using some transformation functions/matrices.

2.4.2.1. RGB Color Space

RGB color space represents an image as three layers of a 2D array in which each layer represents one color. These three layers are stacked to create all visible colors (Gonzalez, Woods et al. 2009) from the Red, Green and Blue. Given an image I of size $M \times N$, the RGB representation can be denoted as:

$$I_{RGB} = \begin{bmatrix} R \\ G \\ B \end{bmatrix} \epsilon \, \mathbb{R}^{3 \times d} \, , d = M \times N \tag{2.15}$$

RGB is the fundamental color space that all other color spaces and models can be derived either linearly, nonlinearly, directly or indirectly from this color space.

2.4.2.2. The Linearly Derived Color Spaces

These color spaces can be derived from the RGB images using a linear transformation matrix. Here we will give some examples of these color spaces:

- The YIQ color space containing three color components, luminance (Y), hue (I), and saturation (Q). This color space is designed for analogue TV, and it can be derived from RGB as follows (Gonzalez, Woods et al. 2009).

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0596 & -0.247 & -0.322 \\ 0.211 & -0.253 & 0.312 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{2.16}$$

- The YCbCr color space used in the digital video (Gonzalez, Woods et al. 2009). Y component represents the luminance, and Cb and Cr represent the color information.

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.481 & 128.533 & 24.966 \\ -37.797 & -74.203 & 112.000 \\ 112.000 & -93.786 & -18.214 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{2.17}$$

- The XYZ color space proposed by e International Commission on Illumination (CIE) (Weeks 1996). The XYZ color space can be derived from RGB as follows.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.607 & 0.299 & 0.000 \\ 00.174 & 0.587 & 0.066 \\ 0.201 & 0.114 & 0.1.117 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{2.18}$$

- The $I_1I_2I_3$ color space proposed by (Ohta 1985) using Karhunen-Loève transform (Jain 1976). The $I_1I_2I_3$ can be obtained as follows:

$$\begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} = \begin{bmatrix} 0.333 & 0.500 & -0.500 \\ 00.333 & 0.000 & 1.000 \\ 0.333 & -0.500 & -0.500 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{2.19}$$

2.4.2.3. The Non-linearly derived Color Spaces

These color spaces are derived using some nonlinear functions. One typical example of these color spaces is the $L^*a^*b^*$ (Weeks 1996) which derived from the XYZ color space as follows:

$$\begin{aligned} L^* &= 116 f(\frac{Y}{Y_n}) - 16 \\ a^* &= 500 \left[f\left(\frac{X}{X_n}\right) - f(\frac{Y}{Y_n}) \right] \\ b^* &= 200 \left[f\left(\frac{Y}{Y_n}\right) - f(\frac{Z}{Z_n}) \right] \end{aligned} \tag{2.20}$$

where $$f(t) = \begin{cases} t^{1/3}, & if\ t > (\frac{6}{29})^3 \\ \frac{1}{3\left(\frac{29}{6}\right)^2} + (\frac{4}{29}), & otherwise \end{cases}$$

2.4.2.4. Normalised Color Spaces

The normalised color spaces nRGB and nXYZ are proposed by (Yang, Liu et al. 2010) to enhance face recognition accuracy by achieving the Double Zero Sum (DZS) as they observed that the color spaces with DZS property such as $I_1 I_2 I_3$ can achieve better recognition accuracy. The nRGB and nXYZ can be achieved as follows:

$$\begin{bmatrix} nR \\ nG \\ nB \end{bmatrix} = \begin{bmatrix} 1.000 & -0.333 & -0.333 \\ 0.000 & 0.666 & -0.333 \\ 0.000 & -0.333 & 0.666 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{2.21}$$

$$\begin{bmatrix} nX \\ nY \\ nZ \end{bmatrix} = \begin{bmatrix} o.6070 & -0.0343 & -0.3940 \\ 0.1740 & 0.2537 & -0.3280 \\ 0.2000 & -0.2193 & 0.7220 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{2.22}$$

2.4.2.5. Training based Color Spaces (Color Models)

These color spaces are obtained from RGB via some learning algorithms in order to obtain better image representation for face recognition. In these color spaces, there is no predetermined transformation matrix. A training process is used to learn a new color space. The transformation matrix may vary based on the training data used in the training phase. We call this kind of color spaces as *color models* because they are modelled specifically for face recognition. Furthermore, there is no specific transformation that can specify any of these models. However, we will use color space and color model interchangeably hereafter in this thesis if no confusion arises. Next we will introduce two types of color models.

- The Uncorrelated Color Space (UCS) (Liu 2008). This color space can be derived from RGB based on PCA criteria. A training set is needed to train the model. Given a training dataset A containing N training samples. The covariance matrix is calculated as:

$$C_A = \frac{1}{N} \sum_{i=1}^{N} (A_i - \bar{A})^T (A_i - \bar{A}) \tag{2.23}$$

where $\bar{A}$ is the mean of all images in the dataset. The transformation matrix can be formulated as:

$$C_A = t^T{}_{UCS}\, \Lambda\, t_{UCS} \tag{2.24}$$

where Λ is the Eigen value matrix.

- The Discriminant Color Space DCS. This color space is proposed by (Liu 2008), and its aim is to discriminate the color component using LDA criteria instead of PCA. A matrix W is used to maximise the following function

$$J(W) = \frac{tr(W^T L_B W)}{r(W^T L_w W)} \tag{2.25}$$

where $L_B\ and\ L_w$ are the between scatter matrix and within scatter matrix respectively.

$$L_B = \sum_{j=1}^{M} (\bar{A}_j - \bar{A})^T (\bar{A}_j - \bar{A}) \tag{2.26}$$

$$L_B = \sum_{j=1}^{M} \sum_{A_i \epsilon p_j} (A_i - \bar{A}_j)^T (A_i - \bar{A}_j) \tag{2.27}$$

where $\bar{A}_j$ is the jth class mean and the Eq.22 can be solved as follows.

$$L_B W = \lambda\, L_w W \tag{2.28}$$

With the learned transformation matrix W, one can obtain the DCS model as below.

$$[D^1, D^2, D^3] = [R, G, B]\, W \tag{2.29}$$

where $D^1, D^2\ and\ D^3$ are the color components of the DCS model.

- Local Discriminant Color Space. This model proposed by (Li, Liu et al. 2012) applied the DCS model locally in pixel level (PLDCS) and block-wise (BWDCS). As different local areas in the face image have different colors and shapes, the best representation for face recognition may vary from one area to another. The DCS procedure is applied on each area separately, and as a result, each block will have its learned transformation matrix (W_{Block}). In the PLDCS, each pixel in the face image has its learned transformation matrix (W_{Pixel}).

- Other trained color spaces. There are many other training-based color spaces proposed for face recognition. Independent Color Space (ICA) (Liu 2008) is obtained utilising the Independent Component Analysis (ICA) (Comon 1994) to achieve independent color components. The Holistic Orthogonal Analysis HOA proposed by (Jing, Liu et al. 2010) uses LDA criteria to attain Orthogonal discriminant transforms for the red, green and blue components.

All above mentioned color models are based on exactly three color components to represent a face image. However, recent work shows the superiority of using multiple colors as in (Li, An et al. 2011) as described below.

2.4.2.6. Multiple Color Fusion Model (MCF)

The MCF color model (Li, An et al. 2011) breaks the restriction of using only three color components to represent a face image. In (Li, An et al. 2011), nine different color spaces are utilised (containing 39 color components). After removing the duplicated and the highly correlated color components, only 22 color components are used in the training process. The training process can be briefly described as follows:

In the first round, each color component is trained for face recognition; the color component that achieves the highest recognition rate is selected as a first component of the MCF model.

The second color component is achieved by adding the remaining color components one by one to the first component and applying face recognition. The best complementary color component is kept as a second color component of the MCF model.

This scenario is repeated to obtain the 3rd, 4th, …, and n^{th} color component until the recognition rate starts to decrease. The obtained MCF model is used for face recognition which can achieve higher recognition rate in comparison with different three channels based color spaces. The basic idea behind the MCF model is that the optimal image representation for face recognition can lie in a space spanned by more than three color components, and the training for face recognition is the best way to derive this representation.

2.5. Summary

This chapter has presented some background related to this thesis. Firstly, PCA and LDA are introduced as examples of subspace face recognition methods. Secondly, the SRC classifier is presented briefly. Thirdly, The Artificial Neural Network (ANN) and Deep Learning neural networks are described. In the last section, different 2D image representations are presented including grey and color images. In this section, different three channels based color spaces are described. Finally, the Hybrid Multiple Color Fusion (MCF) is reviewed as it is the only color model with more than three color channels.

Chapter Three

Face Recognition against Mouth Shape Variations

Facial expression is one of the challenging issues in face recognition as expressions can significantly change the shape and appearance of the whole face as well as locations of different face organs. Most of the work on face recognition against expressions is dealing with the problem from only one perspective, which is the change of facial shape and appearance. The main goal on this issue in existing literature is to distinguish between the neutral face and face with expressions as discussed in (Lee and Kim 2008), where the features extracted from *the Active Appearance Model* (AAM) (Cootes, Edwards et al. 2001) are used to represent facial image, and a query face image is recognised by matching this feature vector to the features of gallery images. A recent method proposed in (Taheri, Patel et al. 2013) splits up the components of facial features and facial expressions by utilising the sparsity and morphological diversity, and it is used both for face recognition and expression recognition. The method proposed by (Ramachandran, Zhou et al. 2005) attempts to neutralise a face image with smiling facial expressions to enhance face recognition. The CANDIDE (Rydfalk 1987) face model is used to stretch, press, and depress the lips and lip corners, then the geometry mesh is transformed to a neutral state, and the wrinkles are removed from the texture. This work has been designed specifically for smiling face rather than for other expressions associated with mouth shape variations. In the current literature, there is no systematic investigation on mouth-shape variations due to expression changes and it is obvious such mouth-shape variations will have a significant impact on face recognition. The geodesic polar representation (Mpiperis, Malassiotis et al. 2007) is proposed for 3D face recognition by considering the opened-mouth case. In this method, the opened-mouth is recognised based on the connectivity between the upper and the lower lips. In the case of opened mouth, the connection between upper and lower lips will be lost. Consequently, the geodesic paths are deformed, and the isometry assumptions are failed if the lips are not connected. This problem is solved by first classifying the mouth as Opened or Closed using the Support Vector Machine (SVM) classifier (Osuna,

Freund et al. 1997), where the AAM (Cootes, Edwards et al. 2001) is used to detect the lips. The geodesic paths are modified for *dis-connected lips*, where the straight geodesic paths are used with *connected-lips*. This method is proposed specifically for 3D face recognition in which the detection of tiny Opened-Mouth cases is highly required. To the best of our knowledge, there are no such investigations on 2D face images in the existing literature.

Based on this observation, the Opened-Mouth problem for 2D face images is investigated in this chapter and it can be considered as a very special case of mouth shapes and facial expressions due to:

- Opened-Mouth changes the locations of the chin and nose.
- Opened-Mouth changes the geometry of the whole face image.

In this chapter, the Opened-Mouth problem in face recognition is introduced, and then an effective solution is proposed for 2D face recognition. The rest of this chapter is organised as follows: In the next section, the significance of the Opened-Mouth problem is presented, and different *Mouth Shape Variations* are explained. In section 3.2 a new approach is proposed to solve this problem. Experimental results and discussions are demonstrated in section 3.3. Finally, the chapter is concluded in section 3.4.

3.1. The Opened-Mouth Problem

The Opened-Mouth case can be considered as a special outcome of different expressions. Before addressing this problem formally, different possible mouth shapes will be explained first.

3.1.1. Possible Mouth Shapes

In this thesis, we categorize mouth shapes as one of the following situations:

1. **Closed Mouth**: the lips are connected, and the teeth are not visible. We denote this shape as *C*.
2. **Smiling Face** (Closed Mouth with teeth): the teeth are visible and connected, and the lips are not connected. We denote this shape as Ct (Closed with teeth).

3. **Opened Mouth**: The lips are not connected (upper and lower lips are apart), and the teeth are not visible. We denote this shape as O.
4. **Opened Mouth with teeth**: The lips are not connected, and the teeth are visible but not connected. We denote this shape as Ot.

In face recognition, the shapes of Closed-Mouth will not change the face recognition accuracy significantly as they often appear in neutral expressions. However, the shapes of Opened Mouth and Opened Mouth with teeth have a significant impact on face recognition as we addressed below, we need to develop an approach to solve such changes.

3.1.2. The Significance of the Opened Mouth Problem (O and Ot)

The Opened-Mouth situations (O and Ot) are different from any other mouth shapes because:

- O and Ot can change the locations of some face organs such as the chin, lips and the nose as shown in Figure 3.1(a).
- They can change the isometry of the whole face, that is, the ratio between the height and the width of a face can be changed when the mouth is opened.
- O and Ot can add an undesirable face area, which is regarded as mouth cavity as highlighted in Figure 3.1(b). This area can be considered as noise in face recognition process.

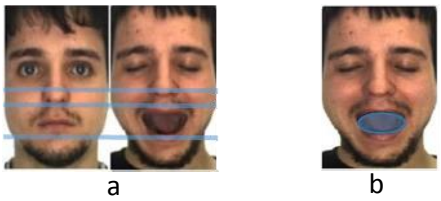


Figure 3.1 .a: Misalignment caused by opened mouth.
b: Undesired information in opened mouth area

In the literature of face recognition, the negative effect of noise and misalignment is very well addressed. So, for better performance, face images need to be aligned and normalised before passing it to a classifier. The Opened-Mouth usually brings changes of face shape (O and Ot), and as a result, the face recognition accuracy can be significantly degraded. Most importantly, the patch based (or block-wise) face recognition methods can be affected more than the holistic methods, as the corresponding blocks or patches will be mismatched, and thus face recognition

may fail. Based on this observation, it is expected that recovery from the Opened-Mouth can significantly improve the recognition accuracy. As shown in Figure 3.2, effective correction of the Opened-Mouth can eliminate the inner mouth area (cavity) and also can correct some misalignments.

The presence of teeth is important to the opened mouth and closed-mouth classification as demonstrated in next section. That is why we use the teeth as an important indicator in these four categories.

Figure 3.2 a: Closed Mouth. b: Opened Mouth.
c: Recovered from Opened Mouth

For further illustration, we examine the shapes of Opened-Mouth and Closed-Mouth. As can be seen in Figure 3.3(a), the landmarks on inner lips form a semi-straight line in the case of Closed-Mouth. On the other hand, those on opened mouth forms two separate curves as in Figure 3.3(b). In fact, the opened mouth changes the geometry and locations of lips, nose, chin, and moustache. Logically, the correction of Open-Mouth should be a right solution to recover the original shapes and locations of these organs, and consequently we expect the face recognition accuracy can be improved.

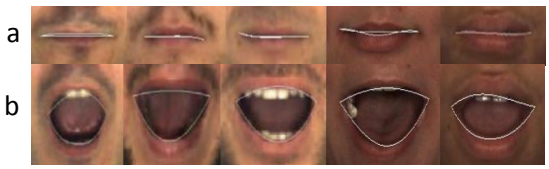


Figure 3.3 a. Inner shape of the Closed-
Mouth. b. Inner shape of the Opened-Mouth

The geometric models like CANDIDE may fail in recovery from Opened-Mouth. Such models use landmarks for the whole face image, which may result in a large accumulative error resulted by landmarks positioning errors, the widely Opened-Mouth and the presence of teeth may complicate the problem. In the next section, a solution for recovery from the opened

mouth is proposed by using only the inner lips landmarks. As to landmark detection, the recently proposed AR model (Liang, Liu et al. 2014) is used.

3.2. Opened Mouth Correction Algorithm

The recovery from opened-Mouth can be divided into two separate stages which are: Mouth shape detection and mouth shape correction. Before performing the correction procedure, the mouth shape must be recognised. As mentioned earlier in this chapter, the mouth shape can be in one of the following cases: **Closed** (C), **Closed with teeth** (Ct), **Opened** (O), and **Opened with teeth** (Ot). After classifying the mouth shape into one of the above-mentioned cases, the correction process is performed in the cases of Ct, O, and Ot. In this chapter, we only consider frontal and cropped face images for simplicity. It is worthy to mention here that we don't use illumination normalisation or any other pre-processing techniques.

3.2.1. Mouth Shape Detection

This is the first step in the process of mouth shape correction. In this phase, the mouth shape is classified to C, Ct, O or Ot by utilising the inner mouth landmarks and teeth. First, some terminologies are defined as below to represent the detection accuracy:

- *False negative detection* indicates that the mouth is opened (O or Ot) but classified as closed (C or Ct). As Ot can be considered as a subclass of O, the incorrect classifications of O$\in$ Ot and Ot $\in$ O are not considered as *False negative detection.*
- *False positive detection* indicates that the mouth is closed (C or Ct) but it is classified as opened (O or Ot).

By detecting the inner mouth landmarks, the mouth shape can be classified to one of the four shapes based on the connectivity between the *upper and lower lips landmarks*, taking the presence of the teeth in the account. We will address this detection problem in the next section.

3.2.1.1. Mouth Landmark Detection and Lips Connectivity

In this step, the AR model (Liang, Liu et al. 2014) is used for landmark detection. This model is an extension one proposed by (Zhu and Ramanan 2012). It increases the number of detected

landmarks as well as the accuracy of the detected landmarks. The AR model extends the tree structure to cover the shapes of facial components. Furthermore, it improves the accuracy of landmarks positioning. For mouth shape detection, only the inner mouth landmarks are used. To decide whether the lips are connected or not, the Middle Inner Landmark of Upper lip (MU) and the Middle Inner Landmark of Lower Lip (ML) are used, the vertical distance D_v between MU and ML is used for this purpose as shown in Figure 3.4.

$$D_v = MU_v - ML_v \tag{3.1}$$

where MU_v and ML_v are the vertical coordinates of the MU and ML respectively.

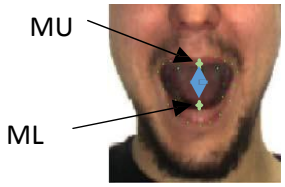


Figure 3.4 The Vertical distance between ML and MU

Obviously the mouth is closed if the $D_v = 0$. However, there are some *false positive detection* cases in landmark detection and we need to take some errors into consideration. From our experimental observation, the error of landmark detection usually cannot exceed 2% of the face image height. Consequently, this ratio will be used as a degree of tolerance for the C class:

$$Mouth\ shape\ =\ \mathrm{C}, \quad \text{if}\ \ D_v \leq \frac{2(hieght(image))}{100}$$

Through extensive experiments, we observed that this tolerance is satisfactory for face recognition.

The landmarks on inner lips are used to estimate all Upper Lip Points (ULP) and Lower Lip Points (LLP). These points draw the inner contour of lips and will be used later in mouth shape correction. Figure 3.5 shows these points.

Figure 3.5 . Estimation of all inner mouth points

3.2.1.2. Teeth Detection

The use of inner mouth landmarks is validated to be a good approach to detecting the mouth shapes. However, this is true only if the teeth are hidden behind the lips because the lips can be separate even in the closed mouth but with smiling cases. Moreover, if the teeth are present, and the mouth is opened, the closing of mouth based on lips landmarks will deform the isometry of face image. Consequently, the present of teeth should be taken into account. First, the teeth are detected, and then they used to classify the mouth shape to one of the previously mentioned four classes. Moreover, the teeth detection is important for us to correct the classes Ct and Ot into the class C while maintaining the geometry of the face. To detect the existence of teeth, we need to use color information as there are no existing algorithms for landmark detection for teeth. we observed that the color of mouth cavity and the color teeth are distinct, and can be separated using a simple segmentation algorithm as described Algorithm 3.1 in this section. Figure 3.6 shows the histogram map of teeth and cavity in the R, G, and B color components of RGB space. The values of the cavity concentrate in the range of [50, 100] where the pixel values of teeth concentrate in the range of [100, 200]. This can be clearly seen in Figure 3.6.

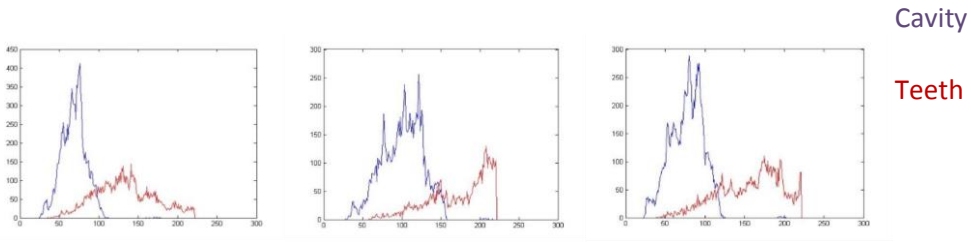


Figure 3.6 Histogram showing R, G, and B components of mouth cavity and teeth

The inner mouth area is segmented into teeth and cavity based on the values of pixels and a pre-trained threshold. The threshold is obtained based on the average of the cropped mouth image (M_{Av}). Our analysis based 30 images in classes Ct, O and Ot shows that the best threshold is 1.4 × M_{Av}. The selected 30 images also cover some lighting variations to cope all possible cases. Keep in mind that the segmentation is conducted on the inner mouth area rather than the whole mouth area. For this purpose, a growing object is used starting from the middle of mouth cavity, growing up and down and it stops when touching the teeth or the inner lips boundaries according to the threshold as illustrated in Figure 3.7. Then, the object is growing into right and left directions to cover all cavity area. When the growing object stops growing, it draws the contour of the mouth cavity; here we denote the upper boundaries of the object as UCP referring to the Upper Cavity Points, and LCP which refers to the Lower Cavity Points. As in detecting the lips connectivity, the vertical distance between the upper teeth and the lower teeth is used to detect the teeth connectivity and mouth shape. These two points are the middle top and middle bottom points of the growing object if they do not match the points MU and ML. In Figure 3.7, they are denoted as TU and TL which refer to Upper Teeth and Lower Teeth.

Figure 3.7 Teeth Detection using a moving object and an appropriate threshold

Mouth Shape Detection Algorithm (Algorithm 3.1)

Input: a cropped frontal face image.

1. *Detect landmarks by using the AR model.*

2. *Find the middle points of upper lip and lower lip and compute the vertical Distance D_v between MU and ML.*

$$D_v = |MU_v - ML_v|$$

3. $if \;\; D_v \leq \frac{2(hieght(image))}{100} \; then, Mouth \; shape \; = \; C, \; exit.$
4. *Use the outer mouth landmarks to crop the mouth image and compute the M_{Av} of the input image.*
5. *Set the threshold to be M_{Av} × 1.4..*
6. *Detect the lower point of middle Upper Teeth TU and the upper point of middle Lower Teeth* TL as described in the previous section.
7. *If TU ≈ MU and TL ≈ ML then, Mouth shape = O, Exit.*
8. *Compute the vertical distance between TU and TL,*

$$D_{vt} = |TU_v - TL_v|$$

9. *If D_{vt} ≈ 0 then mouth shape = Ct, exit.*
10. *Mouth shape = Ot.*

With the above algorithm, we can detect the four shapes for a mouth in addition to all inner lips points and mouth cavity contour. In the next section, we will use these points to correct different mouth shapes to a closed mouth shape. In the proposed correction algorithm we will use the Lower Lips Points (LLP), Upper Lips Points (ULP), Lower Cavity Points (LCP) and Upper Cavity Points (UCP). It is clear that UCP ≈ULP and LLP≈LCP only in the absence of teeth.

3.2.2. Correction Algorithm from Opened Mouth to Closed Mouth

Based on the above-mentioned analysis, the mouth shape can be corrected from the classes O, Ot and Ct to the Class C based on landmarks. In the proposed algorithm below we use terms *stretch, shrink* which refer to the scaling using the cubic interpolation (Gonzalez 2009) which is widely used in digital image scaling.

Mouth Shape Correction Algorithm (Algorithm 3.2)

1. After detecting the landmarks using the AR model, all inner points of the Upper Lips Point (ULP) and Lower Lips Points (LLP) can be obtained.
2. Detect the Upper and lower teeth positions, obtain the UCP and LCP as described previously, and then classify the mouth shapes into C, Ct, O or Ot.
3. Divide the face image into three parts Right (R), Middle (M), and Left (L). Then further divide the Middle area into Upper Middle (UM), Middle Middle (MM), Cavity (CA) (if detected) and Lower Middle (LM) as shown in Figure 3.8.
4. Transform the cases of Ct and Ot into C and O respectively. This can be accomplished by:
    a. Stretching MM columns from ULP to UCP to cover the lower teeth, and
    b. Stretching LM columns from LLP to UCP to remove the upper teeth area.

    After this step the mouth shape will be in O shape or C shape.

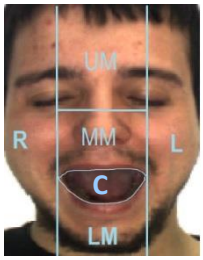


Figure 3.8 Dividing face image into L, R, UM,

5. Correct the LLP and ULP points to LCP and UCP such that, LLP=LCP, ULP=UCP.
6. Stretch the MM area and LM area vertically column by column based on LLP and ULP until reaching the vertical coordinate of MU landmark as illustrated in Figure 3.9.b. At this step, the mouth cavity (CA) is illuminated and the mouth becomes closed. However, the size of chin and lower lip is larger than the normal size and the symmetry of whole face image is not regular.
7. Shrink the LM area vertically according to Eq (3.2) to restore the normal size of chin and lower lip (Figure 3.9.c).

$$Hieght_{LM} = Hieght_{LM} - Hieght_{CA} \quad (3.2)$$

8. To restore the isometry of whole face image, stretch the face image vertically to restore the original image size (Figure 3.9.d).

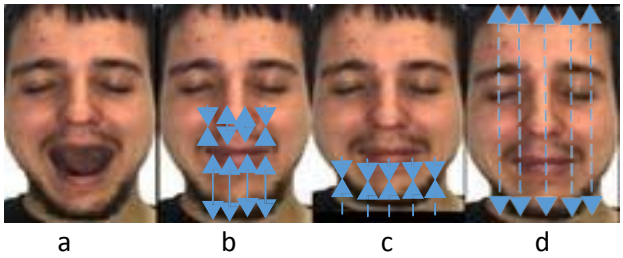


Figure 3.9 a. Detected opened mouth; b: stretching/shrinking LM and MM to eliminate mouth cavity; c: shrinking lower part of the face to recover the original size of chin; d: recovering the original size of face image.

With above correction algorithm, we not only can recover the closed mouth shape, but also restore the neutral shape and natural dimensions of face as a whole, as well as the shapes of different organs and their locations. It is noteworthy to mention that the above algorithm corrects the openned mouth rather than other facial expressions which is out of scope of this work. Figure 3.10 shows some examples of corrected images. As we can see in the experiments in the next section, this correction/normalisation will lead to a significant improvement in facial recognition accuracy.

Figure 3.10 Some samples of recovered face images

3.3. Experiments

In the experiments of this section, the influence of mouth shapes on face recognition accuracy will be investigated and the proposed algorithm for mouth shape correction is evaluated. One important task is to show the improvement of face recognition after mouth corrections. For such purpose, two different face databases are used including the AR database (Martinez 1998) and BU-4DFE database (Lijun Yin, Xiaochen Chen et al. 2008). For simplicity, the PCA (Turk and Pentland 1991) and Regularized LDA (Lu, Plataniotis et al. 2005) are used for dimensionality reduction and the Nearest Neighbor NN is used for classification with Euclidian distance measure.

3.3.1. Databases Setup

In these experiments, the AR and BU-4DFE (Martinez 1998, Lijun Yin, Xiaochen Chen et al. 2008) databases are used to evaluate the proposed method.

3.3.1.1. The AR database

The AR database (Martinez 1998) is a 2D face database containing 2600 face image covering various expressions, lighting conditions and occlusions. Furthermore, the images were taken in two sessions separated by a period of six months. These images belong to 100 subjects, 50

males and 50 females. All images in the database are frontal, cropped and well aligned. For each subject, there are two smiling faces and two Opened-Mouth faces. In our experiments, eight images are used, two neutral, two smiling faces, two Opened-Mouth faces and two faces with other expressions. In the first experiment, the smiling faces are used as query images where the rest six images including the Opened-Mouth faces are used as gallery images. In the second experiment, the Opened-Mouth faces are used as query images where the rest six images including the smiling faces are used as gallery images. Examples of the Opened-Mouth face are shown in Figure 3.11.a.

3.3.1.2. The BU-4DFE Database

The BU-4DFE database (Lijun Yin, Xiaochen Chen et al. 2008) contains video sequences of 3D images. It is mainly designed for facial expressions recognition and tracking. This data base covers different facial expressions including angry, disgust, fear, happy, sad and surprise. The images in the database belong to 101 subjects, 58 female and 43 male with different skin colors. Each subject has sequences of the all mentioned expressions. The total number of images is up to 60000 still images. All images are frontal and upright. In the surprised face images, there are a large number of images with Opened-Mouth. So, we use these images to evaluate the proposed method. For each person, 50 images are used such that, 44 Opened-Mouth images and 6 closed mouth images. The Opened-Mouth images are used as query images in all cases of experiments and the closed mouth images are used as galleries. However in the cases of using less than 6 gallery images, some of closed mouth images are used as query images. Figure 3.11.b shows some examples of the Opened-Mouth images.

a

b

Figure 3.11 a. Samples of Opened-Mouth images from AR database. b. Samples of Opened-Mouth images from BU-4DFE database.

3.3.2. Experiment One

This experiment is applied on the AR database. It can be divided into three stages which include mouth shape detection, mouth shape correction and face recognition. The PCA and rLDA are used for dimensionality reduction and feature extraction where the Nearest Neighbour NN with Euclidian distance measure are used for classification.

3.3.2.1. Mouth Shape Detection

In this step, the all 800 face images in the dataset are passed to the mouth shape detector. For the Opened-Mouth images, 99% of accuracy is achieved with 0% of false positive detection. For the smiling faces, 100% of accuracy is achieved. Even though the detection of the images with Opened-Mouth and teeth (Ot) is 97.5%, the false negative detection is only 0.5% which implies that the rest of 2% misclassified images are assigned to the class O. Detailed results are shown in table 3.1.

Table 3.1. Mouth shape detection for AR database

Correct Class / Classified to	C	Ct	O	Ot
C	100%	0%	1%	0.5%
Ct	0%	100%	0%	0%
O	0%	0%	99%	2%
Ot	0%	0%	0%	97.5%

3.3.2.2. Mouth Shape Correction

In this phase, all images that classified to classes O, Ot and Ct are corrected using the proposed correction method in the section 3.2.2.

3.3.2.3. Face Recognition

In this phase, the effect of corrections from Opened-Mouth and smiling mouth is evaluated for face recognition. For this purpose, the experiment is divided into two parts. In part 1, we conduct face recognition on the corrected smiling faces as query images and the neutral faces as training samples, where the corrected images from the Opened-Mouth are used in part 2. As can be seen in table 3.2, the smiling faces (Mouth with the presence of teeth) doesn't have a significant impact on face recognition where 100% of accuracy is achieved by rLDA using five neutral gallery images. On the other hand the effect of Opened-Mouth on face recognition accuracy is very clear as shown in table 3.3. More importantly, the improvement achieved by the correction from Opened-Mouth is very high (10% to 21.5%).

Table 3.2. Face recognition on corrected smiling faces of AR database in comparison with the original smiling face images

Test : Gallery	Smiling (Original)	Closed (Recovered)
PCA		
2 : 1	80%	79%
2 : 2	96%	96%
2 : 3	97%	96.5%
2 : 4	97.5%	97.5%
2 : 5	98%	85 %
2 : 6	98.5%	98%
rLDA		
2 : 2	99%	99%
2 : 3	99.5%	100%
2 : 4	100%	100%
2 : 5	100%	100%
2 : 6	100%	100%

Table 3.3 Face recognition on corrected Opened-Mouth faces of AR database in comparison with the original open mouth face images.

Test : Gallery	Opened-Mouth (Original)	Closed mouth (Recovered)	Improvement
PCA			
2 : 1	48%	69.7%	21.7%
2 : 2	63%	76%	13%
2 : 3	64.5%	75%	10.5%
2 : 4	68.5%	83%	14.5%
2 : 5	73%	85.5%	12.5%
2 : 6	74%	87%	13%
rLDA			
2 : 2	58%	68%	10%
2 : 3	74.5%	92%	17.5
2 : 4	78%	94%	16%
2 : 5	86%	97.5%	11.5%
2 : 6	87%	97.5%	10.5%

3.3.3. Experiment Two

This experiment is conducted on the BU-4DFE database which contains a larger number of Opened-Mouth face images as described earlier. Here, only images with Opened-Mouth are considered as queries, where smiling faces are not included. By using 44 Opened-Mouth images per subject, this can evaluate our method better than the experiment 1 where the AR database contains only two Opened-Mouth images per subject. The PCA and rLDA are used for feature selection and the Nearest Neighbour NN with Euclidian distance measure is used for classification. As in experiment 1, this experiment is divided into three steps as follows.

3.3.3.1. Mouth Shape Detection

In this step, all images are passed to the mouth shape detector and each image is classified to one of the four classes. Results of the detection process are shown in Table 3.4. Even though there is a small error rate in the detection, this will not affect the recognition accuracy as we will show later in the face recognition step where 100% is achieved by the corrected images using only five neutral gallery images.

Table 3.4 Mouth shape detection for the BU-4DFE database

Correct class / Classified to	C	Ct	O	Ot
C	100%	N/A	2%	1%
Ct	0%	N/A	0%	0%
O	0%	N/A	98%	2%
Ot	0%	N/A	0%	97%

3.3.3.2. Mouth Shape Correction

Mouth shapes of classes O and Ot are corrected into class C as described earlier.

3.3.3.3. Face Recognition

In this phase, the proposed correction method is evaluated by comparing the recognition accuracy of the corrected images with the original ones. The face recognition process repeatedly performed with a different number of neutral gallery images using PCA and rLDA. The significant improvement in face recognition accuracy can be clearly seen from table 3.5. With PCA, up to 13% of improvement over the original images is achieved. Similarly, up to 8.75% of improvement is achieved using rLDA. Furthermore, the error of Opened-Mouth detection doesn't have a significant effect on the recognition accuracy as 100% of accuracy is achieved using rLDA with only 5 neutral gallery images.

Table 3.5 Face recognition on corrected open mouth faces of BU-4DFE database in comparison with the original *opened mouth* face images

Test : gallery	Open mouth (Original)	Closed mouth (Recovered)	Improvement
	PCA		
49 : 1	64.08%	77.57%	13.49%
48 : 2	69.60%	83.81%	14.21%
47 : 3	71.91%	86.32%	14.41%
46 : 4	75.35%	87.50%	12.15%
45 : 5	78.64%	90.07%	11.43%
44 : 6	80.57%	91.11%	10.54%
	LDA		
48 : 2	89.81%	98.56%	8.75%
47 : 3	91.96%	99.32%	7.36%
46 : 4	95.41%	99.89%	4.48%
45 : 5	96.78%	100%	3.22%
44 : 6	96.61%	100%	3.39%

3.4. Efficacy of Opened-Mouth Correction

The proposed Opened-Mouth correction algorithm can be considered as a normalisation step or pre-processing. The effect of misalignment on face recognition is very well known. The negative effect of Opened-Mouth on face recognition lies in its effect on the alignment that is, the Opened-Mouth face cannot be aligned properly. Consequently, the correction from Opened-Mouth is equivalent to image alignment process. Another important point is that the method used in Opened-Mouth correction doesn't need any training steps; the process can be performed straightforward. Although the landmark detector needs training, this training is done only once and then, the resultant model can be used constantly.

3.5. Summary

In this chapter, the problem of mouth shape variations and its effect on face recognition are investigated. The mouth shape is classified into four different shapes which are: Opened mouth (O), Opened mouth with teeth (Ot), closed mouth (C) and closed mouth with teeth (Ct). Firstly, Moth shape detection method is proposed by utilising the recently proposed AR landmark detector. Secondly, the shapes of class O, Ot and Ct are transformed into class C (closed mouth). Finally, the corrected images are passed to face recognition process. The proposed method only needs a small number of landmarks and can be performed straightforward without the use of any training or geometric models. It is also simple, effective and computationally efficient. Through the experiments on the AR database and BU-4DFE database, significant improvement in face recognition accuracy is clearly achieved. Furthermore, experiments show that the small variations like smiling (Ct) don't affect the recognition accuracy.

Chapter Four

Two Directional Multiple Color Fusion for Face Recognition

The use of colour information has made a significant improvement for colour face recognition due to its effective representation over greyscale representation (Torres, Reutter et al. 1999). The RGB color space has been widely used for face recognition. Unfortunately, RGB subspaces are highly correlated (Yang, Liu et al. 2010), and such correlation has a negative impact on face recognition accuracy. Other color spaces such as, YIQ, I1I2I3, XYZ, HSV, HUV, and YCbCr, are derived from the RGB color space for various purposes (Gonzalez 2009). The efficacy of some color spaces for face recognition is shown in (Yoo, Park et al. 2007). On the other hand, some color spaces are designed specifically for face recognition in order to enhance recognition accuracy. These color spaces include Discriminant Color Space (DCS), Independent Color Space (ICS), Uncorrelated Color Space (UCS) (Liu 2008), Holistic Orthogonal Analysis (HOA) (Jing, Liu et al. 2010), normalized RGB **(**nRGB) , normalized XYZ (nXYZ) (Yang, Liu et al. 2010), General Color Discriminant model (GCID) (Yang and Liu 2008). Despite different approaches are developed these color spaces, all of them use exactly three color components and they are derived from the RGB color space in a holistic manner. Some models are patch based like Block Wise Discriminant Color Space (BWDCS) and Pixel Level Discriminant Color Space (PLDCS) (Li, Liu et al. 2012) attempting to assign the proper color representation to each block (or patch) rather than on the holistic face image. These methods are still restricted by using only three color components, which is expected not enough for face recognition. The Multiple Color Fusion (MCF) (Li, An et al. 2011) has made significant change in color face recognition. It is the only attempt to utilize more than three colors to represent face image. The motivation behind the MCF is that the discriminant information may lie in a space spanned by more than three color components. This preferred representation can be achieved via training. Instead of using only three channel based color spaces, the MCF trains multiple color channels for face recognition holistically. LDA is used for feature extraction with the Nearest neighbour classifier for recognition in which the greedy algorithm (Cormen and ebrary 2009) is used to select the proper channels representing the

MCF model. The MCF model shows much better performance compared with all other color spaces and models. Though the MCF model outperforms all other color spaces, it also holistically extracts its components. Research in face recognition has shown the superiority of patch based methods over the holistic methods like in (Li, Liu et al. 2012). In this chapter, a new method named as the Two Directional Multiple Color Fusion (2D-MCF) is proposed (Mustafa M. Alrjebi 2015). Borrowing the idea of MCF, the 2D-MCF model uses more than three color components in a block-wise manner rather than in a holistic way. Instead of training the whole face image for face recognition to attain the model, the 2D-MCF divides face images into a number of blocks, and then each block is trained for face recognition using the MCF criteria separately. Because the number of channels may differ from one block to another, the MCF model is further trained based on the holistic face derived from the representation in the first training step and this training can achieve the best representation of the whole face image. The motivation behind the 2D-MCF model is that different facial areas have different shapes and colors and the best representation of each face area may be different from other facial areas. Research on lip segmentation shows that lip requires a specific color representation for correct segmentation (Gritzman, Rubin et al. 2015).

The rest of this chapter is organized as follows: in the next section, the idea and algorithm of the 2D-MCF model are described. Experiments and discussion are presented in section 4.2. Finally, the chapter is summarised in section 4.3.

4.1. The 2D-MCF Model

The MCF model uses more than three colors for face recognition and it shows a high improvement over other color spaces and models. As has been proven in many of face recognition research that the block-wise or patch based representation and feature selection can further improve the recognition accuracy (Wong, Chen et al. 2011, Li, Liu et al. 2012), it is expected that employing the MCF model in block-wise manner can achieve better recognition rate than the previous holistic MCF. As we will see hereafter in this chapter the use of the MCF model in a block-wise way can achieve much better performance than the holistic MCF.

4.1.1. The 2D-MCF Model

Given training and testing colour dataset that is large enough to achieve a good representation. Firstly, all images are transformed into different color spaces. Nine color spaces are utilised containing 27 color channels. Secondly, the identical and highly correlated color channels are removed (5 channels are removed where 22 channels are used for training). Then, each channel in each image is divided into a number of non-overlapped blocks. Thirdly, the MCF procedure is performed on each block separately leading to recognition performance in each block as well as color representation in each block. In this case, each block may have a different number of selected color channels based on the MCF criteria (Li, An et al. 2011). To solve this issue, the selection of more color channels is continued (based on the performance) in each block until reaching the maximum number of the selected color channels through all blocks. For example, if the maximum number of selected color channels among all blocks is M_1, then, the number of selected channels is extended to this number (M_1) in all blocks by adding more channels as we will describe later in this chapter. At this step, a cubic color map can be achieved. Next, a second round of training is required in order to achieve the best representation for the whole face image. The representation achieved previously with M_1 channels is passed to the MCF procedure again in order to choose the best representation with M_2 layers $M_2 \leq M_1$. Finally the achieved model which is a 3D color map with M_2 channels is used to represent all gallery and query images in the face recognition process. Figure 4.1 illustrates an image with RGB, MCF and 2D-MCF. It is worth mentioning here the feature selection and face recognition are applied on the holistic face image rather than separate blocks.

We will investigate how to choose the number of color components M_1 for each block as well as the number of layers for the whole image M_2 later in this chapter.

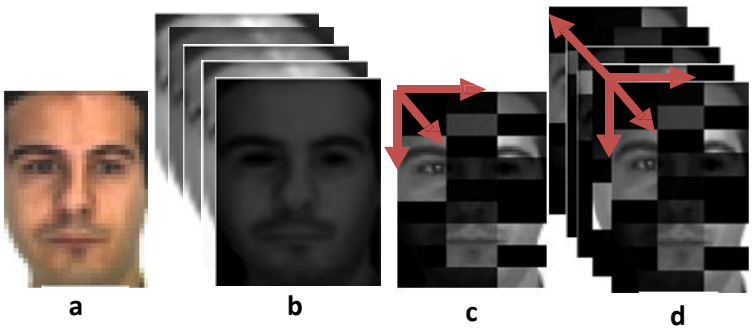


Figure 4.1 a. Original RGB image. b. MCF image in holistic. c. One dimensional –MCF image in blockwise. d. Two dimensional –MCF image

4.1.1.1. Number of Color Components

The 2D-MCF is a training based model, which implies that it needs training and testing dataset to build it. Such dataset should be represented in various color spaces. For this purpose, 9 different color spaces are utilised including RGB, XYZ, YIQ, I1I2I3, YUV, nRGB, nXYZ, YCbCr, and HSV. As some color components are identical such as the channel R in RGB and nRGB, only one of any identical color channels is kept. Furthermore, the same action is taken with the highly correlated color channels like R in RGB and Z in XYZ. The remaining number color channels after this filtering are 22 color channels (or components). These color channels will be used in the training phase to build the 2D-MCF model.

4.1.1.2. Selection Criteria, Feature Selection, and Classification Method in Training the 2D-MCF

The selection of first color channel for each block is based on the performance of each color component (channel) in the recognition accuracy with the given training and testing samples. The greedy algorithm (Cormen and ebrary 2009) is used in the selection process by adding one color channel in each round such that the color component that achieves the best recognition rate in combination with the previously selected channels is selected. This process can provide different blocks with different numbers of color channels/layers with the maximum number of channels among all blocks as M_1

To train different blocks for face recognition, the regularised LDA (Lu, Plataniotis et al. 2005) is used for feature selection and the Nearest Neighbour classifier with Euclidian distance measure is used for classification.

4.1.1.3. Size of Training and testing dataset

To construct a 3D color map, training and testing dataset is needed. Such dataset needs to be large enough to cover different cases. This can make the resultant model useful for general face recognition. Extensive experiments on the AR database show that 360 face images belonging to 45 subjects are large enough for this training purpose. For each subject, 4 images are used as gallery images and 4 images are used as query images. We use 400 images belong to 50 subjects to train the 2D-MCF model. The resultant trained 2D-MCF is effective for face recognition as illustrated in Figure 4.2.

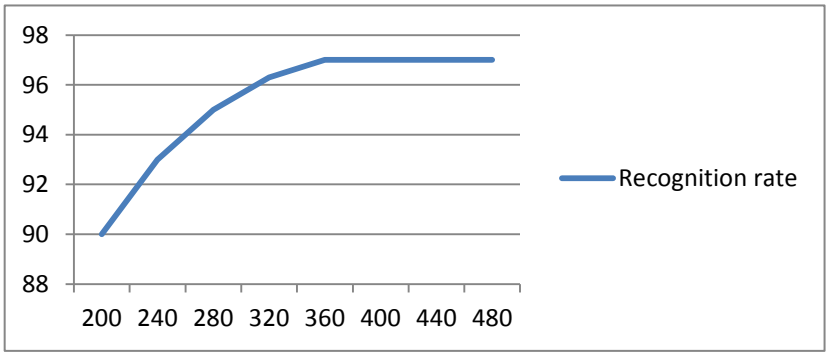


Figure 4.2 Recognition accuracy achieved by 2D-MCF model trained with different sizes of training and testing dataset

4.1.1.4. Number of Blocks

As to number of blocks in an image used in training the 2D-MCF, the use of large number of blocks (or patches) may lead to misalignment problem if the query face images are not well aligned, where the use of small number of blocks may lead to losing the local variances. Both of these cases can cause a failure in the training/testing quality of the 2D-MCF model. In this section, the 2D-MCF model is trained with different number of blocks in order to seek the suitable number of blocks. Experiments conducted on the AR database show that the optimal

number of blocks is (3× 9) or (3 × 10). As shown in table 4.1, the best performance is achieved by using this number of blocks with 400 images belong to 50 subjects (4 gallery images and 4 query images per subject). In the rest of this chapter, the number of blocks we choose is 3×9 blocks.

Table 4.1 Recognition accuracy with different size of blocks trained on the AR database

Vertical size / Horizontal size	6	7	8	9	10	11
	Recognition Rate					
3	94%	94%	95.2%	**97**%	**97**%	95%
4	93%	91.15%	92.50%	94.67%	95.5%	94%
5	92.33%	93.12%	93.5%	94%	94.5%	92%
6	90%	89%	89%	91.13%	92%	91%

4.1.1.5. Normalisation

Because of the inconsistency among different color components in each block, each block in each layer of the 2D-MCF model is normalised to zero mean and standard deviation for better performance. This normalisation is commonly used in the literature of face recognition in order to avoid the domination of one color channel over other channels (Theodoridis 2008).

4.1.1.6. The Proposed Algorithm for 2D-MCF

The following steps describe how to build the 2D-MCF model via using a training and testing dataset.

Algorithm 4.1

1. *Transform the original RGB face images into 9 color spaces which contain 27 color components.*
2. *Remove the identical color components such as R in RGB and nRGB. At this point only 22 color components are retained.*
3. *Divide each color component into 3x9 non-overlapped blocks.*
4. *Each block in each color channel is normalised to zero mean and standard deviation.*

(Normalisation).

5. *Perform MCF algorithm* (Li, An et al. 2011) *with rLDA as feature extraction on each block to obtain the important color components for this block.*
6. *For each block, select the color component that achieves the best recognition rate as a first color component of this block.*
7. *The second color component of each block is the color component that achieves the best accuracy in combination with the first chosen color component.*
8. *Add one more color component each time by using the same criterion in step 7 until the recognition rate starts to decrease. At this point, the optimal representation for this block is obtained.*
9. *The optimal number of color components that achieves the highest recognition rate may vary from one block to another block. In this case, we select the maximum number (M_1) of optimal color components among all blocks.*
10. *With this number M_1, we can extend each block to M_1 color components with the following criteria. We add one more layer iteratively until the number of block layers reaches M_1 and in each iteration, the color component that achieves the best recognition accuracy in combination with the previously selected color components is selected.*
11. *Now we can construct a color map template with M_1 layer holistic face image with each layer including different color components for different blocks as illustrated in Figure 4.1c.*
12. *With the given training and testing dataset, we can now first convert each image into the template and train these converted holistic images to obtain the optimal number (M_2) of layers based on the performance using the rLDA and greedy search algorithm as did in MCF model and then can obtain the final 2D-MCF model as shown in Figure 4.1 (d).*

4.1.1.7. The 2D-MCF Model for Face Recognition

Once the 2D-MCF model is obtained as described previously, it can be used for face recognition as described below. First, all gallery images and query images are transformed into the 2D-MCF representation. The face recognition is applied on the new holistic face images rather than the separate blocks. These 2D-MCF representative images are reshaped into one concatenated vector before face recognition process. The following steps are used for the construction of 2D-MCF representation.

Algorithm 4.2

1. *All the training and testing face images are transformed into different 9 color spaces and the number of color components is reduced to 22 by removing the identical and highly related components, then all images are segmented into the same pre-determined number of non-overlapped blocks.*
2. *Each block in each color channel is normalised to zero mean and standard deviation. (Normalisation).*
3. *All images are reconstructed using the trained color map structure with the optimal number (M_2) of layers.*
4. *Holistically perform face recognition with a given feature extractor and classifier.*

Next, we will conduct extensive experiments on different face datasets to validate the effectiveness of the proposed 2D-MCF model.

4.2. Experimental Setup and Results

In this section, the efficacy of the proposed 2D-MCF model is evaluated by conducting several experiments. All experiments are conducted on frontal facial images from the AR database and Curtin database. The Regularized LDA is us used for feature extraction, where the Nearest Neighbor NN classifier with Euclidian distance measure is used for classification.

4.2.1. Databases and Setup

To evaluate and build the 2D-MCF models, two different databases are utilised including the AR face database and Curtin face database.

4.2.1.1. The AR Database

The AR database is a 2D color database containing 2600 frontal face images for 100 subjects (The AR database is mentioned in chapter 2). In this chapter, 10 images per subject are used covering neutral faces, expressions and occlusions. Some sample images of this dataset are shown in Figure 4.3.a.

4.2.1.2. The Curtin Database

Curtin database (Li, Liu et al. 2014) is a 3D color face database containing 5044 3D face images for 52 subjects (97 images per subject). This database is originally designed for 3D face recognition. All images are taken in a controlled environment using two cameras including a low-resolution 3D camera. Furthermore, this database covers various expressions, illumination conditions, pose variations and occlusions. In this work, the images with expressions and occlusions are utilised. As shown in Figure 4.3.b, 10 images per subject are used covering facial expressions and occlusions.

a

b

Figure 4.3 a. Sample images from the AR dataset.

b. Sample images from the Curtin database

4.2.1.3. Building the 2D-MCF Model

To evaluate the correctness, stability and efficacy of 2D-MCF model, two different 2D-MCF templates are built based on two different databases. The first template named as (Template 1) is built utilising the AR database. For this purpose, 400 images for 50 subjects are used (8 images per subject) covering neutral and different expressions. For each subject, 4 images are used as gallery images where the rest 4 images are used as query images. The second template named as (Template 2) is built based on the Curtin database utilising 416 images for 52 subjects covering neutral and different expressions. As in the first template, 4 images are used as gallery images and the rest 4 images are used as query images. For both Template 1 and Template 2, the rLDA is used for feature extraction and the Nearest Neighbour NN with Euclidian distance measure is used for classification.

4.2.1.4. Number of Color Components and Layers

As mentioned in section 4.1.1.1, 22 color components are used to train the 2D-MCF model. These entire color components come from color spaces derived from the RGB color space. On the other hand, D1, D2 and D3 color components from the DCS model are also used in the process of training the MCF model and this is mainly for fair comparison with the original MCF model because these components have been used in the original MCF.

After training the 2D-MCF model, the highest recognition rate is achieved by using 3 to 5 (M_2) layers. Consequently, 5 layers of the 2D-MCF model are used in the experiments. In contrast, 9 to 10 color components (layers) are selected to represent the MCF model and one can see that the number of layers is significantly reduced due to patch operation. Next, we will investigate the performance of 2D-MCF model.

4.2.2. Experiments

In this subsection, the 2D-MCF model is evaluated by conducting face recognition on the AR and Curtin databases using previously mentioned setup.

4.2.2.1. Experiment One

In this experiment, 500 images for 50 persons from the AR database are used to evaluate the 2D-MCF model. Template1 is used to construct the 2D-MCF representation for all gallery and testing images. The experiment is repeated 10 times and the average recognition rate is compared with the MCF and different color spaces and models. Furthermore, different numbers of gallery images are used as shown in table 4.2. One can see that the proposed 2D-MCF model outperforms all other color spaces and models. As can be seen in table 4.2 the 2D-MCF improves the recognition accuracy by 4.1% to 7.30% over the MCF and other color spaces. In the next experiments, the robustness and efficacy of the 2D-MCF model will be evaluated by applying cross database testing where the 2D-MCF template is trained on one database and the face recognition is conducted on another database.

Table 4.2 Face recognition for 2D-MCF in comparison with different color representations on 50 subjects of AR face database

No. of gallery images / Color space/model	2	3	4	5	6	7	No. of layers
	Recognition Rate %						
RGB	71.05	78.77	79.00	82.08	84.80	87.40	3
XYZ	70.40	77.77	78.10	81.24	84.30	86.93	3
YIQ	79.50	85.86	85.20	87.40	88.65	89.27	3
I1I2I3	75.95	83.57	83.47	86.12	87.40	88.73	3
YUV	71.10	79.06	80.53	83.12	85.15	87.47	3
nRGB	75.88	83.00	83.23	85.60	86.90	88.40	3
nXYZ	72.60	80.69	81.40	83.44	85.40	87.73	3
YCbCr	71.10	79.06	80.53	83.12	85.15	87.47	3
HSV	70.08	76.66	78.53	82.44	85.00	87.00	3
DCS	71.57	79.71	82.45	85.60	86.90	88.15	3
MCF	80.92	86.37	86.60	88.36	89.45	90.23	10
2D-MCF	**88.25**	**91.80**	**91.53**	**93.84**	**95.30**	**94.33**	5

4.2.2.2. Experiment Two

This experiment is conducted on the Curtin database for evaluation purpose. Template 1 is used to represent all training and gallery images. The dataset used in this experiment contains 416 images from the Curtin database belonging to 52 subjects (10 images per subjects). As shown in Figure 4.3.b, the images cover neutral faces, expressions and occlusions. Like experiment 1, the face recognition is repeated 10 times and the average recognition rate is used to compare with different color spaces and models. In each round of experiment, the query images are selected randomly. Results of this experiment are detailed in table 4.3 showing the superiority of 2D-MCF model over all other color spaces and models including the MCF. This experiment shows the stability and robustness of 2D-MCF representation.

Table 4.3 Face recognition for 2D-MCF in comparison with different color representations on 52 subjects of Curtin face database

No. of gallery images / Color space/model	2	3	4	5	6	7	No. of layers
	Recognition Rate %						
RGB	79.93	85.82	87.34	88.35	90.10	90.96	3
XYZ	79.62	85.38	86.41	87.92	89.04	89.74	3
YIQ	85.70	91.04	91.63	91.40	91.88	92.59	3
I1I2I3	86.03	90.82	91.35	91.23	91.60	92.53	3
YUV	84.62	89.73	90.83	91.15	91.97	92.95	3
nRGB	85.96	91.02	90.63	91.50	91.45	92.21	3
nXYZ	85.17	90.33	90.77	90.88	91.54	92.76	3
YCbCr	84.62	89.73	90.83	91.15	91.97	92.95	3
HSV	81.39	88.21	89.46	89.81	90.63	91.60	3
DCS	82.15	89.90	89.33	90.67	91.00	91.33	3
MCF	86.76	91.16	91.67	92.44	92.64	93.59	9
2D-MCF	**90.13**	**94.64**	**95.32**	**95.96**	**96.73**	**97.37**	5

4.2.2.3. Experiment Three

In this experiment, Template 2 is used to represent 1000 images from the AR database belonging to 100 persons. A different number of gallery images is used and the results are illustrated in table 4.4. The experiment is repeated 10 times with a random selection of query images each time. The results of this experiment are detailed in table 4.4 showing improvement of 3.94% to 7.58% over other color spaces and models. This experiment consolidates the reliability, efficacy and robustness of 2D-MCF model. As in experiment 2, the 2D-MCF model has stability in cross database testing.

Table 4.4 Face recognition for 2D-MCF (trained on Curtin database) in comparison with different color representations on 100 subject of AR face database.

No. of gallery images / Color space/model	2	3	4	5	6	7	No. of layers
	Recognition Rate						
RGB	72.63	78.21	80.78	83.44	86.72	88.83	3
XYZ	71.80	77.36	79.82	82.66	86.03	88.37	3
YIQ	78.10	84.63	85.53	87.70	90.12	91.57	3
I1I2I3	76.74	83.14	84.38	87.24	89.98	91.13	3
YUV	73.31	80.21	81.67	84.80	87.85	89.80	3
nRGB	76.81	82.90	83.68	86.90	89.45	90.69	3
nXYZ	75.48	80.97	81.68	84.64	88.40	90.00	3
YCbCr	73.38	80.03	81.87	85.22	87.25	89.93	3
HSV	69.90	77.11	79.23	83.36	86.23	87.77	3
DCS	74.36	80.64	81.85	84.64	88.23	89.90	3
MCF	80.48	86.90	87.63	90.60	92.90	93.63	10
2D-MCF	**88.06**	**92.59**	**92.57**	**95.68**	**97.23**	**97.57**	5

The experimental results show the superiority of the 2D-MCF model over all other color spaces and models. As expected, the local variances among different facial areas are enriched based on 2D-MCF representation and it can achieve better performance consistently in face recognition. This consistency can be seen from experiment two and experiment three where the 2D-MCF models are built with different datasets in different environments, where the training of the model is conducted on one database and the testing is applied on another database (cross database testing). This shows the independency of the 2D-MCF model.

4.3. Summary

In this chapter, a new image representation model named as the Two Directional Multiple Color Fusion (2D-MCF) is proposed for face recognition. This model utilises more than three color components in block-wise. Different color components are fused in each block based on the performance of each color component in face recognition. The principle of MCF model is used to select the color components in each block, and then, the image is reconstructed from different trained blocks. The final model is determined by training the reconstructed images again for face

recognition using the MCF criteria. The idea is to improve the discriminative information in different facial areas rather than in a holistic face image. From the MCF, it is experimentally proven that the discriminant information lie in a space spanned by more than three color channels. However, the optimal representation for face recognition is varied among different facial areas. Experiments on two different databases show the superiority of the 2D-MCF model over the holistic MCF model and other color spaces and models. Even though the 2D-MCF model needs training and testing dataset in the training process, this doesn't consume time in real face recognition systems as this training process can be done offline and then the obtained model is dataset independent and can be used for face recognition online.

Chapter Five

Face Recognition against Occlusions via Color Fusion using 2D-MCF Model and SRC

Many challenges can impede facial recognition process. Occlusion is one of these challenges as it can significantly degrade the recognition accuracy. The occluded area usually doesn't contain any relevant useful information. The problem of occlusion is highly complicated because the occluded area, occlusion color and occlusion shape are unpredictable. Many solutions have been proposed to tackle this problem (Wright, Yang et al. 2009, Luan, Fang et al. 2013, Qian, Luo et al. 2015, Zhao and Hu 2015) for grey images. The partitioned SRC (P-SRC) (Wright, Yang et al. 2009), shows high performance of face recognition against occlusion. In this method, the face image is partitioned into a number of non-overlapped blocks. Each block of the query image is classified using the SRC and the final classification decision is made using a statistical majority voting among all blocks. The core issue of this method is that it does not require any prior information about the existence of occlusions or the position of occluded area. Consequently, this method can be considered as one of the most effective methods of face recognition against occlusion. Another SRC based method for face recognition with occlusion is proposed by (Mery and Bowyer 2014). In this method, an adaptive sparse representation for random patches is used and different patches are assessed based on a score level. For final classification decision, statistical majority voting is used. The reconstruction based method in (Luan, Fang et al. 2013) requires occlusion detection first and then the occluded area is recovered utilising overdetermined equation based on the Principle Components Analysis (PCA). After recovering the occluded area, any classifier can be used for face recognition. As mentioned in (Luan, Fang et al. 2013), the reconstruction based algorithm is effective if the occluded area is up to 40% of the whole face area. In (Nan and Xu 2013), A multi-resolution based adaptive blocking and sparse representation classifier (MBSRC) is introduced to tackle

the problem of face recognition against occlusion. In this method, multi-sized blocks are used to represent all training and gallery images where each image is represented differently by 4×4 blocks, 2×2 blocks, and the holistic face image resulting in total of 21 blocks. Then, the SRC is used to classify each block and the majority voting is used for final classification as did in (Wright, Yang et al. 2009). In the robust nuclear norm with regularised regression (abbreviated as RNR) (Qian, Luo et al. 2015), the error detection and error correction are incorporated in one regularised model instead of using two models. A modular weighted SRC (Zhao and Hu 2015) is proposed for face recognition against occlusion utilising the fisher discriminant information. The method proposed in (Morelli Andrés, Padovani et al. 2014) is based on compressed sensing for occlusion detection and removal. After detecting the occlusion, only the un-occluded pixels are used for face recognition. Other methods for face recognition against occlusion use different occlusion detection techniques to remove the occluded areas before performing face recognition as those proposed in (Ou, You et al. 2013, Priya and Wahida Banu 2014). In the literature of deep learning techniques, many approaches are introduced tackling different difficulties in face recognition such as pose invariant, uncontrolled environment and illumination variations (Zhu, Luo et al. 2013, Sun, Wang et al. 2014, Sun, Wang et al. 2014, Zhuo 2016). However, the problem of occlusion has not been addressed systematically.

It can be seen from the literature of face recognition against occlusion that, most of the techniques need to detect the occlusion first as a step of pre-processing; then, the occluded area will be discarded or recovered before implementing face recognition. Although some occlusion detection based techniques show high performance in face recognition, they are in fact very sensitive to the detection rate, where the recognition accuracy is significantly affected by the detection error (Luan, Fang et al. 2013). In contrast, the P-SRC (Wright, Yang et al. 2009) is the only method that doesn't depend on occlusion detection. One common characteristic among all above mentioned methods for face recognition against occlusion is that, all of them only use grey image representation while focusing mainly on occlusion detection, feature extraction and/or classification method. However, the color information plays a significant role in improving the recognition accuracy (Chengjun 2008, Liu 2008, Chengjun and Jian 2009, Jing, Liu et al. 2010, Li, An et al. 2011, Li, Liu et al. 2014) and its role in face recognition against occlusion needs study. Most importantly, the patch based (or block-wise) color models in last chapter have shown significant improvement in face recognition accuracy (Li, Liu et al. 2012, Mustafa M. Alrjebi 2015), it is necessary to investigate the occlusion issue in face recognition using the color information effectively.

As the occlusion can appear locally covering a local area in a face image, it is expected that the use of block-wise color models (which is locally robust) can effectively improve the recognition accuracy with the presence of occlusion. In this chapter, the 2D-MCF model (Alrjebi., Liu. et al. 2015) is used with the P-SRC (Wright, Yang et al. 2009) to tackle the problem of face recognition with occlusion. By using the P-SRC, the occlusion detection problem can be avoided where the use of 2D-MCF model in image representation can increase the local recognition accuracy in each block. Consequently, it is expected that the recognition accuracy will be significantly improved as the P-SRC is based on local recognition in different face patches.

Based on above discussion, a new approach for face recognition against occlusion is proposed in this chapter. This new method, named as the *Face Recognition against occlusion via Color Fusion using 2D-MCF Model and SRC,* uses the 2D-MCF model which is introduced in chapter 4 for image representation due to its robustness in local representation for face recognition and the P-SRC classification technique. It is clear that, if the recognition accuracy in each block is increased in P-SRC, then, the face recognition accuracy should be increased as well because the P-SRC depends on statistic majority voting among all blocks in a face image. For such purpose, the 2D-MCF model will be slightly modified regarding to the number blocks and the use of greedy method as will illustrated later in this chapter. Moreover, the face recognition will be performed using the P-SRC based on each block, and then, the statistical majority voting is used for final classification decision.

The rest of this chapter is organised as follows: In section 5.1, the Partitioned SRC (P-SRC) is detailed. The proposed 2D-MCF and P-SRC is presented in section 5.2. Experiments and discussions will be illustrated in section 5.3. Finally, the chapter will be summarised in section 5.4.

5.1. The Partitioned SRC (P-SRC)

The Partitioned SRC (P-SRC) (Wright, Yang et al. 2009) has shown high performance in face recognition against occlusion, and also it doesn't require occlusion detection. In this chapter, the P-SRC is used with the 2D-MCF image representation in order to enhance face recognition against occlusion. As the idea of SRC is already introduced in chapter 2, only the P-SRC is described in this section. The idea of P-SRC is to divide the face image into a number of non-overlapped blocks (8 blocks in the original work). Then each block in the query image is

classified to one of the corresponding labels in gallery images using the SRC. At this point, we have different query blocks, each of which is classified to specific corresponding labels in the gallery images. The final step is to apply statistical majority voting among all blocks to classify the query image to a specific class of gallery images. The steps of P-SRC are outlined below.

1. Given gallery and query image, divide each image into 8 non-overlapped blocks.
2. Reshape each 2D block among all images into 1D vector.
3. Classify each block within the query image using the SRC classifier and corresponding blocks in the gallery images.
4. Classify the query image using statistical majority voting among all blocks.

It is worth mentioning that the P-SRC used the grey images in the original paper and the color information is not included.

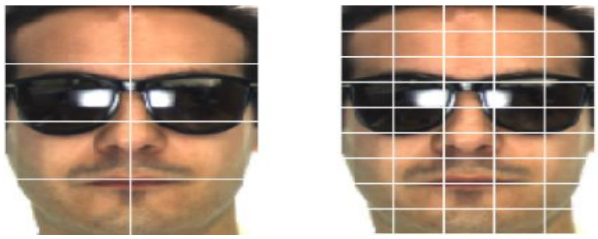
5.2. Face Recognition against Occlusion using the 2D-MCF and P-SRC

As the occlusion appears in a local face area, thus dividing the image into a number of blocks results in two types of blocks which are, occluded blocks and un-occluded blocks. Obviously, the un-occluded blocks will not be affected by the occlusion, where the effect of occluded blocks can be discarded using the majority voting for face classification. Based on this observation, it can be seen that the enhancement of local recognition rate in each block can increase the face recognition rate using the block-wise recognition methods and majority voting.

5.2.1. Number of divided Blocks

The number of blocks is selected based on the performance of P-SRC with occluded query images using the RGB color space. As mentioned above, the blocks can be occluded or un-occluded and only the un-occluded blocks are useful for face recognition. However, the occluded blocks may be only partially occluded, and splitting these blocks into smaller blocks may result in more un-occluded blocks that can be exploited in face recognition. It can be seen from Figure 5.1 that dividing the face image into 8 blocks results in 50% of un-occluded blocks and 50% of occluded blocks (Figure 5.1.a), where dividing the face image into 45 blocks

results in 66.66% of un-occluded blocks and 33.33% of occluded blocks (Figure 5.1.b). This increase of un-occluded ratio can improve the chance of correct recognition using majority voting. As shown in Figure 5.2, the best recognition rate is achieved using 45 to 54 blocks. Consequently, the face image is divided into 45 blocks in the experiments in this chapter. It is worth mentioning that this training is done with occluded test images and un-occluded gallery images to obtain the optimal number of blocks. However, for training the 2D-MCF model, un-occluded gallery and test images are used.



a b

Figure 5.1 Increasing the number of the un-occluded blocks over the occluded blocks by increasing the number of blocks. a: 4 occluded blocks Vs. 4 un-occluded blocks. b: 15 occluded blocks Vs. 30 un-occluded blocks.

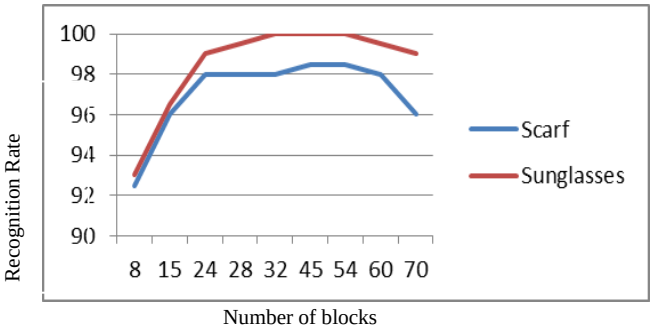


Figure 5.2 Recognition rates with different number of blocks on AR database

5.2.2. The 2D-MCF model

The 2D-MCF model is already detailed in Chapter 4. Here, the 2D-MCF is slightly modified to further increase its performance with occluded face images. To do so, firstly, the number of

blocks or layers is modified based on training on occluded faces instead of un- occluded faces in chapter four. Secondly, the greedy algorithm is modified to achieve the global maximum recognition rate instead of the local maxima in the second training stage when searching for the optimal M_2 layers. It is known that the greedy search is not optimal because it stops at the first local maxima (Cormen and ebrary 2009). Though, it is still the only available choice when the search scope is very large. Fortunately, the search scope in our case is relatively small which allows us to accommodate such modification.

5.2.3. Number of layers

In Chapter 4, the number of layers in 2D-MCF model is determined in the second round of training using the greedy criteria (Cormen and ebrary 2009). However, the greedy search will stop in the first local maxima as it is described previously. As the number of layers M_1 obtained in the first round of experiment is relatively small, and thus we can use the comprehensive greedy search, i.e., we search all layers from layer 1 to layer M_1 to achieve the best recognition performance. Unlike the classical greedy search, the comprehensive greedy search can achieve the global maxima by exploring all possible numbers of layers from 1 to M_1 and then, in this case, M_2 layers ($M_2 \leq M_1$) can be selected such that they will achieve the best recognition accuracy. Figure 5.3 show the difference of comprehensive greedy search in comparison to the classical greedy search on the AR dataset.

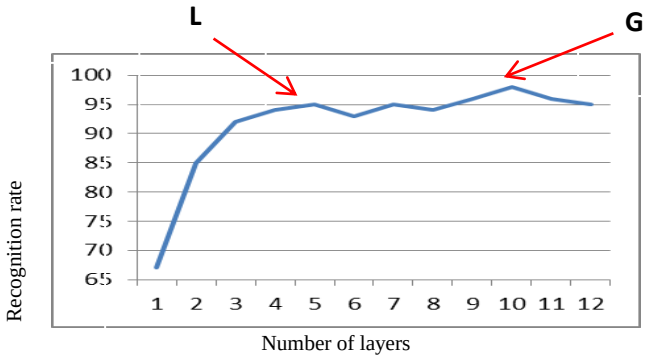


Figure 5.3 Local maxima L using greedy search vs. global maximum G using the comprehensive greedy search

5.2.4. Building up the 2D-MCF Model

Given a training and testing dataset, the 2D-MCF model can be obtained via a training process based on MCF criteria. As detailed in chapter 4, two rounds of training are required. In the first round, M_1 layers for each block are obtained using the rLDA for dimensionality reduction and the Nearest Neighbor classifier NN as classifier. In the second round of training, the optimal M_2 $(M_2 \leq M_1)$ layers representation is obtained by training the M_1 layers again for face recognition in a holistic representation and this best combination is determined by the modified greedy method as illustrated in the subsection 5.2.3 in order to avoid the problem of local maxima. Other factors including the number of color components used in the training and testing dataset and the size of the training and testing dataset are discussed as in chapter 4. The steps of building the 2D-MCF (Algorithm 4.1) model is slightly modified as follows in this case.

Algorithm 5.1

1. *Transform the original RGB face images into 9 color spaces which contain 27 color components.*
2. *Remove the identical color components such as R in RGB and nRGB. Thus, only 22 color components are retained.*
3. *Divide each color component into 45 non-overlapped blocks.*
4. *Each block in each color channel is normalised to zero mean and standard deviation. (Normalisation).*
5. *Perform MCF algorithm with rLDA as feature extraction approach on each block to obtain the important color components for this block.*
6. *For each block, select the color component that achieves the best recognition rate as a first color component of this block.*
7. *The second color component of each block is the color component that achieves the best accuracy in combination with the first chosen color component.*
8. *Add more color components one by one using the same criterion in step 7 until the recognition rate starts to decrease. At this point, the optimal representation for each block is obtained.*
9. *The optimal number of color components that achieves the highest recognition rate may*

vary from one block to another block. In this case, select the maximum number (M_1) of optimal color components among all blocks.

10. *With this number (M_1), extend each block to M_1 color components that achieving the best performance in each block in combination with the previously selected color components as described in Algorithm 4.1.*
11. *Construct a color map template of M_1 layer holistic face image with each layer including different color components in different blocks as illustrated in Figure 5.4.b.*
12. *With given training and testing dataset, convert each image into the template.*
13. *Train these converted holistic images to obtain the optimal number (M_2) of layers based on the performance using the modified greedy search as illustrated in Figure 5.4.c.*

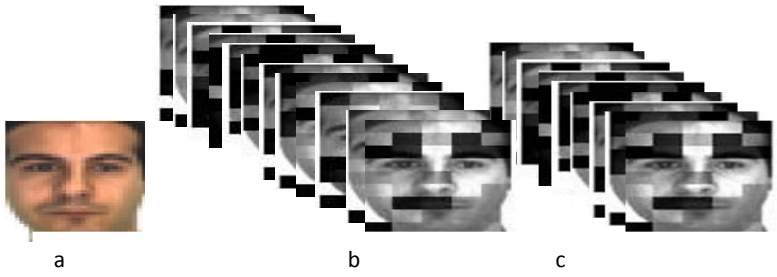


Figure 5.4 a: the original RGB image. b: 2D-MCF image with $\boldsymbol{M_1}$ layers. c: 2D-MCF image with $\boldsymbol{M_2}$ layers.

5.2.5. Face Recognition against Occlusions

For face recognition against occlusions, firstly, all gallery and query images are transformed into the 2D-MCF model with M_2 layers. Secondly, each block in the query image is classified using the SRC classifier. At this point, we have 45 blocks each of which is classified separately. Finally, the statistical majority voting is used to classify the image into a specific class. These steps are outlined in algorithm 5.2.

Algorithm 5.2

1. *Given a training and testing dataset, build up the 2D-MCF model as described Algorithm 5.1*
2. *All the training and testing face images are transformed into different 9 color spaces and the number of color components is reduced to 22 by removing the identical components, then all images are segmented into the same pre-determined number of non-overlapped blocks.*
3. *Each block in each color channel is normalised to zero mean and standard deviation. (Normalisation).*
4. *All images are reconstructed using the trained color map structure with the optimal number (M2) of layers.*
5. *Perform face recognition in each block (occluded or un-occluded) using the SRC classifier. The result of this step is a class vector containing the classification result (class labels) per block.*
6. *Classify the query image using the majority voting such that, image class = mode (class label), where the mode is the* number *or value that appears most often in the class vector.*

Next, we will use some benchmark datasets to do some experiments and validate the proposed approach.

5.3. Experimental Setup and Results

In this section, experiments on various databases are extensively conducted to evaluate the efficiency and correctness of the proposed method. For a fair comparison, firstly, all existing approaches are categorised to one of two categories which are: *occlusion detection based methods* and *direct methods*. The P-SRC (Wright, Yang et al. 2009) and the MBSRC (Nan and Xu 2013) are the only *direct methods* where all other approaches are *occlusion detection based.* The P-SRC and reconstruction based method (Luan, Fang et al. 2013) are originally designed for grey images. However, in these experiments, we will extend them to color images for fair comparison with the proposed approaches.

5.3.1. Databases and Experimental Setup

In these experiments, five different databases are used to evaluate the proposed method. The AR database (Martinez 1998) contains scarf and sunglasses occlusion. The Curtin database (Li, Liu et al. 2014) contains faces occluded by sunglasses and hand occlusion. The Bosphorus database (Savran, Alyüz et al. 2008) contains hand occlusion, glasses occlusion and hair occlusion. Even though the FRGC database (Phillips, Flynn et al. 2005) doesn't contain any occluded images, we added different types of occlusion to some faces in the database to cover different shapes, sizes and colors of occlusions. Also, occlusion is added to some occluded and unclouded images from the AR database to evaluate the proposed method in large size occlusions and different occlusion shapes.

5.3.1.1. The AR Database

As mentioned before in chapter 3 and chapter 4, the AR database contains 2660 images belonging to 100 subjects. In this chapter, the AR database is used for training the 2D-MCF mode using 8 un-occluded images per subject. Furthermore, it also utilised for face recognition against occlusion as described as follows:

- Six un-occluded images per subject are used as gallery images.

The query images can organised as follows:

- Two screaming un-occluded images.
- Two images occluded by sunglasses.
- Two images occluded by a scarf.
- Mask occlusion is added to the screaming images and then they are used again as a different case of occlusion.
- Mask occlusion is added to two neutral images.
- Scarf occlusion is added to the images with sunglasses and they are used as another different case of occlusion.
- Three different masks are added to neutral images covering the upper face area. These masks are different in the shapes and colors.

Figure 5.5 shows these entire gallery and query images. It is worth mentioning that all images used as gallery images are used again as query images when they are added masks as occlusions.

Figure 5.5 The AR database. a. Gallery images. b. Query images with different types of occlusions and screaming faces.

5.3.1.2. The Curtin Database

The Curtin database is a 3D database containing 5044 belonging to 52 subjects. The images in this database are low resolution captured by Kinect camera in the Curtin university lab (Li, Liu et al. 2014). This database covers various conditions including: expressions, lighting conditions, pose variations and occlusion. With regard to occlusions, the Curtin database contains two occluded images per subject, one image occluded by hand and another image occluded by sunglasses. In this chapter, eight un-occluded frontal images are used to train the 2D-MCF model (4 training images and 4 testing images). In the face recognition process using the 2D-MCF model and P-SRC, 6 un-occluded images are used as gallery images and the two occluded images are used as query images. Figure 5.6 shows a sample of the dataset used for face recognition. Only 2D color information is used in the experiments and the depth information is ignored.

a

b

Figure 5.6 Curtin database. a. Gallery images. b. Query images with occlusions

5.3.1.3. The FRGC Database

The Face Recognition Grand Challenge database (FRGC) (Phillips, Flynn et al. 2005) contains more than 50,000 face images captured in the controlled and uncontrolled environment. This database is one of the most challenging databases in the face recognition research community. In this dataset, 9324 images belonging to 222 subjects (42 images per subject) are used for face recognition. For each subject, 15 images are used as gallery images, where the rest 27 are used as query images. As this database doesn't include any occluded images, different types and shapes of occlusion are added to the query images as follows:

- Three images are occluded by sunglasses.
- Three images are occluded by another type of sunglasses.
- Mask occlusion covering the eyes area is added to three images making ~15% of occlusion.
- Mask occlusion on the upper face area is added to three images covering ~50% of the face area.
- Mask occlusion on the mouth area is added to 3 images covering ~ 35% of the face area.
- Scarf occlusion is added to 3 images covering ~ 45% of the face area.
- Three different mask occlusions are added to 9 images covering 50% to 60% of the face area.

Some samples of this occluded query images and the gallery images are shown in Figure 5.7

Figure 5.7. The FRGC database. a. Gallery images. b. Query images with occlusions

5.3.1.4. The Bosphorus Face Database

The Bosphorus database is originally 3D face database containing more than 4500 face images belonging to 105 subjects. Each subject in this database has 29 to 52 images. Regarding to occlusion, this database contains: hand occlusion, hair occlusion and glasses (sunglasses and eyeglasses). In our experiments, twelve images per subject are used as gallery images where four occluded images are used as query images. It is worth mentioning that, only 2D color information is used in the experiments and the depth information are discarded. Figure 5.8 shows some sample images used in the experiments for one subject.

Figure 5.8 The Bosphorus database. a. Gallery images. b. Query images with different types of occlusions

5.3.1.5. Two 2D-MCF Models

In order to construct the 2D-MCF model, a training and testing dataset is required. Two different datasets are used to build up two different 2D-MCF models. The first dataset contains 400 frontal images from the AR database belonging to 50 persons. For each person, 4 images are used as gallery images and 4 images are used for testing. The second dataset contains 400 images from Curtin database belonging to 50 subjects. For each subject, 4 images are used as gallery images and 4 images are used as query images. All images in both datasets are frontal un-occluded images. The 2D-MCF model resulted by training the first dataset is named as *AR-model*, where the second one which based on the Curtin database is named as *Curtin-model*.

5.3.2. Experimental Results and Analysis

In this subsection, face recognition against occlusion is conducted using the proposed 2D-MCF with P-SRC and the results are in comparison with some state of the art approaches. The experiments are categorised based on the datasets used in each experiment. Essentially, the P-SRC and some reconstruction based methods are used for fair comparison with the proposed method. As mentioned earlier in this chapter, both the P-SRC and the reconstruction based methods are designed for grey images. For a fair comparison, we extend these methods directly for RGB representations in order to compare fairly. Furthermore, the reconstruction based

method is supported by the P-SRC classifier and it will be referred as R-P-SRC hereafter in this chapter.

5.3.2.1. Experiment One

This experiment is conducted on a dataset from AR database as illustrated in subsection 5.3.1.1. The *Curtin-model* is used to transform all images to a 2D-MCF representation, and then, face recognition is performed using the P-SRC classifier with majority voting. This experiment is divided into three sub-experiments as follows.

1. The dataset from AR as mentioned in section 5.3.1.1 is used for face recognition with the 2D-MCF and P-SRC in comparison with the P-SRC based on RGB and the reconstruction based method (R-P-SRC). Six un-occluded images are used as gallery images, where the screaming face images and the occluded images are used as query images. The results of this experiment in Table 5.1 show the superiority of the proposed approach over the other two methods in all cases including the screaming faces. This experiment also shows the robustness of the proposed method with some extreme cases such as the occluded screaming faces where the accuracy of 72.5% is achieved. Another challenging case is the high percentage of occlusion when sunglasses and a scarf covering ~80% of the face area and 87% of accuracy is achieved as show in Table 5.1. On the other hand, both the P-SRC and R-P-SRC fail in the screaming occluded faces and they also perform significantly worse than the proposed method with the high percentage of occlusion as shown in Table 5.1 and Figure 5.9. For the occlusion ratio of 30% to 60%, the proposed method still achieves up to 100% of accuracy as shown in Figure 5.9. In summary, all results of this experiment are detailed in table 5.1.

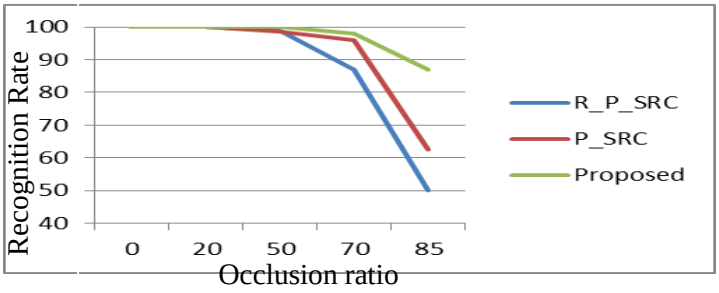


Figure 5.9 Face recognition on AR database with 6 gallery images per subject and different occlusion ratios

Table 5.1 Face recognition on AR database with screaming faces and different occlusion type, size and color

	R-P-SRC	P-SRC	2D-MCF + P-SRC	improvement
Screaming	NA	95%	97%	2%
Side mask	100%	99%	100%	0%
Side mask + Screaming	51.5%	25%	72.5%	21.5%
Scarf	99%	98.5%	100%	1%
Sunglasses	100%	100%	100%	0%
Scarf + Sunglasses	50%	62.5%	87%	24.5%
Upper mask1	87%	96%	98%	2%
Upper mask2	65%	82%	100%	18%
Upper mask3	90%	98%	100%	2%

2. In this experiment, the 2D-MCF with P-SRC is compared with some reported results using the compressed sensing (Morelli Andrés, Padovani et al. 2014) and the robust nuclear norm with regularised regression (RNR) (Qian, Luo et al. 2015). The RNR can achieve 99% and 100% for sunglasses occlusion and scarf occlusion respectively by using 8 gallery images while the compressed sensing methods can achieve 97.46% for sunglasses occlusion and 99.15 for the scarf occlusion with using 8 gallery images. Moreover, both the compressed sensing method and RNR require occlusion detection and this implies that such high recognition accuracy is based on the detection accuracy. In contrast, the 2D-MCF with P-SRC also achieves 100% in both cases with using only 6 gallery images without occlusion detection. As shown in Table 5.2. This experiment shows the superiority of our method over two of the most recent methods on AR database.

Table 5.2. Face recognition on AR database in comparison with some state of the art approaches

	RNR	Compressed Sensing	2D-MCF + P-SRC
sunglasses	99%	97.46%	100%
Scarf	100%	99.15%	100%
Detection based	Yes	Yes	No
#gallery images	8	8	6

5.3.2.2. Experiment Two

This experiment is conducted on Curtin database as mentioned in subsection 5.3.1.2. The *AR-model* is used for transforming the RGB images to the 2D-MCF model and then the P-SRC is used for classification. In face recognition process, 6 frontal images per subject are used as gallery images, and 2 images occluded by hand and sunglasses are used as query images. The results of this experiment show the superiority of the 2D-MCF + P-SRC over the P-SRC and the reconstruction based method as shown in table 5.3. The proposed method improves the recognition accuracy by at least 1.93% for hand occlusion and 3.84% for sunglasses occlusion.

Table 5.3 Face recognition on Curtin face database using different methods

	Hand occlusion	Sunglasses occlusion
R-P-SRC	98.07%	94.23%
P-SRC	94.23%	88.46%
2D-MCF+P-SRC	100%	98.07%
Improvement	1.93%	3.84%

5.3.2.3. Experiment Three

This experiment is conducted on the challenging FRGC face database. The images in this database are not highly clean and most of them have some expressions and lighting variations. 27 images per subject are used as query images after adding occlusion to them and 15 images are used as gallery images where the total number of images used in this experiment is 9324 belonging to 222 subjects. As mentioned in subsection 5.3.1.3, the occlusion is added to the query images where the gallery images are un-occluded. A sample of gallery images and query images with a different type of occlusion is shown in Figure 5.7. For face recognition, all gallery and query images are first transformed to the 2D-MCF model using the *Curtin-model*. Secondly, face recognition process is repeated 5 times using randomly selected 10 gallery images each time and we record the average performance accuracy. Detailed results with comparison with the P-SRC and R-P-SRC are shown in table 5.4. One can see that the proposed 2D-MCF + P-SRC outperforms the P-SRC and R-P-SRC in all cases. For sunglasses occlusion of ~20%, 5.4% improvement is achieved, and more importantly, 25% of improvement is achieved in the case of mask occlusion of ~50%. This experiment shows the robustness of the proposed method against occlusions even in challenging circumstances such as unclear images, lighting variations and expressions.

Table 5.4 Face recognition using different methods on FRGC database

	R + P-SRC	P-SRC	2D-MCF + P-SRC	improvement
Sungalsses1	77.31%	79.27%	86.18%	6.92%
Sunglasses2	77.53%	81.93%	86.97%	5.04%
Upper mask1	59.40%	68.02%	74.25%	6.23%
Upper mask2	58.86%	67.02%	80.88%	13.86%
Lower mask1	71.32%	72.12%	82.18%	10.06%
Lowermask2	64.16%	68.96%	75.03%	6.07%
Upper mask3	61.36%	47.45%	81.63%	20.27%
Upper mask4	55.70%	57.41%	82.48%	25.07%
Upper mask5	67.42%	55.35%	85.08%	17.66%

5.3.2.4. Experiment Four

This experiment is carried out using the Bosphorus database. 1680 frontal images belonging to 105 subjects (16 images per subject) are used. For each subject, 12 images are used as gallery images, where 4 occluded images are used as query images. In order to evaluate the robustness of the proposed, both the *AR-model* and *Curtin-model* are used to represent all gallery and query images. The face recognition process is performed using each model separately. Detailed results in table 5.5 show the robustness of the proposed method against different types of occlusions in comparison with the P-SRC and RP-SRC. Furthermore, the stability of the proposed method can be clearly seen from the results as the recognition accuracy is almost the same when the 2D-MCF models are trained by using different datasets.

Table 5.5. Face recognition on Bosphorus face database using different methods

		Hand Occlusion	Sunglasses Occlusion	Hair Occlusion
R-P-SRC		94.76%	97.14%	93.33%
P-SRC		95.24%	97.14%	92.38%
2D-MCF + P-SRC	*Curtin-model*	97.62%	98.10%	95.24%
	AR-model	98.10%	98.10%	95.24%
Improvement		**2.86%**	**0.96%**	**1.91%**

5.3.2.5. Experiment Five

In this experiment, the use of deep learning technique is investigated in comparison with the proposed method. In the literature of face recognition using deep learning, no work have been reported specifically for face recognition against occlusions. In this section, we adopt a deep neural network named as the Stacked Progressive Auto Encoder (SPAE) (Kan, Shan et al. 2014) to be used for face recognition against occlusions as it can be easily mapped to solve this occlusion problem. The SPAE is originally designed for pose correction from different pose angles to frontal faces. More details on the SPAE are presented in chapter 6. In this section, the SPAE is used to recover the occluded area gradually. As the SPAE requires large training samples, we use the Multi-Pie database (Gross, Matthews et al. 2008) for this purpose. In training the SPAE neural network, 600 images belonging to 200 subjects are used with occlusion of 50% added to the images covering the lower face area. Each layer in the network recovers 10% of the occluded area. So, 5 layers are used to recover the whole 50% of occlusions. Finally, the experiment is performed on the scarf occlusion from the AR database. In this experiment, a single gallery image is used where the PCA and SRC are used for classification. It can be seen from table 5.6 that the SPAE doesn't achieve high recognition rate in comparison with the proposed 2D-MCF + P-SRC. The lack of effective use of color information and the use of holistic face information may be the main reasons of this low recognition rate.

Table 5.6 Face recognition against 50% of scarf occlusion from the AR database using the proposed method in comparison with the SPAE

	Training dataset	PCA	SRC
SPAE	Multi-Pie	77%	78%
2D-MCF + P-SRC	Curtin	96%	98%

5.3.2.6. Sensitivity Analysis to Alignment Variations

As the proposed method is in block-wise, one would expect that it is sensitive to some alignment conditions. Therefore, this experiment is conducted with the same dataset of experiment one and the same setup. The alignment conditions are changed for the query images by rotating the images by 1% and 2% respectively. The results in table 5.7 show that the proposed 2D-MCF with P-SRC is indeed sensitive to the alignment as expected. As shown in table 5.7, the recognition accuracy is degraded by 1% to 3.5% with only a rotation of 1^{O}.

Table 5.7 The effect of 1^0 and 2^0 of rotation on recognition accuracy on AR database

Rotation	0^0	1^0	2^0
Side mask	100%	98%	96.5%
Side mask+ screaming	72.5%	70%	66%
Scarf	100%	98.5%	95%
Sunglasses	100%	99%	96.5%
Scarf + Sunglasses	87%	83.5%	79.5%
Upper mask1	98%	97%	94%
Upper mask2	100%	98.5%	96%
Upper mask3.	100%	98.5%	96%

5.4. Summary

In this chapter, a new method is proposed for face recognition against occlusions. This method is based on the representation by effectively utilising the color information. Unlike most of the existing methods, the proposed method doesn't require any information about the existent of occlusions. The 2D-MCF model which utilises more than three color components in block-wise way is used for image representation and the P-SRC is used for classification with statistical majority voting. Each block in the query image is classified independently, resulting in different blocks classified into different classes. The final classification decision is taken by majority voting. The proposed method shows high recognition accuracy in comparison with some of the state of the art methods. Furthermore, it is stable and robust with noisy images and different types and sizes of occlusions. Furthermore, the proposed method can still achieve high recognition rate even with some extreme cases such as 50% of occlusion as well as extreme facial expressions. The 2D-MCF model is stable even when they are trained on a specific database and then used for classification with another database. Also, the time cost of training the 2D-MCF model can be ignored because the training is only needed once and the resultant model can be used for face recognition with different datasets and environments after training.

Chapter Six

Enhancing Face Recognition against Pose Variations using Multi-resolution Multiple Color Fusion

As seen in the previous chapters, one can significantly improve face recognition performance by using multiple color channels fusion over using only three color channels or single channel. The MCF model (Li, An et al. 2011) achieves high performance in face recognition in comparison with different three channels based color spaces. The 2D-MCF model (Alrjebi., Liu. et al. 2015) further improves the accuracy over the MCF model via focusing on the local face variations in block-wise, where the discriminative information is optimised locally. The 2D-MCF model can also effectively utilised for face recognition against occlusion (Alrjebi, Pathirage et al. 2017) and outperform other state of the art methods as illustrated in chapter 5. In fact, both the MCF and 2D-MCF deal with different color channels equally, i.e., assuming that they all have the same effect on facial recognition. However, research in the field of face recognition confirms that different features should have different contributions to face recognition (Sun, Pan et al. 2008, Chen, Tang et al. 2010, Yule and Chen 2011, Lai, Dai et al. 2015, Zhao and Hu 2015). To represent different contribution for different features, usually different weightings are assigned to different features in feature fusion for face recognition. In addition to that, the 2D-MCF model is designed specifically for frontal faces and it cannot be used with pose variations directly. In this chapter, we propose a new strategy to fuse different color channel contributions into one framework for face recognition against pose variations by using the 2D-MCF model.

First, the importance of image size in terms of image resolution for face recognition has been investigated in (Bilson 1987). In fact, the Multi-resolution spectral features are extracted for face recognition in (Sun, Lam et al. 2013), in which it only attempts to enhance the recognition accuracy using single gallery image. The idea there is to enlarge the training dataset by generating more multi-resolution spectral feature images while reducing the effect of the illumination and facial expressions (Ye, Janardan et al. 2005). Despite a lot of research in

colour face recognition, the importance of color channels and their contribution to face recognition with multi-resolution spectral features have not been addressed. The only initial investigation shows that different color channels can achieve different results in face recognition when using each channel separately. This investigation validates the assumption that weighting for each color channel should be adopted according to its contribution in face recognition. We will use different resolutions of color channels as possible contributions in our approach, which will be described in detail later.

With regard to face recognition against pose variations, many researches have been conducted to tackle this problem in 2D and 3D spaces (Gross, Matthews et al. 2002, Ashraf, Lucey et al. 2008, Asthana, Marks et al. 2011, Castillo and Jacobs 2011, Sharma and Jacobs 2011, Biswas, Aggarwal et al. 2013). Most of these methods focus on pose correction using only grey image information while ignoring the role of the color information. The recently proposed Stacked Progressive Auto Encoder (SPAE) method (Kan, Shan et al. 2014) for face recognition against pose variations shows promising results compared with other different 2D and 3D based methods. This method uses the deep auto encoder to gradually correct face images to the frontal pose. The core idea of this method is that it doesn't need any information about the pose angle in the query image and it is corrected to the frontal pose automatically. In addition to that, the accuracy of this method is much higher than other methods even including some 3D based approaches. Although the SPAE model can automatically correct any pose angles to a frontal pose, it would fail for face images with large pose angles. Moreover, such neural network needs to be trained in the same environment as the cross database testing completely fails as we will show later in this chapter. In this chapter, we expect that the effective fusion of color information can enhance the efficiency and robustness of this method.

Based on the above analysis, a new multi-color Fusion scheme is proposed in this chapter named as the Multi-resolution Multiple Color Fusion (MMCF). This new image representation assigns different weightings for different color channels based on their performance with different resolutions. It is known that the amount and the level of information in the image are based mainly on the image size such that the larger size image contains more details of information. Consequently, the use of different image sizes (resolutions) can represent different weightings. In the MMCF model, various color channels are used each with four different sizes (resolutions or weightings), then the MCF criteria are used to optimise these channels with different resolutions for face recognition (Li, An et al. 2011) . As in the MCF, each color channel can be used only once in the MMCF model, i.e. if any color channel is selected for the

MMCF model, then the other three resolutions of this channel will be ignored. Unlike the 2D-MCF model, the MMCF model is obtained in holistic face image rather than in block-wise way and this holistic implementation makes it suitable for face recognition with pose variations. Details of building MMCF model will be presented later.

Once the MMCF is built, we will use this model to represent gallery and query images for face recognition. As we will show later in this chapter, the proposed MMCF model can achieve very high face recognition rate outperforming the MCF and 2D-MCF. Another advantage of the MMCF model is that it achieves high recognition rate even with small pose variations.

As to the face recognition against pose variations using the SPAE, in this chapter, we further embed the MMCF model in the SPAE for pose correction and face recognition. The idea is to compensate the discriminant information that may be lost in the correction process in SPAE. For this purpose, all images in the training dataset are transformed into different MMCF color channels, and then, the training/correction of the SPAE is applied on each color channel separately. Finally, the corrected images are transformed into the MMCF model before passing them to the classifier. To show the role of color information in this matter, the RGB color space is also used with the SPAE for pose correction and face recognition. We can see that RGB color space can achieve higher recognition rates than grey images. Furthermore, the MMCF model can achieve the best promising results in comparison with the grey representation and the RGB color space. In addition to that, the MMCF model is stable and robust with large pose angles, cross testing database as well as single gallery image case.

In summary, the proposed MMCF model is combined with SPAE deep learning neural network nicely in dealing with face recognition against pose variations. The rest of this chapter is organised as follows, the SPAE is reviewed and presented in the next section. In section 6.2, the MMCF model is explained. The integration of the SPAE with color information (RGB and MMCF model) for pose correction and face recognition is presented in section 6.3. The MMCF model is evaluated with extensive experiments conducted on six different databases.

6.1. The Stacked Progressive Auto Encoder (SPAE)

The shallow Auto Encoder is mainly used to reconstruct the ideal input signals (Bengio 2009). This reconstruction process is achieved through two phases, encoding and decoding. In the encoding phase, an encoder (f) is used to map the input layer to a hidden layer. In contrast, the

hidden layer is mapped to the output layer using the decoder (g). One hidden layer is usually used for this purpose. Inappropriately, such shallow Auto Encoder may fail in face pose correction because of the large pose variations may need more layers to correct. The Stacked Progressive Auto Encoder SPAE (Kan, Shan et al. 2014) solves this problem by stacking a number of shallow Auto Encoders to correct face images with pose variations to a frontal pose gradually. Each shallow progressive Auto Encoder can map images with larger pose angle to those with a smaller pose angle while keeping the images already with small pose angles unchanged. This leads to auto-correct any large pose angle to smaller pose angle without any information about the original pose angle, which is usually to detect via a specially designed pose detector. By stacking the Auto Encoders, the large pose angles can be corrected to a frontal pose. As an example of pose correction, given a face image with pose angle of 45^0, this pose angle can be corrected to the 0^0 (front face) using the SPAE with three hidden layers, the first layer corrects the pose angle from 45^0 to 30^0, then, the second layer corrects the image from 30^0 to 15^0 and finally, the third layer corrects the image from 15^0 to 0^0. Given pose variation range of [-V, V], the number of pose angles (pins) used in the SPAE can be generalised as $2 \times L + 1$, where $(V - 1) > (L \times 15) \leq V$. For example, given pose angle range of $[-45^0, 45^0]$, L=3 and the number of pins is 7 representing the number of pose angles used in the network with L hidden layers. Figure 6.1 shows the architecture of the SPAE with 3 layers, which can correct the angle scope of $[-45^0, 45^0]$ to a frontal face image. It is clear that larger pose variations need more hidden layers.

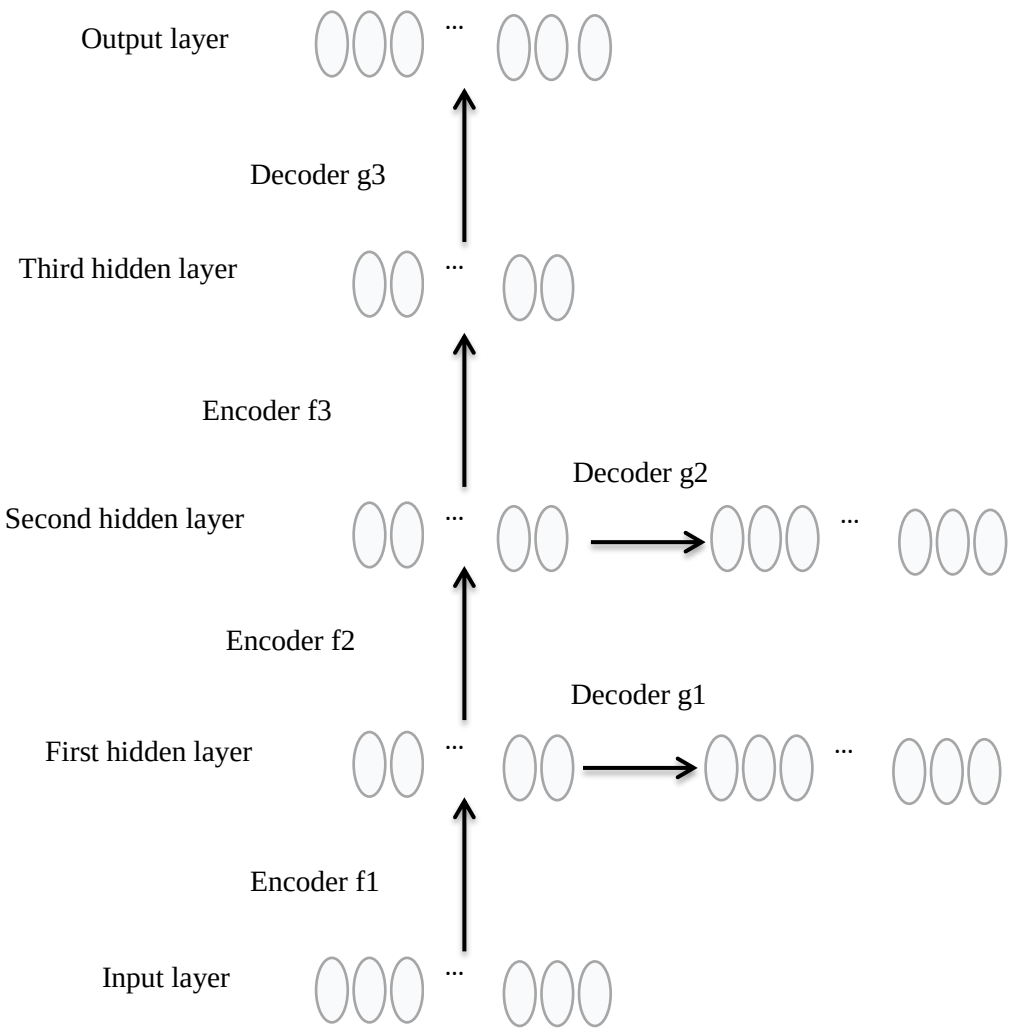


Figure 6.1 The Architecture of the SPAE with three hidden layers

Each hidden layer in the SPACE model attempts to map an image with larger pose angle to one with the next smaller pose angle as shown in Figure 6.2.

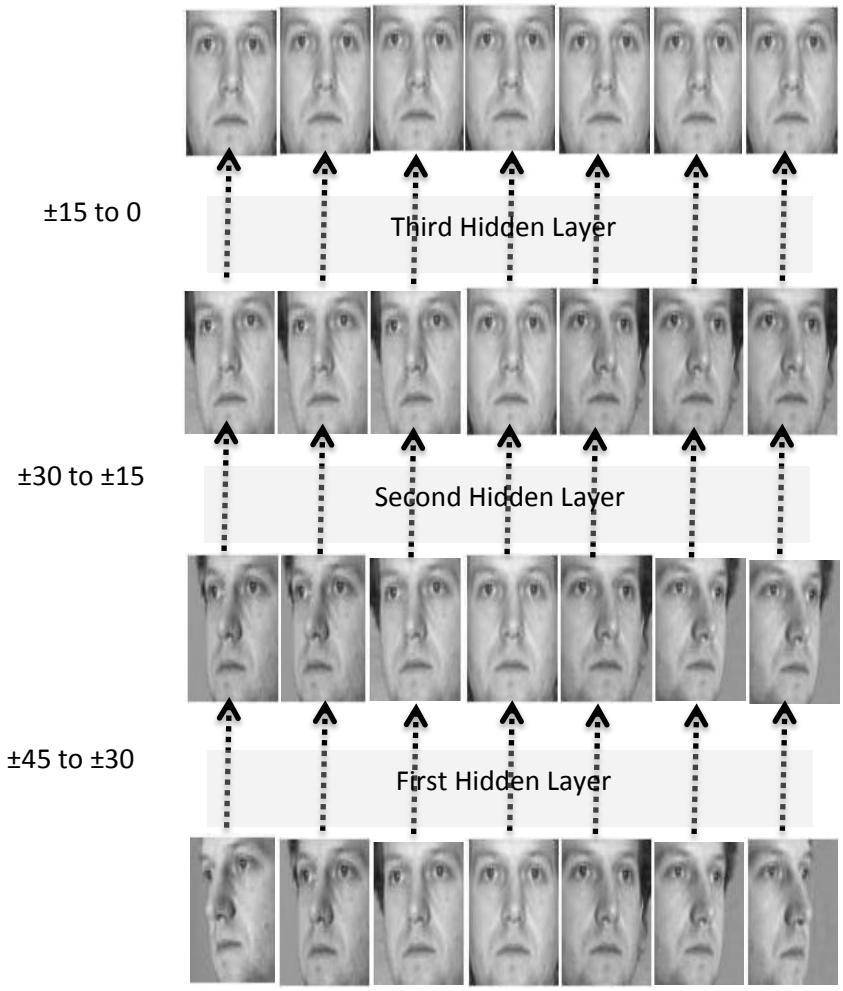


Figure 6.2 The correction from different pose angles to frontal face using the SPAE

Mathematically, the cost function for the SPAE can be formulated as follows:

$$\left[W_k^*, y_k^*, \widehat{W}_k^*, \hat{b}_k^*\right] = \underset{w_k, b_k}{arg} \underset{\widehat{w}_k \hat{b}_k}{min} \sum_{i=1}^{N} \sum_{j \in V} \left\| x_{il} - g_k\left(f_k\left(z_{ij}^{k-1}\right)\right)\right\|_2^2 \tag{6.1}$$

where z_{ij}^{k-1} is the representation of $(k-1)^{th}$ progressive Auto Encoder from the hidden layer for the $x_{i,j}$ sample, $w \in R^{d \times r}$ and $\widehat{w} \in R^{d \times r}$ are linear weightings, $b \in R^{d \times 1}$, $\hat{b} \in R^{d \times 1}$ are bias, r is the number of neurons and l is target pose, which can be calculated as follows:

$$\mathrm{l}_{ij}^{k} = \begin{cases} -\mathbb{P}(k) & if\ l_{ij}^{k-1} < -\mathbb{P}(k) \\ +\mathbb{P}(k) & if\ l_{ij}^{k-1} > \mathbb{P}(k) \\ l_{ij}^{k-1} & if \left| l_{ij}^{k-1} \right| \leq \mathbb{P}(k) \end{cases} \tag{6.2}$$

where $\mathbb{P}$ is the array covering the target poses for each progressive Auto Encoder.

It can be seen from the Eq. 6.1 and Eq. 6.2 that the pose angles can be reduced to $[-\mathbb{P}(\mathrm{k}), \mathbb{P}(\mathrm{k})]$ within the k^{th} order (layer) Auto Encoder. Thus, the pose angle is gradually reduced by stacking a small number of Progressive Auto Encoders until a virtual frontal face is achieved.

Given a dataset with pose variations in the range [-V, V], L progressive Auto Encoders are needed to correct the face images into frontal face. Finally, all layers in the network are optimised to finely tune the whole network according to Eq 6.3. Although Figure 6.1 shows three decoders, The network needs L encoders and only one decoder to correct any pose angle to frontal face. In figure 6.1, the network needs only decoder g_3, where both decoder g_1 and decoder g_2 can be removed.

$$\begin{aligned} &\left[W_k^* \Big|_{k=1}^{L}, b_k^* \Big|_{k=1}^{L}, \widehat{W}_L^*, \hat{b}_L^* \right] \\ &= \underset{W_k \Big|_{k=1}^{L}, b_k \Big|_{k=1}^{L}, \widehat{W}_L, , \hat{b}_L}{\arg\min} \sum_{i=1}^{N} \sum_{j \in \mathbb{V}} \left\| x_{i,0^0} \right. \\ &\left. - g_L(f_L\left(f_{L-1}(\cdots f_1(x_{i,j})\cdots)\right)) \right\|_2^2 \end{aligned} \tag{6.3}$$

The gradient descent algorithm which is widely used for optimisation can be easily used to solve Eq 6.3 using $f_1, f_2, \cdots f_L, g_L$ (Baldi 1995, Costa, Palmisano et al. 2000, Afshar, Brown et al. 2009, Klein, Pluim et al. 2009).

6.2. The MMCF Model

As mentioned previously, the use of multiple colors for face recognition can significantly improve the recognition accuracy. The MCF model shows a significant improvement over all other three channels based color spaces. Furthermore, the 2D-MCF model can further improve the recognition rate over the MCF by utilising the local discriminant information in a block wise way. However, both the MCF and 2D-MCF lack the effective use of color information as

they treat different color channels equally. In addition to that, the 2D-MCF model can be used only with frontal faces. In this section, the Multi resolution Multiple Color Fusion (MMCF) model is proposed. The use of multiple color channels with multiple resolutions/sizes will give each color channel a specific weighting according to its contribution in face recognition. Firstly, the MMCF model is obtained by training a number of color channels each with a number of resolutions for face recognition using the MCF criteria. Secondly, the obtained model is embedded in the SPAE network for pose correction. One can see that the MMCF model can avoid the drawbacks of the MCF and 2D-MCF models as it gives each color channel its actual contribution for face recognition with different resolutions. Next, the steps of obtaining the MMCF model are presented below. To clarify some terminologies in this chapter, the terns *resolution* and *image size* are used throughout this chapter referring to the number of pixels in a single image.

6.2.1. Building the MMCF Model

Given one RGB training and testing dataset, all images in the dataset are transformed into various color spaces which are including: RGB, XYZ, YIQ, YUV, I1I2I3, nRGB, nXYZ, HSV and YCbCr. Here, only color spaces obtained directly from the RGB using transformation matrix are used and we try to avoid any training based color spaces such as DCS on purpose. These nine color spaces with 27 color channels contain some duplicated color channels such as R in RGB and nRGB, as well as some highly correlated color channels like B in RGB and Z in XYZ. Consequently, only un-identical/un-correlated channels are retained as discussed previously. In this step, the remaining number of color channels is 22 color channels. Next, each image in each color channel is resized into four different resolutions which are, $32 \times 32, 24 \times 24, 16 \times 16$ and 12×12 representing four different weightings. Finally, the MMCF model is built as follows.

After preprocessing the data as described above, the training and testing dataset is used to build the MMCF model as outlined below.

1. First, we use the training/testing dataset with four different resolutions for face recognition. Each color channel with each resolution is validated for face recognition separately, and the color channel with a specific resolution that achieves the highest recognition rate is selected as the first channel of the MMCF model with this specific resolution. For example, if the G channel (from RGB) achieves the highest recognition

rate with the size of 16×16, then this channel with this size (resolution) is selected as the first channel (layer) of the MMCF model.

2. The selected channel for all other resolutions is removed in future validation in order to avoid redundant color information.
3. To select the next channel of MMCF model, we use all remaining channels with different resolutions with all possible combinations with the previously selected channel to validate the face recognition performance. In this case, we can select a unique color channel with a specific resolution to achieve the best recognition rate among all the possible combinations. Such combination of different color channels with different resolution is done by concatenating these channels into one vector. Then remove this selected color channel with other resolutions.
4. The steps 2 and 3 can be repeated by adding a new channel each time and removing them from the rest process until the face recognition rate starts to decrease.
5. At this point the training process is stopped and the MMCF model is obtained with all possible combinations of different color channels with different resolutions.

In this training phase, a large training and testing dataset from the AR database and Curtin database is used and the SRC is used for face recognition with 2 gallery images and 4 query images per subject. The training and testing dataset contains 912 frontal images belonging to 152 subjects and each subject has 6 images. Figure 6.3 shows the block diagram of the training phase where table 6.1 shows the selected color channel for the MMCF model with different resolutions.

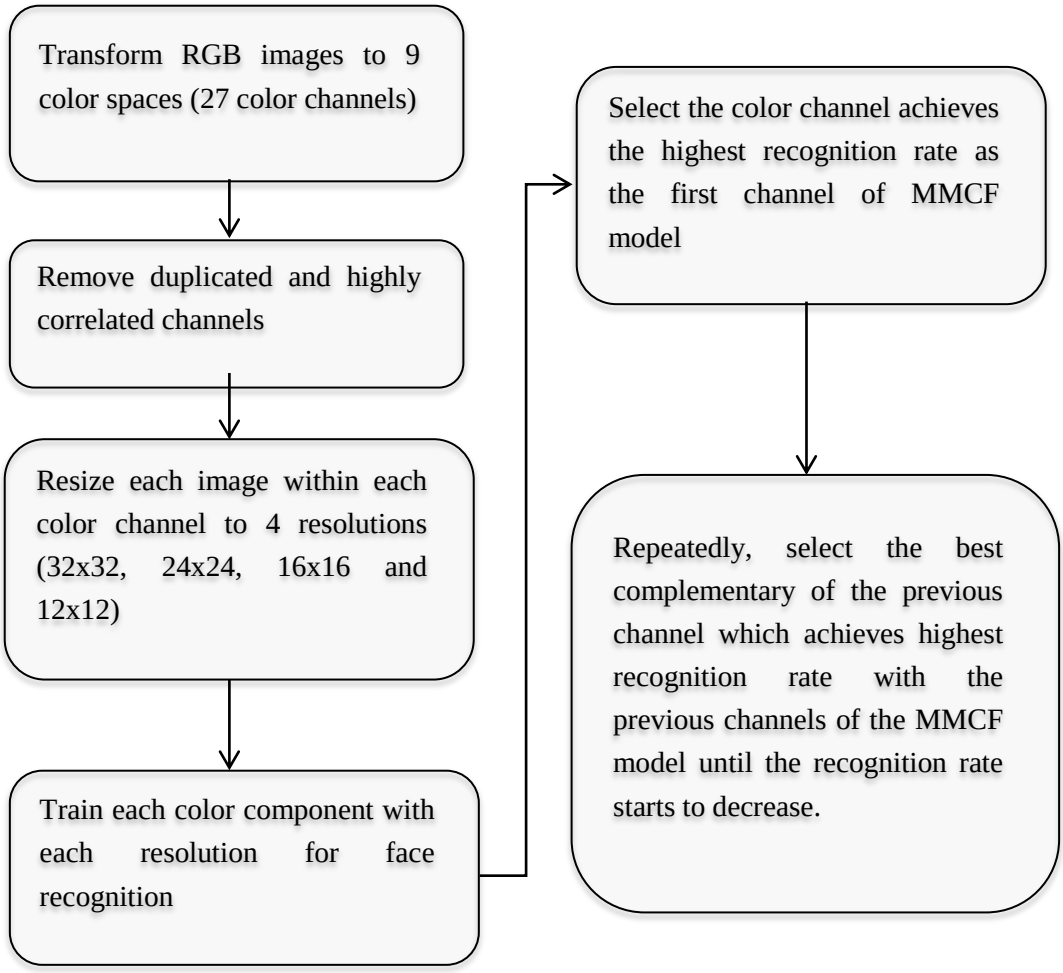


Figure 6.3 Building the MMCF model

Table 6.1 MMCF representation trained on AR + Curtin databases with the SRC classifier.

Color Channel	Color Space	Resolution
V	YUV	24×24
U	YUV	32×32
X	XYZ	12×12
I3	I1I2I3	32×32
B	RGB	12×12
I	YIQ	12×12
Q	YIQ	32×32
nG	nRGB	24×24
nB	nRGB	32×32

From above process, one can see that the contribution of each color channel to face recognition is different with its resolutions. Also such contribution will change according to its possible combinations with pre-selected color channels. The proposed MMCF model considers these differences and can achieve much better performance in comparison with previous approaches. This is the first time that color information with different resolutions is effectively used for face recognition. In next section, we will show how to use this model for face recognition.

6.2.2. Face Recognition using the MMCF Model

Given some gallery and query images in RGB representation, all images are transformed to the MMCF model representation according to the table 6.1, and then the face recognition is performed using a given classifier. Here, the regularized LDA (rLDA) is used for fair comparison with the previous MCF, 2D-MCF and SPAE. As the MMCF model has different

resolutions and they cannot be organised in a 3D matrix properly. So in this chapter, each image in the MMCF representation is built by concatenating the selected different color channels into one vector.

6.2.3. Face Recognition against Pose Variations using the MMCF and SPAE

In this section the MMCF model is integrated with SPAE to further enhance the recognition accuracy against pose variations. As it has been already shown that color information can play a major role in frontal face recognition (Torres, Reutter et al. 1999, Yip and Sinha 2002, Liu 2008, Li, An et al. 2011, Alrjebi., Liu. et al. 2015). As to pose variations in face recognition, we observe that poor recognition rate in this case may be caused by:

- The mismatch between the query image with pose angle and the frontal gallery images. This issue can be solved by pose correction as in the SPAE (Kan, Shan et al. 2014).
- The lack of satisfactory discriminant information especially the color information and we need to fuse color information effectively for this issue.
- The correction process may lead to loss of valuable discriminant information such that the corrected face images don't contain the same amount of discriminant information as those which are originally frontal.

In this case, it is expected that the effective use of color information may compensate this lack of discriminant information, which in turn can significantly improve the recognition accuracy. In this chapter, the SPAE is integrated with color information using the RGB and MMCF instead of only grey images as previously reported. The modification of SPAE according to RGB and MMCF representation is presented in the next two subsections.

6.2.3.1. Face Recognition Based on the SPAE and RGB Images

Research in face recognition shows that the use of color information can improve recognition accuracy over grey images (Torres, Reutter et al. 1999, Yip and Sinha 2002). In this section the SPAE is integrated with RGB images for pose correction and then face recognition. Firstly, the SPAE network is trained using R, G and B color components separately. The results of such training phase are three trained SPAE networks. Secondly, each trained network is used to correct the corresponding color channel in the query images. Finally the corrected RGB query images are used for classification using the rLDA for dimension reduction and the nearest

neighbour classifier. The integration of the SPAE with RGB images is illustrated in Figure 6.4. It is noteworthy mentioning that all images in the training phase and correction phase are resized to 33×33 for a fair comparison with the original SPAE.

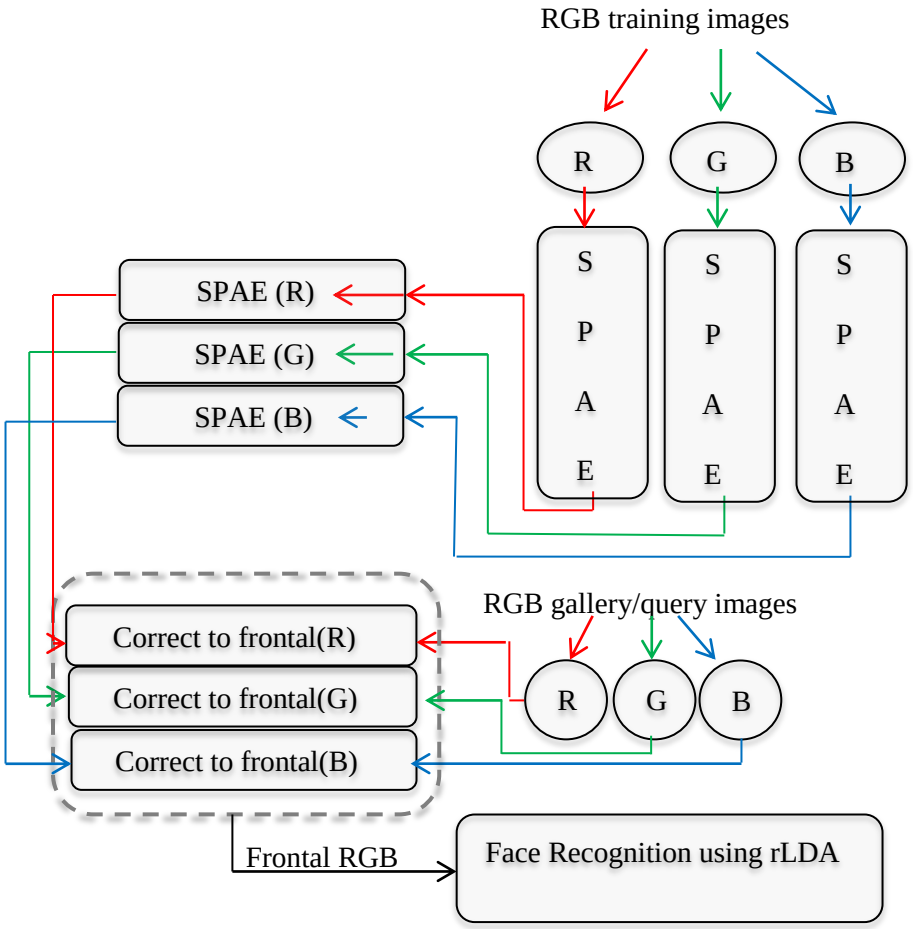


Figure 6.4 SPAE for pose correction with RGB images

One can see this integration process is separate for three color channels and no effective fusion for color information. Next, we will present the integration of the SPAE and MMCF.

6.2.3.2. Face Recognition Based on SPAE and MMCF Model

In this section, the SPAE is integrated with MMCF to enrich the corrected images with valuable discriminant information from the MMCF. First, the SPAE is modified for pose correction with different color channels as follows.

For SPAE training.

- Given RGB training dataset, all images in the dataset are resized to 33×33 and then they transformed to the selected nine color channels from table 6.1.
- The SPAE network is trained for pose correction using each color channel separately with image size of 33×33. The results of such training are nine trained SPAE models corresponding to the nine color channels from table 6.1.

For pose correction and face recognition.

- Given the gallery and query images, all images are transformed to the nine color channels selected in table 6.1. To avoid the domination of one color component over other components, all pixel values in different color channels are rescaled to [0 255] with image size of 33×33 as this image size is used in the original SPAE work. So, this can provide fair comparison with the original SPAE algorithm.
- Each color channel within each image is corrected to frontal face using corresponding trained SPAE network.
- Each corrected image in the dataset is transformed to the MMCF model according to table 6.1.
- Face recognition is performed using rLDA and the Nearest Neighbour NN classifier.

As we will show in this chapter that the MMCF outperforms the RGB color space regarding face recognition. Furthermore, the MMCF model can significantly improve recognition accuracy when it is embedded with the SPAE. It is worth mentioning that we use the same image resolution in the original SPAE as well as the same training dataset and classifier for a fair comparison with the original SPAE. The whole procedure of training the SPAE using the MMCF color channels for pose correction and face recognition is illustrated graphically in Figure 6.5.

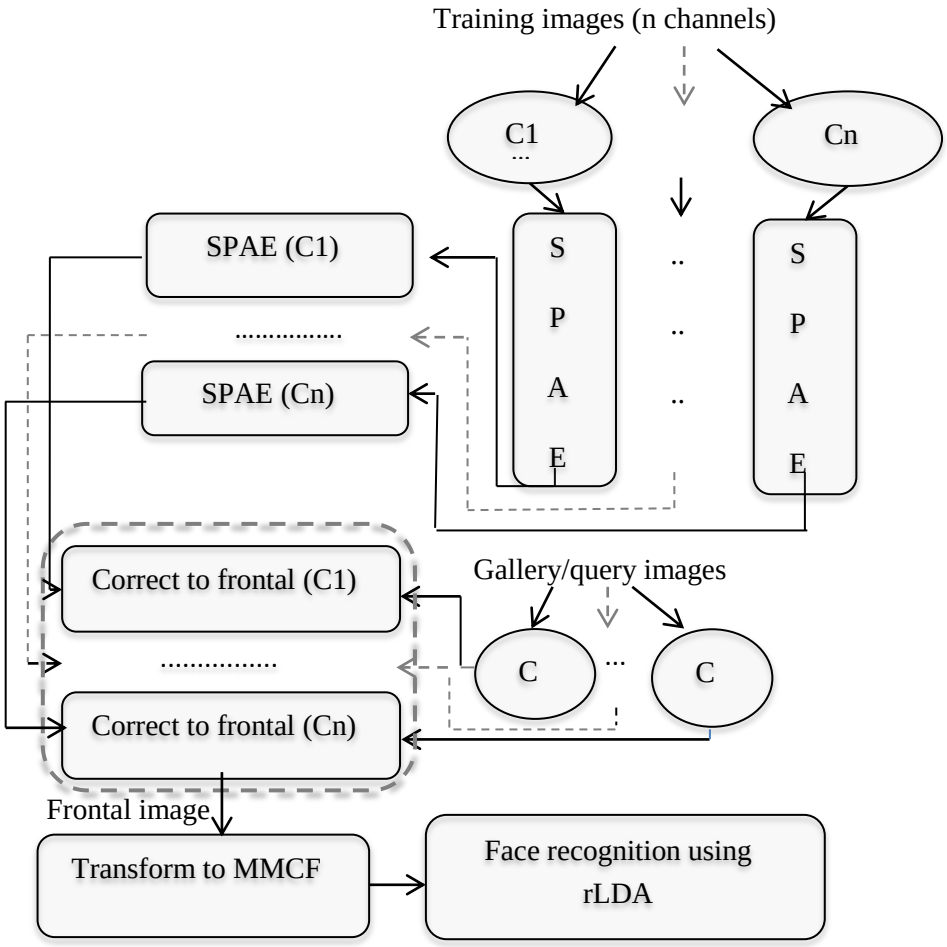


Figure 6.5 SPAE with MMCF for pose correction and face recognition.

One can see from above Figure that the SPAE corrects pose variations for each selected color channels and the proposed MMCF is used for face recognition with multi-clolor and multi-resolution fusion.

6.3. Experimental Setup and Results

In this section, the proposed MMCF model is evaluated for frontal face recognition in comparison with the RGB color space, the MCF model and 2D-MCF models. Six different face databases are employed for this purpose to show the effectiveness of the MMCF in different environments and conditions. The MMCF model is also evaluated for face recognition against

pose variations when integrated with the SPAE utilising the Multi-Pie database and Curtin database. Next, the databases used and experiments are presented in detail. For fair comparison with MCF, 2D-MCF, and the original SPAE, we use the rLDA with Nearest Neighbour for feature extraction and classification.

6.3.1. Experimental Setup

In this subsection, we present the databases used in different stages of our experiments as well as the experimental setup.

6.3.1.1. Databases

To evaluate the proposed MMCF model, we need a number of datasets for training the MMCF model, evaluation of the MMCF model for frontal face recognition, training the SPAE and evaluation of the MMCF model for face recognition against pose variations. For this purpose, the following databases are used.

1. The AR database (Martinez 1998) , which is already mentioned in previous chapters. In these experiments, 8 frontal images per subject are used for training the MCF model and evaluating the MMCF model for face recognition (Figure 6.5.a).
2. The Curtin database (Li, Liu et al. 2014), which is also mentioned in the previous chapters. In the experiments of this chapter, 8 frontal faces per subject are used for training the MMCF model and evaluation of the MMCF model for face recognition; then, 7 images with pose angles are used as gallery or query images for evaluating the MMCF model against pose variation (Figure 6.6.b).
3. The Georgia Tech face database GT Nefian (2007) which contains 750 images belonging to 50 subjects. As shown in Figure 6.6.c, the images in this database cover unregistered (not aligned), facial expressions, pose variations and they are not cropped well.
4. The LFW (Huang, Ramesh et al. 2007) face database, which is the most challenging database, contains faces from an unconstrained environment. In our experiment, 850 images belonging to 85 subjects are used. Figure 6.6.d shows some sample images of one subject from this database.

5. The FERET database (Phillips, Moon et al. 2000). This database contains 1614 frontal images belonging to 807 subjects, where each subject has only two images as shown Figure 6.6.e.
6. The Aberdeen database from Psychological Image Collection (PICs) face database (Hancock 2004). This database contains 261 images belonging to 29 subjects. Each subject has 9 mages covering small lighting condition and some pose variations as shown in Figure 6.6.f.
7. The Multi-Pie face database (Gross, Matthews et al. 2007). This database contains more than 12,000 face images belonging to 237 subjects. This database covers various range of pose angles of $[-90^0, -75^0, -60^0, -45^0, -30^0, -15^0, 0^0, 15^0, 30^0, 45^0, 60^0, 75^0, 90^0]$. In our experiments, the ranges of $[-45^0, -30^0, -15^0, 0^0, 15^0, 30^0, 45^0]$ are used for training the SPAE and evaluating the MMCF model against pose correction. Figure 6.6.g shows some sample images of one subject from this database.

In the experiments, we attempt to cover a large scope of challenges that may find in the real face recognition system including:

- Extreme facial expression in the AR and Curtin databases.
- Poor image registration and poor image cropping in the GT database.
- Single gallery image per subject in the FERET database.
- Pose variations in the Multi-Pie, PICS, GT and Curtin databases.
- The unconstrained environment in the LFW database.

6.3.1.2. Training Datasets and Classifiers

In the training of MMCF model, a large dataset from the AR and Curtin database is used. This dataset contains 912 frontal face images belonging to 152 subjects where, each subject has 6 images (the first six images in Figure 6.1.a and 6.1.b), in which 3 images per subject are used as gallery images and 3 images are used for testing. For training the SPAE, the same dataset with the same setup in the original SPAE is used. This dataset contains 4207 images belonging to 200 subjects, where each subject has 21 images covering pose angles of $[-45^0, -30^0, -15^0, 0^0, 15^0, 30^0, 45^0]$. For face recognition, the rLDA with NN classifier is used for fair comparison with the original SPAE.

Figure 6.6 samples images from different databases.

6.3.2. Creation of the MMCF Model

The MMCF model is built using a large training and testing dataset from the AR and Curtin databases as detailed in section 6.3.1.2. For each subject in the dataset, two images are used as gallery images and 4 images are used as query images, where the SRC is used for classification. Once the MMCF model is built, it can be used to represent all gallery and query images for face recognition.

6.3.3. MMCF Model for Face Recognition

To evaluate the efficacy of MMCF model for face recognition, various experiments on six different databases are conducted as detailed in table 6.2 and table 6.3. These experiments cover various conditions including facial expressions, single gallery image, pose variations, and uncontrolled environment. The superiority of the MMCF model over the RGB color space, the MCF model, 2D-MCF model can be clearly seen from the results in table 6.3.

For frontal face recognition with facial expressions and extreme open mouth from Curtin and AR databases, the MMCF model improves the recognition accuracy by.

- 0.65% over the 2D-MCF model achieving 98.35% where the 2D-MCF achieves 97.70%.
- For frontal face recognition and single gallery image from FERET database, the MMCF model achieves 96.01% where the 2D-MCF model achieves 96.10% which are nearly the same performance.
- On the other hand, an improvement of 1.83% is achieved with unregistered images in the GT database by achieving 85.15% where the 2D-MCF model achieves 83.32%.
- For Uncontrolled environment in the LFW database, a significant improvement of 8.24% is obtained over the 2D-MCF model by achieving 63.53% of accuracy where the 2D-MCF model achieves 55.29%.
- For pose variations and facial expressions in the PICS databases. The superiority of the MMCF can be seen by achieving 98.30% over the 2D-MCF which achieves 91.03%

These experiments show the robustness of the MMCF model against the challenging conditions of poor image quality, extreme expressions and pose variations overcoming the 2D-MCF model and MCF model. The stability of MMCF model against different conditions can be clearly seen from Figure 6.7. Unlike MCF and 2D-MCF model, the MMCF model is stable with different cases of expressions, poses, uncontrolled environment and single gallery image.

Table 6.2 Different databases used for face recognition using MMCF model.

DB	# Images	# Subjects	# Images /Subject	Conditions	# Gallery images	# Query images	Times
GT	750	50	15	Expressions / Misalignment	7	8	5
AR + CURTIN	1216	152	8	Extreme expressions / Open mouth	6	2	1
LFW	850	85	10	Uncontrolled / illumination / Pose / expressions	9	1	10
FERET	1614	807	2	Expressions / single gallery image	1	1	1
PICS	261	29	9	Expressions /Pose	4	5	5

Table 6.3 Face recognition using the proposed MMCF model in comparison with RGB, MCF and 2D_ MCF on 6 different databases

DB	Model	# Channels	Accuracy	Image Size
GT	RGB	3	75.85%	3072
	MCF	11	81.90%	11264
	2D-MCF	7	83.32%	7560
	MMCF	9	85.15%	5680
AR + CURTIN	RGB	3	89.67%	3072
	MCF	11	93.09%	11264
	2D-MCF	7	97.70%	7560
	MMCF	9	98.35%	5680
LFW	RGB	3	54.24%	3072
	MCF	11	60.35%	11264
	2D-MCF	7	55.29%	7560
	MMCF	9	63.53%	5680
FERET	RGB	3	88.96%	3072
	MCF	11	92.93%	11264
	2D-MCF	7	96.10%	7560
	MMCF	9	96.03%	5680
PICS	RGB	3	86.51%	3072
	MCF	11	89.12%	11264
	2D-MCF	7	91.03%	7560
	MMCF	9	98.31%	5680

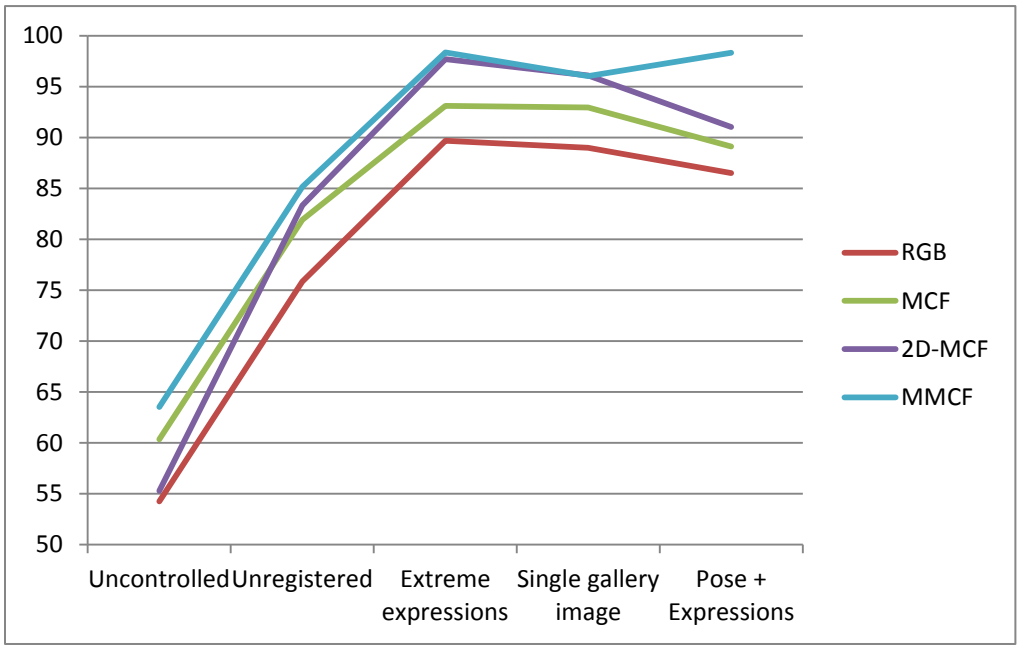


Figure 6.7 Face recognition using MMCF model in comparison with RGB images, MCF model, and 2D-MCF model

6.3.4. Integration of the SPAE with MMCF Model for Pose Correction

As illustrated in subsection 6.2.2, the SPAE pose correction can be integrated with the RGB and MMCF by using the SPAE for correcting different poses to the standard front pose with each color channel separately. Then, the RGB or MMCF is used to represent images based on the corrected color channels for face recognition. For fair comparison with the original SPAE, we use the same dataset used in the original SPAE as detailed in subsection 6.3.1.2. The SPAE neural network contains three hidden layers (or three stacked progressive auto encoders), each of which can correct a given pose to a smaller one as follows: the first layer corrects the pose angles from $(-45^0, 45^0)$ to $(-30^0, 30^0)$; the second layer corrects the pose angles from $(-30^0, 30^0)$ to $(-15^0, 15^0)$, and finally the third layer corrects the pose angles from $(-15^0, 15^0)$ to a standard frontal face.

6.3.5. Face Recognition against Pose Variations

In this subsection, the SPAE based MMCF (SPAE-MMCF)) is evaluated for face recognition against pose variations in comparison with the original SPAE for grey images (SPAE-G) and

the SPAE based on RGB images (SPAE-RGB)). For this purpose, two experiments on two different datasets are conducted as described below.

6.3.5.1. Evaluation on the Multi-Pie Database

This experiment is conducted on a dataset from Multi-pie containing 2,877 images belonging to 137 subjects. This is the same dataset used in the original SPAE (Gross, Matthews et al. 2008) for grey images. In this experiment, three frontal images are used as gallery images, where the rest 18 images which are already corrected to frontal pose are used as query images. The rLDA is used with NN classifier for fair comparison with the original SPAE as it was used in (Kan, Shan et al. 2014). The results of this experiment are shown in table 6.4. It is evidenced from table 6.4 that the SPAE (MMCF) model achieves a significant improvement over the SPAE (G) and the SPAE (RGB) especially with large pose angles and with single gallery image. One can summarise the observations in the following.

1. For small pose variations of $\pm 15^0$ and single gallery image, the SPAE-MMCF improves the recognition accuracy by up to 33.92% over the SPAE-G and up to 33.17% over the SPAE-RGB. This shows the importance of SPAE-MMCF as the SPAE-G and SPAE-RGB are completely fail by only achieving 45.09% to 46.22%, where some acceptable recognition rates are achieved using the SPAE-MMCF achieving from 77.94 to 79.39%. The weakness of using the grey images and RGB images is clearly seen in this case where promising results are achieved with the MMCF model.
2. With the single gallery image and large pose angles, both of the SPAE-RGB and SPAE-G are completely fail as they only achieve lower than 42% (in average) of accuracy with different pose angles. In instead, the SPAE-MMCF achieves an acceptable recognition rate up to of 75.2% which is very significant improvement over SPAE-RGB and SPAE-G. The SPAE-MMCF still achieves promising results in this extreme case of a single gallery image and large pose variation showing the high efficacy of this method.
3. For small pose variations and multiple gallery images, the SPAE-MMCF improves the recognition accuracy by 1.01% over the SPAE-G and also achieves exactly the same recognition rate of the SPAE-RGB. This small improvement can be justified as follows:
    a. The performance achieved by other methods is already very high. Consequently the possibility of improvement is very limited.

b. The correction of small pose angles leads to loss of very small amount of discriminant information which may not have a significant impact on face recognition accuracy.

4. For large pose angles ($\pm 30^0$ and $\pm 45^0$) and multiple gallery images, the SPAE-MMCF can improve the accuracy by up to 12% over the SPAE-RGB) and up to 17% over the SPAE-G. Again, we can see the superiority of the SPAE-MMCF compared with SPAE-G and SPAE-RGB.

Table 6.4 Face recognition on Multi-Pie using SPAE-MMCF in comparison with SPAE-RGB, SPAE-G

Number of galleries	Pose angle / Approaches	-45	-30	-15	+15	+30	+45	Average
		Recognition Rate %						
1	SPAE-G	32.76	39.58	45.09	45.74	38.84	32.60	40.37
	SPAE-RGB	32.84	40.39	45.74	46.22	41.28	34.39	41.60
	SPAE-MMCF	61.96	69.02	77.94	79.39	75.02	65.53	73.38
2	SPAE-G	79.40	92.86	98.62	98.22	94.57	82.89	93.43
	SPAE-RGB	84.91	95.95	99.27	98.54	97.81	90.51	96.42
	SPAE-MMCF	96.92	98.70	99.27	98.54	98.38	97.48	98.46
3	SPAE-G	82.73	93.67	98.78	98.54	95.86	84.43	94.26
	SPAE-RGB	86.62	97.57	99.27	98.54	98.78	94.16	97.66
	SPAE-MMCF	97.81	98.78	99.27	98.54	98.78	97.81	98.64

The efficacy and stability of the SPAE-MMCF can by clearly seen from Figure 6.8, where the SPAE-G and SPAE-RGB are clearly unstable.

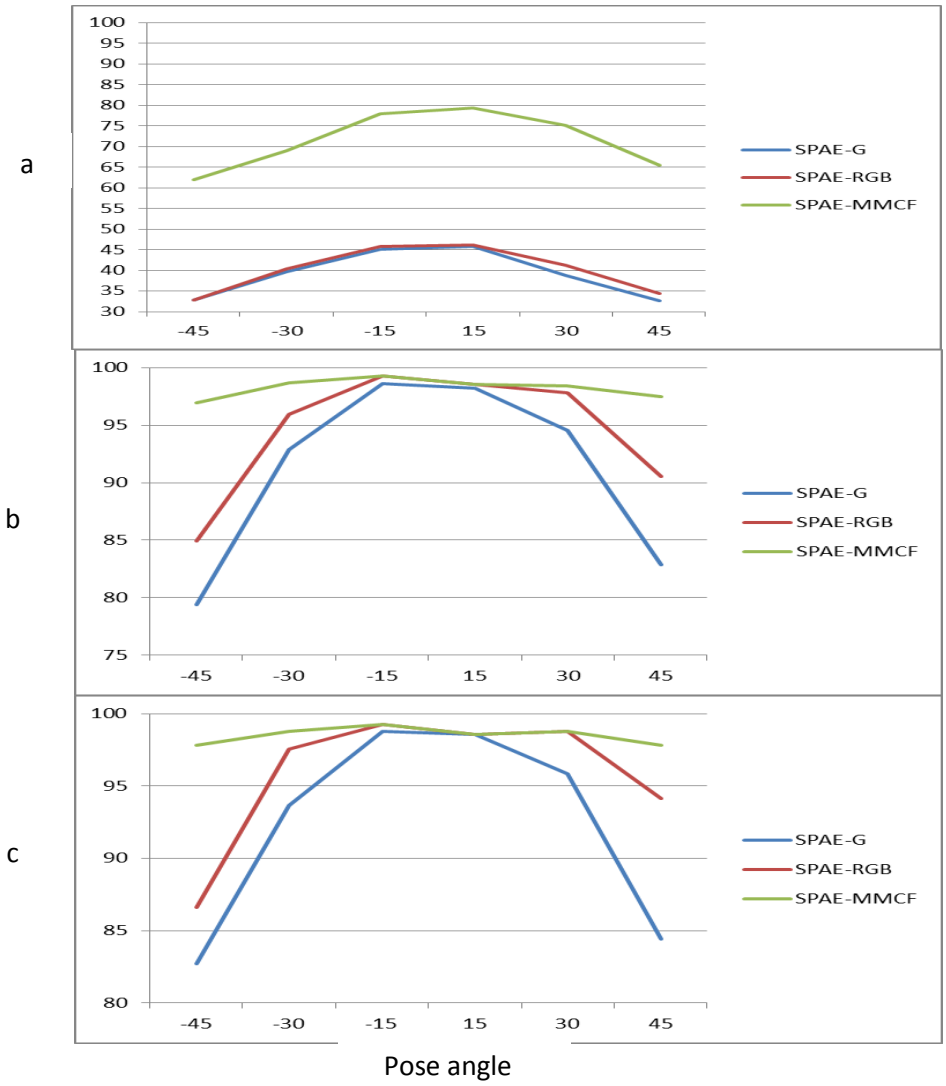


Figure 6.8 a. Face recognition on corrected images from Multipie database using SPAE-MMCF in comparison with SPAE-G and SPAE-RGB with: a: single gallery image, b: 2 gallery images, c: 3 gallery images

6.3.5.2. Evaluation on the Curtin Database

In the previous experiments, the training of SPAE and implementation of face recognition are conducted on the same dataset. In this experiment, the evaluation of the SPAE-MMCF is conducted on the Curtin database for face recognition while the SPAE is trained on the Multi-Pie database. This cross database testing is a challenging problem, which is quite common in real facial recognition systems. Moreover, in this experiment, facial expressions are also considered, which makes the recognition problem more difficult. The dataset used in this experiment contains 676 images belonging to 52 subjects from the Curtin database. Each subject has 13 images in which 6 frontal images are used as gallery images where the rest 7 images with pose variations are used as query images. First of all, all images are corrected to the frontal pose as illustrated earlier, and then, face recognition is performed using rLDA with the NN classifier. As shown in table 6.5, the SPAE-G and SPAE-RGB achieve very low recognition rates from 44.23% to 57.69% for SPAE-RGB and from 34.62% to 42.3% for SPAE-G. However, the SPAE-MMCF has achieved very promising rates with average of 92.31% for pose angles in the range $\pm 30^0$ and 82.85% with pose angle of $\pm 30^0$ plus different facial expressions.

Table 6.5 Face Recognition on corrected images from Curtin database using SPAE-MMCF in comparison with SPAE-RGB and SPAE-G.

	# Gallery images	# Query images	-30^0-	30^0+	Average
Pose + expressions					
SPAE-G	6	6	34.62%	37.82%	36.22%
SPAE-RGB	6	6	44.23%	54.17%	49.20%
SPAE-MMCF	6	6	81.41%	84.29%	82.85%
Pose					
SPAE-G	6	1	42.30%	34.62%	38.46%
SPAE-RGB	6	1	57.69%	45.19%	51.44%
SPAE-MMCF	6	1	94.23	90.38%	92.31%

The efficacy of using the SPAE-MMCF can be clearly seen from Figure 6.9, where the superiority of SPAE-MMCF over the SPAE-G and SPAE-RGB is very obvious.

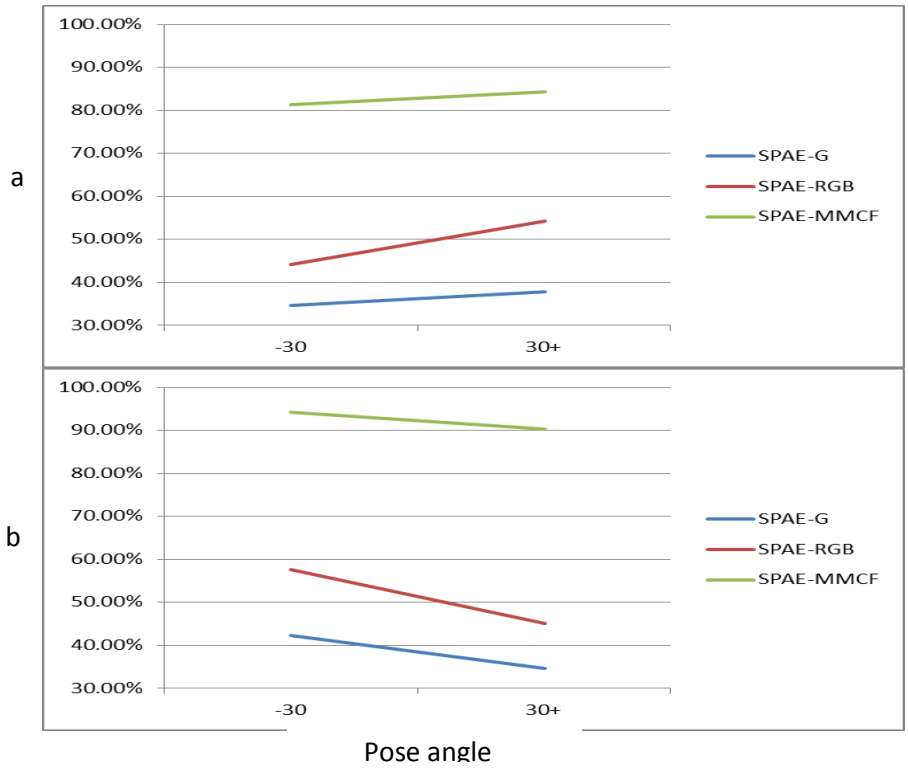


Figure 6.9 Face Recognition on corrected images from Curtin database using SPAE-MMCF in comparison with SPAE-RGB and SPAE-G. a: pose + facial expressions. b: Pose and neutral expression.

From the previous experiments, one can see that the use of grey images for pose correction is not very promising especially with large pose variations, one single gallery image and cross database testing. The use of RGB color information shows better recognition performance in comparison with the grey image when the training and testing samples are from the same environment or dataset. In cross datasets, the SPAE-RGB is still not working properly. So, the use of grey mages and RGB images is impractical in the real face recognition systems against pose variations. On the other hand, the use of SPAE-MMCF shows very high performance in different environments including the cross datasets for training and testing, and the one single gallery image case. The minimum recorded recognition rate of 81.41% is highly acceptable for challenging conditions of large pose angles, cross datasets for training and testing with different facial expressions. These results confirmed our assumptions that the pose variations can lead to

some losses of discriminant information and the effective use of color information in the MMCF can compensate this loss significantly and effectively.

6.4. Summery

In this chapter, a new multi color fusion model named as the Multi-resolution Multiple Color Fusion (MMCF) is proposed for face recognition against pose variations. This new model has utilised multiple color channels with different resolutions. The idea is to select different resolutions for different color channels as contribution weighting based on the value of this channel in face recognition. Technically, each color channel/resolution in the MMCF model is selected based on its performance for face recognition as in the MCF model. The proposed MMCF model outperforms the MCF and 2D-MCF for frontal face recognition. In addition to that, it can be effectively used in pose correction for face recognition. As evidenced in the experiments in this chapter, the pose correction process may lead to a loss of valuable discriminant information and make the SPAE useless in the real face recognition systems. This loss of information can be compensated by effective fusion of color information in the MMCF. The experimental results show that the integration of the MMCF model with SPAE can compensate some valuable discriminant information and make the SPAE work effectively in different environments including large pose angles, facial expressions, cross database testing and one single gallery case.

Chapter Seven

Enhancing Face Recognition against Illuminations using Two Directional Multi-Level Threshold-LBP Fusion

Illumination variation is one of the challenging problems in face recognition. The negative impact of illumination on facial recognition has attracted much attention as one of the main difficulties in face recognition. The importance and difficulty of this problem lie mainly in its apparent negative impact on the accuracy of facial recognition. One main reason is that the intensity of light and its locations affected by the illumination are unpredictable. Furthermore, in extreme illumination variations, the direct use of grey or colored images may lead to a complete failure in facial recognition. Many researches have been conducted trying to remove the effect of illumination. In (Shiguang, Wen et al. 2003), Histogram Equalization is used in combination with Gamma Intensity Correction (GIC) to remove the effect of illumination. Some other methods are based on Quotient Image (IQ) method (Shashua and Riklin-Raviv 2001) such as those proposed in (Wang, Li et al. 2004, Xiaohua, Wei-Shi et al. 2008, Perez and Castillo 2009). The Discrete Cosine Transform (DCT) is also used in logarithmic domain (Log-DCT) (Chen, Meng Joo et al. 2006) for illumination compensation and normalisation in application of face recognition. The gradient faces and weber faces proposed in (Taiping, Yuan Yan et al. 2009, Chen, Shan et al. 2010, Biao, Weifeng et al. 2011) utilize the local pixels information to normalize the image information for robust face recognition. The Logarithmic Total Variation (LTV) (Chen, Wotao et al. 2006) normalizes a face image by factorizing it with an aim to obtain a robust facial structure for face recognition. Some other methods combine variety of elementary features in order to take the advantages of each elementary feature. For example, the Multiscale Local Phase Quantization for Robust Component-Based Face Recognition Using Kernel Fusion of Multiple Descriptors (Chan, Tahir et al. 2013) introduces a Multi-Scale Local Quantisation representation (MLPQ) in combination with Multi-scale Local Binary Pattern (MLBP) to decrease the sensitivity to illumination. The Multi-Scale Logarithm Difference Edge-maps (MSLDE) (Lai,

Dai et al. 2015) fuses 6 edge-maps of local differences in logarithmic domain. The fusion decision in (Lai, Dai et al. 2015) is based on training where different weightings are assigned to the edge-maps. The result of this fusion is finally only one channel representation. This method outperforms other existing methods for face recognition against illumination variations. The main drawback of this method is that it still uses only one channel to represent an image. As demonstrated in chapter 4, chapter 5 and chapter 6, valuable discriminant information may lie in multiple color channels. In contrast, the use of only one channel with any normalisation method may lead to loss of some relevant information in other colors. Other methods such as Local Directional Number Pattern (LDN) (Ramirez Rivera, Rojas Castillo et al. 2013), Enhanced Local Directional Pattern (ELDP) (Zhong and Zhang 2013) and Adaptive Homomorphic Eight Local Directional Pattern (AHELDP) (Faraji and Qi 2015) also utilise only one channel to represent the face image. Consequently, such representation may lead to loss of valuable discriminant information. To the best of our knowledge, the Threshold Local Binary Pattern (Meng, Gao et al. 2010) is the only method that has used more than one channel to represent face images for face recognition with LBP. This method utilises two levels of the Local Binary Pattern (LBP) which are, the classical LBP with threshold=0 and the adopted LBP with threshold=5. However, this method lacks optimal selection of the LBP channels. Moreover, the selection of threshold is based on the subjective observation rather than objective training. Besides, this method is designed for general face recognition rather than face recognition against illumination.

As the use of multiple color channels can improve face recognition accuracy, it is expected that the use of multiple channels resulting from illumination normalisation processes can also enrich the discriminant information for face recognition. In this chapter, a new multi-channel image representation based on LBP is proposed. This new method is based on the Multilevel of Threshold LBP (TLBP), which is obtained from the face image with normalisation using the Discrete Cosine Transform (DCT). The fusion decision is taken based on training after segmenting the image to non-overlapped blocks. The proposed approach is named as the *Two Directional Multilevel Threshold Local Binary Pattern Fusion* (2D-MTLBP-F) as the training procedure is similar to the 2D-MCF proposed in previous chapters. Technically, a face image is first normalised using the DCT in order to remove some illumination, noise and blur. Secondly, the multiple levels of TLBP (Meng, Gao et al. 2010) are obtained by using several scale of thresholds. This step results in a large number of TLBP levels. Finally, this large number of TLBP levels (Channels) is reduced by selecting the face images for face recognition using the previously proposed 2D-MCF criteria. The resultant representation is in fact a Multilayer (Channels) of TLBP that fused in block-wise where the criteria of selecting any TLBP level is

based on its performance in face recognition on different blocks. The idea behind it is to obtain different levels of discriminant information that can improve the recognition accuracy. Instead of fusing multiple TLBP levels (channels) into one channel, they can be fused in a multi-channel representation in order to avoid the loss of some valuable discriminant information. Even though some TLBP levels have fewer details, they can still be used to remove some irrelevant information such as wrinkles and lighting effects.

The rest of this chapter is organised as follows, the next section introduces the LBP and TLBP in some details. In section 7.2, the Discrete Cosine Transform (DCT) is reviewed. The proposed Two Directional Multilevel Threshold Local Binary Pattern Fusion (2D-MTLBP-F) is presented in detail in section 7.3. The proposed method is evaluated with extensive experiments in section 7.4. Finally, the chapter is summarised in section 7.5.

7.1. Threshold Local Binary Pattern (TLBP)

As mentioned in subsection 2.4.1.2, the Local Binary Pattern is one of image descriptors widely used in face recognition. Like all other image descriptors for grey images, the LBP represents an image using only one channel (layer). However, this single channel representation has been modified for two channels in the Threshold Local Binary Pattern (TLBP) (Meng, Gao et al. 2010). The idea is to use another threshold to obtain the LBP in addition to the original LBP with threshold=0. The proposed method uses a threshold = 5 such that the neighbouring pixels will have a value of 0 if (the Neighbouring pixel - centred pixel < 5). The resultant descriptor is a different level of LBP with different level of information as Figure 7.1 illustrates the TLBP in comparison with the LBP.

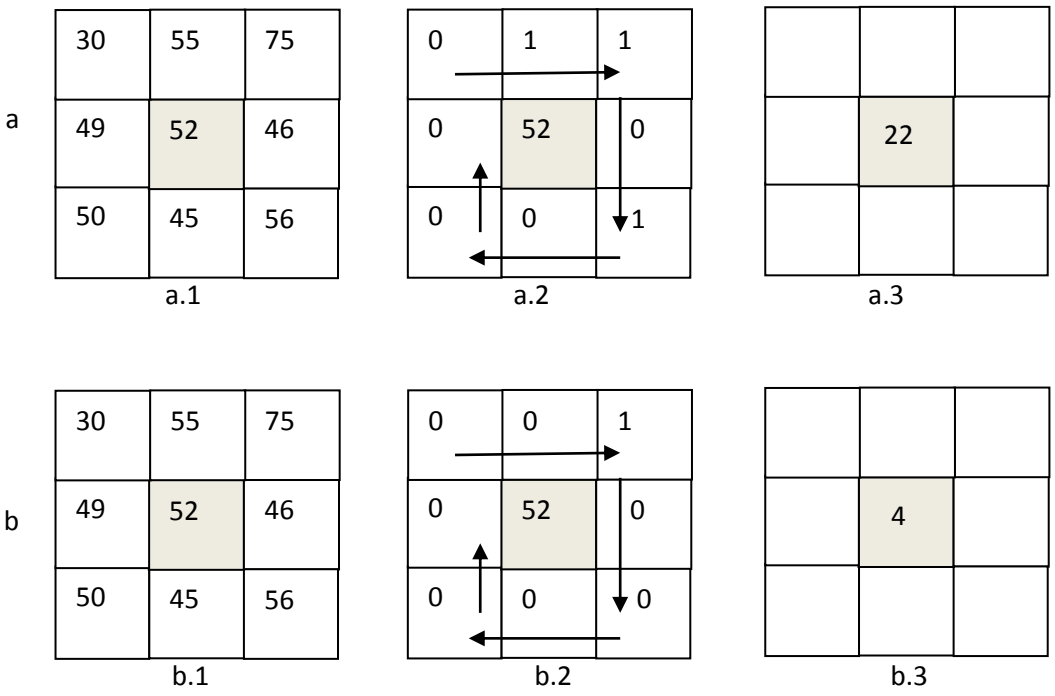


Figure 7.1 a. Pixel-wise LBP.a.1 3X3 block. a.2 Binary values of neighbouring pixels according to LBP criteria. a.3 LBP value of centred pixel. b. Pixel-wise TLBP with Threshold t=5. b.1 3X3 block. b.2 Binary values of neighbouring pixels according to TLBP criteria. b.3 TLBP value of centred pixel.

Eq 2.13 that computes the binary values of the neighbouring pixels Z is modified into TLBP with Threshold t=5 as follows.

$$B_j = \begin{cases} 0 & Z_j - P(x,y) < 5 \\ 1 & Z_j - P(x,y) \geq 5 \end{cases} \quad j = 0,1,\dots,7 \tag{7.1}$$

This equation can be further extended for any threshold value t as follows

$$B_j = \begin{cases} 0 & Z_j - P(x,y) < t \\ 1 & Z_j - P(x,y) \geq t \end{cases} \quad j = 0,1,\dots,7 \tag{7.2}$$

Finally, the decimal TLBP value of the centred pixel can be computed according to Eq. 2.14. As shown in Figure 7.2, different levels of information can be obtained by different levels of TLBP. As we will see later in this chapter, this information can be effectively utilised for face recognition.

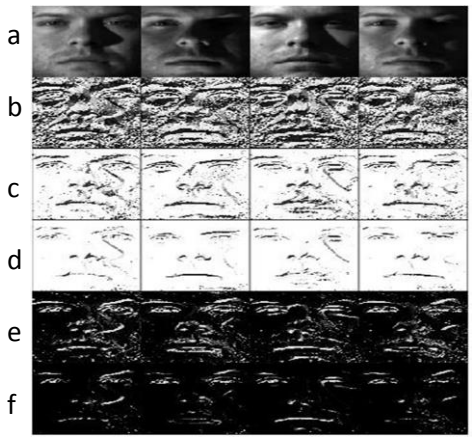


Figure 7.2 . Different levels of information from different TLBP levels. a. Sample images from Extended Yale database. b to f. Different levels of TLBP showing different levels of details. b. t=0. c. t= -10. d. t= -20. e. t=10. f. t=20.

For face recognition task, two levels of TLBP are utilised which are $TLBP_0$ with threshold=0 and $TLBP_5$ with threshold=5 as described below.

The following algorithm is proposed for face recognition (Meng, Gao et al. 2010).

1. Each face image is transformed into two levels of TLBP which are $TLBP_0$ with t=0 and $TLBP_5$ with t=5.
2. The $TLBP_0$ is segmented into 16 non-overlapped blocks and the histogram h_i is obtained for each block, and then the histogram of the whole face image H_0 is obtained by concatenating all histograms together, $H_0 = [h_1, h_2, \ldots, h_{16}]$.
3. The histogram of the $TLBP_5$ H_5 is obtained similarly for the whole face image.
4. The H_0 and H_5 are concatenated to form one vector such that $H= [H_0, H_5]$.
5. Finally, the vector H is used for face recognition with PCA and LDA.

The TLBP is compared with the LBP for face recognition showing a stable improvement in recognition accuracy. Even though the TLBP with two channels shows a stable improvement in face recognition accuracy, it is expected that the use of only two channels of TLBP may not contain sufficient information for facial recognition. Furthermore, the selection of threshold is done subjectively rather than based on objective training. Finally, the proposed TLBP uses the histogram as feature vector and many researches show that the use the LBP image can give much better results than the use of histogram features directly (Yang and

Chen 2013) especially in the presence of illumination variations. By taking above observations in the account, we expect that the use of training-based multi-level TLBP would result in much higher performance in face recognition in general especially when using the TLBP image instead of TLBP histogram as we will show later in this chapter.

7.2. Discrete Cosine Transform (DCT)

The discrete Cosine Transform is widely used in image compression and face recognition as a dimensionality reduction technique (Rao 1990, Clark 1994, Chen, Meng Joo et al. 2006, Aroussi, Amine et al. 2008, Dabbaghchian, Ghaemmaghami et al. 2010). The 2D-DCT transformation can be formulated as follows.

$$C(u,v) = \propto (u) \propto (v) \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) \times cos\left[\frac{\pi(2x+1)u}{2M}\right] \cos\left[\frac{\pi(2y+1)v}{2N}\right] \qquad (7.3)$$

To recover the original image from the 2D-DCT representation, the inverse 2D-DCT is applied as follows.

$$f(x,y) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} \propto (u) \propto (v) C(u,v) \times \cos\left[\frac{\pi(2x+1)u}{2M}\right] \cos\left[\frac{\pi(2y+1)v}{2N}\right] \qquad (7.4)$$

where $\propto$ (u) and $\propto$ (v) can be computed as follows:

$$\propto (u) = \begin{cases} \frac{1}{\sqrt{M}}, & u = 0 \\ \sqrt{\frac{2}{M}}, & u = 1, 2, \dots, M-1 \end{cases}$$

.

$$\propto (v) = \begin{cases} \frac{1}{\sqrt{N}}, & v = 0 \\ \sqrt{\frac{2}{N}}, & u = 1, 2, \dots, N-1 \end{cases}$$

where $f(x, y)$ is the pixel value in the special domain and $C(u, v)$is the pixel value in 2D-DCT frequency domain.

The DCT image can be divided into three bands, which are Low frequency band (L) which contains information that highly correlated with illumination, Middle frequency band (M) which contains the relevant information that can be effectively used to reconstruct the structure of the original image, and finally, the High frequency band (H) which contains noise and trivial variations. These sub-bands are shown in Figure 7.3. It is clear that, the relevant information for face recognition lies in the middle frequency band. Consequently, this band is usually used for image normalisation as well as DCT feature vector for classification, where the high frequency band and low frequency band are often ignored. Next we describe the image normalisation process based on DCT.

For image normalisation:

- The image is transformed to frequency domain using 2D-DCT with Eq. 7.3.
- The high/low bands are removed by assigning zero values to them.
- The original image is reconstructed according to Eq. 7.4.

For face recognition:

- The image is transformed to frequency domain using 2D-DCT using Eq. 7.3.
- The middle band is used as DCT features vector for face recognition using a suitable classifier.

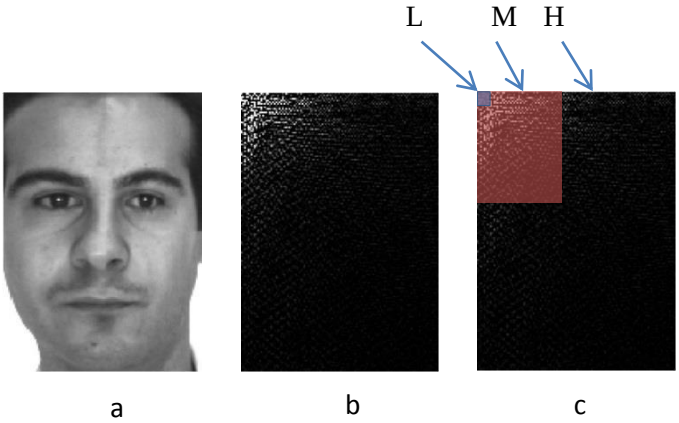


Figure 7.3 a. A typical face image; b. The DCT transformed image. c. The division of the DCT coefficients into low, middle and high frequencies

In this chapter, the 2D-DCT will be used for face image normalisation before applying the Threshold LBP. Such normalisation can effectively enhance image quality by reducing the effects of noise and illumination. Next, the Two Directional Multilevel Threshold Local Binary Pattern Fusion (2D-MTLBP-F) is introduced.

7.3. The Two Directional Multilevel Threshold Local Binary Pattern Fusion (2D-MTLBP-F)

In this section, the 2D-MTLBP-F is proposed. This method is based on multilevel of the TLBP where the fusion decision is taken locally in a block-wise manner. The selection of each TLBP level in each block is taken based on the performance of this level in face recognition. For this purpose, the criteria of 2D-MCF are used. As in the 2D-MCF, a training and testing dataset is required for training the proposed 2D-MTLBP model. The goal is to obtain the best representation for face recognition against illumination variations. The DCT is used as a pre-processing step in order to reduce some noise and illumination effects, and then the 2D-MTLBP model is trained for face recognition. Finally, the obtained model is used to represent

all gallery and query images for face recognition. In the following subsections, the whole procedure is described in details.

7.3.1. DCT Normalisation

This step can be considered as a pre-processing step where all images in the training and testing dataset (in the training phase) and all gallery/query images (in face recognition process) are normalised using the DCT. The DCT has been used in combination with LBP for efficient face recognition by either fusing the LBP features with DCT features (Dasari, Kraleti Srinivasa et al. 2012) or by obtaining the DCT features from the LBP representation. In this work, the DCT is used to enhance the image quality before performing the model training or face recognition.

7.3.2. Multi-Level TLBP Representation

In this step, the number of TLBP levels that useful for face recognition is obtained. First, using all possible TLBP levels is impractical because this leads to very large size of image as there are total 509 levels from -255 to +255. On the other hand, using a very small number of TLBP levels may be not sufficient for providing the relevant discriminant information. This problem can be solved via training in pre-processing step. The goal of this step is to obtain the optimal number of the TLBP levels that can be effectively used for face recognition. In this pre-processing scenario, the 2D-MTLBP-F is trained repeatedly several times. In each round, the number of TLBP levels used in the training is increased and the resultant template is built after each round. The steps of obtaining this TLBP level number is outlined in the algorithm 7.1.

Algorithm 7.1.

1. *Set t=5.*
2. *Transform DCT image into 2×t+1 TLBP levels (from –t to +t). Eq (7.2) is used to obtain binary values of neighbouring pixels, where Eq (2.14) is used to obtain TLBP decimal values. Here, TLBP levels with (-t) and (+t) are called* ***Boundaries*** *of MTLBP.*

3. *Train a 2D-MTLBP-F model using the MTLBP levels obtained in step-2. The result of this step is a 2D-MTLBP-F representation containing the selected levels from MTLBP. If this selection does not contain one of the Boundaries of MTLBP, go to step 4.*
4. *Update boundaries as follows: t=t+5 then go to step 2. (This expands the number of levels by adding extra more 10 levels).*
5. *Add the original DCT image to MTLBP. The maximum number of levels obtained by previous steps becomes the number of MTLBP levels + 1.*

Based on this proposed pre-training, 61 levels of TLBP are selected to be used in training the 2D-MTLBP –F model from *t=-30 to t=30.* As the normalized image (DCT image) may contain relevant but different information, it is added to these 61 TLBP levels and the total becomes 62 Levels or channels. Figure 7.4 shows a sample image from the extended Yale database with these 62 channels.

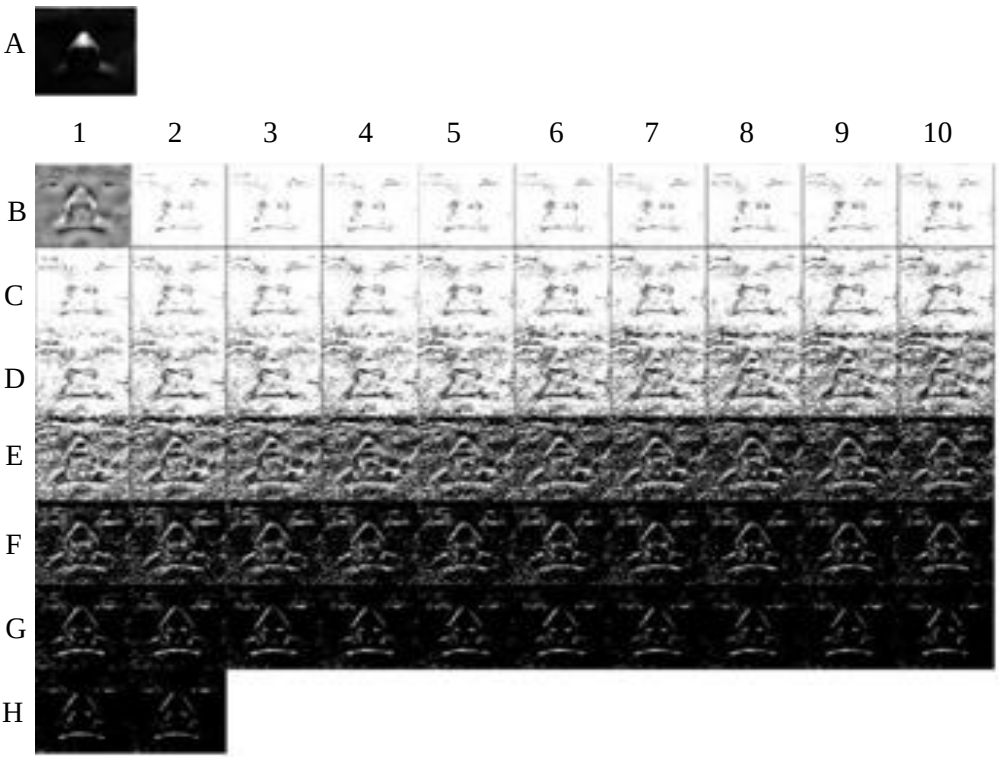


Figure 7.4 A. Illuminated grey image. B1. DCT image. B2 to H2 with 61 levels of TLBP

7.3.3. Training the 2D-MTLBP-F Model

Even though the images obtained in the previous subsection are likely to have the required discriminant information for face recognition, the direct use of such representation may lead to low performance in face recognition process because of following reasons.

- The resultant number of images is very large which consumes time and large memory space in face recognition if all these images are used.
- The neighboring TLBP levels are highly correlated. Such correlation may decrease the recognition accuracy.

To avoid these shortcomings, a new representation is trained for face recognition in a block-wise manner using the similar idea in 2D-MCF, and the only difference is to use the MTLBP channels instead of different color channels. The goal of such training is to:

- Reduce the number of the images in face recognition process and,
- Optimize the image representation and provide the best possible representation for face recognition.

The training of 2D-MTLBP-F model is illustrated in Figure 7.5.

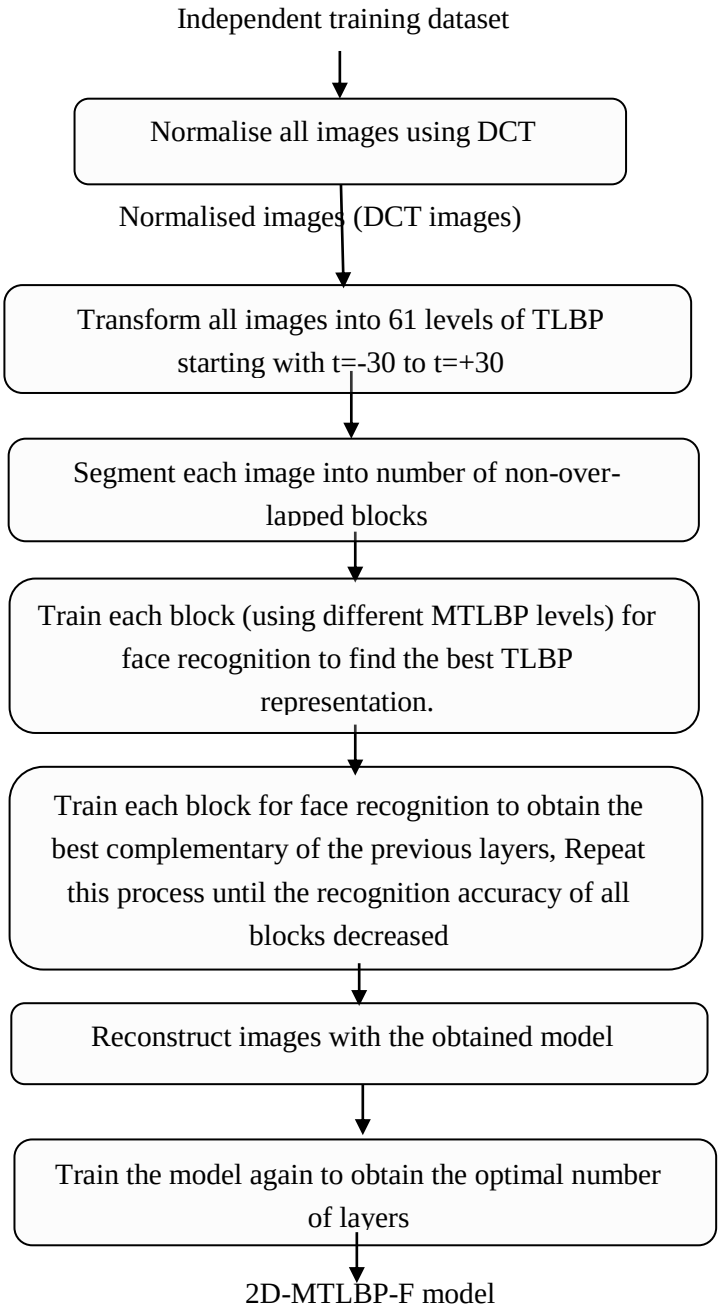


Figure 7.5 Training the 2D-MTLBP-F model

The result of this training process is the 2D-MTLBP-F model with only a small number of layers which is quite small in comparison with the 62 layers of full MTLBP representation. On the other hand this model can achieve high performance as we will show later in this chapter.

7.3.4. Face Recognition

Once the 2D-MTLBP-F model is obtained as described in the previous subsection, it can be used for face recognition. All gallery and query images are transformed into the 2D-MTLBP-F model, and then the face recognition is performed using the SRC classifier (Wright, Yang et al. 2009). As the LBP image performs better than its histogram bins in face recognition (Yang and Chen 2013), it is used with the SRC classifier without any dimensionality reduction for face recognition. As we will see in the experiments, the 2D-MTLBP is evaluated by comparing it with the The Multi-Scale Logarithm Difference Edge-maps (MSLDE) (Lai, Dai et al. 2015) which adopted the SRC for classification. Consequently, we use the SRC classifier for fair comparison. The process of face recognition using the 2D-MTLBP-F is stated in Figure 7.6.

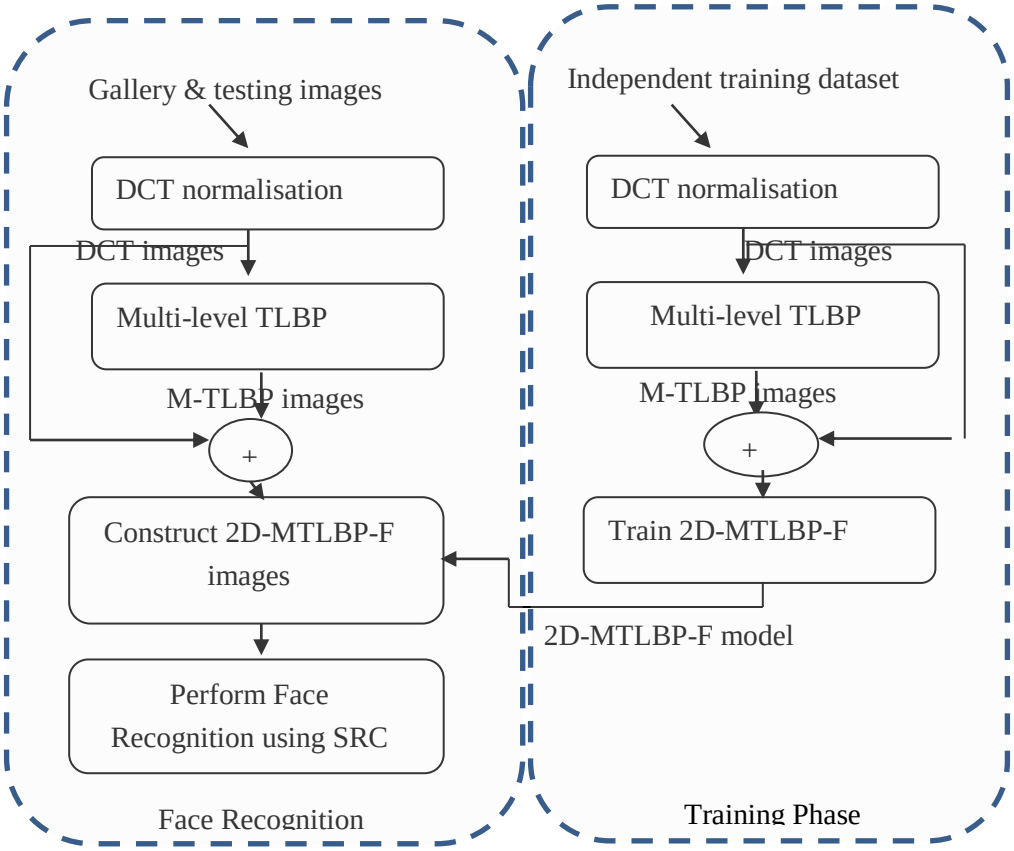


Figure 7.6 The structure of the whole system of the proposed 2D-MTLBP-F method

7.4. Experimental setup

In order to evaluate the efficacy of the proposed 2D-MTLBP-F, extensive experiments are conducted in this section utilising five different databases. The proposed 2D-MTLBP-F is

compared with the original TLBP (Meng, Gao et al. 2010) because it is the basis of the 2D-MTLBP. The recently proposed MSLDE (Lai, Dai et al. 2015) is already compared with different other methods including the Log-DCT (Chen, Meng Joo et al. 2006), the Gradient faces (Taiping, Yuan Yan et al. 2009, the Weber faces (Biao, Weifeng et al. 2011), The LTV (Chen, Wotao et al. 2006), the AHELDP (Faraji and Qi 2015), the LDN (Ramirez Rivera, Rojas Castillo et al. 2013), and the ELDP (Zhong and Zhang 2013). Consequently, it is sufficient for us here to compare the 2D-MTLBP-F with the MSLDE as the MSLDE outperforms other methods.

7.4.1. Databases and Experimental Setup

In these experiments, five different databases are used to evaluate the proposed 2D-MTLBP-F. This includes datasets from the AR database (Martinez 1998), the Multi-Pie database (Gross, Matthews et al. 2008), the Curtin database (Li, Liu et al. 2014), the Extended Yale database (Georghiades, Belhumeur et al. 2001) and the FRGC database (Phillips, Flynn et al. 2005). Detailed information about the datasets used in the experiments is shown in table 7.1 where sample images from different dataset are shown in Figures 7.7, 7.8, 7.9, 7.10 and 7.11.

Table 7.1 Different datasets used in the experiments.

	#imgs	# Subs	# Imgs per sub	Gallery	Test	Conditions
AR	1400	100	14	8	6	Brightness
Ex-Yale	2414	38	59-64	7	52-57	Darkness
Curtin	2236	52	43	8	35	Lighting/ Expressions
Multi-Pie	9960	249	40	8	32	Lighting/ expressions
FRGC	7992	222	36	8	28	Lighting/ expressions/uncontrolled

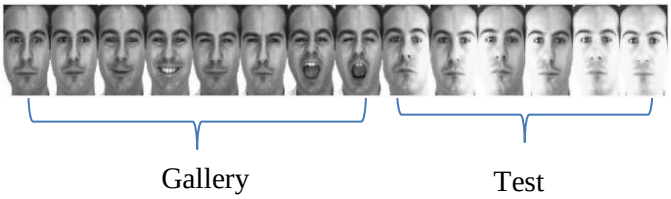


Figure 7.9 Sample Of gallery and test images from AR database

Figure 7.7 Sample Of gallery and test images from Curtin database with variations in lighting and expressions

Figure 7.8 Sample images of one subject from Extended Yale-B Database showing the illumination variations.

Figure 7.10 Unregistered images of one subject from Multi-Pie database with variations in illumination and expressions.

Figure 7.11 Sample images of one subject from the FRGC database with different illumination and expressions.

7.4.2. Building the 2D-MTLBP-F Model

In order to evaluate the robustness of the 2D-MTLBP-F model, three different templates of the 2D-MTLBP-F are built based on three different datasets from different environments. These three templates are used later to represent gallery and test images in face recognition process. Table 7.2 details these three templates and the datasets used to train each template. To train each template from table 7.2, all images in the corresponding dataset are first normalised using the DCT, and then, the dataset is divided into gallery and query images. The 2D-MTLBP-F template is trained as described earlier in this chapter using the rLDA for feature extraction and the Nearest Neighbour NN with Euclidian distance for classification.

Table 7.2 Datasets used for training different 2D-MTLBP-F templates.

	Template	#Imgs	#Sub	Imgs/Sub	Gallery	Test
AR	Tem1	400	50	8	4	4
Curtin	Tem2	572	50	11	4	7
EX-Yale	Tem3	380	38	10	4	6

7.4.3. Experimental Results

In this subsection, the 2D-MTLBP-F model is evaluated for face recognition utilising the 5 datasets from table 7.1. Three different templates (Tem1, Tem2, and Tem3) are obtained by training on three different databases, and they are used to represent the query and gallery images for face recognition. This can better evaluate the robustness of the 2D-MTLBP-F model.

7.4.3.1. Experiment One

This experiment is conducted on the AR database. As shown in Figure 7.7, 8 un-illuminated images are used as gallery images, where 6 illuminated images are used as query images. The face recognition is performed using the SRC classifier for fair comparison with the MSLDE. The results are presented in table 7.3 showing the superiority of the 2D-MTLBP-F over the

TLBP and MSLDE. To evaluate the stability of the 2D-MTLBP-F model, Tem2 and Tem3 are also used to represent the images achieving 100% of accuracy in both cases. This shows the stability of the proposed method when trained on different training and testing datasets.

Table 7.3 Face recognition results using the proposed method in comparison with different methods on AR face database.

times	Grey image	TLBP	MSDLE	2D-MTLBP-F	Template
1	77.50%	93.22%	99.67%	100%	Tem2
				100%	Tem3

7.4.3.2. Experiment Two

This experiment is conducted on the Curtin database as detailed in table 7.1. This database contains more challenging cases of illumination and facial expressions. All gallery images are un-illuminated, where the query images are illuminated with some facial expressions. All images are transformed to the 2D-MTLBP-F model using Tem1 and Tem3 and then the face recognition is performed using the SRC classifier. It is evidenced from the results in table 7.3 that the proposed method is stable against different training datasets and robust against the challenging cases when the query image is illuminated and has facial expressions. Comparing with the TLBP and MSLDE, the 2D-MTLBP-F method improves the recognition accuracy by ~5% over the MSLDE and ~11% over the TLBP.

Table 7.4 Recognition rates achieved by the 2D-MTLBP, MSLDE, TLBP and grey images conducted on Curtin face database.

times	Grey image	TLBP	MSDLE	2D-MTLBP-F	Template
1	28.67%	85.11%	91.23%	96.26%	Tem1
				96.42%	Tem3

7.4.3.3. Experiment Three

This experiment is conducted on the extended Yale database which contains 59 to 64 images per subject. All images are neutral expression and different lighting variations as shown in Figure 7.9. This experiment is divided into two sub-experiments as follows.

Firstly, 7 un-illuminated images per subject are used as gallery images, where the rest 25 to 75 images are used as query images. The reason of this selection is that the gallery images in real systems usually taken in controlled environment with neutral lighting conditions. The SRC is used as a classifier where Tem1 and Tem2 are used to represent the 2D-MTLBP-F images. The results in table 7.5 show the superiority of the 2D-MTLBP over other methods. Also, the stability of the 2D=MTLBP can be clearly seen where the results achieved by 2 different templates are nearly the same.

Secondly, face recognition is performed 5 times using 7 gallery images chosen randomly where the rest images are used as query images. Here, any gallery image or query image can be either illuminated or un-illuminated. The 2D-MTLBP still achieves high recognition rate with both Tem1 and Tem3 templates overcoming the other methods. The results shown in table 7.6 are the mean of the 5 rounds of experiments. This show the robustness of the 2D-MTLBP-F even when the gallery images are not clear in addition to the stability in the case of cross database testing.

Table 7.5 Results of face recognition of the proposed method in comparison with other methods on the Extended Yale B database with 7 neutral gallery images

Times	Grey image	TLBP	MSDLE	2D-MTLBP-F	Template
1	32.50%	92.67%	98.79%	99.68%	Tem1
				99.77%	Tem2

Table 7.6 Recognition rates for different method on Extended Yale B database with different illumination conditions for gallery and test images

times	Grey image	TLBP	MSDLE	2D-MTLBP-F	Template
5	29.00%	82.33%	96.20%	98.67%	Tem1
				98.70%	Tem3

7.4.3.4. Experiment Four

In this experiment, the Multi-pie database is used to evaluate the 2D-MTLBP-F model. Having a large number of images that belong to a large number of people makes the problem more challenging. In addition to that, the images are not well aligned. The face recognition is repeated 5 times using 8 randomly selected gallery images and the rest 35 images are used as query images. The SRC is used for classification while the images are transformed into the 2D-MTLBP-F model using Tem1 and Tem3. The results of this experiment are shown in table 7.8. This experiment show the efficiency of the 2D-MTLBP-F method with challenging circumstances of large dataset, illuminated, un-registered images and cross database testing.

Table 7.7 Recognition rates for different method on Multi-Pie database.

times	Grey image	TLBP	MSDLE	2D-MTLBP	Template
5	87.33%	89.21	93.57	97.97	Tem1
				98.10	Tem3

7.4.3.5. Experiment Five

In this experiment, 7992 images belonging to 222 subjects from the challenging FRGC database are used to evaluate the 2D-MTLBP-F model. The images are taken in un-controlled environment with illumination and facial expressions. The face recognition is repeated 5 times with 8 randomly selected gallery images and 28 images query images each round. The SRC is used for classification where Tem2 is used for 2D-MTLBP-F representation. As shown in table 7.8, the 2D-MTLBP-F improves the recognition rate by 4.41% over the MSLDE and 9.44% over the TLBP. The 2D-MTLBP-F is robust even in un-controlled environment, large dataset and cross database testing.

Table 7.8 Recognition rates for different method on FRGC database with different illumination conditions for the gallery images

times	Grey image	TLBP	MSDLE	2D-MTLBP-F	Template
5	79.28	85.23	90.26	94.67	Tem2

7.4.3.6. Experiment Six

I order to show the significance of training procedure for2D-MTLB-F, this experiment is conducted on the AR and Curtin databases. The 2D-MTLBP-F is compared with full layers MTLBP which contains 62 channels and with randomly selected 6 layers of MTLBP. The gallery and query images are selected as in table 7.1 where the SRC is used for classification. As shown in table 7.9, the full MTLBP and randomly selected 6 layers of MTLBP do not achieve a high recognition rate. The reason is that the neighbouring layers in full MTLBP are highly correlated which can degrade the recognition rate. Similarly, the random selection of channels may fall into the same trap, in addition to the fact that its chance of achieving the best representation for face recognition is very small.

Table 7.9 Recognition rates achieved by the proposed trained method in comparison with full 62 layers of MTLBP and randomly selected layers from MTLBP

Dataset	Full MTLBP	Random 5 layers of MTLBP	2D-MTLBP-F
AR	92.50%	90.23%	100%
Curtin	81.41%	78.67%	96.26%

7.5. Summary

In this chapter, a multi-channel image representation named as the 2D-MTLBP-F is proposed for face recognition against illumination. This method is based on the Threshold Local Binary Pattern (TLBP). Using a multichannel of TLBP, 62 channels are achieved. As this representation is very high in size and the adjacent channels are highly correlated, a training procedure for face recognition is used in block-wise as in the 2D-MCF model with the difference of using TLBP channels instead of color channels. The proposed 2D-MTLBP-F is evaluated using different 5 databases and the results are very promising. The experiments show the superiority of the 2D-MTLBP-F over the MSLDE method and the TLBP method. The idea is to optimise the MTLBP information to gain the relevant discriminant information. The proposed 2D-MTLBP-F model achieves high recognition rates with different databases, different environments, and different condition. Moreover, the time spent on the training phase can be ignored because we need to conduct the training only once, and then the resultant model can be used with different datasets and environment.

Chapter Eight

Conclusions and future works

This thesis investigates the role of image representation in automatic 2D face recognition. Tackling various challenging problems in automatic face recognition, we proved that effective face image representation is an essential part of successful face recognition system. Starting with the problem of extreme open mouth which deforms the face image registration, we show the importance of recovering the closed mouth case for better face recognition accuracy as different other normalisation methods cannot solve this issue. The main concentration of this thesis is the importance of effective use of multiple channels information. In this regard, we extended the MCF model into the 2D-MCF model showing the importance of the local variances rather than the holistic variances. In addition to that, we integrated the 2D-MCF model with the partitioned SRC for face recognition against occlusion. In addition to the 2D-MCF model, we investigate the role of the weighted multiple color fusion proposing the Multi-resolution Multiple Color Fusion MMCF signifying its preference on both MCF and 2D-MCF, especially when integrated with the SPAE deep network for pose correction and face recognition. Finally, for grey image which contains only one channel, we proved that the use of multi-channel based on the Threshold Local Binary Pattern can expressively enhance the recognition accuracy in the presence of variant illuminations. Overall, we proved that the discriminant information gained by multiple channels cannot be extracted from any single/three channel(s) representation even when using modern approaches such as SRC classifier and deep networks. This finding can be clearly seen through the integration of the multi-channel image representation with these modern approaches where the multi-channel representation plays essential role in the success of the face recognition. This research can be summarised and concluded with some future direction in the following subsections.

8.1. Mouth shape correction

This part can be considered as re-registration of face image as this problem is unique and it has a clear negative effect on face recognition accuracy. We can conclude from chapter three that the correction of mouth shape variation can significantly improve the recognition rate because this correction realigns the face image where the alignment of this case is impossible without this correction. This method utilises the AR model for landmark detection where only mouth landmarks are used. The opened mouth and smiling mouth are corrected to a closed mouth shape by stretching/shrinking the upper and the lower areas of the mouth cavity. Another advantage of such correction is that it is easy to implement and is straight forward procedure, meaning it doesn't need any training steps or training datasets. Even though this correction method employs the training based AR model for landmark detection, such training is performed offline and the obtained model can be used for mouth shape recovery for any frontal face image.

8.2. Two Directional Multiple Color Fusion 2D-MCF

This multi-color is introduced in chapter 4 and its integration for face recognition against occlusion is presented in chapter 5. This model is an extension of the MCF model in block-wise rather than for the holistic face image. As different face areas are different in shape and color, the best representation of each area is gained locally and then the whole face image is reconstructed from different blocks. In this model, each block in the face image has its own representation of multi-color channels based on MCF criteria where each block is trained locally for face recognition employing the greedy search to optimise its color representation for face recognition. This model shows better recognition accuracy compared with the MCF and other three channels based color spaces. In addition to the high accuracy of this model with frontal face images, it succeeded with cross testing dataset when the model is trained on a specific database and then it can be used for face recognition on another database. This makes the use of this model straightforward such that we need to train the model only once, and then it can be used for face recognition for any dataset. In chapter 5 the 2D-MCF model is integrated with the partitioned SRC (P-SRC) for face recognition against occlusion. For this purpose, all gallery and query images (with occlusion) are transformed to the 2D-MCF model and then the P-SRC is used for classification with statistical majority voting. Comparing with the original P-

SRC, reconstruction based method and some other state of the art methods, the integration of 2D-MCF model with P-SRC achieves outstanding improvement in face recognition accuracy by up to 21%. In addition to that, we adopt a reconstruction algorithm based on deep SPAE network and compare the results with our model. The results show the superiority of the 2D-MCF with P-SRC over this algorithm. We believe that the most important factor in this success is the model's dependence on optimal local color representation rather than the holistic color optimization. This local color optimization can improve the local recognition accuracy in different blocks and as a result this improvement will improve the holistic face recognition when using the statistical majority voting.

One of future work suggestion is to replace the greedy criteria by more efficient optimisation method. Another possible improvement can be obtained by using patch based on landmark detection rather than a fixed block size as this can avoid the misalignment problem. For face recognition against occlusions, deep learning technique can be used to reconstruct different color channels, and then the 2D-MCF model can be used for image representation and face recognition.

8.3. Multi-resolution multiple channel representation

In chapter 6, we introduced a new weighted multiple channel representation for face recognition against pose variations named as the Multi-resolution Multiple Color Fusion MMCF. In this model, various number of color channels is used to represent face images, where different color channels can have different resolutions (or sizes). The image size (or resolution) is considered as a weight as different sizes can provide different levels of information. After converting the RGB images to different color channels from different color spaces, each color channel is resized to four different sizes, and then the MCF criteria is used to optimize the number of channels (with selected size) for face recognition. The idea is to give each color channel a specific weight according to its cooperation in face recognition. This new multi-channel representation can improve the recognition accuracy over the MCF and the 2D-MCF as detailed in chapter 7. As the 2D-MCF is based on frontal faces, it cannot be used for face recognition against pose variations. In contrast, the MMCF model acts on the holistic face image rather block-wise. Consequently, we integrate it with the SPAE deep network for pose correction and then for face recognition. Comparing with the original SPAE and SPAE based on RGB images, the integration of SPAE and MMCF achieves higher recognition rate

especially with large pose angles. Moreover, the original SPAE and SPAE for RGB images are completely failed with single gallery image and cross database testing where the SPAE with MMCF achieves an acceptable recognition rate of 82.85% to 92.31% while cross database testing and 61.96% to 79.39% for single gallery image. Again, this new model shows the importance of effective use of multiple channels for face recognition.

For future work, the MMCF can be adopted in two directional for frontal face recognition as the 2D-MCF and again with efficient search criteria rather than the greedy method. Also, the MMCF model can be trained in frequency domain instead of special domain using Discrete Cosine Transform DCT or Discrete Wavelet Transform DWT. Finally, the MMCF model can be utilised for face recognition against occlusions with occlusion detecting and reconstruction. The occlusion is detected and reconstructed for different MMCF color channels, and then the MMCF model is built for efficient face recognition.

8.4. Multiple Channels for Grey Images

In chapter 7, we introduced a new multi-channel image representation for face recognition against illumination variations based on grey images. The proposed model named as the Two Directional Multi-level Threshold Local Binary Pattern Fusion MTLBP-F utilises multiple levels of Threshold Local Binary Pattern TLBP in two directions to produce multiple LBP channels, where the 2D-MCF criteria is used to optimise this model for face recognition. The idea is inspired from the 2D-MCF and the Threshold Local Binary Pattern TLBP. Firstly, the face images are normalised using the DCT then the TLBP is applied on each image with multiple levels of threshold resulting in 61 channels of the TLBP. This large number of channels is optimised for face recognition locally using 2D-MCF criteria resulting in multiple channel model containing 9 layers (channels). The idea is to obtain various levels of discriminant information from different TLBP levels. The optimisation of these TLBP levels results in high performance model for face recognition against illumination variations. The experiments on 5 different face databases show the superiority of the proposed method over the resent MSLDE method and the TLBP achieving 1% to 5% of improvement in recognition accuracy. As expected, the use of multiple channels enriches the face image with invaluable discriminant information which significantly increases the recognition accuracy.

In future, we will investigate another multiple channels based on grey images such as the Logarithmic Total Variation LTV with different direction. Also, a multiple channels based on

holistic face image can be used with weights assigned to different channels as in MMCF in chapter 8. Finally, the greedy search can be replaced by more efficient optimisation methods for more improvement.

Bibliography

Afshar, P., Brown M. and Wang H. (2009). Gradient descent optimisation for ILC-based stochastic distribution control. ICCA 2009. *IEEE International Conference on Control and Automation : 1134-1139.*

Ahonen, T., Hadid A. and Pietikainen M. (2006). Face Description with Local Binary Patterns: Application to Face Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(12): 2037-2041.

Ahonen, T., Hadid A. and Pietikäinen M. (2004). Face Recognition with Local Binary Patterns. *Computer Vision - ECCV 2004: 8th European Conference on Computer Vision*, Prague, Czech Republic, May 11-14, 2004. Proceedings, Part I. T. Pajdla and J. Matas. Berlin, Heidelberg, *Springer* : 469-481.

Aitkenhead, M. J. and McDonald A. J. S. (2003). A neural network face recognition system. *Engineering Applications of Artificial Intelligence* **16**(3): 167-176.

Ali A., Hussain S., Haroon F., Hussain S. and Khan M. (2012). Face Recognition with Local Binary Patterns. *Bahria University Journal of Information & Communication Technology* **5**(1): 46-50.

Alrjebi M. M., Liu W., Li L. (2015). Two Directional Multiple Colour Fusion for Face Recognition. *International Conference on Computers, Communications, and Systems* (ICCCS). Kanyakumari, India, IEEE **:** 171-177.

Alrjebi, M. M., Pathirage. N., Liu W. and Li L. (2017). Face recognition against occlusions via colour fusion using 2D-MCF model and SRC. *Pattern Recognition Letters* **95**: 14-21.

Aroussi, M. E., Amine A., Ghouzali S., Rziza M. and Aboutajdine D. (2008). Combining DCT and LBP Feature Sets For Efficient Face Recognition. *3rd International Conference on Information and Communication Technologies: From Theory to Applications* : 1-6.

Ashraf A. B., Lucey S. and Chen (2008). Learning patch correspondences for improved viewpoint invariant face recognition. *IEEE Conference on Computer Vision and Pattern Recognition :* 1-8.

Asthana A., Marks T. K., Jones M. J., Tieu K. H. and Rohith M. (2011). Fully automatic pose-invariant face recognition via 3D pose normalisation. *International Conference on Computer Vision*, Nov. 2011, pp.937-944**:** 937-944.

Baldi, P. (1995). Gradient descent learning algorithm overview: A general dynamical systems perspective. *IEEE Transactions on Neural Networks* **6**(1): 182-195.

Bengio, Y. (2009). Learning deep architectures for AI. *Foundations and trends in Machine Learning* **2**(1): 1-127.

Biao, W., Weifeng L., Wenming Y. and Qingmin L. (2011). Illumination Normalisation Based on Weber's Law With Application to Face Recognition. *Signal Processing Letters*, IEEE **18**(8): 462-465.

BILSON, A. J. (1987). *Image Size And Resolution In Face Recognition* (Order No. 8713350). Available from ProQuest Dissertations & Theses Global. (303627954).

Biswas S., Aggarwal G., Flynn P. J. and Bowyer K. W. (2013). Pose-Robust Recognition of Low-Resolution Face Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **35**(12): 3037-3049.

Candes E. J., Romberg J. K. and Tao T. (2006). Stable signal recovery from incomplete and inaccurate measurements. *Communications on pure and applied mathematics* **59**(8): 1207-1223.

Candes E. J. and Tao T. (2006). Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE transactions on information theory* **52**(12): 5406-5425.

Castillo C. D., Jacobs D. W. (2011). Wide-baseline stereo for face recognition with large pose variation, *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 537-544.

Chan C. H., Tahir M. A., Kittler J. and Pietikäinen M. (2013). Multiscale Local Phase Quantization for Robust Component-Based Face Recognition Using Kernel Fusion of Multiple Descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35**(5): 1164-1177.

Chen H., Tang Y. Y., Fang B. and Zhang T. (2010). Wavelet-Based Multi-Level Image Matching with Detail Measure Weight for Face Recognition under Varying Illumination. *Chinese Conference on Pattern Recognition* (CCPR).

Chen J., Shan S., He C., Zhao G., Pietikainen M., Chen X. and Gao W. (2010). WLD: A Robust Local Image Descriptor. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **32**(9): 1705-1720.

Chen T., Wotao Y., Xiang Z., Comaniciu D. and Huang T. S. (2006). Total variation models for variable lighting face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(9): 1519-1524.

Chen W., Meng E. and Shiqian W. (2006). Illumination compensation and normalisation for robust face recognition using discrete cosine transform in logarithm domain. *IEEE Transactions on Systems*, Man, and Cybernetics, Part B (Cybernetics) **36**(2): 458-466.

Chengjun, L. (2008). Learning the Uncorrelated, Independent, and Discriminating Color Spaces for Face Recognition. *IEEE Transactions on Information Forensics and Security*,**3**(2): 213-222.

Chengjun L. and Jian Y. (2009). ICA Color Space for Pattern Recognition. *IEEE Transactions on Neural Networks*,**20**(2): 248-257.

Clark, R. (1994). Destined to be the definitive text. *Computer Communications* **17**(11): 815-816.

Comon, P. (1994). Independent component analysis, A new concept? *Signal Processing* **36**(3): 287-314.

Cootes T. F., Edwards G. J. and Taylor C. J. (2001). Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*,**23**(6): 681-685.

Cormen T. H. and ebrary I. (2009). Introduction to algorithms / Thomas H. Cormen and I. ebrary. Cambridge, Mass., Cambridge, Mass. : *MIT Press*.

Costa M., Palmisano D. and Pasero E. (2000). Gradient descent in feed-forward networks with binary neurons. 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS *International Joint Conference on Neural Network* 1: 311-316.

Dabbaghchian S., Ghaemmaghami M. P. and Aghagolzadeh A. (2010). Feature extraction using discrete cosine transform and discrimination power analysis with a face recognition technology. *Pattern Recognition* **43**(4): 1431-1440.

Dasari H., Srinivasa R. and Chittipotula S. (2012). Performance Evaluation on the Effect of Combining DCT and LBP on Face Recognition System. *International Journal of Modern Education and Computer Science* **4**(11): 21-32.

Deng, L. (2014). A tutorial survey of architectures, algorithms, and applications for deep learning. *APSIPA Transactions on Signal and Information Processing* **3.**

Deng, L. and Yu D. (2014). Deep Learning: Methods and Applications. *Foundations and Trends® in Signal Processing* **7**(3–4): 197-387.

Donoho, D. L. (2006). For most large underdetermined systems of linear equations the minimal ?? 1 -norm solution is also the sparsest solution *Communications on pure and applied mathematics* **59**(6): 797-829.

Etemad K. and Chellappa R. (1997). Discriminant analysis for recognition of human face images. *Lecture Notes in Computer Science.* **1206:** 125-142.

Faraji M. R. and Qi X. (2015). Face recognition under varying illumination based on adaptive homomorphic eight local directional patterns. *IET Computer Vision* **9**(3): 390-399.

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

Gonzalez R. C., Woods R. E. and Eddins S. L. (2009). Digital Image processing using MATLAB / Rafael C. Gonzalez, Richard E. Woods, Steven L. Eddins. S.I.], [S.I.] : Gatesmark Pub.

Gritzman A., Rubin D. and Pantanowitz A. (2015). Comparison of colour transforms used in lip segmentation algorithms. *Signal, Image and Video Processing* **9**(4): 947-957.

Gross R., Matthews I. and Baker S. (2002). Eigen light-fields and face recognition across pose. *Fifth IEEE International Conference on Automatic Face Gesture Recognition*. USA**:** 3-9.

Gross R., Matthews I., Cohn J., kanada T. and Baker S. (2007). The cmu multi-pose, illumination, and expression (multipie) face database, *Carnegie Mellon University Robotics Institute* TR-07-08, Tech. Rep.

Gross, R., I. Matthews, J. Cohn, T. Kanade and S. Baker (2008). Multi-PIE**:** 1-8.

Haddadnia, J. and M. Ahmadi (2004). N-feature neural network human face recognition. *Image and Vision Computing* **22**(12): 1071-1082.

Hadid, A., M. Pietikainen and T. Ahonen (2004). A discriminative feature space for detecting and recognizing faces. Proceedings of the 2004 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (CVPR) **2**.

Hancock, P. (2004). The psychological image collection at stirling (pics). *University of Stirling.*

Hecht-Nielsen, R. (1988). Theory of the backpropagation neural network. *Neural Networks* **1**(Supplement-1): 445-448.

Huang D., Shan C., Ardabilian M., Wang Y. and Chen L. (2011). Local Binary Patterns and Its Application to Facial Image Analysis: A Survey. *IEEE Transactions on Systems*, Man, and Cybernetics, Part C (Applications and Reviews) **41**(6): 765-781.

Huang G. B., Ramesh M., Berg T.and Learned-Miller E. (2007). Labeled faces in the wild: A database for studying face recognition in unconstrained environments, Technical Report 07-49, *University of Massachusetts,* Amherst.

Jain, A. (1976). A Fast Karhunen-Loeve Transform for a Class of Random Processes. *IEEE Transactions on Communications* **24**(9): 1023-1029.

Jian Y. and Chengjun L. (2008). A discriminant color space method for face representation and verification on a large-scale database. *19th International Conference on Pattern Recognition* : 1-4.

Jing X., Liu Q., Lan C., Man J., Li S. and Zhang D. (2010). Holistic orthogonal analysis of discriminant transforms for color face recognition. *IEEE International Conference on Image Processing* :3841-3844.

Kan M., Shan S., Chang H. and Chen X. (2014). Stacked Progressive Auto-Encoders (SPAE) for Face Recognition Across Poses. *IEEE Conference on Computer Vision and Pattern Recognition* : 1883-1890.

Klein S., Pluim J. P., Staring M. and Viergever M. A. (2009). Adaptive stochastic gradient descent optimisation for image registration. *International journal of computer vision* **81**(3): 227.

Lai Z. R., Dai D. Q., C. X. Ren and Huang K. K. (2015). Multiscale logarithm difference edgemaps for face recognition against varying lighting conditions. *IEEE transactions on image processing* : a publication of the IEEE Signal Processing Society **24**(6): 1735.

Lawrence S., Giles C. L., Chung T. and Back A. D. (1997). Face recognition: a convolutional neural-network approach. *IEEE Transactions on Neural Networks,* **8**(1): 98-113.

Lee H. S. and Kim D. (2008). Expression-invariant face recognition by facial expression transformations. *Pattern Recognition Letters* **29**(13): 1797-1805.

Li B., An S., Liu W. and Krishna A. (2011). The MCF Model: Utilizing Multiple Colours for Face Recognition. *Sixth International Conference on Image and Graphics* :1029-1034.

Li B., Liu W. Q., An S. and Krishna A. (2014). Robust face recognition by utilizing colour information and sparse representation. *International Journal of Pattern Recognition and Artificial Intelligence* 28 (3): 1456004-1-1456004-27.

Li B. Y., Liu W., An S., Krishna A. and Xu T. (2012). Face recognition using various scales of discriminant color space transform. *Neurocomputing* **94**: 68-76.

Liang, A., Liu W., Li L., Farid M. R. and Le V. (2014). Accurate Facial Landmarks Detection for Frontal Faces with Extended Tree-Structured Models. 22nd *International Conference on Pattern Recognition* : 538-543.

Yin L., Chen, X., Sun, Y., Worm, T., and Reale, M. (2008). A high-resolution 3D dynamic facial expression database. *8th IEEE International Conference on Automatic Face & Gesture Recognition* : 1-6.

Liu C. J. (2008). Learning the uncorrelated, independent, and discriminating color spaces for face recognition. *IEEE Trans. Inf. Forensic Secur*. **3**(2): 213-222.

Lu J., Plataniotis K. N. and Venetsanopoulos A. N. (2005). Regularization studies of linear discriminant analysis in small sample size scenarios with application to face recognition. *Pattern Recognition Letters* **26**(2): 181-191.

Luan X., Fang B., Liu L., Yang W. and Qian J. (2013). Extracting sparse error of robust PCA for face recognition in the presence of varying illumination and occlusion. *Pattern Recognition* **47**(2): 495-508.

Martinez A. M. (1998). The AR face database. CVC Technical Report **24**.

Meng J., Gao Y., Wang X., Lin T. and Zhang J. (2010). Face Recognition Based on Local Binary Patterns with Threshold. *IEEE International Conference on Granular Computing* **:** 352-356.

Mery D. and Bowyer K. (2014). Face recognition via adaptive sparse representations of random patches**.** *IEEE International Workshop on Information Forensics and Security* **:** 13-18.

Micheli-Tzanakou E. (2011). Artificial Neural Networks: An Overview. Network: *Computation in Neural Systems* 22(1-4) : 208-230.

Ming Z. and Fulcher J. (1996). Face recognition using artificial neural network group-based adaptive tolerance (GAT) trees. *IEEE Transactions on Neural Networks* **7**(3): 555-567.

Morelli A., Padovani S., Tepper M. and Berlles J. (2014). Face recognition on partially occluded images using compressed sensing. *Pattern Recognition Letters* **36**: 235-242.

Mpiperis I., Malassiotis S. and Strintzis M. G. (2007). 3-D Face Recognition With the Geodesic Polar Representation. *IEEE Transactions on Information Forensics and Security*, **2**(3): 537-547.

Nan D. and Xu Z. (2013). A Multi-resolution Based Adaptive blocking and Sparse Representation Approach for Face Recognition under Occlusion. *International Journal of Advancements in Computing Technology* **5**(5): 813-813.

Nefian A. V. (2007). Georgia Tech face database; http://www.anefian.com/research/face_reco.htm.

Ohta Y. (1985). Knowledge-based interpretation of outdoor natural color scenes: . *Computer Vision, Graphics and Image Processing*, **31**(0734-189X): 392-392.

Osuna E., Freund R. and Girosi F. (1997). Training support vector machines: An application to face detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **:** 130-136.

Ou W., You X., Tao D., Zhang P., Tang Y. and Zhu Z. (2013). Robust face recognition via occlusion dictionary learning. *Pattern Recognition* **47**(4): 1559-1572.

Pathirage C. S. N., Li L., Liu W. and Zhang M. (2015). Stacked Face De-Noising Auto Encoders for Expression-Robust Face Recognition. *International Conference on Digital Image Computing: Techniques and Applications (DICTA)* : 1-8.

Perez C. A. and Castillo L. E. (2009). Illumination compensation for face recognition by genetic optimization of the Self-Quotient Image method. *ISOT 2009 - International Symposium on Optomechatronic Technologies* : 322-327 322-327.

Phillips P. J., Flynn P. J., Scruggs T., Bowyer K. W., Jin C., Hoffman K., Marques J., Jaesik M. and Worek W. (2005). Overview of the face recognition grand challenge. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 1**:** 947-954.

Phillips P. J., Moon H., Rizvi S. A. and Rauss P. J. (2000). The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on pattern analysis and machine intelligence* **22**(10): 1090-1104.

PietikãInen, M. (2010). Local Binary Patterns. *Scholarpedia* **5**(3): 9775.

Priya, G. and R. Wahida Banu (2014). Occlusion invariant face recognition using mean based weight matrix and support vector machine. *Sadhana* 39(2):303-315.

Qian J., Luo L., Yang J., Zhang F. and Lin Z. (2015). Robust nuclear norm regularized regression for face recognition with occlusion. *Pattern Recognition* 48(10) : 3145-3159.

Ramachandran M., Zhou S. K., Jhalani D. and Chellappa R. (2005). A method for converting a smiling face to a neutral face with applications to face recognition. *IEEE International Conference on Acoustics, Speech, and Signal Processing* 2 : .ii/977-ii/980.

Rivera A., Castillo J. and Chae J. (2013). Local Directional Number Pattern for Face Analysis: Face and Expression Recognition. *IEEE Transactions on Image Processing* **22**(5): 1740-1752.

Rao, K. R. (1990). Discrete cosine transform : algorithms, advantages, applications / K.R. Rao, P. Yip. Boston, Boston : *Academic Press*.

Rydfalk, M. (1987). CANDIDE, a parameterized face. Sweeden, *Consulting Psychologists Press Inc*.

Savran A., Alyüz N., Dibeklioğlu H., Çeliktutan O., Gökberk B., Sankur B. and Akarun L. (2008). Bosphorus database for 3D face analysis. *European Workshop on Biometrics and Identity Management, Springer* : 47-56.

Schmidhuber J. (2015). Deep learning in neural networks: An overview. *Neural Networks* **61**: 85-117.

Sharma A. and Jacobs D. W. (2011). Bypassing synthesis: PLS for face recognition with pose, low-resolution and sketch. *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)* : 593-600.

Shashua A. and Raviv T. (2001). The quotient image: class-based re-rendering and recognition with varying illuminations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23**(2): 129-139.

Shiguang S., Wen G., Bo C. and Debin Z. (2003). Illumination normalisation for robust face recognition against varying lighting conditions. 2003 *IEEE International SOI Conference. Proceedings* (Cat. No.03CH37443).

Sun S., Pan Q. and Li J. (2008). Research on Double-weight Face Recognition Model. 2008 *Third International Conference on Pervasive Computing and Applications* **2**: 724-728.

Sun Y., Wang X. and Tang X. (2014). Deep Learning Face Representation by Joint Identification-Verification. *Advances in Neural Information Processing Systems* **3**: 1988-1996.

Sun Y., Wang X. and Tang X. (2014). Deep Learning Face Representation from Predicting 10,000 Classes. *IEEE Conference on Computer Vision and Pattern Recognition* : 1891-1898.

Sun Z. L., Lam K M., Dong Z. Y., Wang H., Gao Q. W. and Zheng C. H. (2013). Face recognition with multi-resolution spectral feature images. *PloS one* **8**(2): e55700.

Taheri S., Patel V. M. and Chellappa R. (2013). Component-Based Recognition of Facesand Facial Expressions. *IEEE Transactions on Affective Computing*, **4**(4): 360-371.

Taiping Z., Yan T., Bin F., Zhaowei S. and Xiaoyu L. (2009). Face Recognition Under Varying Illumination Using Gradientfaces. *IEEE Transactions on Image Processing* **18**(11): 2599-2606.

Theodoridis S. (2008). Pattern Recognition. Burlington, Burlington : *Elsevier Science.*

Torres L., Reutter J. Y. and Lorente L. (1999). The importance of the color information in face recognition. Proceedings 1999 *International Conference on Image Processing* (ICIP)3: 627-631.

Turk M. and Pentland A. (1991). Eigenfaces for Recognition. *Journal of Cognitive Neuroscience* **3**(1): 71-86.

Wang H., Li S. Z. and Wang Y. (2004). Face recognition under varying lighting conditions using self quotient image. *Sixth IEEE International Conference on Automatic Face and Gesture Recognition* : 819-824.

Weeks A. R. (1996). Fundamentals of electronic image processing / Arthur R. Weeks, Jr. Bellingham, Wash. : New York, Bellingham, Wash. (1000 20th St. Bellingham WA 98225-6705 USA) : SPIE.

Wong Y., Chen S., Mau S., Sanderson C. and Lovell B. C. (2011). Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition. Workshops (CVPRW), *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* : 74-81.

Wright J., Yang A. Y., Ganesh A., Sastry S. S. and Yi Ma S. S. (2009). Robust Face Recognition via Sparse Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **31**(2): 210-227.

Xiaohua X., Wei-Shi Z., Jianhuang L. and Yuen P. C. (2008). Face illumination normalisation on large and small scale features. 2008 *IEEE Conference on Computer Vision and Pattern Recognition :* 1-8.

Yang B. and Chen S. (2013). A comparative study on local binary pattern (LBP) based face recognition: LBP histogram versus LBP image. *Neurocomputing* **120**: 365-379.

Yang J. and Liu C. (2008). Color Image Discriminant Models and Algorithms for Face Recognition. *IEEE Transactions on Neural Networks,* **19**(12): 2088-2098.

Yang J., Liu C. and Yang J. y. (2010). What kind of color spaces is suitable for color face recognition? *Neurocomputing* **73**(10): 2140-2146.

Yang J., Liu C. and Zhang L. (2010). Color space normalisation: Enhancing the discriminating power of color spaces for face recognition. *Pattern Recognition* **43**(4): 1454-1466.

Ye J., Janardan R. and Li Q. (2005). Two-Dimensional Linear Discriminant Analysis. *Neural Information Processing Systems* : 1569–1576.

Yip A. W. and Sinha P. (2002). Contribution of color to face recognition. *Perception* **31**(8): 995-1003.

Yoo S., Park R. H., Sim D.G.and Dong N. g. (2007). Investigation of Color Spaces for Face Recognition. *Conference on Machine Vision Applications (MVA)* : 106-109.

Yule D. and Chen L. (2011). Face recognition through regional weight estimation**.** *18th IEEE International Conference on Image Processing* **:** 1761-1764.

Zhao S. and Hu Z. P. (2015). A modular weighted sparse representation based on Fisher discriminant and sparse residual for face recognition with occlusion. *Information Processing Letters* **115**(9): 677-683.

Zhong F. and Zhang J. (2013). Face recognition with enhanced local directional patterns. *Neurocomputing* **119**: 375-384.

Zhou L., Liu W., Lu Z. and Nie T. (2014). Face recognition based on curvelets and local binary pattern features via using local property preservation. *Journal of Systems and Software* 95 : .209-216.

Zhu X. and Ramanan D. (2012). Face detection, pose estimation, and landmark localization in the wild. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) :* 2879-2886.

Zhu Z., Luo P., Wang X. and Tang X. (2013). Deep Learning Identity-Preserving Face Space. 2013 *IEEE International Conference on Computer Vision* 113-120.

Zhuo T. (2016). Face recognition from a single image per person using deep architecture neural networks. *Cluster Computing* **19**(1): 73-77.