FACE RECOGNITION USING PZMI, ANN AND ANT COLONY ALGORITHMS

MILAD MIRI

FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY

UNIVERSITY OF MALAYA
KUALA LUMPURE

2018

FACE RECOGNITION USING PZMI, ANN AND ANT COLONY ALGORITHMS

MILAD MIRI

DESSERTATION SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SOFTWARE ENGINEERING

FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY

UNIVERSITY OF MALAYA

2018

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Milad Miri

Registration/Matric No: WGC140017

Name of Degree: Master of Software Engineering

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):
FACE RECOGNITION USING PZMI, ANN AND ANT COLONY ALGORITHMS

Field of Study: Image Processing

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor ought I reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every right in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date: 26th March 2018

Subscribed and solemnly declared before,

Witness's Signature Date

Name:
Designation:

ABSTRACT

Face recognition system is part of facial image processing applications, which is one of the biometric methods to identify people by the features of the face. This system has many usages in security system and also can be used for authentication, person verification, video surveillance, preventing crime, and security activities. Usually, most of the standard face recognition systems contain four sections: face detection, feature extraction, feature selection, and classification. Although there are many barriers for each part of this system, many algorithms are also created to tackle these limitations. Algorithms developed for face recognition are tightly related to the rate of extracted face features. The huge redundant number of extracted features can reduce the performance of face recognition system drastically and increase the time to complete the whole process surprisingly. So, it is important to choose a proper combination of algorithms that not only diminishes the number of selected features which reduce the executing time of the system, but also improves the rate of efficiency and performance of face recognition. This study applies a new set of combination, which is Discrete Wavelet Transform (DWT) and Pseudo Zernike Moment Invariant (PZMI) for feature extraction with Ant Colony Optimization (ACO) in collaboration with Artificial Neural Network (ANN) that is experimented for the first time in the face recognition domain. ORL database has been employed as the primary dataset. The accuracy rate resulted from the system is 88.25% for PZMI+ACO+ANN and 81.34% for DWT+ACO+ANN. This research provides a new opportunity for researchers to develop face recognition system further. Researchers should be aware that the real-world conditions can be different and unpredictable as compared to the lab conditions. Online face recognition system has limitations which can motivate them to investigate more rigorously in this area.

ABSTRAK

Sistem pengecaman wajah terdiri daripada sebahagian daripada aplikasi pemprosesan imej wajah yang merupakan salah satu kaedah biometrik untuk mengenal pasti seseorang menggunakan ciri-ciri pada wajah. Sistem sebegini digunakan secara meluas dalam sistem-sistem keselamatan dan ia juga boleh digunakan untuk pengesahan, penentusahan individu, pengawasan video, mencegah jenayah, dan aktiviti keselamatan. Kebanyakan sistem pengecaman wajah standard mengandungi empat bahagian iaitu pengesanan wajah, penyarian ciri, pemilihan ciri dan pengelasan. Walaupun setiap bahagian dalam sistem ini menghadapi pelbagai batasan, pelbagai algoritma turut dicipta untuk menangani batasan-batasan tersebut. Algoritma yang dibangunkan untuk pengecaman wajah berkait rapat dengan kadar ciri wajah yang disari. Faktor ini boleh mengurangkan prestasi sistem pengecaman wajah dan meningkatkan tempoh untuk menyelesaikan proses pengecaman wajah pada keseluruhannya. Maka, adalah penting untuk memilih gabungan algoritma yang tepat yang tidak hanya mengurangkan bilangan ciri yang dipilih, yang akan mengurangkan tempoh pelaksanaan sistem, tetapi juga meningkatkan kadar kecekapan dan prestasi pengecaman wajah. Tujuan penyelidikan ini dijalankan adalah untuk mencadangkan gabungan teknik baru untuk mencapai matlamat tersebut. Kajian ini menggunakan set gabungan baru, iaitu Discrete Wavelet Transform (DWT) dan Pseudo Zernike Moment Invariant (PZMI) untuk pengekstrakan ciri dengan Pengoptimuman Ant Colony (ACO) dengan kerjasama Rangkaian Neural Buatan (ANN) yang bereksperimen untuk yang pertama masa dalam domain pengiktirafan wajah. Pangkalan data ORL telah digunakan sebagai dataset utama. Kadar ketepatan yang dihasilkan daripada sistem adalah 88.25% untuk PZMI + ACO + ANN dan 81.34% untuk DWT + ACO + ANN. Penyelidikan ini memberi peluang baru kepada para penyelidik untuk membangunkan sistem pengenalan wajah lebih jauh. Penyelidik perlu sedar bahawa keadaan dunia sebenar boleh berbeza dan tidak dapat

diramalkan berbanding dengan keadaan makmal. Sistem pengecaman wajah atas talian mempunyai batasan yang boleh memberi motivasi kepada para penyelidik untuk menyiasat dengan lebih teliti dalam bidang ini.

ACKNOWLEDGMENT

I would like to express my deepest appreciation to supervisor Dr. Mumtaz Begum Binti Peer Mustafa for the continuous support of my study, for her patience, motivation, and guidance which helped me in all the process of doing this research.

I would like to thank my committee members and all of the Faculty of Computer Science & Information Technology (Software Engineering Department) members and staff for their help and support.

My final thanks go to my parents Mohammad Miri and Marzieh Hashemi and to my lovely wife Fatemeh Kalantari for her patient and continuous support in every step of this study.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

(AF)	-	Artificial Fish
(ABC)	-	Artificial Bee Colony
(ACO)	-	Ant Colony Optimization
(ANN)	-	Artificial Neural Network
(CNN)	-	Convolutional Neural Network
(DCT)		Discrete Cosine Transform
(DFT)	-	Discrete Fourier Transform
(DWT)	-	Discrete Wavelet Transform
(GMM)	-	Gaussian Mixture Modelling
(HMM)	-	Hidden Markov Models
(KNN)	-	K-nearest Neighbor
(LDA)	-	Linear Discriminant Analysis
(MSE)		Mean Squared Error
(PSO)	-	Particle Swarm Optimization
(PZMI)	-	Pseudo Zernike Moment Invariant
(RBF)		Radial Basis Function
(SVM)	-	Support Vector Machine

CHAPTER 1: INTRODUCTION

1.1 Overview

Face recognition systems are perceived as a subset of image processing applications which are considered as one of the unique biometrics to detect people by the features of the face (Bagherian & Rahmat, 2008). The term of biometric refers to some of the unique parameters in the human body such as voice, fingerprint, palm, face, signature, and iris. Nowadays, the recognition of individuals from their face is a mission so ordinary and straightforward that nobody even notices how many times it is performed every day. Interestingly, during the last decade, numerous face modeling techniques have been progressed. Nonetheless, face recognition has some advantages over other biometric methods. This system is natural and easy to use and it does not lead to disturbing the people (Almohammad & Mahmoud, 2013). It is also considered as a non-collaborative biometric system since it can quickly gather the facial data from cameras or webcams without the assistance of the people. These systems have significant usage in security systems and some security activities such as video surveillance, verifying the identity of people and blocking crime. In real-world condition, a face recognition system has complexity due to image processing problems stemming from handling the complicated and vast effects of occlusion, illumination, and shape of the images on the live systems.

1.2 Research Background

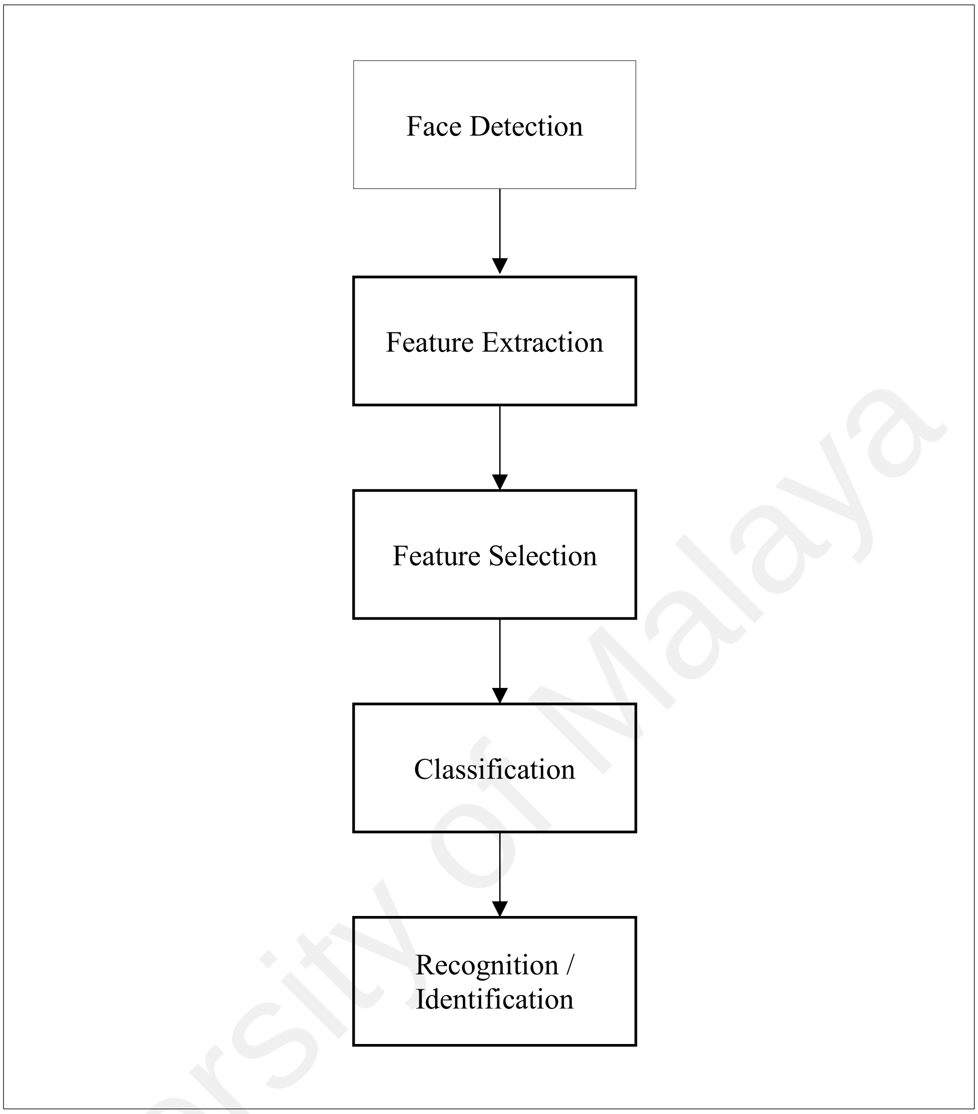


Figure 1.1: Order of different part in a face recognition system(Bagherian & Rahmat, 2008)

Face recognition systems mostly include a procedure of four stages, namely face detection, feature extraction, feature selection and recognition or classification stage which are shown in Figure 1.1. Initially, an input image could be captured from a camera or webcams. Detecting the shape of the face in a video or image is the beginning step of a face recognition system. The next step is making a vector of features by extracting multiple features from the input image. These features should hold unique data about each person in the database so the system would be able to identify the individual based on the extracted features. The input images are taken from devices like cameras, might not be appropriate for recognition due to having noise or illumination circumstances. Reducing the feature vector by utilizing some algorithms to enhance the accuracy rate and lessen the execution time is one of the roles of feature selection stage. The final step is classification where the system should identify an undiscovered sample. The classification section employs several recognition algorithms to classify and recognize the given images. These face images usually have some accepted attributes such as the same size or resolution of the picture. Recognition algorithms are usually applied on the standard datasets. There are a variety of algorithms for each stage which are collected in Table 1.1 based on the studies were conducted before.

In the past, there have been many attempts to improve the performance of a face recognition system. Multiple kinds of research and studies have been conducted to obtain the highest accuracy result for the recognition system. Several combinations of algorithms have been applied, and each generating a different rate of recognition accuracy. Lately, Principal Component Analysis (PCA) has been considered as one of the most popular feature extraction algorithm which is adopted in face recognition system research. In a research by Eskandari et al. (2014), PCA is applied as the feature extraction algorithm and Radial Basis Function (RBF) as the primary classifier. Their obtained result for accuracy was not slightly more than 82% (Eskandari & Toygar, 2014). In another study, PCA was applied with Naive Bayesian Classifier. In this study, the accuracy rate surprisingly decreased to 78% (Ouarda, Trichili, Alimi, & Solaiman, 2013). On the other hand, several studies have reported higher rates of accuracy. For example, Latha et al. (2009) experimented the combination of PCA, and K-Nearest Neighbors (k-NN) and the result attained is almost 92% of accuracy rate (Latha, Ganesan, & Annadurai, 2009). 98.5% accuracy rate was obtained for the combination of DWT, ACO, and Nearest-Neighbor (Kanan, Faez, & Hosseinzadeh, 2007), 94.37% for the sequence of DWT, Firefly, and Nearest-Neighbor (Agarwal & Bhanot, 2015), and 90.5% accuracy rate for the combination of DWT, GA, and KNN (Lv, Wu, & Liu, 2014). Hence, variation in algorithm combinations can produce a range of different accuracy rates.

Table 1.1: The summary of algorithms for each part of a recognition system

Section	**Algorithm's Name**	**Reference**
Feature extraction	• Principal Component Analysis (PCA) • Linear Discriminant Analysis (LDA) • Discrete Wavelet Transform (DWT) • Discrete Fourier Transform (DFT) • Pseudo Zernike Moment Invariant (PZMI)	(Eskandari & Toygar, 2014) (Zhu, 2001) (Kanan et al., 2007) (Kaur & Kaur, 2012) (Kanan & Faez, 2005a)
Feature selection	• Ant Colony Optimization (ACO) • Artificial Fish (AF) • Particle Swarm Optimization (PSO) • Artificial Bee Colony (ABC)	(Rao & Rai, 2016) (Cheng, Li, & Bao, 2016) (Nadhir, Wahab, Nefti-meziani, & Atyabi, 2015) (Kaur, Panchal, & Kumar, 2013)
Classification Algorithms	• Support Vector Machine (SVM) • Hidden Markov Models (HMM) • Artificial Neural Network (ANN)	(Foruzan, Scott, & Lin, 2015) (Ho & Chellappa, 2013) (Foruzan et al., 2015)

	• KNN	(Lv et al., 2014) (Latha et al., 2009) (Kanan et al., 2007)
	• Radial Basis Function (RBF)	(Omer & Khurran, 2015) (Eskandari & Toygar, 2014)

1.3 Research Problem

As studies show, all face recognition systems are susceptible to occlusion, image adjustment, nature of the image and image condition. Moreover, skin color, sexuality, face accessories like glasses influence the performance of the detection (Wong, Lam, & Siu, 2001). On the other hand, algorithms developed for face recognition are tightly related to perfecting the rate of the extracted face features (Agarwal & Bhanot, 2015). Irrelevant and redundant features not only degrade the performance of the system but also increase the execution time for completing the whole process consequently (Agarwal & Bhanot, 2015). Accordingly, realizing an accurate and efficient sequence of algorithms in a face recognition system is challenging and intricate work.

Feature extraction is crucial in extracting the facial images which influence the performance of face recognition. There are several algorithms for feature extraction such as Discrete Wavelet Transform (DWT), Pseudo Zernike Moment Invariant (PZMI), Discrete Cosine Transform (DCT) and Discrete Fourier Transform (DFT). Usage of these algorithms can have a direct impact on accuracy rate. For example, the accuracy rate was 84.4% using DWT, while, interestingly, in the same environment it was enhanced to 93% when PZMI was used for feature extraction (Kanan & Faez, 2005b).

The features extracted in the previous stage pass to the feature selection component. The principal goal of this component is to diminish the size of the given dataset as much as possible by eliminating the undesired and redundant features. Consequently, the execution time of a face recognition system is decreased considerably. Classification is the most crucial part of a face recognition system. The system can produce accuracy rate by training and testing the feature vectors which were rendered and prepared in the former steps. Several well-known algorithms including Support Vector Machine (SVM), Radial Basis Function (RBF), Hidden Markov Models (HMM), Artificial Neural Network (ANN) and Graph Matching are usually applied in this part of the system.

Various combination of feature extraction, feature selection and classifiers will yield varying accuracy results. As a sample in a study conducted by Agarwal et al. (2015) 94.357% accuracy result was obtained for the sequence of DCT for feature extraction, Firefly for feature selection and Nearest-Neighbor as the primary classifier. Farmanbar et al. (2016) performed another study with the same classifier but different kinds of algorithms for feature extraction and selection. They applied LBP as the central feature extraction and BSA as the feature selection algorithms. The result achieved could not exceed more than 85% of accuracy rate which is almost 10% less than the rate reported by Farmanbar et al. (2016).

Therefore, as it was said previously, the different sets of the algorithms can produce a varied range of accuracy rates. Consequently, there is a lack of experiment on the combination of some of the feature extraction, feature selection and classifiers algorithms that can lead to higher rate of recognition accuracy.

1.4 Research Objectives

The primary intention of this research is to deliver a new sequence of algorithms in a face recognition system domain which can generate an acceptable accuracy rate and execution time corresponding with some of the existing studies.

The objectives of this study are described as follows:

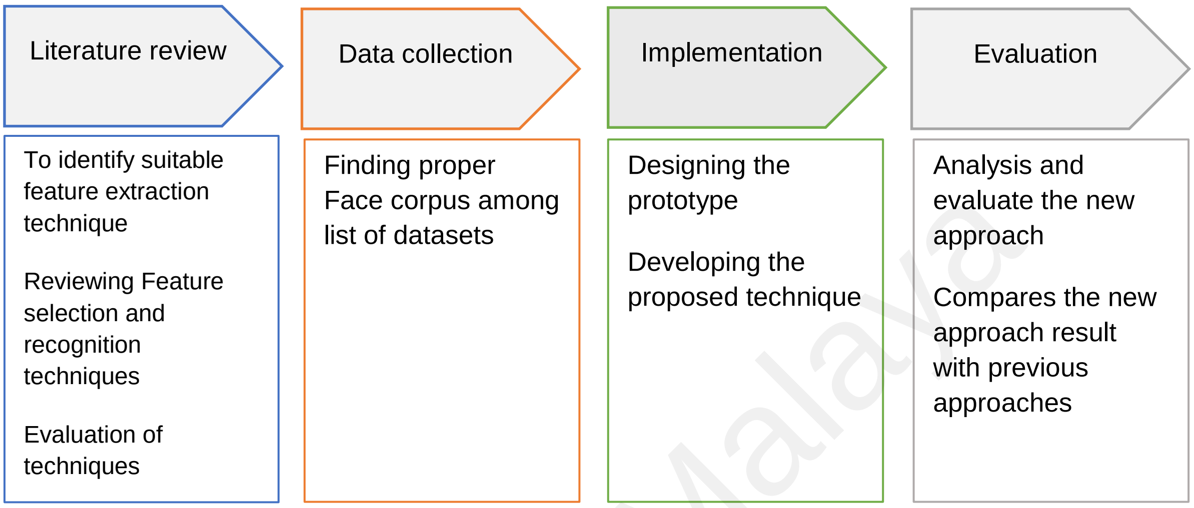
1) To identify the existing algorithms and analyze the performance of the combination algorithms for feature extraction, feature selection and classification.
2) To develop a face recognition system using the proposed combination of algorithms for improving the performance of face recognition system.
3) To evaluate and compare the performance of the developed face recognition system with the performance of existing benchmark algorithms.

1.5 Research Scope

The main effort in this research is dedicated to attaining and adopting a new sequence of algorithms for feature extraction, selection, and classification that can achieve a satisfactory accuracy rate in face recognition system. As each standard datasets have the standard size and type of image, face detection part is usually evaluated independently. Moreover, this research uses the standard dataset to compare its results with the other benchmark algorithms.

1.6 Research Methodology

This section provides a brief introduction to the research methodology applied in this research. There are five major steps in this research methodology, which includes literature review, data collection, design and implementation, developing the proposed system and performing the evaluation as described below:



1.6.1 Literature Review

The first step of this study is to investigate the existing literature that concentrated on the face recognition systems. The purpose of the review is to recognize the appropriate algorithm and approach to be employed in this research for the development of a new sequence of algorithms for each part of the system. This part comprises three critical stages including identifying feature extraction algorithms, reviewing feature selection and classification algorithms and finally, providing evaluation, advantages, and disadvantages of each one to perform a thorough background study about each algorithm.

1.6.2 Data Collection

Finding a list of the most well-known datasets widely used in other studies in this domain is the primary goal of this part. A summary of all concerns for deciding to select the appropriate dataset is listed below.

- The dataset should be free of charge and easy to access.
- The dataset should be very common for the researchers.
- The dataset should contain reasonable number of images.

1.6.3 Design and Implementation

This phase arranges a framework by choosing appropriate algorithms for the system. Moreover, developing and implementing the mentioned prototype must be done in this phase. The evaluation of this part is listed below.

- Completing a review of conventional methods and approaches used in the existing face recognition systems.
- Comparing the performance of the existing face recognition systems to recognize proper feature extraction, feature selection, and classification methods.
- Implementing the selected methods in developing a novel combination of algorithms for face recognition system.

1.6.4 Developing the Proposed System

To develop the proposed system initially some sample data should be prepared, so for this step, it is common to use the existing dataset to produce a comparable result, then it is needed to train some of this sample data and finally, the result is captured. MATLAB version R2016a is used as the main tool to develop the proposed system.

1.6.5 Evaluation and Results

The prototype developed in the previous step will be experimented in this part. The new approach will be thoroughly analyzed and evaluated. Eventually, it will be compared with the results of other studies to discover the effectiveness of the components and features of this approach.

1.7 Outline of the Dissertation

The structure of the rest of this dissertation is as follows:

Chapter 2 Reviews the existing literature that is associated with face recognition systems. Accordingly, performing a comprehensive review of the mixtures of algorithms widely used in recent research is conducted.

Chapter 3 The main purpose of this chapter is planning before performing the experiment. The planning includes determining what suitable algorithms and steps to be applied in the experiment.

Chapter 4 Describes the development of the proposed system. Also, discussions regarding the performance of each adopted method and algorithm applied during the experiment are included in this chapter.

Chapter 5 Discusses the result provided by the proposed system and compares the achievement of the entire system with the result of other studies.

Chapter 6 Provides a summary and conclusion of the research and discusses about limitations and possible future research for further improvement of the face recognition system.

CHAPTER 2: LITERATURE REVIEW

2.1 Overview

This chapter presents a review of the literature on various algorithms for each particular part of a face recognition system. Initially, some background in face recognition has been presented. Then, the most recent methods for each section will be discussed. After that, the popular face recognition databases are shown, and then, results of some most well-known combination of algorithms in face recognition systems will be introduced. Finally, in conclusion, the main points of the chapter are summarized.

2.2 Face Recognition System

Face recognition is a biometric recognition which is widely used in different technologies. The term 'biometrics' refers to the features which are related to human characteristics. Table 2.1 presents a comparison of various biometric attributes.

As can be seen, although among all biometrics the fingerprint has an average outcome regarding the usability, it has the highest rate of long-standing stability. Face and voice have the highest degree of user acceptance among the others. Although fingerprint recognition has been used for a long time, the popularity of face recognition has increased surprisingly in the last 20 years (Agarwal & Bhanot, 2015). Besides, some biometric systems utilize a behavioral pattern, but face recognition distinguishes a person on physiological attributes. In addition, this method is non-invasive which means that it does not need a person to be isolated from a group to be monitored; therefore, a direct contact with a user is not necessary and also it is quite inoffensive and acceptable (Darestani, Sheikhan, & Khademi, 2013).

Table 2.1: The comparison of different biometric recognition.

Biometric	Usability	Long term stability	User acceptance	Variability
Face	Medium	Medium	High	Orientation of the head, situation of lighting, illness and etc.
Voice	High	Medium	High	Illness, age, stress, environment.
Iris	Low	High	Low	Poor lighting, eye position.
Fingerprint	Medium	High	Medium	Dryness, sensor noise, Dirt, Bruises.
Hand	High	Medium	Medium	Injury, age.

Face recognition systems are classified into several types. Regarding pose invariant, it can be grouped into two main classes: global approach and component-based approach. In the global approach, the system applies a single feature vector for the whole face image but in component-based approach, it employs a compensation approach by finding a flexible geometrical relation between the components in the classification stage. The latter approach aligns the image to be insensitive to translation and rotation. Generally, in all procedures, the following steps are implemented:

1. An image is captured by the sensor (which is known as face detection)
2. The given picture requires some pre-actions to be prepared and normalized (feature extraction and selection).

3. A comparison between the images in dataset and normalized image will be performed (which is known as classification)

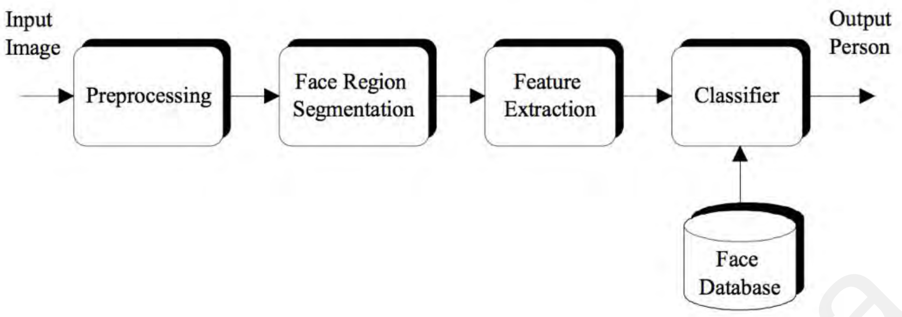


Figure 2.1: Diagram of a face recognition system (Kanan et al., 2007)

The face recognition system is very susceptible to occlusion, image adjustment, and image quality and image position. Moreover, several variables alter the detection performance, including wearing glasses, having different skin color and gender, and facial emotions (Wong et al., 2001).

2.2.1 Facial Expression

There are many external elements which might affect the performance of the system. For example, wearing glasses, growing beard or mustache and having emotional expressions like smiling or scowling may modify the natural condition of the face and might deteriorate the outcome of the recognition. These are some challenges which are expected to overcome by the system.

2.2.2 Illumination

It is confirmed that the position of the light source produces a shadow on the face and in some circumstances, it can alter the style of the face entirely or will create some highlights in a face which are so bright or dark. This issue might influence the efficiency of the system and make it very challenging to identify some facial features.

2.2.3 Head Pose

In order to produce a reasonable accuracy rate, the system should capture numerous information about the individual's face. To achieve this goal the position of the head is essential for the system. For example, pictures which are taken from a direct view is desired for the system. In these images, users look straight into a camera. Hence, the orientation of the head might have an immediate effect on the efficiency.

2.2.4 Occlusions

Human beings are smart enough to be able to know another human who wears scarf or sunglasses. Unlike humans, the automated face recognition system is not that intelligent. Sometimes an unimportant occlusion in the input image is a critical challenge for the application. These occlusions have a notable effect on the performance of the system and can lower the accuracy rate drastically.

2.3 Algorithms in Face Recognition System

2.3.1 Feature Extraction Algorithm

Extracting redundant features from images is a vital segment of face recognition. Hence, selecting the right feature extractor algorithm is a principal function to produce a high rate of acceptance. Usually, feature extraction algorithms are classified in two different models; feature-based approach and holistic approach (Darestani et al., 2013).

2.3.1.1 Feature-based Approaches

Feature-based approaches are those which are based on extracting fundamental and geometrical facial features. For example, the pattern of the mouth, nose, eyes and the distance of them from each other are distinguished by this type of approach. Although, redundant information in the image might not affect these methods; they are susceptible to the unpredictability of face appearance and environmental conditions. Linear discriminant analysis (LDA) is the most robust and practical algorithm in this class (Nabatchian, Abdel-Raheem, & Ahmadi, 2008). Feature-based approaches are divided into two parts:

a) Geometric feature based matching

This algorithm is based on the calculation of a group of regular feature extracted from the picture of a face. The whole configuration can be described as a vector. This vector signifies the position and the size of central facial features like the eyes, mouth, nose, eyebrows, chin and the boundary of the face. The benefit of this algorithm is that it can effortlessly overwhelm the problem of occlusion. As the major problem of these algorithms, it is declared that efficiency of these algorithms is meager (Masupha, Zuva, Ngwira, & Esan, n.d.).

b) Elastic bunch graph

This algorithm is based on dynamic link pattern. A graph for an individual face is formed utilizing a set of fiducial points on the face. Each fiducial point is a connection of a fully coupled graph and is marked with the Gabor filters' response. Each curve is named by the distance within correspondent fiducial points. (Masupha et al., n.d.) The summary of all advantages and disadvantages of feature-based approaches is presented in Table 2.2.

Table 2.2: The Pros and Cons of Feature-based approaches (Masupha et al., n.d.)

Advantages	Disadvantages
They are stable in orientation, size and lighting.	Feature-based algorithms have an absence of discrimination capability.
It is fast and efficient.	Auto-detection is very troublesome in this approach.

2.3.1.2 Holistic Approaches

Holistic approaches or deterministic approaches are those who examine face images as a two-dimensional holistic pattern. Due to considering features as global in the whole vision, irrelevant elements like the pattern of the background and other unnecessary textile in the picture might influence the feature vectors and generate an inaccurate outcome. DFT, DCT, DWT, PZMI, HMI, BMI are the most prominent algorithms in this group (Darestani et al., 2013). The summary of all advantage and disadvantages of holistic approaches are listed in Table 2.3.

Table 2.3: The Pros and Cons of Holistic approaches (Masupha et al., n.d.)

Advantages	Disadvantages
By focusing on some specific parts of the picture they will not remove any data from the images.	Since this approach does not ignore any information from the image, it is required to start with the underlying assumption that all the pixels in the image are equally important and this drains the system resource.
Producing slightly better result than the other algorithm.	This method usually needs loads of system resources during the implementation.
	The effectiveness of this procedure is not excellent, especially in an extensive system.

- ***Principal Component Analysis (PCA)***

Principal Component Analysis (PCA) is one of the most successful and well-known algorithms used in image identification and compression for extracting feature and reproducing data. The primary goal of PCA is to minimize the massive dimensionality of feature vector to the smaller set of elements which is needed to represent the data efficiently (Gonzalez & Woods, 2002). The summary of all advantage and disadvantages of PCA are presented in Table 2.4.

Table 2.4: The Pros and Cons of Principal Component Analysis(PCA)(Masupha et al., n.d.)

Advantages	Disadvantages
It is easy and efficient as the PCA reduces the dimension size of an image in a short period of time. It has high correlation between the training data and the recognition data	The outcome depends singularly on some circumstances. For example, some factors like lighting can reduce its correctness drastically.

- ***Pseudo-Zernike moments invariants (PZMI)***

Pseudo-Zernike moments invariants (PZMI) is one of the most prominent feature extractors which is highly stable against rotation, scaling, and translation. This algorithm plays a remarkable role in classifying images because the efficiency rate of the classifiers is notably based on the relevance of the features. Moment functions capture global elements and thus are fitting in face recognition domain. There are various examples of moments including geometric, complex, radial and orthogonal. Geometric moments are widely employed in image processing; however, these moments are not optimal concerning data redundancy. Some moment functions exhibit natural invariance properties including invariance to translation, rotation or scaling. It is very sensitive to the pattern features so it can be easily applied to the pattern recognition systems. In 1962 Hu (Ming-Kuei Hu, 1962) announced algebraic moment invariants (HMI). Later in 1981 the other improved version of the algorithms which is called regular moment invariants (RMI) (Reddi, 1981) was announced. It was the easiest and perhaps the most prominent moment invariants. Later bamieh moment invariants (BMI) was presented which had small feature vectors; therefore, it was more efficient than others (Bamieh & De Figueiredo, 1986). Zernike and pseudo Zernike orthogonal polynomials are the basis of the zernike moment invariant (ZMI) and pseudo zernike moment invariant (PZMI) (Wallin & Kübler, 1995). In a research which compared all the above feature extraction algorithms, PZMI had the highest result. Although BMIs were the fastest, they could not produce high recognition accuracy (Nabatchian et al., 2008).

- ***Discrete Wavelet Transform (DWT)***

Discrete Wavelet Transform (DWT) is a wavelet transform performed in the discrete domain. It has mother and daughter wavelets with which multi-level breakdown is performed and spectral representation of samples is obtained. DWT is frequently utilized in image compression, recovery, and de-noising applications. DWT comprises three steps applying a signal, decomposition, and reconstruction. It is also named as discrete wavelet transform and inverse discrete wavelet transform. Lately, Discrete Wavelet Transform has been applied many times in face recognition systems (Jana & Sinha, 2014) (Patil, Nayak, & Jain, 2015). Usually, in DWT a Haar is used on images. After passing the input images from DWT signals, the output would contain four sections which are one approximation band called LL band which is made from low frequency and this part contains the most important information about images and three detailed bands called LH, HL and HH bands which are made from high frequency(Patil et al., 2015). For example, after applying 3 levels of decomposition 9 different frequency bands will be produced which are shown in Figure 2.2.

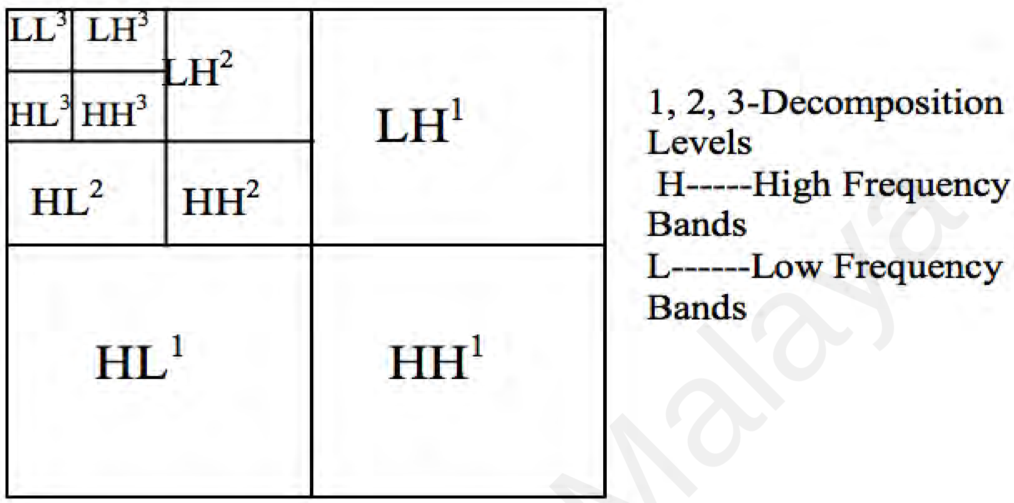


Figure 2.2: Decomposition of DWT after 3 levels (Sihag, 2011)

Table 2.5: The Pros and Cons of DWT (Dond, Sun, & Xu, n.d.; Sihag, 2011)

Advantages	Disadvantages
Since data are spread into several components, so it is easier to be filtered. Having much more flexibility.	Having many complexity It is required much more resource regarding memory and CPU.

2.3.1.3 Summary of feature extraction algorithms

The most striking result to emerge from feature extraction section is summarized as below. Generally, there are two main approaches in feature extraction domain which are feature-based and holistic (Bagherian & Rahmat, 2008; S, B, & B, 2012). In a study by Darestany et al. (2013), it is shown that holistic approach has slightly better result than feature-based approach. Consequently, some of the most famous algorithms in the holistic approach were introduced which are mainly derived from the moment invariants or wavelet transform. For those algorithms which are moment invariants-based PZMI produced better result than the other moment derivatives algorithms (Nabatchian et al., 2008). On the other hand, in the second category DWT outperforms the other wavelet-based algorithms (A. Kaur & Kaur, 2012). Therefore; according to significant results of DWT and PZMI in comparison with other algorithms in their category, these two algorithms are selected to use as the main feature extractor.

2.3.2 Feature Selection Algorithm

Feature selection has been in the center of attention for quite some time and has played an essential role in a standard face recognition system. With having considerable databases in a machine learning algorithms, new challenges occur, and novel and proper approaches to select suitable features are required (Dash & Liu, 1997). In many applications, the size of a dataset is so vast that learning might not operate as well. Therefore, it is required only to extract some features from images for performing the recognition. Unluckily, some of the extracted features are redundant or irrelevant. Thus, they are not proper to be introduced to the system. Consequently, an inappropriate feature is not able to help the system to generate a robust result, and redundant elements merely add an overload to the system (Biodiversity, Shannon, & Shannon, 2010). Therefore, reducing the number of redundant features minimizes the execution time of a face recognition system considerably (Dash & Liu, 1997). Diminishing the unnecessary features helps to have a better insight into the underlying concept of a real-world classification problem. These algorithms usually explore the whole solution space for the best result, and it is considered as the core advantages of them. Recently, in some novel research, these algorithms have been combined and the suggested hybrid method produced moderate results (Sen & Mathur, 2016). The following research has been conducted on some of the existing feature selection algorithms.

2.3.2.1 The BAT Algorithm

The BAT algorithm is developed by Xin-She Yang in 2010. It is based on the behavior of bats in the nature (Fouad, Zawbaa, Gaber, Snasel, & Hassanien, 2016). The honor of the echolocation of micro-bats can be compiled as follows:

I. Each practical bat flies randomly with distinct rapidity with a varying frequency or wavelength and loudness.

II. As it explores and attains its victim, it adjusts frequency, pulse discharge rate, and loudness.

III. Exploration is enhanced by a local casual position.

IV. Collection of the best remains until regular stop criteria are met.

This algorithm utilizes a frequency-tuning method to measure the dynamic performance of a swarm of bats, and the offset between investigation and exploitation can be tested by tuning algorithm-dependent parameters (Fouad et al., 2016).

Table 2.6: The Pros and Cons of BAT (Fouad et al., 2016).

Advantages	Disadvantages
Finding the solution is almost guarantee. BA uses parameter control, Frequency tuning, and Automatic zooming.	The fine adjustment in parameters does affect the convergence rate of the optimization process.
It performs well for systems with large-scale problems.	It Is performance is widely based on the number of the parameters in algorithms.

2.3.2.2 Artificial Fish Algorithm (AF)

Artificial Fish-Swarm Algorithm which is mainly based on the swarm intelligence algorithms. This method has a slightly better optimization rate than others. It is motivated by the natural social life of the fish. Naturally, the fish always attempt to protect their colonies and accordingly illustrate an intelligent action which is the main reason for creating this algorithm. Searching for food, immigration and dealing with dangers all happen in a social form and interactions between all fish in a group will result in an intelligent social behavior. This algorithm has many advantages including high merging speed, adaptability, fault sensitivity and high efficiency (C. Cheng et al., 2016).

Table 2.7: The Pros and Cons of Fish-Swarm(C. Cheng et al., 2016).

Advantages	Disadvantages
High convergence speed	High time complexity
High accuracy and flexibility	There is no stability among global and local search.
	It is not smart enough to experience of the movement of the group members for its next movement.

2.3.2.3 Artificial Bee Colony (ABC)

Artificial Bee Colony was one of the most recently established algorithms by Dervis Karabogain in 2005, motivated by the intelligent operation of honey bees. It is as simple as Particle Swarm Optimization (PSO) and Differential Evolution (DE) algorithms and accepts only standard control parameters such as colony size and maximum cycle number. ABC as an optimization tool presents a population-based search procedure in which the artificial bees modify individuals called food positions with time, and the bee aims to discover the places of food sources with high nectar amount and finally the one with the highest amount of nectar. In ABC system, artificial bees fly around in a multidimensional search space, and some (employed and onlooker bees) choose food sources depending on the background of themselves and their nest-mates and adjust their positions. Some (scouts) fly and pick the food sources randomly without using experience. If the nectar amount of a new source is higher than that of the previous one in their memory, they memorize the original situation and forget the previous one. Thus, ABC system combines local search methods, carried out by operating and observer bees, with global search algorithms, managed by onlookers and scouts, attempting to balance the examination process (Loubière, Jourdan, Siarry, & Chelouah, 2016).

Table 2.8: The Pros and Cons of Artificial Bee Colony(ABC) (Loubière et al., 2016)

Advantages	Disadvantages
Simplicity, flexibility and robustness.	Getting trapped into several local optima.
Ease of hybridization with other optimization algorithms.	Using fixed parameters and they do not change with the time.
Ease of implementation with basic mathematical and logical operations.	ABC is not that smart to remember the path which lead to a good result. So in the next move it will try the other path regardless of the good path which find out before.

2.3.2.4 Particle Swarm Optimization (PSO)

Particle Swarm Optimization is a population-based stochastic optimization algorithm developed by Dr. Eberhart and Dr. Kennedy in 1995, motivated by social behavior of bird flocking or fish schooling. PSO shares many communities with evolutionary calculation algorithms such as Genetic Algorithms (GA). The system is initialized with a population of random solutions and explorations for optima by refreshing productions. However, unlike GA, PSO has no development operators such as crossover and mutation. In PSO, the potential solutions, called particles, fly within the problem location by following the current optimum particles. Each particle holds path of its coordinates in the difficulty area which is correlated with the best solution (fitness) it has produced so far. (The robustness assessment is also saved.) This value is called pbest. Another "best" value that is followed by the particle swarm optimizer is the best value, gained so far by any particle in the neighbors of the particle. This location is called lbest. When a particle takes all the population as its topological neighbors, the best value is a global best and is called gbest (Couceiro & Ghamisi, 2016).

Table 2.9: The Pros and Cons of Particle swarm optimization (PSO) (Couceiro & Ghamisi, 2016)

Advantages	**Disadvantages**
In big application, PSO in a shorter time are able to solve the problem than ABC.	It produces unnecessary fluctuation of particles when could solve the problem.
	High accuracy which results from more sophisticated finite element formulation.

2.3.2.5 Ant Colony Optimization Algorithm (ACO)

Ant Colony Optimization was invented in the 1990s by M. Dorigo. For combinatorial problems which are considered as hard for optimizing, ACO has applied meta-heuristic to find the optimum solution which is good enough with reasonable computation time and cost for them. It uses some computational agents (which can be simulated as real ants) and dynamic memory structure to construct the process for each agent. It takes ethnologists approach to find the shortest path to the optimum point by means of the most passed path. It uses a positive feedback loop to increase probability of finding optimum solution (Engelbrecht, 2005; Yang, 2010).

Basically, ants have always communicated to find the shortest path to the food. When each ant passes a way to the food, it puts pheromone on the ground which guides other ants to find the way to the food. More pheromone in a path, more probable the shortest path to the food. ACO is used for image processing like image compression, image segmentation, and image edge detection. It is also used for optimization of continuous problems, so it can be used for various applications of image processing which show continuous behavior (Kaur, Agarwal, & Rana, 2011).

Table 2.10: The Pros and Cons of Ant Colony Optimization(ACO) (Kaur et al., 2011)

Advantages	Disadvantages
Positive Feedback accounts for rapid discovery of good solutions Efficient for Traveling Salesman Problem and similar problems In dynamic application it can also be applied	Theoretical analysis is difficult.

In conclusion, the Nature-inspired meta-heuristic algorithms have gained popularity because of their ability to deal with nonlinear global optimization problems (H. Kaur et al., 2013). We have briefly reviewed the some popular naturally inspired algorithms and their application in feature selections.

2.3.2.6 Summary of feature selection algorithms

The result arises from this part can be concluded as follow. Firstly, by reducing the number of redundant features the execution time of a face recognition system minimizes considerably (Dash & Liu, 1997). So, removing the unnecessary features helps to have a better insight into the underlying concept of a real-world classification problem. Then, it was discussed that there is a huge interest in biological system behaviors to develop meta-heuristic algorithms (Kaur et al., 2013). Consequently, several algorithms in this category were introduced such as Particle Swarm Optimization (PSO), Artificial Bee

Colony (ABC), Artificial Fish Algorithm (AF), Bat Algorithm (BA), and Ant Colony Optimization (ACO). Finally, to select the proper feature selection algorithm it is required to compare the results of these algorithms together regardless of type of the dataset and classifier which were applied. In a study conducted by Abd-Alsabour & Randall (2010), ACO outperformed PSO and Genetic Algorithms. In other study by Sen & Mathur (2016), ACO could produce better results than ABC. Therefore; having a better result in compare to other algorithms provokes this study to use ACO as the main feature selection algorithms.

2.3.3 Feature Classification Algorithm

This section discusses the main algorithms in face recognition system. The main algorithms include Eigenface, Artificial Neural Networks (ANN), Dynamic Link Architecture (DLA), Hidden Markov Model (HMM), Support Vector Machine (SVM), template matching, and graph matching. Also, these algorithms have been analyzed in term of facial representation. Furthermore, advantages and disadvantages of each algorithm are presented.

2.3.3.1 Eigenface

Eigenface is the most conventional approach for face recognition systems. Sometimes it is referred to as Karhunen-Loève expansion. Eigenface has been applied to represent a feature of the face (Kirby & Sirovich, 1990; Sirovich & Kirby, 1987). Basically, each face can be represented by a small summation of weights which are obtained by projection of edge picture. A face identification and detection has been proposed by Kirby-Sirovich algorithm (Rahman, Rahman, Safar, & Kamruddin, 2013). Mathematically, an eigenface is the eigenvector of the covariance matrix some set of the face images. This is also known as a principle component of the face distribution. By ordering the eigenvectors in various variation among the faces, a face can be expressed by a combination of linear eigenfaces. There is a possibility to use the largest eigenvalues and best eigenvectors which have M-dimensional space. Although a study has shown 96%, 85%, and 64 % face identification for lighting, orientation, and size variation, illumination is important for the performance of face recognition with eigenface (Kirby & Sirovich, 1990). A new method was developed by computing covariance matrix of three images which is taken in various lighting conditions to reduce the effect of illumination (Zhao & Yang, 1999). This method is extended by integrating eigenface to eigenfeature including eyes, mouth, and nose (Pentland, Moghaddam, & Starner, 1994). Actually, the eigenfeature is composed of eigengenes, Teignmouth, and Eigen nose which is immune to any appearance changing. The main advantages of eigenface are simplicity and practically. However, invariance to scale and lighting is the main disadvantage of eigenface algorithm.

2.3.3.2 Graph Matching

Another approach for face recognition is known as graph matching. Elastic graph matching applies Dynamic Link Structure (DLS) to recognize an object based on the closest graph which is found in the database. Actually, DLS is an extended version of ANN. Each stored object in the database is represented by multi-resolution vectors which are labelled. Elastic graph matching formulates objects for recognition based on cost function which is optimized by stochastic. The main advantage of this algorithm is the superiority of the recognition performance over other face recognition algorithms in terms of invariance and rotation. However, computational complexity, which is increasing the matching cost, is the main disadvantage of this algorithm (von der Malsburg, 2014).

2.3.3.3 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a strong mathematical tool for pattern recognition systems. SVM always finds the best hyper-link where it has the maximum distance to both classes. This hyper-link is known as Optimal Separating Hyperplane (OSH). SVM not only is useful for limited training data but also can use a decomposition algorithm like Quadratic Programming (QP) to guarantee the optimality of face classification (Lin, 2001). Sometimes a binary tree is used with SVM to improve the face recognition performance (Srinivasan, 2015). SVM is used as a classifier with eigenface features to improve face recognition rate. Another PCA face recognition approach has been compared to a SVM-based approach, with SVM outperforming PCA (Phillips, 1998). The main advantages of SVM are relevant discriminatory capability, the ability to generalization with over-trained data, less training time, and employing different kernel for mapping data to higher space and also in comparison with other classification methods it is shown that the SVM achieves a higher level of classification accuracy than other classifiers, but it can only be used with high-dimensional and small datasets(Pal & Mather, 2005).

2.3.3.4 Artificial Neural Networks (ANN)

Generally, ANN is the most attractive and efficient method for face recognition. A single layer adaptive ANN known as WISARD is presented for face recognition and it is based on a separate network for each individual (Stonham, 1986). The architecture of ANN affects the performance of the system.

Depending on the application different ANNs have been applied. For example, for face detection purpose convolutional NN or multilayer perceptron (Sirovich & Kirby, 1987), and for face verification, multi-resolution pyramid structure is applied (Weng, Ahuja, & Huang, 1993). A self-organizing map (SOM) can provide topological space by quantization of images. However, dimension reduction is necessary for quantization of the image samples. The convolutional network can provide invariance to rotation, scale, translation, and deformation partially by using a large set of layers hierarchically. Probabilistic Decision-Based Neural Network (PDBNN) is also applied to improve the recognition performance by using modular structure (Lin, Kung, & Lin, 1997). The

PDBNN is so efficient for three main reasons. First, it can be applied for face detection by using clustered image. Second, it works as eye localizer which finds the positions of eyes for producing feature vectors. Third, it is used for face recognition by dividing the network into the N different sub-nets which are assigned to each person specifically. The likelihood of each sub-net is computed by using a mixture of Gaussian density which provides a flexible and complex model for face recognition. The PDNN has two phases including training, and decision. Hence, using ANN and statistical approaches, it is easy to implement the PDNN in a parallel way to improve the learning time (Chen, Shu, Chen, & Ge, 2014). The main advantage of ANN is its recognition rate and efficiency. However, with increasing the number of subjects, the demand for computing is increased. Furthermore, a multiple images model is required to reach the optimum training condition. The ANN is widely used in medical imaging field and stock market prediction.

Medical Imaging Field: Artificial Neural Networks (ANNs) has a significant role in the medical imaging field. For example, it is applied in the system to analyze and diagnose some organs from medical images since it is not easy to distinguish (Raja & Rajagopalan, 2014).

Stock Market Prediction: Stock market would be affected by many factors and the rate would go up and down daily. Since ANN can examine and sort a lot of information quickly, so it can be used to predict prices (Abhishek, 1992).

There are different architectures of ANN models for face recognition systems. The main ANN algorithms for face recognition includes Retinal Connected Neural Network (RCNN), Rotation Invariant Neural Network (RINN), Principal Component Analysis with ANN (PCA & ANN), Fast Neural Networks (FNN), Convolutional Neural

Network (CNN), Evolutionary Optimization of Neural Networks, Multilayer Perceptron (MLP), Back Propagation Neural Networks (BPNN), and Cascaded Neural Network.

a) Retinal Connected Neural Network (RCNN)

Retinal Connected Neural Network (RCNN) is developed by arbitrating among many neural networks to enhance face recognition performance. For training process, an algorithm of bootstrap has been applied to reduce the false detection. This algorithm avoids using non-face images for training. Basically, RCNN can help to differentiate between non-face and face images. Figure 2.3 presents RCNN for face recognition systems.

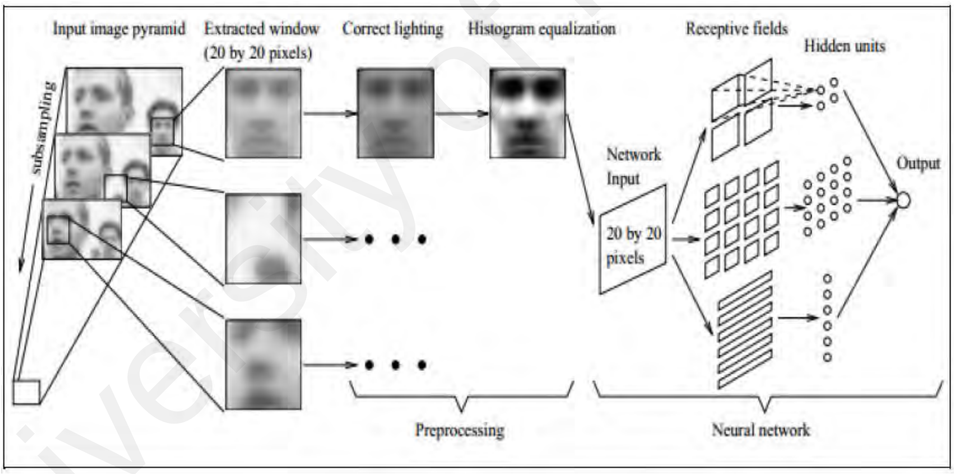


Figure 2.3: RCNN for face recognition system (Rowley, Baluja, & Kanade, 1998).

b) Rotation Invariant Neural Network (RINN)

Rotation Invariant Neural Network (RINN) has been developed to detect face in various rotation angels. This model applies router neural networks for normalization of the image plan orientation and then it is fed to multiple neural networks to process the facial image. Figure 2.4 shows the main processes in RINN model.

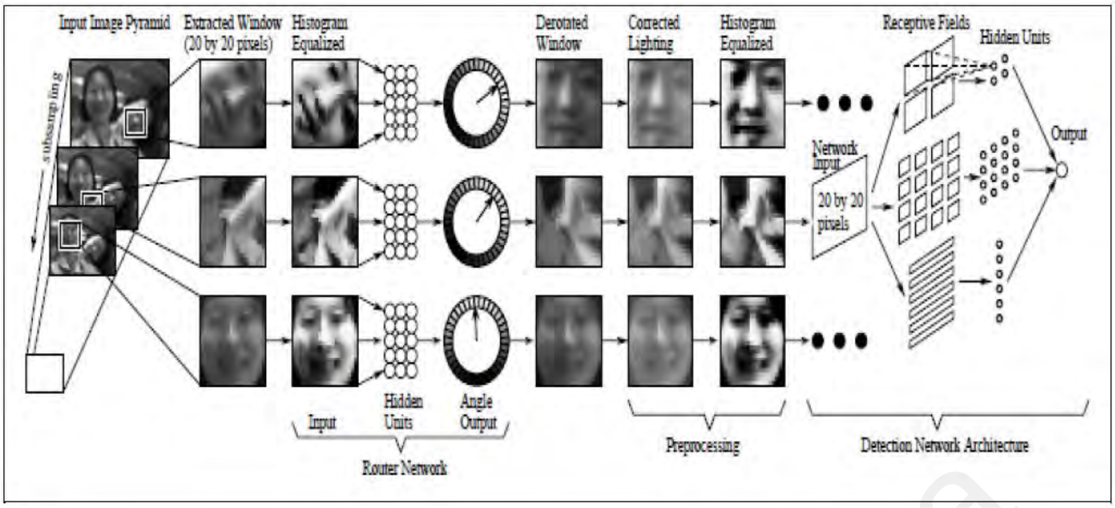


Figure 2.4: RINN for face recognition system (Rowley et al., 1998).

c) Fast Neural Networks (FNN)

Fast Neural Networks (FNN) has been developed for real-time face recognition systems due to less computational time and complexity. For this reason, every image is segmented to various numbers of sub-images. Then, each of these sub images are fed into the FNN.

d) Polynomial Neural Network (PNN)

Polynomial Neural Network (PNN) has been used in binomials projection to map the local image into a space by applying PCA. Due to large computational complexity, this model has seldom been used for face recognition systems.

e) Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) has been applied for robust face recognition system because of its properties for resisting against rotation, translation, and scale invariance. However, this model is a rule-based model which needs strong effort to develop. Figure 2.5 shows different steps for CNN model.

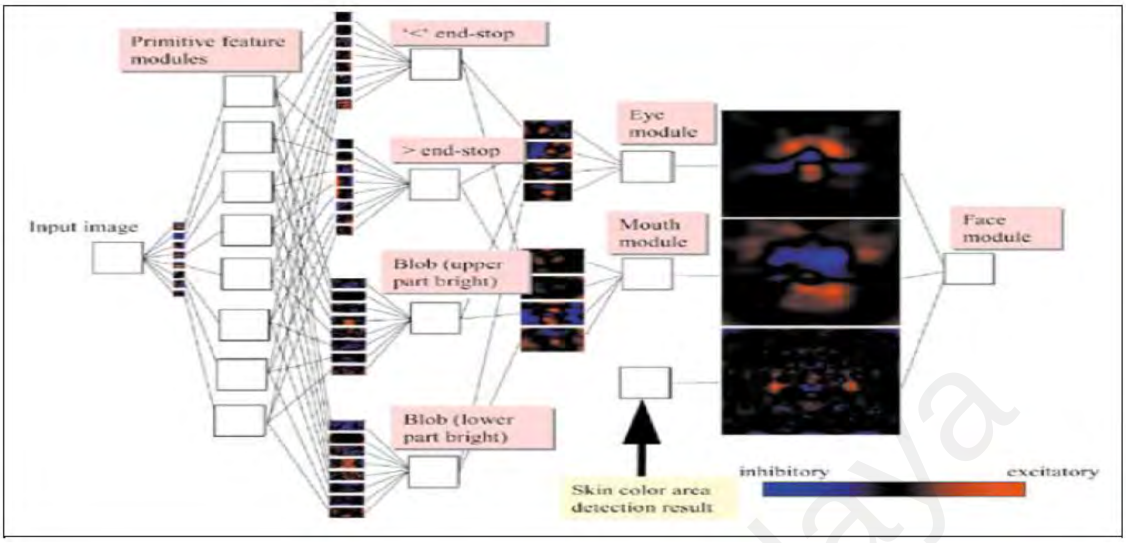


Figure 2.5: CNN for face recognition system (Matsugu, Mori, Mitari, & Kaneda, 2003).

f) Back Propagation Neural Networks (BPNN)

Back Propagation Neural Networks (BPNN) has been used for three face representations including eigenfaces, pixel, and partial profile independently to train the model. Each BPNN is trained based on Gaussian Mixture Model (GMM) for segmenting the image. Figure 2.6 shows BPNN for face recognition systems.

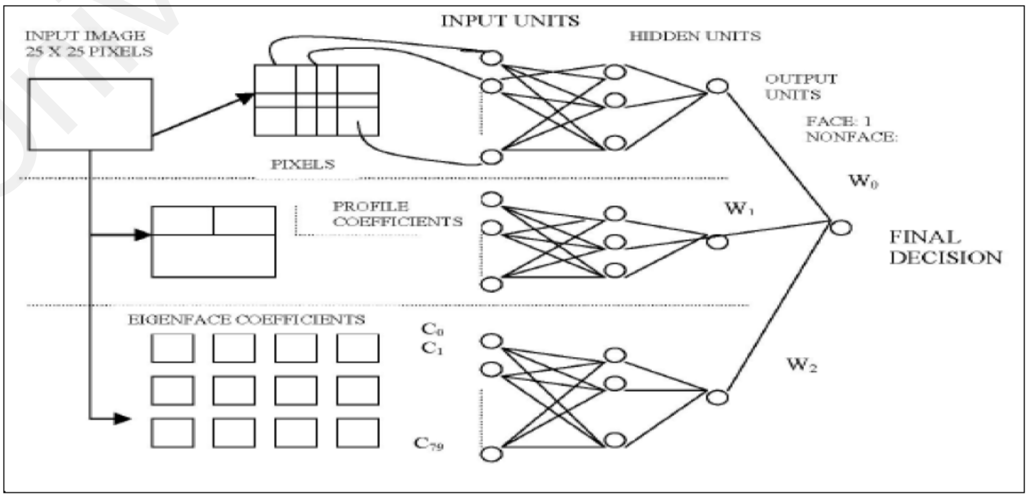


Figure 2.6: BPNN for face recognition systems (Bojkovic & Samcovic, 2006).

2.3.3.5 Summary of classification algorithms

As a conclusion of this part, some of the most famous classifiers were discussed then ANN as one of the robust classifiers was introduced. Artificial Neural Networks (ANNs) play an essential role in the medical imaging field, including medical image analysis and computer-aided diagnosis, because objects such as lesions and organs in a medical image may not be represented into an accurate equation easily (Raja & Rajagopalan, 2014). In the other study about measuring the expression levels of thousands of genes from DNA microarrays, ANN produces more accurate result than SVM (Dond et al., n.d.). ANN not only have a great result in medical field but also in stock market prediction field has a significant result since ANN can examine and sort a lot of information quickly, so it can be an effective tool for stock market prediction (Abhishek, 1992). Therefore; according to the efficient result of ANN in the previous study in other fields and to experiment a new combination of ACO as an optimizer for the classifier, it was decided to use ANN in the study.

2.3.4 Face Corpuses

In this section, various face databases are introduced. Basically, each database has special properties which have been designed for specific purposes. Face databases have a variety of illumination angles, colors, poses, and face occlusions but they do not annotate about its pose angles which may be considered as a limitation of them. For designing a good experimental setup, having some knowledge about characteristics of each face database can improve our fairness. Based on the study by (Black Jr, Gargesha, Kahol, Kuchi, & Panchanathan, 2002), four types of illumination including daylight, fluorescent, incandescent, and skylight have been used in face databases. Some of these face databases are publically available for download over the internet (Black Jr et al., 2002).

2.3.4.1 The ORL / AT&T Database

The AT&T face image database (formerly known as the ORL database) contains a set of face images taken between April 1992 and April 1994 from 40 people. There are 10 different images of each person taken at different times. Some of the images have varying illumination, facial expressions like open or closed eyes or glasses, or have different emotion expressions like smiling or frowning. Some of the images may also have rotations up to 20 degrees. All of the images have a dark homogeneous background and the subjects are in an upright, frontal position. The size of each image is 92×112 pixels, with 256 grey levels per pixel. Some samples of the images in this database are shown in Figure 2.7 (Roychowdhury & Emmons, 1991).

Figure 2.7: Sample picture from ORL dataset(Roychowdhury & Emmons, 1991).

2.3.4.2 The FERET Dataset

The FERET database of images includes a large number of images and subjects with different variations. The FERET program ran from 1993 through 1997 to develop automatic face recognition capabilities that could be employed to assist security, intelligence and law enforcement personnel in the performance of their duties. The FERET database of images consists of 14051 eight-bit grayscale images of human heads with views ranging from frontal to left and right profiles. The size of the pictures is 256×384 pixels. The pictures are from 1209 people, taken at different time, illuminations, and facial expressions(Roychowdhury & Emmons, 1991).

Figure 2.8: Sample picture from FERET dataset(Roychowdhury & Emmons, 1991).

2.3.4.3 The Yale B Database

The Yale B Face database contains 5760 images from 10 individuals. Each subject has been represented under 576 observing positions which are nine poses and 64 illuminations per pose. Single light sources have been employed in different angles for the illumination variations. The Extended Yale B database is an expanded version of the Yale B Database with 28 more subjects and includes 21888 single light source images of 38 subjects each seen under 576 viewing conditions similar to the Yale B database.

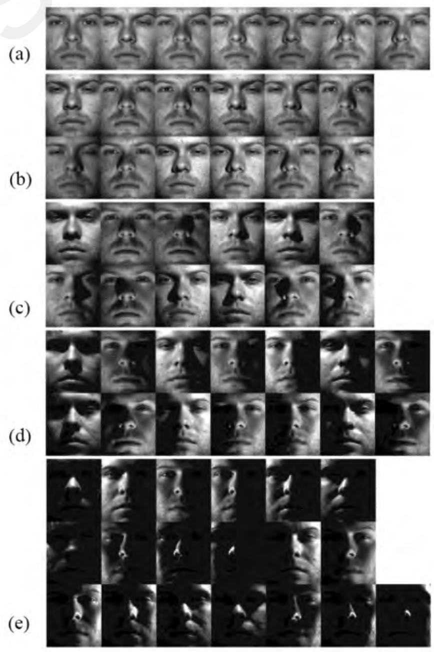


Figure 2.9: Sample picture from Yale dataset(Roychowdhury & Emmons, 1991).

Table 2.11: A comprehensive summary of datasets(Roychowdhury & Emmons, 1991)

Database name	Number of individual	Description	Limitation
AT&T (formerly ORL)	40 individual Each has 10 images	Always dark background Faces have been captured in different times and different illuminations	The inconsistency of illumination conditions. Lack of information about lighting, head and rotation conditions.
Oulu Physics	125 individuals Each has 16 images	Colour image, four various lighting conditions, four different cameras, grey background,	Lack of lighting angle Lack of posing angle and distance during face collection.
XM2VTS	295 individuals Each has 4 images	4-month period for each face data collection session, seriously differences of individuals, it is video and audio database.	Lack of any description about face capturing information.
Yale	15 individuals Each has 11 images	Variation in lighting position, emotion and spectacle	Small number of subjects Positions of lighting sources aren't described precisely No posing angle variation Environmental light is not describe
MIT	16 individuals Each has 27 images	Variation in head orientation, lighting, and scaling	The variations are not so extensive and not precisely described Many moving is available between images.
UMIST	20 individuals 564 images	Various posing angles, genders, and races	Posing angle is not exactly described. No information about illumination, colour, and direction is described.
The FERET	1000 individuals 14051 images	Variation in face frontal and time capturing. It has recorded during many years.	It does not have a large amount of posing variation. No lighting information is given.
Kuwait University face database (KUFDB)	50 individuals 250 images	Various lighting, size, rotation, and facial expression	Small number of individuals No any information about acquisition factors.

A comprehensive information about the existing datasets was presented in Table 2.11. As we can see each of them contains different number of pictures. Some have dark background and some color images.

2.4 Existing Face Recognition System Using Various Combined Algorithms

In this section, results of previous investigations are presented to have a better understanding of comparisons among face recognition algorithms. Table 2.12 presents a summary of comparisons among different face recognition algorithms that use ORL as their main dataset. As can be seen clearly, each of the face recognition algorithms is optimum and it is directly depending on the face database.

Table 2.12: Comparison among face recognition algorithms.

	Feature Extraction	**Feature Selection**	**Classification**	**Evaluation Method**	**Performance (recognition accuracy) and Result**
1	DWT (Kanan et al., 2007)	Ant Colony Optimization(ACO)	Nearest Neighbour Classifier	MSE(Mean squared error)	98.5%
2	PZMI HMI BMI ZMI TZMI NZMI (Nabatchian et al., 2008)	---	Artificial neural networks (ANN)		PZMI = 95%
3	PZMI	genetic algorithm	Radial Basis Function		DWT = 84.4%

	DWT (Faez, 2005)	(GA)	(RBF) neural networks		PZMI = 93%
4	DCT DWT (Agarwal & Bhanot, 2015)	Firefly	Nearest Neighbour Classifier		94.375%
5	DCT (G. Cheng, Shi, Zhu, & Gong, 2011)	Binary Particle Swarm Optimization (BPSO) Genetic Algorithm (GA)	Euclidean distance classification	Comparison Charts (Number of Features) (Training Time) (Recognition Rate)	BPSO = 94.5% BPSO > GA
6	CT-PCA DWT-PCA (Darestani et al., 2013)	Particle Swarm Optimization (PSO)	Multi-Layer Perceptron (MLP)	Confusion Matrix	1-CT-PCA has higher detection rate 2-CT-PCA method has shorter time training than DWT-PCA
7	-1D-PCA -2D-PCA (Kashem, Akhter, Ahmed, Alam, & Ntroduction, 2011)	----	K-Nearest Neighbour Classifier(KNN) Support Vector Machines(SVM)	Comparison Charts	1-Final results indicate that 2DPCA has a better performance in terms of accuracy and complexity.
8	ZMI Wavelet Transform Features (Haddadnia, Ahmadi, & Raahemfa, 2003)	----	SVM	Comparison Charts	1-DWT is faster than ZMI about 0.078 images per seconds. (around 11 times)

9	Local Binary Pattern (LBP) (Omer & Khurran, 2015)	BSA	Nearest Neighbour Classifier	Comparison Charts	85%
10	FCT PZMI (Kanan & Faez, 2005b)	----	Radial Basis Function (RBF) neural networks		99.3%
11	Local Binary Pattern (LBP) sub pattern-based PCA (spPCA) modular PCA (mPCA) (Eskandari & Toygar, 2014)	Histogram equalization (HE) mean variance normalization (MVN)	Radial Basis Function (RBF) neural networks		LBP = 91.5% spPCA = 83.5% mPCA = 82.5% PCA = 82%
12	Discrete Wavelet Transform (DWT) (Khadhraoui, Ktata, Benzarti, & Amiri, 2016)	modified particle swarm optimization (MPSO)	Euclidean distance		98.33%

In Table 2.12, a summary of all benchmark algorithms has been presented together. It is also categorized based on the feature extraction techniques. As it is shown below, the benchmarks which use DWT or PZMI as their feature extraction algorithm have the highest results for accuracy. But among them the combination of DWT, ACO as feature selection and nearest neighbor as classifier show reasonable result; however, it takes a long time to get this result (about 1320 seconds or 22 minutes). Using DWT, firefly and nearest neighbor also provides a reasonable result but the dataset used is a small one. In case of PZMI as the second feature extraction algorithm, the combination of PZMI, genetic algorithms and ANN provides a result slightly more than 90% but getting 120 features and in 20 minutes is not a good result. In the worth case of combinations, PCA and GA as the main feature extraction and selection with Naïve Bayesian Classifier could only produces 78.75% accuracy which is not acceptable for this study. In this study, we assume that a different set of algorithms would produce a better result.

Table 2.13: Summary of the performance of the existing benchmark algorithms.

Feature Extraction	**Feature Selection**	**Classification**	**Recognition Accuracy**	**Limitations**
DWT	ACO	NN (Nearest Neighbor)	95.5%, 42 features in 1320 sec	Taking a long time
	GA	RBF	84.4%	Accuracy not good
		KNN	90.5%	
	Firefly	NN (Nearest Neighbor)	94.37%	Using small dataset
	PSO	Euclidean distance	94.7%	
PZMI	GA	RBF	93%	
		ANN (Neural Network)	90.87%, 120 features, 20 min	Taking many feature
DCT	Firefly	NN	91%	Using small dataset
	BPSO	NN	94.5%	
PCA	GA	Naïve Bayesian Classifier	78.75%	Accuracy is low
	PSO	MLP	90%	

2.4.1 Evaluation method

The performance of the existing face recognition systems is measured by using the following statistics:

- ***Recognition Accuracy***

which is result of the number of correctly classified samples divided by the number of all classified samples(Patil et al., 2015) (Utsumi, Matsumoto, & Iwai, 2009).

$$RA = \frac{The\ Number\ of\ correctly\ matched}{Totall\ number\ of\ classified\ samples}$$

2.5 Summary

In this chapter, we discussed the most conventional face recognition algorithms. The benefits and drawbacks of each one is critiqued comprehensively. Furthermore, several face recognition databases are presented. The most relevant studies in feature selection were also discussed. In order to develop a robust methodology, this information is necessary to have a better understanding of the following chapters. As a result, our primary contribution in next section would be developing a face recognition system by finding a suitable combination of feature extraction, selection, and classification algorithms.

CHAPTER 3: RESEARCH METHODOLOGY

3.1 Overview

This chapter presents the research design and methodology to be adopted to fulfill and achieve all the research objectives. In this section, the proposed method is also presented. The process flow of this research is depicted in Figure 3.1. The research methodology adopted in this research contains several essential steps including the literature review, corpus selecting, and development of the face recognition system, evaluation, and discussion of the significant findings.

Stage I: Stage one is the preliminary focus on identifying the problem and establishing the fitting objectives for the research. A critical review of other publications and published academic papers in the face recognition domain was conducted to determine the scope of the study and significance of the problem.

Stage II: Stage two of the research mainly concentrated on having a survey on face recognition system algorithms. Recognizing various parts of the system, collecting more information about the existing algorithms and methods for each element of the system is the principal purpose of this stage. Eventually, performances of the methods were inspected and collected in a list as a benchmark and reference for the study.

Stage III: Stage three of the research is focused mainly on proposing a unique combination for the system. The central intent is to converge on the functionality of each algorithm individually and organize them for the usage in the face recognition system.

Stage IV: Developing the proposed method is carried out in this section. The performance of the proposed system will be analyzed and compared with the existing benchmark.

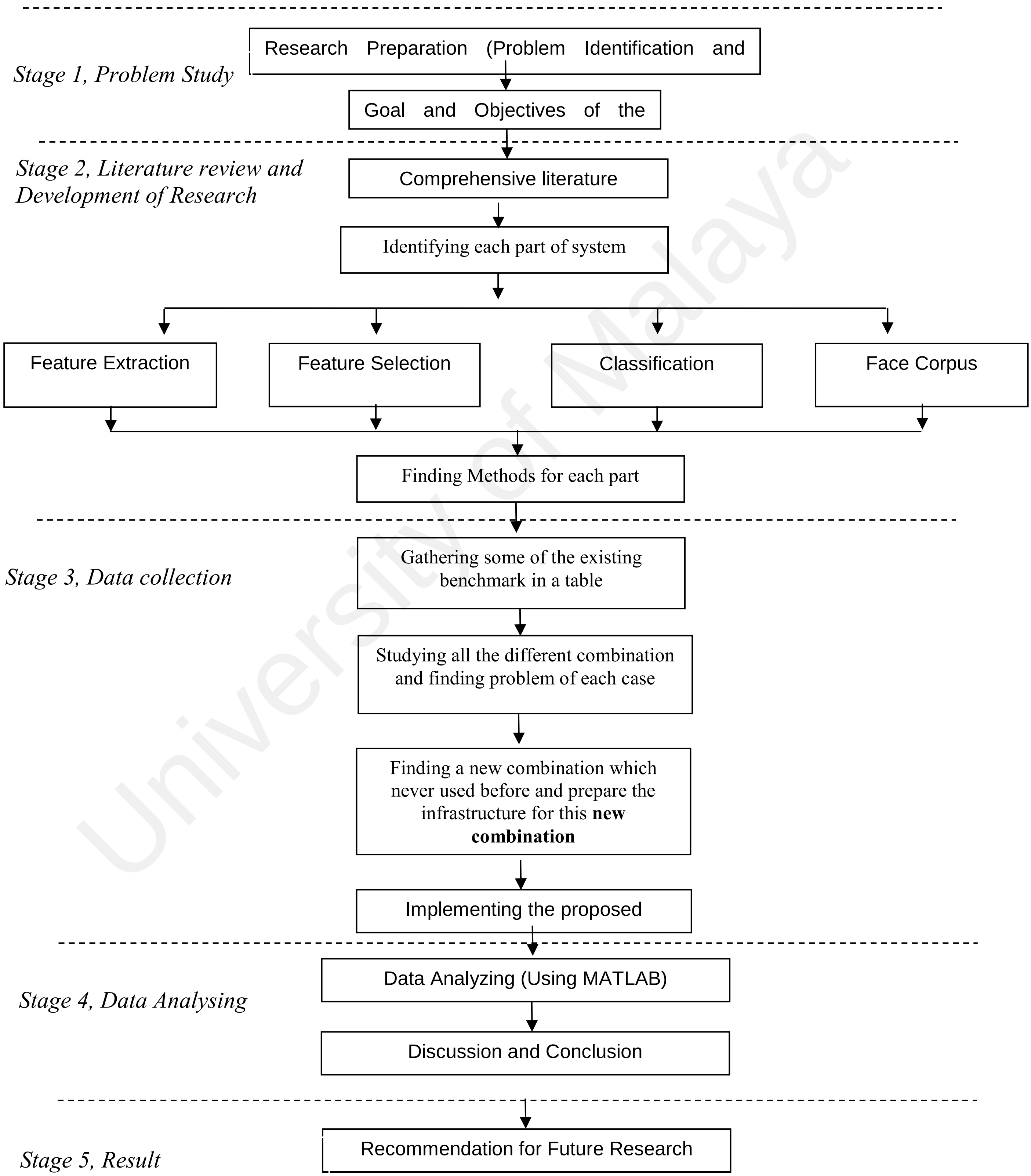


Figure 3.1: Overall Structure of Research Methodology of this research

3.2 Problem Identification and Solution

Feature selection has usually been applied to diminish the dimensionality of the feature set for face recognition systems. Therefore, feature selection is the central part of many face recognition systems. With having a massive database in a machine-learning algorithm, new difficulties will arise, and novel and proper procedures to choose valuable features are in demand (Dash & Liu, 1997). In many applications, the dimension of the dataset is so vast that learning might not operate as well. Hence, subtracting these unwanted features is a necessity for the system. Unfortunately, many of these features are entirely or partially redundant or irrelevant to the target theory. Currently, much attention is on biological system behaviors to develop meta-heuristic algorithms. Particle Swarm Optimization (PSO), Artificial Bee Colony (ABC), Artificial Fish Algorithm (AF), Bat Algorithm (BA), and Ant Colony Optimization (ACO) are some of the most up-to-date examples. In this section, the selected algorithms for the proposed system are considered separately. Then, ACO as one of the most efficient feature selectors is presented for the system. During the classification phase, the parameter computed from the feature selection stage is used to classify and recognize a specific picture. Researchers have usually employed several types of classifiers in developing face recognition systems. Up to now, there is no clear evidence proving that which classifier is the best for the system. Currently, ANN is known as one of the most reliable classifiers in many contexts like in medical and stock prediction field. Therefore, it could provide a good result in other types of domain (Raja & Rajagopalan, 2014).

3.3 Development of Proposed Method for Face Recognition System

As explained in the earlier section, every algorithm has its own advantages and disadvantages. According to Table 2.12 and 2.13 in chapter 2, DWT, PZMI, ACO, and ANN could produce satisfactory results individually or in collaborating with other algorithms. Therefore, this study intends to propose and set these algorithms together. Figure 3.2 shows the overall processes for the proposed face recognition system. So, the primary proposed combination in this study is to utilize DWT and PZMI as central feature extractors, ACO as the primary feature selector and ANN as the classifier. Consequently, the experiment and to observe the result of this combination on ORL dataset will be discussed in chapter 5 in details. Hence, the set of DWT/PZMI, ant colony optimization, and ANN is proposed to be experimented for the first time in the face recognition domain.

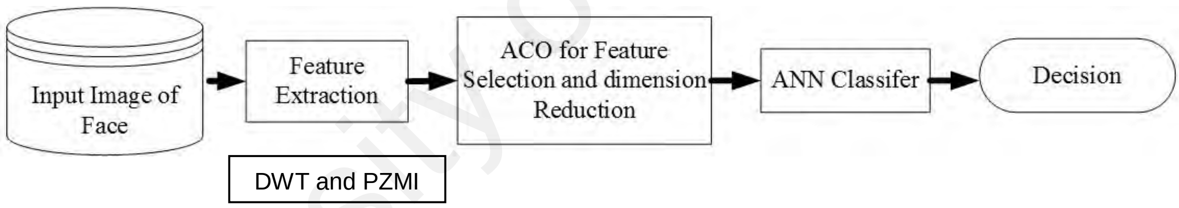


Figure 3.2: Block diagram of the proposed methodology.

3.3.1 Dataset Selection

The initial step towards the development of a face recognition system is to choose an appropriate database. ORL is a standard dataset that is very popular and has been employed in many kinds of research (Haddadnia et al., 2003; Kanan et al., 2007; S. Kaur et al., 2011). This face dataset holds 400 images of 40 people. There are ten distinct images for each person in the size of 112x92 pixels. In some cases, the images have been taken at different times. In this dataset, some images have various kinds of facial expression, and appearance. For example, some images contain open or closed eyes and smiling or frowning. Moreover, in some cases the individuals wear glasses in the images. All images were taken against a dark homogeneous background. Images are in an up-right frontal position with tolerance for some side movements. There are also some variations in scale (Jalled, 2017). ORL database has sufficient samples to examine the correctness or the accuracy of the face recognition research. Therefore, this study used ORL dataset as its central repository of sample data.

Feature Extraction

Extracting unnecessary features from images is a vital segment in a recognition system. Hence, picking the right feature extractor is a crucial task to produce a high rate of accuracy. High-grade feature extractors should be efficient, invariant, robust and accurate in order to enhance the recognition rate. For example, in some of the time-critical applications, these algorithms should be susceptible and accurate to improve the accuracy. As it was shown in Table 2.12, there was a significant difference between the result of the studies which applied DWT and PZMI as the feature extraction algorithms and the others. Based on this indication, this study aims to use and examine these two algorithms. These two algorithms are explained below.

3.3.1.1 DWT

The wavelet first mentioned by Haar in 1909. Then later, Mallet could make a notable mutation in the digital signal processing by formulating the “pyramidal algorithms.” Next Daubechies used Mallet's outcome to develop a set of wavelet functions that are the foundation of wavelet application today. Figure 3.3 presents the overall feature extraction process for DWT (Kanan & Faez, 2005b). As it is shown, three levels of DWT with Haar wavelet function are implemented to decompose the image to approximation coefficients and detail coefficients. Then, the feature vector is built by the approximation coefficients section. The output of this action contains four sections which are one approximation band called LL band which is constructed by low frequency, and three specific bands called LH, HL and HH bands which are made by high frequency (Patil et al., 2015).

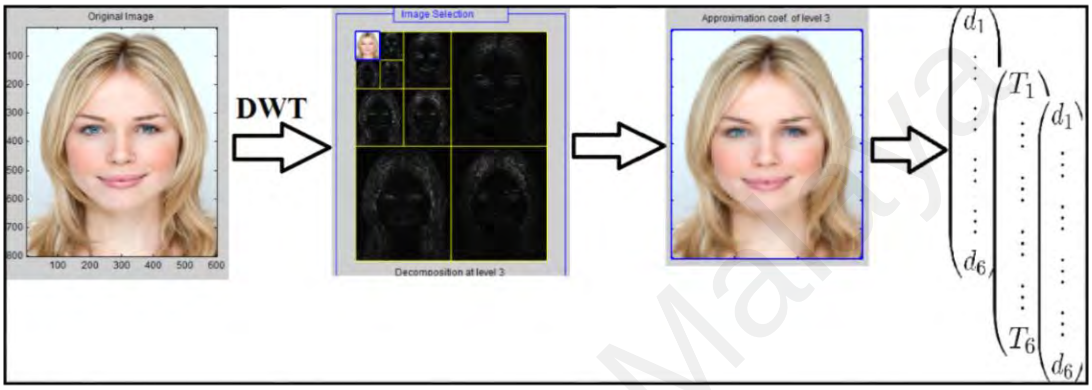


Figure 3.3: Feature extraction - DWT(A. Kaur & Kaur, 2012)

A set of wavelet basis functions is formulated as Equation (1)

$$\psi_{ab}(x) = \frac{1}{\sqrt{a}}\,\psi\left(\frac{x-b}{a}\right) \quad (1)$$

3.3.1.2 PZMI

Pseudo-Zernike moments invariants (PZMI) is one of the most remarkable feature extraction algorithms that are invariant to rotation, scaling, and translation which has an outstanding performance in analyzing images because the rate of efficiency of the classifier is directly related to the rate of the correctness of these extracted futures. Moreover, PZMI has been extensively utilized in the area of computer vision and robotics such as object identification algorithms. It is very sensitive to the pattern features so it can be easily applied to the pattern recognition systems. Zernike and pseudo-Zernike orthogonal polynomials are the basis of the Zernike moment invariant

(ZMI) and pseudo-Zernike moment invariant (PZMI) (Wallin & Kübler, 1995). Among all type of the moment's algorithm, Pseudo Zernike Moment Invariants can be employed to represent an image with the lowest level of redundancy information.

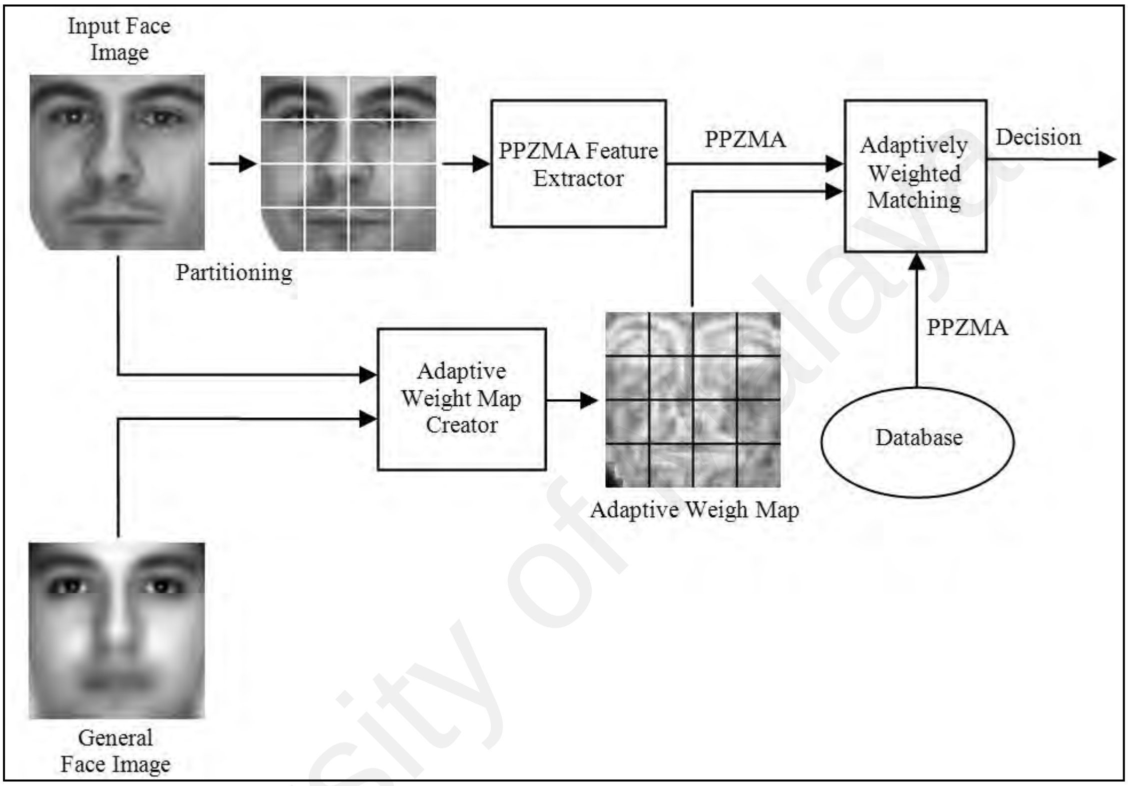


Figure 3.4: Feature Extraction– PZMI (Kanan & Faez, 2005a)

Pseudo Zernike Moment Invariants is expressed as Equation (2) (Faez & Farajzadeh, 2007)

$$V_{nm}(x, y) = R_{nm}(x, y)\exp(jm\ \tan^{-1}\frac{x}{y}) \qquad (2)$$

Where

Vnm = The Zernike function

Rnm = The Zernike radial polynomials

The Zernike radial polynomials will be analyzed at each pixel position given that Zernike moments stated in polar coordinates. By using a square-to-circular image transformation Zernike polynomial mainly requires being computed once for all pixels mapped to the same circle. Figure 3.5 shows the schematic of square-to-circular image transformation(Srinivasan, 2015).

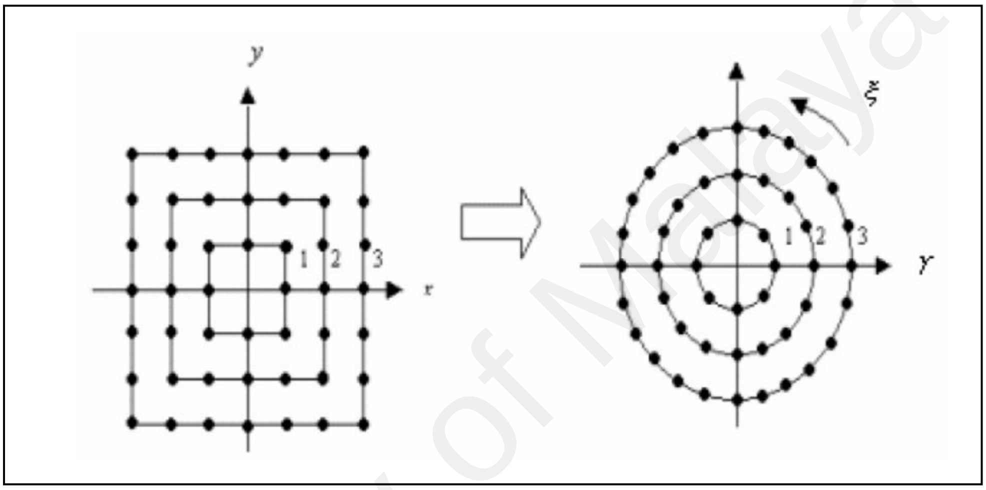


Figure 3.5: square-to-circular image transformation(Jana & Sinha, 2014)

3.4 Feature Selection

Feature selection has usually been utilized to lessen dimensionality of the feature set in a face recognition system. Therefore, feature selection is the vital part in these systems. Due to having an enormous database in the machine learning domains, new challenges transpire, and it requires innovative and proper approaches to select suitable features (Dash & Liu, 1997). In many applications, the size of a dataset is extremely big that learning might not operate before deducting these undesired features. Unfortunately, many of features extracted from the images are entirely or partially unnecessary or irrelevant to the target theory. Currently, much attention is paid to biological systems behavior to develop meta-heuristics algorithms. Invasive Weed Optimization (IWO)

Bees algorithm (BA), Glowworm Swarm Optimization (GSO), Firefly Algorithm (FA), Bat Algorithm (BA), and Ant Colony Optimization (ACO) are the most recent examples. Nevertheless, in this section, ACO as one of the most efficient and practical feature selectors is discussed.

3.4.1 Ant Colony Optimization (ACO) Algorithm

ACO algorithm can properly reformulate feature selection algorithm as discussed in chapter 2. The main steps in ACO algorithm for feature selection are as follows:

1. Initialization
2. Estimating the number of required ants (n)
3. Setting initial value of pheromone intensity for the first trial of any feature
4. Defining the maximum amount of iteration (i)

3.4.1.1 Production of ants and assign a feature for each ant

Each ant has been assigned to a feature in order to visit all the features one by one and to find a general solution. To estimate the value of this solution a measurement is required. Mean Square Error (MSE) estimator is used to evaluate the solution found by ants. If MSE during the iterations decreases then the ant's work will be terminated.

3.4.1.2 Evaluate the selected feature's subset for each ant

The significance of the selected feature is evaluated for each ant using performance of the classifier. After that, the solutions which produce lower MSE rate are arranged in an ascending order to be chosen based on ACO algorithms. This can help to avoid unlimited running of a loop.

3.4.1.3 Updating the ant's pheromone

The pheromone intensity of the ants, which are selected in step 3, are updated.

Go to step 2

The overall frameworks have been presented in Figure 3.6. The details of each step have been briefly discussed above.

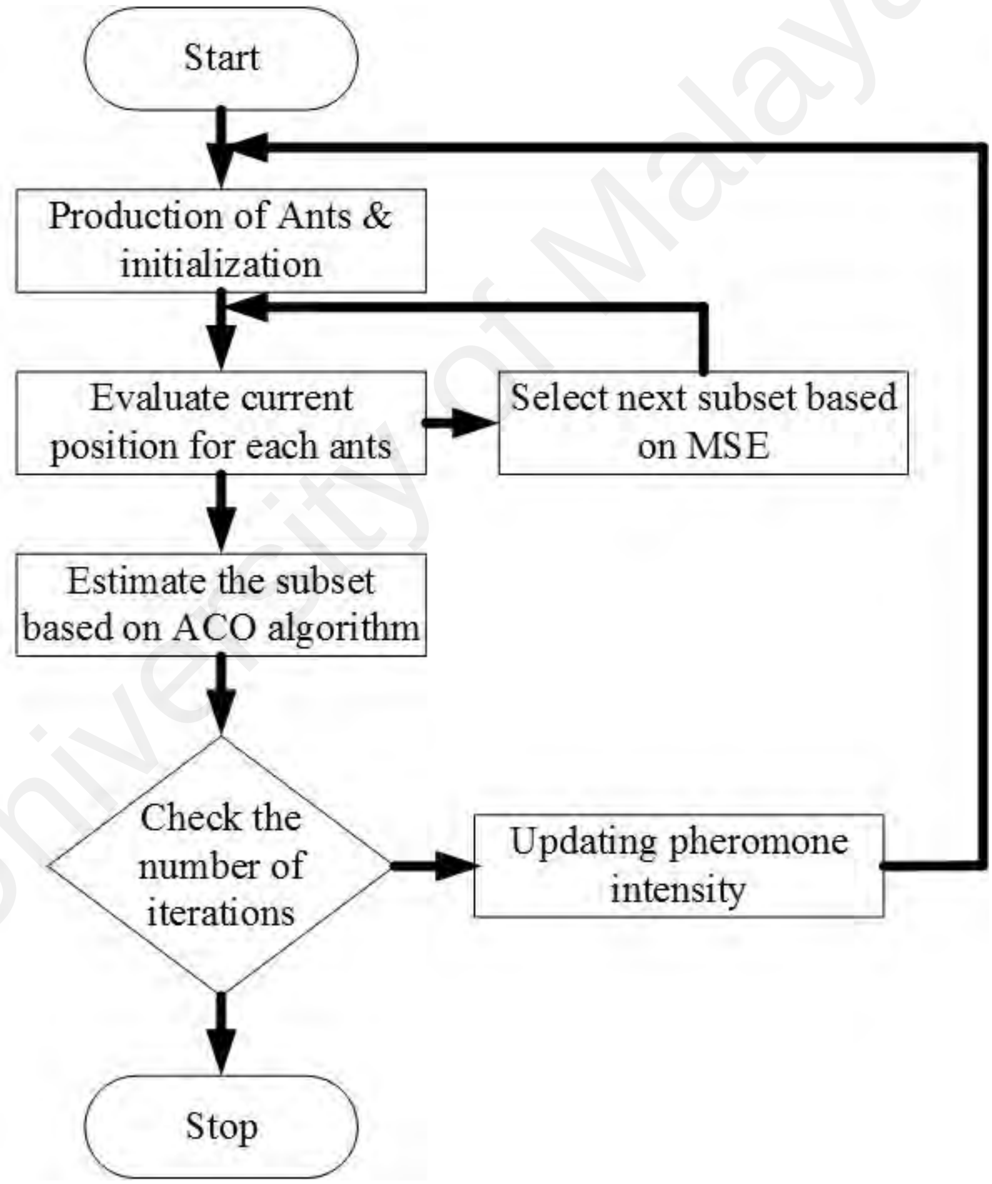


Figure 3.6: Standard framework for ACO(Sen & Mathur, 2016)

3.5 Classification

3.5.1 ANN

ANN is a nonlinear classifier based on the neurons network. Stimulus propagation through each direction of the network can fire the neurons to classify the parallel problem way. The main paradigms with environment include unsupervised, supervised, and reinforcement learning. Figure 3.7 shows architecture of a nonlinear neuron. As seen, each neuron has various parts such as neuron input, weights, a transfer function, threshold and activation function.

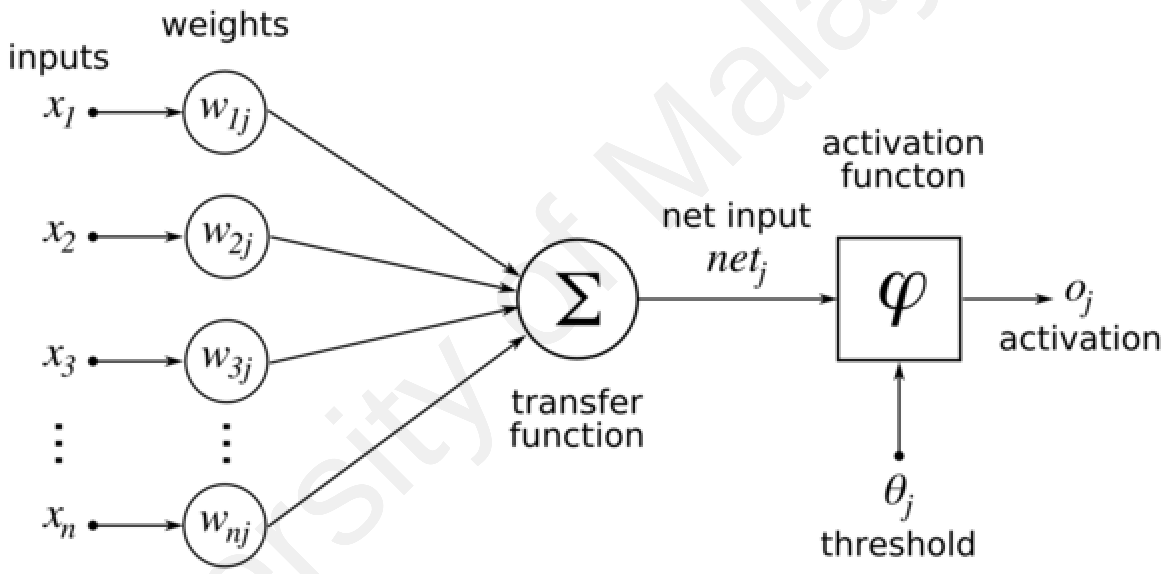


Figure 3.7: Architecture of a nonlinear neuron (Afroge, Mamun, & Mat, 2015)

A multi-layer feed-forward network is trained by gradient descent for estimating a classification function for problems by using some enrolment data based on pairs (x,t). The input (vector X) pattern is fed into the network, then a target or desired output is generated. Therefore, the gradient is computed based on training set overall by summation the gradients for each input vector. In the following, computation of the gradients has been described based on a single input vector. It must be mentioned that the weight from unit j to unit i is denoted by wij.

Basic Definitions is expressed as equation (3)

$$\delta_j = f'_j(net_j) \sum_{i \in P_j} \delta_j \, \omega_{ij} \qquad (3)$$

For calculating the error for unit j, all of the posterior nodes errors must be computed (forming the set Pj). Due to lack of cycle in neural networks, output error is fed back to the network. For this reason, the reverse manner is simply applied using propagated forward. In order to develop a feed forward networks, every node should contact to the next layer. For this reason, a back-prop algorithm based on matrix notation is applied without applying more graph which is generally used as matrix form.

3.6 Evaluation Method

In this research, two methods of evaluation have been used. First, the performance of each feature extraction algorithm as one of the leading parts of the system is measured. Second, the accuracy of the entire proposed system is calculated. The evaluation method which has been employed is calculated as follow:

- Recognition Accuracy

which is the result of the number of correctly classified samples divided by the number of all classified samples (Patil et al., 2015) (Utsumi et al., 2009).

$$RA = \frac{Number\ of\ correctly\ matched}{Totall\ number\ of\ classified\ samples} \times\ 100\%$$

3.7 Summary

Choosing the right research methodology facilitates a successful completion of the research by highlighting vital activities and considering concepts and theories. In this chapter, a novel combination of algorithms is proposed to be utilized in the face recognition domain. Before the development of the system, it is expected to detect and determine the proper algorithms and methods for each particular component of the system. Therefore, DWT and PZMI were selected for feature extraction part due to having a high benchmark result. Moreover, ACO and ANN also generate a powerful benchmark in other domains. Hence, the set of DWT/PZMI, ant colony optimization, and ANN is proposed to be experimented for the first time in the face recognition domain.

CHAPTER 4: EXPERIMENTAL DESIGN

4.1 Overview of

This chapter is composed as follows: Firstly, the main reasons for choosing the dataset is addressed. Secondly, the process of implementing the feature extraction algorithm is thoroughly explained. Then, implementation of ACO as feature selector and ANN as classifier is reviewed.

4.2 Experimental Setup

The experiments are implemented under the Mac OS EL Capitan operating system using MATLAB version R2016a and on Intel (R) CPU Core i5 2,6 GHz with 16.0 GB RAM. As it is shown in Table 4.1, four types of experiment are conducted in this study is listed. At first, the accuracy rate of DWT+ANN and PZMI+ANN is obtained then ACO as a feature selector is applied and the accuracy rate is examined again.

Table 4.1: Different type of experiments in this study

Experiment	Feature Extraction	Feature Selection	Classifier
1	DWT	------	ANN
2	PZMI	------	ANN
3	DWT	ACO	ANN
4	PZMI	ACO	ANN

Figure 4.1: Configuration of the system

4.3 Dataset

ORL dataset contains ten separate images of 40 people which taken at different times between April 1992 and April 1994 at the lab. Some of the images have differing lighting, facial appearances like open or closed eyes or wearing glasses, or have different emotional expressions like smiling or frowning. Some of the images may also have turned up to 20 degrees. All of the images have a dark homogeneous background, and the subjects are in an upright, frontal position. The size of each image is 92×112 pixels, with 256 grey levels per pixel. These images were grayscale which is shown in Figure 4.2. To create a face recognition system all 40 images of individuals have been used.

4.4 Feature Extraction based on PZMI and DWT

As discussed in chapter 3, two kinds of feature extraction algorithm have been employed to analyze and develop the feature vectors for the system. Those are Pseudo Zernike Moment Invariant (PZMI) and Discrete Wavelet Transform (DWT).

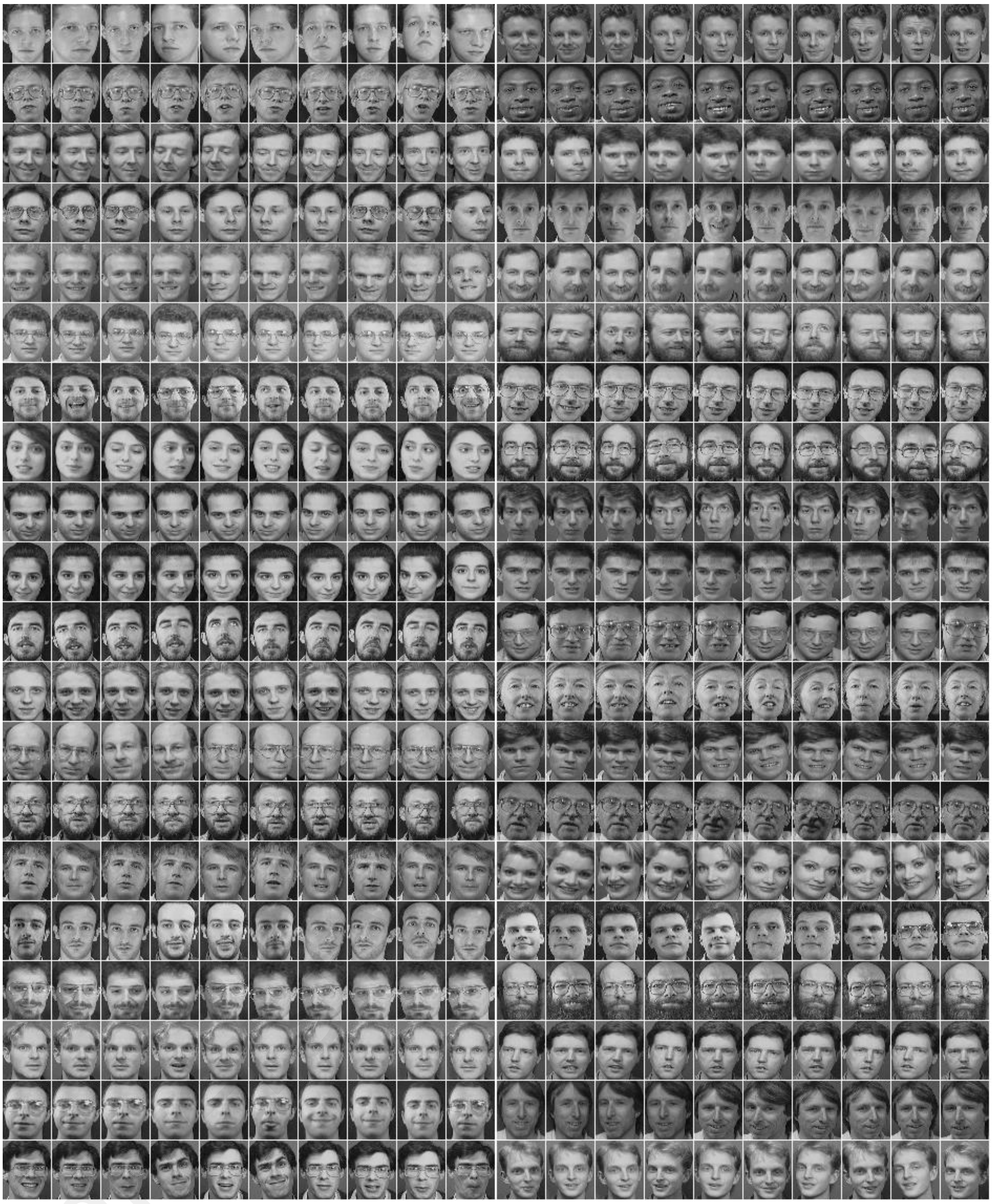


Figure 4.2: 400 images of ORL database

4.4.1 Preprocessing Step

Due to having a grayscale dataset and enhance the contrast of each picture a pre-processing action is required which is utilizing histogram equalization on images. Figure 4.3 is shown the result of this procedure. As it is depicted, the right-side images not only have more contrast than the picture beside it but also it is simpler to extract the features.

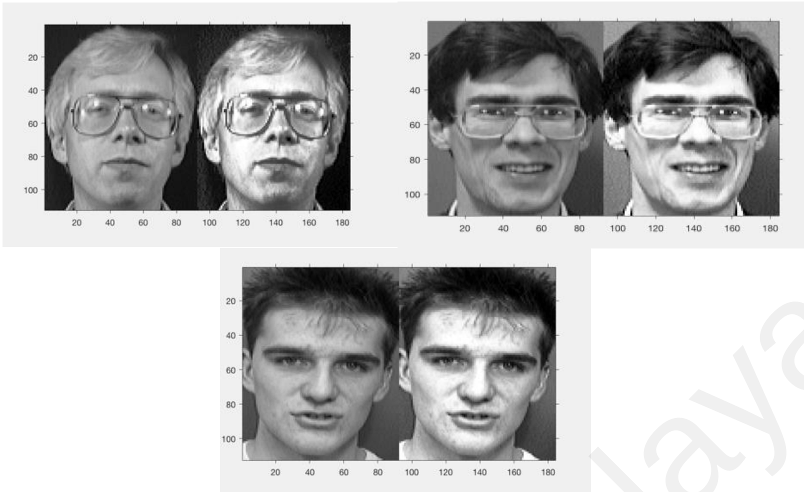


Figure 4.3: Pre-processing step after applying histogram

4.4.2 PZMI

As it was argued before to obtain the features, some pre-processing action is needed for the system. Extra pre-action is also required to implement PZMI which is to transform the digital picture to a binary picture. Consequently, the system can perform a mapping table from the square into a polar coordinator that is represented in Figure 3.4. Also, to implement PZMI with the order of n, there are n+1 feature elements which consist of one feature vector of order n with all m repetitions (where m<n). All moment orders from 1 to 20 were arranged to extract different PZMI features from the facial image. As a sample, eight-moment orders are displayed in Figure 4.4. It is noticeable that in 8 orders 25 features should be extracted from each image.

Zernik moment code which has the role to transform from the square-to-circle model is displayed in Figure 4.5.

	0	1	2	3	4	5	6	7	8
0	M00	_							
1	_	M11							
2	M20	_	M22						
3	_	M31	_	M33					
4	M40	_	M42	_	M44				
5	_	M51	_	M53	_	M55			
6	M60	_	M62	_	M64	_	M66		
7	_	M71	_	M73	_	M75	_	M77	
8	M80	_	M82	_	M84	_	M86	_	M88

Figure 4.4: Figure of 8 order moments

```
function [Z, A, Phi] = Zernikmoment(p,n,m)

N = size(p,1);
x = 1:N; y = x;
[X,Y] = meshgrid(x,y);
R = sqrt((2.*X-N-1).^2+(2.*Y-N-1).^2)/N;
Theta = atan2((N-1-2.*Y+2),(2.*X-N+1-2));
R = (R<=1).*R;
Rad = radialpoly(R,n,m);    % get the radial polynomial

Product = p(x,y).*Rad.*exp(-1i*m*Theta);
Z = sum(Product(:));        % calculate the moments

cnt = nnz(R)+1;             % count the number of pixels inside the unit circle
Z = (n+1)*Z/cnt;            % normalize the amplitude of moments
A = abs(Z);                 % calculate the amplitude of the moment
Phi = angle(Z)*180/pi;      % calculate the phase of the mement (in degrees)
```

Figure 4.5: Zernik moment code

4.4.3 DWT

After applying histogram equalization, three levels of DWT with Haar wavelet function implements to decompose the image to approximation coefficients and details coefficients. Next, the feature vector is built by the approximation coefficients section. According to figure 3.2 after injecting the input images to the system, DWT apply a signal on the images. The output of this action contains four sections which are one approximation band called LL band which is constructed by low frequency, and three specific bands called LH, HL and HH bands which are made by high frequency (Patil et al., 2015). The most significant part of the system is LL band which contains the most valuable information about images. For example, after employing 3 level of decomposition, nine different frequency bands will be generated. As it is shown in Figure 4.6, the size of the feature vector is 14×12=168.

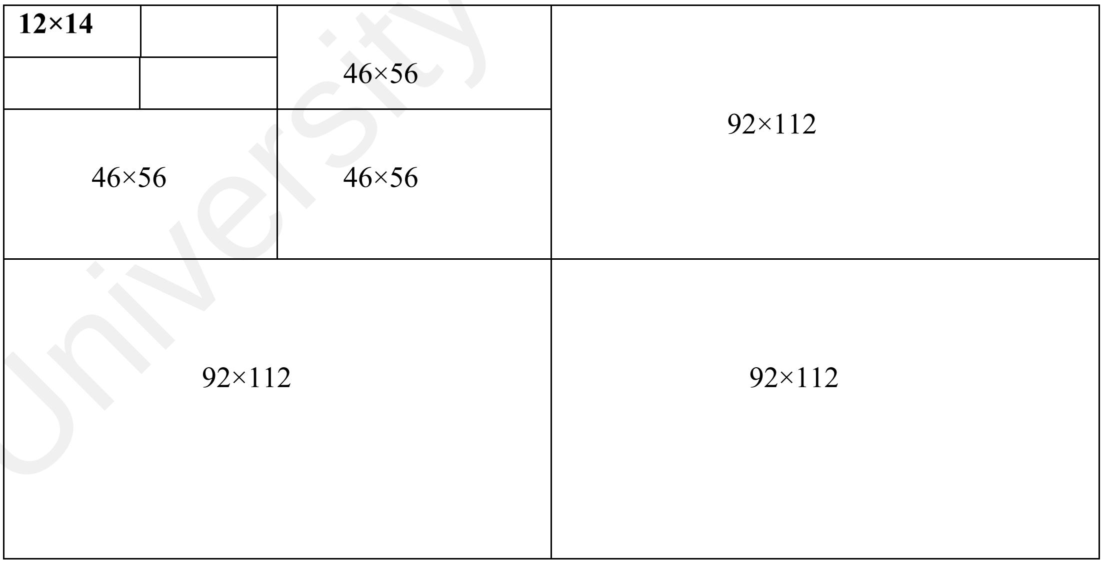


Figure 4.6: Three levels DWT

The size of the original images is 92x112 pixels. After the first level of DWT, some redundant features will be disappeared from the input image. By implementing 3 level of Haar wavelet function as it is shown in Figure 4.7, the input image is small and fitting enough for the next step. The most valuable information for the system is included in this tiny picture.

Figure 4.7: Applying DWT in our dataset

4.5 Establishing a Feature Vector

After implementing both feature extractor algorithms using the ORL dataset, the required feature vector is constructed that can be utilized during the entire process of the implementation and examination. A sample view of our dataset is provided in Figure 4.8.

	21	22	23	24	25	26	27	28	29	30	31	32
1	-9.5305	162.0466	-152.5144	170.2922	-75.9612	-171.4641	-157.8186	170.5115	-54.1379	-91.9780	-83.4219	-130.9511
2	130.2400	-159.1026	-45.3332	33.3047	64.1033	123.8985	-40.9408	109.5032	-55.7515	-9.9079	-28.5186	-179.2417
3	68.1033	-29.0439	18.8228	45.4427	-26.3690	-163.5348	89.2598	163.2142	-167.9907	-79.6585	176.5893	-57.7776
4	95.5182	-24.2318	35.4402	-158.7441	144.5216	-72.1547	-5.2593	166.7039	110.2602	32.7139	-22.5674	-0.2107
5	138.5609	21.9333	-31.1807	-66.8448	43.1104	176.5913	-0.1514	108.2790	85.5892	7.0548	-154.0197	160.2767
6	106.2418	64.8709	-112.9937	130.9501	-56.0037	87.1031	38.3558	-60.7992	138.6600	33.3920	62.7225	-25.6852
7	-117.6974	-59.5200	81.4004	120.6166	8.6637	144.1159	128.9804	161.6144	110.4269	44.5456	166.2684	-6.8314
8	-108.3963	-15.1452	48.1170	-131.1177	178.0567	118.9014	-45.1000	-161.0445	144.0024	-140.1106	22.2485	133.1495
9	-173.7931	-22.4693	-164.5793	19.3311	-168.7901	-27.4416	-43.7420	33.6012	-179.9944	-161.2374	97.8576	-132.0408
10	-82.4727	143.7211	-167.6569	-174.9868	-99.3404	-27.8353	-3.7993	-24.9361	56.9741	155.2760	-95.7386	71.9953
11	-121.1045	107.2864	-14.7864	-82.1875	-108.5117	158.9592	153.6520	-163.8719	67.3131	152.8401	164.5828	-50.7082
12	90.8308	128.2969	-28.2121	-133.5525	-53.5461	153.7519	-179.0255	170.1407	170.0077	168.2926	-73.9076	-145.8472
13	-78.3117	55.2203	-70.1255	-61.6479	-105.7559	164.1136	163.8077	-141.5179	-147.4293	168.6073	165.3905	-17.5734
14	82.5968	125.6305	7.7515	174.7696	-47.4236	86.6173	-117.7360	-170.9373	174.7296	25.3618	-42.1105	-141.4720
15	-74.0013	-14.6061	-0.2380	-57.7595	-120.5538	60.2045	34.4931	20.0662	168.5303	-83.9879	-162.9524	-54.6079
16	111.5900	133.5722	-127.6436	171.3244	-83.4788	-177.5722	-172.4773	85.0458	-169.2700	-152.4630	-106.3001	-172.1998
17	-75.8117	-11.6033	-84.0285	36.8302	-99.4771	173.5353	174.6816	-131.9492	3.8832	149.0180	142.4908	1.9429
18	-59.7893	112.5504	-53.3427	-37.9687	-101.9810	173.9913	172.2467	-160.1435	-93.1439	167.1820	167.9917	-11.4620
19	-77.6475	155.3553	6.8913	163.1957	-113.8150	-7.1472	-106.4396	177.2151	86.0138	39.8315	91.1518	-133.0601
20	-78.8678	-0.7953	-80.5678	-29.7656	99.3782	171.2403	171.5517	-107.3774	114.7972	139.3116	124.2465	25.6274
21	-126.3559	-156.0383	-163.6215	49.8641	-81.1727	-12.1437	-17.1960	51.7733	105.0862	-127.1570	-5.9358	30.6756
22	-59.3974	132.2869	-90.6552	-18.8570	73.3262	-26.7774	71.9454	-12.4215	-125.1353	-156.9411	134.4398	-88.5164
23	-98.8121	109.6646	-3.3600	15.0995	70.9691	-44.8152	163.0798	-171.8012	-150.7535	-164.3705	138.9770	-72.0479

Figure 4.8: Sample extracted feature dataset of PZMI

4.6 Feature Selection based on ACO

As it was discussed in section 3.4.1, an initialization part to implement the ACO is required. Some variable and constant needs to be set to begin the implementation part. These variables are named the number of required ants, pheromone intensity and a maximum number of iteration. Figure 4.9 shows the values set for this study.

```
%% ACO Parameters

MaxIt = 10;          % Maximum Number of Iterations

nPop=50;             % Population Size (Archive Size)

nSample=1;           % Sample Size

q=0.5;               % Intensification Factor (Selection Pressure)

zeta=1;              % Deviation-Distance Ratio

```

Figure 4.9: Initial variables for ACO

The flowchart of ACO already represented in Figure 3.4. To implement the ACO for this study each section of the mentioned flowchart is transformed to MATLAB code. Initially, it is required to evaluate and estimate the subset of the ACO. Then after all M ants finish their tours, the pheromone trails need to be updated. To simulate the computational algorithm to the real life of the ants following steps are done to increase the accuracy. To avoid accumulation of high densities of pheromone as the algorithm progresses, the pheromone is assumed to evaporate over time, pheromone values previously stored on arcs are reduced by a constant factor. The reduction factor is called the pheromone evaporation rate. Some part of the transforming code of the update pheromone is shown in Figure 4.10.

```
%% Step5 Update phromone ....
for j=1:Population_ants
    for l=1:m
        for i=1:num_feature
            deltatrail(i)=0;
            if (s(IX(j),l)==i)
                Errorg = max(SO(1:k))-E(IX(j));
                Errorhg=0;
                for h=1:k
                  if ( (max(SO(1:k))-E(IX(h))) > Errorhg )
                    Errorhg = max(SO(1:k))-E(IX(h));
                    end
                end
                if (Errorhg ~= 0)
                  deltatrail(i)=Errorg/Errorhg;
                else
                  deltatrail(i)=1;
                end;
            end;
            trail(i)= rho * trail(i) + deltatrail(i);
        end;
    end;
end;
```

Figure 4.10: Updating pheromone in ACO

4.7 Classification

In this research, a ten-fold cross validation method is used. This algorithm divides the data into training sets and test sets. In the ten subsamples, nine parts are used as training data and the remaining one sample is retained for testing as the validation data. This method ensures that all of the data has an equal chance to be tested. Therefore, it is assumed that for the enrolment phase, to apply 9 images for training and the 1 image for testing. Therefore, the total amount of testing is 40 images. During training networks, typically, three different subsets are considered to separate the data. The first subset is the training collection, which is practiced for calculating the gradient and refreshing the network measurements and biases. The next subset is the validation set. The error on the validation set is observed throughout the training process. The other one is the test set error which is not used during training, but it is used to compare different patterns. These three subsets are presented in Figure 4.11.

```
for i=1:num_feature  % m  variable ...
    A = [A,Train(:,s(j,i))];
    B = [B,Test(:,s(j,i))];
    C = [C,VAL(:,s(j,i))];
end
```

Figure 4.11: Three subsets of training, validation and test error in our code

As it was recommended for this system, one image set aside for training part. Consequently, the first step of ACO operates on the feature vector which prepared in last part and developing a new training subset. The training part is based on the standard formula for artificial neural network:

net = train (net, input, output);

After updating the pheromone, the system attempts to train one more time to update the training, validation and test error subsets. Finally, in the last part when the training completes, it calculates the accuracy rate based the training part. This rate is achieved by the number of the correctly matched divided by the total number of classified samples. Figure 4.12 shows some part of the code for classification part.

```
y = predict(mdl,B);
y_str=arrayfun(@num2str, y, 'Uniform', false);
Test_class_str=arrayfun(@num2str, Test_class, 'Uniform', false);
index = cellfun(@strcmp,y_str,Test_class_str);
E_best = sum(index)/length(y);
disp('==========Features and accuracy classifiers==========');
disp(s(MinErrorIndex,1:m));
disp(1-E_best);
```

Figure 4.12: Calculating the recognition rate

4.8 Summary of the Experimental Design

This chapter provided parameter specifications of the algorithms used in our experiments and the source code of our implementation in MATLAB. At first, the implementation of the DWT and PZMI as two main feature extractors were explained. Then, the implementation of the ACO was discussed, and finally the classification algorithm reviewed. The details were delivered supplying more information about the proposed methodology and facilitate researchers to reproduce experiments.

CHAPTER 5: RESULTS AND DISCUSSION

5.1 Overview

In this chapter, the results which have obtained from the proposed combination are summarized and will be presented in details. Various tables and charts are supporting the outcome of this study. This chapter is organized as follow: The first part shows the results gained from examining DWT and PZMI for extracting the features from the input images to develop the feature vectors. In the second part, all the experimental results are gathered in some tables and will be discussed. The last part is dedicated to the comparison between some of the existing benchmark results regarding the classification accuracy rate.

5.2 Performance of Feature Extraction Experiment

Figure 5.1 presents the feature extracting duration in milliseconds. As it was shown in the former chapter, both feature extractors had similar training and testing set. The outcome of this experiment shows PZMI is slightly slower than DWT regarding extracting the features and establishing the data vector. This function takes 0.836 seconds for PZMI and 0.747 seconds for DWT; therefore, DWT is around $\frac{0.836}{0.747} \cong 1.11$ times or 11% faster than DWT. Due to having more complexity in the process and doing many iterations, PZMI takes more time than DWT to perform extracting the features from images.

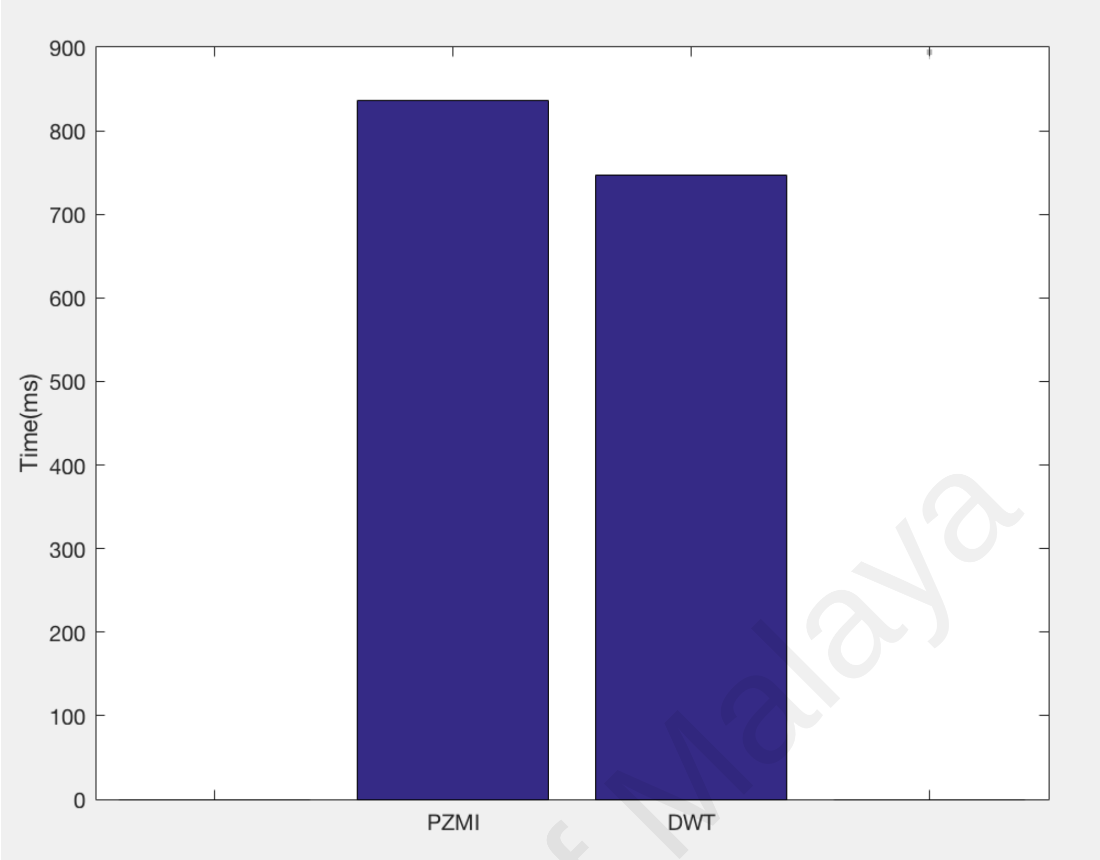


Figure 5.1: A comparison between feature extracting times

5.3 Performance of Feature Selection Experiment

In this part, the performance of the applying ACO on produced feature vectors by PZMI and DWT from the last part is explained. In this study, MSE has been adopted as an estimator to estimate the performance of ACO on the feature vectors which prepared by DWT and PZMI. Mean Square Error is an estimator to measure the differences between two images and as the rate is closer to zero as it is a more beneficial result. Table 5.1 presents MSE rate for both feature extraction methods in connection with ACO. The result captured after applying ACO on PZMI is 0.5 and for DWT is 1.75, and this confirms that the selected features from PZMI vector is more beneficial than DWT and we can predict here that after performing the classification PZMI should produce higher accuracy rate than DWT.

Table 5.1: MSE rate of the feature selection.

Method	MSE
ACO + DWT	1.75
ACO + PZMI	0.5

5.4 Performance of Classification Experiment

It was required to accomplish some actions before the classification part was started. Those were done in the previous steps including the development of datasets and selecting the features. The next step was to examine and to measure the accuracy rate.

Initially, the effect of training set on recognition performance of the system has been reviewed. Figure 5.2 presents the recognition performance for a various number of images which were applied for training. As can be seen, whenever the number of images per individual is increased, the recognition rate is also increased consequently. Furthermore, the increasing number of images for training for the whole population can improve the rate of recognition.

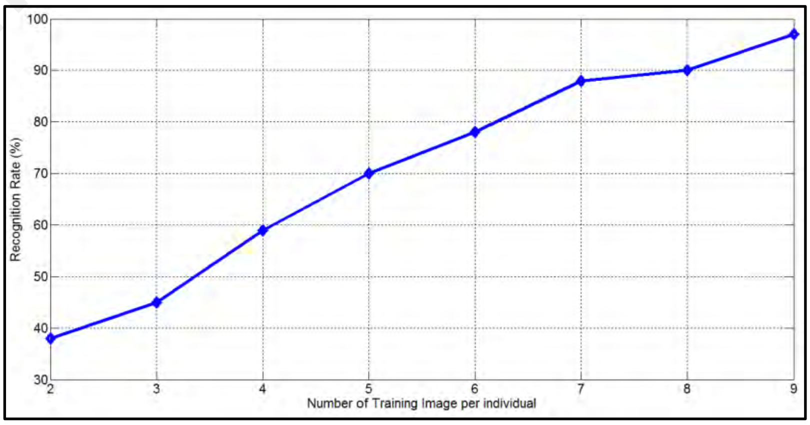


Figure 5.2: Recognition rate (%) for different number of training images per individual.

Table 5.2 shows the recognition rate for the classification part one time for the combination of DWT and ANN and the other time for the combination DWT, Ant Colony and ANN. As it is presented in table 5.2, the highest accuracy rate which system could record for DWT+ANN is 73.47% while this rate after applying ACO increased to 81.90%.

Table 5.2: Classification Accuracy without and with ACO for DWT+ANN

Experiment	Feature Extraction	Feature Selection	Basic Classifier	Execution Time	Classification Accuracy
1	DWT	----	ANN	249	69.03%
2	DWT	----	ANN	278	72.66%
3	DWT	----	ANN	241	70.25%
4	DWT	----	ANN	***269***	***73.47%***
5	DWT	----	ANN	262	71.02%
6	DWT	----	ANN	259	69.11%
7	DWT	----	ANN	253	70.78%
8	DWT	----	ANN	267	72.74%
9	DWT	----	ANN	255	71.68%
10	DWT	----	ANN	249	68.33%
1	DWT	ACO	ANN	364	80.45%
2	DWT	ACO	ANN	369	81.08%

3	DWT	ACO	ANN	***383***	***81.90%***
4	DWT	ACO	ANN	391	81.14%
5	DWT	ACO	ANN	365	80.37%
6	DWT	ACO	ANN	384	79.86%
7	DWT	ACO	ANN	388	80.64%
8	DWT	ACO	ANN	372	79.47%
9	DWT	ACO	ANN	389	79.36%
10	DWT	ACO	ANN	371	80.90%

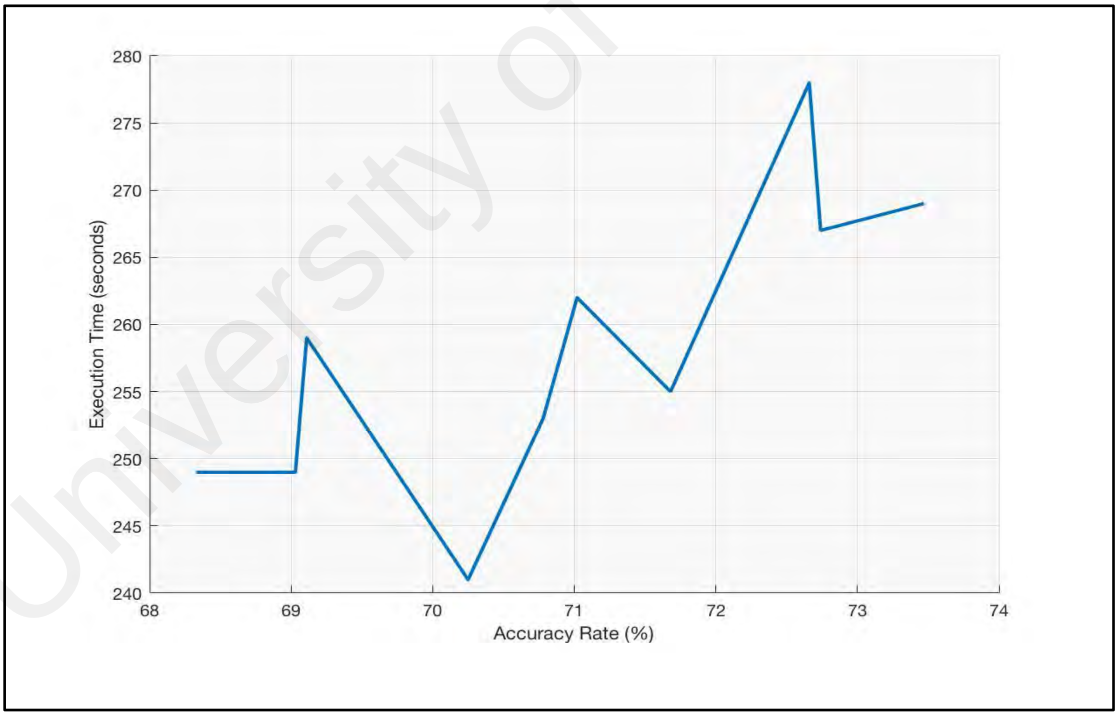


Figure 5.3: Diagram of accuracy rate and execution time for DWT+ANN

As it is shown in Figure 5.3, the diagram has an ascending trend, and this means as accuracy rate increases the execution time also increases accordingly. After employing

ACO execution time has increased drastically. The reason might be having an elaborate implementation for ACO which requires extra actions before selecting the features. As a result, although ACO helps to enhance the accuracy, it raises the total execution time consequently. According to Table 5.2, Figure 5.3 and Figure 5.4 accuracy rate rises significantly from 73.47% to 81.90% and the entire execution time grows from 269 seconds to 383 seconds after practicing ACO on the combination of DWT+ANN.

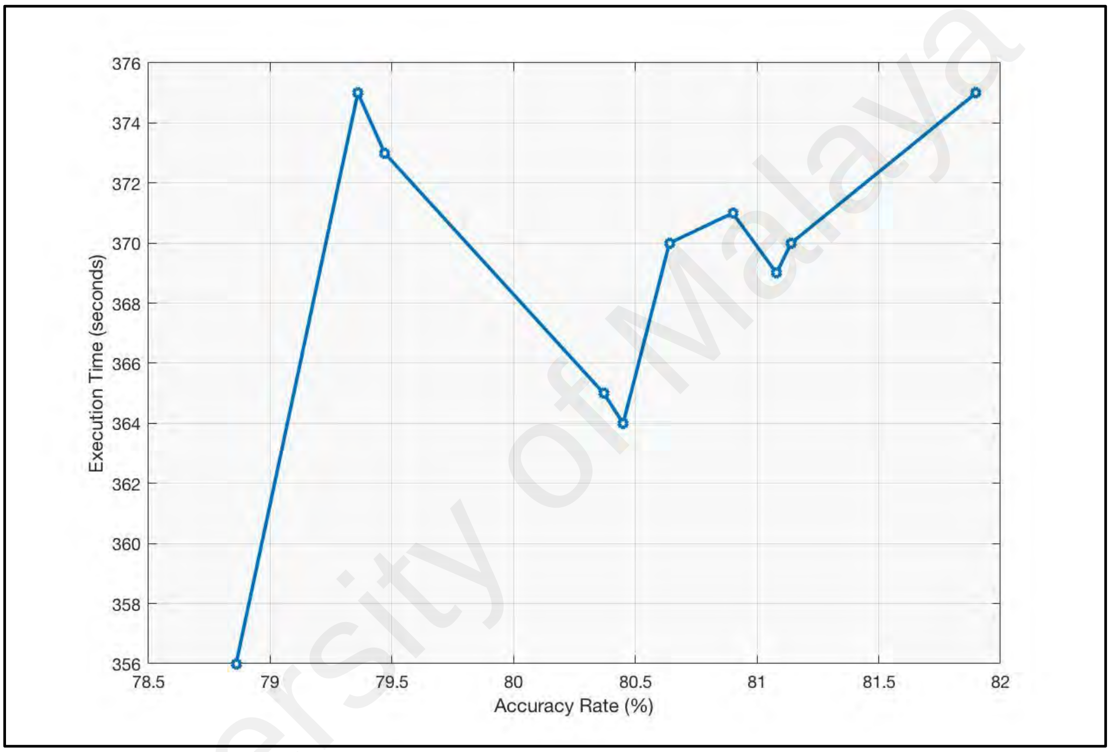


Figure 5.4: Diagram of accuracy rate and execution time for DWT+ACO+ANN

Experimental results for PZMI+ANN and PZMI+ACO+ANN is collected in Table 5.3 and illustrated in Figure 5.5 and Figure 5.6.

Table 5.3: Classification Accuracy without and with ACO for PZMI+ANN

Experiment	Feature Extraction	Feature Selection	Basic Classifier	Execution Time	Classification Accuracy
1	PZMI	----	ANN	263	73.87%
2	PZMI	----	ANN	266	72.66%
3	PZMI	----	ANN	260	74.67%
4	PZMI	----	ANN	271	75.83%
5	PZMI	----	ANN	259	75.17%
6	PZMI	----	ANN	266	74.91%
7	PZMI	----	ANN	261	73.22%
8	PZMI	----	ANN	259	72.74%
9	PZMI	----	ANN	***263***	***76.10%***
10	PZMI	----	ANN	268	75%
1	PZMI	ACO	ANN	385	87.11%
2	PZMI	ACO	ANN	374	86.92%
3	PZMI	ACO	ANN	384	87.84%
4	PZMI	ACO	ANN	386	88.57%
5	PZMI	ACO	ANN	390	88.16%
6	PZMI	ACO	ANN	389	87.88%
7	PZMI	ACO	ANN	388	87.35%

8	PZMI	ACO	ANN	384	87.25%
9	PZMI	ACO	ANN	***<u>392</u>***	***<u>88.73%</u>***
10	PZMI	ACO	ANN	383	87.35%

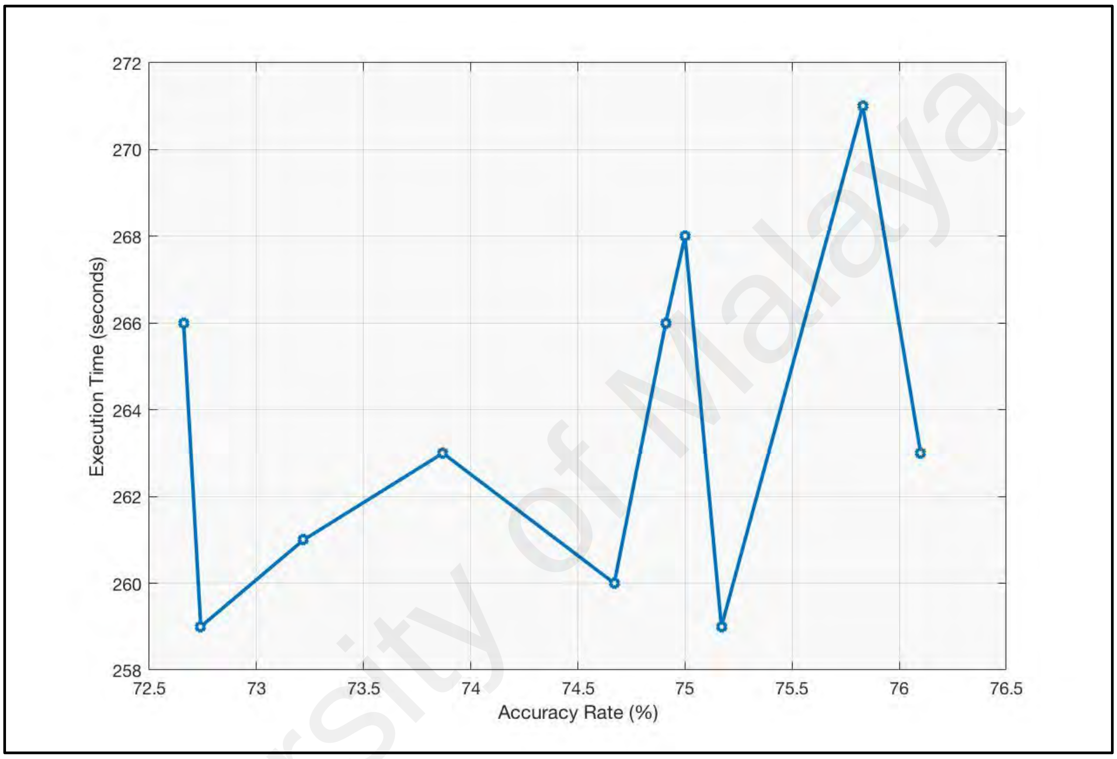


Figure 5.5: Diagram of accuracy rate and execution time for PZMI +ANN

As it is shown in Figure 5.5, the diagram is fluctuating. The minimum execution time is 259 seconds which the system could achieve 75.17% accuracy. Due to the complexity of ACO for selecting the features, the execution time has increased significantly after implementing ACO as it was expected. Figure 5.6 depicts the trend of the accuracy rate with executing time after applying ACO in our system. According to Table 5.3, Figure 5.5 and Figure 5.6 accuracy rate grows from 76.10% to 88.73% and the complete execution time increases from 263 seconds to 392 seconds after utilizing ACO on the combination of PZMI+ANN.

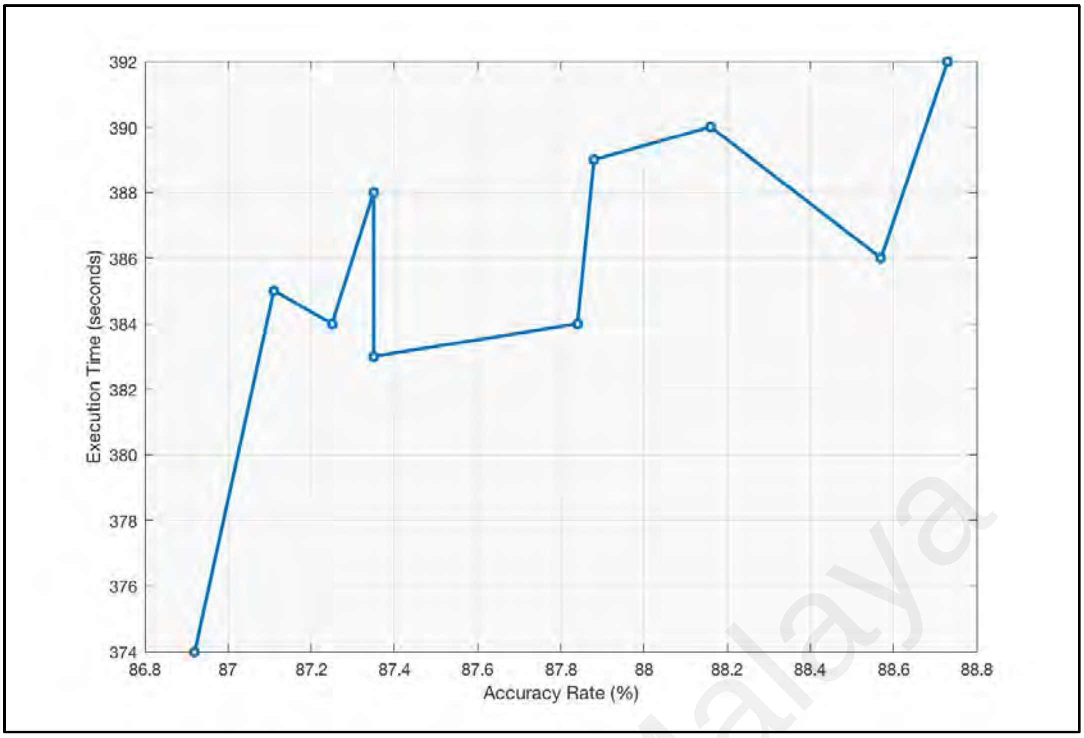


Figure 5.6: Diagram of accuracy rate and execution time for PZMI+ACO+ANN

As it was explained regarding the usage of ACO in the system, although it causes a growth in execution time, it enhances the accuracy rate drastically. Hence, it merits to adopt this algorithm in face recognition system. Table 5.4 and Figure 5.7 shows the summary of this result.

Table 5.4: Accuracy result with ACO for PZMI and DWT

Classification Method	**Repetition**									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
ACO +DWT	80.45%	81.08 %	***81.90*** %	80.37 %	81.14 %	79.86 %	80.64 %	79.47 %	79.36 %	80.90%
ACO +PZMI	87.11%	86.92 %	87.84 %	88.57 %	88.16 %	87.88 %	87.35 %	87.25 %	***88.73*** %	87.35%

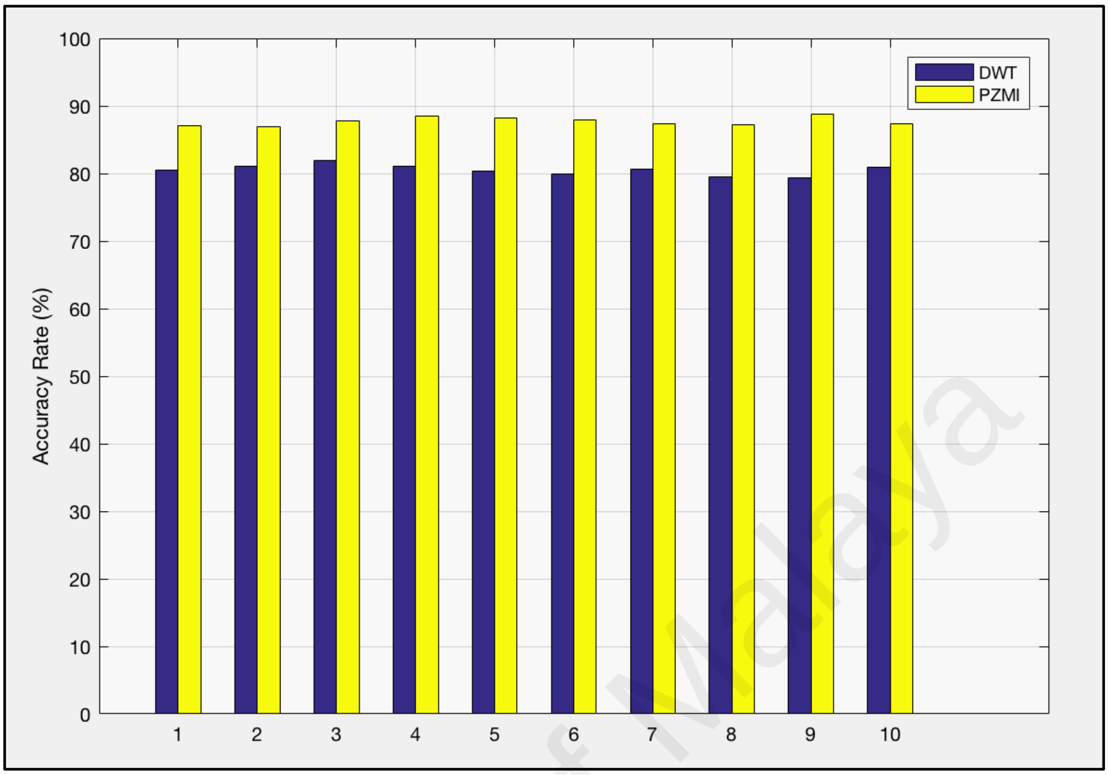


Figure 5.7: Comparison chart for result of classification

According to the result in Table 5.4, ACO+PZMI could extract more valuable features than ACO+DWT to inject to the system. Figure 5.8 displays the error rate of PZMI while moment's progress. It seems that whenever the moments are extended, the recognition error rate is reduced. However, after the order of 10, the growth of the moments could not diminish the recognition error rate severely.

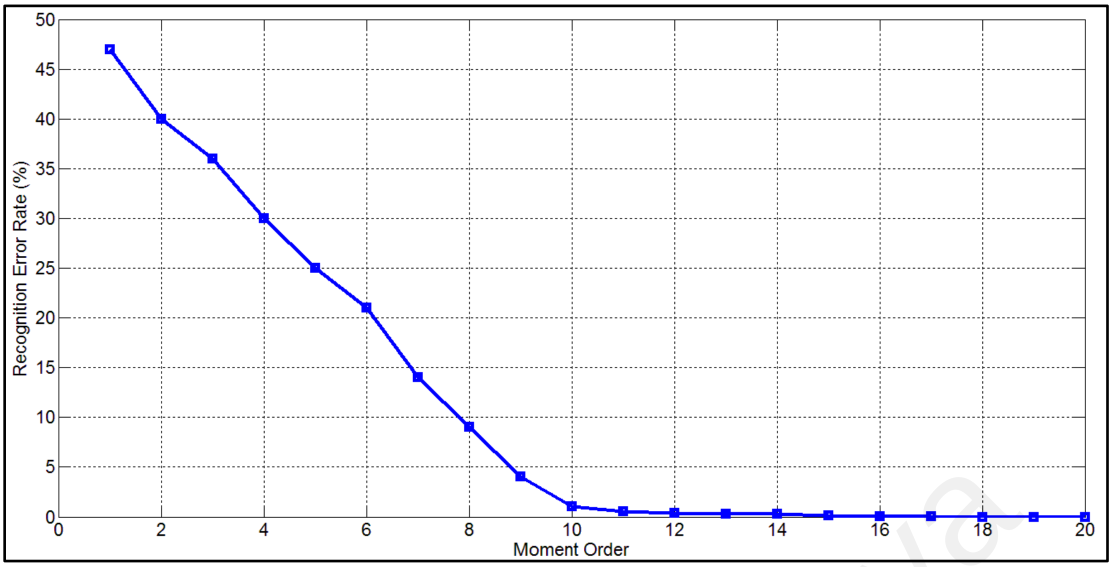


Figure 5.8: The effect of changing moments of PZMI in respect to recognition error rate.

5.5 Discussion

Table 5.5. The final result of the study

Method	Accuracy Rate (%)	Execution Time (seconds)
DWT + ACO + ANN	81.90%	383
PZMI + ACO + ANN	88.73%	392

As it was displayed and explained in figures and tables, this research examined DWT and PZMI as one of the well-known feature extraction algorithms, then ACO was applied on the extracted vectors, and finally the classification part was performed. The outcome of this examination is summarized below:

DWT is approximately 11% faster than PZMI regarding the extraction of the features, moreover; the highest accuracy rate that the system recorded for DWT+ANN is 73.47% in 269 seconds while it increased to 81.90% in 381 seconds after applying ACO. On the other hand, accuracy rate for PZMI grew from 76.10% to 88.73% after ACO was applied but the execution time increased from 263 seconds to 392 seconds. In addition, the sequence of PZMI+ACO+ANN is more reliable than DWT+ACO+ANN regarding the accuracy rate. This rate is $\frac{88.73}{81.90} \cong 1.083$ times or 6.83% higher than DWT. According to Table 5.5, PZMI+ACO+ANN took more time than DWT+ACO+ANN to accomplish the recognition. It took 392 seconds for PZMI+ACO+ANN and 383 seconds for DWT+ACO+ANN to perform this action. Consequently, a comparison among the result of the combination of algorithms proposed in this study and the results of the existing study in this domain have collected in table 5.6.

Table 5.6: A comparison with other benchmarks.

	PZMI+ACO+ANN (88.73%)	DWT+ACO+ANN (81.90%)
PCA+GA+NaiveBayesian (78%)	Improved (10.73%)	Improved (3.9%)
DWT+GA+RBF (84%)	Improved (4.73%)	Deteriorated (-2.10%)
LBP+BSA+NN (85%)	Improved (3.73%)	Deteriorated (-3.10%)
LBP+PCA (82%)	Improved (6.73%)	Equal (-0.10%)
PZMI+GA+RBF (93%)	Deteriorated (-4.27%)	Deteriorated (-11.10%)
DWT+FF+NN (94.375%)	Deteriorated (-5.645%)	Deteriorated (-12.475%)

As it is shown in table 5.6, the proposed combination could obtain higher accuracy rate than a system which employed PCA+GA+ Naive Bayesian by almost 11% for PZMI+ACO+ANN and roughly 4% for DWT+ACO+ANN.

5.6 Summary

Initially, in this chapter, the performance and outcome of the proposed combination of this research have been discussed. Then two feature extractor algorithms which are widely used in this domain examined. After that, the influence of the ACO on the system was measured and finally the result compared with other combinations and benchmarks.

CHAPTER 6: CONCLUSION AND SUGGESTIONS FOR FUTURE RESEARCH

6.1 Overview

This chapter summarizes the significant findings of this research towards the development of a new sequence of algorithms in face recognition systems domain. This chapter also revisits the research objectives and how those objectives are met in this study. Furthermore, this chapter mentions the limitations as well as the possible future research direction.

6.2 Research Problems and Identified Solutions

This thesis principally has concentrated on experimenting a new sequence of algorithms in face recognition system domain. DWT and PZMI as feature extractors, ant colony optimization as feature selector and finally artificial neural network as the classifier are proposed and examined over the ORL face database as a new combination in this domain. In these systems, feature selection usually plays an essential role in promoting the accuracy rate; therefore, it should be efficient since choosing the best facial subset not only increases the recognition performance but also decreases the time surprisingly.

6.3 Research Objectives Revisited

6.3.1 Objective1:

To identify the existing algorithms and analyze the performance of the combination algorithms for feature extraction, feature selection and classification.

The first objective points to investigating in the current algorithms for each piece of a face recognition system. This objective was accomplished by conducting a thorough literature review on some of the most influential algorithms and reviewing advantage and disadvantage of each one distinctly. Besides, in this research PZMI and DWT have been reviewed in details as two of the most remarkable feature extraction algorithms in this domain. Additionally, Ant Colony, Bat, PSO, Firefly and some other algorithms as the main feature selector as well as ANN, SVM, HMM and RBF as classifiers are discussed. A survey was conducted on the various research in which ORL was employed as the dataset. Specific information was extracted and displayed in a table including the accuracy rate of each study, the execution time in some cases, the algorithm used for each section of the system and the evaluation method.

6.3.2 Objective 2:

To develop a face recognition system using the proposed combination of algorithms for improving the performance of face recognition system.

After discovering various algorithms for each specific element of the system, it was determined to select DWT and PZMI as the chief feature extractors, ant colony optimization as the feature selector and ANN as the classifier. These algorithms could all lead to exceptional results in combination with other algorithms or in a different domain. Hence, the set of DWT/PZMI, ant colony optimization, and ANN is proposed to be experimented for the first time in the face recognition domain. The result of the system is measured by accuracy rate evaluation method.

6.3.3 Objective 3:

To evaluate and compare the performance of the developed face recognition system with the existing benchmark.

The outcome of this objective is summarized as follows. At first, DWT is roughly 11% faster than PZMI regarding the feature extracting. In this research, ANN was practiced one time with raw extracted feature and the other time with feature selected by ACO. The highest accuracy rate reported for DWT+ANN was 73.47% in 269 seconds while this rate improved to 81.90% after applying ACO but in 381 seconds. Surprisingly, the accuracy rate for PZMI increased from 76.10% to 88.73% after ACO was employed. However, the execution time rose from 263 seconds to 392 seconds. To conclude, PZMI+ACO+ANN took more time than DWT+ACO+ANN to achieve recognition by 392 seconds for PZMI+ACO+ANN and 383 seconds for DWT+ACO+ANN.

6.4 Research Contribution

This thesis offers the following insights and contributions:

Pseudo Zernike Moment Invariant (PZMI) and Discrete Wavelet Transform (DWT) were implemented for this study. They were also compared concerning time and recognition performance. It was shown that each feature extraction method has a specific advantage as compared to another one.

Ant colony optimization as a feature selector offers a way to reduce the number of selected features of the facial images when applying the feature extraction algorithm.

For this reason, the most discriminated feature subset is selected by ACO method. This feature selection algorithm not only can enhance the recognition performance of the image recognition systems but also it can save time and memory for the process of feature extraction.

The developed face recognition technique was implemented by applying Artificial Neural Networks (ANN). This technique is the most common feature modeling for face recognition systems. The recognition error for this modeling technique is negligible. Furthermore, any new face image can easily be added to the face system model.

A face recognition system was developed by a new combination of feature extraction, ant colony feature selection, and Artificial Neural Network (ANN). Furthermore, the results of each method were fully discussed.

6.5 Research Limitations and Suggestions for Future Research

Through this thesis, some of the issues raised can be further investigated. The gap between the theoretical and real implementation always exists. However, this thesis can provide a new opportunity for researchers to develop and improve face recognition systems further. Researchers should be aware that real conditions can be different and unpredictable compared to lab conditions. Online face recognition systems have limitations which can motivate them to investigate more rigorously in this area.

The developed face recognition system based on DWT and PZMI, ACO and ANN, are implemented as proof-of-concept only. The optimization of the developed face recognition algorithm is possible. Based on this thesis some of the suggestions and recommendations are offered as follows:

The weighted function curve in the ant colony feature selection needs to be improved. For example, the facial content affects the final performance which is estimated by face recognition system. Facial feature information is also distributed in some temporary structures which have not been done yet such as temporal-frequency transition and duration.

A perfect feature subset is needed which is often a big challenge problem in face recognition systems. Therefore, a novel and robust feature selection can contribute toward face recognition applications. The robustness of the developed face recognition systems needs to be analyzed in other feature modeling types such as HMM, GMM and VQ. This issue can provide opportunities for researchers to integrate it with various applications.

The real-time implementation with less computation and low costs is needed for the system in the real environment. A combination of face detection and face recognition can improve the performance.

REFERENCES

Abhishek, K. (1992). Stock Prediction using Artificial Neural Networks. *Science Journal*, *8*(2), 1–14.

Afroge, S., Mamun, A., & Mat, A. (2015). PCA and Back-P ropagation Neural Netw work Based Face e Recognition System, 21–23.

Agarwal, V., & Bhanot, S. (2015). Firefly inspired feature selection for face recognition. In *2015 Eighth International Conference on Contemporary Computing (IC3)* (pp. 257–262). IEEE. https://doi.org/10.1109/IC3.2015.7346689

Almohammad, M. S., & Mahmoud, T. A. (2013). Human Identification System based on Face using Active Horizontal Levels (AHLs) Feature. *International Journal of Computer Applications (0975*, *61*(20), 27–32.

Bagherian, E., & Rahmat, R. W. O. K. (2008). Facial feature extraction for face recognition: a review. In *2008 International Symposium on Information Technology* (Vol. 2, pp. 1–9). IEEE. https://doi.org/10.1109/ITSIM.2008.4631649

Bamieh, B., & De Figueiredo, R. (1986). A general moment-invariants/attributed-graph method for three-dimensional object recognition from a single image. *Robotics and Automation, IEEE Journal of*, *2*(1), 31–41. https://doi.org/10.1109/JRA.1986.1087034

Biodiversity, K., Shannon, T., & Shannon, T. (2010). Measuring biodiversity of ecological communities.

Black Jr, J. A., Gargesha, M., Kahol, K., Kuchi, P., & Panchanathan, S. (2002). Framework for performance evaluation of face recognition algorithms. In *ITCom 2002: The Convergence of Information Technologies and Communications* (pp. 163–174). International Society for Optics and Photonics.

Bojkovic, Z., & Samcovic, A. (2006). Face detection approach in neural network based method for video surveillance. In *Neural Network Applications in Electrical Engineering, 2006. NEUREL 2006. 8th Seminar on* (pp. 44–47). IEEE.

Chen, B. J., Shu, H. Z., Chen, G., & Ge, J. (2014). Color face recognition based on quaternion Zernike moment invariants and quaternion BP neural network. In *Applied Mechanics and Materials* (Vol. 446, pp. 1034–1039). Trans Tech Publ.

Cheng, C., Li, H.-F., & Bao, C.-H. (2016). Hybrid Artificial Fish Algorithm to Solve TSP Problem. In *Proceedings of the 6th International Asia Conference on Industrial Engineering and Management Innovation* (pp. 275–285). Springer.

Cheng, G., Shi, C., Zhu, K., & Gong, K. (2011). The Application of Binary Particle Swarm Algorithm in Face Recognition. *2011 Seventh International Conference on Computational Intelligence and Security*, (1), 1229–1233. https://doi.org/10.1109/CIS.2011.272

Couceiro, M., & Ghamisi, P. (2016). Fractional-order Darwinian PSO. In *Fractional Order Darwinian Particle Swarm Optimization* (pp. 11–20). Springer.

Darestani, M. R. Y., Sheikhan, M., & Khademi, M. (2013). Face recognition using contourlet-based features and hybrid PSO-neural model. In *The 5th Conference on*

Information and Knowledge Technology (pp. 1818–186). IEEE. https://doi.org/10.1109/IKT.2013.6620061

Dash, M., & Liu, H. (1997). Feature selection for classification. *Intelligent Data Analysis*, *1*(3), 131–156. https://doi.org/10.3233/IDA-1997-1302

Dond, X., Sun, G., & Xu, G. (n.d.). Discrete Wavelet Transform Based Feature Extraction for Tissue Classification Using Gene Expression Data, (1).

Engelbrecht, A. P. (2005). Fundamentals of Computational Swarm Intelligence. *Dana*. https://doi.org/978-0470091913

Eskandari, M., & Toygar, Ö. (2014). Fusion of face and iris biometrics using local and global feature extraction methods. *Signal, Image and Video Processing*, *8*(6), 995–1006. https://doi.org/10.1007/s11760-012-0411-4

Faez, K. (2005). PZMI AND WAVELET TRANSFORM FEATURES IN FACE RECOGNITION, 2690–2694.

Faez, K., & Farajzadeh, N. (2007). A performance comparison of the ZM, PZM and LM in the face recognition system in presence of salt-pepper noise. *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, *5*, 4197–4201. https://doi.org/10.1109/ICSMC.2006.384793

Foruzan, E., Scott, S. D., & Lin, J. (2015). A comparative study of different machine learning methods for electricity prices forecasting of an electricity market. *North American Power Symposium (NAPS), 2015*, (October), 1–6. https://doi.org/10.1109/NAPS.2015.7335095

Fouad, M. M., Zawbaa, H. M., Gaber, T., Snasel, V., & Hassanien, A. E. (2016). *A fish detection approach based on BAT algorithm*. *Advances in Intelligent Systems and Computing*. https://doi.org/10.1007/978-3-319-26690-9_25

Gonzalez, R. C., & Woods, R. E. (2002). *Digital Image Processing, 2nd Edition*. *Leonardo*. https://doi.org/10.2307/1574313

Haddadnia, J., Ahmadi, M., & Raahemfa, K. (2003). an Effective Feature Extraction Method for Face Recognition, 917–920.

Ho, H. T., & Chellappa, R. (2013). Pose-Invariant Face Recognition Using Markov Random Fields, *22*(4), 1573–1584.

Jalled, F. (2017). Face Recognition Machine Vision System Using Eigenfaces. *International Journal*, *2*(2), 1–3. Retrieved from http://arxiv.org/abs/1705.02782

Jana, D., & Sinha, K. (2014). Wavelet Thresholding for Image Noise Removal. *International Journal on Recent and Innovation Trends in Computing and Communications*, *2*(6), 1400–1405.

Kanan, H. R., & Faez, K. (2005a). PZMI and wavelet transform features in face recognition system using a new localization method. In *Industrial Electronics Society, 2005. IECON 2005. 31st Annual Conference of IEEE* (p. 5–pp). IEEE.

Kanan, H. R., & Faez, K. (2005b). ZMI and wavelet transform features and SVM classifier in the optimized face recognition system. *Proceedings of the Fifth IEEE International Symposium on Signal Processing and Information Technology*, *2005*, 295–300. https://doi.org/10.1109/ISSPIT.2005.1577112

Kanan, H. R., Faez, K., & Hosseinzadeh, M. (2007). Face Recognition System Using Ant Colony Optimization-Based Selected Features. In *2007 IEEE Symposium on Computational Intelligence in Security and Defense Applications* (pp. 57–62). IEEE. https://doi.org/10.1109/CISDA.2007.368135

Kashem, M. A., Akhter, N., Ahmed, S., Alam, M., & Ntroduction, I. (2011). Face Recognition System Based on Principal Component Analysis (PCA) with Back Propagation Neural Networks (BPNN), *2*(6), 1–10.

Kaur, A., & Kaur, J. (2012). A survey on comparison of DCT and DWT of Image Compression Techniques. *International Journal of Engineering Research and Development ISSN*, *1*(4), 49–52.

Kaur, H., Panchal, V. K., & Kumar, R. (2013). A novel approach based on nature inspired intelligence for face feature extraction and recognition. *2013 Sixth International Conference on Contemporary Computing (IC3)*, 149–153. https://doi.org/10.1109/IC3.2013.6612179

Kaur, S., Agarwal, P., & Rana, R. S. (2011). Ant Colony Optimization : A Technique used for Image Processing, *4333*, 173–175.

Khadhraoui, T., Ktata, S., Benzarti, F., & Amiri, H. (2016). Features Selection Based on Modified PSO Algorithm for 2D Face Recognition. *Proceedings - Computer Graphics, Imaging and Visualization: New Techniques and Trends, CGiV 2016*, 99–104. https://doi.org/10.1109/CGiV.2016.28

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.

Latha, P., Ganesan, L., & Annadurai, S. (2009). Face recognition using neural networks. *Signal Processing: An International …*, (3), 153–160. https://doi.org/10.1109/IJCNN.2015.7280401

Lin, C.-J. (2001). On the convergence of the decomposition method for support vector machines. *IEEE Transactions on Neural Networks*, *12*(6), 1288–1298.

Lin, S.-H., Kung, S.-Y., & Lin, L.-J. (1997). Face recognition/detection by probabilistic decision-based neural network. *IEEE Transactions on Neural Networks*, *8*(1), 114–132.

Loubière, P., Jourdan, A., Siarry, P., & Chelouah, R. (2016). A sensitivity analysis method for driving the Artificial Bee Colony algorithm’s search process. *Applied Soft Computing*. https://doi.org/10.1016/j.asoc.2015.12.044

Lv, X., Wu, J., & Liu, W. (2014). Face image feature selection based on gabor feature and recursive feature elimination. *Proceedings - 2014 6th International Conference on Intelligent Human-Machine Systems and Cybernetics, IHMSC 2014*, *2*, 266–269. https://doi.org/10.1109/IHMSC.2014.166

Masupha, L., Zuva, T., Ngwira, S., & Esan, O. (n.d.). Face Recognition Techniques , their Advantages , Disadvantages and Performance Evaluation, (i).

Matsugu, M., Mori, K., Mitari, Y., & Kaneda, Y. (2003). Subject independent facial expression recognition with robust face detection using a convolutional neural network. *Neural Networks*, *16*(5–6), 555–559. https://doi.org/10.1016/S0893-6080(03)00115-1

Ming-Kuei Hu. (1962). Visual pattern recognition by moment invariants. *IEEE Transactions on Information Theory*, *8*(2), 179–187. https://doi.org/10.1109/TIT.1962.1057692

Nabatchian, A., Abdel-Raheem, E., & Ahmadi, M. (2008). Human Face Recognition Using Different Moment Invariants: A Comparative Study. *2008 Congress on Image and Signal Processing*, 661–666. https://doi.org/10.1109/CISP.2008.479

Nadhir, M., Wahab, A., Nefti-meziani, S., & Atyabi, A. (2015). A Comprehensive Review of Swarm Optimization Algorithms, 1–36. https://doi.org/10.1371/journal.pone.0122827

Omer, A. E., & Khurran, A. (2015). Facial recognition using principal component analysis based dimensionality reduction. *2015 International Conference on Computing, Control, Networking, Electronics and Embedded Systems Engineering (ICCNEEE)*, 434–439. https://doi.org/10.1109/ICCNEEE.2015.7381408

Ouarda, W., Trichili, H., Alimi, A. M., & Solaiman, B. (2013). Combined local features selection for face recognition based on Naïve Bayesian classification. *Hybrid Intelligent Systems (HIS), 2013 13th International Conference on*, 240–245. https://doi.org/10.1109/HIS.2013.6920489

Pal, M., & Mather, P. M. (2005). Support vector machines for classification in remote sensing. *International Journal of Remote Sensing*, *26*(5), 1007–1011. https://doi.org/10.1080/01431160512331314083

Patil, J. P., Nayak, C., & Jain, M. (2015). Palmprint recognition using DWT, DCT and PCA techniques. In *2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)* (pp. 1–5). IEEE. https://doi.org/10.1109/ICCIC.2015.7435677

Pentland, A., Moghaddam, B., & Starner, T. (1994). View-based and modular eigenspaces for face recognition. In *CVPR* (Vol. 94, pp. 84–91).

Phillips, P. J. (1998). Support vector machines applied to face recognition.

Rahman, A., Rahman, M. N. A., Safar, S., & Kamruddin, N. (2013). Human Face Recognition: An Eigenfaces Approach. In *International Conference on Advances in Intelligent Systems in Bioinformatics. Atlantis Press*.

Raja, V. S., & Rajagopalan, S. P. (2014). A COMPARATIVE ANALYSIS OF OPTIMIZATION TECHNIQUES FOR ARTIFICIAL NEURAL NETWORK IN BIO MEDICAL APPLICATIONS, *10*(1), 106–114. https://doi.org/10.3844/jcssp.2014.106.114

Rao, P. D., & Rai, S. (2016). Ant Colony Optimization Algorithm for Improving Efficiency of Canny Edge Detection Technique for Images, *2*(6), 350–355.

Reddi, S. S. (1981). Radial and Angular Moment Invariants for Image Identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-3*(2), 240–242. https://doi.org/10.1109/TPAMI.1981.4767087

Rowley, H. A., Baluja, S., & Kanade, T. (1998). Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(1), 23–38. https://doi.org/10.1109/34.655647

Roychowdhury, S., & Emmons, M. (1991). A S URVEY OF THE T RENDS IN F ACIAL AND E XPRESSION R ECOGNITION D ATABASES AND METHODS.

S, S. S., B, K. A., & B, P. K. (2012). Face Recognition Using Principal Component Analysis and Linear Discriminant Analysis on Holistic Approach in Facial Images Database, *2*(12), 15–23.

Sen, T., & Mathur, H. D. (2016). A new approach to solve Economic Dispatch problem using a Hybrid ACO–ABC–HS optimization algorithm. *International Journal of Electrical Power & Energy Systems*, *78*, 735–744. https://doi.org/10.1016/j.ijepes.2015.11.121

Sihag, R. (2011). Wavelet Thresholding for Image De-noising. *International Conference on VLSI, Communication & Instrumentation(ICVI)2011*, (Icvci), 20–24.

Sirovich, L., & Kirby, M. (1987). Low-dimensional procedure for the characterization of human faces. *Josa a*, *4*(3), 519–524.

Srinivasan, M. (2015). SV-M/D: Support Vector Machine-Singular Value Decomposition Based Face Recognition. In *Advances in Intelligent Informatics* (pp. 61–76). Springer.

Stonham, T. J. (1986). Practical face recognition and verification with WISARD. In *Aspects of face processing* (pp. 426–441). Springer.

Utsumi, Y., Matsumoto, Y., & Iwai, Y. (2009). An efficient branch and bound method for face recognition. In *Signal and Image Processing Applications (ICSIPA), 2009 IEEE International Conference on* (pp. 156–161). IEEE.

von der Malsburg, C. (2014). A vision architecture. *arXiv Preprint arXiv:1407.1642*.

Wallin, Å., & Kübler, O. (1995). Complete sets of complex Zernike moment invariants and the role of the pseudoinvariants. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *17*(11), 1106–1110. https://doi.org/10.1109/34.473239

Weng, J. J., Ahuja, N., & Huang, T. S. (1993). Learning recognition and segmentation of 3-d objects from 2-d images. In *Computer Vision, 1993. Proceedings., Fourth International Conference on* (pp. 121–128). IEEE.

Wong, K.-W., Lam, K.-M., & Siu, W.-C. (2001). An efficient algorithm for human face detection and facial feature extraction under different conditions. *Pattern Recognition*, *34*(10), 1993–2004. https://doi.org/10.1016/S0031-3203(00)00134-5

Yang, X.-S. (2010). A new metaheuristic bat-inspired algorithm. *Nature Inspired Cooperative Strategies for Optimization (NICSO 2010)*, 65–74.

Zhao, L., & Yang, Y.-H. (1999). Theoretical analysis of illumination in PCA-based vision systems. *Pattern Recognition*, *32*(4), 547–564.

Zhu, M. (2001). Feature extraction and dimension reduction with applications to classification and the analysis of co-occurrence data, (June), 1–137. Retrieved from http://search.proquest.com.ezproxy.cul.columbia.edu/docview/304728120/abstract/B59F30759A1641ABPQ/1?accountid=10226