TOWARDS PRACTICAL FACE RECOGNITION SYSTEM EMPLOYING ROW-BASED DISTANCE METHOD IN 2DPCA BASED ALGORITHMS

WALED HUSSEIN MOHAMMED AL-ARASHI

UNIVERSITI SAINS MALAYSIA

2014

TOWARDS PRACTICAL FACE RECOGNITION SYSTEM EMPLOYING ROW-BASED DISTANCE METHOD IN 2DPCA BASED ALGORITHMS

by

WALED HUSSEIN MOHAMMED AL-ARASHI

Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy

February 2014

ACKNOWLEDGEMENTS

All praise and glory be to Allah, for giving me the opportunity to pursue the PhD research at Universiti Sains Malaysia. Then, I would like to express my gratitude to my supervisor, Assoc. Prof. Dr. Shahrel Azmin Suandi, for his continuous support and guidance in my research work. This thesis would never have been written the way it is without his generous help and support.

I wish to thank my co-supervisor Dr. Haidi Ibrahim, for his assistance in this work. I am also very grateful to the staff and faculty of the School of Electrical and Electronic Engineering who have provided assistance during my study.

I take this special occasion to give my greatest gratitude to my father, my mother and my family for their support and to wish them best. I thank my wife for her patient, her encouragement, and her love filled with tenderness.

I would like to express my appreciation to the wonderful Malaysian people who make me feel that I am in my home country.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

2d2DPCA Two directional Two-dimensional Principal Component Analysis

2DPCA Two-dimensional Principal Component Analysis

ABC Artificial Bee Colony

AMD Assembled matrix distance

AI Artificial Intelligent

COA Chaos Optimization Algorithm

CS-LBP Center Symmetric Local Binary Pattern

CMC Cumulative Match Characteristic

D-LBP Direction Local Binary Pattern

DiaPCA Diagonal Principal Component Analysis

EER Equal Error Rate

EPC^{2}A Extended Projection-combined Principal Component Analysis

FAR False Acceptance Rate

FLD Fisher Linear Discriminant

FP False Positive

GRowkNN General Row *k* nearest neighbor

GA Genetic Algorithm

ICA Independent Component Analysis

ICGA Improved Chaos Genetic Algorithm

ID-LBP Improved Direction Local Binary Pattern

ILBP Improved Local Binary Pattern

l_1 **NNCH** l_1-norm Nearest Neighbor Convex Hull

LBP Local Binary Pattern

LDA Linear Discriminant Analysis

LDP Local Derivative Pattern

LFW Labeled Faces in the Wild

LLBP Local Line Binary Pattern

LTP Local Ternary Patterns

MBLBP Multi-scale Block Local Binary Pattern

MPCA modular Principal Component Analysis

MRTD Machine Readable Travel Documents

MSE Mean Square Error

NN Nearest Neighbor

NNCH Nearest Neighbor Convex Hull

ORL Olivetti Research Laboratory

PC^{2}A Projection-combined Principal Component Analysis

PCA Principal Component Analysis

PCA-L1 Principal Component Analysis based on l_1-norm maximization

PIE Pose, Illumination, and Expression

R1-PCA Rotational Invariant l_1-norm Principal Component Analysis

ROC Receiver Operating Characteristic

RowAMD Row assembled matrix distance

RowkNN Row k nearest neighbor

SCface Surveillance Cameras Face Database

SPCA Symmetrical Principal Component Analysis

SpPCA Sub-pattern Principal Component Analysis

TP True Positive

VR Verification Rate

LIST OF SYMBOLS

$\mathbf{A}$ Random matrix

$\bar{\mathbf{A}}$ Average image

$\mathbf{A}_i$ The i^{th} training face image

$A_i^{(c)}$ The c^{th} column vectors of $\mathbf{A}_i$

$\bar{A}_i^{(c)}$ The c^{th} column vectors of $\bar{\mathbf{A}}_\mathbf{i}$

$A_i^{(r)}$ The r^{th} row vectors of $\mathbf{A}_i$

$\bar{A}_i^{(r)}$ The r^{th} row vectors of $\bar{\mathbf{A}}_\mathbf{i}$

c Number of classes

$\mathbf{C}$ Covariance Matrix

$\mathbf{D}$ Alternative covariance Matrix

d The number of columns of a feature matrix

d_{AMD} Assembled matrix distance

d_c Column Euclidian distance

d_e 2D Euclidian distance

$d_{p,q}$ p-q distance

d_v Volume distance

E Expected value

$\mathbf{e}_i$ The i^{th} eigenvector of the alternative covariance Matrix

$\mathbf{G}_c$ Covariance matrix of alternative 2DPCA

$\mathbf{G}_r$ Covariance matrix of 2DPCA

$\mathbf{G}_t$ So-called covariance matrix

k Number of selected eigenvectors

m Number of rows of a matrix

M Number of training face images

n Number of columns of a matrix

p Control variable

q Control variable

$\mathbf{S}$ Independent component variables

$\mathbf{S}_b$ Between-class matrix

$\mathbf{S}_w$ Within-class matrix

$\mathbf{S}_x$ Covariance matrix of the projected feature vectors of the training images

tr Trace

$\mathbf{V}$ Selected eigenvectors matrix

$\mathbf{v}_i$ The i^{th} eigenvector

$\mathbf{W}$ Selected eigenvectors matrix of FLD subspace

$\boldsymbol{W}$ Mixing matrix

$\boldsymbol{X}$ Set of face images

$\mathbf{X}$ n-dimensional unitary column vector

$\mathbf{X_{opt}}$ Selected eigenvectors of the so-called covariance matrix $\mathbf{G}_t$

$\bar{\boldsymbol{x}}$ Average face image vector

$\boldsymbol{x}_i$ Vector of face image

$\bar{\boldsymbol{x}}_j$ Mean vector of class j

$\boldsymbol{y}_i$ Feature vector

$\mathbf{Y}$ Projected feature vector

y_c^i Specific column vector of a train feature matrix

y_c^t Specific column vector of a test feature matrix

$\mathbf{Y}_i$ Feature matrix of a train image

$\mathbf{Y}_t$ Feature matrix of a test image

$\mathbf{Z}$ Covariance matrix of alternative 2DPCA

λ_i The i^{th} eigenvalue

μ_i The i^{th} eigenvalue of the alternative covariance Matrix

$d_{p,q}$ p-q distance

q Control variable

KE ARAH SISTEM PENGECAMAN MUKA PRAKTIKAL MENGGUNAKAN KAEDAH JARAK BERASASKAN BARIS DALAM ALGORITMA BERASASKAN 2DPCA

ABSTRAK

Pengecaman muka secara automatik telah menjadi satu topik fokus penyelidikan dalam beberapa dekad ini. Ini adalah kerana kelebihan pengecaman muka dan potensi permintaan kepada kawalan keselamatan yang tinggi dalam aplikasi komersil dan penguatkuasaan undang-undang. Walaubagaimanapun, disebabkan sifat semulajadi muka, ianya tertakluk kepada beberapa variasi. Sehubungan itu, pencarian satu sistem pengecaman muka yang baik masih lagi menjadi satu topik penyelidikan yang aktif sehingga ke hari ini. Banyak kaedah telah dicadangkan untuk mengatasi masalah variasi muka ini. Di antara kaedah-kaedah ini, kaedah subruang adalah antara kaedah yang paling popular dan berkesan. "Eigenface" atau kaedah Analisa Komponen Prinsipal (PCA) adalah merupakan satu kaedah yang dianggap sebagai salah satu teknik yang paling berjaya dalam kaedah subruang ini. Salah satu daripada lanjutan PCA yang paling penting adalah Dua Dimensi PCA (2DPCA). Walaubagaimanapun, ciri-ciri 2DPCA adalah berasaskan matriks dan bukannya vektor seperti dalam PCA. Oleh itu, kaedah pengiraan jarak yang berbeza telah dicadangkan untuk mengira jarak antara ciri matriks ujian dan ciri matriks latihan. Semua kaedah terdahulu menangani masalah klasifikasi adalah secara matematik tanpa mempertimbangkan ciri matriks dan imej muka. Selain itu, prestasi sistem dalam aplikasi praktikal bergantung kepada bilangan "eigen" yang dipilih. Sebagai penyelesaian kepada isu-isu yang dinyatakan

di atas, empat kaedah jarak baharu telah dicadangkan dalam tesis ini, yang berasaskan kepada baris matriks ciri bagi algoritma 2DPCA. Melalui eksperimen menggunakan lapan pangkalan data muka, peningkatan mereka berbanding dengan kaedah jarak terdahulu telah ditunjukkan. Sebagai tambah, kaedah berasaskan tekstur sebagai langkah pra-pemprosesan juga dicari dan ia digunakan bagi menangani kesan perubahan pencahayaan. Selain itu, satu lagi isu aplikasi praktikal, berkaitan pemilihan data latihan teroptimum juga telah diselesaikan. Kaedah ini menyelesaikan persoalan berapa banyak data dari galeri yang patut dimasukkan dalam peringkat latihan bagi mendapatkan keputusan klasifikasi terbaik dengan menggabungkan Algoritma Genetik (GA) dengan PCA. Dengan menggunakan tiga pangkalan data muka, Keputusan mendedahkan yang kaedah kajian ini mempunyai prestasi lebih tinggi danipada prestasi PCA dari segi ketepatan dan masa klasifikasi.

TOWARDS PRACTICAL FACE RECOGNITION SYSTEM EMPLOYING ROW-BASED DISTANCE METHOD IN 2DPCA BASED ALGORITHMS

ABSTRACT

Automatic face recognition has been a focus research topic in past few decades. This is due to the advantages of face recognition and the potential need for high security in commercial and law enforcement applications. However, due to nature of the face, it is subjected to several variations. Thus, finding a good face recognition system is still an active research field till today. Many approaches have been proposed to overcome the face variations. In the midst of these techniques, subspace methods are considered the most popular and powerful techniques. Among them, eigenface or Principal Component Analysis (PCA) method is considered as one of the most successful techniques in subspace methods. One of the most important extensions of PCA is Two-dimensional PCA (2DPCA). However, 2DPCA-based features are matrices rather than vectors as in PCA. Hence, different distance computation methods have been proposed to calculate the distance between the test feature matrix and the training feature matrices. All previous methods deal with the classification problem mathematically without any consideration between feature matrices and the face images. Besides, the system performance in practical applications relies on the number of eigenvectors chosen. As a solution to the above mentioned issues, four new distance methods have been proposed in this thesis, which are based on the rows of a feature matrix of 2DPCA-based algorithms. Through experiments, using eight face databases, their improvements compared to the previous distance methods are demonstrated. In addition, texture-based methods as a preprocess step are also investigated and used to suppress the effect of illumination variations. On top of this, another practical application issue, which is related to selecting an optimized training data, has also been solved. This method solves the question of how much data from the gallery that should be included in the training stage to achieve better classification results by associating Genetic Algorithm (GA) to PCA. Experimental results using three face databases reveal that the proposed method outperforms PCA in terms of accuracy and classification time.

CHAPTER 1

INTRODUCTION

1.1 Introduction

The problem of face recognition is investigated in this thesis. This chapter introduces the problem considered in this thesis and thesis objectives. The contributions and the organization of the thesis are then described.

1.2 The Face Recognition Problem

Biometric technologies in the current era society have become a vital side in their lives because of increasing security demands. Among these technologies, face recognition has several advantages. It is user-friendly, its facial features have got the highest compatibility scores among the other biometric technologies (MRTD, 2010) and harmonious with human visual perception. In fact, one of the natural capabilities of humans is remembering and recognizing an enormous number of faces during their lifetime without a considerable effort. Due to this capability, many researchers in different areas, such as psychophysicists, neuroscientists and engineers, have studied diverse characteristics on human and machine face recognition.

One of the important characteristics of human perception that may lead to a good machine face recognition system, which has been investigated, is whether the face recognition is done globally or locally (Zhao et al., 2003). Besides, the issue whether the facial features have the same significance in recognition perception or not has also been studied. Some experiments demonstrate that identifying familiar faces with the existence of various lighting conditions is a very hard task (Johnston et al., 1992).

Due to the advantages of face recognition and the potential need to a high security in commercial and law enforcement applications, automatic face recognition has been a research topic in past few decades (Zhao et al., 2003) and many approaches have been proposed to overcome the face variations. However, due to the natural of the face as the three-dimensional object, it is a subject to several variations. Thus, finding a good face recognition system is still an active research field till today. The main sources of difficulties in automatic face recognition are the pose and illumination variations. Another difficulty that usually encounter in practice is whether all the available images should be included in the training stage or maybe not the foremost choice for building an effective face recognition system (Martínez and Kak, 2001).

The development of face recognition systems in the state-of-the-arts shows that subspace methods are considered the most popular and powerful techniques (Jafri and Arabnia, 2009; Lu, 2003; Zhao and Chellappa, 2006; Li and Jain, 2004; Rao and Noushath, 2010). This is due to the strength of their mathematical model, which reveals the underlying discriminative data, solves the curse-of-dimensionality problem, reduces the system's memory and computational requirements (Li and Jain, 2004; Rao and Noushath, 2010; Zuo et al., 2009). Among these approaches, eigenface or Principal Component Analysis (PCA) method (Turk and Pentland, 1991; Kirby and Sirovich, 1990) is considered as one of the most successful techniques in subspace methods (Turk, 2001; Grudin, 2000; Pentland, 2000; Kim et al., 2002; Moon and Phillips, 2001; Tjahyadi et al., 2006; Meng and Ke, 2008; Gupta et al., 2010; Radha and Pushpalatha, 2010; Zeng et al., 2011), because of the ease of implementation, its reasonable performance level (Moon and Phillips, 2001; Tjahyadi et al., 2006), effectiveness in large databases (Zhao and Chellappa, 2006) and it is less sensitive to different training data set (Jafri and Arabnia, 2009).

Due to the advantages of subspace and PCA-based methods mentioned above and based on the state-of-the-arts related to the face recognition, the research presented in this thesis is highly motivated to concentrate on the techniques based on the PCA coupled with face recognition problems. PCA goal is to derive a lower dimensional subspace from a set of training face images so that it maximizes data variance (Turk and Pentland, 1991). Thus, the subspace that is obtained from PCA carry out most of the data information in small space.

Since its initial proposal in face recognition, many researchers have focused on enhancements of PCA and have different directions. One of the most important extensions of PCA is Two-dimensional PCA (2DPCA) (Yang et al., 2004) which changes the way of PCA calculation and deals with original images matrices directly without any conversion as in PCA. However, 2DPCA-based features are matrices rather than vectors as in the PCA. Hence, different distance computation methods have been proposed to calculate the distance between the test feature matrix and the training feature matrices. However, all previous methods deal with the classification problem mathematically without any consideration between feature matrices and the face images. Despite of all previous issues, the number of coefficients of the feature matrix relies on the number of eigenvectors chosen. These number of eigenvectors have a significant influence on the performance of the algorithm classification (Rao and Noushath, 2010), which affects on classification accuracy, time and storage capacity. Furthermore, it is difficult and laborious calculations are needed to classify an image into the correct class with varying number of eigenvectors in practical applications (Rao and Noushath, 2010). Instead, automatic methods, which determine the best number of eigenvectors, are used (Draper et al., 2002). Accordingly, a robust classification method, which is not influenced too much with selecting an imprecise number of eigenvectors, is essential to keep the accuracy in an acceptable value compared with the maximum accuracy. Nonetheless, more attention to these feature matrices must be spent when classification methods are considered with, as well as the relationship between the 2DPCA based methods and features extracted.

As a solution to the above mentioned problems, sequence of new robust classification methods based on the rows of the feature matrix of 2DPCA-based techniques have been proposed. The successive experiments show clearly their enhancement compared with the previous classification methods regarding the issues mentioned above. In addition, the relationship between the 2DPCA-based methods and their features is explored and further investigated.

Another issue related to PCA-based techniques in practice is that the distribution of training images in the image space is unexpected and the underlying distribution of different classes is not known in advance (Martínez and Kak, 2001). Therefore, the available training data may be unsuitable or may be not the foremost choice for building an effective face-recognition system. Thus, finding the most suitable training data from the available ones is demanded in practical face recognition applications which may lead to the enhancement of their performance in terms of classification time and accuracy.

To handle this matter, a new method that associates Genetic Algorithm (GA) to PCA has been proposed to search the most suitable training data from the available ones. By using GA combined with PCA, the best underlying distribution for classification can be determined that enhance the performance of PCA in terms of classification time and accuracy.

1.3 Thesis Objectives

Based on the aforesaid problems in section 1.2, this research investigates the difficulties of practical face recognition applications with the presence of several face recognition variations. The aim of this work is to improve the existence face recognition algorithms in terms of recognition rate and classification times. The leading objectives of this thesis are:

1. To investigate existing 2DPCA-based face recognition algorithms with the presence of various face variations related to practical face recognition applications.
2. To propose a robust classification method on 2DPCA-based techniques for face recognition.
3. To optimize the training process which finds the most suitable training data from the available ones to build a better face recognition system.

1.4 Thesis Contributions

The contributions of the thesis can generally be categorized into two main aspects and summarized as follows:

1. Four new distance methods in 2DPCA-based algorithms have been proposed as follows:

    i. Row distance method is introduced in 2DPCA-based algorithms, which is based on the rows of a feature matrix of 2DPCA-based algorithms. It takes the advantages of the multiplication of the feature matrix of a 2DPCA-based method and face properties. It distinctly demonstrates the stability of the Row distance accuracies with different numbers of eigenvectors compared with other distance methods. This stability has a great influence on the performance of classification methods in practical applications.

    ii. RowAMD distance method is an extension to the Row distance method with a control variable p. It shows a clear improvement compared with the Row distance method along with different 2DPCA-based algorithms and different databases with different face variations and problems.

    iii. RowkNN distance method is another extension to the Row distance method with 2DPCA algorithm. It has a slight improvement compared with Row distance method when the number of training images is increased or a well alignment method is used.

    iv. GRowkNN distance method is the general form of the RowkNN distance method with two control variables p and q. A precise and appropriate selection of these two control variables will result in better result.

2. A new method that associates Genetic Algorithm (GA) to PCA-based algorithm has been proposed to search the most suitable training data for classification from the available ones. It solves a practical application issue, which is the difficulty of ascertaining whether or not the available training data is appropriate for the recognition system. In addition, it enhances the performance of PCA in terms of accuracy and classification times.

1.5 Scope of thesis

The work in this thesis have been conducted within the following scope:

1. Face images used in the experiments are only 2D images.
2. The experiments focus only in frontal pose within approximately $\pm 45^\circ$.
3. All experiments are conducted using bench mark databases and there is no self developed database used in the work.
4. The thesis only considers PCA-based and 2DPCA-based algorithms.

1.6 Outline of the Thesis

The remainder contents of this thesis are organized into four chapters as outlined below:

Chapter 2 comprises of four main sections. The first section generally gives a background information about face recognition techniques associated with their pros and cons. The concept of subspace methods and the mathematical background are then explored in the second section. More details about PCA, which are related to this work directly, are also given in the same section. 2DPCA-based algorithms are described in detail. Besides, the distance computation methods, which are proposed in the state-of-the-arts related to 2DPCA-based methods, are explained with their pros and cons. Then, methods to evaluate face recognition systems are reviewed. Finally, the chapter is summarized.

Chapter 3 introduces eight public face databases, which are used in the thesis. Then, it describes the relationships between the distance methods and feature matrices generated by 2DPCA algorithm. Four new distance methods, namely Row, Row Assembled Matrix Distance (RowAMD), Row *k* nearest neighbor (RowkNN) and General Row *k* nearest neighbor Euclidian distance methods, are presented. The performance of the proposed four distance methods is then evaluated with eight face databases compared with the state-of-the-arts distance methods. Consequently, the final discussion regarding the experiment results is presented. Finally, the chapter is summarized.

Chapter 4 introduces the problem of ascertaining whether or not the available training data is appropriate for the recognition system. It also describes the solution of the problems mentioned in this Chapter and gives a description of genetic algorithm. Experimental results with three face database are presented in later part of this chapter. Then, the final discussion regarding the experiment results is then presented. Finally, the chapter is summarized.

Chapter 5 concludes the work of this thesis and summarizes its contributions. Different directions for future work in face recognition systems are then suggested.

CHAPTER 2

BACKGROUND AND LITERATURE REVIEW

2.1 Introduction

The purpose of this chapter is to give background information about face recognition in general and the recent techniques, which are related to the work in this thesis. Different research directions on face recognition and their pros and cons are introduced in Section 2.2. The concept of subspace methods and related mathematical background is presented in Section 2.3. More details about PCA, which is directly related to this work, are also given in the same section. Then, 2DPCA-based algorithms are described in details in sections 2.4. In addition, the distance computation methods, which are proposed in the state-of-the-arts to 2DPCA-based methods, are investigated in section 2.5. After that, face recognition systems evaluation methods are described in Section 2.6. Finally, the chapter is summarized in 2.7.

2.2 Face Recognition

Biometric technology applications are growing because of the increasing needs of security demands in different life aspects. Six different biometric technologies are considered by Hietmeyer (Heitmeyer, 2000); finger, hand, face, voice, eye, and signature. Among these technologies, face recognition has several advantages. It is natural, non-intrusive, and easy to use. The main advantage of the face technology is that it can be captured invisibly at a distance. Besides, facial features achieve the highest compatibility with a Machine Readable Travel Documents (MRTD) (MRTD, 2010) system based on several evaluation factors such as enrollment, renewal, machine requirements, and public perception. Due to these reasons along with its potential for continuously increasing law enforcement and commercial applications, it has been a research topic for decades.

However, face as a three-dimensional object is subjected to different variations such as, camera noise, illumination, pose and facial expression. These variations degrade face recognition systems' efficiency. Many approaches have been proposed to overcome the face variations and can be categorized into three main general strategies (Zhao et al., 2003; Tolba et al., 2005; Guan, 2012; Jafri and Arabnia, 2009):

- **Appearance-based or Holistic-based approach**

  In this approach, the whole face region is fed into a recognition system as a raw input. Then, the global features of faces are used for recognition. As such, a small number of features describes the global information of faces, which are extrapolated outrightly from the pixel information (Turk and Pentland, 1991; Belhumeur et al., 1997; Bartlett et al., 2002; Phillips, 1998; Lin et al., 1997).

- **Feature-based approach**

  Here, local features and prior information about face geometry are first obtained. These features encompass eyes, nose, mouth, chin, and head outline. Then the spatial relations among the parts are computed. This information is then used for classification (Cox et al., 1996; Lades et al., 1993; Wiskott et al., 1997; Nefian and Hayes III, 1998; Lawrence et al., 1997).

- **Hybrid-based approach**

  This approach follows the human perception system, which utilizes both previous approaches. Hence, local features and the whole face region are both used for face recognition (Pentland et al., 1994; Penev and Atick, 1996; Lanitis et al., 1995; Yong et al., 2006; Kisku et al., 2011; Singh et al., 2012).

In principle, feature-based methods are less sensitive to variations in illumination and viewpoint (Zhao et al., 2003; Tolba et al., 2005; Guan, 2012; Jafri and Arabnia, 2009). However, the facial features required for methods using this approach are difficult to be extracted automatically and still not robust enough (Zhao et al., 2003; Jafri and Arabnia, 2009; Lu, 2003; Zhao and Chellappa, 2006). In addition, feature-based methods cannot withstand with noise and occlusion variations (Jafri and Arabnia, 2009; Lu, 2003; Zhao and Chellappa, 2006; Li and Jain, 2004). This is because the extraction of local features is an extremely difficult task.

In contrast, though holistic-based methods are more sensitive to these variations, few modifications or enhancements with several algorithms compensating such variations will generally lead to better results compared to feature-based methods (Jafri and Arabnia, 2009). Besides, holistic-based methods can be applied to low resolution or poor quality images (Lu, 2003). Due to these reasons, the holistic-based approach is a more preferable approach for dealing with face recognition systems (Jafri and Arabnia, 2009; Li and Jain, 2004). Two main categories belong to holistic-based approach; statistical dimensionality reduction and Artificial Intelligence (AI) approaches (Jafri and Arabnia, 2009). AI approaches employ machine learning and neural networks techniques to recognize a face (Jafri and Arabnia, 2009). Good results have been reported using AI (Jafri and Arabnia, 2009). However, statistical subspace methods are considered the most popular and powerful techniques (Jafri and Arabnia, 2009; Lu, 2003; Zhao and Chellappa, 2006; Li and Jain, 2004; Rao and Noushath, 2010). This is due to the strength of their mathematical models, which reveals the underlying discriminative data, solves the curse-of-dimensionality problem, reduces the system's memory and computational requirements (Li and Jain, 2004; Rao and Noushath, 2010; Zuo et al., 2009). Thus, they are more suitable in practical applications.

2.3 Subspace Methods

In this section, the importance of using subspace based methods in face recognition system is introduced. Besides, three popular subspace methods are described.

2.3.1 Overview

A face image is represented as a high dimensional m-by-n pixel array, which is recorded as an array of intensity values using sensors. Broadly, the face image is a high dimensional in nature. This high dimensional array can be represented as a point in mn-dimensional vector space. This space is called image space or more specifically face space. The simplest way of recognition task is to perform a direct comparison in a high dimensional space between input face image and other face images in the database. However, such approach is sensitive to face variations and it is computationally intensive (Jafri and Arabnia, 2009). Besides, applying parametric methods to this high dimensional space makes the estimation task ill-posed. This is because the number of parameters increases with high dimensional spaces and concurrently the number of face images is usually much lower.

Nonparametric methods also face the same problem (Li and Jain, 2004). These methods need a very high number of samples to describe the underlying distribution of face data efficiently (Li and Jain, 2004). These problems are conventionally called curse-of-dimensionality (Jafri and Arabnia, 2009; Li and Jain, 2004; Rao and Noushath, 2010). To counter the curse-of-dimensionality, statistical techniques can be used to analyze the distribution of the face image data in image space, and derive a much lower subspace which retains the most features of the face images (Lu, 2003; Rao and Noushath, 2010). Thus, a lower dimensional feature space can be used for face recognition, which represents face images efficiently and effectively. In this way, the computational burden is reduced significantly and other problems of curse-of-dimensionality are alleviated dramatically. Subspace methods have been successfully applied in face recognition (Mohanty et al., 2008; Zuo et al., 2009; Rao and Noushath, 2010; Jiwen and Yap-Peng, 2013; Imran et al., 2013).

Few popular subspace methods are discussed here, one of which is the Principal Component Analysis (PCA). It will be described in more details because it is used in this work. Besides, Linear Discriminant Analysis (LDA) and Independent Component Analysis (ICA) are also explained in brief.

2.3.2 Principal Component Analysis (PCA)

In this section, PCA algorithm is described in detail. Then, different directions of its improvement are summarized.

2.3.2.1 Algorithm

PCA goal is to derive a lower dimensional subspace from a set of training face images so that it maximizes data variance (Turk and Pentland, 1991). The subspace is known as principal components or eigenfaces. More precisely, let $\boldsymbol{X}$ represents a set of face images as follows:

$$\boldsymbol{X} = [\boldsymbol{x}_1 \boldsymbol{x}_2 \ldots\ldots\ldots \boldsymbol{x}_M]$$

where $\boldsymbol{x}_i$ is a vector of face image with dimension N and M is the number of face images. The vector of face image is formed by concatenating the columns or rows of the image. The typical method of calculating the principal component is to find the eigenvectors and eigenvalues of the covariance matrix $\mathbf{C}$ (Turk, 2001) as in Equation (2.1):

$$\mathbf{C} = \sum_{i=1}^{M} (\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^T \tag{2.1}$$

where $\boldsymbol{x}_i$ is a vector of face image and $\bar{\boldsymbol{x}}$ is the average face image vector. The eigenvectors and the eigenvalues can be calculated from Equation (2.2):

$$\mathbf{C}\mathbf{v}_i = \lambda_i \mathbf{v}_i, \qquad i = 1, \ldots, N \tag{2.2}$$

where $\mathbf{v}_i$ is the i^{th} eigenvector and λ_i is the corresponding eigenvalue which reflects the variance of the images. However, determining N eigenvalues and eigenvectors is an impractical solution for a typical face image size. Practically, there are $(M-1)$ non-zero eigenvalues. Hence, it is more convenient to work with $M \times M$ matrix for eigenvectors calculation. This is can be done by rewriting the covariance matrix as given in Equation (2.3):

$$\mathbf{D} = \sum_{i=1}^{M} (\boldsymbol{x}_i - \bar{\boldsymbol{x}})^T (\boldsymbol{x}_i - \bar{\boldsymbol{x}}) = \boldsymbol{X}^T \boldsymbol{X} \tag{2.3}$$

The eigenvector problem is solved using Equation (2.4):

$$\mathbf{D}\mathbf{e}_i = \mu_i \mathbf{e}_i, \quad i = 1, \ldots, M \tag{2.4}$$

where $\mathbf{e}_i$ and μ_i are the eigenvectors and the eigenvalues of the covariance $\mathbf{D}$.

By substituting the value $\mathbf{D}$ we obtain:

$$\boldsymbol{X}^T \boldsymbol{X}\mathbf{e}_i = \mu_i \mathbf{e}_i, \quad i = 1, \ldots, M \tag{2.5}$$

Multiply both sides by $\boldsymbol{X}$ we get:

$$\begin{aligned} \boldsymbol{X}\boldsymbol{X}^T \boldsymbol{X}\mathbf{e}_i &= \mu_i \boldsymbol{X}\mathbf{e}_i \\ \mathbf{C}(\boldsymbol{X}\mathbf{e}_i) &= \mu_i (\boldsymbol{X}\mathbf{e}_i) \end{aligned} \tag{2.6}$$

That is, the eigenvectors and the eigenvalues of matrix $\mathbf{C}$ in Equation (2.4) are obtained as follow:

$$\begin{aligned} \mathbf{v}_i &= \boldsymbol{X}\mathbf{e}_i \\ \lambda_i &= \mu_i \end{aligned} \tag{2.7}$$

The eigenvectors corresponding to the largest eigenvalues reflect the most variance of the face images. In this way, a subspace that represents the image space in minimum Mean Square Error (MSE) manner can be yielded. After selecting the suitable number of eigenvectors, the centered gallery face images are projected onto the eigenspace using Equation (2.8):

$$\boldsymbol{y}_i = \mathbf{V}^T(\boldsymbol{x}_i - \bar{\boldsymbol{x}}) \tag{2.8}$$

where $\boldsymbol{y}_i \in \mathbb{R}^m$ is the feature vector, $\mathbf{V}$ is the selected eigenvectors matrix and m is the number of selected eigenvectors. When a probe face image appears, the centered probe face image is projected onto the same eigenspace and the nearest gallery face image is chosen as its match using Euclidean distance.

2.3.2.2 Improvement on PCA

The eigenface or PCA method (Turk and Pentland, 1991; Kirby and Sirovich, 1990) is considered as one of the most successful techniques in subspace methods (Turk, 2001; Grudin, 2000; Pentland, 2000; Kim et al., 2002; Moon and Phillips, 2001; Tjahyadi et al., 2006; Meng and Ke, 2008; Gupta et al., 2010; Radha and Pushpalatha, 2010; Zeng et al., 2011), because of the ease of implementation, its reasonable performance level (Moon and Phillips, 2001; Tjahyadi et al., 2006), effectiveness in large databases (Zhao and Chellappa, 2006) and it is less sensitive to different training data set (Jafri and Arabnia, 2009). Despite its merits, PCA still demand improvement. Hence, many researchers have focused on improvement of PCA. These improvements have different directions that can be summarized as follows:

a. Selecting the appropriate numbers of eigenvectors

The number of the eigenvectors' selected is a critical issue in terms of PCA performance. Some studies by Moon and Phillips (1998); Gupta et al. (2010); Kirby (2000); Meng and Ke (2008); Draper et al. (2002); Li et al. (2010); Tjahyadi et al. (2006); Gomathi and Baskaran (2010); Satone and Kharate (2013) have addressed this problem. These works can be divided into two main approaches. The first approach concentrates on removing the last eigenvectors. The second one takes the other side and removes the first eigenvectors.

Moon and Phillips (1998) found that removing the last 40% of the eigenvectors, which comprises the least variance of images, improves the performance of PCA. However, Gupta et al. (2010) in their experimental results show that the last 85% of the eigenvectors can be removed. This should be associated with a threshold value of 80% as the maximum distance between face images. In Kirby (2000), an energy dimension is defined, which is 90% of the accumulated energy of non-zero eigenvalues. It also defines a stretching dimension, which is the ratio of a selected eigenvalue over the maximum eigenvalue. The eigenvectors, having a stretching dimension value of less than or equals to 0.01, are neglected. The Kaiser criterion is also used. It removes all the eigenvectors corresponding to the eigenvalues with values less than one (Meng and Ke, 2008).

Another way to select the best eigenvectors is by using the eigenvalues curve, all eigenvalues with a gentle slope are discarded (Meng and Ke, 2008). It is found that when the slope is less than 0.01, the eigenvectors after that slope can be removed. Draper et al. (2002) have ordered the eigenvectors according to what is called like-image difference values. Like-image difference value is the ratio of the summation of the different images of the same class over the eigenvalue. To find the optimal subspace, Tjahyadi et al. (2006) presents eigenvalue relative errors. The relative errors between the current eigenvalue and the next eigenvalue of 60% of the total non-zero eigenvalues are calculated. Then a subspace is selected depending on these relative errors.

Genetic algorithm (GA) is also used in Gomathi and Baskaran (2010) to select the most appropriate eigenvectors for classification. Li et al. (2010) used improved chaos genetic algorithm (ICGA), which consolidates GA merits in global searching and the sturdy local searching ability of the chaos optimization algorithm (COA), to extract the best eigenvectors for classification. Satone and Kharate (2013) used a genetic algorithm to select the most revelent eigenvectors and the entropy of eigenvectors on wavelet subband.

It is also found that large eigenvalue may carry irrelevant information of face images and it is affected by another condition such as lighting (Moon and Phillips, 1998). Hence, removing the first eigenvector may improve the performance of PCA. The proposed methods improve the performance of PCA in terms of time duration and accuracy by eliminating eigenvectors containing noise.

b. Distance method

The original Eigenface method uses Euclidean distance as a measuring tool. Perlibakas (2004) stated 16 different measuring distance methods that can be used with PCA. It is found that Mahalanobis distance and Euclidean distance superior than other distance methods (Draper et al., 2002; Perlibakas, 2004). All these measuring distance methods use the nearest neighbor (NN) rule between two points. However, the performance of nearest neighbor depends on the available face images per class, which are often small compared to all possibilities of testing data variations. To alleviate this problem, the distance between a point and feature line links two points in eigenspace of the same class replaces the point-to-point distance (Li and Lu, 1999). The results show improvement in terms of accuracy compared to distance methods, which uses nearest neighbor approach.

Another approach which avoids point-to-point distance has been proposed by Zhou and Shi (2009). Each class of the training data is represented by a convex hull that estimates the class distribution. Then l_2-norm is used to calculate the distances between the probe face image and the convex hulls. Experimental results show that a better performance can be obtained using the nearest neighbor convex hull (NNCH) approach. Zhou et al. (2009) have applied l_1-norm with nearest neighbor convex hull (l_1 NNCH) instead of l_2-norm. The results show better performance compared with previous distance methods including NNCH.

c. Preprocess

PCA as a statistical technique works well when the training samples per class are sufficient to build an accurate eigenspace. However, its performance deteriorates when it works with frontal view face recognition. To address this issue, generating new images can detract from the severity of this problem. Using projection-combined principal component analysis (PC^2A) (Wu and Zhou, 2002), new images could be generated using the idea of projection map of the first order to the original images. After generation of the combined projection images using vertical and horizontal projection map, PCA is applied to the new images. Experimental results show an improvement in terms of accuracy and at the same time offer reduction in eigenvectors numbers. Chen et al. (2004) has extended PC^2A (EPC^2A) and used first and second order of projection map. In addition, the generation images of both orders and the original images are used for training PCA to maintain more information about training images. EPC^2A offers further improvements compared with PC^2A in terms of accuracy and number of eigenvectors. However, the values of combination parameters that are used in PC^2A and EPC^2A are a critical point in these methods. Choosing the appropriate values, which give the best accuracy and decrease time cost, is a trade off and is done manually.

Facial symmetry can be also used to generate new images. Symmetrical PCA (SPCA) (Yang and Ding, 2002) generates an odd and even decomposition images from the original images and their mirrors. Then, PCA is applied on the odd and even decomposition images separately. The best features are selected from both eigenspaces. However, only slight improvements are reported using SPCA.

In addition, differences between the same class images in existence of illumination variations are larger than the differences between classes (Zhao et al., 2003; Heusch et al., 2005; Phillips, 1998; Kim and Kittler, 2005; Moses et al., 1994). Hence, PCA-based approaches are sensitive to illumination variations, which have a greater impact on their performance. Recently, many approaches have been suggested to overcome the problem of illumination variation regarding face recognition (Chen et al., 2006;

Shao and Wang, 2009; Wang et al., 2009; Zhang and Samaras, 2006; O'Toole et al., 2007; Tan and Triggs, 2010; Mian, 2011; Bozorgtabar et al., 2012). Among the approaches, Local Binary Pattern (LBP) (Ojala et al., 1996) has become a popular technique for face representation, because it is invariant to monotonic grayscale transformations.

The LBP descriptor assigns a binary string or a decimal number to a pixel of an image by thresholding the intensity values of the eight neighborhood pixels with the value of the central pixel using 3×3-kernel matrix (More details in Appendix A.1). Since then, a number of LBP variants are offered (Tan and Triggs, 2010; Jin et al., 2004; Heikkilä et al., 2009; Xiaosheng and Junding, 2009; Junding et al., 2010; Liao et al., 2007; Zhang et al., 2010; Jabid et al., 2010; Petpon and Srisuk, 2009). In Jin et al. (2004), Improved LBP (ILBP) is done by giving the largest weight to the central pixel. As the central pixel always has more information than its neighbor pixels. The ILBP operator also reveals the local shape by redefining the threshold, which is the mean of a 3×3 patch.

To produce more compact binary pattern, a Center Symmetric Local Binary Pattern (CS-LBP) (Heikkilä et al., 2009) modifies the description of interest regions. Here, only 4-pairs of center-symmetric pixels are compared. Consequently, the coding number is reduced considerably. Nevertheless, important texture information contains in the central pixel are discarded, which is considered as one of the drawbacks of CS-LBP. Besides, choosing an adaptable threshold is a burdensome job (Xiaosheng and Junding, 2009). To conquer these issues, an improvement to CS-LBP, called Direction LBP (D-LBP) (Xiaosheng and Junding, 2009), has been proposed. The local pattern is classified by D-LBP descriptor based on the relation of the center pixel and the center-symmetric pixels, i.e., the pairs of the opposed pixels in a circular neighborhood. Looking at the previous texture methods, a common property can be noticed. They consider only the gray disparity between the pixels in a local region, whereby noise creates a great effect in the calculation (Junding et al., 2010). In order to cope with this problem, Junding et al. (2010) introduced an improvement to D-LBP (ID-LBP), which considers the relation between the center-symmetric pixels and the local gray mean.

Another direction of LBP improvement is introduced in Liao et al. (2007), which is called a Multi-scale Block Local Binary Pattern (MBLBP). It avoids the locality of LBP descriptor by replacing the single pixel computation with comparison to average gray-values of a block of sub-regions. Hence, more information of image representation is captured.

Recently, a high-order local pattern descriptor, called Local Derivative Pattern (LDP) (Zhang et al., 2010), is proposed for face recognition. It encodes the directional pattern features held in a particular region by extracting high-order local information. Another approach to overcome the drawbacks of LBP, which is also more suitable for face recognition, is proposed in Jabid et al. (2010). A new local feature descriptor, called Local Directional Pattern (LDiP), is introduced. Its descriptor produces local features by computing the edge response values in eight directions for each pixel and generating a code from the relative strength magnitude.

A more discriminant and less sensitive to noise in uniform regions descriptor, called

Local Ternary Patterns (LTP) (Tan and Triggs, 2010), is introduced. In Petpon and Srisuk (2009), a novel face representation method for face recognition, called Local Line Binary Pattern (LLBP), is proposed. It summarizes the local spatial structure of an image by thresholding the local window with binary weight and introduces the decimal number as a texture presentation. The basic idea of LLBP is to compute horizontal and vertical line binary code separately and its magnitude so that the change in image intensity can be captured (More details in Appendix A.2).

d. PCA implementation

As mentioned in Section 2.3.2, principal components are calculated using Jacobi's method for eigenvalue decomposition of a covariance matrix. However, the complexity of Jacobi's method is proportional to the number of feature-vectors or samples used, which requires around $O(N^3 + N^2M)$ computations (Golub and Van Loan, 1996). To reduce the computational complexity, a number of methods have been proposed for computing PCA transformation (Reddy and Herron, 2001; Schilling and Harris, 2000; Roweis, 1998). Recently, a computationally fast method, which uses a fixed-point algorithm, was proposed (Sharma and Paliwal, 2007). The eigenvectors are founded without diagonalizing a symmetric matrix. It was demonstrated that fast PCA reduces the computational times of eigenvectors' calculation with very close MSE compared with original PCA.

Computational complexity of PCA is highlighted more in real time applications when updating new face images dynamically is an important issue. Kokiopoulou and Saad (2005) have proposed an efficient implementation of PCA without eigenvalue calculations using the polynomial filtering technique. The numerical results show a clear improvement in terms of speed and storage with a very close performance compared with PCA. Although the previous algorithms show an enhancement in terms of speed and storage compared to Jacobi's method, there is no guarantee that the same performance can be obtained.

It is well known that PCA tries to find a subspace, which represents the image space with a minimum Mean Square Error (MSE) between original matrices and the reconstructed matrix using l_2-norm. Hence, it is sensitive to outliers because large errors squared take control of the sum. To minimize outliers' influence, l_1-norm or the least absolute value has been proposed to deal with this problem (Ke and Kanade, 2005). However, l_1-norm is a rotational variant algorithm (Ng, 2004), which is not preferable in orthogonal transformation models as in PCA.

Taking the merits of the two norms, a rotational invariant l_1-norm PCA (R1-PCA) has been introduced (Ding et al., 2006). The results show a little improvement compared to PCA. The main drawback of R1-PCA is that a subspace calculation is highly dependent on the dimension m to be found, i.e, the subspace when m is equal to one may not be obtained when m is equal to two and so on (Kwak, 2008). Furthermore, the R1-PCA algorithm uses successive iterations to find the subspace, which takes longer time to converge when applied to a large dimensional space like face images (Kwak, 2008). To resolve this weakness, Kwak (2008) has proposed a new PCA method based on l_1–norm maximization called PCA-L1. In this method, instead of maximizing the variance, which is based on l_2-norm, l_1-norm maximization in feature space is used to gain the advantage of l_1-norm with rotational invariant PCA. It is shown that PCA-L1 gives the least reconstruction error compared with PCA and R1-PCA and has less time