

FACE RECOGNITION USING PRINCIPAL COMPONENT ANALYSIS
WITH SUPPORT VECTOR MACHINE CLASSIFIER

ANDY LOW FOO HWA

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

BORANG PENGESAHAN STATUS TESIS

JUDUL: FACE RECOGNITION USING PRINCIPAL COMPONENT ANALYSIS (PCA) WITH SUPPORT VECTOR MACHINE (SVM) CLASSIFIER

SESI PENGAJIAN: 2014/2015

Saya ANDY LOW FOO HWA

(HURUF BESAR)

mengaku membenarkan tesis (PSM/~~Sarjana/Doktor Falsafah~~) ini disimpan di Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dengan syarat-syarat kegunaan seperti berikut:

1. Tesis dan projek adalah hakmilik Universiti Teknikal Malaysia Melaka.
2. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. ** Sila tandakan (/)

 SULIT (Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)

 TERHAD (Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)

 TIDAK TERHAD

(TANDATANGAN PENULIS)

(TANDATANGAN PENYELIA)

Alamat tetap: 15, HALA SINFAR, TAMAN SINFAR, 31650, IPOH, PERAK

Tarikh: _____

PROFESOR MADYA DR ABDUL SAMAD BIN SHIBGHATULLAH

Nama Penyelia

Tarikh: _____

CATATAN: * Tesis dimaksudkan sebagai Laporan Akhir Projek Sarjana Muda (PSM)

** Jika tesis ini SULIT atau TERHAD, sila lampirkan surat daripada pihakberkuasa

FACE RECOGNITION USING PRINCIPAL COMPONENT ANALYSIS
WITH SUPPORT VECTOR MACHINE CLASSIFIER

ANDY LOW FOO HWA

This report is submitted in partial fulfilment of the requirements for the Bachelor of
Computer Science (Artificial Intelligence)

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2015

DECLARATION

I hereby declare that this project report entitled
**FACE RECOGNITION USING PRINCIPAL COMPONENT ANALYSIS
WITH SUPPORT VECTOR MACHINE CLASSIFIER**

is written by me and is my own effort and that no part has been plagiarized without
citations.

STUDENT : _____ Date: _____
(ANDY LOW FOO HWA)

SUPERVISOR : _____ Date: _____
(PROF.MADYA DR ABDUL SAMAD BIN
SHIBGHATULLAH)

DEDICATION

I would like to thank my beloved parents, lecturers and my friends for supporting me and inspired me throughout the completion of this project.

I would also like to thank to my supervisor, Prof. Madya Dr Abdul Samad Bin Shibghatullah who guide me all along the way to complete this project. He always comments and corrects my work as well as sharing his experience to me. All of my friends are being supportive and encourage me at all time.

Lastly, to my beloved parents, lecturers, supervisors and friends, thanks to you all for supporting me and motivating me all the time.

ACKNOWLEDGEMENTS

First and foremost, I would like to take this opportunity to thank my supervisor, Prof. Madya Dr Abdul Samad Bin Shibghatullah for assisting and guiding me throughout the completion of this project.

I would also like to thank my parents for supporting me mentally and financially to complete this project as well as my studies in UTeM. Lastly, I would like to thank my friends that shared their knowledge and experience with me in order to help me complete this project.

ABSTRACT

An efficient biometrics identification system is important to identify the identity of an individual. Biometrics includes human face, fingerprint, iris and others. In this project, human face recognition has been chosen to be used in the system. Principal Component Analysis (PCA) has been selected to be the algorithm working with the system. However, due to the fact that PCA itself is not a good classifier so a classifier is picked to classify the data which is Support Vector Machine (SVM) classifier. Basically, the main idea of PCA is to reduce the high dimensional image space to a lower one, as known as dimensionality reduction. During the process of reducing dimension, each of the images is being treated as an $n*n$ matrix and the algorithm will calculate the mean face of all the individuals; subsequently create a lower dimensional face space. When a sample test image is feed into the system to perform matching, the test image will be projected onto the face space and perform matching. Then, SVM classifier will classify which class the individual belongs to. In order to test the system, an Adaboost classifier has been picked to compare the accuracy. As a result, SVM classifier outperforms the Adaboost classifier. In general, the graphical user interface is user friendly and user just need to train the system, browse a test image and then perform recognition. Although there are some weaknesses of this system, it can still be improved by hybrid with other efficient algorithm to achieve a more accurate result. Further research needs to be carry out in order to improve and enhance the efficiency and performance of the system.

ABSTRAK

Sistem pengenalan biometrik yang cekap adalah penting untuk mengenal pasti identiti individu. Biometrik termasuk muka manusia, cap jari, iris dan lain-lain. Dalam projek ini, pengiktirafan muka manusia telah dipilih untuk digunakan dalam sistem. Analisis Komponen Utama (PCA) telah dipilih untuk menjadi algoritma bekerja dengan sistem. Walau bagaimanapun, disebabkan kepada fakta bahawa PCA sendiri bukanlah satu pengelas yang baik, satu pengelas telah dipilih untuk mengklasifikasikan data iaitu pengelas Sokongan Mesin Vektor (SVM). Pada asasnya, idea utama PCA adalah untuk mengurangkan ruang dimensi imej yang tinggi kepada yang lebih rendah, boleh juga dikenali sebagai pengurangan dimensi. Semasa proses pengurangan dimensi, setiap imej dirawat sebagai $n * n$ matriks dan algoritma akan mengira purata muka semua individu; seterusnya mewujudkan ruang muka dimensi yang lebih rendah. Apabila satu imej dimasukkan ke dalam sistem untuk melaksanakan sepadan, imej tersebut akan diunjurkan ke ruang muka dan melaksanakan proses sepadan. Kemudian, SVM pengelas akan mengelaskan individu tersebut ke kelas yang paling dekat. Untuk menguji sistem, satu pengelas Adaboost telah dipilih untuk membandingkan ketepatan. Hasilnya, SVM pengelas didapati melebihi performa pengelas Adaboost. Secara umum, Graphical User Interface (GUI) system ini mudah diguna dan pengguna hanya perlu melatih system tersebut, memasukkan satu imej dan kemudian melaksanakan proses sepadan. Walaupun terdapat beberapa kelemahan dalam sistem ini, ia masih boleh diperbaiki dengan hibrid dengan algoritma lain yang cekap untuk mencapai keputusan yang lebih tepat. Kajian lanjut perlu dijalankan untuk memperbaiki dan meningkatkan kecekapan dan prestasi sistem

TABLE OF CONTENTS

CHAPTER PAGE	SUBJECT	
	DECLARATION	ii
	DEDICATION	iii
	ACKNOWLEDGEMENTS	iv
	ABSTRACT	v
	ABSTRAK	vi
	TABLE OF CONTENTS	vii
	LIST OF TABLES	x
	LIST OF FIGURES	xi
CHAPTER 1	INTRODUCTION	
	1.1 Introduction	1
	1.2 Problem Statement	3
	1.2.1 Lightning condition	3
	1.2.2 Face orientation	4
	1.2.3 High dimensional facial images	4
	1.3 Objective	4
	1.4 Scope	5
	1.5 Project Significance	5
	1.6 Expected Output	6
	1.7 Conclusion	6
CHAPTER 2	LITERATURE REVIEW	
	2.1 Introduction	8
	2.2 Facts and Findings	9
	2.2.1 Domain	10
	2.2.2 Related Research	11
	2.2.3 Techniques	15
	2.2.3.1 Principal Component Analysis (PCA)	16
	2.2.3.2 Linear Discriminant Analysis (LDA)	24
	2.2.3.3 Elastic Bunch-Graph Matching (EBGM)	31
	2.3 Project Methodology	39
	2.3.1 Preliminary Studies	40

2.3.2 Data Preparation	40
2.3.3 System and Algorithm Design	41
2.3.4 System Analysis	41
2.3.5 Computerized Model	41
2.3.6 Testing	42
2.3.7 Result Analysis	42
2.4 Project Requirements	42
2.4.1 Software Requirements	43
2.4.2 Hardware Requirements	43
2.5 Project Schedule and Milestone	43
2.6 Conclusion	45
CHAPTER 3 METHODOLOGY AND DESIGN	
3.1 Introduction	47
3.2 Problem Analysis	48
3.2.1 Algorithm Flow	48
3.2.2 System Flow	49
3.2.3 Algorithm Analysis	50
3.2.3.1 Image Representation	50
3.2.3.2 Mean Image	51
3.2.3.3 Covariance Matrix	52
3.2.3.4 Eigenvectors and Eigenvalues	53
3.2.3.5 Projection of face images onto face space	56
3.2.3.6 Recognition	57
3.3 Requirement Analysis	58
3.3.1 Data Requirements	58
3.3.2 Functional Requirements	59
3.3.3 Non-functional Requirements	59
3.3.4 Others Requirements	60
3.4 Conclusion	61
CHAPTER 4 DESIGN THE PROPOSED TECHNIQUE	
4.1 Introduction	62
4.2 High-Level Design	62
4.2.1 System Architecture	63
4.2.2 User Interface Design of the system	63
4.2.2.1 Input Design of the system	64
4.2.2.2 Technical Design	65
4.2.2.3 Output Design	67
4.3 Detailed Design	68
4.4 Conclusion	69
CHAPTER 5 IMPLEMENTATION	
5.1 Introduction	70

5.2 Software Development Environment Setup	70
5.3 Software Configuration Management	71
5.3.1 Configuration environment setup	71
5.3.2 Version Control Procedure	73
5.4 Implementation Status	74
5.5 Conclusion	75
CHAPTER 6 TESTING	
6.1 Introduction	76
6.2 Test Plan	76
6.2.1 Test Organization	77
6.2.2 Test Environment	77
6.2.3 Test Schedule	78
6.3 Test Strategy	78
6.3.1 Classes of Tests	79
6.4 Test Implementation	81
6.4.1 Experimental/ Test Description	81
6.4.2 Test Data	85
6.5 Test Results and Analysis	86
6.5.1 PCA+ SVM	86
6.5.2 PCA+ Adaboost	88
6.6 Conclusion	90
CHAPTER 7 CONCLUSION	
7.1 Observation on Weaknesses and Strength	91
7.2 Propositions of Improvement	92
7.3 Project Contribution	93
7.4 Conclusion	93
REFERENCES	95
APPENDICES A	97

LIST OF TABLES

TABLE	TITLE	PAGE
2.1	Project Schedule and Milestone	43
3.1	Threshold value	57
5.1	Implementation status	74
6.1	Performance measure	89

LIST OF FIGURES

FIGURE	TITLE	PAGE
2.1	Original human facial image (training set)	17
2.2	Eigenfaces created from the original image set	17
2.3	Mean face	17
2.4	Eigenfaces with specific eigenvalues	22
2.5	Projection of input image onto facespace	22
2.6	Faces under various occlusion such as lightning, facial expression and glasses	25
2.7	Example of scatter matrix of two classes	26
2.8	Scatter matrices of PCA and LDA	27
2.9	Steps in LDA	29
2.10	Fiducial points on human face	32
2.11	Gabor jets on elastic grids	32
2.12	Face Bunch Graph (FBG)	34
2.13	Jets extracted at the noted of the face graph	34
2.14	Creation of jet by Gabor Wavelet Transform	36
2.15	Project Methodology	39
3.1	Algorithm flow	48
3.2	System flow chart	49
3.3	Mean face of 3 individuals	51
3.4	Difference of each face with the mean face	51
3.5	Face space (eigenfaces)	55
4.1	GUI of the system	63
4.2	Interface of “Browse image” function	64
4.3	Input image in GUI	65
4.4	Trained images (Face space)	66
4.5	Testing Results	66
4.6	Recognized face	67

4.7	Matched face class	67
5.1	MATLAB R2015a	71
5.2	Starting MATLAB R2015a	71
5.3	Create new script	72
5.4	New script	72
5.5	Rename and save the '.m' file	72
6.1	Functionality testing	80
6.2	Step by step process	82
6.3	Mean faces	82
6.4	Browse image	83
6.5	Recognition	84
6.6	Classified class	84
6.7	Steps of voting strategy	85
6.8	Training samples by SVM classifier	87
6.9	Testing results by PCA+ SVM+ one-vote strategy	87
6.10	Training samples by Adaboost classifier	88
6.11	Testing results by PCA+ Adaboost+ one-vote strategy	89
6.12	Recognition rate	89

CHAPTER I

INTRODUCTION

1.1 Introduction

In computer vision, there are many applications being developed such as security system, object detection, face recognition, human-computer interaction, criminal analysis and others. In verification and identification of a person identity, there are several recognition techniques to do so such as thumbprint, voice, iris, face and other significant human part that define a person. However, face recognition is one of the most interesting topic and being focus by many researchers. We human can easily recognize someone in our daily life easily by remembering and recognizing their features on their face such as eyes, nose, lips and others. This recognizing ability is natural and without awareness for a human being.

When it comes to computer vision, it is not an easy task because computer needs a learning algorithm to extract the face features and match it with a database of images in order to identify a person. It is quite complicated to recognize a human face by computer vision because human face may vary from time to time due to the face shape distortion for example glasses, beard, hair condition and other factors. Since features have to be extracted from human face to proceed with recognition task, some

of the features or informations might not be important and have to be eliminated in order to reduce the time in processing these informations. Thus, the efficiency of the system will be increased. Nowadays, face recognition system has been widely deployed in certain places for security purposes such as airport, offices, police station etc. For example, a robust and efficient face recognition system is needed for an airport security management to quickly identifying and retrieving a person identity.

In face recognition, the images of human face is dimensionality high and requires a longer time to perform the classification task. This problem can be solved by reducing the dimension of the face images thus Principal Component Analysis (PCA) is being recommended because it can significantly reduces the dimension of human faces. The PCA is being introduced by M.A Turk and Alex P.Pentland in 1991 and they have developed a near real time Eigenface system for face recognition by computing the Euclidean distance. While reducing the dimension of facial images, the deviation of the facial images are being retain at the same time. Basically, the PCA method is based on information theory approach that fractionate the facial images into a set of characteristic feature images as known as Eigenfaces and these Eigenfaces are considered as the principal components of the initial training set of images. There are a few steps when performing face recognition, firstly a new input image is being projected into the “face space”, a subspace that spanned by the eigenfaces. Then, the face are being classified by comparing its location in the face space with the locations of known person.

In the face recognition system by using PCA method, the database of facial images are being trained and learned in order to recognize new input faces. A set of training images are being trained so that recognition can takes place under different viewing situation such as upright, 45°, profile view and etc. However, there is a method that almost similar as PCA which is Linear Discriminant Analysis (LDA).

Both of the two methods use the same concept which is reducing the dimension of facial images but LDA is less sensitive to illumination and facial expressions. However, PCA is much better than LDA in terms of computational cost. Besides that, PCA is simple, fast, big learning capacity and is less sensitive to small changes in the facial image.

1.2 Problem Statements

It is difficult to detect and recognize a human facial images when the image is rotated, scaled and change in illumination. Besides that, the dimension of facial images is always high and it is difficult to extract the feature vectors from the face and perform recognition task.

1.2.1 Lightning condition

Images may affect by the illumination variation. By changing the direction of illumination, the coordinate and shape of the shadows may varied thus creating noisy data. Besides that, images may be taken under different condition such as day time or night time especially during night time, the light intensity is so low and weak. Low light intensity may reduce the efficiency of recognition process. Light intensity may also cause by taking images with different cameras as well as video recorder. Different hardware may use different settings for taking pictures thus affecting the quality of images.

1.2.2 Face Orientation

Face orientation is one of the main challenge in face recognition process. Human face maybe taken in different situation. For example, a human face might not be captured in a consistent position or the face may rotated to a certain angle. Failed in capturing a human face completely may leads to difficulty in extracting features from the image and matching.

1.2.3 High dimensional facial images

Most of the human facial images is high dimensional and the human face consists of a lot of important features or information that are needed to be recognized. It will consumes a lot of time in extracting features from a human face in a high dimension facial images. The computational time for classification can be reduced by reducing the high dimensional facial images and eliminating those features that are irrelevant on a human face.

1.3 Objective

1. To investigate the most suitable and effective method used to recognize human face.
2. To design the face recognition system using Principal Component Analysis (PCA) technique.
3. To develop the prototype of face recognition system by using PCA approach.

1.4 Scope

The scope for this project covered the approach to be used which is the Principal Component Analysis (PCA). This algorithm recognize human faces by comparing the characteristics and features of the face to a known person. It reduces the dimensional of facial images by treating the images as two-dimensional (2-D) instead of three-dimensional (3-D). Then, the face images will be projected onto the face space that spanned by the eigenvectors or eigenfaces and weighted sum of the eigenface features are being compared to the known faces in order to recognize a new face. Database of images can be imported for training and testing purpose. For training, the public databases of face images are downloaded from the Yale Face Database and the Yale Face Database B for further investigation. Besides that, the images can be in RGB scale and it will be transform into gray scale for further processing. The input image and training images can be choosed from a file or by taking a picture by initializing the camera tool.

1.5 Project Significance

Based on the study and observation, face recognition using PCA approach can perform reliable matching between a new human face and a database of known individuals because the database of images will be trained and significant features (eigenfaces) will be extracted in order to perform recognition. It is significant for video surveillance such as face scan in an airport. During the face scanning, it can quickly matches the faces accurately and retrieve a person's information. As a result, security level in the airport can be improved. Besides that, in an access control system such as granting permission or authorisation to control access, the system needs a highly

efficient face recognition software in order to correctly verifying faces. In this case, PCA can perform well by correctly matching the face within a large database of face images in a shorter computation time. As a result, the performance of access control system will be increased. By recognizing facial images efficiently, criminal identification can be retrieved in a fast manner. Subsequently, the criminal analysis rate will be improved significantly.

1.6 Expected Output

The face recognition system by using PCA method will be implemented in various application that is involved in computer vision such as pedestrian detection, client verification, criminal identification and others. Besides that, a simple, high efficiency and user friendly system prototype will be developed as the user interface is easier to understand and easier to use. Furthermore, the PCA method will perform robustly by increasing the recognition rate and reduces the computation time. Besides that, PCA approach is more computationally cheaper compared to others face recognition algorithm thus it will be cost effective.

1.7 Conclusion

In this chapter, introduction of this project is explained. The problem statements that are discussed in this project include the difficulty to detect and recognize a human face when the human face is being rotated, under different illumination condition and high dimensional of facial images. Then, the objective in this project is being observed,

that is to investigate the most suitable and effective method use to recognize human face. Next is to design the face recognition system using Principal Component Analysis (PCA) approach. After that, is to develop the prototype of PCA-based face recognition system.

Based on the study and observation, face recognition using PCA approach can perform reliable matching between a new human face and a database of known individuals because the database of images will be trained and significant features (eigenfaces) will be extracted in order to perform recognition. Experiment will be carried out and results will be recorded. Enhancement of the performance of the system is necessary in order to reduce the computation power and time as well as the false match of faces.

A user friendly prototype of PCA-based face recognition system will be developed and will be contributed to the society. Security level and verification process in any location and department can be increased by using a system with PCA method because of the extracted features that are less sensitive to small changes in the facial images and taking the advantage that the facial images are taken from a front view.

As a conclusion, the PCA method has been widely used by many researchers and it is a suitable technique to apply in face recognition system because of its' excellent performance in reducing high dimensional to a low dimensional of facial images. Although the performance is affected by illumination condition and face positioning, the facial images are taken under controlled condition. Next chapter will be focussing on literature review of this project.

CHAPTER II

LITERATURE REVIEW AND PROJECT METHODOLOGY

2.1 Introduction

In chapter II, literature review and project methodology will be discussed. Literature review will cover the background of the proposed technique and other previous researcher's works. The project methodology will further discuss the proposed technique which is Principal Component Analysis (PCA).

In this project, there are a lot of resources for me to refer to, for example books and journals from library, e-library, technical papers from internet and others. The Principal Component Analysis (PCA) is introduced by Karl Pearson who is a mathematician and statistician of University College London King's College, Cambridge. During his researching years, he has been honored by forming the mathematical statistics discipline and he has contributed his knowledge on biometrics such as pattern recognition as well as face recognition. His research is then continued and developed by Harold Hotelling in 1930 where the PCA method has been used as a data analysis tool and model prediction tool. Basically, PCA is deal with covariance matrix, eigenvalue decomposition, computing eigenvectors, finding Euclidean distance and others.

However, there are only limited sources related to the PCA method. This is because this technique was discovered by Karl Pearson and other researchers are just improving the algorithm. There are several techniques that are related to the field of face recognition before the PCA technique such as Linear Discriminant Analysis (LDA), Gabor Wavelet Transform and etc. Then, these techniques have their strengths and weaknesses which will eventually leads to comparison of their performance on recognizing rate and will be discussed later.

Throughout the literature review, there are a lot of useful knowledge and researches to help me in carrying out this project as well as investigating a more suitable method in face recognition. It is a critical process to look for the facts and findings as the resources are limited or insufficient. However, a number of past projects have been completed regarding the PCA method which will guide me throughout the completion of this project.

2.2 Facts and Findings

Facts and findings is the central point to understand the flow and purpose of this face recognition system. There are various kinds of research that are involved in the field of object recognition as well as face recognition that have been done by other researchers.

First of all, facts and findings will discuss about the algorithm of PCA method for example the main idea of PCA method and how it is being implemented. Then, other approaches in face recognition also will be discussed.

2.2.1 Domain

There are several types of domain for recognition tasks that is related to this project for example surveillance camera, access control, personal identification and others. In terms of recognition, the most frequently used part includes face, palm print and fingerprint. In this project, face recognition has chosen to be discussed. Surveillance means that observing activities, behavior and situation of human for the purpose of security, law and enforcement and others. In this case, CCTV camera with embedded face recognition function is suitable to use as an observation electronic device in any places such as pedestrian walkways, housing area and others.

For access control and personal identification, these two domains fall into the same type which is authentication. In authentication, face recognition act as an important medium to allow access to certain system by verifying the identity of a client. If the client is an authorize personnel, he or she will be granted access to the system whereas access denied if the client is an imposter. An accurate and robust face recognition system is needed in order to prevent fault authentication and to reduce redundancy of the system. In this case, face recognition using PCA technique is proposed to overcome this problem.

In recognizing human faces, there are 3 major research categories; firstly is facial characteristic which is the ability of human in recognizing the face features of other person. In our daily life, we human can recognize a number of human faces easily by remembering their features on the face and this action are being performed without our awareness. The second category is identifying human face by extracting the feature vector from a human silhouette. The third category will be extracting feature vectors from a human face at front view.