

UNIVERSITY OF GHANA



DEPARTMENT OF STATISTICS

HEAD TILT CLASSIFICATION USING FFT-PCA/SVM ALGORITHM

JOHN KWAME ESSEL

(10302819)

This Thesis is submitted to the University of Ghana, Legon in partial fulfillment of the requirement for the award of MPHIL Statistics degree.

July 2018

DECLARATION

I hereby declare that this submission is my own work towards the award of the M.Phil. degree in Statistics and that, to the best of my knowledge, it contains no material previously published by another person nor material which had been accepted for the award of any other degree of the university or any other university, except where due acknowledgement had been made in the text.

John Kwame Essel
(ID: 10302819) Signature Date

Certified by:
Dr. L. Asiedu
(Principal Supervisor) Signature Date

Certified by:
Dr. F. O. Mettle
(Co-Supervisor) Signature Date

ABSTRACT

The complexity of a face's components originates from the constant variations in the facial component that occur with respect to time. Notwithstanding these variations, humans recognize a person very easily using physical characteristics such as face, voice, gait, etc. Automatic face recognition deals with extracting these essential features from an image, placing them into a suitable representation and performing some kind of recognition on them. This study presents an assessment of the performance of Fast Fourier Transform and Principal Component Analysis with Support Vector Machines (FFT-PCA/SVM) under the constraint of head tilting. 80 head pose images from 10 individuals were extracted from the Massachusetts Institute of Technology Database (2003-2005). Classification rate and runtime were adopted as the numerical evaluation methods to assess the performance of the study algorithm. All computations were done using MATLAB. An earlier study asserted that the higher the degrees of head-pose the larger the Euclidean distance and that above 20° , the Euclidean distances become profoundly larger compared to the 4° head-pose. The study therefore divided the data set into two classes; Lower (below 20°) and Upper (20° and above) and performed an SVM classification. Results from the classification showed that 16° head tilts were more likely to be in the upper class than the lower class. This indicates that with the classification, the Euclidean distance becomes larger from 16° and beyond. It was observed that the SVM classifier performed better in the upper class (20° and above) than that of the lower class (below 20°). It is therefore recommended that further studies be conducted to understand the poor performance of the model in the lower class (below 20°).

DEDICATION

This thesis is dedicated to the Almighty God, my mother Hannah Essel, Mr. Charles Asare, Comfort Asare, Theophilus Mensah and the entire family.

ACKNOWLEDGEMENTS

To God, the Father, the giver and maker of vision. Thank you for a dream come true.

I would like to express my heartfelt appreciation and gratitude to my supervisors Dr. L. Asiedu and Dr. F. O. Mettle for their support. From the initial stages (choosing a topic) of this thesis, to the last day of submission, I owe an immense obligation of gratitude to my supervisors. The inexhaustible supply of help, sound advice, encouragement, careful guidance and most importantly patience has been invaluable in the completion of this thesis.

I am highly indebted to my good friends Ransford Ekow Baidoo, Abdul S. Badi, and Angelina Ampomah for all the support they gave me in getting my dream in education come true.

Finally, I would like to thank my roommates and colleagues for their constant encouragement without which this assignment would not have been possible.

Thank you all for the immeasurable support. May the Almighty richly bless you.

TABLE OF CONTENT

DECLARATION i

ABSTRACT ii

ACKNOWLEDGMENT iii

DEDICATION iv

ABBREVIATIONSv

LIST OF TABLES vi

LIST OF FIGURES vii

LIST OF PLATES viii

CHAPTER ONE 1

INTRODUCTION 1

 1.1 Background of the Study 1

 1.2 Statement of the Problem 5

 1.3 Objectives of the Study 7

 1.4 Significance of the Study 7

 1.5 Brief Methodology 9

 1.6 Scope of the Study 10

 1.7 Organization of the Study 10

CHAPTER TWO 11

REVIEW OF LITERATURE 11

2.1 Introduction	11
2.2 Review of Works on Facial Recognition	11
2.2.1 Face Recognition Approaches	14
2.2.2 Face Recognition Using Template Matching	14
2.2.3 Face Recognition Using Geometric Features.....	15
2.3 Pre-processing Techniques in Face Recognition	16
2.4 Feature Extraction Methods	17
2.5 Face Recognition by SVM	18
2.5.1 Theoretical Approach	18
2.5.2 Empirical Approach	19
2.6 Face Recognition by PCA	23
2.6.1 Theoretical Approach	23
2.6.2 Empirical Approach	24
2.7 Singular Value Decomposition (SVD)	27
2.8 Summary	28
CHAPTER THREE	29
METHODOLOGY	29
3.1 Introduction	29
3.2 Data Acquisition	29
.3 Classification Procedure	30
3.3.1 Processing Stage	31
3.3.2 Mean Centering	32
3.3.3 Fast Fourier Transform	34

3.4 The Process of SVD	36
3.4.1 Properties of SVD	37
3.4.2 Theorem	37
3.5 PCA.....	37
3.5.1 Principal Component Analysis (PCA)	37
3.5.2 Feature Extraction Technique	38
3.5.3 Classification Stage	40
3.5.4 Euclidean Distance	41
3.6 SVM.....	42
3.6.1 Support Vector Machines	42
3.6.2 Multi-Class Recognition	46
3.7.1 Numerical Evaluation	48
CHAPTER FOUR	49
DATA ANALYSIS	49
4.1 Introdect.....	49
4.1.1 Data Description	49
4.2 Result	50
4.2.1 Pre-processing Stage	50
4.2.2 Feature Extraction Stage	51
4.2.3 Training of SVM Model	57
4.2.4 Classification Score	58
4.3 Posterior Probabilities	59

4.3.1 Model Accuracy	60
4.3.2 Classification Rate	60
4.3.3 Classification Curve	62
CHAPTER FIVE	63
SUMMARY, CONCLUSION AND RECOMMENDATION	63
5.1 Introduction	63
5.2 Summary	63
5.3 Conclusion and Recommendation	64
REFERENCE	65

LIST OF FIGUREES

Figure 1.1: The Classification Process..... 9

Figure 3.1: Research Design..... 31

Figure 3.2: Flow chart of PCA..... 42

Figure 3.3 SVM classification using hyperplanes..... 43

Figure 3.4: Flow chart of SVM Classification..... 47

Figure 4.1 Proportion of variance explained by principal components..... 55

Figure 4.3 Plot of Train Data..... 58

Figure 4.5 Classification Curve..... 62

LIST OF PLATES

Plate 3.1: Original Frontal Images of Dataset.....30

Plate 4.1: Example of Train Images..... 49

Plate 4.2: Examples of Test Images..... 50

LIST OF TABLES

Table 1.1: Application Areas of Face Recognition.....	8
Table 4.1 Euclidean Distance.....	56
Table 4.2 Classification Scores.....	58
Table 4.3 Posterior Probabilities.....	59
Table 4.4 Model Accuracy.....	61

ABBREVIATIONS

ANN.....	Artificial Neural Network
BPNN.....	Back Propagation Neural Network
DFT.....	Discrete Fourier Transform
DWT.....	Discrete Wavelet Transform
FFT.....	Fast Fourier Transform
GMM.....	Gaussian Mixture Model
ICA.....	Independent Component Analysis
IDFT.....	Inverse Discrete Fourier transform
LDA.....	Linear Discriminant Analysis
NC.....	Nearest Centre
NFL.....	Nearest Feature Line
NMC.....	Nearest Mean Classifier
NMF.....	Non-Negative Matrix Factorization
ORL.....	Oto-Rhino Laryngologie
PCA.....	Principal Component Analysis
SMO.....	Sequential Minimal Optimization

SVD..... Singular Value Decomposition

SVM..... Support Vector Machines

CHAPTER ONE

INTRODUCTION

1.1 Background

Biometric technologies have become all important, especially in the twenty-first century. This is greatly because they can be used for automated identification and verification, which in most cases is performed on finger prints, iris, face, hand geometry and voice recognition. Amongst all the forms of identification or verification, face recognition has been proven to be more robust and also requiring less user interaction among these technologies. Thus, face recognition has come to the fore rapidly as a lively area of research in the biometric field. It is known from our own personal experiences, that the human face shows a phenomenal level of robustness against rotation about the vertical axis. As a result, if one becomes familiar with a person from only a single photograph, it is nonetheless not too difficult to see that person even from disparate views which has never been seen before. This, according to Marques' (2010), is primarily because the human face recognition is a dedicated process in the brain. That is, human brains have specialized nerve cells which respond to certain local characteristics of a scene, such as lines, edges, angles or movement. These nerve cells according to Wagner (2012) operate twenty-four (24) hours after birth. The process of automated facial recognition is therefore all about extricating the meaningful features from a face, putting them into a convenient representation and conducting some kind of classification on them (Wagner, 2012). Consequently, the goal of parroting such a skill immanent in human beings by machines can be extremely beneficial although the thought of developing an intelligent and self-learning system may need the supply of adequate information to the system. This is due to the fact that although facial recognition is a dedicated process in the brain, it is quite clear that for a relatively large database, it is very difficult and error prone to visually match a face

to its various pose changes and vice versa. Notwithstanding the fact that humans, to a large extent are good in recognizing faces which are known, they are not so dexterous when it comes to dealing with unknown images in large quantities. However, computers, having almost inestimable computational speed and memory, are able to deal with such human flaws. To a large extent, the algorithms help to ease the problem of trying to identify a particular face by its pose visually. The real-life applications (e.g. human/computer interface, surveillance, secure access) and the accessibility of affordable and hefty hardware have also led to the development of facial recognition systems for commercial purposes.

Among the most challenging problems in the area of image enquiry and understanding is facial recognition. Facial recognition systems are basically categorized into two groups, which are verification and identification. For face verification, there is a 1:1 matching that weighs up face images against template face images. Facial identification on the contrary is a 1: N situation that juxtaposes probe face images against every single one of the image templates in a face database (Hasan, Jouhar & Alwan, 2012). Facial recognition poses a very arduous hurdle as a result of the divergent face poses, considerable variations in light direction (illumination), varied facial expressions, aging (changing the face over time) and Occlusions (like cosmetics, hair, glasses).

Generally, to identify an individual from an unfamiliar face, the unknown face is normally compared with the known faces from a face database. The inquisitiveness of researchers in the issue of facial recognition has grown expeditiously in recent times due to the wide range of law enforcement and commercial applications of face recognition. Areas of application include but not limited to crowd surveillance, driver's license, criminal investigations, passport checks, entrance control, witness face reconstruction, and ATM machines. The ever-soaring demand for

surveillance-based technologies, predominantly resulting from the activities of terrorists and drug peddlers has a significant repercussion on the rising interest in the area of facial recognition.

Researchers have shown much interest in the problem of face recognition in recent years. In the group of such researchers include psychophysicists, engineers and neuroscientists studying this famous situation in separate fields and from different dimensions. Lots of researches are being perceived by these researchers for the last five decades to address the myriad of issues in the challenging field of facial recognition. Since the 1960s, lots of research work has been carried out in this field and this has led to vast improvement in both the technologies and the algorithms, which have profoundly augmented a computer's capability to recognize the same person in multiple images. For many years, people were using physical human attributes, for example, gait (movement), facial expression, voice, height, weight, shadow and many more mechanisms to recognize an individual. The set of facial features that relate to an individual constitutes their respective identity. With the introduction of computers in the nineteenth century, the scientific community became interested in using computer software applications to identify or verify a person's identity. This led to face recognition studies which provides an individual facial image with targeted facial characteristics (such as the mouth, eye, ear and nose) and tends to measure a face image at a time.

The techniques of face recognition have continued to attract remarkable curiosity from researchers partially due to the fact that facial recognition is non-invasive with a sense of primary identification (Pentland & Choudhury, 2000). A major propelling factor for facial recognition is the innumerable applications that a resilient and an efficient system addresses; for instance, security systems relying on biometric data, missing children identification, criminal identification, voter identification and user-friendly human-machine interfaces. A classic illustration of the latter categorization is smart

rooms, which normally employ cameras and microphones arrays to discover the arrival of humans, determine their identity and then acts based on a set of preferences which are predefined for each individual (Asiedu, Adebajji, Oduro & Mettle, 2015).

Regardless of the successes of most of such systems in contrived situations, the broad tasks of facial recognition continue to be faced with numerous challenges like pose, illumination and facial expression (Heisele, Ho, Wu & Poggio, 2003). These challenges in facial recognition persists due to the fact that the human face is complex and rigid and that images can be acquired from different viewpoints. The complexity of the human face's characteristics originates from continual changes in the facial characteristics that occur over time. According to Heisele et al., (2003) varying the head pose of an individual consequently leads to significant imbalances in the images of an individual's face. The basic and foremost implication of such an assertion is that, despite the fact that it is considerably robust, face recognition is not excellently invariant to pose and viewpoint variations. Stiefelhagen (2004) posited that when the head tilt of an individual is known, it gives significant leads regarding visual focus of attention. Since the face images captured by many digital devices are somehow tilted, many researchers have advocated for extensive research work directed towards the constraint of varying head poses. Sahu and Tiwari (2015) corroborated that developing an algorithm for a face recognition system is quite a challenging task since human faces are multifaceted, multi-dimensional and react to visual stimuli. Modelling human interactions and applications like intelligent environments demand the estimation of head poses from face images.

The head pose of an individual is generally ascertained by three angles, which includes the slant, pan and tilt. The slant angle constitutes inclination of the head with respect to the image plane, while the pan and tilt angles shows the vertical and horizontal inclinations of a face. Various

researchers have developed various face recognition algorithms or techniques to deal with the phenomenon of pose variation. Some of these include but not limited to Artificial Neural Networks (ANN), Independent Component Analysis (ICA), Discrete Wavelet Transform and Principal Component Analysis with Singular Value Decomposition (DWT-PCA/SVD), Support Vector Machines (SVM), Linear Discriminant Analysis (LDA), Garbor Wavelet, Fast Fourier Transform and Principal Component Analysis with Singular Value Decomposition (FFT-PCA/SVD), just to mention a few. All these designs (algorithms) give various values for the evaluation procedure, be it the average run time or the recognition rate. However, current attention of researchers is drawn to finding a comparatively efficient and consistent algorithm in recognizing face images under varying head pose.

It is a very significant objective in facial recognition, to be able to measure the algorithm on the basis of consistency and efficiency in identifying or verifying images. Most of the previous methods of evaluation focused only on the algorithm's runtime, recognition rate, the algorithms sensitivity and descriptive statistics as the fundamental technique of measuring the performance of a recognition algorithm (Asiedu et al., 2015). This study, therefore seeks to assess the performance of support vector machines (SVM), combined with principal component analysis and singular value decomposition face classification algorithm, under the constraint of varying head poses, where FFT is used as a pre-processing technique.

1.2 Problem Statement

Weyrauch, Heisele, Huang & Blanz (2004) posited that, variations in the frontal view pose predominantly occasion imbalances in the position of the components of the face and that necessitated flexibility of the geometric model. Murugan, Arumugam, Rajalakshmi & Manish (2010) stated that after many years of invention of various software packages and exploration, face

recognition remains a challenging field of study. The authors argued that face recognition is sensitive to variation based on an individual image pose position. Different authors have used different descriptors and classifiers with different approaches, for the effective face recognition systems, to overcome the stated limitation.

Although a lot of work have been done to find a suitable algorithm to deal with the varying head constraint, few works has been done to evaluate a classification algorithm under the constraint of varying head poses. Philips (1999) applied support vector machines to image recognition by developing an algorithm to compare with the well-known principal component analysis. The comparison of the two algorithms was done based on images obtained from the Facial Recognition Technology (FERET) database and assessment of the performance was done on the back of verification and identification scenarios. The results obtained showed that support vector machines performed better (77-78%) as compared to 54% for PCA, on the basis of identification.

Guo, Li & Chan (2001) concluded that SVM achieves higher accuracy as an algorithm for face recognition. Asiedu, Adebajji, Oduro and Mettle (2016) also concluded in their work that, the recognition rates of PCA/SVD and FFT-PCA/SVD are 88.49% and 89.29% respectively, under the constraint of variable facial expressions. Various algorithms have been proffered for the purpose of recognizing images, which have their rates of recognition to be above 90% under conditions that are expedient. (Jain, Ross & Prabhakar, 2004; El-Bakry & Mastorakis, 2009, Zhao, Chellapa, Philips & Rosenfield, 2003; Chellapa, Sinha & Philips, 2010).

From the few examples above it is clearly observed that different algorithms when applied to different data and different constraints produce different results in terms of performance. Hence there is no clear evidence which algorithm is best under what constraint. Also, little work has been

done which combines the two approaches; Principal Component Analysis and Support Vector Machines (PCA/SVM). Thus, the study tries to fill the gap by assessing the performance of support vector machines, combined with principal component analysis (FFT-PCA/SVM), under varying head tilts.

1.3 Objectives

This study seeks to assess the classification performance of support vector machines (SVM), combined with Principal Component Analysis algorithm under varying head poses. More specifically, the study seeks to

- Train an SVM model for classification using Matlab
- Assess the algorithm's (model) performance under the constraint of head poses through the computation of the classification rate and runtimes.
- Compute the accuracy of the model under various head tilt classifications.

1.4 Significance of The Study

Developing an efficient and resilient face recognition technique is worthwhile due to the vast number of application areas. These include criminal identification, access management, law enforcement, information security and entertainment or leisure. There has been an increase in the interest of people in facial recognition in recent times for differing reasons. These include but not limited to the rise in security matters to the general public, how identity verification has become very important in a technology-driven world and the exigency of analyzing images and modeling methods in data management.

Key among the factors responsible for driving facial recognition is the never-ending advantages that a strong and an efficient recognition procedure presents; for instance, identifying missing children, voter identification and verification, security set-ups which are dependent on biometric

data and user-friendly human-machine interfaces. A classic illustration of the latter categorization is smart rooms, which usually make use of cameras and sometimes microphones assemblage to discern human presence, reach a decision on their identity and acts thereafter based on a predisposition established in advance for each situation. Also, in banking ATMs, the technique is applied to identify and verify a customer’s identity. A few other areas of application are as shown in Table 1.1

Table 1.1: Some Application Areas of Face Recognition System

Areas	Application of Face Recognition
Access Management	Secure Access Authentication, Personal Identification (National ID, Passport, Drivers’ License) Welfare fraud, Voter registration, Immigration, Multimedia communication (synthetic faces)
Information Security	Intranet security, Application security, File encryption, Data Piracy, TV-Parental control, Personal device log on, Internet access, Medical records, Accessing security, Securing trading terminal
Biometric	Advanced video surveillance, Forensic reconstruction of face remains, CCTV, Shop lifting, Portal control
Criminal Justice System	Post-event analysis and forensics.
Residential Security	Alert homeowners of approaching personnel
Entertainment, Leisure	Human-robot interaction, Human-computer-interaction and home video games

According to Duda, Hart & Stork, (2001), the present-day breakthrough image analysis, pattern recognition and machine learning has made it very practicable to come up with automatic image recognition techniques to deal with such applications. Improving the recognition procedure or rate

warrants an improvement in the areas of information security, biometric identification, access management and law enforcement. This study, therefore seeks to assess the performance of two combined algorithms which are mostly used: PCA and SVM for classification. Also, since face recognition is multifaceted phenomenon and no one research is capable of addressing it in full, this work will serve as a catalyst to breed interest and further research into the other aspects of biometrics.

1.5 Brief Methodology

The study will employ a secondary database which would be extracted from the Massachusetts Institute of Technology (2003-2005) Database. This database consists of various head tilts of individuals (both male and female) in a morphable model. The head tilt image data would serve as an input for the proposed model. From literature it has been shown that face recognition is essentially executed in three sequential phases, namely the pre-processing phase, extraction (dimension reduction) of facial features phase and classification (recognition) phase. The procedure is as demonstrated in Figure 1.1



Figure 1.1: The Classification Process

In order to assess the classification performance of the model, numerical evaluations such as classification rate and the average runtime would be employed. The classification rate of the algorithm is stated as the ratio of the total number of face images correctly classified by the algorithm to the total number of face images in the test dataset for each experimental run. On the

other hand, the runtime of the algorithm basically constitutes the amount of time it takes for the algorithm to execute the task of recognition.

1.6 Scope

This study will rely on recognizing (classifying) tilted face images in the adopted database under varying head pose, to evaluate the algorithm (model) under consideration.

1.7 Organization of the Study

This study contains five main chapters. In chapter one, the concept of face recognition is introduced. In Chapter two, a comprehensive theoretical framework, review of the related literature in the field which serves as the basis for the theoretical framework for the study is discussed. Chapter three presents a detailed outline of the two algorithms; SVM and PCA, and how they are applied to the constraint of varying head poses under study. In Chapter four, interpretive critique and analysis of the outcome of the study are presented. The final chapter (five) of this work gives the conclusion of the entire study by stating the way forward for further study based on the major findings made in the study.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

The focus of this chapter is to assess various literature in the area of Facial Recognition and Classification, SVM and PCA/SVD. Major theories, arguments, methodologies, approaches and controversies in the existing literature on the subject of this study are discussed in this chapter. Also, critical attention and focus will be on the empirical and theoretical appraisal of articles, journals and related issues on face recognition using SVM and PCA/SVD.

2.2 Review of Works on Face Recognition

The history of facial recognition can be traced to the late nineteenth century where Darwin (1872), initiated the earliest study on facial expression of emotions in both man and animals. This was followed by Galton (1892), who also worked on profile-based biometrics aimed at personal identification. His study discovered an independent feature suitable for hereditary investigation that aided personal identification. However, the study was limited only to parentage and next of kingship, which failed to address the variation in similarity image, pose, illumination, occlusion, etc. Automated face recognition was developed around the 1950's. Bruner and Tagiuri (1954) pioneered the earliest research studies into the automated facial recognition system in a psychology research. In the 1970s, Kelly (1970) and Kanade (1973) presented seminars in which semi-automated face recognition technique were adopted to map a person's face image onto a global template using a facial template scheme. However, their method required high computational time, because locations and measurements were done manually. Literature has revealed that after the studies of Kelly (1970) and Kanade (1973), research work related to face recognition system became dormant until the mid-1980s, when researchers concentrated on automated face

recognition techniques which employed facial component (such as the eye, nose and mouth) to recognize an individual image. Majority of the work in automated face detection and face recognition concentrated on detecting individuals' facial attributes like the eyes, nose, mouth etc., and describing a model based on the size, position and relationship among such facial characteristics, with some emanating from correlation-based procedures, neural networks and shape matching from range (Patel, Chellapa, Sankaranarayanan & Chen, 2015). It was during this era also, that a roughly absolutely statistical method led to the premier facial recognition system that is fully automated.

The work by Kohonen (1989) remains one of the famous exemplification of facial recognition systems. Kohonen in his work was able to show how simple neural networks could execute face recognition for a situated and normalized facial image. He also employed a computer image descriptor to approximate eigenvectors of the image autocorrelation matrix. Such eigenvectors are referred to as the Eigen faces. Kohonen's system seemed good because of the need for a definite and accurate alignment and normalization but the system was not practically successful. Other researchers, since then, have tried other recognition schemes rooted on edge inter-feature distances and other neural network methods. In as much as a few of the approaches were very successful on smaller databases of aligned facial images, not one of them was able to successfully solve the more pragmatic challenge of large database with unknown scale and position of the face (Kohonen, 1989).

Principal component analysis was introduced as an algebraic manipulation to face recognition problem by Kirby and Sirovich (1990). This made the direct computation of Eigen faces simple and very easy. The limitation of this technique was that, in order to correctly code face images that are suitably aligned and normalized, it required less than 100 values. In 1991, Turk and Pentland

found out and also demonstrated that while using the technique of Eigen faces, the residual error could be utilized to recognize faces in jumbled natural imagery and to ascertain the exact scale and location of faces in the image. Turk and Pentland again demonstrated that this method of facial detection coupled with localizing faces with the Eigen face recognition method could lead to achieving an authentic real-time recognition of faces in an environment that is extremely restrained. Although this approach was to a certain extent restricted by environmental constraints, it still generated notable interests in advancing the developments of automated technologies in facial recognition.

Research in the area of facial recognition received a major boost after the work by Turk and Pentland (1991) on Eigen faces. Some of the successes chalked in the area of facial recognition comprise: Fisher face approach, which employed Linear Discriminant Analysis (LDA) after performing PCA to obtain a high precision; the utilization of local filters like the Gabor jets that yield facial characteristics which are more efficient and effective and the drawing of AdaBoost learning based cascade classifier designed for actual time facial detection. The techniques for facial recognition have advanced remarkably since that period when the procedure of Eigen faces was put forward. In strained environments, for instance, situations in which facial appearance, lighting and pose are managed, automated facial recognition outclass human recognition achievement, particularly where the data base has a sizeable number of faces. Notwithstanding, the process of automated face recognition is still faced with numerous issues when the face images are obtained in unrestrained situations. For the past four decades, researchers have concentrated on making face recognition approaches wholly automatic, by tackling the setbacks like the localization of faces in given images and extrication of characteristics. The attempt to design classifiers for successful facial recognition algorithms under various constraints has seen tremendous improvement but for

ageing and angular image (Zhao et al., 2003). Various techniques have been developed in separate fields like Statistics, Engineering, Computer Science, Psychology, etc. Notable among these are PCA, ICA, LDA, Neural Network and their derivatives (Asiedu et al., 2015).

2.2.1 Face Recognition Approaches

A greater number of the approaches that were proffered at the initial periods of computer vision were not seen to be a success. However, nearly all of contemporary techniques to the problem of facial recognition have, to a large extent been very creditable. From literature, facial recognition algorithms are generally grouped into two classes:

- Image template-based
- Geometric feature-based.

2.2.2 The Template Matching Approach

The template-based methods (Appearance-based method) utilize the facial feature of the entire face to perform face recognition. Template-based methods calculate a measure of correlation between original faces and a set of template models to assess the face identity (Guillamet & Vitria, 2002). With the template-based technique, the face is regarded as a two-dimensional array of intensity values which are then juxtaposed using an appropriate measure like the Euclidean distance, where each template denotes the entire face (Brunelli and Poggio 1993). Some of the examples of the template-based approach include the Linear Discriminant Analysis (LDA), Non-Negative Matrix Factorization (NMF), Independent Component Analysis (ICA), Principal Component Analysis (PCA), and the Singular value Decomposition (SVD). These algorithms are typically dimensional reduction (feature extraction technique) algorithms which seek to lessen the dimensions of the face space, thereby bringing out only important points on the face for recognition

and identification purposes. In facial recognition works, the dimensions of the face images are generally huge and thus needs a substantial amount of computational time (runtime) for recognition (classification) (Thakur, Sing, Basu, Nasipuri & Kundu, 2008). One problem associated with this method is computational intricacy. An additional drawback is with the description of the templates. Because the recognition procedure will have to accommodate some dissimilarity between the template and test images, the tolerance might average out the dissimilarities that will make individuals (images) distinct (Tolba, El-Baz & El-Harby, 2006).

2.2.3 Face Recognition Using Geometrical Features

The geometric (feature-based) methods analyze the localized facial components and establish a geometrical relationship amongst them. The geometric facial features include the eye brow, mouth, chin, cheeks, etc. This approach basically involves the calculation of a set of geometrical characteristics such as the length and width of the nose, the position of the mouth and the shape of the chin etc. from the facial images under consideration. This set of characteristics obtained are subsequently compared with that of known faces. An appropriate metric like the Euclidean distance is used to determine the nearest match. A classic illustration of such method is the Elastic Bunch Graph Modeling algorithm. It is known from literature that the geometric approach employs a set of fiducial points for image recognition (Marques', 2010). Notwithstanding the fact that the image cannot be observed in detail, the comprehensive feature configuration can be extricated for recognition. An advantage of using the feature-based (geometrical) method is that, recognition is very feasible even with noisy images and at resolution levels that are low. The main disadvantage, on the other hand is that automated extraction of the feature-based characteristics is extremely rigid. According to Brunelli and Poggio (1993), the automatic extraction of the geometrical features is also sensitive to rotation and scaling of faces in the image plane. Consequently, this

approach is less often used since there are developed algorithms that use the whole face (template-based approach). There is no doubt that all existing techniques have limitations. More efforts are thereby needed to ensure that there is an improvement in the facial recognition approaches, specifically in the vast range of environmental constraints that exist in reality (Tolba et al., 2006).

2.3 Pre-processing Techniques in Face Recognition

It is an established fact, that facial recognition algorithms have made important breakthroughs in recent times, but there is still the need for improvement. The areas that make use of pre-processing methods are those with a high potential for improvement. The quality of an image, no doubt plays a very crucial part in improving face recognition rate. All things being equal, an image with a good quality will provide a better recognition rate, as opposed to a noisy image. Current research works on facial recognition have proved that devices that are utilized in capturing images like the video camera, also affects the rate of recognition as they capture noisy and low-quality images as a result of their poor quality. When such situations occur, the images that have been captured tends to have contrast variation, noise, low resolution and differing brightness throughout the image. Undoubtedly, it is not very easy to extricate the characteristics from the noisy images which tends to diminish the recognition rate. Thus, pre-processing of facial images is done before extracting the features so as to overcome the challenges that arises from noisy and low-quality images.

Pre-processing of images has the potential of extensively maximizing the trustworthiness of any optical inspection. A lot of techniques for filtering that further increase or decrease the information obtained from an image also allow for faster and easier evaluation. Camera images can thus be optimized with just a few clicks. According to Anila and Devarajan (2012), pre-processing of

images tends to reduce the processing time while increasing the possibility of an accurate matching. Different preprocessing mechanisms exist that can be utilized to ensure the enhancement of the images that are captured and thereby improve the rate of recognition. Peng & Xue (2005) asserts that the quality of an image and the rate of recognition can be greatly improved by the use of techniques such as filtering, image resizing, image normalization, histogram equalization, de-noising and accurate image detection. Therefore, pre-processing of images is done before features are extracted from the image. There exist a lot of preprocessing techniques (algorithms) such as Mean Centering, Discrete Wave Transform (DWT) and Fast Fourier Transform (FFT). For instance, Asiedu et al., (2015) compared FFT-PCA/SVD with PCA/SVD, where the constraint was varied facial expressions and concluded that the former algorithm had low variation and a high recognition rate than the latter algorithm. This was attributed to the preprocessing technique and therefore suggested FFT as a mechanism to be employed for de-noising. The study as a result employed FFT as a pre-processing mechanism.

2.4 Feature Extraction Methods

Raw face images usually take a longer period of time to recognize, which basically results from the fact that it has a huge number of pixels. There is therefore the need for one to reduce the number of pixels to increase recognition. This is what is known as dimensionality reduction, which has the primary purpose of saving time for the deep step of decision. Extraction of features basically refers to the transformation of face space into feature space. The image database of the feature space is represented by a smaller number of characteristics from the original face (Haykin, 2009). Turk and Pentland (1991) asserts that the most commonly used method of doing this is through the application of the Eigenfaces algorithm. Bellakhdhar, Loukil, and Abid (2013) have stated that face recognition algorithm performance depends on how it extracts feature vector and recognize a

facial image correctly. The feature extraction method is a vital tool in determining the performance of recognition rate of a face recognition algorithm. Some of the widely used methods of feature extraction include; Independent Component Analysis (ICA), Neural Networks, Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) etc. In this section, the study will principally focus on PCA/SVD as the face dimensional reduction (feature extraction) mechanism

2.5 Support Vector Machines (SVM)

2.5.1 Theoretical Approach

SVM is a widely used binary classification method. With classification systems that are based on SVMs, the classification is carried out by drawing a hyper-plane that yields the greatest segregation of the data points. Such a hyper-plane is termed as the optimal separating hyperplane. With this, the bound of the hyperplane will be at a maximum. With a set of observations belonging to two distinct classes, SVM identifies a hyperplane that provides the largest fraction of separation of points which belongs to the same class, and at the same time maximize the distance from either of the classes to the hyperplane. The vectors which are nearest to the hyperplane are referred to as the support vectors. When dealing with classification situations, SVMs present a way for image recognition. The technique is normally applied to the space that is derived subsequent to the feature extraction procedure and the image space can be the original appearance space.

The SVM algorithm takes as an input a set $\{(x_i, y_i)\}$ of labeled training data, where each x_i is the data and $y_i = -1$ or 1 is the label. A set of N support vectors s_i is obtained as the output of an SVM algorithm, with α_i as coefficient weights, y_i as class labels and b as a constant. For a linearly separable dataset, the decision surface can be stated as;

$$w \cdot x + b = 0, \tag{2.1}$$

where

$$w = \sum_{i=1}^N \alpha_i y_i s_i \quad (2.2)$$

This construction can be extended to the situation where the separation surfaces are non-linear.

Every observation in the input space is non-linearly projected onto a high-dimensional feature space. Zagouras, Economou, Macedonas & Fotopoulos, (2007) asserts that such a mapping is executed using kernel functions. Every single observation x in the input space is projected onto another point $s = \Phi(x)$ of a higher dimensional space, known as the feature space, such that observations are differentiated by a hyperplane. The indispensable characteristic such a method is that the projection $\Phi(\cdot)$ is contingent on the fact that the dot product of two observations in the feature space $\Phi(x) \cdot \Phi(y)$ is able to be captured as a kernel function $K(x, y)$. Thus, the decision boundary will have the equation:

$$\sum_{i=1}^N \alpha_i y_i K(s_i, x) + b = 0 \quad (2.3)$$

Here too, α_i and b are coefficients and also the solutions of a quadratic programming problem.

2.5.2 Empirical Approach

Philips, (1999) compared an SVM-based technique with a PCA-based algorithm on a set of images obtained from the FERET database. The research took into consideration the dissimilarities between two faces of the same individual and the variations between the faces of separate individuals. The two classes were used as inputs to the SVM algorithm and performance was evaluated for both identification and verification situations. It is a well-known fact in face recognition that real application depends on the algorithm's ability to generalize from a training set to images independent of the training set. Comparing the two algorithms, it turned out that for

identification; the performance of SVM was better (77-78%) than that of PCA (54%). Verification performance was also measured, and SVM had an error rate of 7% while PCA had the error rate to be 13%. Considering identification, the error rate of SVM was half (22-23%) as compared to that of PCA, which was 46%. This, according to Philips indicated that SVM was more efficient in using information in the face space as compared to the PCA. The research work however, did not take into consideration the constraint of pose variation.

Guo et al., (2001) also compared SVM based recognition with two other algorithms: the standard eigenface and the nearest feature line (NFL) approaches. Evaluation of the algorithms was done using two databases, the Cambridge ORL and a compound database. The error rates for the various algorithms considered were SVM (3.0%, 8.79%), NC (5.25%, 15.14%) and NFL (3.125%, 9.72%) for the two databases respectively. The SVM based algorithm outperformed the rest as seen in the error rates reported above and as a result can achieve high accuracy. Although the constraints included variable expression, and pose, the mode of assessment was only numerical and thus the need for this work to add to that a statistical assessment.

Heisele et al (2003), compared the component-based approach with two global approaches and facial recognition and assessed their performance on the merit of robustness against pose changes. With the initial global system, a sole SVM was employed for each individual in the system while the database was clustered and view specific SVMs were trained for each person in the second. With the component-based method, facial components were extracted and presented in a single vector which would then be classified by SVMs with a linear kernel. However, a single SVM was trained for every individual in the database while a set of SVM classifiers were trained for clustered images of an individual for the two global approaches respectively. Extensive tests were performed on a test set which contained images rotated up to 40°. The conclusion was that, despite

the fact that more powerful classifiers (non-linear rather than linear) were used for the global system as compared to that of the component-based system; the latter performed better than the two global systems. They therefore concluded that it is very useful to employ the face components instead of the entire face as the former makes the task of face recognition significantly simpler. The challenge that was associated with the system was that, although there was some level of pose invariance, the classifier was unable to deal completely with the comprehensive array of poses (from frontal to profile views)

According to Déniz, Castrillon & Hernandez, (2003), two other powerful and significant techniques which are being used recently are the SVM and the ICA. Thus, they combined these two techniques and performed experiments using two databases of which they concluded that the recognition rates achieved was high. This conclusion was arrived at after they compared the performance of ICA/SVM with others such as ICA/NMC and PCA/SVM on the two databases used: Yale Face Database and AR Face Database. It was observed from the experiments that the performance of PCA/SVM was very close to that of ICA/SVM and that SVM was not very sensitive to the representation space. As a result of this observation and the fact that PCA is a fast feature extraction technique as compared to ICA, they posited that it is practically best to combine PCA and SVM.

This assertion of combining PCA and SVM was also corroborated by Pal and Foody (2010). According to the duo, series of experiments performed with two sets of hyperspectral data did not support the argument that SVM was insensitive to the data dimensionality. They found out that the accuracy of classification by the method of SVM was highly dependent on the dimensionality of the data. They therefore concluded that with regards to face recognition, better classification accuracy would be achieved when PCA is combined with SVM.

Frolov and Sadykhov (2008) also employed support vector machines (SVM) as a face recognition technique. The input feature vector that contained the most significant coefficient was employed for the training and classification. This was very possible after they had reduced the data space using the discrete wavelet transform algorithm which they assert that enabled them to select the coefficients which were most significant to describe the area of interest. To test the algorithm based on support vector machines, 611 images of size 256-by-384 pixels were used and the conclusion was that their system (SVM) was 84.28% valid. Their work, however, only took into consideration low image quality as a constraint but failed to consider other issues including illumination and pose constraints of which this work seeks to explore.

Olivares-Mercado, Aguilar, Toscano-Medina, Nakano & Meana, (2011) posited that support vector machines (SVM) generally performed better than the Gaussian Mixture Model (GMM), especially with a relatively small training set. They compared the two models for face recognition and verification, with the constraints being occlusion, variable facial expressions and illumination surroundings. Verification and Evaluation of the two systems were compared for various feature extraction methods such as the Gabor-based, Wavelet-based and Eigenphases. Using two data sets, one with occlusion (using scarf and sunglasses) and the other without, they found out that the identification performance decreased for the one with occlusion as compared to the other without occlusion under the various feature extraction methods. This according to them was as a result of the fact that there was not much information about the effects of occlusion. Hence it was observed that when the database is constant and also not too large, the SVM is most preferred but GMM is to be used in case of large database that changes with time.

2.6 Principal Component Analysis (PCA)

2.6.1 Theoretical Approach

Principal Component Analysis can be mathematically described as an orthogonal linear transform which maps to a new coordinate system so that the maximum variability by any projection of data will be on the first coordinate (referred to as first principal component), the variability on the second coordinate, in that order. Thus, PCA is theoretically thought of as the optimal transform for any dataset in least square terms. The Eigenface technique is one of the mostly used algorithms in facial recognition as a result of its straightforwardness and easy implementation. PCA is an approach for dimension reduction and feature selection and is also termed as the Karhunen–Loève procedure. From literature, it can be seen that PCA is one of the popularly accepted approaches in the field of image detection. PCA was proposed first by Karl Pearson in 1901 and in 1933 by Harold Hotelling to change any given set of variables which are practicably orthogonal into a smaller set of variables which are uncorrelated. (Turk & Pentland, 1991). The fundamental objective of a face recognition system is to construct and convert two-dimensional facial image into one-dimension vector of values consisting of compact principal component of the feature space. This idea implies that, PCA determines the vectors which most account for the face image distribution within the whole image space and seeks to bring out a subspace where the variance is maximised (Moghaddam & Pentland, 1997). PCA makes use of an orthogonal projection to transform a set of variables which are correlated to a new set of variables that are uncorrelated.

The PCA procedure finds the directions which have the maximum variance in the data, known as principal components. It is not only for dimensionality reduction, but it also keeps some of the variability in the data and produce a very solid characterization of an image (Turk & Pentland,

1991). Thus, PCA can also be described as low dimensional subspace of an image in the shortest period of time. Generally, the main purpose of the PCA approach is to change a face image into a smaller set of characteristic features referred to as the eigen face which represents the principal components of the original train data. The process of using PCA for facial recognition which expresses the large 1-D vector pixels created from the 2-D face images into the compact principal components of the feature space is known as the eigen space projection (Kim, 2000). The eigen space is calculated by ascertaining the eigen vectors of the covariance matrix which is gotten from the set of images (vectors).

In facial recognition systems, the PCA is seen as a statistical method that is adopted to explain the variance-covariance structure of an image dataset. Based on the variance-covariance matrix generated from the image dataset, the PCA technique is employed to compute the eigenvalues and eigenvectors respectively, and to project the original image datasets onto a reduced dimensional feature space. This involves a numerical procedure that transforms an orthogonal basis vector of a set of correlated variables into an uncorrelated form called principal component, which maximises the scatter of all the projected samples.

2.6.2 Empirical Approach

Majority of scholarly works in the field of image recognition have employed PCA as dimension reduction mechanism. The study considered a few of them which are discussed below.

Kirby and Sirovich (1987) are known to be the leading people to have employed PCA to economically denote face images. They were able to show that given any distinct image, it can be thoroughly represented along the standard face picture coordinate space, and so every face image can be rebuilt using a smaller set of eigen pictures and the corresponding transformations along each eigen picture.

Nayar and Murase (1995) also utilized principal components of numerous perspectives of a lone object to visualize the high-dimensional manifold expounded by changes as a result of illumination conditions and rotation in depth. The pose of the object would then be ascertained depending on its position on the manifold. The study also employed PCA in a like manner to scrutinize the distribution of a pose in a high-dimensional space.

According to Chung, Nairn, Murata & Brautigam, (1999) it was better to have a combination of Gabor filters and PCA. Their approach consisted of two sections, where Gabor filters were utilized in the extraction of facial characteristics from the primary image on fiducial points that are predefined in the first scenario. When it came to the second phase, PCA was employed to group the facial features optimally. The motivation for their suggestion of combining the two techniques was to repress the deficiency in PCA. Their argument was that whenever the unprocessed images are employed as matrices for PCA, the eigen space cannot mirror the correlation of the image so well, since the primary faces have distortions resulting from in-depth rotation, in-plane, contrast and illumination variability.

In 2008, Nicholl and Amira adopted the PCA (eigenfaces) approach in face recognition. They presented DWT/PCA to automatically calculate the most discriminative coefficients in a face image. The most discriminative coefficient was determined based on their respective intra-class and inter-class variations of the training set of eigenface weighted vectors. Even though the method improved the performance of face recognition algorithm, the study failed to address the pose variation and illumination in the training face image datasets.

Abdullah, Wazzan, and Bo-Saeed (2012) employed PCA (eigenvalues) as a data representation and dimension reduction technique in solving face recognition problems. The objective of their

enquiry was to measure the out-turn of run time in the performance of facial recognition algorithms. The outcome showed that for small database size the eigenfaces approach has no relationship with time complexity in face recognition. However, this approach has drawback of high computation time especially for large database size

Paul and Al Sumam (2012) focused on applying statistical approach to construct a face recognition system. They employed the PCA approach to minimize the high-dimensional space to an intrinsically low- dimensional space. In order to do that, the dataset was divided into train and test sets. The eigenvalues (eigenvectors) were deduced from the covariance matrix of the training dataset and the eigenfaces measured the combination (linear) of the weighted eigenvectors. Recognition was achieved by extending a face onto the sub-space covered by the eigen faces. The test images were projected onto the sub-space spanned by the eigen faces and classification was performed afterwards by evaluating the least possible Euclidean distance. However, this study failed to address the problem of real time in recognition rate.

Dhoke and Parsai (2014) performed facial recognition with PCA and Back Propagation Neural Network (BPNN) for identification and verification purposes. Their study adopted PCA as a dimensional reduction technique (feature extraction) and BPNN was used for recognition and classification. The results revealed that their face image recognition and classification was computationally very fast and provides high precision rate.

Asiedu et al., (2015) employed statistical techniques to evaluate the performance of PCA/SVD and FFT-PCA/SVD, under the constraint of varying expressions. The study findings indicated that FFT-PCA/SVD relatively exhibited low variability and a better recognition rate than the

PCA/SVD algorithm. The researchers posited that FFT is a viable noise filtering technique that should be employed in the pre-processing stage of face recognition.

Goyal and Batra (2016) adopted PCA with neural network in a face recognition system and applied photometric normalization for comparison. The study results indicated that Euclidean distance classifier gives a high precision using the primal face images, although employing histogram equalization methods on the image does not yield appreciable effect to the performance of the face recognition approach under unconstrained surroundings.

2.7 Singular Value Decomposition (SVD)

Singular value decomposition can be considered from three different angles which are mutually compatible. Firstly, it can be seen as an approach used for converting variables which are correlated to uncorrelated variables so as to reveal all the relations among the primary data points. Secondly, SVD is a technique that is used for locating and ordering the dimensions through which observations display the greater variability. This ties in to the third and final way of considering SVD, which is that, as soon as the greatest variability has been identified, it becomes feasible to identify the best estimation of primary data items with fewer dimensions (Baker 2005). According to Hong (1991), the characteristics of an image can be categorized into four namely, the transform coefficient, statistical pixel, visual and the algebraic characteristics. The algebraic characteristics consist of the intrinsic features of a face, having an excellent stability. According to Hong, the algebraic characteristics of the image are genuine characteristics in object recognition such as facial recognition, and thus suggested a recognition method which is SVD based. The efficiency and effectiveness of the Singular Value Decomposition has been tested in (Hong, 1991) and (Yong-Qing, 1991) respectively. In Hong (1991), 42.47% was recorded as the error rate and this

was attributed to the statistical limitations of small samples. Cheng (Yong-Qing, 1991) also suggested a technique facial recognition which was formed on the statistical model of a small sample whereby the singular values were used as features of a face. Hong constructed an optimal discriminate transformation to convert the primary space of singular value (SV) vectors into a new space with a considerably lower dimension as opposed to the original space, so as to reduce the consequence that would arise as a result of small samples. To test this approach, 64 images of eight individuals were used and a very good ability of discrimination was achieved with an accuracy rate of 100% (Yong-Qing, 1991).

2.8 Conclusion

Currently, the problems of face recognition continue to attract more research effort from the scientific community, as a result of the diverse implementation of facial recognition techniques in the real world. Although facial recognition is known to be a difficult task for frontal face images, certain algorithm has been proposed under constrained environments. Literature has revealed that the eigen face technique is most prominent in tackling face recognition problem. However, most of the performances of face recognition algorithms are limited to constrained environments such as head pose, occlusion, etc. Most of the previous works done on face recognition relied only on a single frontal-pose model (Chen, Patel, Sankaranarayanan & Challepa, 2015) possibly normalizing images through frontalization (Taigman, Yang, Ranzato & Wolf 2014). Also, literature has revealed that there is little work done at assessing the performance of the combination of PCA and SVM for classification with head tilts as the constraint. From the discussions above, it is evidently clear that both SVM and PCA/SVD are powerful techniques employed in the field of facial recognition. This study therefore seeks to evaluate the classification performance of FFT-PCA/SVM algorithm on varying head tilts.

CHAPTER THREE

METHODOLOGY

3.1 Introduction

The chapter is dedicated to the methodology that was employed in this study. It focused on a detailed and comprehensive understanding of FFT, PCA, SVD and SVM. Among the aspects that came under discussion included the source of data, the methodologies used in the image processing, methods of analysing the available data, theories and concepts used in order to obtain the study objectives. The study employed three (3) phases in the data analysis of the facial recognition procedure comprising: FFT as a noise filtering in the pre-processing phase, PCA/SVD as feature extraction (dimension reduction technique) phase and SVM for training the classification model. Also, SVM model was used to classify a new image into the appropriate class or group.

3.2 Data Acquisition

The study employed a secondary database which was extracted from Massachusetts Institute of Technology (2003-2005) database. This database is made up of the varying head tilts of individuals' images.

Plate 3.1 shows the original images of individuals used in the study. The head tilts of each individual were captured along $0^\circ, 4^\circ, 8^\circ, 12^\circ, 16^\circ, 20^\circ, 24^\circ, 28^\circ$ and 32° rotation of the head.



Plate 3.1: Display of the ten Original Frontal Images of the dataset

Source: [Weyrauch, Huang, Heisele & Blanz (2004)]

3.3 Classification Procedure

The research concentrated on employing FFT-PCA/SVD and SVM classification technique on a database from MIT under the constraints varying head tilts. The database comprising of facial images (in both the training and testing datasets) served as input images for the face classification module. The inputted facial images are operationally transformed into uniform dimension and compatible format for image processing. The research design for FFT-PCA/SVD is illustrated in Figure 3.1

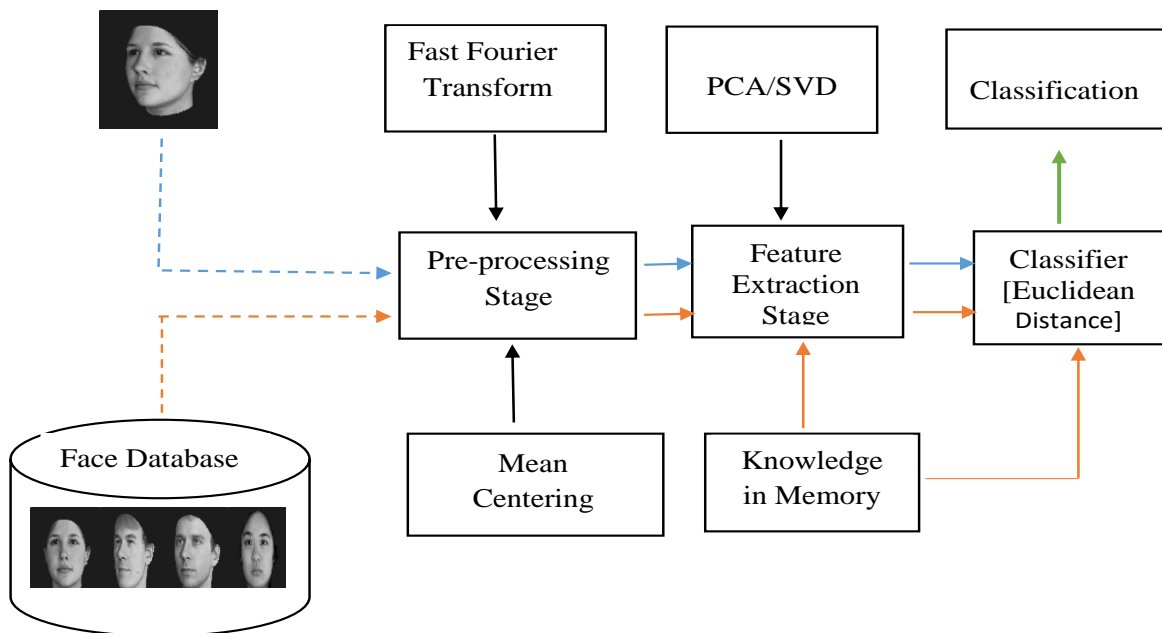


Figure 3.1 Research Design for FFT-PCA/SVD.

3.3.1 Pre-processing Stage

The principle behind using a pre-processing tool is to increase the quality of the face image in order to achieve a good recognition performance of the algorithm. Pre-processing stage is a useful phase in face image representation and serves as a noise removal mechanism as stated by Asiedu et al. (2015). Pre-processing stage is an effective way of suppressing the unwanted distortion of image feature for further processing. This helps to drastically minimize the level of noise contained in an image data set and also makes the estimation process simpler and better conditioned for an improved recognition rate. From the suggestion of Asiedu et al., (2015), the study employed the Mean Centring and Fast Fourier Transform (FFT) at pre-processing stage as de-noising/filtering technique before extracting features from face images. The study adopted two stages in the pre-processing phase namely:

- ❖ Resizing of the face image
- ❖ Denoising of the face image

3.3.1.1 Resizing of the face image

Resizing involves reshaping the images to a preferred size. Each face image from the database was imported into the Matlab software. The face images were resized into a uniform dimension of 100×100 . Reddy, Babu & Kishore (2011) stated that the resizing of face image helps to reduce the lighting effect associated with visual images thus makes them best suited for PCA in image processing. The study adopted this procedure to lessen the mathematical complexity in the feature extraction of a face image.

3.3.1.2 Image De-noising

Many researchers (Yang, Kriegman & Ahuja, 2002; Zhang, Verma, Tjondronegoro & Chandran 2018 etc.) have revealed that face images by nature exhibit the characteristics of a Gaussian noise due to the presence of illumination variations. For a face image in the dataset to be de-noised (filtered), the study adopted a pixel-based filtering technique. In this study, mean centring and Fast Fourier Transform (FFT) were applied to eliminate noise in the face image and retain vital facial feature for recognition. FFT according to Asiedu et al., (2015) has proven to be a viable tool for removing noise in an image data.

3.3.3 Mean Centring

The Mean Centred face is deemed as the arithmetic mean of the training image vectors at each pixel point, having its dimensional size to be $S \times 1$.

Let the image matrix H_k be defined as

$$H_k = (h_{klm}); \quad l, m = 1, 2, \dots, p; \quad k = 1, 2, \dots, n,$$

$$\mathbf{H}_k = (h_{k1}, h_{k2} \dots h_{kp}), \text{ where } h_{km} = (h_{k1m}, h_{k2m}, \dots, h_{kpm})'$$

$$\mathbf{X}_k = (h'_{k1}, h'_{k2}, \dots, h'_{kp})', \quad (3.10)$$

where p is the order of the image matrix and n represents the number of face images to be trained.

From equation (3.10), let \mathbf{X}_k be a column vector of dimension S given by

$$\mathbf{X}_k = (X_{kl})_{S \times 1} \quad (3.11)$$

where the X_{kl} replaces the h_{klm} position-wise. The pre-processing phase is carried out based on the sample $X = (X_1, X_2, \dots, X_n)$, whose members constitute vectorised form of each image. The mean centring thus becomes a simple and easy pre-processing phase, performed by the difference in the mean. Let the training set be denoted by $X = (X_1, X_2, \dots, X_n)$.

Now $\bar{h}_k = E(X_k)$ of the data $(X_k, k = 1, 2 \dots n)$ through the data

$$\bar{h}_k = \frac{1}{S} \sum_{k=1}^S X_{kl} \quad (3.12)$$

$$\bar{h}_k = \frac{1}{S} \sum_{l=1}^S \sum_{m=1}^S X_{klm} \quad (k=1, 2 \dots n)$$

where $S = (p \times p)$, length = (row of image \times column of image) of the image data, H_k .

Let \bar{X}_k be defined as a constant column vector with order $(r \times r)$ having all members equivalent to that of $\bar{h}_k, (k = 1, 2 \dots n)$.

The mean (average) centring is represented by $Y = (y_1, y_2, \dots, y_n)$. Subtracting the mean image from the training image we obtain the equation below.

$$Y = X - \bar{X}_k \quad (3.13)$$

3.3.4 Fast Fourier Transform

A mechanism for noise reduction that is used in image pre-processing is the Fast Fourier Transform (FFT). The FFT is a computationally efficient algorithm, which is employed in calculating the Discrete Fourier Transform and its inverse (IDFT). That is, FFT is an algorithm that basically does the same thing as the DFT but in much lesser time. The FFT is faster than the DFT and is able to reduce the computational burden from $O(N^2)$ arithmetic operations to $O(N \log N)$ operations (Glynn, 2007). The FFT is faster as compared to the DFT because it has the ability to minimize the number of computations needed to examine an image data set. Thus, in dealing with large image datasets, it is prudent to employ FFT other than DFT.

In the process of performing FFT during image pre-processing, the first step is to compute the DFT. For a column vector h_{km} , the DFT is mathematically denoted as

$$h_{kgm}^* = DFT\{h_{km}\} = \sum_{f=0}^{p-1} h_{km} e^{-i\left(\frac{2\pi gf}{p}\right)}, \quad (3.14)$$

where, $g = 0, 1, \dots, p-1$, $j = 1, 2, \dots, n$ and $i = \sqrt{-1}$.

h_{km} is the k^{th} column of the image matrix H_k . For an image matrix of order 4, $p = 4$ and $g = 0, 1, 2, 3$. The DFT becomes;

$$h_{k0m}^* = h_{k0m} e^{-0.i\pi/2} + h_{k1m} e^{-0.i\pi/2} + h_{k2m} e^{-0.i\pi/2} + h_{k3m} e^{-0.i\pi/2}$$

$$h_{k1m}^* = h_{k0m} e^{-0.i\pi/2} + h_{k1m} e^{-1.i\pi/2} + h_{k2m} e^{-2.i\pi/2} + h_{k3m} e^{-3.i\pi/2}$$

$$h_{k2m}^* = h_{k0m} e^{-0.i\pi/2} + h_{k1m} e^{-2.i\pi/2} + h_{k2m} e^{-4.i\pi/2} + h_{k3m} e^{-6.i\pi/2}$$

$$h_{k3m}^* = h_{k0m} e^{-0.i\pi/2} + h_{k1m} e^{-3.i\pi/2} + h_{k2m} e^{-6.i\pi/2} + h_{k3m} e^{-9.i\pi/2}$$

Thus,

$$\begin{bmatrix} h_{k0m}^* \\ h_{k1m}^* \\ h_{k2m}^* \\ h_{k3m}^* \end{bmatrix} = \begin{bmatrix} e^{-0.i\pi/2} & e^{-0.i\pi/2} & e^{-0.i\pi/2} & e^{-0.i\pi/2} \\ e^{-0.i\pi/2} & e^{-1.i\pi/2} & e^{-2.i\pi/2} & e^{-3.i\pi/2} \\ e^{-0.i\pi/2} & e^{-2.i\pi/2} & e^{-4.i\pi/2} & e^{-6.i\pi/2} \\ e^{-0.i\pi/2} & e^{-3.i\pi/2} & e^{-6.i\pi/2} & e^{-9.i\pi/2} \end{bmatrix} \begin{bmatrix} h_{k0m} \\ h_{k1m} \\ h_{k2m} \\ h_{k3m} \end{bmatrix}$$

and

$$\begin{bmatrix} h_{k0m}^* \\ h_{k1m}^* \\ h_{k2m}^* \\ h_{k3m}^* \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -i & -1 & i \\ 1 & -1 & 1 & -1 \\ 1 & i & -1 & -i \end{bmatrix} \begin{bmatrix} h_{k0m} \\ h_{k1m} \\ h_{k2m} \\ h_{k3m} \end{bmatrix}$$

Next after this is to calculate the Inverse Discrete Fourier Transform (IDFT), which is given as

$$h_{km} = IDFT\{h_{km}^*\} = \frac{1}{p} \sum_{f=0}^{p-1} h_{km}^* e^{-i(\frac{2\pi gf}{p})} \quad (3.15)$$

$g = 0, 1, \dots, p-1$, $j = 1, 2, \dots, n$ and $i = \sqrt{-1}$.

When $p = 4$, the IDFT is given by

$$\begin{bmatrix} h_{k0m} \\ h_{k1m} \\ h_{k2m} \\ h_{k3m} \end{bmatrix} = \frac{1}{4} \begin{bmatrix} e^{0.i\pi/2} & e^{0.i\pi/2} & e^{0.i\pi/2} & e^{0.i\pi/2} \\ e^{0.i\pi/2} & e^{1.i\pi/2} & e^{2.i\pi/2} & e^{3.i\pi/2} \\ e^{0.i\pi/2} & e^{2.i\pi/2} & e^{4.i\pi/2} & e^{6.i\pi/2} \\ e^{0.i\pi/2} & e^{3.i\pi/2} & e^{6.i\pi/2} & e^{9.i\pi/2} \end{bmatrix} \begin{bmatrix} h_{k0m}^* \\ h_{k1m}^* \\ h_{k2m}^* \\ h_{k3m}^* \end{bmatrix},$$

and

$$\begin{bmatrix} h_{k0m} \\ h_{k1m} \\ h_{k2m} \\ h_{k3m} \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & 1 & -i \\ 1 & 1 & -1 & 1 \\ 1 & -i & 1 & i \end{bmatrix} \begin{bmatrix} h_{k0m}^* \\ h_{k1m}^* \\ h_{k2m}^* \\ h_{k3m}^* \end{bmatrix}$$

The images that have been transformed basically comprises of two parts; real and imaginary. The real parts of these images are extracted for the process of feature extraction whereas the imaginary parts are ignored and treated as noise.

3.4 The Process of Singular Valued Decomposition (SVD)

Singular Value Decomposition can be thought of as a numerical tool for analysing matrices. It is analogous to the well-known theory of matrix diagonalization. Singular Value Decomposition is concerned with the transformation of variables that are correlated into a set of uncorrelated ones which best display all the correlations among the original observations. SVD can similarly be viewed as a dimensionality reduction or data reduction method, and this is primarily because after identifying the most variation in the original data items, it is also able to find the approximation that best reduces the data items with fewer dimensions (Baker, 2005). The study considers SVD of a matrix \mathbf{X} defined as the factorization of \mathbf{X} into the product of three matrices. Let \mathbf{X} represent a $f \times g$ real valued-data of rank h , without loss of generality $f \geq g$, and hence $h \leq f$. Let \mathbf{A} and \mathbf{B}^T be orthonormal matrices where $\mathbf{A} = f \times g$ and $\mathbf{B} = g \times g$. Thus, supposing that matrix \mathbf{X} is real implies also that the singular values are basically real numbers, and thus matrix \mathbf{A} and matrix \mathbf{B} are also real. The equation for Singular Valued Decomposition of \mathbf{X} can be quoted as:

$$\mathbf{X} = \mathbf{A}\mathbf{\Sigma}\mathbf{B}' \quad (3.16)$$

where $\mathbf{\Sigma}$ is a matrix having diagonal entries such as $\sigma_{jj} \geq 0, \forall_j = 1, 2, 3, \dots, g$ and $\sigma_{11} \geq \sigma_{22} \geq \sigma_{33} \geq \dots, \geq \sigma_{gg}$.

Equation 3.17 display the matrix representation of the SVD,

$$\begin{array}{cccc}
 \mathbf{X} & & \mathbf{A} & & \mathbf{\Sigma} & & \mathbf{B}' \\
 \begin{bmatrix} x_{11} & \cdots & x_{1g} \\ \vdots & \ddots & \vdots \\ x_{f1} & \cdots & x_{fg} \end{bmatrix} & = & \begin{bmatrix} a_{11} & \cdots & a_{1g} \\ \vdots & \ddots & \vdots \\ a_{f1} & \cdots & a_{fg} \end{bmatrix} & \begin{bmatrix} \sigma_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{fg} \end{bmatrix} & \begin{bmatrix} b_{11} & \cdots & b_{1g} \\ \vdots & \ddots & \vdots \\ b_{f1} & \cdots & b_{fg} \end{bmatrix} \\
 f \times g & & f \times h & & h \times h & & h \times g
 \end{array} \tag{3.17}$$

3.4.1 Properties of SVD

The properties of SVD which were employed in this study are presented as follow:

- ❖ The singular values $\sigma_{11}, \sigma_{22}, \sigma_{33}, \dots, \sigma_{gg}$ are unique.
- ❖ $\mathbf{X}'\mathbf{X} = \mathbf{B}\mathbf{\Sigma}'\mathbf{A}'\mathbf{A}\mathbf{\Sigma}\mathbf{B}'$, and since \mathbf{A} orthogonal matrix, $\mathbf{A}'\mathbf{A} = \mathbf{I}$ such that \mathbf{I} is the identity matrix. Hence, $\mathbf{X}'\mathbf{X} = \mathbf{B}\mathbf{\Sigma}'\mathbf{\Sigma}\mathbf{B}'$ and this demonstrates the uniqueness of the SVD. The matrix \mathbf{A} and the matrix \mathbf{B} are calculated from the eigenvector of $\mathbf{X}'\mathbf{X}$ and $\mathbf{X}\mathbf{X}'$ respectively.
- ❖ The rank of the matrix \mathbf{X} is the same as the number of its non-zero singular values.

Theorem 3.1

Let \mathbf{X} be a matrix with dimension $f \times g$, ($f \geq g$).

Then

- ❖ The matrices $(\mathbf{X}'\mathbf{X})_{g \times g}$ and $(\mathbf{X}\mathbf{X}')_{f \times f}$ are symmetric with real and nonnegative eigenvalues.
- ❖ If λ is a nonzero eigen value of $\mathbf{X}'\mathbf{X}$ having the eigen vector \mathbf{X} , then λ too is an eigen value of $\mathbf{X}\mathbf{X}'$ with \mathbf{X} as the correspondent eigen vector. In other words, $\mathbf{X}'\mathbf{X}$ and $\mathbf{X}\mathbf{X}'$ have the same nonzero eigenvalues.

Proof 3.1

- ❖ $(\mathbf{X}'\mathbf{X})' = \mathbf{X}'(\mathbf{X}')' = \mathbf{X}'\mathbf{X}$ and so $\mathbf{X}'\mathbf{X}$ is symmetric. Similarly, $\mathbf{X}\mathbf{X}'$,

$$(\mathbf{X}\mathbf{X}')' = (\mathbf{X}')'\mathbf{X}' = \mathbf{X}\mathbf{X}'$$

Let \mathbf{y} be an eigenvector of $\mathbf{X}'\mathbf{X}$ corresponding to a nonzero eigenvalue λ . Then

$$\mathbf{X}'\mathbf{X}\mathbf{y} = \lambda\mathbf{y} \quad (3.18)$$

Multiplying through equation (3.18) on the left by \mathbf{y}' yields

$$\mathbf{y}'\mathbf{X}'\mathbf{X}\mathbf{y} = \lambda\mathbf{y}'\mathbf{y}$$

$$(\mathbf{X}\mathbf{y})'\mathbf{X}\mathbf{y} = \lambda\|\mathbf{y}\|^2$$

$$\|\mathbf{X}\mathbf{y}\|^2 = \lambda\|\mathbf{y}\|^2$$

$$\geq 0$$

Hence,

$$\lambda = \frac{\|\mathbf{X}\mathbf{y}\|^2}{\|\mathbf{y}\|^2} \geq 0$$

This holds for $\mathbf{X}\mathbf{X}'$.

❖ Multiplying through equation () on the left by \mathbf{X} gives

$$\mathbf{X}\mathbf{X}'(\mathbf{X}\mathbf{y}) = \lambda(\mathbf{X}\mathbf{y})$$

That is, $\mathbf{X}\mathbf{y}$ is an eigenvector of $\mathbf{X}\mathbf{X}'$ corresponding to the eigenvalue λ , where

\mathbf{y} is an eigenvector of $\mathbf{X}'\mathbf{X}$ corresponding to the eigenvalue λ . Hence $\mathbf{X}'\mathbf{X}$ and

$\mathbf{X}\mathbf{X}'$ have the same nonzero eigenvalue λ .

3.5.1 Principal Component Analysis (PCA)

In face recognition, facial images have very high dimensions which require a great deal of time for classification. For instance, two-dimensional $k \times k$ grayscale image will span $p = k^2$ -dimensional vector space. Thus, a (256×256) pixel image will lie in 65536-dimensional space, which is far too high for any calculation. In this study, the PCA technique adopted primarily focused on data reduction and interpretation which explains the variance-covariance structure of

any given set of variables though they are fewer linear combination of these variables. That is, PCA makes it possible to identify the salient directions in which the image data set varies. This procedure changed high dimensional image in every two-dimensional picture into a dimensional vector. The study concentration lies in the components that explain most of the information required. It is evident from literature that one of the simplest and effective approaches used in the extraction of facial features of face images during face recognition is the eigenface approach, which is basically based on PCA. This approach is able to convert face images into a smaller set of facial characteristics, which constitute the central components of the initial training dataset.

3.5.2 Feature Extraction Technique

After the pre-processing stage, features like the eyes, nose, ears and mouth are extricated from the face images. In this study, a dimensional reduction technique was employed after using FFT at the pre-processing phase. Thus, a feature extraction method related to dimension reduction technique was employed. The extracted features of the face image are expected to have important information through the input data. This implies that the necessary problem can be executed by a minimized representation in place of the comprehensive original data. This process means decreasing the amount of resources needed to label a large dataset. In the feature extraction phase, the study adopted PCA/SVD approach as the technique for data reduction.

Mathematically, the mean centred images from equation (3.13) are then subjected to PCA, which pursues a set of n orthogonal vector e_k , that best explains the distribution of the image data.

The t^{th} vector e_t is selected so that

$$\lambda_t = \frac{1}{n} \sum_{k=1}^n (e_k' Y_k)^2 \quad (3.19)$$

is maximized with respect to the orthogonal restrictions

$$\sigma_{kt} = \begin{cases} e_k' e_t = 1, & \text{if } k = t \\ 0, & \text{otherwise} \end{cases} \quad (3.20)$$

The vector e_t and the scalar λ_t represent eigenvectors and eigenvalues respectively, which are deduced through the covariance matrix

$$C = \frac{1}{n} Y Y' \quad (3.21)$$

where the matrix $Y = (y_1, y_2, \dots, y_n)$.

Let the matrix $U = (u_1, u_2, u_3, \dots, u_n)$ where $u_k, k = 1, 2, 3, \dots, n$ represents the k th eigen vector of C , which is the covariance matrix. Suppose that D (dimension $n \times n$) is the diagonal matrix with entries $(\lambda_k, k = 1, 2, 3, \dots, n)$ which are the eigen values having a corresponding eigen vectors $u_k, k = 1, 2, 3, \dots, n$.

The dimension of $N^2 \times N^2$ is huge and computing the eigenvalues and eigenvectors is a very complex task for distinctive image sizes. For instance, 100×100 create a covariance matrix of size 10000×10000 . From literature, it can be inferred that the vectors e_k and scalars λ_t are obtained through the computation of the eigenvalues of $Y'Y$ respectively.

$$\mu_k \alpha_k = Y'Y \alpha_k \quad (3.22)$$

Where μ_k and α_k constitutes the eigenvalues and eigenvectors respectively of the covariance matrix $Y'Y$.

By post-multiplying both by Y , we can deduce that

$$\mu_k (Y \alpha_k) = Y Y' Y \alpha_k \quad (3.23)$$

which implies that the initial $n - 1$ eigen vectors e_k and eigen values λ_k of YY' are given as $Y\alpha_k$ and μ_k respectively. Thus $Y\alpha_k$ is normalized so that it becomes equivalent to e_k and covariance matrix ranks cannot exceed $n - 1$ (the negative one was deducted as a result of the average vector n).

Hence;

$$e_k = \sum_{k=1}^n y_k \alpha_k \quad (3.24)$$

where α_k and y_k are the column matrices of U and Y respectively and $k = 1, 2, 3 \dots, n$.

The principal component of the trained set is deduced by;

$$\gamma_k = e_k'(x_k - \bar{h}) \quad (3.25)$$

$$\beta = (\gamma_1, \gamma_2, \gamma_3, \dots, \gamma_n) \quad (3.26)$$

The huge dimension of the correlated face images is finally reduced to smaller uncorrelated fundamental dimensions which exhibit advantageous features of the image set.

3.5.3 Classification Stage

This segment gives a stage by stage approach to classifying an unknown face in the trained database. After the formulation of the representation of each face, the final phase will be the need to classify the unknown face into the appropriate class.

An unfamiliar image is passed from the stage below, before classification. The face image dataset contains 10 individuals straight image pose (0°) and the varying head tilt of each individual were captured along $4^\circ, 8^\circ, 12^\circ, 16^\circ, 20^\circ, 24^\circ, 28^\circ$ and 32° rotation of the head and their frontal face component are drawn out and kept in the database. Therefore, when an unfamiliar image is placed

in a dataset, the face image undergoes a pre-processing stage and dimensional reduction process and compares its component to every face class kept in the database for classification.

3.5.4 Euclidean Distance

The rule which is most common in face recognition is the use of Euclidean distance. It basically calculates the distance from the query image and database images. Following the procedures in the dimensional reduction phase, a new image from the test set is changed into its components (eigen face).

Firstly, the unfamiliar image is juxtaposed with the mean face image and the difference is multiplied by every eigen vector from e_k .

Each value represents a weight and is saved on a vector γ_k .

Recall, from equation (3.24) and (3.25) $\gamma_k = e_k'(x_k - \bar{h})$ and $\beta = (\gamma_1, \gamma_2, \gamma_3, \dots, \gamma_n)$.

β_k accounts for the impact of every eigen face in depicting the face image by taking the eigen faces as a basis set for facial images.

The k -dimension Euclidean distance between them is computed as stated below;

$$\tau_k = \|\beta - \beta_k\| \quad (3.27)$$

where β_k represents the vector describing the k^{th} image class and β is a feature space of the database image. In this case, τ_k is chosen as the distances at which a test image is recognised in the trained image database.

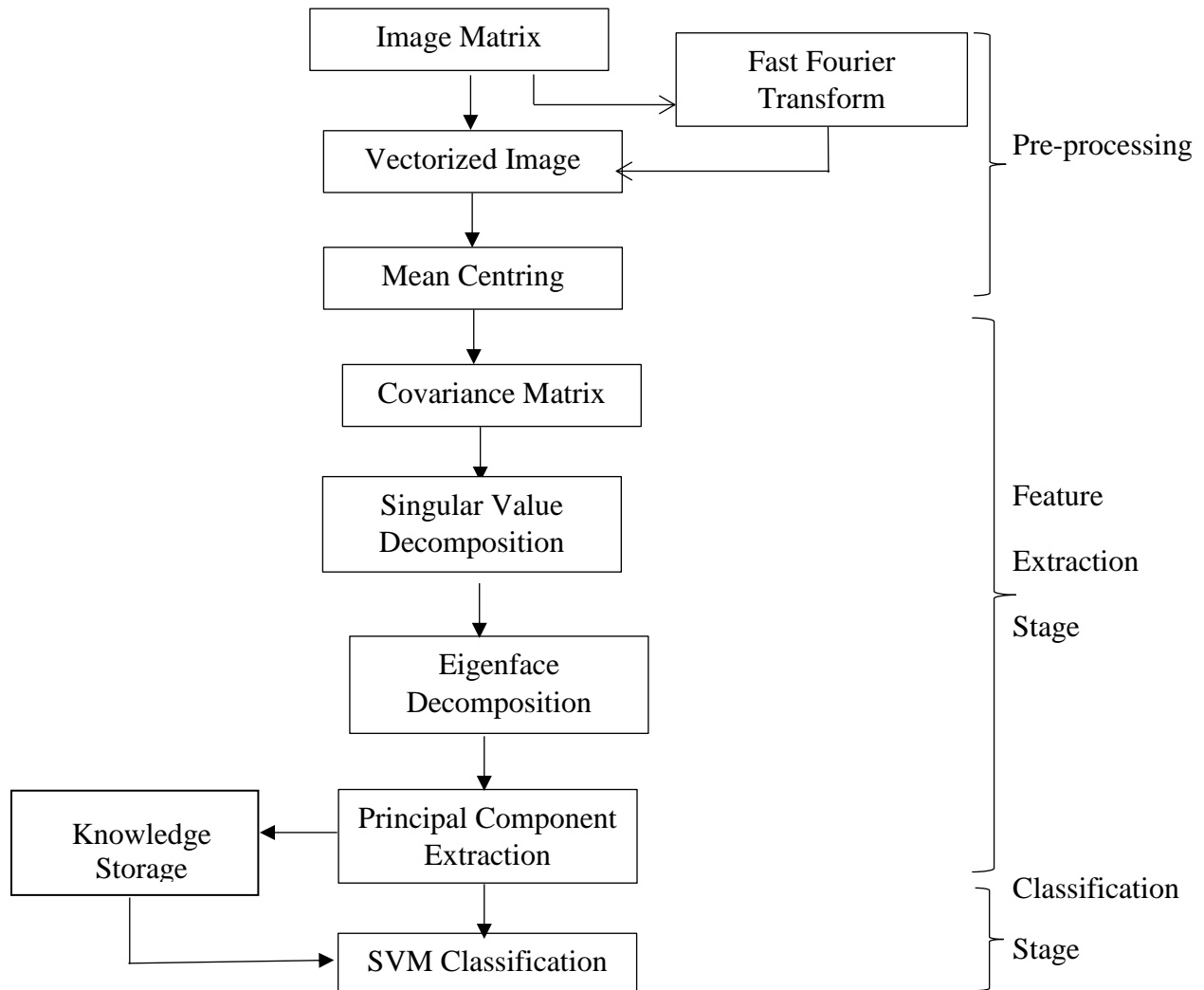


Figure 3.2. Summary of the Face Classification Process

3.6.1 Support Vector Machines (SVM)

Support vector machine, according to Vladimir and Filip (1998) is a ubiquitous constructive learning process rooted in statistical theory. Far from normal neural network and statistical methods, the SVM approach does not seek to control model complication by keeping the number of features relatively small. Rather, the dimensionality of a k-space can be very large with SVM since the complexity of the model is controlled separately from its dimensionality.

Vapnik (1998), proffered SVM as a newly discovered type of feed forward network that is particularly effective for pattern recognition. Instinctively for a set of observations belonging to two groups, the SVM identifies the hyperplane that divides the biggest possible fraction of the same class on one hand, while maximizing the distance from either class to the hyperplane. According to Vapnik (1998), this hyperplane is known as the Optimal Separating Hyperplane (OSH) which minimizes the risk of misclassifying, not only the examples in the training set but also the unseen examples of the test set.

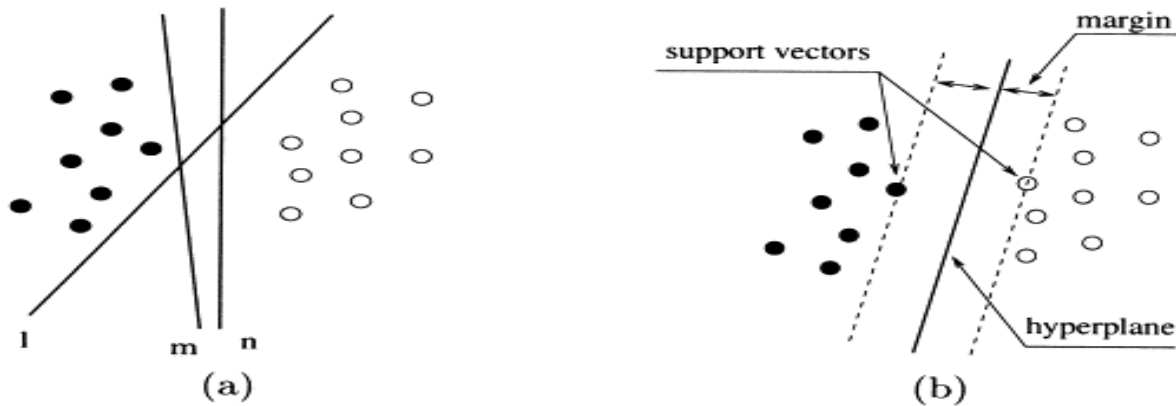


Figure 3.3. Classification between two classes using hyperplanes

Consider a two-class problem; the objective will be to divide the two classes by a function which is instigated from available examples. For instance, in Figure 3.3 (a), there exist many likely linear classifiers that will distinctively divide the data. However, only one as observed in Figure 3.3 (b) will maximize the margin. Intuitively, the expectation will be that this boundary generalizes well as against the other likely decision boundaries as seen in Figure 3.3(a). Optimal separation of the vectors by a hyperplane is achieved if it can be done in a way such that the margin is maximum and also without error. In canonical form, the optimal separating hyperplane would have to satisfy;

$$y_i[(w \cdot x_i + b)] \geq 1, \quad i=1, \dots, n \quad (3.28)$$

The distance from the hyperplane to any point \mathbf{x} will be given by

$$d(\mathbf{w}, \mathbf{x}) = \frac{|\mathbf{w} \cdot \mathbf{x} + b|}{\|\mathbf{w}\|} \quad (3.29)$$

The hyperplane that minimizes the equation below is the one that optimally separates the data.

$$\Phi(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\|^2 \quad (3.30)$$

To solve the optimization problem of (3.30) under the constraint of (3.28), there is need of the Lagrange function;

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^n \alpha_i \{y_i [(\mathbf{w} \cdot \mathbf{x}_i) + b] - 1\} \quad (3.31)$$

such that the Lagrange multipliers are the α_i s. Minimization of the Lagrangian will be done with respect to w , b and it will be maximized subject to $\alpha_i \geq 0$. The primal variables w and b will disappear with the following conditions

$$\frac{\partial l}{\partial w} = 0 \quad (3.32)$$

$$\frac{\partial l}{\partial b} = 0 \quad (3.33)$$

and $\alpha \leq 0$.

From equation (3.32) and equation (3.33) we obtain

$$\mathbf{w} = \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i \quad (3.34)$$

$$\sum_{i=1}^n \alpha_i y_i = 0 \quad (3.35)$$

Now, substituting equations (G) and (H) into (D), results in

$$L_D \equiv \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \alpha_i \alpha_j y_i y_j x_i x_j \quad (3.36)$$

where the primal variables are taken out. When α_i is determined, $w = \sum_{i=1}^n \alpha_i y_i x_i$ is obtained and any new point x can be classified with

$$\begin{aligned} f(x) &= \text{sign}(w \cdot x + b) \\ &= \text{sign}((\sum_{i=1}^n \alpha_i y_i x_i) \cdot x + b) \\ &= \text{sign}(\sum_{i=1}^n \alpha_i y_i (x_i \cdot x) + b) \end{aligned} \quad (3.37)$$

The discussion so far has been focused on the assumption that the training data is linearly separable. In situations where the separating surface of the classes is not linear, the data points are mapped to a high dimensional space, where the data points will now be linearly separable. Assuming that the transformation is $\Phi(\cdot)$, then the resulting equation in the high dimensional space will be

$$L_D \equiv \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \alpha_i \alpha_j y_i y_j \Phi(x_i) \Phi(x_j) \quad (3.38)$$

When perfect separation is not achieved, then the data points that cross the boundaries are penalized with a finite constant C . Thus, non-negative slack variables $\xi_i \geq 0$ are introduced such that

$$\begin{aligned} y_i [(w \cdot x_i + b)] &\geq 1 - \xi_i, & i=1, \dots, n & \quad (3.39) \\ \xi_i &\geq 0 \quad \forall i \end{aligned}$$

Now when the penalizing term is added to the objective function, the OSH is ascertained by minimizing

$$\Phi(w, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \quad (3.40)$$

Subject to equation (3.39)

Introducing α and β as the Lagrange multipliers, the Lagrangian becomes

$$L(w, b, \xi_i, \alpha, \beta) = \frac{1}{2} \|w\|^2 + \sum_{i=1}^n (C - \alpha_i - \mu_i) \xi_i - (\sum_{i=1}^n \alpha_i y_i x_i^T) w - (\sum_{i=1}^n \alpha_i y_i) b + \sum_{i=1}^n \alpha_i$$

The ξ_i 's and the Lagrange multipliers do not appear in the wolfe dual problem which

$$\begin{aligned} \text{maximize} &= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \alpha_i \alpha_j y_i y_j x_i x_j \\ &w, b, \xi \end{aligned}$$

subject to $0 \leq \alpha_i \leq C$.

$$\sum_{i=1}^n \alpha_i y_i = 0$$

The solution to this problem is very analogous to the perfectly separable situation, save for a moderation of the boundary of the multipliers. The sole difference from that of the perfectly separable situation is that α_i will be bounded above by C other than ∞ . The solution still remains as;

$$w = \sum_{i=1}^n \alpha_i y_i x_i$$

In this research work the focus was only on the use of the linear classifier, so there is no further discussion on the non-linear decision surfaces.

3.6.2 Multi-Class Recognition

Discussion so far has been on the basic SVM classification for only two classes. A combination of two class SVMs will lead to multi-class pattern recognition. According to Rifkin (2002), there are several approaches for solving m -class problems with binary support vector machine classifiers, but the two popular once are the one-against-all and the pairwise (one-against-one). For the one-against-all procedure, m SVMs are trained, where every one of them separates one class from all other classes. For the pairwise situation, $m(m-1)/2$ machines are trained and each SVM separates a pair of classes. From literature, there exists no theoretical analysis of the two approaches on the

basis of classification performance. When it comes to the training performance, the one-against-all strategy is mostly preferred because only m SVMs are trained as opposed to the pairwise where there is the need for $m(m-1)/2$ SVMs. The run-time complexity of the two approaches is very alike; there are m evaluations of the one-against-all while there are $m-1$ evaluations for the pairwise strategy. In 2002, a comparison of several multi-class strategies favoured the one-against-all strategy because it was very simple and had excellent classification performance (Rifkin, 2002). This research work also settled on the one-against-all technique as it seems to be at par with other techniques concerning the rate of classification and also it demands that only m SVMs are trained.

Flow chart of SVM

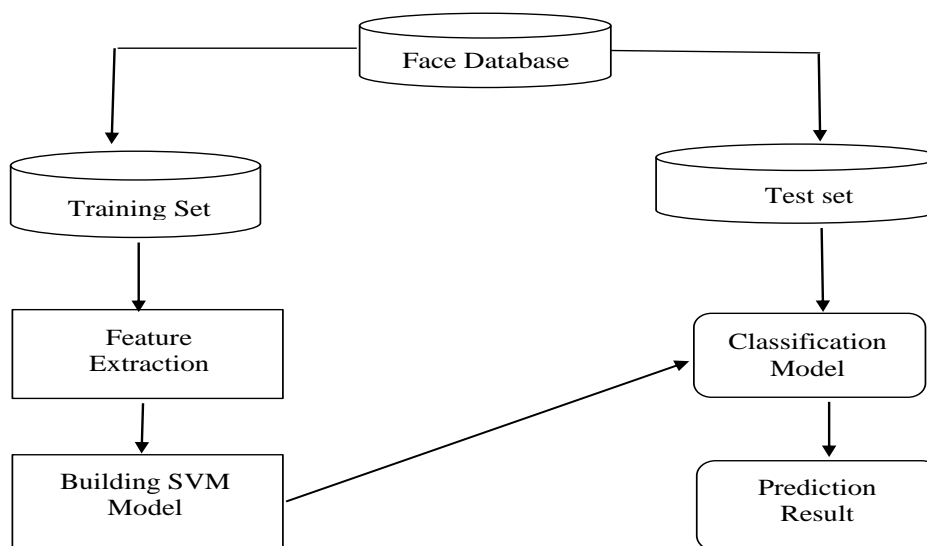


Figure 3.4 Summary of the Classification Process with SVM

3.7.1 Numerical Evaluation: Recognition (Classification) Rate

Computationally, the evaluation of face recognition algorithm is usually done through the recognition rate and run time (computational time). The recognition rate of an algorithm is expressed as the ratio of the total summation of face images recognized (classified) by the

algorithm to the summation (total) of face images in the test dataset for each experimental run. The performance of this recognition has many measurement standards.

Mathematically, the rate of recognition is expressed

$$\text{Recognition Rate} = \frac{\text{Number of Face Images Recognized } (n_r)}{\text{Number of Face Images Present } (n_t)} \times 100 \quad (3.41)$$

Thus, the average recognition rate as stated by Thakur et al., (2008) is numerically given as:

$$R_{avg} = \frac{\sum_{j=1}^6 n_{cls}^j}{q \times n_{tot}} \times 100 \quad (3.42)$$

where q represents total total number of experimental runs, n_{tot} represents the number of faces under test in each run and n_{cls}^j represents the number of correct faces which were recognized in the j^{th} run. Therefore, the mean error rate is computed as

$$ER_{avg} = 100 - RR_{avg} \quad (3.43)$$

The precision of a face recognition algorithm is an essentially vital part in the analysis process of image processing. The highest similarity measure is the lowest distance of a resultant image.

CHAPTER FOUR

DATA ANALYSIS

4.1 Introduction

This chapter gives a comprehensive computational work of the study. The subject area focused on face classification algorithm on the available database using SVM and FFT-PCA/SVD under the constraint of varying head poses. The chapter further explains the rationale behind every method used and the outcome of the analysis.

4.1.1 Data Description

The data were extracted from Massachusetts Institute of Technology (MIT) Database (2003 2005). The study generates 3Dface model based on ten training images of each individual with eight angular ($4^\circ, 8^\circ, 12^\circ, 16^\circ, 20^\circ, 24^\circ, 28^\circ, 32^\circ$) image constraints. Eighty (80) head pose images from 10 individuals were captured into the study database. Twenty (20) face images (4° and 32° head pose) from 10 individuals were captured into the train image database. Noting that, both the train and test images were resized to 100×100 -dimensional units.

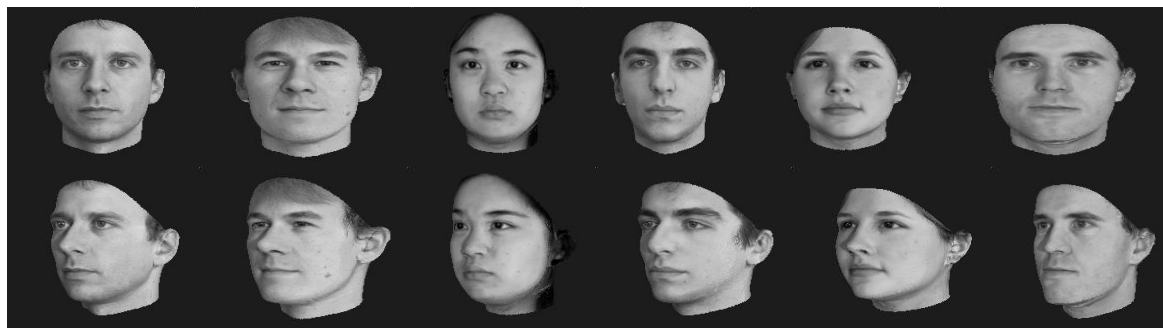


Plate 4.1 Example of Train Image (4° and 32°)

Source: [Weyrauch, Huang, Heisele & Blanz (2004)]

Sixty (60) head pose from 10 individuals across ($8^\circ, 12^\circ, 16^\circ, 20^\circ, 24^\circ, 28^\circ$) were captured into the test image database

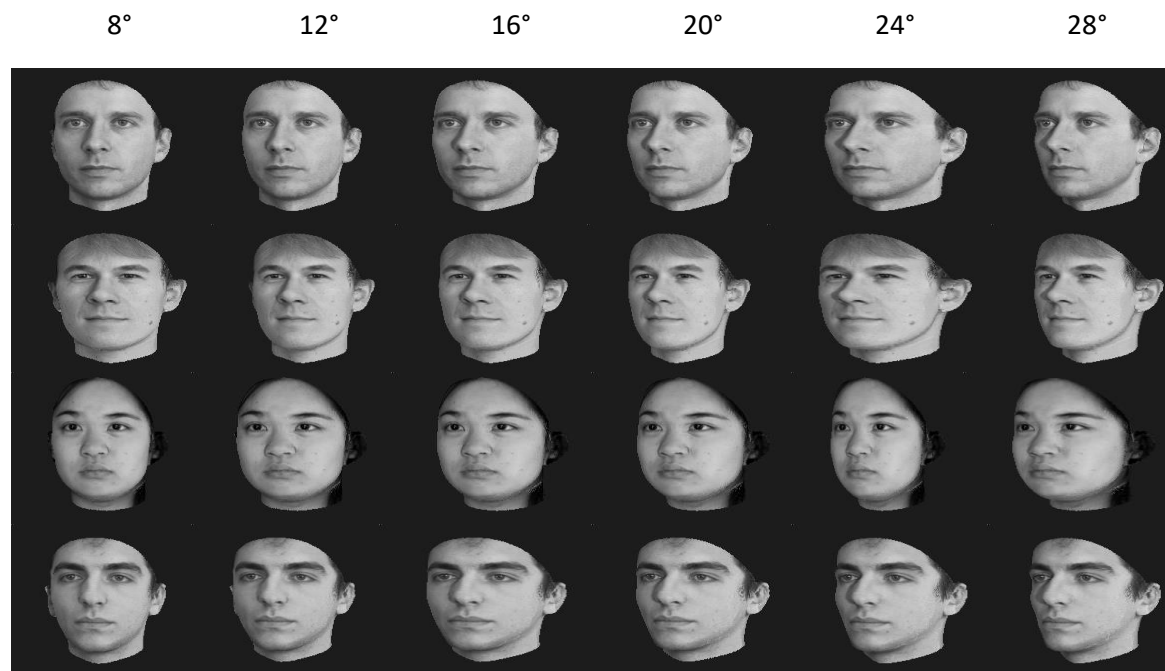


Plate 4.2: Example of the test image under the various angular pose

Source: [Weyrauch, Huang, Heisele & Blanz (2004)]

4.2 Result

As stated earlier in the methodology, the classification process is divided into three stages comprising pre-processing, feature extraction and classification.

4.2.1 Preprocessing Stage

Prior to performing the feature extraction, a preprocessing stage is applied to remove noise artifact constructed in the database before collecting the sample image values for analysis. Reviewed literature has revealed that the preprocessing stage significantly increases the consistency of a visual examination and increases the recognition performance of the algorithm. During the preprocessing stage, the study applied Fast Fourier Transform on the images.

Thus, FFT serve as the noise reducing mechanism, making the computation process simpler and better conditioned for recognition (classification). More so, the FFT technique is easy to implement on the computer (Asiedu et al., 2015).

The process of FFT began with the calculation of the Discrete Fourier Transform (DFT) and the Inverse Discrete Fourier Transform (IDFT) which yielded real parts and imaginary parts. The real components of the transformed images were extracted for the feature extraction stage whereas the imaginary components were discarded as noise.

4.2.2 Feature Extraction Stage

The feature extraction starts after the image decomposition ends. A test face dataset information is calculated according to the described process and its principal components are extracted.

Procedures

The study displays the result of the first six train image in the database;

Step One

Consider six trained images in the database Recall from (3.11);

Given the set of training images $X_k = (X_{kl})_{s \times 1}$, and $X = (X_1, X_2, \dots, X_n)$,

Let the image matrix H_k be defined as

$$\begin{aligned} H_k &= (h_{klm}); \quad l, m = 1, 2, \dots, p; \quad k = 1, 2, \dots, n, \\ H_k &= (h_{k1}, h_{k2}, \dots, h_{kp}), \quad \text{where } h_{km} = (h_{k1m}, h_{k2m}, \dots, h_{kpm})', \\ X_k &= (h'_{k1}, h'_{k2}, \dots, h'_{kp})' \end{aligned}$$

Step Two

Vectorise the train images and reshape. The trained images are reshaped from 2D to 1D so as to construct the face database such that each row depicts the trained image of an individual. The dimension of the constructed row vector is given as 1 x 10000.

The process in steps 1 and 2 are repeated for all face images in the training dataset.

Step Three

Compute the mean of the reshaped train images.

Step Four

Determine the mean centred images.

Step Five

The covariance matrix for the trained image is given as;

$$C = \begin{bmatrix} 29.4152 & -2.0518 & 2.4406 & 1.0352 & 17.7978 & 9.4755 \\ -2.0518 & 48.1708 & 2.4156 & -10.5491 & -4.9199 & -25.3641 \\ 2.4406 & 2.4156 & 41.4037 & -5.9578 & 4.7693 & 9.0433 \\ 1.0352 & -10.5491 & -5.9578 & 36.9821 & 1.5685 & 16.7911 \\ 17.7978 & -4.9199 & 4.7693 & 1.5685 & 30.4888 & 10.2756 \\ 9.4755 & -25.3641 & 9.0433 & 16.7911 & 10.2756 & 61.9324 \end{bmatrix}$$

Next, was to run a singular value decomposition (SVD) of the matrix $C = USV'$ to ascertain its eigenvalues and their corresponding eigenvectors, as seen from equation (3.16). This will split the covariance matrix C into two orthogonal matrices U and V and a diagonal matrix S .

Step Six

The Singular Value Decomposition of the covariance matrix C is;

$$U = \begin{bmatrix} -0.2232 & -0.0838 & 0.1740 & 0.3621 & -0.3268 & 0.1589 \\ 0.0704 & -0.6494 & -0.0068 & 0.0950 & 0.4363 & 0.2943 \\ -0.1764 & -0.2018 & 0.4565 & -0.5981 & 0.1685 & -0.3417 \\ -0.1578 & 0.2442 & -0.4806 & 0.1978 & 0.5261 & 0.0058 \\ -0.2242 & -0.0773 & 0.1408 & 0.2839 & -0.4285 & -0.0009 \\ -0.3641 & 0.4127 & -0.2469 & -0.3986 & -0.0952 & 0.0598 \end{bmatrix}$$

$$S = \begin{bmatrix} 233.9100 & 0 & 0 & 0 & 0 & 0 \\ 0 & 81.5831 & 0 & 0 & 0 & 0 \\ 0 & 0 & 43.8176 & 0 & 0 & 0 \\ 0 & 0 & 0 & 42.9688 & 0 & 0 \\ 0 & 0 & 0 & 0 & 33.1752 & 0 \\ 0 & 0 & 0 & 0 & 0 & 26.9746 \end{bmatrix}$$

$$V = \begin{bmatrix} -0.2232 & -0.0838 & 0.1740 & 0.3621 & -0.3268 & 0.1589 \\ 0.0704 & -0.6494 & -0.0068 & 0.0950 & 0.4363 & 0.2943 \\ -0.1764 & -0.2018 & 0.4565 & -0.5981 & 0.1685 & -0.3417 \\ -0.1578 & 0.2442 & -0.4806 & 0.1978 & 0.5261 & 0.0058 \\ -0.2242 & -0.0773 & 0.1408 & 0.2839 & -0.4285 & -0.0009 \\ -0.3641 & 0.4127 & -0.2469 & -0.3986 & -0.0952 & 0.0598 \end{bmatrix}$$

The diagonal values of the matrix S are the eigenvalues of the covariance matrix C and their corresponding unit eigenvectors are the elements of the matrix U . Here, the eigenvalues are arranged in descending order.

Step Seven

The eigenvalue for the train image is deduced from the solution of the equation $|C - \lambda I|$, where C is the covariance matrix, I is the identity matrix and λ is the eigenvalue. The result of the eigenvalue is arranged in the order of magnitude

$$\text{eigenvalues} = [233.91 \quad 81.5831 \quad 43.8176 \quad 42.9688 \quad 33.1752 \quad 26.9746]$$

The first eigenvalue accounts for an approximately 51% of the variance in the dataset, the first two eigenvalues account for approximately 68% of the variance in the dataset, while the first three eigenvalues together explained for an approximately 78% and the first four eigenvalues together an approximately 87%.

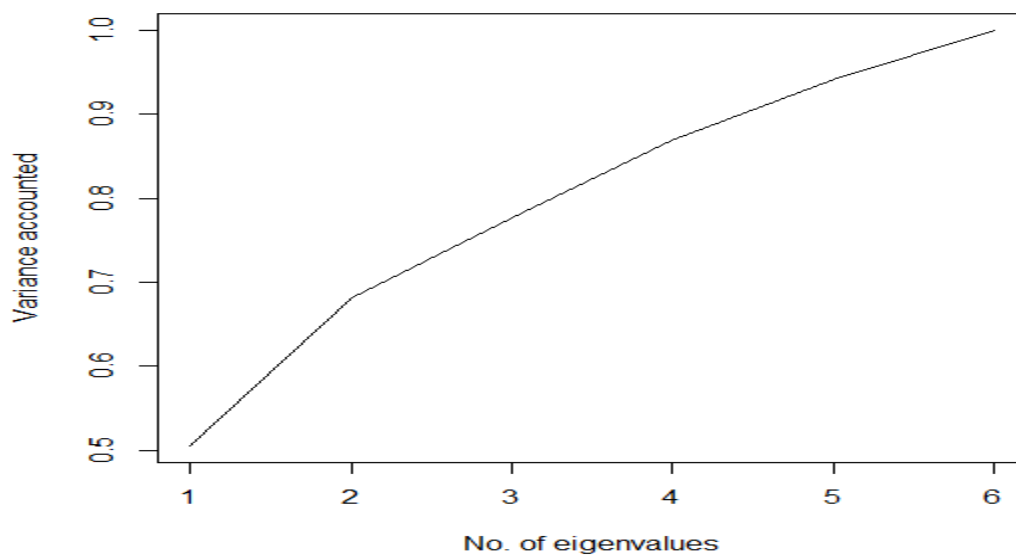


Figure 4.1 Proportion of Variance accounted by the principal components

The eigenvector is derived from $c\bar{v} = \lambda\bar{v}$ where \bar{v} is the eigenvector.

$$Eigenvectors = \begin{bmatrix} -0.2232 & -0.0838 & 0.1740 & 0.3621 & -0.3268 & 0.1589 \\ 0.0704 & -0.6494 & -0.0068 & 0.0950 & 0.4363 & 0.2943 \\ -0.1764 & -0.2018 & 0.4565 & -0.5981 & 0.1685 & -0.3417 \\ -0.1578 & 0.2442 & -0.4806 & 0.1978 & 0.5261 & 0.0058 \\ -0.2242 & -0.0773 & 0.1408 & 0.2839 & -0.4285 & -0.0009 \\ -0.3641 & 0.4127 & -0.2469 & -0.3986 & -0.0952 & 0.0598 \end{bmatrix}$$

Step Eight: Eigenface

The eigenfaces are computed as;

$$\text{Eigen Face} = \text{Eigen vector} \times \text{mean centered}$$

Step Nine: (Principal Component Analysis)

The principal component of the trained is given as

$$pc = \begin{bmatrix} -52.2096 & 16.4774 & -41.2578 & -36.9041 & -52.4339 & -85.1603 \\ -6.8382 & -52.9811 & -16.4658 & 19.9187 & -6.3070 & 33.6713 \\ 7.6235 & -0.2990 & 20.0027 & -21.0594 & 6.1675 & -10.8175 \\ 15.5591 & 4.0840 & -25.7011 & 8.4977 & 12.1973 & -17.1267 \\ -10.8407 & 14.4758 & 5.5889 & 17.4540 & -14.2164 & -3.1568 \\ 4.2854 & 7.9390 & -9.2161 & 0.1576 & -0.0256 & 1.6123 \end{bmatrix}$$

4.3 Euclidean Distance

The Euclidean distance for the various head tilts from their straight position is display in the table (4.1).

Table 4.1 Euclidean Distances of the Ten Images

4°	8°	12°	16°	20°	24°	28°	32°
6.4443	2.549	18.168	23.378	27.844	31.701	35.22	35.949
4.4414	10.076	15.876	22.391	28.646	35.618	42.119	48.482
4.3553	9.4	13.538	17.704	22.29	27.102	32.069	37.282
5.2552	9.1987	12.778	16.498	19.351	21.447	23.21	24.929
3.0838	6.7859	10.124	13.448	16.81	20.324	23.814	27.004
4.9905	10.953	16.75	21.843	27.048	31.356	35.068	35.306
5.6938	11.289	17.016	22.753	28.061	32.712	37.098	40.486
5.9992	12.095	17.701	22.625	26.747	30.066	32.485	30.894
6.3476	14.544	22.222	30.575	39.372	47.93	54.955	61.346
5.0754	12.674	20.95	29.404	38.097	46.683	54.701	57.684

Source: Researcher's Computation using Matlab

From Table (4.1), it is obvious that as an individual increases the tilt position of the head, the Euclidean distance increases. According to Asiedu, Mettle, Nortey and Yeboah (2017), the higher the degrees of head-pose, the larger the Euclidean distances and that above 20°, the recognition distances become profoundly larger.

4.4 Support Vector Machines

Based on the above conclusion, the study conducted an SVM classification by grouping the dataset into two classes namely Lower (below 20°) and Upper (20° and above). To train an SVM algorithm for classification purposes, twenty (20) images (4° and 32°) of the 10 individuals were used and the remaining sixty ($8^\circ - 28^\circ$) were used for testing as shown in plate 4.1 and plate 4.2 respectively. The lower class (below 20°) was assigned the class labels of negative one, while the upper class (20° and above) was assigned positive one as labels. Since the two classes were linearly separable, a linear kernel was adopted in the SVM algorithm. The training was done using the classifier `fitsvm` in matlab. The trained dataset and the corresponding support vectors are shown in Figure 4.3

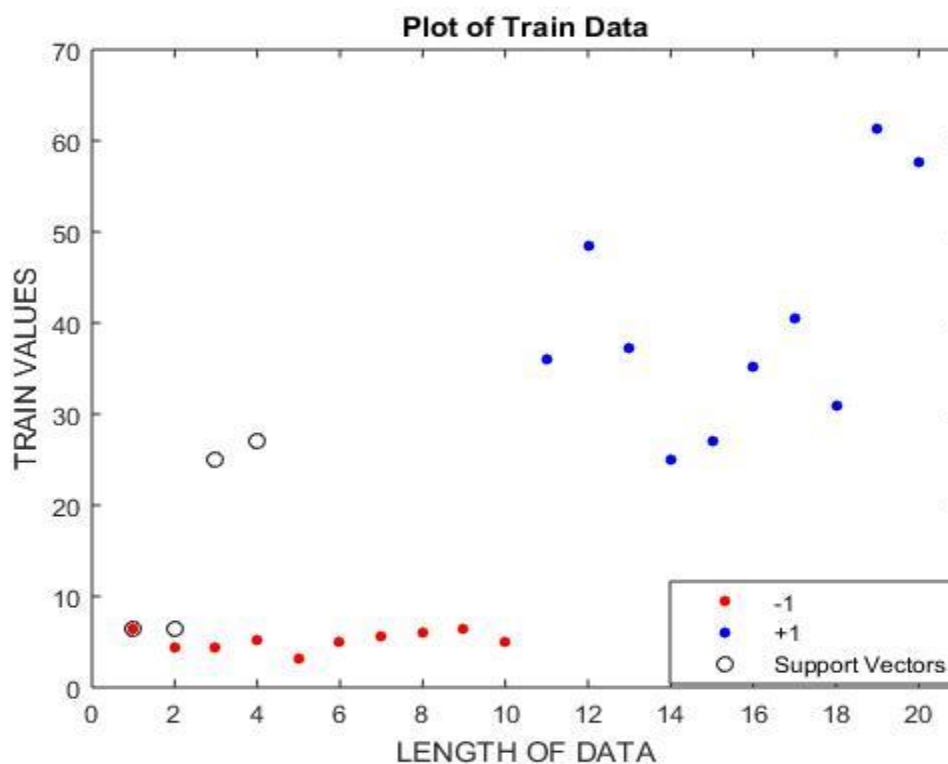


Figure 4.3 Plot of Train Data and Support Vectors

Source: Matlab Output

4.5 Classification Score

The classification score of an SVM algorithm, for classifying an observation is the signed distance from the observation to the decision boundary, ranging from $-\infty$ to $+\infty$. When an observation is predicted to the correct class, a positive sign is obtained for that class while a negative sign means the contrary. The results of the test are presented in the table of raw scores below.

Table 4.2 Raw Scores from Model Prediction (Classification)

8°		12°		16°		20°		24°		28°	
L	U	L	U	L	U	L	U	L	U	L	U
0.4	-0.4	-0.14	0.14	-0.65	0.65	-1.08	1.08	-1.45	1.45	-1.80	1.80
0.64	-0.64	0.08	-0.08	-0.55	0.55	-1.15	1.15	-1.83	1.83	-2.46	2.46
0.7	-0.7	0.30	-0.30	-0.01	0.01	-0.54	0.54	-1.01	1.01	-1.49	1.49
0.72	-0.72	0.38	-0.38	0.02	-0.02	-0.26	0.26	-0.46	0.46	-0.63	0.63
0.96	-0.96	0.63	-0.63	0.31	-0.31	-0.01	0.01	-0.35	0.35	-0.69	0.69
0.55	-0.55	-0.01	0.01	-0.5	0.50	-1.00	1.00	-1.42	1.42	-1.78	1.78
0.52	-0.52	-0.03	0.03	-0.59	0.59	-1.10	1.10	-1.55	1.55	-1.97	1.97
0.44	-0.44	-0.1	0.1	-0.58	0.58	-0.98	0.98	-1.30	1.30	-1.53	1.53
0.21	-0.21	-0.54	0.54	-1.35	1.35	-2.2	2.2	-3.03	3.03	-3.71	3.71
039	-0.39	-0.41	0.41	-1.23	1.23	-2.07	2.07	-2.91	2.91	-3.68	3.68

Source: Researcher's Computation using Matlab

From the table, it is clearly evident that the SVM model correctly classified all the observations of 8° into the lower class. However, only four (4) and two (2) observations were correctly predicted by the model into the lower class for 12° and 16° respectively. From 20° and above, the model correctly classified all into the upper class as expected. The study also considered the posterior probabilities of classifying an observation into either of the two classes, and the result is as presented in table 4.2.

Table 4.3 Posterior Probability Scores from Model Prediction (Classification)

	8°		12°		16°		20°		24°		28°	
	L	U	L	U	L	U	L	U	L	U	L	U
	0.5	0.5	0.5	0.5	0.5	0.5	0.0	1.0	0.0	1.0	0.0	1.0
	0.5	0.5	0.5	0.5	0.5	0.5	0.0	1.0	0.0	1.0	0.0	1.0
	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.0	1.0	0.0	1.0
	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1.0
	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1.0
	0.5	0.5	0.5	0.5	0.5	0.5	0.0	1.0	0.0	1.0	0.0	1.0
	0.5	0.5	0.5	0.5	0.5	0.5	0.0	1.0	0.0	1.0	0.0	1.0
	0.5	0.5	0.5	0.5	0.5	0.5	0.0	1.0	0.0	1.0	0.0	1.0
	0.5	0.5	0.5	0.5	0.5	0.5	0.0	1.0	0.0	1.0	0.0	1.0
	0.5	0.5	0.5	0.5	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0
	0.5	0.5	0.5	0.5	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0
Overall	0.5	0.5	0.5	0.5	0.4	0.6	0.1	0.9	0.1	0.9	0.0	1.0

Source: Researcher's Computation using Matlab

From the above table it can be observed that for the 8° and 12°, there was an equal chance for the model to classify an individual into either of the groups, although the expectation would have been for a greater if not absolute value for the lower class. It is worth noting that though the model correctly predicted all the observations of 8° into the lower class, the posterior probabilities suggests an equally likely situation. The posterior probability of classifying an individual increased toward the upper class from 60% (16°) to an absolute of 100% (28°). The mean posterior probability of classifying an observation into the lower class is 0.27 as compared to 0.73 for the upper class. Thus, it can be rightly deduced that it was more likely for the model to classify observations into the upper class than the lower class.

The accuracy of the classification is as presented below.

$$\text{Rate of Accuracy} = \frac{\text{Number of Face Images Correctly Classified } (n_r)}{\text{Number of Face Images in the Test set } (n_t)} \times 100$$

Table 4.4 Accuracy of Model Predictions (Classifications)

	8°	12°	16°	20°	24°	28°
Correct	10	4	2	10	10	10
Wrong	0	6	8	0	0	0
Accuracy	100%	40%	20%	100%	100%	100%
Misclassification	0%	60%	80%	0%	0%	0%

From the above table, it is clear that out of the thirty (30) classifications in the lower class (8°, 12° and 16°), only sixteen were correctly classified as compared to all thirty (30) in the upper class (20°, 24° and 28°) being correctly classified.

4.6 Classification Rate

The average classification rate as stated by Thakur et al., (2008) is numerically given as:

$$R_{avg} = \frac{\sum_{j=1}^6 n_{cls}^j}{q \times n_{tot}} \times 100$$

The total number of correct classification is $\sum_{j=1}^{10} n_{cls}^j = 46$, the total number of experimental runs $q = 10$ and the total number of face images in a single experimental run $n_{tot} = 6$

The average (mean) classification rate of the model is

$$R_{avg} = \frac{46}{60} \times 100 = 76.7\%$$

The average error rate (misclassification) of the classification model is

$$ER_{avg} = 100 - R_{avg}$$

$$ER_{avg} = 100 - 76.7\% = 23.3\%$$

This means that the FFT-PCA/SVM algorithm has a classification rate of 76.3% and a misclassification rate of 23.3%.

4.7 Classification Curve

The study displayed the graphical representation of the classification rate with respect to the various head tilts.

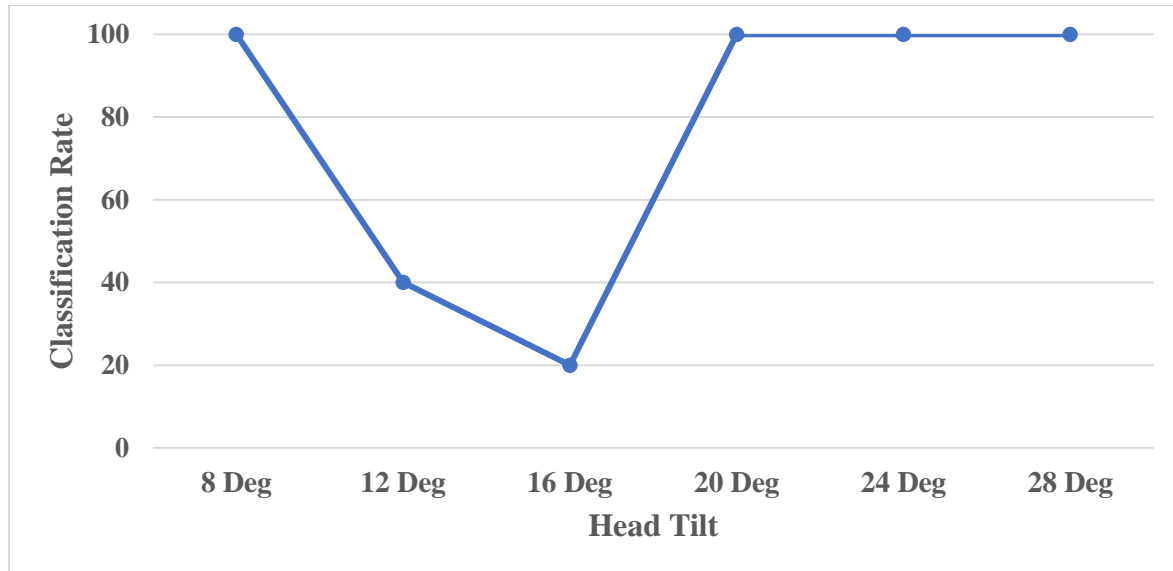


Figure 4.5: Displayed the classification curve

From the graph in Figure 4.5, it can be deduced that in adopting an FFT-PCA/SVM algorithm, the classification rate declines for the head tilt of 12° to 16° and increases for 20° and above. Thus, it can be inferred that the FFT-PCA/SVM algorithm correctly classifies observations in the upper class (20° and above) than it does for the lower class (below 20°). Also, the classification showed that 16° head tilts were more likely to be in the upper class than the lower class. This indicates that with the FFT-PCA/SVM classification, the Euclidean distance becomes larger from 16° and beyond.

CHAPTER FIVE

CONCLUSION AND RECOMMENDATION

5.1 Introduction

This is the final chapter of the study and also makes recommendations for future thesis work to be done in this field.

5.2 Summary

The study provided a concrete application of face classification algorithm using Fast Fourier Transform and Principal Component Analysis and Support Vector Machines (FFT-PCA/SVM) under the constraint of varying head tilt. The study used ten images with an angular constraint from Massachusetts Institute of Technology database (2002-2005).

From the numerical evaluations in the previous chapter, the classification rate of Support Vector Machines with Principal Component Analysis and Singular Value Decomposition (PCA/SVD) as feature extraction techniques and Fast Fourier Transform as a noise removal technique at the pre-processing stage (FFT-PCA/SVM) is 76.7% and a mean error rate of misclassification is 23.3%.

5.3 Conclusion and Recommendation

The study successfully implemented the Support Vector Machines and Principal Component Analysis (PCA/SVM) algorithm by using Fast Fourier Transform as a noise removal mechanism during the face image pre-processing stage. Adopting the study numerical evaluation, it can be concluded that, FFT-PCA/SVM has an appreciable performance when used to classify images under various head tilts. However, in the case of this study, the algorithm performed better in classifying observations in the positive (20° and above) class than the negative (below 20°) class

Future studies should be carried out with other classifiers as well as other noise removal mechanisms. Also, future studies can be conducted to understand the poor performance of the model in the lower class. Another area for future research could be a study on other constraints such as ageing and occlusions.

REFERENCES

- Abdullah, M., Wazzan, M., & Bo-Saeed, S. (2012). Optimizing face recognition using pca. arXiv preprint arXiv:1206.1515.
- Anila, S., and N. Devarajan (2012). "Preprocessing technique for face recognition applications under varying illumination conditions." *Global Journal of Computer Science and Technology* 12, no. 11-F.
- Asiedu L., Adebajji A.O., Oduro, F., Mettle, F.O. (2015). Statistical evaluation of face recognition techniques under variable environmental constraints. *International Journal of Statistics and Probability*. 4(4):93-111.
- Asiedu, L., Adebajji A.O., Oduro, F., Mettle, F.O. (2016) Statistical Assessment of PCA/SVD and FFT-PCA/SVD on Variable Facial Expressions. *British Journal of Mathematics & Computer Science* 12(6): 1-23, Article no. BJMCS.22141
- Asiedu L Mettle FO, Nortey ENN and Yeboah SE. (2017). Recognition of Face Images under Angular Constraints using DWT-PCA/SVD Algorithm. *Far East Journal of Mathematical Sciences (FJMS)* 102 (11), 2809 – 2830.
- Baker, K. (2005). Singular value decomposition tutorial, Ohio
- Bellakhdhar, F., Loukil, K., & Abid, M. (2013). Face recognition approach using Gabor Wavelets, PCA and SVM. *International Journal of Computer Science Issues*, 10(2), 201-207.

- Brunelli, R., & Poggio, T. (1993). Face recognition: Features versus templates. *IEEE Transactions on pattern analysis and machine intelligence*, 15(10), 1042-1052.
- Bruner, J. S., & Tagiuri, R. (1954). The perception of people. In "Handbook of Social Psychology, vol. II. (Lindzey, G., ed.): Cambridge, Mass: Addison-Wesley.
- Chellapa R., Sinha, P. & Phillips, P. J. (2010), Face recognition by computers and humans. *Computer Magazine*, Vol. 43, Issue 2, February 2010, pp. 46-55, ISSN: 0018-9162.
- Chen, J.C., Sankaranarayanan, S., Patel V.M., and Chellappa R. (2015), "Unconstrained face verification using fisher vectors computed from frontalized faces," 2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS), Arlington, VA, pp. 1-8.
- Chung, H., Nairn, A.C., Murata, K., and Brautigam, D.L. (1999). Mutation of Tyr307 and Leu309 in the Protein Phosphatase 2A Catalytic Subunit Favors Association with the $\alpha 4$ Subunit Which Promotes Dephosphorylation of Elongation Factor-2. *Biochemistry* 38 (32), 10371-10376
- Darwin, C. (1872). *The expression of the emotions in man and animals* John Murray: London.
- Déniz, O., Castrillon, M., and Hernandez, M. (2003). *Pattern Recognition Letters* 24, 2153–2157
- Dhoke, P., & Parsai, M. (2014). Matlab based face recognition using PCA with back propagation neural network. *Int. J. Innov. Res. Comput. Commun. Eng*, 2(8).
- Duda O. R; Hart E. P. & Stork G. D. (2001). *Pattern Classification*, Wiley-Interscience, ISBN: 0-471-05669-3, United States of America.

- El-Bakry, H.M., & Mastorakis, N. (2009). Personal identification through biometric technology, AIC'09 Proceedings of the 9th WSEAS international conference on Applied informatics and communications, pp. 325-340, ISBN: 978-960-474-107-6, World Scientific and Engineering Academy and Society (WSEAS) Stevens Point, Wisconsin, USA
- Frolov, I., and Sadykhov, R. (2008), "Experimental system for face identification based on support vector machines", in Conference on Information systems and technologies, Minsk
- Galton, F. (1892). Finger prints: Macmillan and Company.
- Glynn EF (2007). Fourier analysis and image processing. Stowers Institute for Medical Research
- Guillamet, D., & Vitria, J. (2002). Non-negative matrix factorization for face recognition. CCIA, 2, 336-344
- Guo, G., Li, S.Z., and Chan, K. (2001). Support Vector Machines for Face Recognition. Image and Vision Computing Volume 19, Issues 9–10, Pages 583-710
- Goyal, S., & Batra, N. (2016). Using 3G Dongle, Recognition and Reporting based on Pervasive Computing. International Journal of Computer Science and Information Security, 14(10), 907.
- Hasan, H. M., Jouhar, W. A. A., & Alwan, M. A. (2012). Face recognition using improved FFT based radon by pso and pca techniques. International Journal of Image Processing (IJIP), 6, 26-37.

- Heisele, B., Ho, P., Wu, J., and Poggio T. (2003). “Face Recognition with Support Vector Machines: Global versus Component-based Approach”, *Computer Vision and Image Understanding* 91, 6–21
- Haykin, S., (2009). *Neural networks and learning machines* (Vol. 3): Pearson Upper Saddle River, NJ, USA.
- Hong, Z. (1991). Algebraic feature extraction of image for recognition. *Pattern Recognition* 24(3), 211-219.
- Jain, A.K.; Ross, A. & Prabhakar, S. (2004). An introduction to biometric recognition. *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 14, Jan. 2004, pp. 4-20, ISSN: 1051-8215
- Kanade, T. (1973). *Computer Recognition of Human Faces*, Birkhauser, Basel, Switzerland and Stuttgart: Germany.
- Kelly, M. D. (1970). *Visual identification of people by computer*: Stanford Univ Calif Dept Of Computer Science.
- Kim, K. (2000). *Face recognition using principal component analysis*, USA: University of Maryland
- Kirby, M., & Sirovich, L. (1990). Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Transactions on pattern analysis and machine intelligence*, 12(1), 103-108.

- Kohonen, T. (1989). Self-organization and associative memory (Vol. 8): Springer Science & Business Media
- Marques, I., (2010). Face Recognition Algorithms; Proyecto Fin de 1-5.
- Moghaddam, B., & Pentland, A. (1997). Probabilistic visual learning for object representation. IEEE Transactions on pattern analysis and machine intelligence, 19(7), 696-710.
- Murase, H. and Nayar, S. (1995). Visual learning and recognition of 3d objects from appearance. Int. Journal of Computer Vision, 14:5-24.
- Murugan, D., Arumugam, S., Rajalakshmi, K., & Manish, T. (2010). Performance evaluation of face recognition using Gabor filter, log Gabor filter and discrete wavelet transform. International journal of computer science & information Technology (IJCSIT), 2(1).
- Nicholl, P., & Amira, A. (2008). DWT/PCA face recognition using automatic coefficient selection. Paper presented at the 4th IEEE International Symposium on Electronic Design, Test and Applications, 2008. DELTA 2008.
- Olivares-Mercado, J., Aguilar, G., Toscano-Medina, K., Nakano, M., and Meana H.P., (2011). GMM vs SVM for Face Recognition and Face Verification, Reviews, Refinements and New Ideas in Face Recognition. Available from: <http://www.intechopen.com/books/reviews-refinements-and-new-ideas-in-face-recognition/gmm-vs-svm-for-face-recognition-and-face-verification>
- Pal, M. and Foody, G.M. (2010). Feature selection for classification of hyperspectral data by SVM. IEEE Transactions on Geoscience and Remote Sensing, 48(5), pp. 2297-2307

- Patel, V.M., Gopalan, R., Li, R., and Chellappa R. (2015), "Visual Domain Adaptation: A survey of recent advances," in *IEEE Signal Processing Magazine*, vol. 32, no. 3, pp. 53-69.
- Paul, L. C., & Al Sumam, A. (2012). Face recognition using principal component analysis method. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, 1(9), pp: 135-139.
- Peng DL, Xue A (2005). Degraded image enhancement with applications in robot vision. *Proceedings of IEEE International Conference on Systems, Man and cybernetics*.2:1837–42.
- Pentland A. and Choudhury T. (2000). 'Face Recognition for Smart Environments' in *Computer*, vol. 33, no. 2, pp. 50-55. IEEE Computer Society Press Los Alamitos, CA, USA
- Phillips, J., (1999). *Proceedings of the 1998 conference on Advances in neural information processing systems*,803-809. MIT Press Cambridge, MA, USA.
- Rifkin, R., (2002). Everything old is new again: a fresh look at historical approaches in machine learning. Ph.D. thesis M.I.T.
- Reddy, K. R. L., Babu, G., & Kishore, L. (2011). Face recognition based on eigen features of multi scaled face components and artificial neural network. *International Journal of Security and Its Applications*, 5(3), 23-44.
- Sahu, G., & Tiwari, A. (2015). Training Free Face Recognition and Cropping using Skin Identification Algorithm. *International Journal of Scientific & Engineering Research*, 6(3).

- Sirovich, L., & Kirby, M. (1987). Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A*, 4(3), 519-524
- Stiefelhagen, R (2004). Estimating Head Pose with Neural Networks - Results on the Pointing04 ICPR Workshop Evaluation Data. In: *Pointing 2004, ICPR, Visual Observation of Deictic Gestures*, Cambridge, UK.
- Taigman, Y., Yang, M., Ranzato M., and Wolf L. (2014). "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, pp. 1701-1708
- Thakur, S., Sing, J. K., Basu, D. K., Nasipuri, M., & Kundu, M. (2008). Face recognition using principal component analysis and RBF neural networks. Paper presented at the First International Conference on Emerging Trends in Engineering and Technology, 2008. ICETET'08.
- Tolba A.S., El-Baz A.H. and El-Harby A.A. (2006). 'Face Recognition: A Literature Review' *International Journal of Signal Processing* 2;2
- Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1), 71-86.
- Vladimir C. & Filip M. (1998). *Learning from Data: Concepts, Theory, and Methods*, Wiley Inter-Science, ISBN: 0-471-15493-8, USA
- Vapnik, V.N. (1998). *Statistical Learning Theory*, Wiley, New York,

- Wagner, P. (2012). Face recognition with python. Tersedia dalam: [www. bytefish. de](http://www.bytefish.de) [diaksespada 16 Februari 2015].
- Weyrauch, B., Heisele, B., Huang, J., & Blanz, V. (2004). Component-based face recognition with 3D morphable models. Paper presented at the Conference on Computer Vision and Pattern Recognition Workshop, 2004. CVPRW'04.
- Yang, M-H., Kriegman, D.J., & Ahuja, N. (2002). Detecting Faces in Images: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24, 34-58.
- Yong-Qing, C. (1991). Human face recognition method based on the statistical model of small sample size. *SPIE Proc: Intell. Robots and Computational Vision*, 1607, 85-95
- Zagouras, A., Economou, G., Macedonas, A. and Fotopoulos, S., (2007). An application study of manifold learning-ranking techniques in face recognition. In *Multimedia Signal Processing, 2007. MMSP 2007. IEEE 9th Workshop on*, pp. 445-448.
- Zhang, L., Verma, B., Tjondronegoro, D., and Chandran, V. (2018). Facial Expression Analysis under Partial Occlusion: A Survey. *ACM Computing Surveys (CSUR)*. Volume 51 Issue 2, June 2018 Article No. 25 ACM New York, NY, USA
- Zhao W.; Chellappa R.; Phillips P. J. & Rosenfeld A. (2003). Face recognition: A literature survey. *ACM Computing Surveys (CSUR)*, Vol. 35, Issue4, December 2003, pp. 399-459.