DECEMBER, 2016

M.Sc. IN ELECTRONICS AND COMPUTER ENGINEERING

SHERKO BAPER SAEED

HASAN KALYONCU UNIVERSITY
GRADUATE SCHOOL OF
NATURAL & APPLIED SCIENCES

FACE RECOGNITION SYSTEM BASED ON PCA-WAVELET AND SUPPORT VECTOR MACHINES

M. Sc. THESIS
IN
ELECTRONICS AND COMPUTER ENGINEERING

BY

SHERKO BAPER SAEED

DECEMBER 2016

FACE RECOGNITION SYSTEM BASED ON PCA-WAVELET AND SUPPORT VECTOR MACHINES

M.Sc. Thesis

in

Electronics and Computer Engineering

Hasan Kalyoncu University

Supervisor

Prof. Dr. Celal Korasli

Co-Supervisor

Dr. Laith Rezouki Fliah

By

Sherko Baper Saeed

DECEMBER 2016

REPUBLIC OF TURKEY
HASAN KALYONCU UNIVERSITY
GRADUATE SCHOOL OF NATURAL & APPLIED SCIENCES
COMPUTER ENGINEERING DEPARTMENT

Name of the thesis: FACE RECOGNITION SYSTEM BASED ON PCA-WAVELET AND SUPPORT VECTOR MACHINES
Name of the student: Sherko Baper Saeed
Exam date: December 21, 2016

Approval of the Graduate School of Natural and Applied Sciences

Prof. Dr. Mehmet KARPUZCU
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Prof. Dr. Celal KORASLİ
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Prof. Dr. Celal KORASLİ
Supervisor

Examining Committee Members	Signature
Prof. Dr. Celal Korasli	.....
Asst.Prof. Dr. Laith Rezouki Fliah	.....
Asst.Prof. Dr. Abdulhafez ABDUL HAFEZ	.....
Asst.Prof. Dr. M. Fatih HASOGLU	.....
Assoc.Prof. Dr. Ahmet ALKAN	.....

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Sherko Baper Saeed

Signature

ABSTRACT

Face recognition can represent a key requirement in various types of applications such as human-computer interface, monitoring systems as well as personals identification etc.

In this thesis, different types of methods were used for implementing and testing face recognition system. These methods are first way introduced in one face recognition system and the combination of these methods together support the recognition system and give better results compared to other methods.

In the first part, a combination of PCA and Wavelet Feature Extraction methods are used to obtain the important features and reduce the dimensions of the face image. In the second part, SVM classifier is used to classify the features of image and K-Nearest Neighborhood is applied to identify it. In the third part, Classification performance under different kernel types is examined to demonstrate the classification performance of SVM, in addition, to compare the SVM classifier result with artificial neural network classifier.

Finally, the performances of the system under various conditions are tested. For a more comprehensive comparison, two face image databases are used to test the performance of the system. The experimental results proved the efficiency and reliability of the system and the results enhancement of 5% by using the SVM classifier with polynomial Kernel Function compared to use feed forward Backpropagation neural network classifier.

Keywords: Face recognition, Support vector machines, Feed forward backpropagation neural network and K-nearest neighborhood.

ÖZET

Yüz tanıma, insan-bilgisayar arabirimi, izleme sistemleri ve kişisel tanımlama gibi çeşitli uygulamalarda önemli bir gereksinimi temsil edebilir.

Bu tezde, yüz tanıma sistemini uygulamak ve test etmek için farklı yöntem türleri kullanılmıştır. Bu yöntemler ilk önce bir yüz tanıma sisteminde tanıtılır ve bu yöntemlerin kombinasyonu tanıma sistemini destekler ve diğer yöntemlere kıyasla daha iyi sonuç verir.

Birinci bölümde önemli özellikleri elde etmek ve yüz imgesinin boyutlarını azaltmak için PCA ve Wavelet özellik çıkarımı yöntemlerinin bir kombinasyonu kullanılır. İkinci bölümde, imge özelliklerini sınıflandırmak için SVM sınıflandırıcısı kullanılır ve K-en yakın mahalle uygulanarak tanımlanır. Üçüncü bölümde SVM sınıflandırıcı sonuçlarını yapay sinir ağı sınıflandırıcısı ile karşılaştırmak için SVM'nin sınıflandırma performansını göstermek için farklı çekirdek türleri altındaki Sınıflandırma performansı incelenmiştir.

Sonunda, sistemin çeşitli koşullardaki performansları test edilir. Daha kapsamlı bir karşılaştırma için, sistemin performansını test etmek için iki yüz görsel veritabanı kullanılır. Deney sonuçları, sistemin etkinliği ve güvenilirliği üzerinde kanıtlanmış ve SVM sınıflandırıcısını, polinom çekirdeği işleviyle birlikte, ileriye yayılma Yaylanma yayılımı sinir ağı sınıflandırıcıya kıyasla% 5 oranında sonuç geliştirme ile karşılaştırmıştır.

.

Anahtar Kelimeler: Yüz tanıma, Destek vektör makineleri, İleriye dönük ilerleme propagasyon nöral ağ ve K-en yakın komşuluk.

To My Brothers and Sisters

To My Wife

To My Son

ACKNOWLEDGEMENT

First and foremost, I would like to thank the Allah who helped me to finish this work and without his help and mercy, this work is not completed.

Secondly, I would like to sincerely thank my supervisor (**Prof. Dr. Celal Korasli**) for his advice and help him to me, during my work in this thesis.

Special thanks to my university **Hasan Kalyoncu** for accepting M.Sc. program.

I would like to say thanks to (**Dr. Laith R. Flieh Al-Jumaily**) who provide everything to help me to finish my thesis in the limited time.

My special thanks to my brothers for support and encouraging me to continue my research.

Finally, I would like to thank my son (**Lehat**) and my wife for all your love, understanding, support and encouragement.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SYMBOLS/ABBREVIATIONS

ANNs	Artificial Neural Networks
DR	Dimensionality Reduction
GA	Genetic Algorithm
LDA	Linear Decremented analysis
MLP	Multilayer Perceptron
PCA	Principal Component Analysis
RBF	Radial Base Kernel Function
SVM	Support Vector Machine
BPNN	Back-Propagation Neural Networks
FFNN	Feed Forward Neural Networks
RBNN	Radial basis neural network
LFD	local frequency descriptor
DCT	Discrete cosine transform
LPQ	local phase quantization
CWT	Continuous Wavelet Transform
DWT	Discrete Wavelet Transform

CHAPTER 1
INTRODUCTION

1.1 General

In recent years, biometric technologies have become very thriving field because of the huge development in computing power. There are a number of factors such as reduced both of cost and size, increased accuracy, and ease of use. All these factors have worked together to make biometrics an increasingly feasible solution for securing access to information. This biometric technologies can be used for many fields as security, to reduce fraud and even can be used as an empowering technology. There are many the biometric applications used in the different field. Figure (1-1) shows a number of biometric systems.

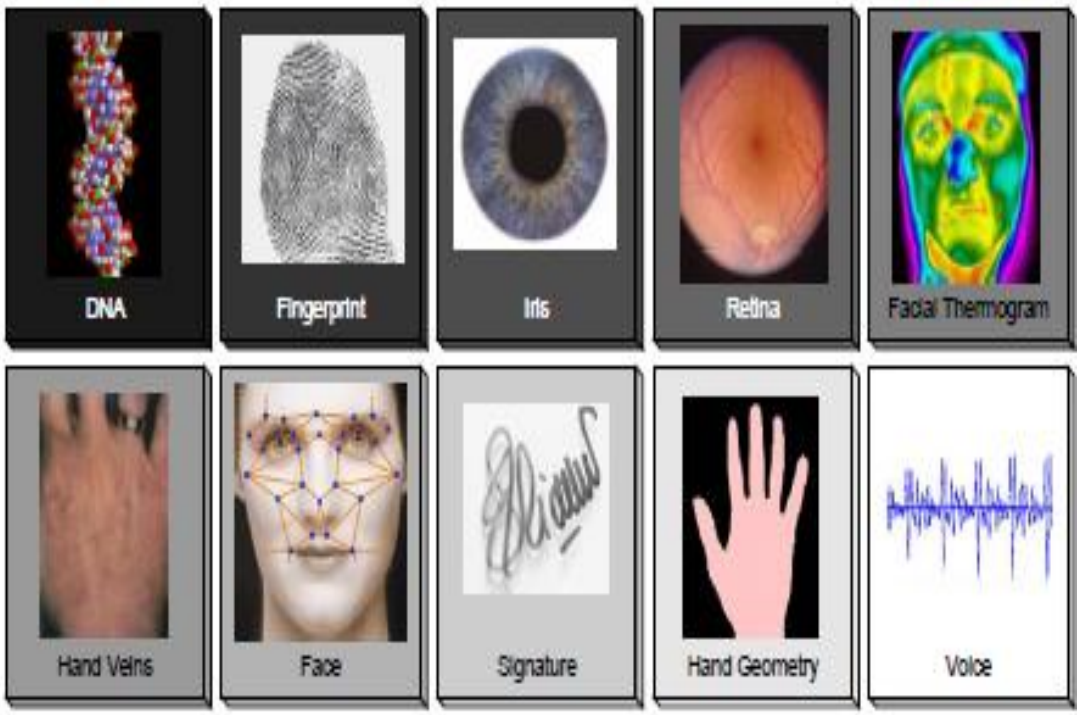


Figure (1-1): Different types of biometric system [1].

Facial recognition is considered as one of the biometric systems and this field as a topic of research have increased in recent years. Face recognition system uses biometric information which takes from human but the use of this system is easier

than in another biometric system such as fingerprints iris and signature, etc., especially with the uncooperative people. Face recognition systems are usually used in security systems to monitor people and to discover any security breach as an example face recognition system is applied in airports for monitoring human behaviour, therefore, face recognition system is used in many security application such as avoiding crime, monitoring through video and also for verification of the identity of persons [2].

This type of system is considered as one of the intricate problems in the field of image processing and this is related for many of reasons such as the effect of illumination on the image and many of other reasons that can affect the quality of the image. Face recognition system uses two techniques and these techniques are face detection and recognition. Face detection is applied to determine the position of the faces in the image and disregard any another object. The algorithm of recognition that is applied to recognize the face in any images depends on the structure properties and these properties are widely applied in these types of applications [2].

Any biometric system has many of powerful and impairment points and the choice of the system usually, depends on the type of application that used and therefore it is not expected that a single system can meet all needs effectively.

1.2 Advantage of Biometric Systems

There are a number of advantages of biometrics systems [2] compared to other traditional systems. These advantages are listed below:

a) Increased the security.
b) Increased the convenience.
c) Increased the accountability.

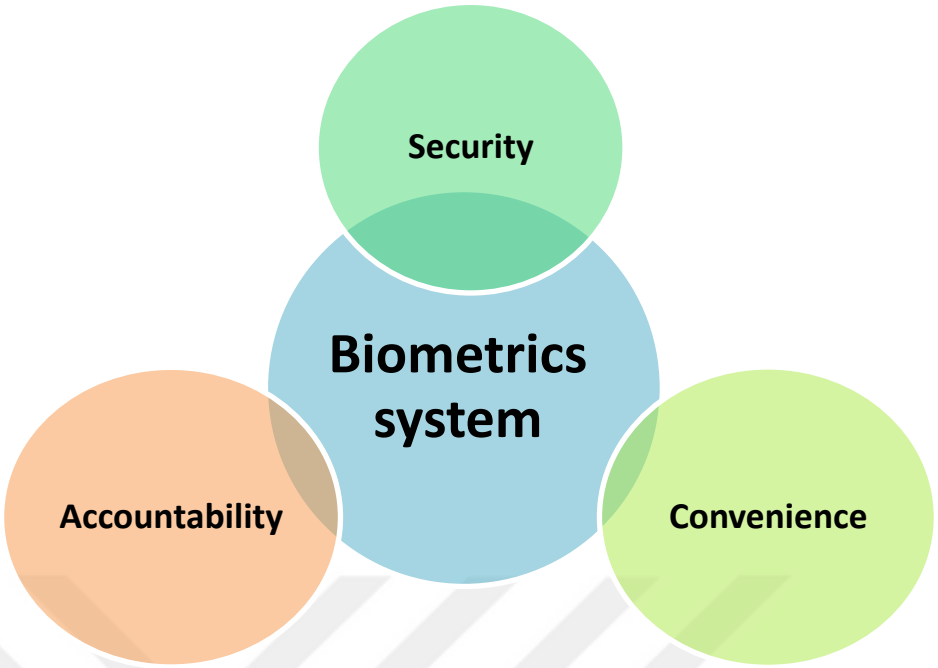


Figure (1-2): Benefits of Biometrics System [2].

a) Increase the security: Biometric systems can provide greater security rate compared to traditional authentication systems, and this means that the resources are available only to authentication users and they are kept protected from unauthorized users. Passwords and PIN numbers can be guessed or can be stolen. Many users using password can break it so that an unauthorized user may be able to break the account with little effort. From another point, it is possible stolen or estimate a password or PIN code, In contrast of biometric data, it difficult to anyone to steal this data or estimate it some conditions, some of biometric systems can be infringed its security or attacked it, With development of technology, these types of systems are becoming very robust and it difficult to anyone to breached it because it is impossible to find two people have the same face, fingerprint, or same voice.

b) Increased the convenience: One of the passwords problems is that they are easily forgotten or lost. Often computer users change passwords more than one time, and this raises the possibility of forgotten passwords. Tokens and cards can be forgotten or losing it also; though keeping them at the safe location. Because biometric data is difficult or not impossible to forget, they can provide more convenience compared than traditional methods such as Password or PIN code.

Sales points have begun that provide many of applications for a customer using biometrics systems during the purchasing processes, which abolished the need for using cards.

c) Increase the accountability: Due to increased consciousness of security in the institutions as well as in applications facing customers, grown the need to audit capabilities and write clear and accurate reports. The use of biometrics to secure computers or another device provides a high degree of security of information that stored in a computer or other devices [2].

1.3 The Problem Statement

In the recent years, Face recognition has become one of the interesting area of research in computer vision field, exactly after increasing consciousness of security in the institutions. In face recognition, the simple process of face recognition system should go through image data retrieval, face detection, preprocessing, feature extraction, feature matching and face recognition. However, some of the researches focus on the part of face recognition system, such as face detection, faces recognition, or algorithms that used in extract feature or classify it. In addition, there are many of factors that effect on the image and make it a challenging for any system, such as complex background noise, lightening condition, poses variation, and other factors; and most of the researches only focus on. Hence, this work focuses on introduced a combination of techniques that can that provide a robust face recognition system in addition to high recognition performance and it can be able to overcome to most of problems that meet any system.

To do that each of PCA-Wavelet is used to extract the most important features and decompose the size of the image to reduced the time that needed to recogniztion. Futhermore, SVMs classifier with different kernelsfunction is used to obtain the best classification results between the different classes and K-NN is used to matching the features to obtain the nreast image to input one. Finally, to give reliability to the system, two types of the databases are used to evaluate the accuracy of the recognition system. Moreover comparison between its results with the proposed system is implemented

1.4 Outlines of The Thesis

The thesis is stated as the following:

- Chapter 1: Provides general information about the biometrics systems and explain the advantage of these systems. In addition, thesis objectives and thesis organization are presented.
- Chapter 2: Presents an introduction of two and three dimension face recognition system and a brief literature review of the related work done in the fields of face recognition. In all surveys that intrdouced, a brief account of early work is presented first, before gradually moving into more comprehensive descriptions of prominent recent works.
- Chapter 3: Gives background theory of the work depended on problem statement of this thesis. Wavelet transform and pinciple componet analysis have been discussed extensively. Additionally, Artifical neural network, support vector machine and K-Nearest Neighbor are presented.
- Chapter 4: Presents the architecture of the system, which is based on a principle components analysis and wavelet method with. A concise overview of the system is first presented, followed by a detailed description of the four main processes: preprocessing, feature extraction, classification and identification. At the final, several experiments on the performance of the system and comparative analysis with other popular techniques are presented.
- Chapter 5: Give conclusion of the thesis by summarizing the overall work and major contributions accomplished, and indicating possible future works of this work.

CHAPTER 2

LITERATURE REVIEW AND TYPES OF DATABASES

2.1 Introduction

All types of research on this topic can be divided into two types of system [3]:

1. Two-dimensional (2D) face recognition system.
2. Three-dimensional (3D) face recognition system.

The two dimension recognition can be implemented with two-dimensional image data while the three dimensions can be achieved by three-dimension of image data and each of these systems have numbers of advantages and disadvantages points.

The way to obtain a two-dimensional image is easier and less expensive compared to three-dimension image. On the other hand, the problem of two-dimensional images is sensitive to the change of light. In contrast, three-dimensional images are not affected by the light change. Also, to model the surface of the face is simple by using three dimensions. In contrast, the two-dimensional images do not contain the depth data.

Figure (2-1) describes the all of techniques that are used in face recognition. Face recognition is considered as one of the dialectical issues in a pattern recognition field. To solve this problem two stages are needed. These two-stage are training and learning. For two-dimensional, a number of methods such as Linear/Nonlinear Projection methods and to solve these problems the methods such as Neural Networks can be utilized.

The methods of analysis of the Linear Projection techniques are principal component analysis, linear discriminant analysis, and spectral feature analysis,..etc.and For the Nonlinear Projection techniques are as kernel principal component analysis, back-

propagation Neural Networks (BPNN), Radial basis neural network (RBNN), feedforward neural network (FFNN) as well as Multi-Layer Cluster neural network.

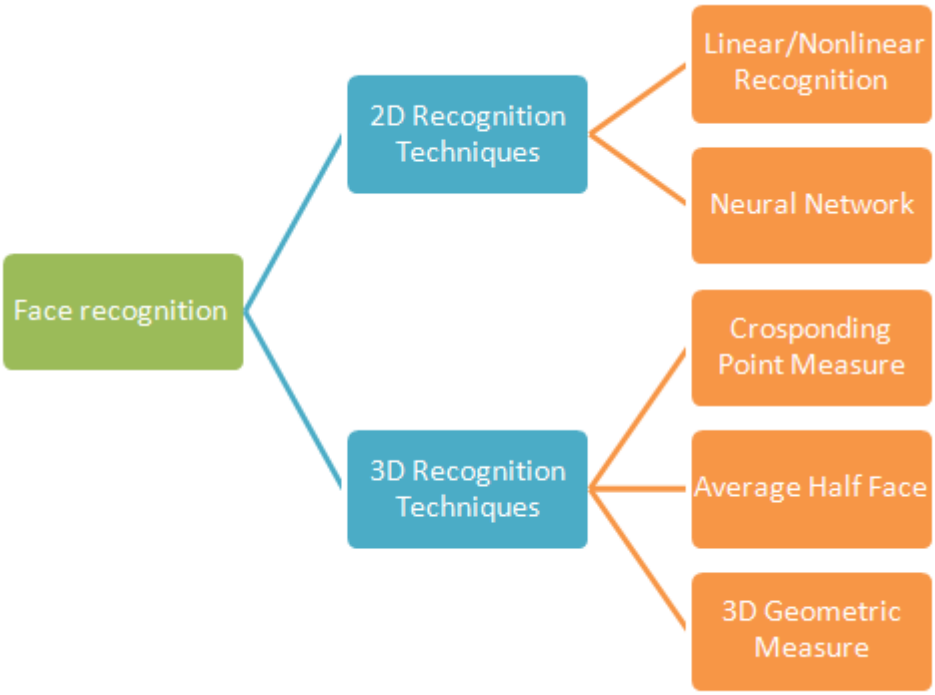


Figure (2-1): The types of face recognition techniques [3]

The two-dimensional Linear and Nonlinear projection techniques are applied to produce features vector for each image of the person's face, and then use these feature vectors are used to classify the input face image of the person inside the database. This technique is very effective to reduce the size of the images, to enhancement the input image and to reduce the effect of the lighting of the image before the process of recognition.

Enhancement of the images is also known as normalization techniques . The propose of the usage of these techniques is to find the difference in illumination between both side of (right and left) an image and to average the feature extraction illumination if the difference between the two side is large.

It is done by using one of feature extraction techniques such as with Gabor Wavelets, PCA, LDA or ICA. Finally, one of the classification method such as SVM and ANN are be applied to classify the input face image [3].

2.2 Literature Review

Masoud Mazloom et al. (2005). Introduced a hybrid approach for face recognition by handling three issues together. In this work, for preprocessing and feature extraction, a combination of wavelet transform and PCA was used while during the classification phase, a multi-layer perceptron neural network (MLP) was applied for strong decision in the presence of various facial changes. The experiments that done on the Yale database proved that the combination of each of Wavelet, PCA, and MLP provide good performance, The results provided the lowest overall training time, the lowest redundant data, and the highest recognition rates when compared to other methods [4].

Yen-Shao Chen et al. (2008) .presented face recognition system and this system was designed for the work in a mobile robot. In fact, The recognition system contains from two subsystems. The first subsystem is aimed to detect a face in the image and the second one was aimed to recognize faces in the image. In the face detection subsystem, facial color filter, connected component labeling procedures, an ellipse mask, and normalize face blocks for the later faces recognition procedures were used. While in face recognize system, the PCA method with Euclidean distance was employed to recognize the face. The experimental results prove that the proposed real-time face recognition system for mobile robots have fast and precise recognition abilities [5].

Gupta S. et al. (2010). Presented an approach for face recognition with Eigenfaces using principal component analysis (PCA). The research has been focused on the effects of taking the number of significant Eigenfaces. Experimental results showed that only 15% of Eigenfaces with the largest eigenvalues were sufficient for the recognition of a person and also prove that if the minimum Euclidian distance of the test image from other images was zero, then the test image completely matched the existing image in the database if minimum Euclidian distance was non-zero but less

than the threshold value, then it is a known face but having a different face expression else it is an unknown face [6].

Lei Z.et al. (2011).supported an effective local frequency descriptor (LFD) for low-resolution face recognition, by building upon the ideas behind local phase quantization (LPQ) and exploring both blur-invariant magnitude and phase information in the low-frequency domain. The proposed descriptor is more descriptive than LPQ with more comprehensive information. In addition, a statistical uniform pattern definition method is introduced to improve the efficiency of the proposed descriptor. Experimental results on FERET and a real video database show that LFD is effective and robust for low-resolution face recognition [7].

Kaur S. et al. (2013). Proposed a noble face recognition algorithm which integrates the principal component analysis (PCA); back-propagation neural network (BPNN) and discrete cosine transform (DCT) to improve the performance of face recognition algorithms. PCA is used to reduce the dimensionality of the face image and the recognition is done by the BPNN for efficient and robust face recognition. DCT is an accurate and robust face recognition system used in compression due to its compact representation power. Various faces images are considered in this research work to evaluate the performance of the proposed algorithm. The proposed algorithm is efficient and fast because DCT reduces the amount of time required to recognize a face [8].

Javed A. et al. (2013). Proposed method to recognize a person by comparing the characteristics of face to those of known individuals. In this work, the focus is on frontal two-dimensional images that are taken in a controlled environment i.e., the illumination and a constant background. Principal Component Analysis (PCA) technique has been used in the proposed system for face recognition. The purpose is to compare the results of the technique under the different conditions and to find the most efficient approach for developing a facial recognition system [9].

Oommen A. A., et al. (2014). Introduced a face recognition system by using Principal Component Analysis (PCA) method to reduce the data dimension of the image. It is based on the approach that breaks the face images into a small set of

characteristic feature images. These “eigenfaces” are the principal components of the initial data set of face images. Recognition is done by comparing the input face image with the faces in the data set through distance measuring methods. Here the face recognition system is developed using Matlab and it recognizes the input face from a set of training faces [10].

Rajath Kumar M. P. et al. (2015). Presented a face recognition system using Principal Component Analysis (PCA) with Back Propagation Neural Networks (BPNN) is analysed. A neural based algorithm is presented to recognize the frontal views of faces. The multi-variate data set of face image is reduced using the PCA technique. BPNN is used for training and learning, leading to efficient and robust face recognition. Experiments and testing were conducted over Olivetti Research Laboratory (ORL) Face database. Results indicate that PCA based execution is faster while the recognition accuracy suffers and vice versa for BPNN, thus suggesting a system incorporating both techniques is preferred [11].

Can Eyupoglu et al. (2016). Presents face recognition system based color. In this paper, k-Nearest Neighbors (k-NN) is used in order to classify color face images. Firstly, the classification is performed using only k-NN classifier. After that Principal Component Analysis (PCA) and k-NN classifier are used together. In addition, these two methods are implemented for different color space models and k values. Finally, the experiment results are compared with each other [12]**.**

2.3 Face Recognition Database

When one of the researchers proposes a new method. The researcher always needs to evaluate the performance of his method and compare the results to show the efficiency of his method. Therefore for researchers a standard test database is required to compare directly the results. While there are many databases in use currently, the choice of an appropriate database to be used should be made based on the task given (aging, expressions, lighting etc). Another way is to choose the database specific to the property to be tested such as how algorithm behaves when given images with different lighting condition or images with different facial

expressions. On the other hand, any method to be concluded needs to be trained with more images per class.

In this work, the following databases are chosen for consideration:

Yale Face Recognition Database (Yale University), United States of America and ORL Face Recognition database (Cambridge University Computer Laboratory) United Kingdom.

2.3.1 Yale Database

This is the first database used to calculate the performance and efficiency of the proposed method of recognition. This database is very useful because the image of this database takes under different lighting condition and face expression. It consists of fifteen classes and these images are in gif format with eleven images for each class and each image is taken under various facial expression and light condition such as with glasses, without glasses, happy, sad, normal and sleepy. Figure (2-2) shows some sample face image of Yale database face image [13].

Figure (2-2): Some samples of Yale database [13].

2.3.2 ORL Database

ORL Database is one the most famous database which consists of a number of images that captured between 1992 and 1994 at the Computer Laboratory and this database was applied in the many of a face recognition projects.

In this database, all images are in PGM format and the dimension of the every image in the database is 92x112 pixels and with 256 gray levels per pixel. The images are put in 40 folders (one folder each person) and the each folder hold name (Sx), where "' x " refers to the person number. In each of these folders, there are ten different images of each person and each image hold name (Y.pgm), where Y is referred to the number of the image of each person and it is between 1 and 10.These images were taken at different times, facial expressions and also at different lighting condition. Figure (2-3) displays a number of images of ORL database [14].

Figure (2-3): Some samples of ORL database [14].

CHAPTER 3

THEORETICAL BACKGROUND

3.1 Overview

This chapter introduces the theoretical background of this work. In Section 3.2, the wavelet transform technique is presented and in section 3.3, the techniques of image discrimination for representing the face image are introduced. In Section 3.4, the classification methods such as Support vector machine and k-nearest neighborhood are discussed.

3.2 Wavelet Transforms

Wavelet transforms are very effective methods and specifically in image compression and many other types of data. Because many coefficients of wavelet transform method for a standard image approach zero or become very small coding of the coefficients is easy. For this reason, wavelet transforms can consider as a very helpful tool for image compression. The primary benefit of wavelet transforms compared to other decomposition methods is that the basis functions that related with a wavelet decomposition typically have short and long support. The short support can effectively act on sharp transitions as the edges of the image however long support is operative in representing slow variations of the image. These points make wavelets perfect to represent the most of the low-frequency signals that consist of mixed with a little number of sharp transitions. The basis functions in the most classical transforms method have supported in the whole image and thus representing edges and slow variations are difficult [15].

3.2.1 The Continuous Wavelet Transforms

The Continuous Wavelet Transform or also know by CWT is represented by equation (3-1), where x(t) is the signal, and $\gamma\ t$ refers to the basis function. the wavelet function that applied for transformation which become the mother wavelet over shifting and dilation compression.

$$X_{WT}(\tau.s) = \frac{1}{\sqrt{s}} \quad x\ t\ \cdot \gamma^{*}(\frac{t-\tau}{s}) \qquad (3\text{-}1)$$

The mother wavelet applied to create the basis functions. The parameter τ in the equation (3-2) is associated with the position of the wavelet function as it is shifted over the signal. So, it matches with the time information in the wavelet transform. The scale parameters can be represented as |1/frequency| and associated to the information's frequency. Scaling is either expands the signal or compresses the signal.

Large scales are aimed to spread the signal and by this way, the hidden information in the signal will appear, while small scales or high frequencies are aimed to compress the signal and thus only the global information about the signal will appear. For more information, the wavelet transform is done together with the convolution operation applied to both the signal and the basis function. In many of practical applications, the analysis above is very helpful. The Wavelet Series is achieved by discretizing CWT.

This aids in the computation of CWT using computers and is obtained by sampling the time-scale plane. The sampling rate can be changed accordingly with scale change without violating the Nyquist criterion. Nyquist criterion states that, the minimum sampling rate that allows reconstruction of the original signal is 2ω radians, where ω is the highest frequency in the signal. Therefore, as the scale goes higher (lower frequencies), the sampling rate can be decreased thus reducing the number of computations [16].

3.2.2 The Discrete Wavelet Transforms

The discrete wavelet transform (DWT) is a linear transformation. Its operating length is an integer power of two, transforming it into a numerically different vector of the same length but It is a tool that separates data into different frequency components, and then studies each component with a resolution matched to its scale. DWT is computed with a cascade of filtering followed by a factor 2 subsampling. Figure (3-1) shows the DWT Tree.

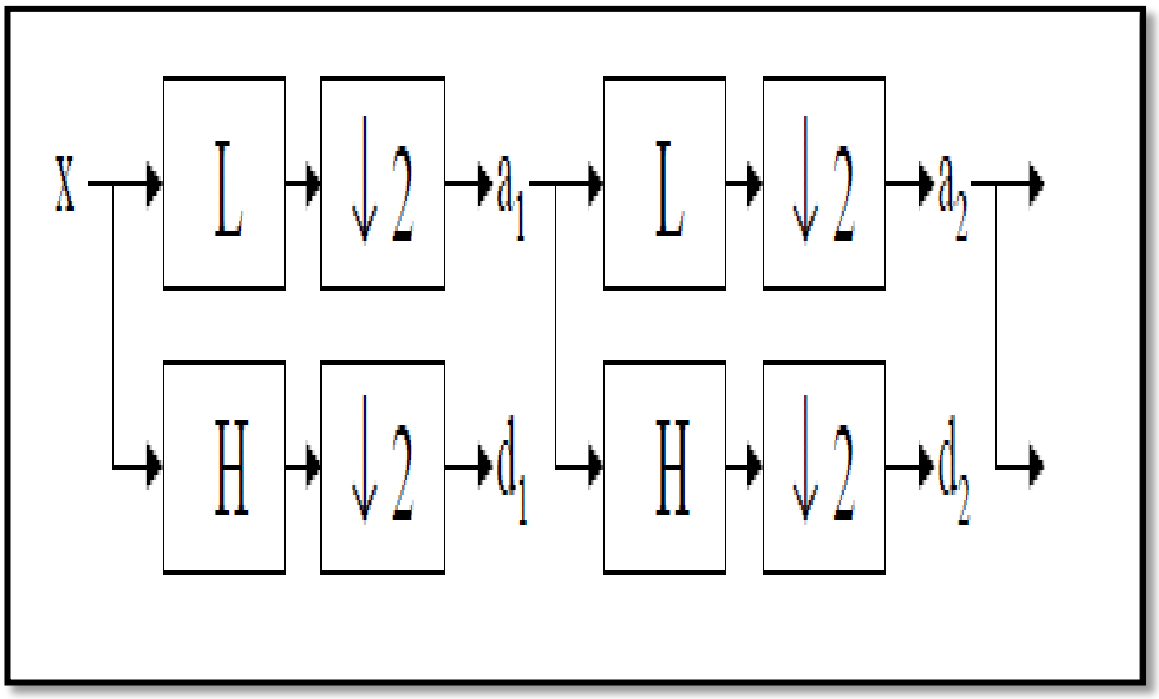


Figure (3-1): DWT Tree [16]

Where symbols H and L in the figure refer to the both high pass and low pass filters and symbol (↓ 2) refer to subsampling. The outputs of two filters are obtained by equations (3-2) and (3-3):

$$a_{j+1}[p] = \sum_{n=-\infty}^{+\infty} l\ n-2p\ a_j\ [n] \tag{3-2}$$

$$d_{j+1}[p] = \sum_{n=-\infty}^{+\infty} h\ n-2p\ a_j\ [n] \tag{3-3}$$

In the equation (3-3) $\boldsymbol{a_j}$ parameter are applied for next scale of the transform while $\boldsymbol{d_j}$, parameter is known as wavelet coefficients and it is used to define output of the transform. The ***l*[n]** and ***h[n]*** parameters represent the coefficients of low and high-pas filters respectively One can assume that on scale ***j***+1 there is only half from number of a and d elements on scale ***j***. This causes that DWT can be done until only two $\boldsymbol{a_j}$ elements remain in the analyzed signal these elements are called scaling

function coefficients. DWT algorithm for two-dimensional pictures is similar. The DWT is performed firstly for all image rows and then for all columns.. The wavelet decomposition of two-dimensional images is shown in figure (3-2).

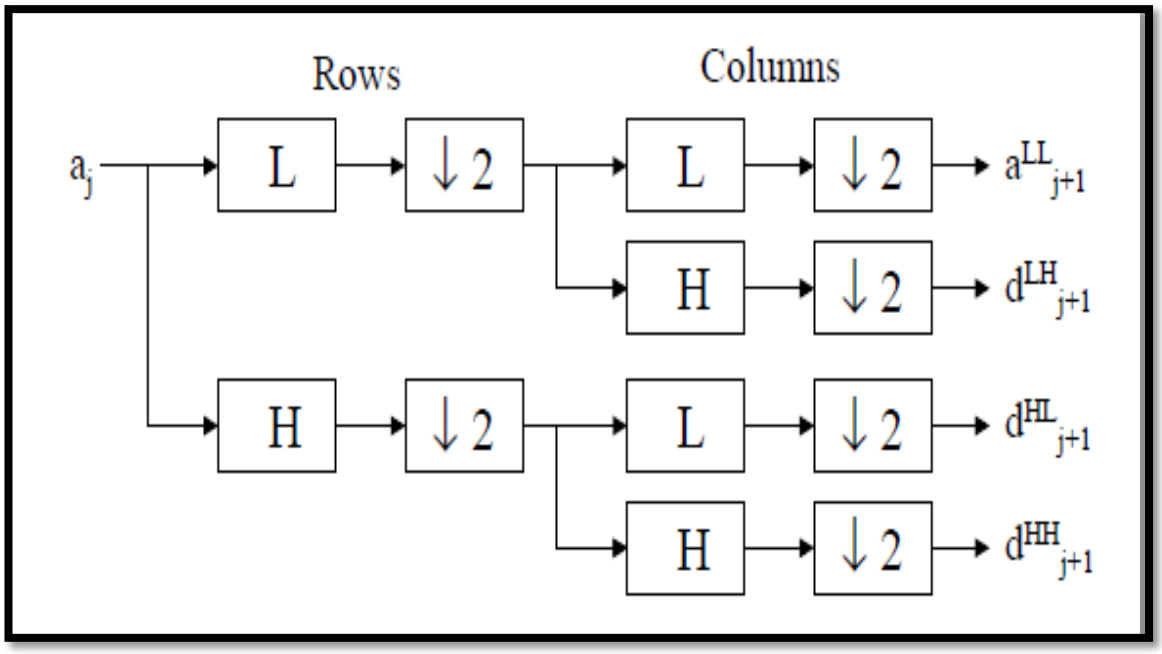


Figure (3-2): Wavelet decomposition of two-dimensional [17]

One of DWT advantages is the function representation using a number of scales. So that, function that obtained by the wavelets can be analyzed at different levels of resolution. The second advantage of DWT is that invertible and also it can be orthogonal. Wavelets transform is also very effective for analysis the textures recorded with various resolution. One of the problems of nuclear magnetic resonance (NMR) imaging it is that it needs a long time of acquisition. This led to artifacts caused by patient movements and this cause must be averted [17].

3.3 Techniques of Image Discrimination

There are many techniques for discriminating. These techniques are dependent on types of image applications such as in face recognition or detection, iris, ear, figure print analysis etc. These techniques take a very important part in the analysis of the image as well as in features extraction. In image analysis, the main objective to be used the reduction in the size of data dimensions. There are a large

number of methods can be applied in this research area. But in this present work Principle Component Analysis or PCA is applied [18].

3.3.1 Principle Component Analysis

This technique is also called Karhunen Lo`eve transformation. It is one of the linear transformation technique that is used to obtain features from data or to decompress the information It is used to take global structures from the high-dimensional information set as well as it is applied to decompose dimensionality of the data and take unique features from the faces picture. This technique can be used also to distinguish patterns in data and express the data so as in order to highlight their differences and similarities between them. This technique also provides a successful way for dimensional decomposition of the data.

The image of the face image is required to be represented in a low dimension subspace.This is done by using one of the discrimination techniques such as the Principal Component Analysis. This subspace is created by using the eigenvectors or principal components of the array that created from the image of the face image in the dataset. Mathematically, eigenfaces represent the eigenvectors that derived from the covariance matrix of the image dataset and the eigenvectors represent the significant variations through the faces in the dataset and every image of the face in the training dataset has a unique position in the subspace. In this technique, every image will be tested projected onto this feature space.

The variation of faces image resulted from a different level of illumination is much more than the variation that generated from the identity of the person. Therefore eigenface method is applied to reduce the dimension of the image. Where PCA gives directions of projection that aim to maximize the total scatter across all faces image in the dataset [19].

3.3.1.1 PCA Algorithm

At first, the training set is generated. This set consists of a number of images (M) and each image is to be transformed to the vector (N^2). The step after generating the training set has calculated, the mean of images as this is given by equation (3.4):

$$\Psi = \frac{1}{M} \sum_{n=1}^{M} \Gamma_i \tag{3-4}$$

Γ_i in the Equation (3-5) refers to the image in the training group. Then the normalization is implemented to the trained set to achieve the zero mean that describes the degree of deviation between the mean of image and image and this is described in:

$$\phi_i = \Gamma_i - \Psi i = 1, 2 \ldots \text{M.} \tag{3-5}$$

The covariance matrix *C* of the training set can be obtained using Equation (3.6):

$$\Psi = \frac{1}{M} \sum_{n=1}^{M} \phi_i \phi_i^T \tag{3-6}$$

İf a vector u satisfies the condition in equation (3-7), the eigenvector of the covariance matrix C is considered a non-zero vector.

$$C\, u_k = v_k\, u_k \tag{3-7}$$

Where v_k refers to the corresponding eigenvalues.

The size of matrix *C* with the dimension of N X N is large and need much computational time and size memory to obtain results. To solve this problem, we consıder the matrix *L* that has small size compare than the original covariance matrix *C*. Equation (3-8) describes how the matrix C can be decomposed:

$$C = AA^T \tag{3-8}$$

where A represents zero means matrix $[\phi_1, \phi_2 \ldots, \phi_i]$ and the eigenvectors v_i can be consideration according to the following:

$$A\, A^T v_i = \mu_i v_i \tag{3-9}$$

Now, by multiply each side of equation (3-10) by *A*, we can obtain the following:

(3-10)

$$A\,A^T A v_i = \mu_i A v_i$$

From Equation (3-10) we can see that, the Av_i represnts the eigenvectors of the covariance matrix C. The matrix *L* with dimension (MXM) can be rewritten as show in the equestion (3-11):

$$L = A\,A^T \tag{3-11}$$

Now, we obtained the eigenvectors v_i of smaller covariance matrix *L* that has dimension *M* and this way, the linear combinations of the training data of the images to the form of Eigenfaces u_i can be obtained from the equation:

$$u_i = \sum_{k=1}^{M} v_{ik}\phi_k \tag{3-12}$$

This eigenvector that has symbol u_i is also called as eigenfaces because when we transfer these eigenvectors form a vector of length (N) to a two-dimensional matrix with dimension (NX N) and show it. it is shown as the human face [19].

3.4 Support Vector Machines

It is one of important learning algorithms that are applied in many of applications and it is currently represented one of the very used methods in many application fields. SVM has been applied in many of applications in many various fields such as medical diagnosis, text or image classification and also categorization, spam categorization, detection and recognition of an object, face detection, verification and recognition, bioinformatics, signal processing, prediction, information, and image retrieval.

SVMs have been developed in the in the inverted direction of the neural network, as SVMs have been evolved depended on the sound theories for implementation and experimentation, whilst neural network has been evolved depended on intuition and heuristic rules, which it come from the applications and spacious experimentation. SVMs is like to neural network possess, the capability to approximate of any

multivariate function to different accuracy and thus, SVMs are very appropriate for modeling complex systems and nonlinear system, [20].

3.4.1 Advantages of SVM Based Classifiers

One of this advantages of SVMs is that it needs only a small training set in the training stage because only a few number of vectors are chosen as support vectors. Hence, the computational cost is reduced and it provides a better generalization. Finally, the SVMs solution is unlike to that of neural networks because it does not depend on the starting initial conditions [21].

3.4.2 Basic SVM Formulation

SVMs is used to make classification among two classes to give a decision surface.This decision surface is represented by maximum distance that can depend on it to separate closest points in the classification's class. The closest points are known as support vectors data. Figure (3-3) describes the hyperplane of support vector separating the data.

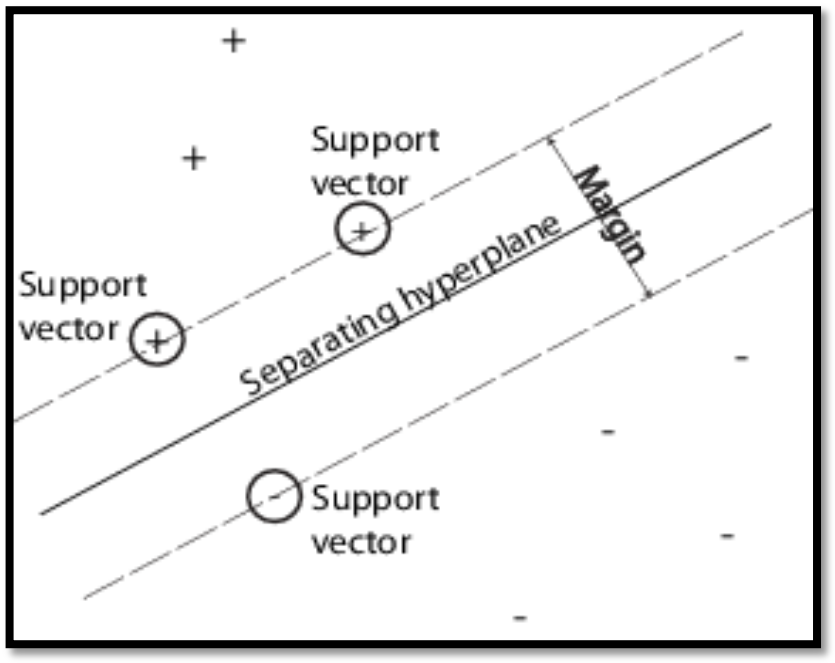


Figure (3-3): The support vector and the separating hyperplane [21]

It can be considered that SVMs is maximum margin classifiers. To facilitate the condition that input data can be supported to linearly separable. Under this supposition. Any two classes can by obtained the linear hyperplane between them. Figure (3-10) describe the good and bad separation between two classes.

Let us assume that a binary classification, the input data x_i where i=1,2,3,.., Land $y = \mp 1$ for the two class, y = f(x) is:

$$f(x) = sign(w \cdot x + b) \quad (3\text{-}13)$$

Where sign refers to label of class.Where symbol $w \cdot x$ represnted the inner product and symbol b represnt the bias of the function and point x take position directly on the hyperplane and who condition is achieved, i.e,

$$w \cdot x + b = 0 \quad (3\text{-}14)$$

We can obtain the label y_i of a data by sloving equestıon (3-15).

The points x_i on the both side of the hyperplane must achieve the following conditions:

$$x_i \cdot w + b > 0 \quad (3\text{-}15)$$

$$x_i \cdot w + b < 0 \quad (3\text{-}16)$$

Equations (3-15) and (3-16) can be rewritten as the following:

$$y_i(x_i + b \geq 0) \quad (3\text{-}17)$$

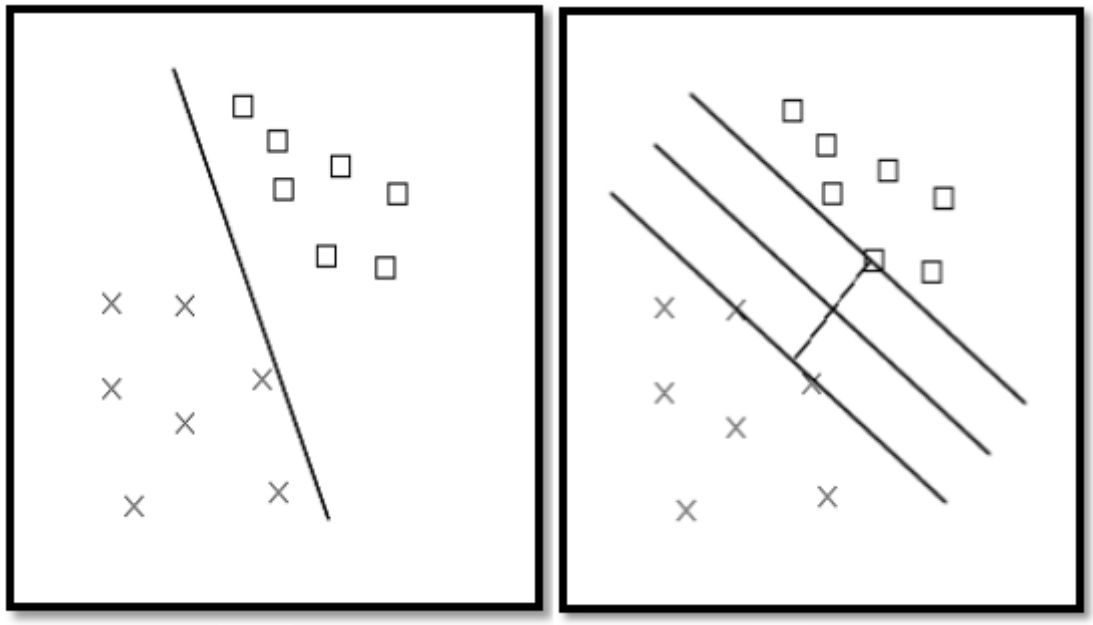


Figure (3-4): (A): Random separation, (B): The maximized separation [21]

Using the formulation of Lagrangian, the SVMs prediction can obtain using following equation:

$$f(x) = \sum_{i=0}^{m} y_i \alpha_i < x, x_{si} > +b \qquad (3\text{-}18)$$

Where ***m*** represents the number of support vectors, $\mathbf{x_i}$ is a support vector and $\boldsymbol{\alpha_i}$ is corresponding Lagrange multiplier and finally, a sign of ***f(x)*** is classified every test vector.

We can say that the output of classification is correct if equation (3-18) holds for all input points, then there is one ideal separating hyperplane in offer. On the ideal hyperplane, the margin among two data is maximized. As shown in the figure (3-3), the margin among two sets of data is represented by the distances (d+ and d-) between them. Figure (3-4) shows the two separating lines on of them is ideal and another is random[22].

3.4.3 Kernel Function: We can expand the solution of equation (3-19) to solve the nonlinear separating hyperplanes problem and that can be applied by mapping the input space into a high dimensional space from $x \rightarrow \emptyset(x)$ .The primary key of this

mapping is the function $\emptyset$. This function is subject to the condition that the dot product of the this functions $\emptyset\ x_i\ \ and\ \emptyset(y_j)$ can be formulated as a kernel function $k(x_i, x_j)$. Then the decision function in (3-10) will becomes as:

$$f\ x\ = \sum_{i=0}^{m} y_i\, \alpha_i\, k\ x, x_{si}\ + b \qquad (3\text{-}19)$$

There are many types of SVMs kernel functions and these kernels are depend on different algorithms. In the next section, the most important four types of SVMs kernel function will be introduced [23]:

i) **Linear Kernel Function**

Linear Kernel function is a famous method and simple one. The first kernel function that applied to classify the data is:

$$k(x, x_{i)} = (x \bullet x_i\ ) \qquad (3\text{-}20)$$

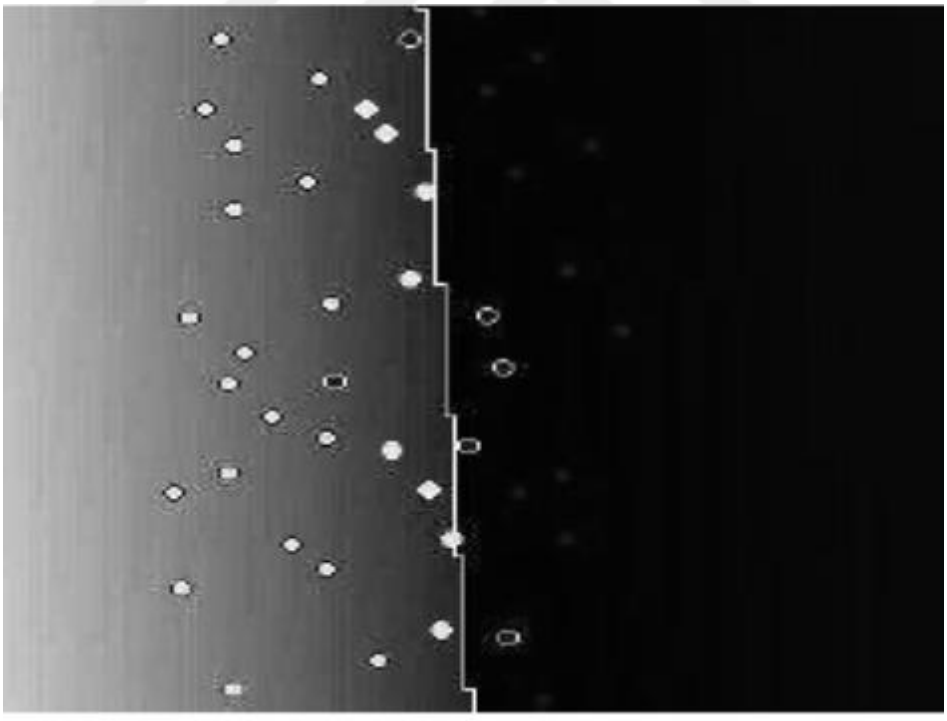


Figure (3-5): Linear Kernel [23]

ii) **The polynomial Kernel Function**

The polynomial kernel function that is applied for non-linear modeling is

$$k(x, x_{i)} = [(x \bullet x_i\ ) + 1]^q \qquad (3\text{-}21)$$

This kernel is usually more effective than the first one because it can avoid the problems associated with the Hessian zero.

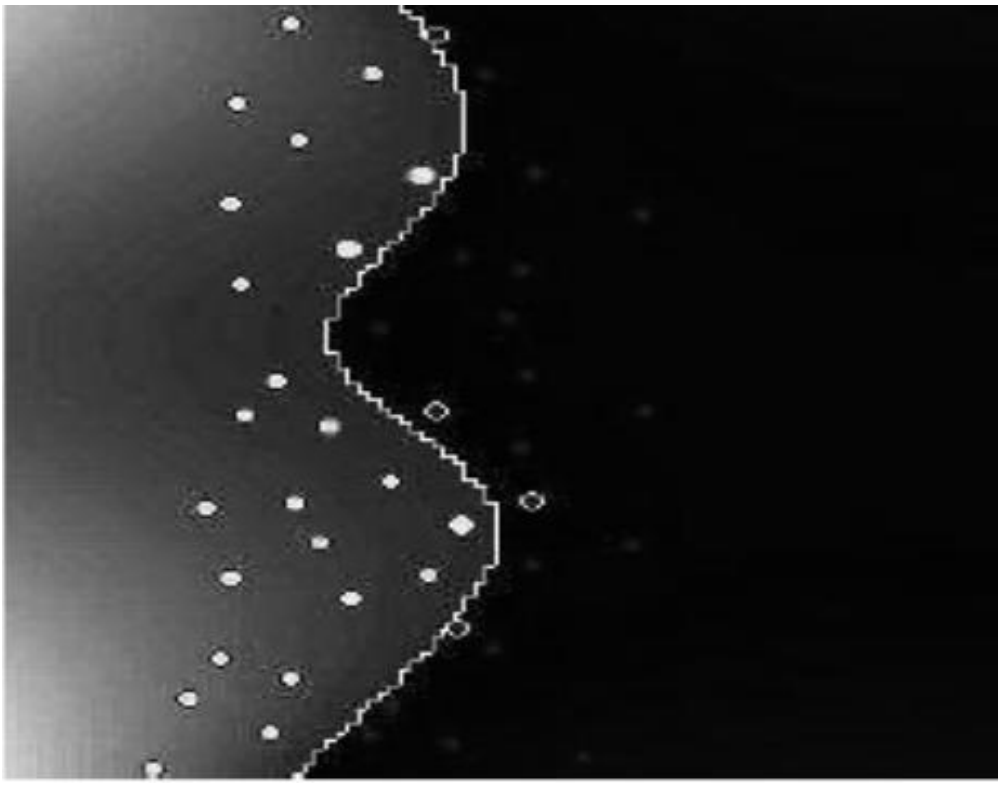


Figure (3-6): Polynomial kernel [23]

iii) Radial Base Kernel Function (RBF)

Radial basis functions are the third types of the kernel has the Gaussian of the form as,

$$k(x, x_i) = exp\ \ -\frac{|x - x_i|^2}{\sigma^2} \qquad (3\text{-}22)$$

Where x_i is the input data [23].

iv) Multi-Layer Perceptron

It is on of the long established method. It consists of a single hidden layer expressed as

$$k\ x, x_i\ = tanh(v\ \ x \bullet x_i\ \ + c) \qquad (3\text{-}23)$$

The mapping function and feature space depend on the kernel functions. Hence, the dimensions of the sub-space will be changed according to kernel functions and their parameters and this change in complexity can determine the dimension of maximum vapnik chervonenkis dimension[24].

3.5 Artificial Neural Network

Artificial Neural Network (ANN) is a part of artificial intelligence field that aims to imitative the conduct of the brain of a human and nervous system. This system is described as a parallel computational paradigm, containing adaptive processing units connect with each other closely. One of the important features of neural networks method is their adaptive nature of adaptive. This characteristic makes the ANNs method to be used in many applications in engineering, scientific,....etc. to solve nonlinear problems [25].

ANN is also used in specific applications, such as pattern recognition, or classification and all of this done through the learning process. It can also be applied to extract pattern or distinguish complicated patterns that human or another ordinary computer cannot trace. The neural network training can give knowledge in the specific field to be analyzed and exactly in the field that used it to solve any problem that may appear in future [26].

In the list below some reasons that make from ANN is an important training method:

i) **The ability to adaptive of the learning:** ANN has the ability to learn from a given data or can learn from previous experiences and-and implement its task depended on it.

ii) **The ability to the self-organization:** ANN has the ability to build it information that presentation through the learning process stage.

iii) **The ability to real time operation:** The operations of ANN is implemented simultaneously.

iv) **Fault tolerance via redundant information coding:** Any damage that occurs at the certain level of the ANN won't affect the whole operation of the net.

3.5.1 Neural Network Types

The types of neural network that create by earlier researchers are fewest.There are several conditions that can be applied to choose the types of neural network that can be used from these conditions: the learning algorithm, and the types of data that used in the network. From this types: Feed forward neural network, perceptron neural network and dynamic neural network.

There are a little number of neural network models have to take the interest of researchers. Related on this, the development of the models has been implementing depended on the application that needed to be built. İn this part from the chapter, the commonly learning algorithms and different models of the neural network will be presented [26].

3.5.2 Learning Algorithm

Learning can be defined as that the ability of processing unit to change the input/output as a response to change that occurs in its environment. Usually, activation of the algorithm is chosen in the development stage of the neural network, in order to adjust the value of the neuron weights that associated with inputs in order to change their behavior. During the training stage, one of the methods is needed to adjust weight dependency on the process that is implemented in the input/output. There are many learning algorithms is designed to the model of neural network. The learning algorithm can be supervised or unsupervised and both of them will be discussed in the next section [26].

i) Supervised Learning

All input of the pattern that is applied to train stage in the network is related to an output of the pattern, which is the target or the desired pattern. During the learning stage, a trainer (sometimes called a teacher) is supposed to be present during this process, Then a comparison among each of the network's output results and the correct output results is done to calculate the error. The computed error is used to the update parameters of the network [26].

ii) Unsupervised Learning

In this method of learning, the target of the network is not given during the learning stage of the network. Of this kind, there is no trainer to given the desired patterns and in this way, the system can learn based on itself [26].

iii) Reinforced Learning

In this kind of learning, a trainer during the learning stage does not give the correct answer but only determine if the computed output of the network is correct or not. The data that given during the learning stage helps the network during the learning stage. For a correct answer, a reward is given and in contrast, a penalty is given for the wrong answer [26].

3.5.3 Data Types

In this section, two types of data that used in neural network will be discussed: [26]

A. **Category:** For this type of data a finite number of possible values are taken, and there are many statutes falling into each category. The variables of categorical may have only the symbolic values such as (male, female, red, white color) and this symbolic values must be converted to numbers before being applied to the network.

B. **Quantitative:** The variables can be defined as numerical measurements of some feature such as distance in the center, for this kind of data; some of the arithmetic relations between the measurements are reflected analogously and the feature of the objects are computed.

3.5.5 Advantages and Disadvantages of Neural Network

In this section, some advantages and disadvantages of ANN technique are introduced. The neural network is applied when the rules or standards for finding an answer are not obvious, therefore, sometime neural networks are known as black boxes.

One of the important advantages of using a neural network is the ability of adaptive to new scenarios and they can accept error and it can deal with the noise that appears. However, the time that needed a neural network for training is considered as the biggest disadvantage of this method as well as a need for large training sample to train. It is difficult to demonstrate the output results of what's happened inside neural network [27].

3.6 K-Nearest Neighbor

K-Nearest Neighbor is also called KNN is one of the most important algorithms. It is used in various of applications such as vision, proteins, computational geometry, and graphs. It also considered as one of the best ten algorithms in data mining.

KNN is considered as one of the non-parametric learning algorithms. This mean that it doesn't produce any supposition on the implicit data distribution. This is very helpful in the real world application, many of the practical data does not conform the classical theoretical supposition made such as linearly separable, gaussian mixtures and etc .

In spite of that, KNN is considered as a simple algorithm. it does not need data for training because this method doesn't have training stage or it has very minimal training stage. This means the training phase is pretty fast.

The KNN data in the metric space can be scalars or maybe vectors with multi-dimensional. So, the data in feature space, they have a concept of distance and this does not necessarily represent a Euclidean distance as it is famous.

Every of the training set is contained a group of vectors and every vector is related to the class label. In another word, The class will be taken (+1) for positive class represented and (–1) for negative classes represented. in spite of that, KNN has abilities to work with a number of classes equally.

The symbol "k" refers a number of neighbors that take in classification stage. This number is an odd number. For two classes, case k is equal to1.

Although the classification is the main application of KNN method but it can be used to implement estimation also it is a non-parametric classification method, therefore it has the ability to estimate of the arbitrary distributions.

KNN method isn't as other methods that used hypercube or kernel functions, instead of it is made to estimate at point x but it is centered of hypercube at point x. The estimate of the density is shown in equation (3-24):

$$p\ x\ = \frac{k/n}{v} \tag{3-24}$$

where n is referred to the total number of class and symbol $\boldsymbol{v}$ is referred to the hypercube volume. From the equation above, we can see that the numerator is constant and the effect of the volume is only on the density. Let's consider that the density at point x has the highest value. by this way, to obtain the k points that near to x points become very easy because these points are very near to x where assumed of high density. This led to that the volume of the hypercube is small and the values of density are high. On contrary to this, Let's us consider that value of the density at x point is low, then, the hypercube that necessary to encompass kNN is large and thus, the ratio is small [28].

3.6.1 KNN for Classification

In the section, the usage of KNN as a classifier will be explained. To do that, two datasets will be employed. One of them is the training set and the another is the testing set. The goal of this algorithm is to obtain the class label of the new point. The algorithm has various conduct depended on the value of k parameter.

i) The first hypothesis: the value of k parameter is equal to one (k=1)

To explain the first hypothesis. Let consider that x is the point to be labeled and to obtain the point that closest to it. Let consider the closest point y. In this way, the nearest neighbor rule asks to assign the label of y to x. This appears very simple, but this way is correct only when the number of points is not very large.

But in other situations when the number of data points is very big, the label of x and y may be same. To simplest the explain – Let's assume that the potentially biased coin. Then maybe the head will appear in next call. The same argument can be applied in this situation.

Now, Let's us apply another case here – suppose that every point are represented in a D-dimensional plane and the number of points is very huge. This led to conclude that the density of the plane on the each point of data is high. In other description, in the each subspace, a large number of points is offered and consider x point is in the subspace and is has many of neighbors but in this time consider y be the nearest neighbor to the point x.

We can say that, If the point x point and y are very close together, then the following can be proposed; the probability of point x and point x related to the one class is very large, then by using decision theory, both points related to the one class.

The reference [29] provides a very good argument about the rule of Nearest Neighbor. One of the important points is to obtain a very small error bound. This bound is described in question as:

$$p^* \leq p \leq p^*(2 - \frac{c}{c-1}p^*) \qquad (3\text{-}25)$$

where is the Bayes error rate, the symbol c represents the number of classes and the symbol p represents the error rate of NearestNeighbor. The result of equation (3-19) is that if the number of points very huge then the error rate of nearest neighbor becomes less than twice of the Bayes error rate.

i) The Second hypothesis: the value of k parameter is equal to two (k=2)

In general, we aim to obtain k nearest neighbor and then make a majority voting. As we mentioned previously, the value of k is odd when there is two number of classes. Let's assumed that k is equal 5 and there are three instances of class1 and two instances of class 2. In this situation, KNN method will label the new point as class1.

One of KNN techniques that it does not provide 1 to every neighbor and another common technique provides weight to each point according to the distance that calculated.

In another way, the values of each weight of point are inversely proportional to the distance of the point that will classify. From the statement given above the neighboring points have higher values compared than the ultimate points [29].

CHAPTER 4

FACE RECOGNITION SYSTEM DESİGN

4.1 Introduction

Face recognition can represent a key requirement in various types of applications such as human-computer interface, monitoring systems as well as personal identification etc. Therefore it was necessary to provide a robust face recognition system to become reliable for the success of these types of applications.

The function of face recognition is very complex and it is not easy as they appear. The human face is very different from each person to other. There are many factors such as the sweat, age and many of other characteristics that represent a challenge to any researcher in this type of study.

In this chapter, Four methods are introduced for face recognition. These methods are first way introduced in same face recognition system. The combination of these methods supports the recognition system and give better results compared to the others methods.

The goal of this chapter is to introduce face recognition systems that has provided the following:

- Capture features from face image using each of PCA and Wavelet transform techniques.
- Classify the image of the face using SVMs classifier.
- Identify the image of a face using KNN technique.
- Compare the results of the method that used to identify with feed forward back propagation artificial neural networks.
- Design and implementation graphics user interface (GUI) to make the use of program easy to the end user.

4.2 System Software

There are various types of programs that can be used in different applications fields such as system control, image processing, and artificial intelligence… etc. In this thesis, the MATLAB R2015b software is applied to create the different part of the algorithm.

This software is one of famous software that produced by MathWorks company is implemented to develop an algorithm using different types of language that can be used in programming such as C++, java ...etc. and also it can connect it with a number of applications board such as Arduino. This software is easy to use and it has many toolboxes that make the creation of the algorithm as well as it has powerful help documentation. In this thesis, many types of Matlab toolbox are applied to generate the algorithm such as machine learning, image processing, and artificial neural network toolbox.

4.3 Face Recognition System Block Diagram

In the last ten year, there are various types of face recognition methods that have been introduced and has been applied. But the ability of the human to recognize human through face after many years is excellent fantastic as well as his ability to recognize many of people until after the effect of time on their appearance.

As we said above,the ability of a human to recognize faces is very robust despite the large changes that may appear on shapes of people during the years of their lives but this ability different from one to another and in sometimes the human has a problem in recognition and in other times he loses the ability to recognition. Therefore, the need to a system can do this job becomes very significant in various types of applications such as human-computer interface, monitoring systems as well as personal identification. In another hand, until now, the ability of people to recognize a human face is not clear to scientists the spite the existence of many studies for many years in this topic.

Face recognition system that is introduced in this thesis consists of the following steps: preprocessing, feature extraction, classification and in the final step

identification. The stages of the system are shown in the Figure (4-1). The structure of a proposed system is shown in figure (4-2) and it will be introduced in the next sections.

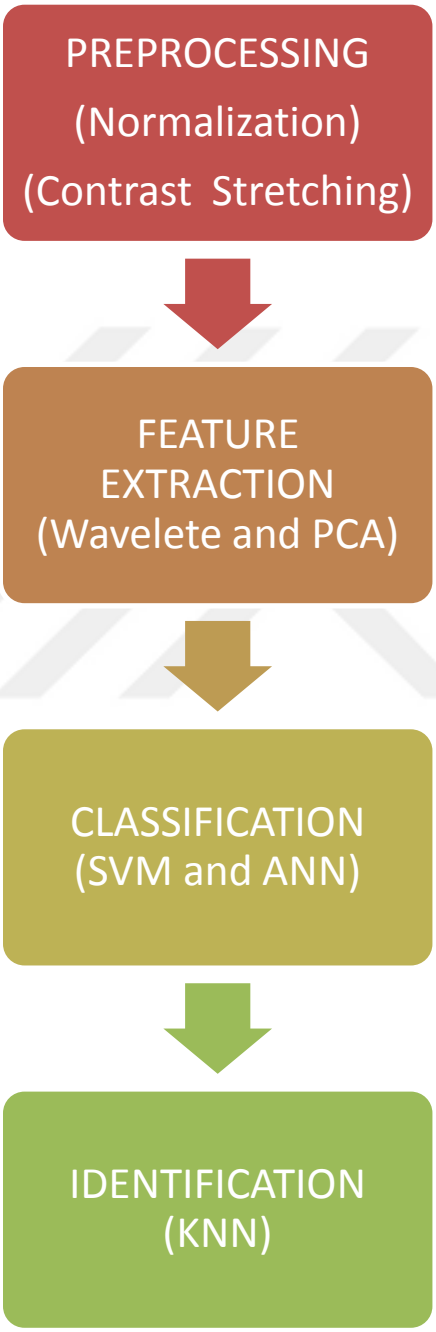


Figure (4-1): The stages of face recognition system

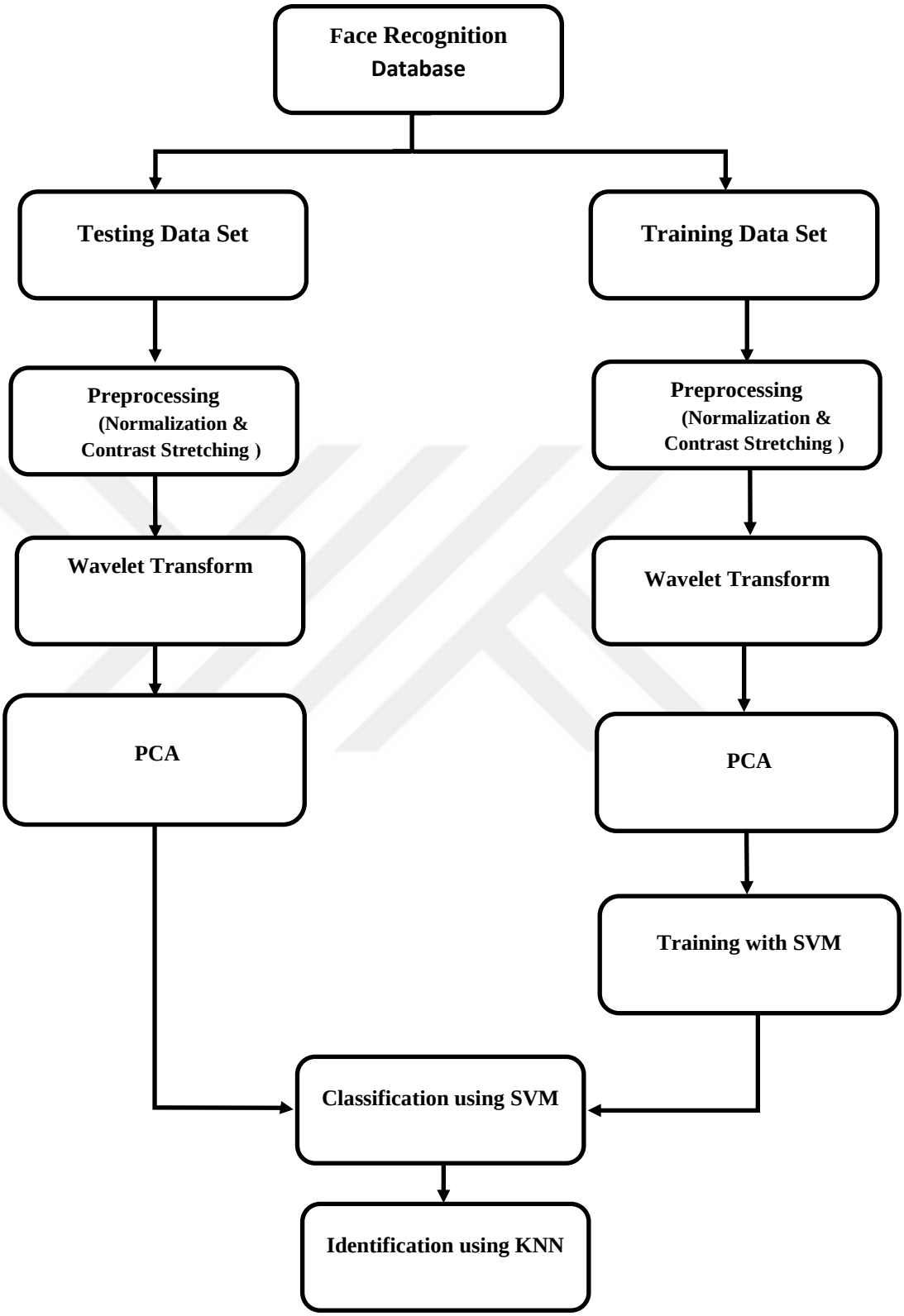


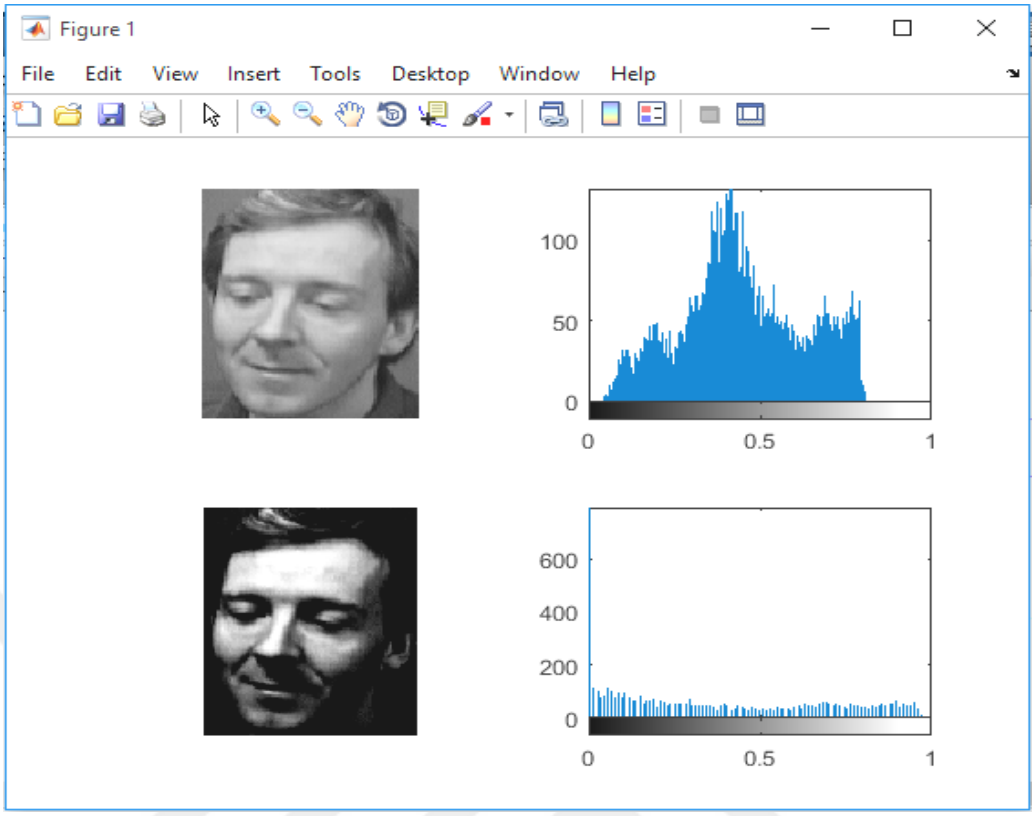
Figure (4-2): The structure of a proposed system

4.4 Preprocessing

It is the first process. It is applied to improve the image and also to improve the recognition rate through remove the effect of the light on and also to correct the gradient of illumination in the image. In the following the two preprocessing techniques that applied at this stage to get the good results:

i) **Normalization Size of the Image:** Two types of database (YALE and ORL) are used in this thesis and each database has different size of an image, therefore this technique is used to make all images to have the same size.This is because the next stages have an arithmetic operation such as multiplication and division of array. For all these reasons, the size normalization is very significant.

ii) **Contrast Stretching:** it is also known as histogram stretching. It is considered as one of the techniques that used to enhance an image. The goal of these techniques has improved the contrast in an image through stretch the range of intensity values. In the other word, this process is applied by raising the dynamic range of the gray levels in the image. For example, in an 8-bit system, the 256 gray levels can be shown only but if the number of gray levels in the captured image prevalence through a lesser range, the images can be improved by extended this number to a wider range. From the experiments, the image of ORL is better from YALE because the quality of ORL image is better than the quality of the YALE image. Figure (4-3) shows the two example of a contrast-stretched image in both YALE and ORL database.



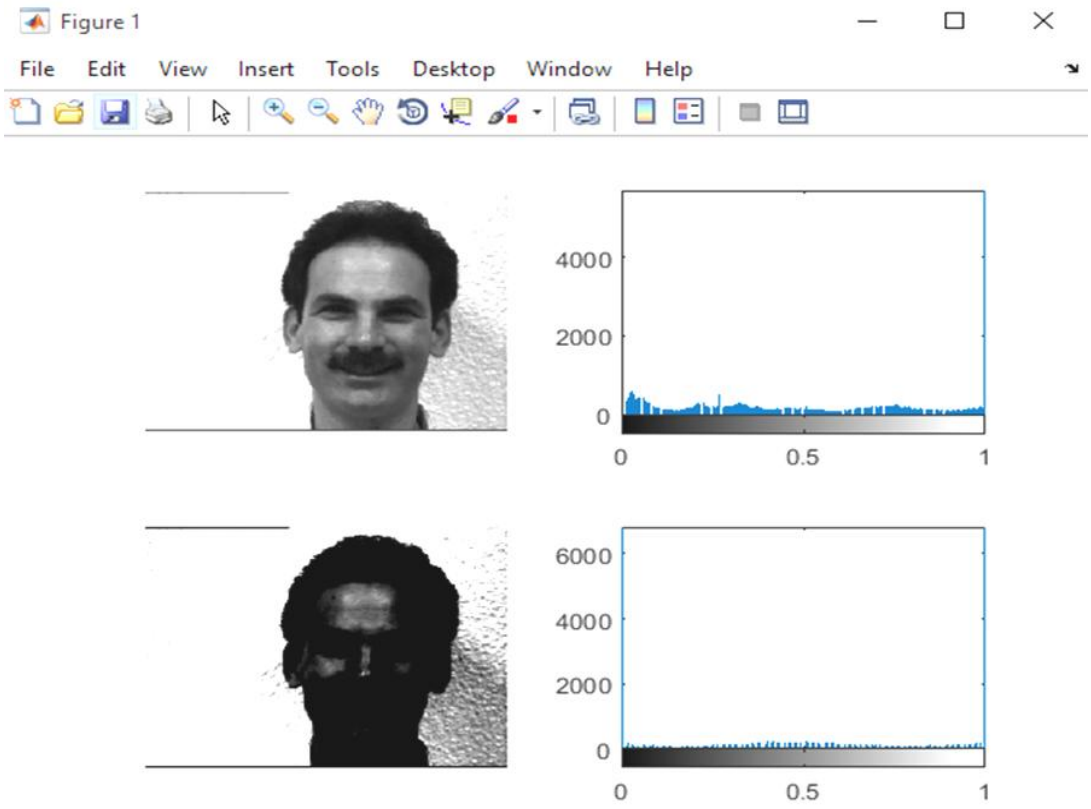


Figure (4-3): The contrast-stretched image in the YALE and ORL database[23].

4.5 Feature Extraction

The second stage of this system is feature extraction. It is used to obtain the significant features from the picture of the face and then these features are applied to the next stage of processing. Without this stage, the function of recognition and classification becomes very hard and very difficult. This is because that the picture consists of many features and most of these features cannot be applied to implement classification.

In this work, each of PCA and wavelet transform techniques is used to obtain the significant features that are important to identify and classify the face pictures. After preprocessing stage, the face picture I(x,y) a two-dimensional matrix (NxN) is then transformed to a vector and of dimension N^2. Thus the picture with dimension 64 x 64 will transform to a vector of dimension 4096. But this vector cannot be applied directly to recognize directly because the face of human are very comparable from each other and another reason the 4096-dimension space is big and it is needed much time to process. Therefore, Wavelet and PCA techniques are applied to obtain a better feature from face picture and also to enhance the capability of the recognition rate.

4.5.1 Wavelet Transform

It is the first step of this stage and it is used wavelet transform to remove the noise from the image with saving the quality of an image. The second benefit is to obtain the feature from the image and also to decompose the size of the image and by this way thus reduced the time of the process.

The output of the wavelet transform on an image, it consists of four- subbands. These subbands are: Low-Low (LL), High-Low(HL), Low-High (LH), and High-High (HH). The Low Low sub-band gives the most important features of the image and High Low and Low High subbands give the horizontal edges and vertical edges and finally, the High High subband gives the diagonal edges. The results that obtained after applying the wavelet transform is shown in figure (4-4).

When this technique is applied to the image with dimensions (112x92) once, the dimensions of the four subbands becomes 56x46.

we can conclude from the previous paragraph that after applying a stage of the wavelet transform, the size of the image decreases the half. In this work, two-level the wavelet transform is used.

(a)

(b)

Figure (4-4): (a): The original Image. (b): The result after used one-level of the wavelet transform.

4.6 Classification Stage

A classification is the third stage of this recognition system. It is used after each of previous stages is performed. It is a very significant stage. The goals of this stage are to classify the inter images based on the information obtained through the training stage. There is a number of classification methods that can be used to do this job. In this recognition system, support vector machine (SVMs) are considered as.

SVMs are one of most important learning algorithms which are applied in many of applications. In this work, we also aim to make a comparison between different kinds of the SVMs kernel to analyze its results and determine any kernels of them can give a good result with high accuracy rate and also we implement a comparison between a feed-forward backpropagation neural network FFBPNN and SVMs to ensure the reliability of our methods.

4.6.1 SVM Classifier

SVM is considered as a learning algorithm method depended on statistical learning algorithms. It is used to classify the feature of the data by finding a maximum marginal hyperplane that separates between two classes. It can also be applied for linear and non-linear classification using different types of kernels functions. The primary goal of the SVM is to obtain a plane between different types of features and by using this plane, we can separate these features. There are many types of program that provide toolboxes for SVMs, for example, Open CV and Matlab.

In this thesis, several toolboxes are used for statistical and machine learning operations. These toolboxes offer many of statistic methods that can be applied to solve many problems in different tasks. Two stages must be provided to obtain the recognition. These two stages are training and test.

The training stage requires to setup training image matrix. Then target vector related to the training data and kernel functions are generated. The training data matrix consists of features that are obtained from the previous stage (the preprocessing stage and feature extraction stage) and each column in this matrix refer to the image of the person.

The training set is divided into two sets; the first set refers to authorized people and the second set refers unauthorized people. The target is related to the persons in the training set. This is done by labeling label ‘1’ the persons that they want in the system to be accepted (authorized person) and label‘0’ for unauthorized people. We must notice that a number of columns in target matrix are equal to the number of images in the training set.

After creating each training set and target set. The kernel function is must be chosen for any types of classification. Many of experiments are implemented for different types of the kernel functions such as linear, quadratic, polynomial, RBF and map kernels functions to determine the best results. The figures from (4-5) to (4-15) show the classification of data in YALE and ORL database using different kernel function and the experiment results explain these figures.

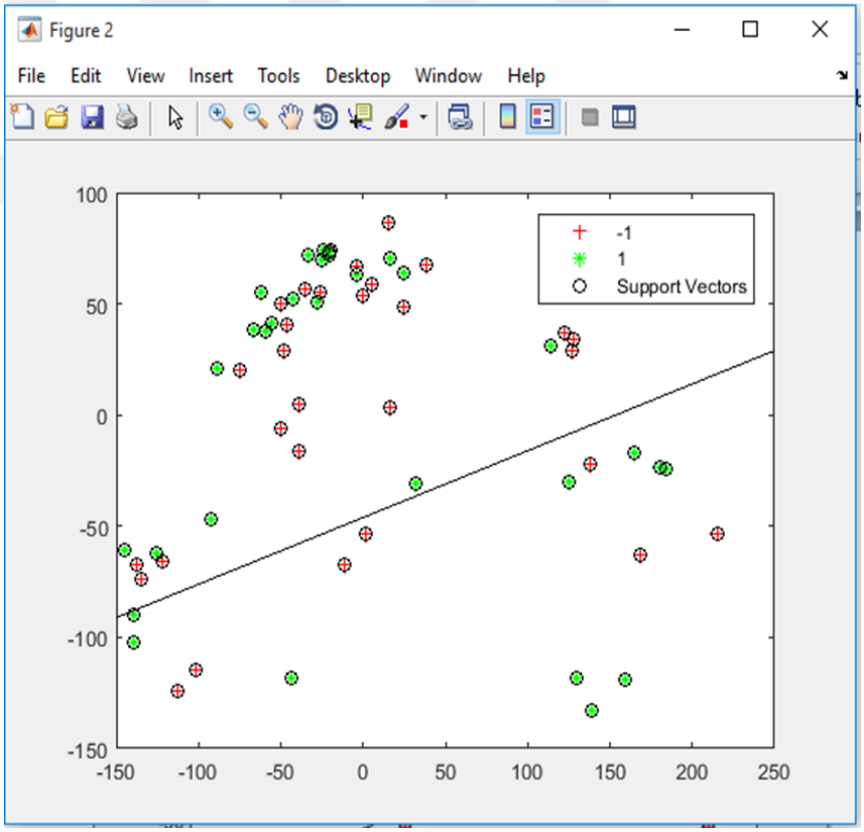


Figure (4-5): Linear kernel in YALE database

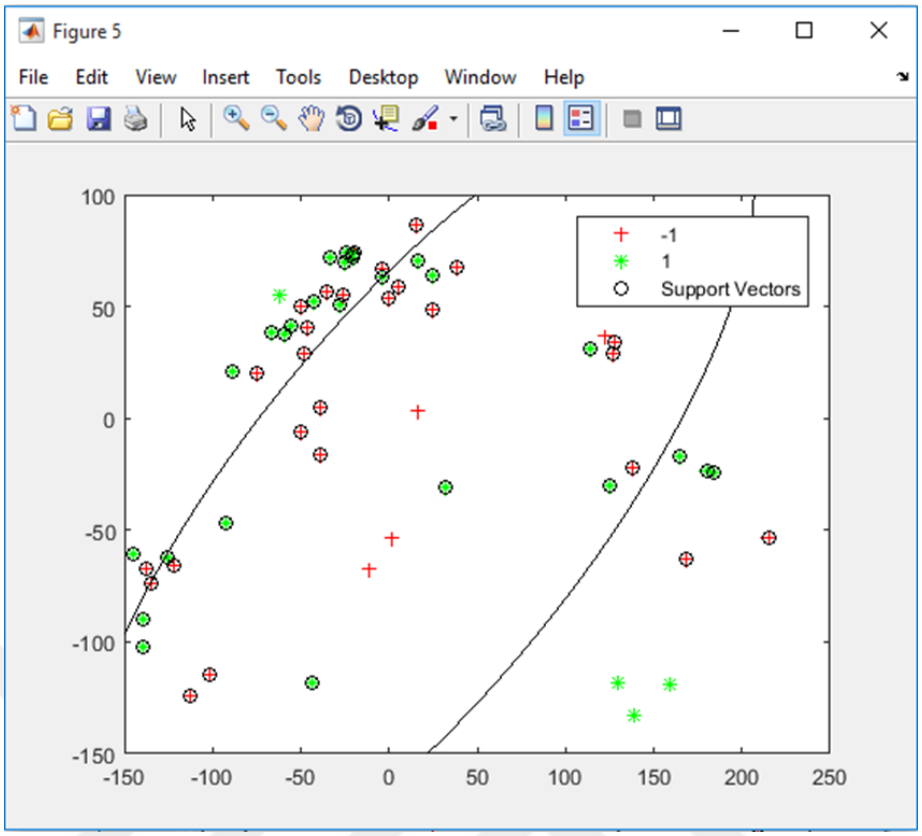


Figure (4-6): Quadratic kernel in YALE database

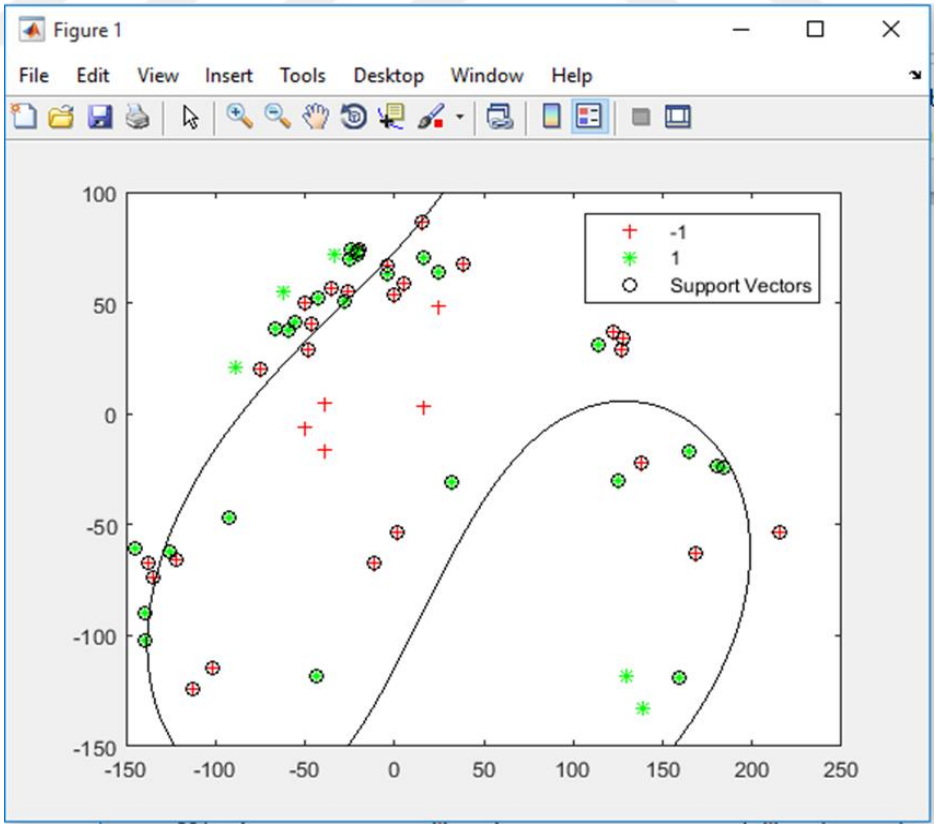


Figure (4-7): Polynomial kernel in YALE database

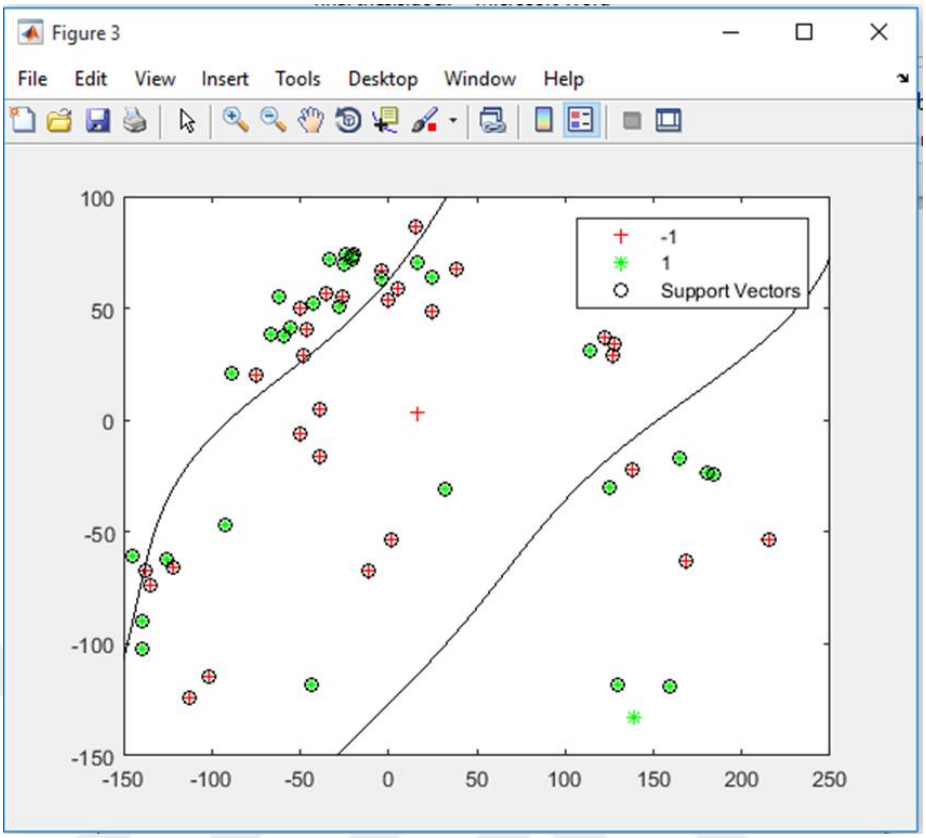


Figure (4-8): Gaussian Radial Basis Function kernel in YALE database

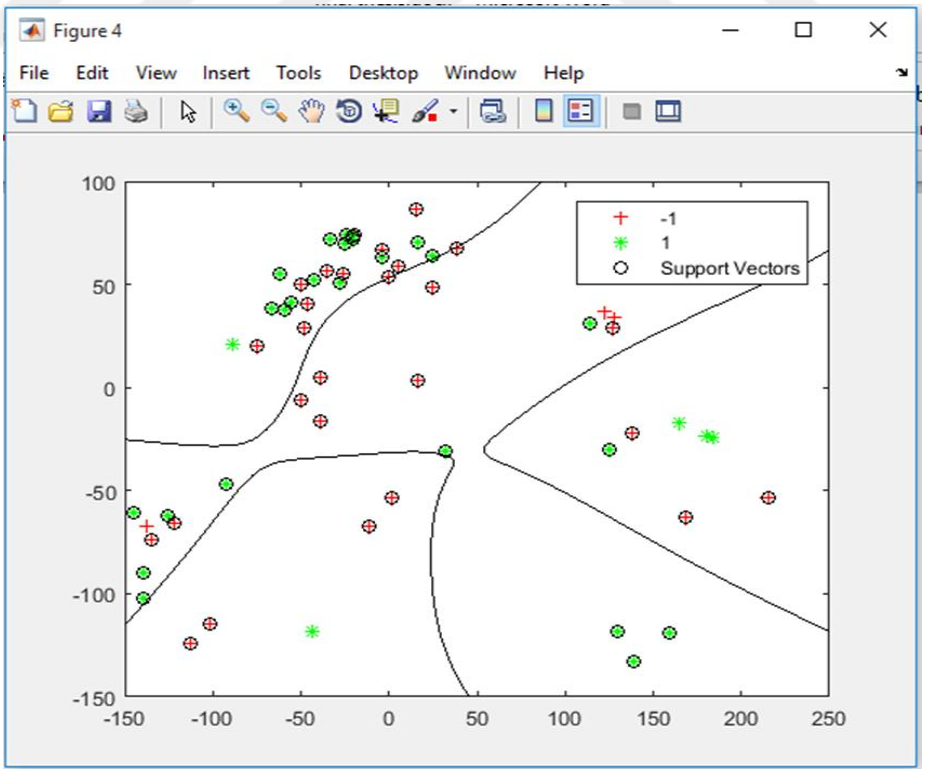


Figure (4-9): Multilayer Perceptron kernel in YALE database

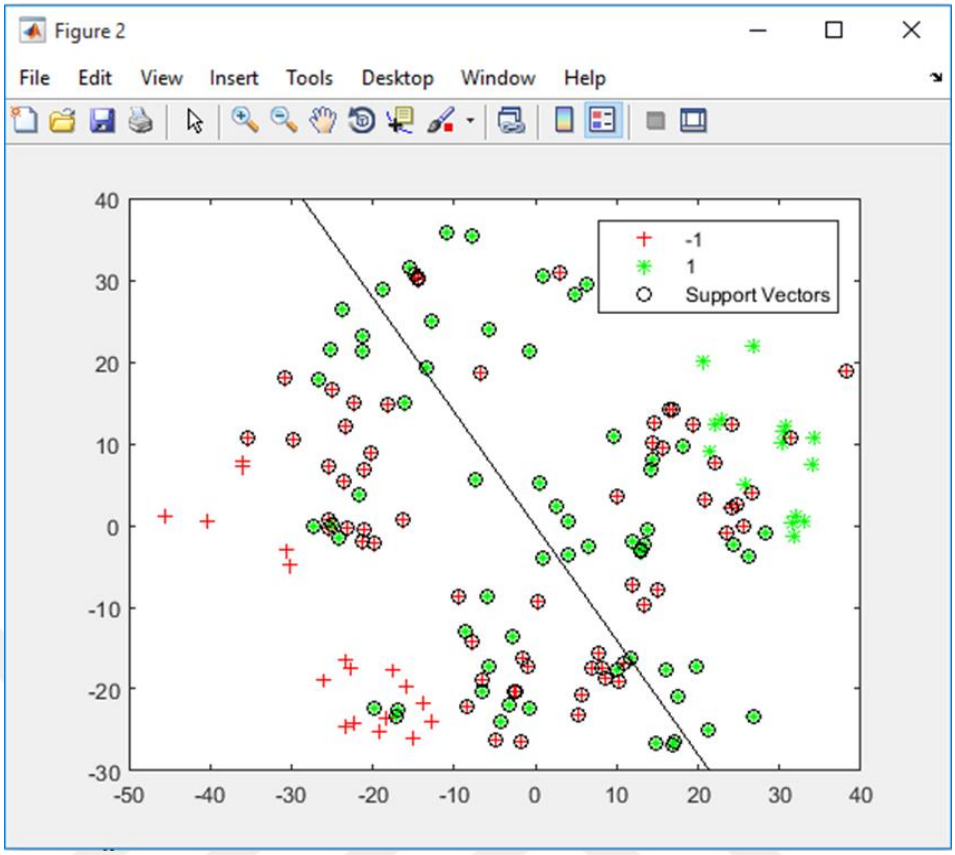


Figure (4-10): linear kernel in ORL database

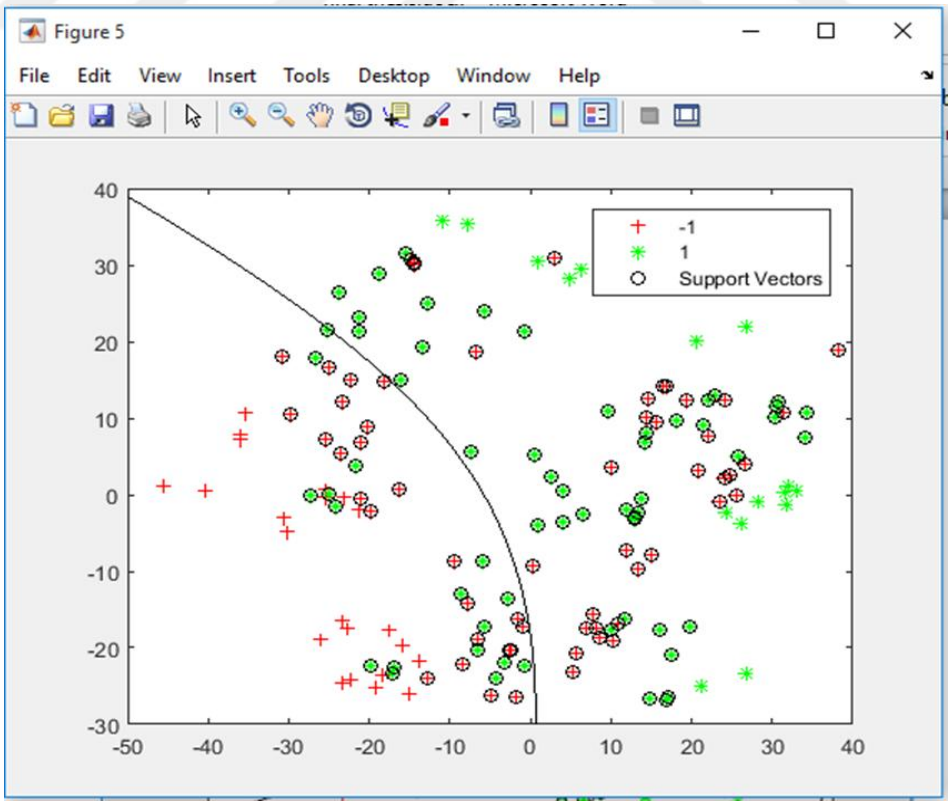


Figure (4-11): Quadratic kernel in ORL database

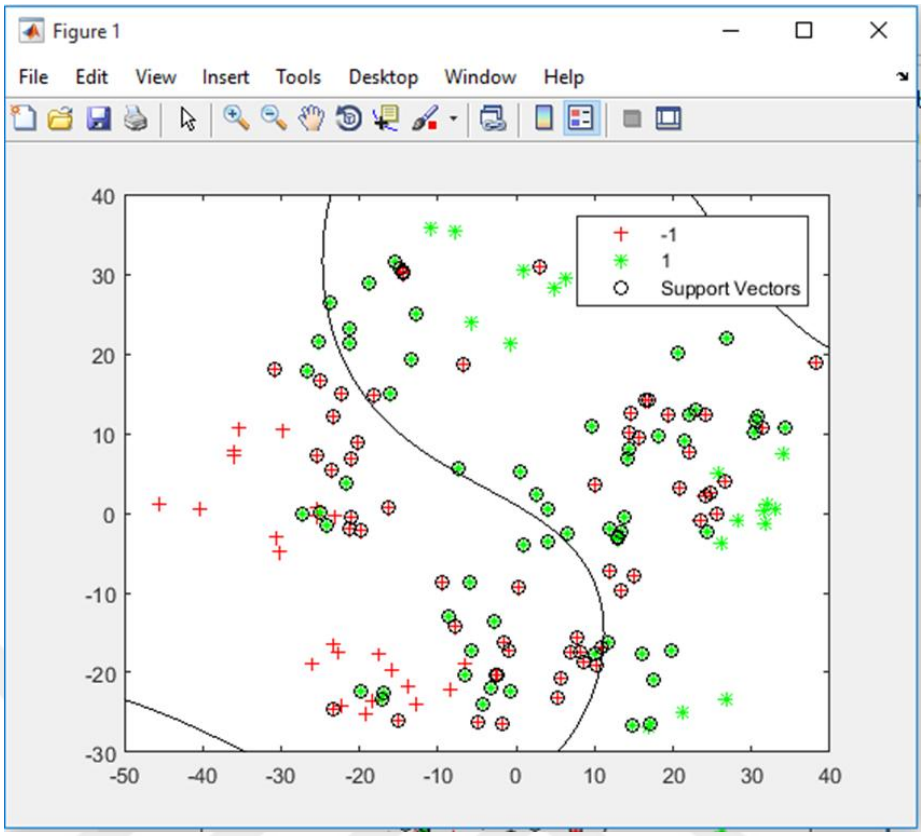


Figure (4-12): Polynomial kernel in ORL database

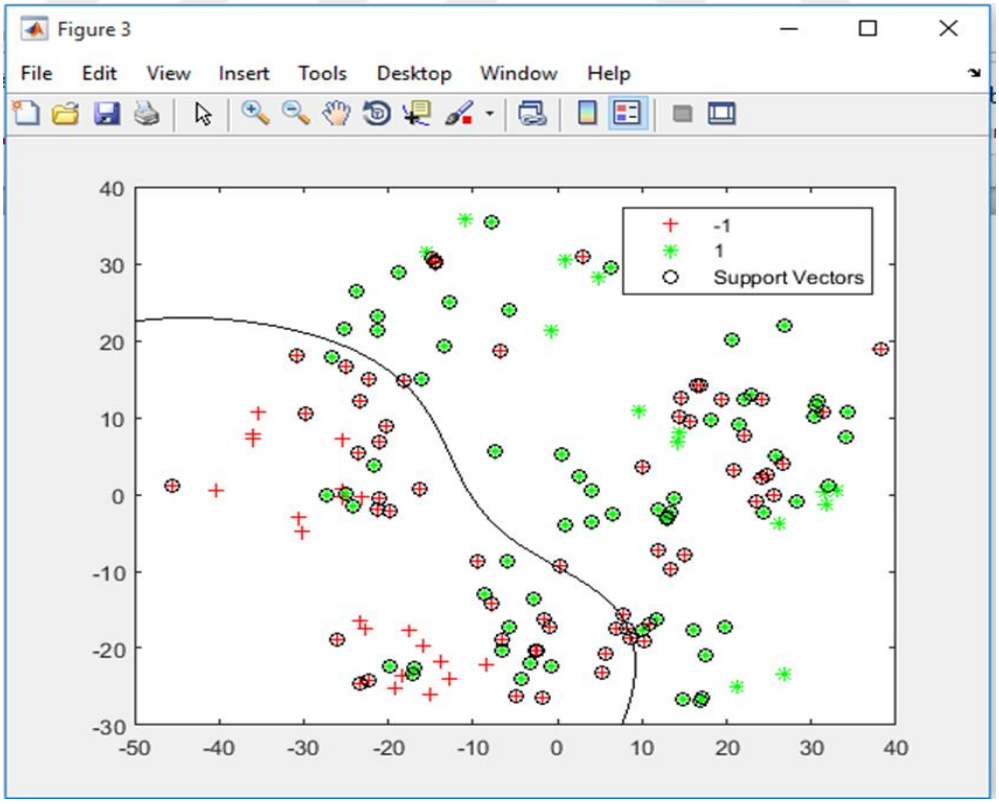


Figure (4-13): Gaussian Radial Basis Function kernel in ORL database

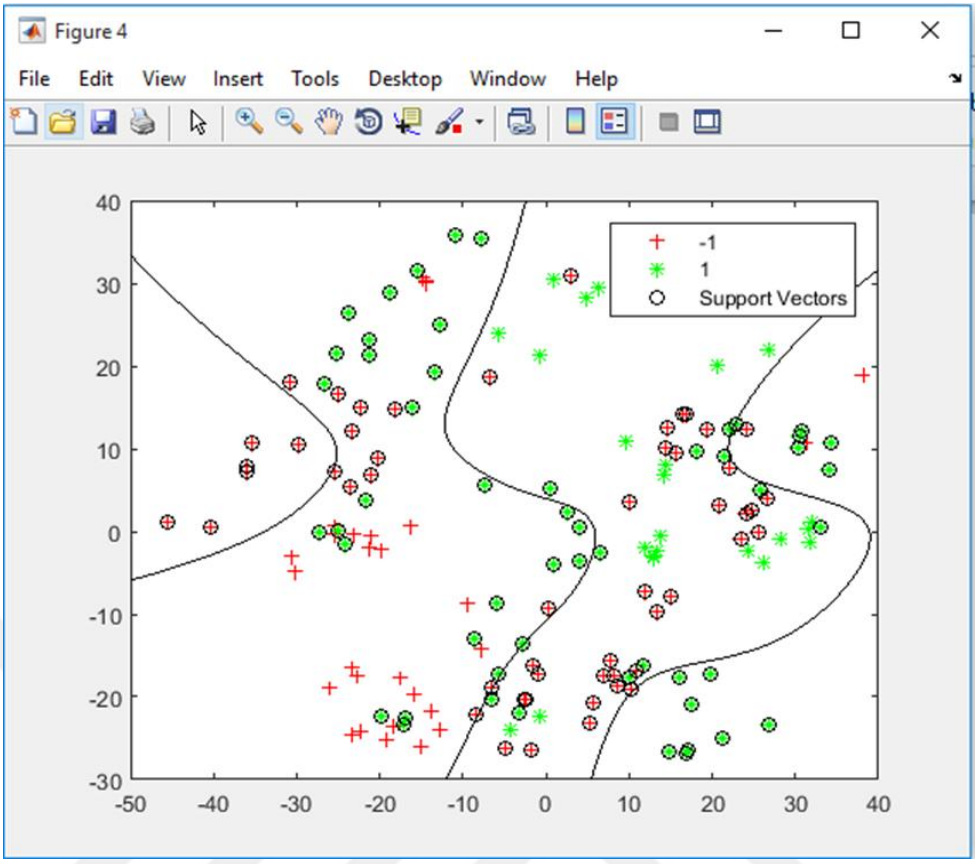


Figure (4-15): Multilayer Perceptron kernel in ORL database

4.6.2 Neural Network Classifier

The neural network is used as a classifier in order to compare its results with the results of SVM. Feed forward back propagation neural network is one of the important types of the artificial neural network. In general, the input data feed on input layer to output layer and the output result are compared to the true results are updated the result based on this comparison. The network that used consists of two layers with a sigmoid transfer function in each layer.

In generally, the two-layers network is applied for pattern recognition applications. To generate the structure of the network, the designer must determine the number of layers and also the number of neurons in the input, hidden and output layers. Then the designer must train the network by introducing the training data matrix and the target matrix generated with SVM classifier.

Training matrix consists of the features that come from the previous stage. After that, the target vectors is created and its values are related to the image of the person in the dataset. The target vectors are explained in the previous paragraph.

After the training matrix and target vector are generated, the training stage network can be implemented. Feed forward Back propagation neural network is applied to train the network. It has two stages: the feed forward and back propagation. In the feed forward stage, the data comes from input to output to obtain the result and calculate the results. Then, it goes from the output layer to input layer with updating the weight. The training performance of our network goals the errors are reached to 1e-17. This error is between calculated output and sample output data, and it is used to create an adjustment to the weights. After training the network is completed, Neural Network is ready to classify any input image.The network consists of three layers (input, hidden and output) with 21 in the input layer,10 neurons in the hidden layer and one neuron in the output layer. The neural network training performance of both Yale and ORL database are displayed in the figures (4-15) and (4-16).

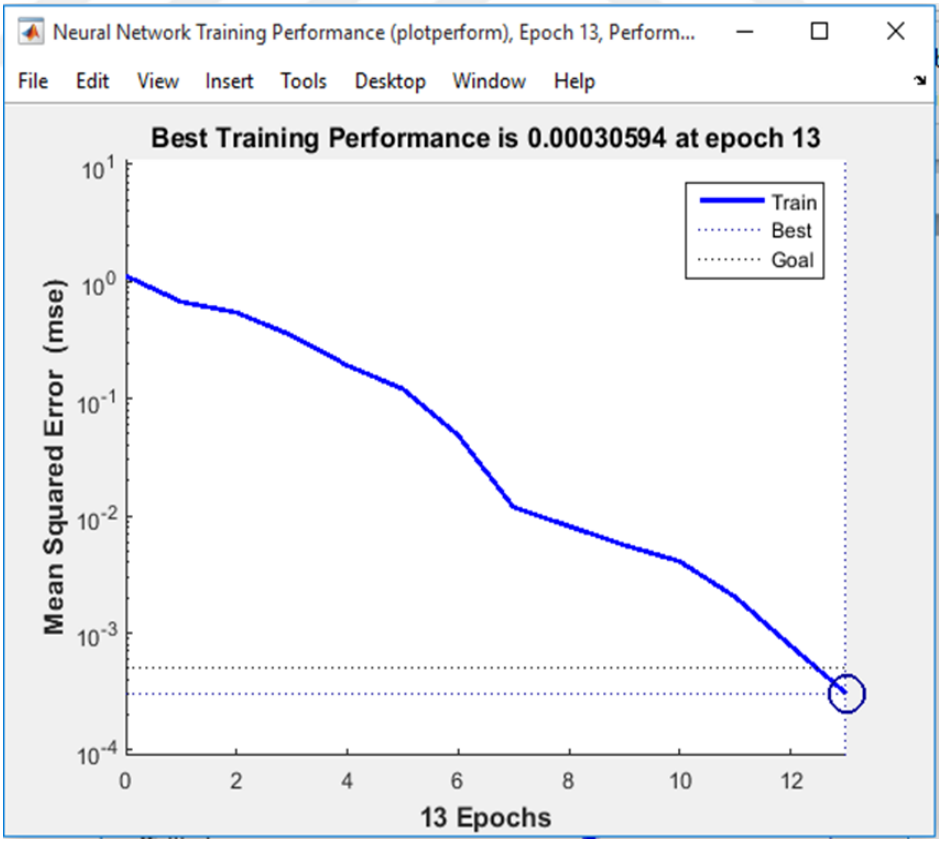


Figure (4-15):Neural network training performance yale database

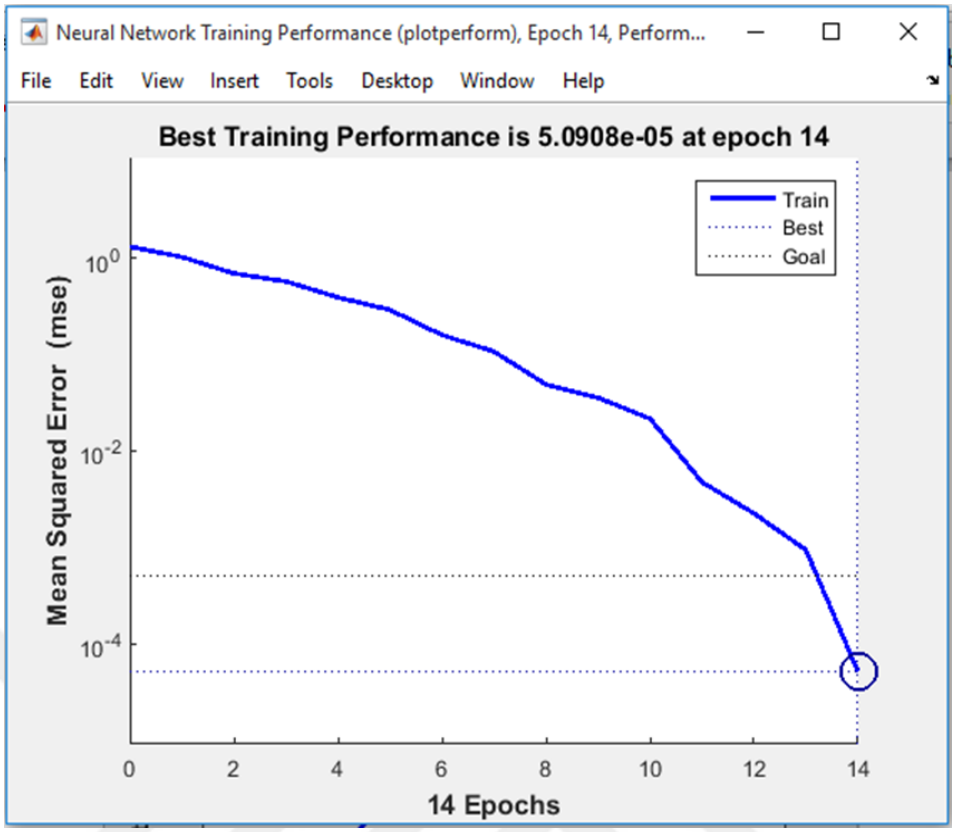


Figure (4-16): The neural network training performance of ORL database

Figure (4-17) describes the training window. In this figure all the information about the network is appeared and also the concerning the training of the network such as the network diagram, details about the algorithm, the number of the epoch, the training time,the performance of the network as well as this windows can be used to plot performance of the figure, regressing figure. Figure (4-17) and Figure (4-18) show the training window for both YALE and ORL database.

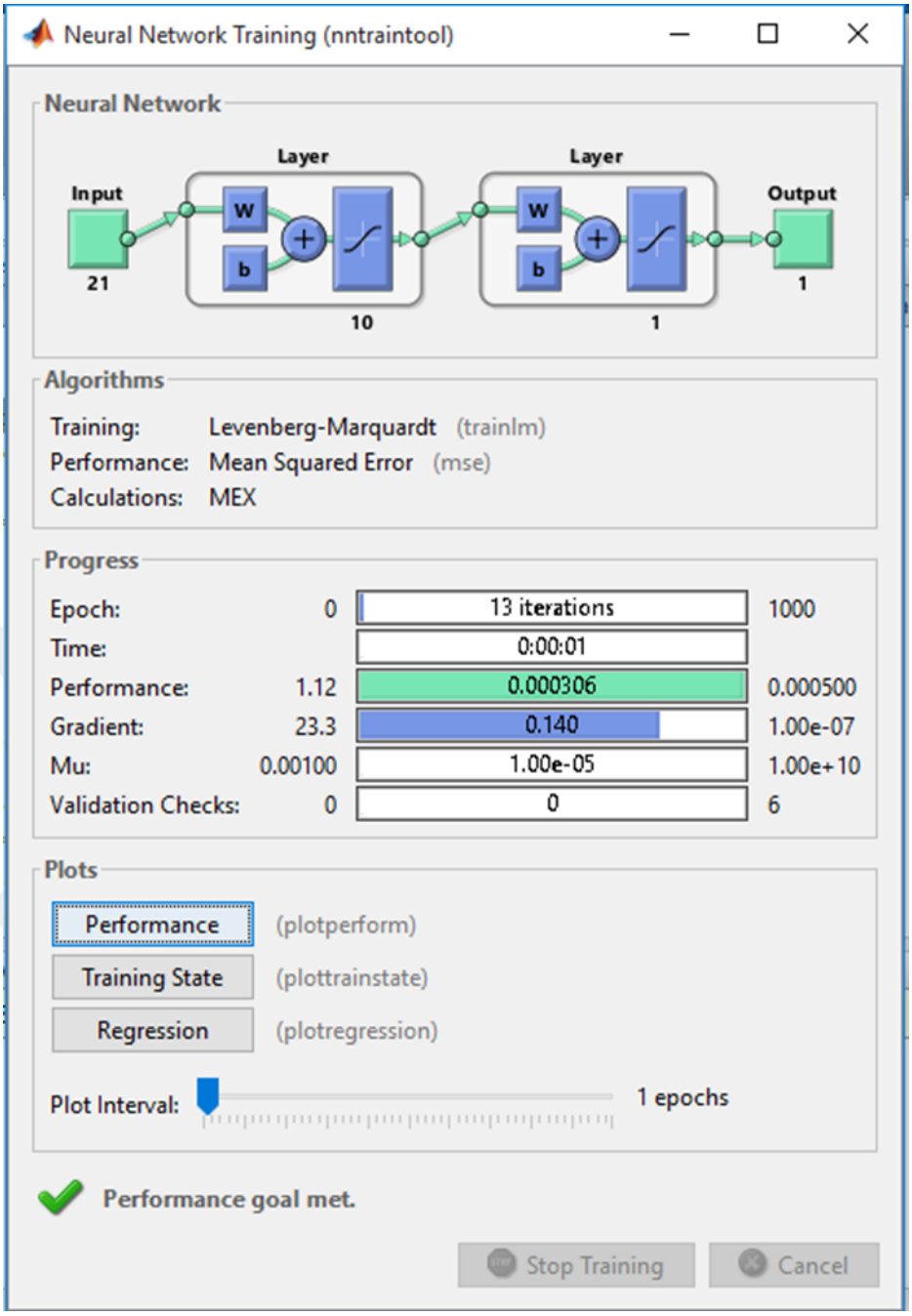


Figure (4-17): Train window of Yale database

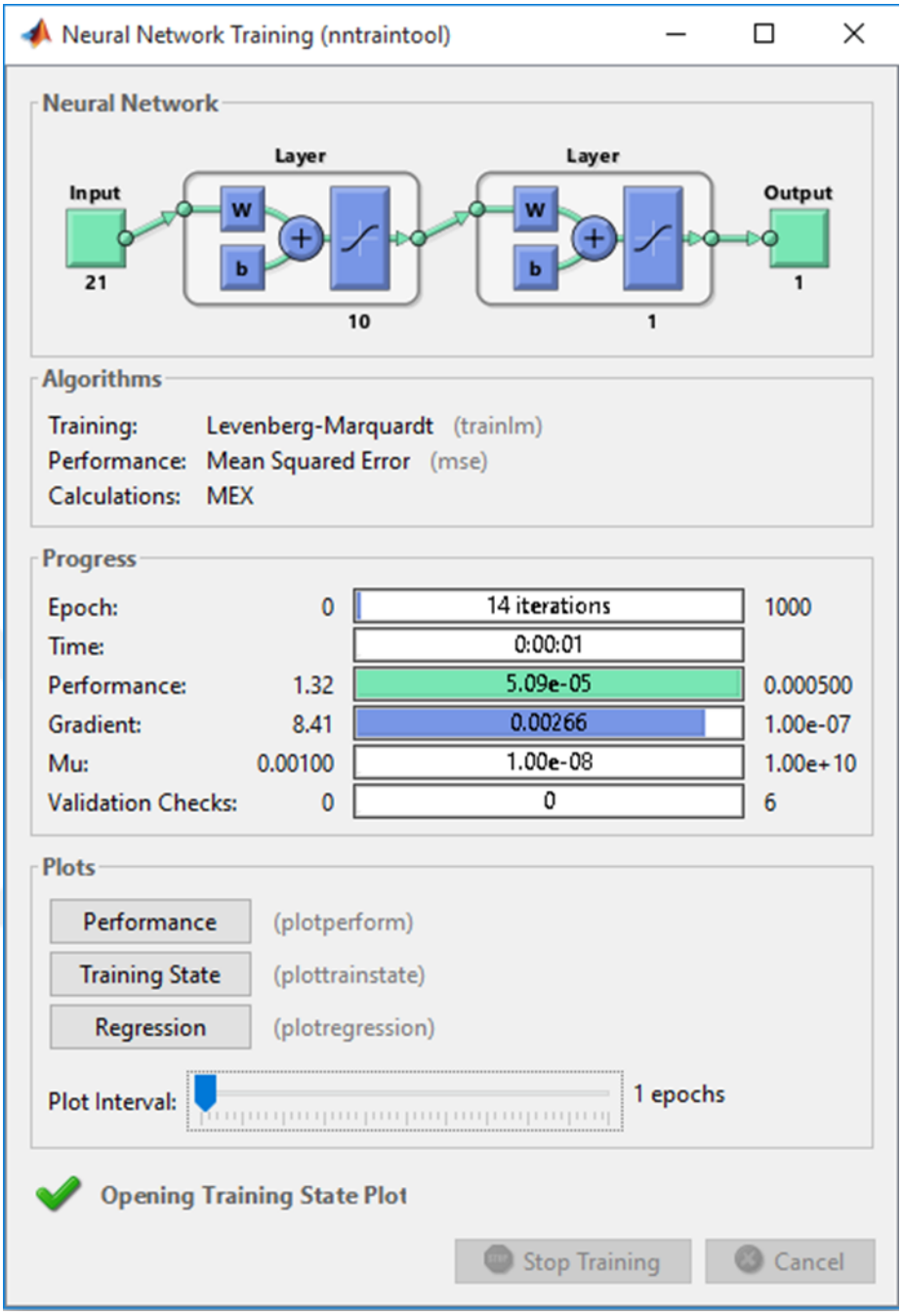


Figure (4-18) : Train window of ORL database

Figure (4-19) and Figure (4-20) describe the regression plot. This plot shows the relationship between the outputs of the network. The plots show the training, validation, and testing data. The dashed line in each plot refers to that (the perfect result – outputs = targets) while the solid line represents the best fit of the linear regression line between outputs and targets.

İn Figure (4-19), R value is an indication of the relationship between the outputs and targets. If R = 1, this indicates that there is an exact linear relationship between outputs and targets. If R is close to zero, then there is no linear relationship between outputs and targets.

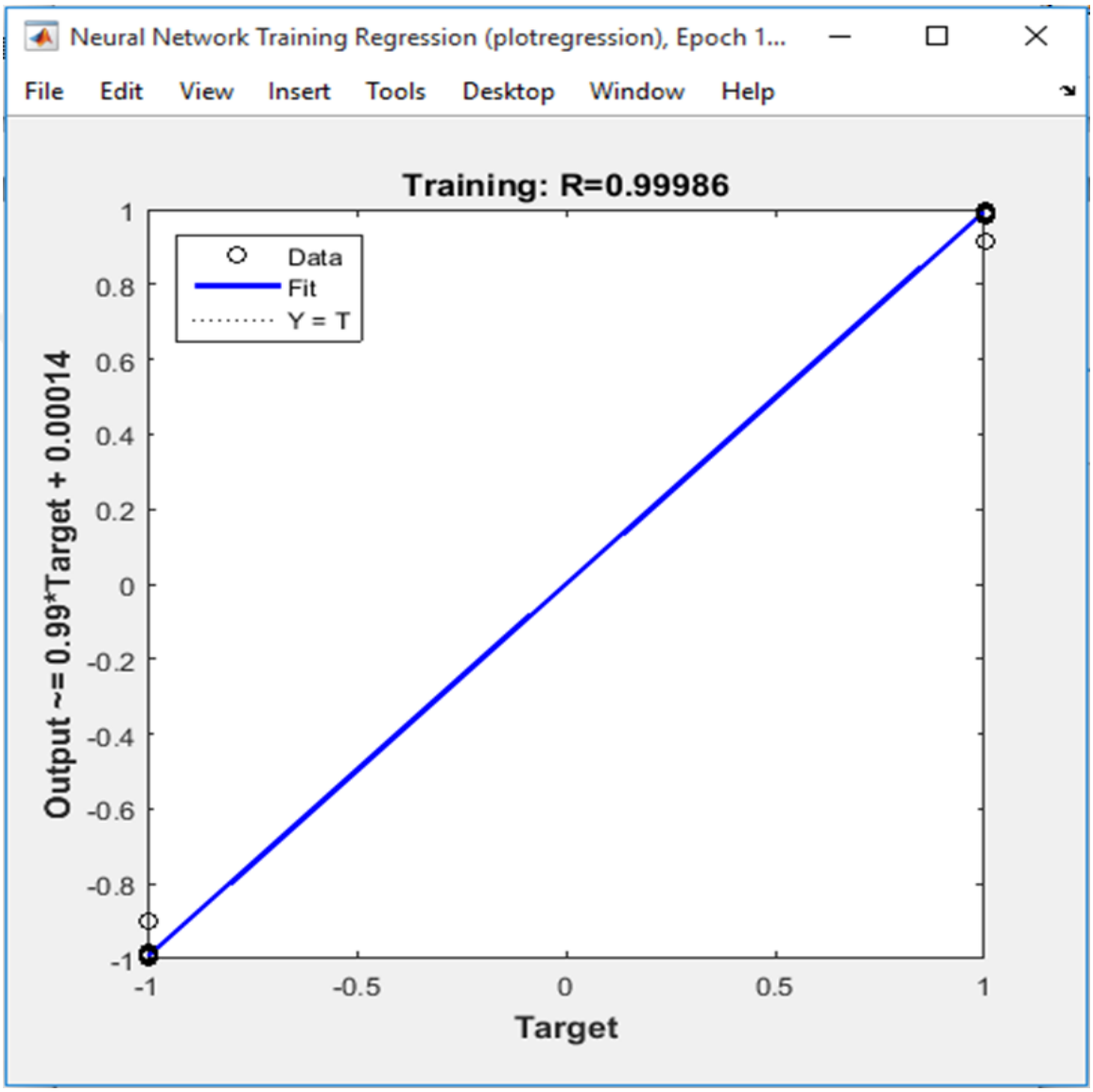


Figure (4-19): Regression plot of Yale database

From the Figure (4-19) and Figure (4-20), the training data indicates a good fit and The validation and test results also show R values are greater than 0.9. This means, there is no linear regression line between outputs and targets.

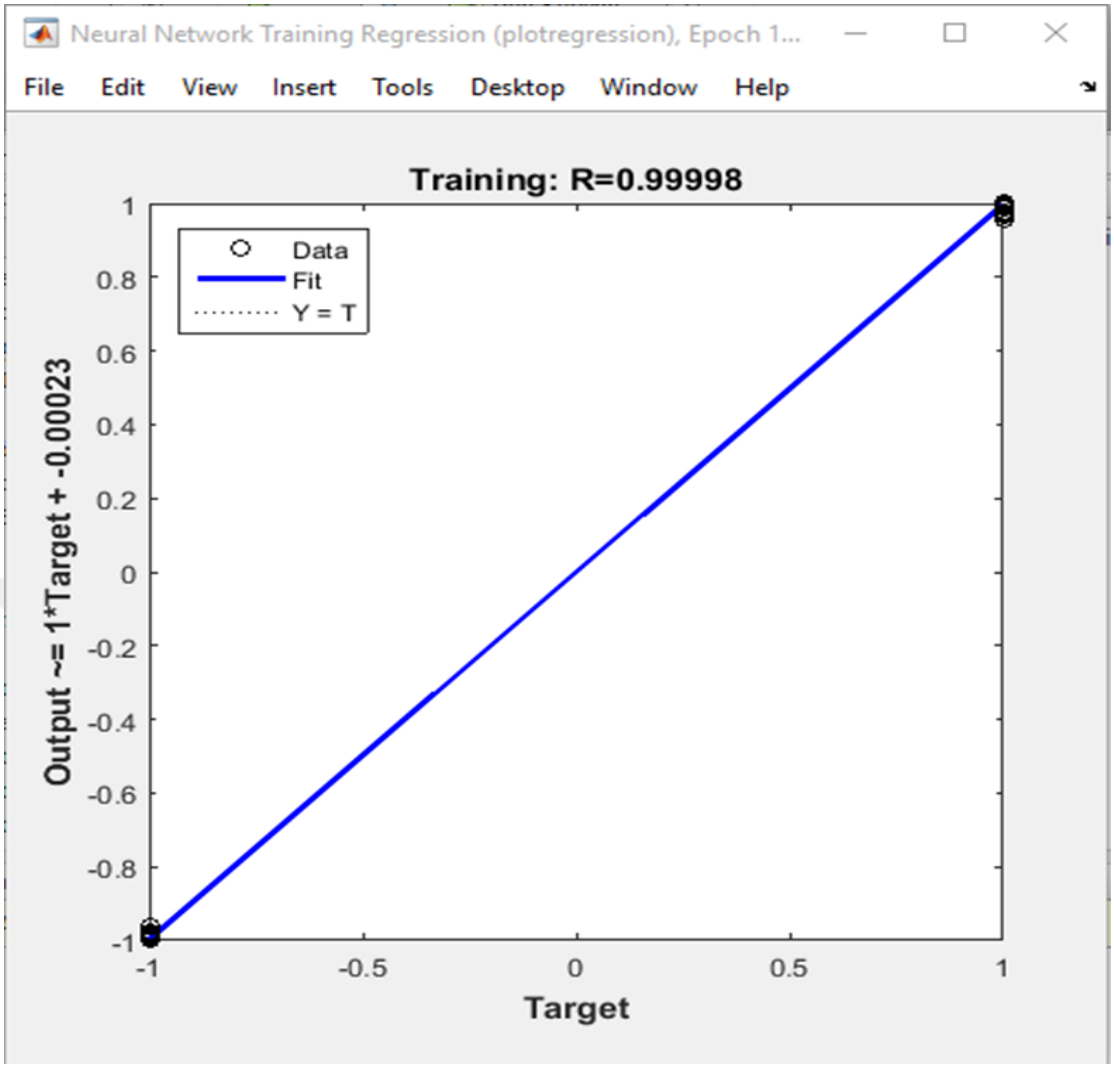
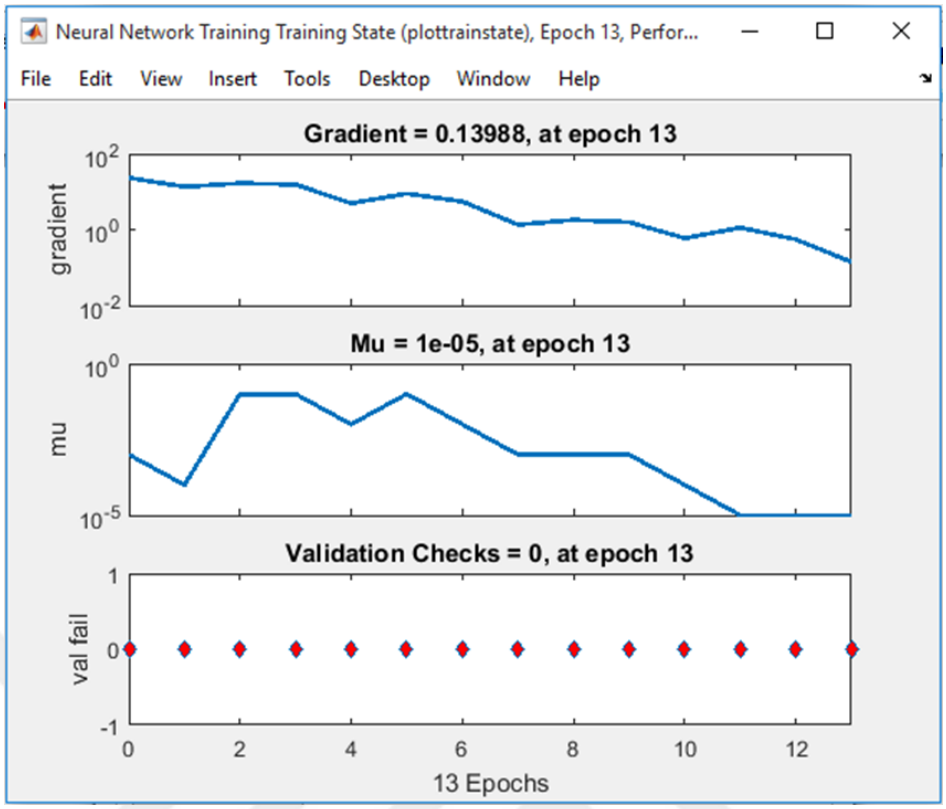
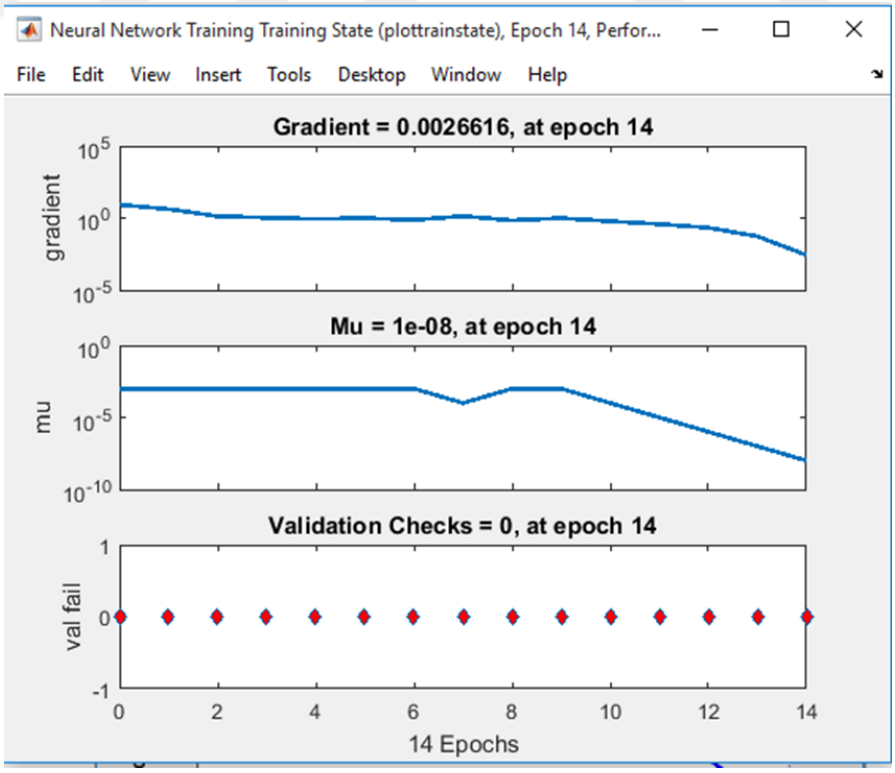


Figure (4-20): Regression plot of ORL database

Figure (4-21) and Figure (4-22) describes variation in gradient coefficient with respect to a number of epochs. The final value of gradient coefficient at epoch number 5000 is 0.0028313 which is approximate near to zero. Minimum the value of gradient coefficient better will be training and testing of networks. Also from Figure (4-21) and Figure (4-22) can be seen that gradient value goes on decreasing with increase in a number of epochs while the third graph represents the number of validations checks in each epoch.



Figure(4-21): Neural Network Training State Plot of YALE database



Figure(4-22): Neural Network Training State Plot of ORL database

4.7 Identification

The final stage of the system is identification. The aim of this stage is to find the nearest image for the new image that classified before using SVM or ANN. K-Nearest Neighbor is used to implementing that. KNN is not like SVM or ANN. it does not need to the training phase and therefore all the data is needed for the testing phase.

The data required for testing consists of a set of vectors and class labels associated with each vector only. In this system, the number of neighbors (k) is set to one.

4.8 Experimental Results

A number of experiments on the two different databases are used. The first experiment is applied on the Yale database and the second experiment is applied on the ORL database. In the every database, two experiments are done. First experiment is compared between different SVMs kernels to show any kernel to give best results and the second experiment is between SVM classifier and neural network classifier to calculate the performance of the classifier that used the system.

4.8.1 Yale Experiment

Yale database is used in this experiment. This database contains fifteen subjects and each subject has eleven images taken under various condition. This means that the database consists of 165 images.

Figure (4-23): Samples of Yale Database [13].

This database is applied because it consists of a number of images captured at a various condition this includes lighting condition, with glasses, without glasses, happy, sad, normal and sleepy, therefore it can be applied to test the recognition rate of the proposed system and it is also very widely used by many of researchers.

In our experiment, the database is divided into two groups, the first group is for training and the second group is for the testing. The training group is applied to learn the system and the training group consists of five images of the first 12 classes. The testing group is applied to calculate recognition rate of the proposed system and it consists of the other five images of the same 12 classes as well as all images of the last three class. This means that 40% of the database is applied for training and 60% of the image is applied for the test. Figure (4-25) describe the results of the SVMs classifier using various kernel function in YALE database.

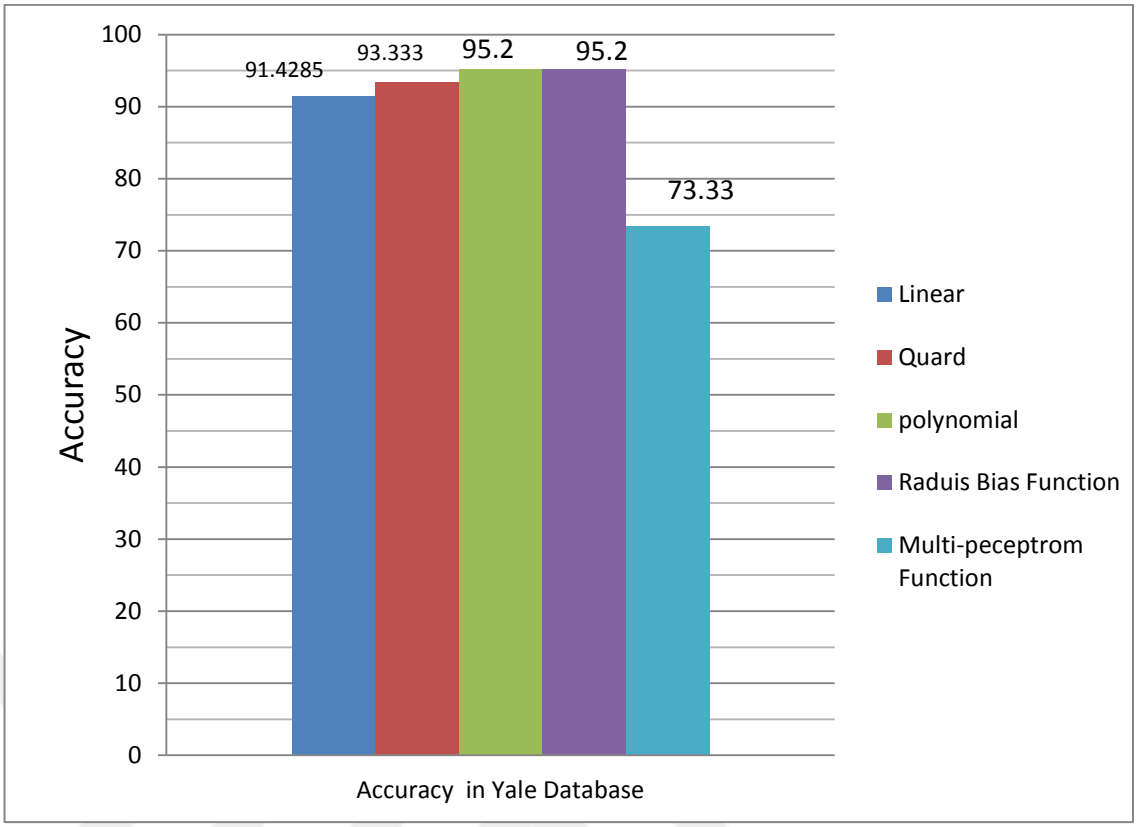


Figure (4-24): The results of the SVMs classifier using different kernel function in YALE database.

In The second experiment, FFBPNN is applied as classifier instead of the SVMs. This is done to make a comparison between the results of SVMs and ANNs classifiers as well as to calculate the performance of the system. From experiment results, we can note that the highest accuracy rate is obtained by using polynomial function and radius bias function while the lowest one is obtained by using the multi-perceptron function. Figure(4-26) describe the recognition rate results of the SVMs classifier using polynomial kernel function and artificial neural network.

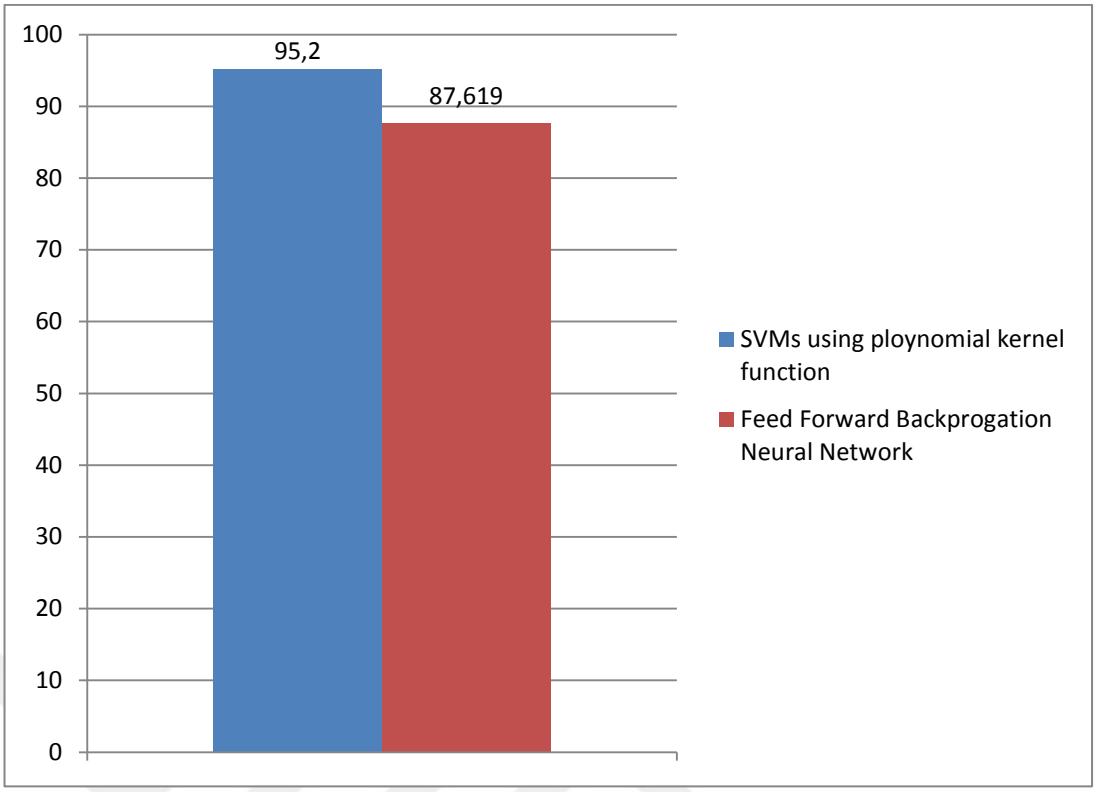


Figure (4-25): The accuracy results of the SVMs classifier using polynomial kernel function and artificial neural network in Yale database

4.8.2 ORL Experiment

ORL database is used in the second experiment. This database consists of 40 classes with ten images for each class. This mean database consists of 400 images.

In our experiment, the database is divided into two groups, the first group is for training and the second group is for the testing. The training group is applied to learn the system and the training group consists of five images of the first 32 classes. The testing group is applied to calculate recognition rate of the proposed system and it consists of the other six images of the same 32 classes as well as all images of the last eight class. These three class are used for testing only. This means that 40% of the database is applied for training and 60% of the image is applied for the test. Figure (4-27) describe the results of the SVMs classifier using different kernel function in ORL database. From the experiment results, we can see that the highest accuracy rate is obtained by using polynomial function while the lowest one is obtained by using multi-perceptron function.

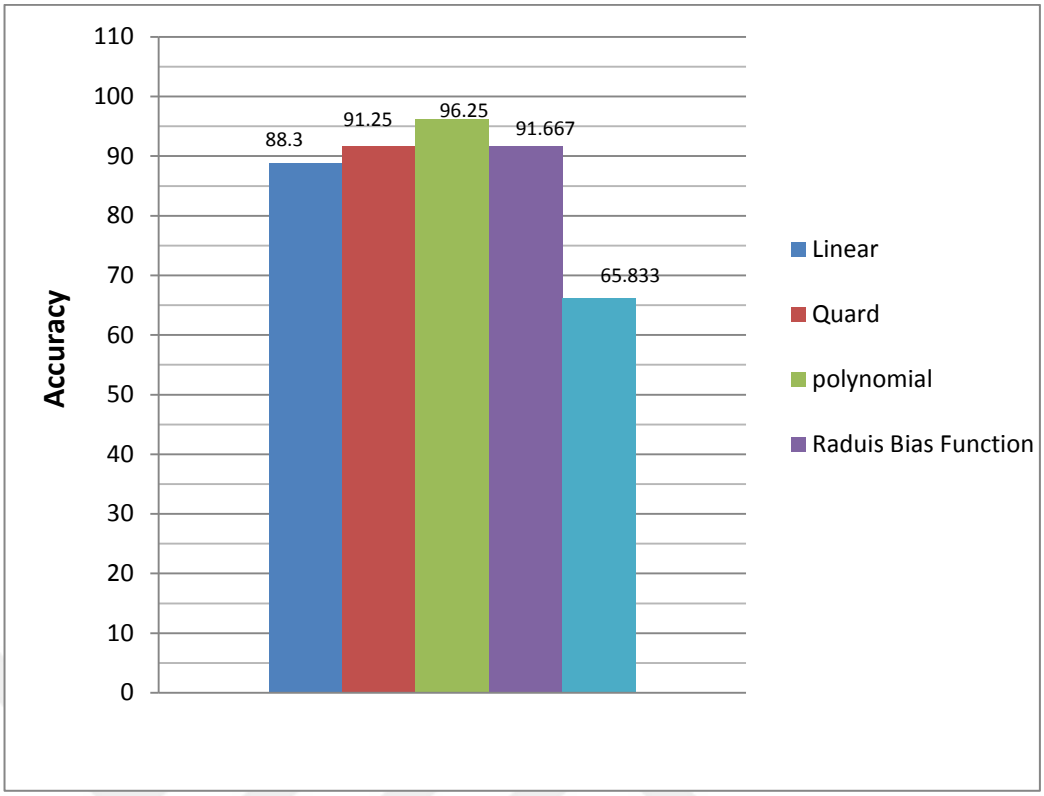


Figure (4-26): The results of the SVMs classifier using different kernel function in ORL database.

In The second experiment, FFBPNN is applied as classifier instead of the SVMs.This is done to make a comparison between the results of SVMs and ANNs classifiers as well as to calculate the performance of the system. Figure(4-28) describe the recognition rate results of the SVMs classifier using polynomial kernel function and feedforward backpropagation neural network classifier. From the figure (4-28), we can note that the results of SVMs are better from ANN classifier.

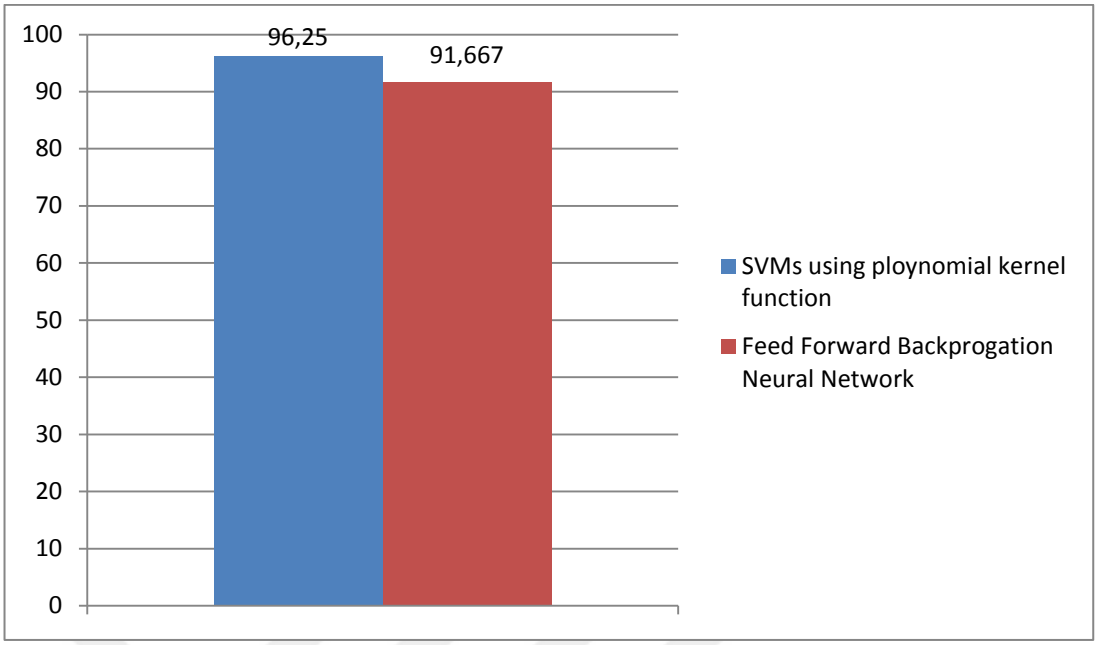


Figure (4-27): The accuracy results of the SVMs classifier using polynomial kernel function and artificial neural network

In addition to all that, Comparison between our proposed system and PCA-LDA and SVM [29] on the Yale database is implemented. Table (4-1) shows the comparison between them. As well as to this table, there are other two main differences. The first one is that our system has identification stage using KNN but PCA-LDA and SVM [29] used classification stage only without identification stage. The second difference is that our proposed system use PCA method and other system used PCA-LDA method.

Table (4-1): Comparison Between differences proposed method

Methods	Accuracy
PCA-Wavelet and SVM	95.2
Wavelet PCA-LDA and SVM [29]	97.5

While table (4-2) shows the comparison between our proposed system based on ANN classifier (BPNN classifier) and PCA-BPNN [30] methods on the Yale database.

Table (4-2): shows the comparison between PCA-Wavelet, ANN classifier, and PCA-BPNN

Methods	Accuracy
PCA-Wavelet and ANN	87.619
PCA-BPNN [30]	85.7

4.9 Graphics User Interface

Graphics User Interface is also known as GUI. This GUI is designed to help the end user to use this system. This GUI consists of five pushbuttons: (i). Select Image, (ii). Classify using SVM, (iii). Classify using ANN, (iv). Help, and (v). Exit. Figure (4-29) show the GUI designed of this system. The details of this pushbutton are as the following:

To illustrate the present work:

Select Image pushbutton: It is applied to select an image from any folder on the computer.

Classify using SVM pushbutton: It is applied to classify the selective image based on support vector machine classifier.

Classify using ANN pushbutton: It is used to classify the selective image using artificial neural network classifier.

Help pushbutton: It is used to give support to the new user.

Exit pushbutton: It is used to close the window of GUI.

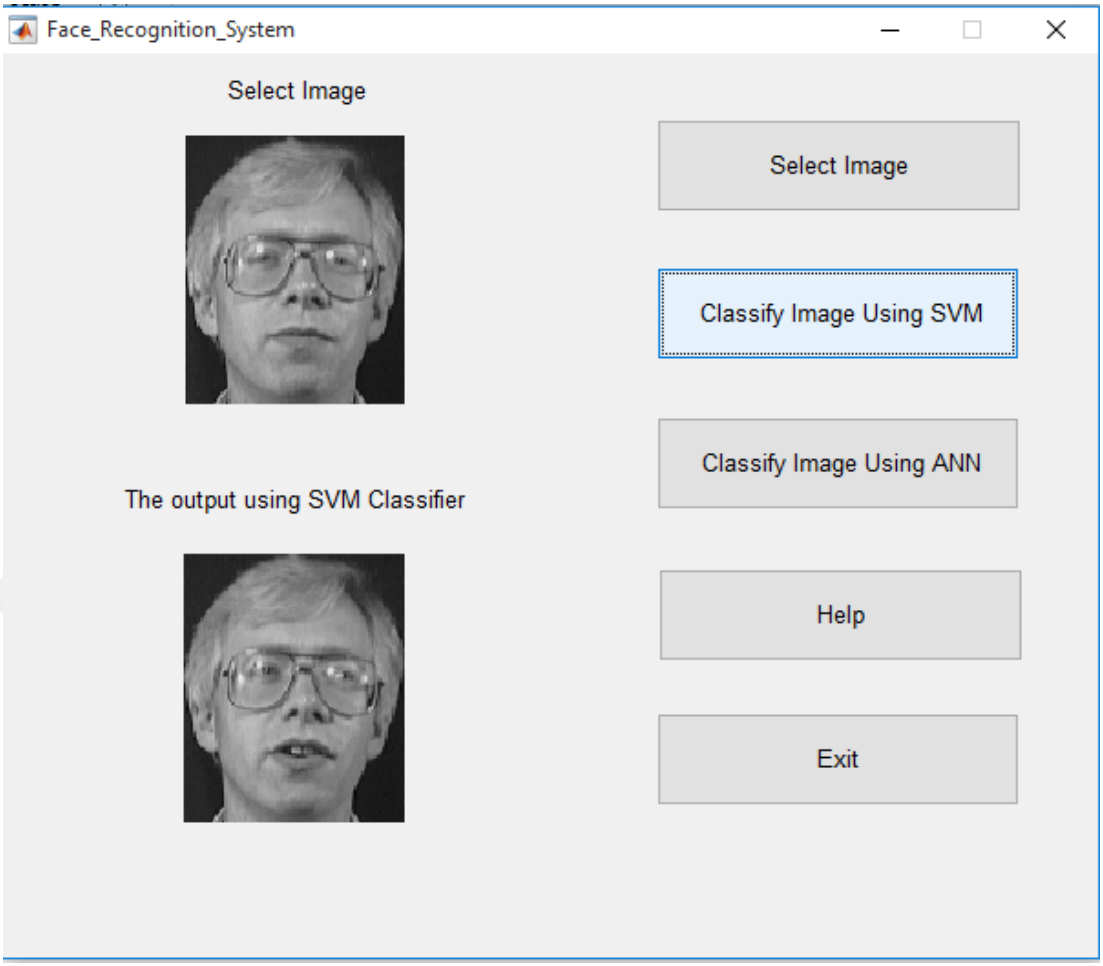


Figure (4-28): The Face Recognition GUI

4.10 Summary and Discussion

The present recognition procedure consists of four main steps; preprocessing, feature extraction, classification, and identification.

i. Preprocessing performs image enhancement part, which composes of white balance correction and reduces the effect of noise on the face image.

ii. Feature extraction technique is very significant to obtain features from the standard face image. Feature extraction part is composed of two methods principle component analysis and wavelet transform. These two methods extract most import feature and without these methods, the recognition system becomes very difficult and impossible execute.

iii. The classifier is considered as a key point of recognition part. The output of classification stage is applied to classify the features that taken from image. Two comparisons are executed. First one is between different kernels of SVMs to choose any one of them is best. The comparison leading to the conclusion that polynomial kernel function gives the best results compare comparing to other kernel SVMs function The second one is between SVMs and feed forward neural network is used to see who of them is suitable for classification and the results prove that SVMs is more suitable than ANN.

iv. Finally, K-nearest neighborhood is used to identify the person.

Algorithm is developed on MATLAB environment. The final results come persons image is achieved from the database then the person is recognition. if the image is not found in the database, the output of the system is “Person is not recognized”. The system has been found to produce results and the performance of the system is agreeable with some of the results in the literature.

CHAPTER 5

CONCLUSIONS AND FUTURE WORKS

5.1 Conclusion

Face recognition systems are one of computer vision application which applied in many security application such as avoiding crime, monitoring through video and also for verification of the identity of persons.

This work introduces a description of the design and implementation of face recognition system. In this system combines of wavelet transform with PCA method is applied for obtained feature from the image and SVM is applied to classify these features and KNN to identify the input face image. SVM is one of the new methods and uses it open the way to compare the results of proposed system with other types of a classifier that can be used such as neural networks.

The comparison between two types of classifiers proved that SVM classifier is more accurate than FFBNN classifier in recognition. The proposed system is tested by using two types of databases to guarantee from the power of the system.

According to the results, the following conclusions can be stated:

- Applying Wavelet- PCA feature extraction algorithms enhanced the recognition rate of the system compared to the using PCA only.
- The experimental results proved the efficiency and reliability of the system and the results enhancement of 5% by using the SVM classifier with polynomial Kernel Function compared to use feed forward Backpropagation neural network classifier.
- Using KNN in identification stage give the system the ability to identify the person in a fast and accurate way.

5.2 Suggestions of Future Works

A number of future works that can be applied to improvements this system can be summarized as:

1. Using other methods to obtain the features such as independent component analysis.
2. Adding feature selection methods as genetics algorithm method to select feature after feature extraction stage.
3. Implement the proposed system as a real-time system by adding the camera as acquisition part.
4. Develop the system to work with three-dimension images.

REFERENCES

1 Drahanský Martin. Fingerprint Recognition Technology - Related Topics. Saarbrücken: Lambert Academic Publishing, 2011.

2 Eastwood Shawn C., Shmerko Vlad. P., Yanushkevich Svetlana, Drahanský Martin and Gorodnichy Dmitry," Biometric-Enabled Authentication Machines: A Survey of Open-Set Real-World Applications", IEEE Transactions on Human-Machine Systems, Washington Institute of Electrical and Electronics Engineers, vol. 46, no. 2, pp. 2168-2291. 2016.

3 Antonio Rama Calvo, Francesc Tarrés Ruiz, Jürgen Rurainsky and Peter Eisert,"2D-3D Mixed Face Recognition Schemes", Recent Advances in Face Recognition, Kresimir Delac, InTech,2008.

4 Masoud Mazloom and Shohreh Kasaei,"Face Recognition using Wavelet, PCA, and Neural Networks",Proceeding of the First International Conference on Modeling, Simulation and Applied Optimization,February, 2005.

5 Yen-Shao Chen, Ming-Yuan Shieh and Guei-Hua Chen," Design and Implementation of A Real-Time Face Recognition System for Mobile Robot", Journal of Communications and Information Technology, Vol.1, Nov., 2008, pp. 89-92.

6 Sheifali Gupta, O.P.Sahu, Rupesh Gupta and Ajay Goel, "A Bespoke Approach for Face-Recognition Using PCA", International Journal of Computer Science and Engineering, Vol. 02, No. 02, 2010,pp.155-158.

7 Lei Z., Ahonen T., PietikainenM.andStan Z,"Local Frequency Descriptor for Low-Resolution Face Recognition",2011 IEEE International Conference on Automatic Face & Gesture Recognition", March, 2011,pp.1-6.

8 Kaur S. andSinghVirk R., "DCT Based Fast Face Recognition Using PCAand ANN",International Journal of Advanced Research in Computer and Communication Engineering,Vol. 2, Issue 5, May, 2013.

9 Javed a., "Face Recognition Based on Principal Component Analysis",I.J. Image, Graphics and Signal Processing,Vol.2 ,2013,pp.38-44.

10 Oommen A., C.Senthil Singh and M.Manikandan,"International Journal of Research in Engineering and Technology",Vol.: 03, Issue 01,2014.

11 Rajath Kumar M. P., Keerthi Sravan R. and K. M. Aishwarya," Artificial neural networks for face recognition using PCA and BPNN",TENCON 2015 - 2015 IEEE Region 10 Conference, 1-4 Nov. 2015.

12 Can Eyupoglu," Implementation of color face recognition using PCA and k-NN classifier",NW Russia Young Researchers in Electrical and Electronic Engineering Conference (EIConRusNW), 2016 IEEE, 2-3 Feb. 2016.

13 http://cvc.cs.yale.edu/cvc/projects/yalefaces/yalefaces.html

14 http://www.cl.cam.ac.uk/Research/DTG/attarchive:pub/data/att_faces.zip

15 Michael David, (2002),"Reversible Integer – To Integer Wavelet Transforms for Image Coding", Thesis, University of British Columbia.

16 Alexander D. Poularikas,(2000)," The Transforms and Applications Handbook", Second Edition, CRC Press.

17 Kociołek, A. Materka, M. Strzelecki P. Szczypiński," Discrete wavelet transform derived features for digital image texture analysis", International Conference on Signals and Electronic Systems, pp. 163-168, Sep., 2001.

18 Zhang David," Biometric Image Discrimination Technologies" Idea Group Publishing, 2006.

19 M.N.Shah., Radi H.R., S.Muniroh., "Face Recognition Using Principle Discernment Analysis and Linear Discernment Analysis", International Journal of Electrical & Computer Sciences ,Vol.12 No.05,Oct.,2012.

20 Jasneet Singh Taneja, (2006)" Analysis of E. coli Promoters Using Support Vector Machine, Thesis, Thapar Institute of Engineering & Technology.

21 Pascal Paysan, (2004)" Stereovision based Vehicle Classification Using Support Vector Machines" Thesis, Massachusetts Institute of Technology (MIT).

22 Martin Michael schulz, (2002)"Facial Expression Recognition with Support Vector Machine", Thesis, University of Applied Science Konstanz.

23 Zhang David," Biometric Image Discrimination Technologies" Idea Group Publishing, 2006.

24 Saeed B. Niku,(2011) "Introduction to Robotics", Second Edition, John Wiley & San, Inc.

25 Michael Negnevitsky,(2005),” Artificial Intelligence A Guide to Intelligent Systems”, Second Edition, Pearson Education Limited

26 Mirza Cilimkovic, "Neural Networks and BackPropagation Algorithm", Institute of TechnologyBlanchardstown,Dublin,Ireland.

27 Xindong Wu · Vipin Kumar · J. Ross Quinlan · Joydeep Ghosh ,"·Top 10 algorithms in data mining",Knowledge and Information Systems, Volume 14, Issue 1,2008, pp 1–37.

28 Richard O. Duda, Peter E. Hart, David G. Stork, Pattern Classification, 2nd Edition, Wiley, 2000.

29 Dr. Thair A. Salh and Mustafa Z. Nayef, " Face Recognition System based on Wavelet, PCA-LDA and SVM", Computer Engineering and Intelligent Systems, Vol.4, No.3, 2013.

30 P.Latha, Dr.L.Ganesan & Dr.S.Annadurai," Face Recognition using Neural Networks", Signal Processing: An International Journal (SPIJ) Volume (3),Issue (5).