STATUS OF THESIS

Title of thesis	FACE CLASSIFICATION FOR AUTHENTICATION APPROACH BY USING WAVELET TRANSFORM AND STATISTICAL FEATURES SELECTION METHOD

I NADIR NOURAIN DAWOUD JADALAH

hereby allow my thesis to be placed at the Information Resource Center (IRC) of Universiti Teknologi PETRONAS (UTP) with the following conditions:

1. The thesis becomes the property of UTP.

2. The IRC of UTP may make copies of the thesis for academic purposes only.

3. This thesis is classified as

☐ Confidential

☑ Non-confidential

If this thesis is confidential, please state the reason:

____

The contents of the thesis will remain confidential for ____________ years.

Remarks on disclosure:

____

	Endorsed by
____________________	____________________
Signature of Author	Signature of Supervisor
Permanent Address:	
University of Science & Technology	Dr. Brahim Belhauari Samir
Faculty of Electrical an Electronic Engineering	
P.O. Box. 30 Omdurman, Sudan	
Date:____________________	Date:____________________

UNIVERSITI TEKNOLOGI PETRONAS

FACE CLASSIFICATION FOR AUTHENTICATION APPROACH BY USING WAVELET TRANSFORM AND STATISTICAL FEATURES SELECTION METHOD

By

NADIR NOURAIN DAWOUD JADALAH

The undersigned certify that they have read, and recommend to the Postgraduate Studies Programme for acceptance this thesis for the fulfillment of the requirements for the degree stated.

Signature: ______________________________

Main Supervisor: Dr. Brahim Belhaouari Samir

Signature: ______________________________

Co-Supervisor: Assoc. Prof. Dr. Josefina Janier

Signature: ______________________________

Co-Supervisor: Assoc. Prof. Dr. Vijanth Sagayan

Signature: ______________________________

Head of Department: Assoc. Prof. Dr.Nor Hisham Bin Hamid

Date: ______________________________

FACE CLASSIFICATION FOR AUTHENTICATION APPROACH BY USING WAVELET TRANSFORM AND STATISTICAL FEATURES SELECTION METHOD

By

NADIR NOURAIN DAWOUD JADALAH

A Thesis

Submitted to the Postgraduate Studies Programme
as a Requirement for the Degree of

MASTER OF SCIENCE

ELECTRICAL AND ELECTRONIC ENGINEERING DEPARTMENT

UNIVERSITI TEKNOLOGI PETRONAS

BANDAR SRI ISKANDAR

PERAK

AUGUST 2011

DECLARATION OF THESIS

FACE CLASSIFICATION FOR AUTHENTICATION APPROACH BY USING WAVELET TRANSFORM AND STATISTICAL FEATURES SELECTION METHOD

I NADIR NOURAIN DAWOUD JADALAH

hereby declare that the thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at UTP or other institutions.

	Witnessed
______________________	______________________
Signature of Author	Signature of Supervisor
Permanent Address:	Name of supervisor
University of Science & Technology Faculty of Electrical an Electronic Engineering P.O. Box. 30 Omdurman, Sudan	Dr. Brahim Belhauari Samir
Date:______________________	Date:______________________

ABSTRACT

This thesis consists of three parts: face localization, features selection and classification process. Three methods were proposed to locate the face region in the input image. Two of them based on pattern (template) Matching Approach, and the other based on clustering approach. Five datasets of faces namely: YALE database, MIT-CBCL database, Indian database, BioID database and Caltech database were used to evaluate the proposed methods. For the first method, the template image is prepared previously by using a set of faces. Later, the input image is enhanced by applying n-means kernel to decrease the image noise. Then Normalized Correlation (NC) is used to measure the correlation coefficients between the template image and the input image regions. For the second method, instead of using n-means kernel, an optimized metrics are used to measure the difference between the template image and the input image regions. In the last method, the Modified K-Means Algorithm was used to remove the non-face regions in the input image. The above-mentioned three methods showed accuracy of localization between 98% and 100% comparing with the existed methods. In the second part of the thesis, Discrete Wavelet Transform (DWT) utilized to transform the input image into number of wavelet coefficients. Then, the coefficients of weak statistical energy less than certain threshold were removed, and resulted in decreasing the primary wavelet coefficients number up to 98% out of the total coefficients. Later, only 40% statistical features were extracted from the hight energy features by using the variance modified metric. During the experimental (ORL) Dataset was used to test the proposed statistical method. Finally, Cluster-K-Nearest Neighbor (C-K-NN) was proposed to classify the input face based on the training faces images. The results showed a significant improvement of 99.39% in the ORL dataset and 100% in the Face94 dataset classification accuracy. Moreover, a new metrics were introduced to quantify the exactness of classification and some errors of the classification can be corrected. All the above experiments were implemented in MATLAB environment.

ABSTRAK

Tesis ini terdiri daripada tiga bahagian, ciri - ciri muka berpusat, pemilihan ciri -ciri dan proses klasifikasi. Tiga kaedah dicadangkan untuk mengesan kawasan berkenaan yang mengandungi gambar muka dalam bentuk imej, dua daripadanya adalah berdasarkan kaedah persamaan paten dan kaedah pengumpulan dalam kumpulan. lima data kumpulan yang dinamakan, YALE kandungan data, MIT-CBCL kandungan data, Indian kandungan, BioID kandungan dan Caltech kandungan Digunakan untuk menguji kaedah yang dicadangkan. Imej template disediakan sebelum ini dengan menggunakan set muka. Kemudiannya, imej tersebut akan dinaik taraf dengan menggunakan 'n-means kernel' untuk mengurangkan ganguan pada imej. Selepas itu 'Normalized Correlation (NC)' digunakan untuk mengira keyakinan perhubungan diantara imej template dan imej rantau. Pada kaedah kedua, selain dari menggunakan 'n-means kernel', matrik yang dioptimiskan digunakan untuk mengira perbezaan diantara imej template dan imej rantau. Pada keadah yang terakhir, 'modified k-means' algorithma digunakan untuk membuang kawasan yang tidak berkenaan. Ketiga - tiga kaedah diatas menunjukkan ketepatan kawasan diantara 98% hingga 100%. Pada bahagian kedua tesis, 'Discrete Wavelet Transform (DWT)' digunakan untuk mengubah imej kemasukan kepada bilangan wavelet. Kemudian, statistik wavelet yang lemah daripada sesetengah threshold akn dibuang dan mengakibatkan kepada penurunan bilangan wavelet sehingga 98%. Kemudian, hanya 40% ciri - ciri statistik diekstrak daripada ciri - ciri tenaga dengan menggunakan pemboleh ubah varian matrik. Semasa eksperimen ORL kandungan data digunakan untuk menguji kaedah yang dicadangkan. Akhir sekali, 'Cluster-K-Nearest Neighbour (C-K-NN)' telah dicadangkan untuk mengklasifikasikan muka kemasukkan bedasarkan sampel latihan. Keputusan menunjukkan peningkatan yang penting iaitu 99.39% dalam ORL dataset dan 100% dalam Faca94 dataset. Lebih dari itu, kami telah mencadangkan satu matrik bilangan yang boleh membetulkan kesalahan. Semua eksperiment diatas dijalan dengan menggunakan persekitaran MATLAB.

ACKNOWLEDGEMENT

I thank God the Almighty, on giving me the strength as opportunity to achieve what had been done in this work. I would like to thank my supervisors Dr. Brahim Belhaouari Samir, AP. Dr. Josefina Janier and AP. Dr. Vijanth Sagayan for their help and supervision in conducting this research. I greatly appreciate their care, dedication and constructive criticism. I have truly benefited and enjoy working with them.

I would like to thank the management and authorities of Universiti Teknologi PETRONAS for the generous financial support, providing different administrative and facilities throughout my stay with Electrical and Electronic Department.

I would like to express my sincere appreciation to my friends and my country mates for their support during the progress of this work.

I also would like to extend my appreciation to all of my family members for their understanding and inspiration for me to finish this project.

Dedication

To the memory of my father

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATION

AFKM	Adaptive Fuzzy K-Means
AR	Aleix Martinez & Robert Benavente
ASM	Active Shape Model
ATM	Automated Teller Machine
AWGN	Additive White Gaussian Noise
CCA	Curvilinear Component Analysis
C-K-NN	Cluster-K-Nearest Neighbor
DC	Direct Current
DCT	Discrete Cosine Transform
DNA	Deoxyribonucleic Acid
DWT	Discrete Wavelet Transform
ED	Euclidian Distance
FA	Factor Analysis
FBI	Federal Bureau of Investigation
FCM	Fuzzy C-Means
FERET	The Face Recognition Technology
FGFCM	Fast Generalized Fuzzy C-Means
FSVM	Fuzzy Support Vector Machine
FT	Fourier Transform
HMM	Hidden Markov Model
IC	Independent Component
ICA	Independent Component Analysis

JAFFE	The Japanese Female Facial Expression
K-NN	K-Nearest Neighbor classifier
KPCA	Kernel Principle Component Analysis
LDA	Linear Discriminant Analysis
LSAD	Locally scaled Sum of Absolute Differences
LSSD	Locally scaled Sum of Squared Differences
MIT-CBCL	Massachusetts Institute of Technology- The Center for Biological & Computational Learning
MKM	Moving K-Means
MLP	Multilayer Perceptrons
MPCA	Multi-linear Principle Components Analysis
NC	Normalized Correlation
NCC	Normalized Cross Correlation
NN	Neural Networks
NNC	Neural Networks Classifier
ORL	Olivetti Research Laboratory
OSAD	Optimized Sum of Absolute Differences
OSSD	Optimized Sum of Square Difference
PC	Principle Component
PCA	Principle Component Analysis
PDM	Point Distributed Model
PDF	Probability Density Function

PIN	Personal Identification Number
PPCA	Patch Principle Component Analysis
RBF	Redial Basis Function
ROI	Region of Interest
SAD	Sum of Absolute Differences
SFSM	Statistical Features Selection Method
SHD	Sum of Hamming Distances
S-LNMF	Selective Local Non-negative Matrix Factorization
SSD	Sum of Square Difference
SSS	Small Sample Size
STFT	Short Time Fourier Transform
SVM	Support Vector Machine
TC	Trust Coefficients
UMIST	University of Manchester Institute of Science and Technology
WEE	Wavelet Energy Entropy
WT	Wavelet Transform
ZNCC	Zero-mean Normalized Cross Correlation
ZSAD	Zero-mean Sum of Absolute Differences
ZSSD	Zero-mean Sum of Squared Differences
2DLDA	2 Dimensional Linear Discriminant Analysis

CHAPTER 1

INTRODUCTION

1.1 Introduction

Since the beginning of creation, the universe is run according to the specific system in high accuracy. Every day the science reveals the new aspect of this system, which the human began and it has remained until now. In general, we find that the regulations are divided into two types, the first type are public systems that govern the lives of people as general or in specific range and the second type of the regulations are private systems which are setup to organize the work of public systems to facilitate human life. Because these systems fall under the person who creates, modifies or invalidate them and with existing advanced technology, there is a growing need to make a sophisticated mechanism to authenticate and identify individuals to decide who is authorized to enter the system in confident way to ensure the security of the system. It is well known that there are two main traditional types of automatic individual's authentication approaches that have been widely used and these are ***knowledge-based approach and token-based approach*** [1]. The systems that depend on the first approach use "something you know" to verify the person such as Personal Identification Number (PIN) or password and the second approach use "something you have" to verify the person such as ID card, keys, credit card and passport. It appears from the above-mentioned approaches, which rely on storing the password in his\her mind for example or keep the credit card in his\her pocket may be forgotten or lost are part of the drawbacks of these approaches [2]. Hence there is a need to invent an idea to authenticate the individual by using human inherent attributes that can't be lost, forgotten, or guessed by impostors. Therefore, we should add new components called "biometric components" [3] to increase the security rate and to obtain perfect authentication system. At least two of these components (knowledge-based components, token-based components or biometric components) should be available.

1.2 The Concept of Biometrics

The term ***"biometrics"*** refers to the area of developing the statistical methods and mathematical operations to analyze biological data [4]. The meaning of ***"biometrics"*** in the field of computer science refers to the method of automatic recognition of the individuals based on their physical and behavioral traits [5]. Historically, there were some ideas which are the nucleus of the recent biometric systems. In the mid-19th of century, the chief of the criminal identification division of the police department in Paris Alphonse Bertillon invented a way to verify identities [6]by using some calculations and measurements on the individual body to get to know the identity of the offender and the idea was gaining popularity. The fingerprint is one of the first types of measurements and were stored and then compared with fingerprints found at the crime scene to get to the identity of the perpetrator. Concerning about crime scenes, in the early part of 20th century two journals have been established [1] and there were ***Biometrika*** created in 1900 by Karl Pearson and the second was ***Biometric Society*** in 1947. There were also applications of biometrics in other areas such as analyzing the results of laboratory experiments that were made of different agricultural crops to study the effect of treatments on these crops. Recently, ***Biometrics*** methods have been widely used in the field of the law, and some examples are fatherhood determination, forensic, illegal aliens…etc. Also there is large number of biometrics applications in other areas of civilian life such as system control access, prison security, Automated Teller Machine (ATM) and so on. Figure 1.1 has shown some examples of biometrics recently used.

Figure 1-1**:** Examples of biometrics

1.2.1 Biometric Authentication

The new technology of authentication relies on utilizing the individual's biometrics to verify their identity. It is preferred to the old traditional methods "***knowledge-based approach and token-based approach***" because many advantages. For example, biometrics cannot be forgotten, lost or guessed unlike the passwords; also it cannot be stolen by a thief. Moreover, Biometrics can be utilized in large numbers of applications as compared to the traditional methods. We can make several comparisons between biometrics and the traditional method to show the advantage of biometrics over these approaches [3]. In any facility, like building or office the surveillance system needs to check the authorized individual that needs to enter or not. In the traditional approaches if the individual has (card, password) and he\she lost or forgot this (card, password) for any reason firstly he\she cannot enter the building added to this if anyone found or guessed this (card, password) this means obtaining an entry permit, in contrast, in biometrics systems there is no possibility because the biometrics are features of the individual/s. To check if the individual has a criminal record, the traditional approaches cannot be used but we can use biometrics approach. One more advantage of using biometrics is the working mechanism. In traditional approach we just insert the card or the password then we can enter to the system while in biometrics we can enroll many individuals by storing their biometrics, later when the individuals claim to enter the systems the stored biometrics can be recalled and compared with individuals biometrics recently captured. Now, it's clear to us that biometrics authentication approach overcomes the drawback of the traditional authentication approaches.

1.2.2 Application of Biometrics Authentication

In the all applications of authentication we have to answer one of three questions, either ***"is this individual who he\she claims to be?"*** or ***"who is this individual?"*** or ***"is this individual in the database?"*** .In the case of the former question, the system uses the input biometrics of the individual to verify the authenticity of claimed identity. It compares these input biometrics with enrolled biometrics associated with the claimed identity of the individual himself then the system can accept or reject the claim, this operation called by ***"verification"*** problem [7] because we compare two biometrics of the same individual. In the last two questions, systems compare the input biometrics with all enrolled biometrics associated of all individuals in the database. Then the system can identify the person or just confirm if the individual belongs to database or not, this operation is called by ***"classification"*** problem [7] because we compare the biometrics of the claimed identity with all individuals. Both, verification and classification problems are recognition problem.

The physical and behavioral traits of the human have been widely used to verify the authenticity of claimed individuals, many applications used the face as biometrics also fingerprint and voice. The uses of fingerprint biometrics started in 1860 in India by Sir William James Herschel. He was introduced a method of identifying criminals by their fingerprints. The first person, who provided a scientific concept of the fingerprint, was Francis Galton in 1892, but there was a big problem in the matching process because it was done manually, making it difficult and time consuming. The automatic biometrics just stated about 40 years ago. The Federal Bureau of Investigation (FBI) developed a new method to identify fingerprint in the late 1960s. The eighth decade of the twentieth century has seen a revival in fingerprint recognition also a new emerged technologies have been used such as retina, iris, and voice. Later, face recognition [4]gain more attention in this area because it's more reliable than the previous biometrics. Also there is a huge number of biometrics that has been proposed recently to satisfy the applications conditions. It is clear that the first application of using the biometrics were on criminal investigation and law enforcement, but now we have other areas based biometrics authentication. These areas comprises controlled access systems which includes single sign-on, network access, resources access, web security, transaction security, and logon to applications. Also biometrics has been widely used in national security of countries against the sabotage attempts, international criminals and terrorist organizations to identify these people specifically after the attacks on September 11, 2001. Now, all the modern companies use the biometrics to monitors staff in terms of work schedules (attendance-leave) and the performance of employees. Recently, a new Automated Teller Machine (ATM) became available to bank customers allowing them to perform all their banking transactions, eliminating the trouble of going to banks. The machine verify the authenticity of customer identity by using their biometrics. Health institutions and social insurance companies, take advantage of biometrics technique to prevent fraud in the allocation of money for welfare benefits. In addition, biometrics is used in e-commerce, the new technology of e-government, cellular phone, e-banking, and phone banking. Moreover, the operation of storing the biometrics of individuals and passport checking in the airports and seaports helped to reduce the proportion of illegal immigration of the people from developing countries to the well-developed countries which show the growing importance of biometrics authentication area. At present, researchers are still working to develop these biometrics systems. There are a new biometrics systems introduced by companies and universities and researchers to increase our security in life that will cover all social and economic areas [3].

It can be seen that the identification of large number of individuals by using automated biometrics system is a challenging task. Therefore, there is a need to design a reliable and highly accurate biometrics system. As a result of the growing needs of the biometrics systems, there are many automatic biometrics systems that have been introduced and also number of systems still under development. Speech recognition and fingerprint identification systems are widely used during the middle of last century, facial biometrics became an interest area of research to fix the limitation of the previous biometrics and it gives good result in many applications as well [6]. Further, we will discuss their limitations.

1.3 Facial Biometrics

From all type of the body biometrics, facial biometrics is very important because we can recognize the individual through easily. Moreover, it contains a huge amount of information. Scientifically, it has been confirmed that an infant can recognize and distinguish his/her mother face from the other people after less than two days from the birth [5]. Furthermore, the people recognize each other through their faces for many reasons: the face biometrics is the best part of the body that gives sufficient answer about the identity, facial biometrics is the clearest biometrics for the eye. Finally, in most cases, the inability of the rest of the body parts to give sufficient information about the identity of the other people gives the advantage to the face biometrics.

Face recognition has caught attention of many researchers in last few decades. Numerous of the research efforts have been made on wide range of facial biometrics [8]. This biometrics has been extracted from different situations of the face images starting from still image to capturing the face biometrics from video image and also from image with clutter background. These research efforts encounter many challenges to design robust recognition systems because of changes in illumination, pose, background color, viewpoint of face image acquired and also the changes on the person face appearance from time to time (glass, hair, beard-mustache, makeup,….etc), also the facial expressions (sad, happy, surprised, wondering, depression,….etc) is challenging , Figure 1.2 shows some of these challenges. Furthermore, identifying twins is an important challenge for any recognition system. In 2003, two researchers Michael and Alexander Bronstein designed a new system that can distinguish between the twins [5], and this new technology is based on 3D face maps.

Actually, a face authentication system is face recognition system and it is divided into two parts: ***(i) enrollment (registration) (ii) matching (classification)***. The former includes the registration of individual's traits which are used later as templates for classification; and this can be done by storing these traits in known database. The latter is comparing between the stored templates and new traits captured from the individuals through system sensors to verify the person's identity.

Figure 1-2: Face challenges

In any face biometrics authentication system, there are five main steps:

1. Create database.
2. Face localization.
3. Features extraction.
4. Matching process (classification).
5. Decision.

Actually, the database creation is a separated step from the other steps the recognition system because it concerning on constructing the system library which will contain a set of images from the all individuals that will use the system. While the other steps happen after the start using the system and it deal with current users. In the localization step, first we have to differentiate between detection, localization and the concept of authentication. The detection means determining whether there is any face in the image and its location. In the localization, we are sure there is face in the image then we need to locate the place. In the authentication system there is only one face in the image, where in such systems it must be a camera that was used to capture the face images of the person. Normally, the face size is smaller than the image size, which causes problems in the verification, so we should locate the face and remove the rest of the image parts.

In the third step, facial features will be extracted from located face. In this step, we determine the most important features that can represent the face very well and invariance to lighting conditions, rotation, view point, expressions, pose and so on. This step increases the system’s accuracy.

A robust classifier is a very important task in authentication systems. The matching process will be done in step four to compute the similarity between extracted features and stored template features in the database. The importance of this step comes from the need for the high accuracy level of the classification in order to classify the extracted features to the right template feature that gives reliable result.

In the last step, based on the obtained results from the classifier the system can decide whether accept or reject the claim (verification) or it can identify the person (identification). The matching result is passed and uploaded in percentage and the system designer determines this percentage.

1.4 Motivation of work

As has been reported previously, identity authentication of individuals can be done by one or more of three approaches which are “knowledge-based” approach, “token-based” approach, and “biometrics” approach. Definitely, each one of these approaches has advantages and disadvantages but the combination between them will increase the accuracy of the systems. The biometrics system is the most secure way over the other method because it makes use of unique physical and\or behavioural trait of the individual. It cannot be forgotten or guessed by others like “knowledge-based” approach, and also it cannot get lost or damaged like “token-based” approach. Furthermore, biometrics approach is based on matching between the stored template of individual biometrics trait and the individual itself, where it is not possible to lend, hand over, or give the biometrics trait to someone else as in the other authentication approaches. In contrast of automatic applications, biometrics is undesirable for some spatial applications like manual applications. Moreover, biometrics cannot rely on perfect match between stored template and new acquired biometrics trait because there is a percentage of variation in characteristics and traits resulting from natural variability and health conditions due to passage of time. On the contrary, we find that this change does not happen where they remain full compliance with the stored value in the database. Therefore, we found that biometrics can only give us an estimation of the similarity between the new biometrics and the stored template in database. These natural variations and ageing which changes the biometrics traits, makes the biometrics technology a challenging task and thereby increasing the need of complicated algorithms to cancel the effect of these variations. Beside the variation in biometrics, another challenge is the selection of biometrics methodology such as face, finger print, iris, and… etc because in some cases these traits maybe do not exist which means not all individual have prominent traits. So we need to select the most clear and invariant trait for ease recognition. Also, the acceptance of the selected biometrics for the users is very important to make them feel assured of the usage system. The user should know systems' instructions. In addition, biometrics systems takes a longer time than the other authentication approaches in the matching operation because of relatively complex analysis of physical and behavioural trait stored in the template due to complicated algorithms.

1.4.1 Problem statement

Recently, human face became an area of interest for research to construct biometrics systems. Many face biometrics systems have been developed and introduced in different fields for commercial and private applications and it was proven to be successful. However, the selection of invariant facial features of the faces with highly variation conditions such as (illumination, expressions, poses and clutter background) represent a challenging task in recognition approaches. Many facial features have been proposed to represent the human face such as eyes, nose, ear and mouth. In some cases, these features are not present because of occlusion or injury. For some features extraction methods, the used algorithms are too complex. So there is need to simplify the algorithm or proposed simple one.

Nowadays, the facial features extraction based on statistical methods have been widely used which overcame the problems of the old methods. However, new drawbacks have been floated in the surface in these methods like features redundancy as in Wavelet Transform (WT) and Independent Components Analysis (ICA), small sample size as in Linear Discriminant Analysis (LDA) or the expression and rotation affection as in Principle Component Analysis (PCA). Therefore, in the selection of the appropriate features for classification there are a number of criteria's must be taken into account includes: invariant against the face variations, simple and quickly extraction algorithm, small sample dimensions and small memory size.

1.5 Objective

The main objective of this work is to increase recognition accuracy and efficiency of the authentication systems. To achieve this main objective, the following specific objectives were taken into account:

- To identify and develop some methods for face localization in order to locate the region of interest (ROI) face position in the input image.
- To develop a new feature extraction and selection method from multi-resolution representation mainly Discrete Wavelet Transform.
- To investigate the application of Cluster-K-Nearest Neighbor classifier to the face recognition problem in order to increase the accuracy and decrease the classification time.

1.6 Scope of work

This work will concentrate on template matching approaches to develop theory for new face localization method. Moreover, automatic region of interest (ROI) detection methods will be studied to propose a new detection method. Also it will study the advantages and disadvantages of features extraction and selection methods to develop an invariant and reliable features extraction method that can represent the face with small amount of data. Then in matching step the popular classifiers will be studied to propose robust and accurate classifier.

1.7 Work contribution

In this work we have verified a number of methods that are necessary for any recognition system. We have made improvements and enhancements to existing methods and in some cases, new methods are proposed. The specific contributions of this work are summarized as follows:

1. Three ROI methods for face localization with various pose, illumination, expressions and clutter background are presented. Two of these methods using template matching approach based on similarity measurements metrics which are:

- Template matching based n-mean-kernel-normalized correlation to measure the similarity.

- Template matching based optimized metrics to measure the difference.

In the third method, a new face localization method based on clustering concept is proposed and it utilize K-Means modified algorithm to distinguish between face and non-face regions.

2. New method for features selection from multi-resolution representation of discrete wavelet transform using the integration between statistical energy and variance modify metric of the wavelet coefficients is proposed to reduce the coefficient redundancy and select only the significant features.

3. A new robustly and efficiently classifier for facial features is investigated.

4. A new metrics to quantify the exactness of classification response is introduced with eventual correction of some classification errors.

1.8 Thesis Outline

The rest of the thesis is organized as follows:

Chapter 2 is a literature review on biometrics systems. Brief introductions of some face recognition systems are also included. Face localization problem are included and a comparison between localization methods are provided as well. In addition, some of features extraction and selection methods are presented. Moreover, a set of popular classifiers have been introduced in this chapter.

Chapter 3 presents the methodology of proposed techniques. The chapter is divided into three parts. In the first part, uses pattern (template) matching approach based on three types of optimized of similarity measurements metrics namely Normalized Correlation (NC) with n-mean kernel, Optimized Sum of Absolute Difference (OSAD) and Optimized Sum of Square Difference (OSSD) metrics in order to locate the ROI in the input images which contains the face. In addition, this part presents the automatic ROI detection method based on K-Means modified Algorithm and its mathematics theory included. In second part, first presents brief description of Discrete Wavelet Transform (DWT), followed by presenting the mathematical explanation of the proposed features extraction and selection method which invariant to several of facial variations, illumination and clutter background. The last part, explains the mathematical theory of the investigated classifier. In addition to that, a new idea of auto-correction and how to quantify the exactness of classification is included in part.

Chapter 4 presents the results and the localization accuracy rate obtained by applying each of three methods proposed in the first part of chapter 3. Moreover, this chapter presents results of classification accuracy rate obtained by applying proposed features extraction method in the second part of the previous chapter. The chapter gives the comparison between the proposed method and some other extraction methods using popular classifier. Finally, the results obtained by using the proposed classifier from part three in chapter 3 are compared with number of existing classifiers results and the results of the error correction metric is presented.

Chapter 5 gives the conclusion of the work and recommendations for the future work.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Normally, the individuals recognized each other through their physical or behavioral trait such as face, voice and gait. Nowadays, it became difficult to use human themselves for all areas of the life, so efforts were made to use the machine rather than human for recognition. Using machines to recognize the individual appears to be difficult unlike the human. The reason behind this problem lies in the difficulty of providing a clear explanation for machine of how the human mechanism can identify the other. Generally, the recognition can be classified into two main categories identification and authentication. The identification is to determine the identity of a person from set of people stored in database, while the authentication is to verify the authenticity claim of person's identity. Both categories make use of one or more of human biometrics such as face, speech, iris, fingerprint, and hand geometry to recognize the individuals. There are considerable numbers of biometrics systems that have been widely used for different applications such as system control access, prison security and Automated Teller Machine (ATM) and so on [6]. Choosing the appropriate biometrics is an important challenge since not only one human traits can work effectively for all applications that it could satisfy the following conditions: trait should existed in each person, should not be identical with the trait of someone else, should always be exist, does not change with the passage of time, and can be acquired easily as well. Moreover, there are some spatial conditions for biometrics systems comprising of the following: implementation should be accurate and fast, the biometrics scanning devices should be acceptable for users, and should be robust against fraud operations.

2.2 Biometrics Systems

Biometrics system is a pattern recognition system where its works on locating the biometrics data of the individuals in the input image and then set of features that will be extracted from these data and finally it will be matched with the stored classes in the system's database. The mechanism of biometrics systems can be classified into two classes either verification or identification. The verification, system will match the extracted features of the individual with the stored features in the system's database of the person who claimed to be. Such system, will verify the identity of the person through his biometrics and there will be one to one matching in order to determine whether the claim is true or not. This class usually prohibit the use of same identity by a number of individuals [4]. In the identification, the system will match the extracted features of the individual with the stored data in the system's database of all persons. In such system it will verify individual's identity by one to many matching in order to determine his identity. This class usually prohibits the use of many identities by a single individual [4]. Biometrics systems consist of four main stages, as follows:

1. Enrollment Stage, which is the preparing stage and works on collecting set of biometrics features for each user in the system with different situation then storing these features in the system database for matching process later. During the enrollment, firstly the biometrics traits of the users are scanned to give a characteristic representation of the user. A set of features data will be captured from the biometrics in different times and stored as biometrics features in order to increase the quality checker. In addition, these templates should be updated over the time to reduce the effects of biometrics variations.

2. Biometrics Detection Stage, it works on determining the location of biometrics data in the input images by using some algorithms. There are various types of algorithms which will be selected based on the application and the type of biometrics.

3. Features Extraction Stage, it works on extracting a set of features data from the detected biometrics. These extracted features should give robust representation and invariant features of the system users.

4. Matching Stage, it works on comparing the extracted features with stored features in order to compute matching similarity. This similarity is to determine whether the claimed identity of the user is true or not.

Now its clear for us the objective behind each above stages and job of each one in any biometrics system. Figure 2.1 shows the steps of both pathways (verification, identification) and the difference between them.

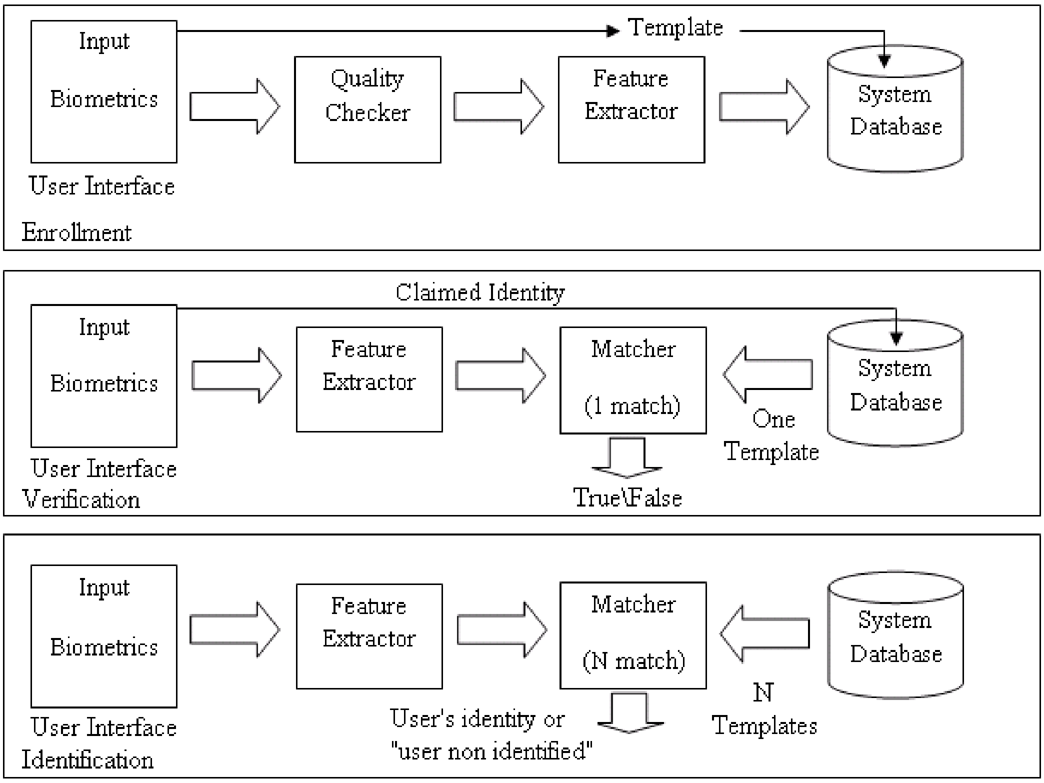


Figure 2-1: Block diagrams of enrollment, verification and identification

The huge number of human biometric traits were proposed by researchers to design a recognition systems for various applications and several purposes [3]. However, we cannot use only one trait from these biometrics for all applications due to variation in the strengths and weaknesses of them, but it will be selected based on the conditions of the application [2]. For more detail, a brief explanation of the most popular biometric is given below.

- ***Fingerprint:*** It is one of the first biometric traits used to identify the people through their fingers and its result was good [6]. Fingerprint formation started during the first seven months of the mother pregnancy in the form of a series of ridges and valleys in front of the finger and it's different between the twins even between the fingers of the person himself. Classification of fingerprints is a one of the important tasks in any recognition system, and it's a very difficult problem because of the small inter-class variation, and high intra-class variation. Moreover, the understandable properties of fingerprints and the presence of noise make the classification process more difficult. The wide review of the exiting fingerprint approaches that have been applied in the recognition systems to reduce the classification problems can be found in [9].

- ***Ear:*** Researchers proved that ear can be used as human biometrics to distinguish between the people, by applying some measurements on the ear and make comparison between the numbers of people. Human ear has many advantages over the other biometrics such as invariant against expressions, very small variation during the life time, and ease for data acquisition. However, in applications requiring high accuracy is not satisfactory because of its effects by variation in the lighting and also the variation of the size and pose in image and it needs to extract the ear features in standard position to increase the recognition accuracy [10].

- ***Palm-print:*** It's the same like fingerprint and it is in the form of a series of ridges and valleys but here we took the whole human hand. On the other hand, palm-print has larger area than fingerprint and due to that its more distinguishable between the individuals [11]. In addition, palm-print has more than one advantage over fingerprint which are useful for distinguishing such as wrinkles and principle lines. These addition features can help the biometrics scanner to capture the palms data even were lower resolution scanner [12].

- ***Hand Geometry:*** In this type of systems a number of measurements will be calculated from human hand such as size, palm area, finger length and width and the shape as well. These systems are very simple and easy to implement but it's expensive. One more problem in such systems it work bad in large population database and there are many approaches proposed to enhance such systems performance [13].

- ***Iris:*** It's the first stage in the human visual process and different from person to person even the twins. It's the ring region which is bounded by the white of the eye. These types of systems can be used for large population database but it's very expensive to establish.

- ***Gait:*** It's the way how the person walks and it is consider as a behavioral trait which can be effected during the long period of the time due to an injury and weight change. Gait systems can be used for slow application because it takes a long time for recognition operation. Start by capturing the gait data first and then a number of measurements on person movements will be done to match with the stored in the database. However, gait system will be the best choice in the low resolution better than the other biometrics [14].

- ***Voice:*** This type of systems can be classified that either physical, behavioral traits or both of them. Physical trait is way of how the person pronouns the word which different from each other and it's not applicable for the high accurate application. While in behavioral traits it is rely on the corresponding lips movements to person sound. A comparison between some voice techniques can be found in [15].

- ***Deoxyribonucleic Acid (DNA):*** its one of newer biometrics techniques to verify and identify the individuals. It's a very effective method and it's widely used in forensic. In fact, DNA is the same for the identical twins and it's very expensive due to using some chemical components.

- ***Face:*** It's the most commonly used biometrics in the recognition applications. There are many approaches in face recognition systems such as knowledge-base approach, features-base approach and appearance-based approach which use different methods to recognize the human faces [8].

Table 2.1 shows a comparison between the above biometrics and also some addition biometrics [6]. This comparison is based on numbers of factors and it is observed that no single biometrics can be used for many applications. In the table H, M and L indicate High, Medium and Low respectively.

Table 2-1**:** Comparison between a number of biometrics technologies based on different factors

Human Biometrics	Universality	Permanence	Collectability	Acceptability	Performance	Distinctiveness	Circumvention
Fingerprint	M	H	M	M	H	H	M
Ear	M	H	M	H	M	M	M
Palm-print	M	H	M	M	H	H	M
Hand Geometry	M	M	H	M	M	M	M
Iris	H	H	M	L	H	H	L
Gait	M	L	H	H	L	L	M
Voice	M	L	M	H	L	L	H
DNA	H	H	L	L	H	H	L
Face	H	M	H	H	L	L	H
Signature	L	L	H	H	L	L	H

2.3 Face Recognition

Both, authentication and identifications problem is recognition problem and extensive survey of face recognition approaches and techniques can be found in [8]. Recently, face recognition using appearance-based approach has acquired high attention which gives high recognition rate comparing with the other recognition approaches. The reason behind this superiority relies in using the whole face as input and utilizes a set of face samples to learning the classifiers. These training sets increase the accuracy rate in both categories (authentication, identification). Moreover, this approach uses statistical techniques to extract the most relevant features from the input faces to distinguish between the individuals [16]. Nowadays, there are number of efficient features extraction techniques such as Principle Components Analysis (PCA) and Independent Component Analysis (ICA) which are widely used to extract the facial features. Hyun et al. [17]used mixture-of-Eigenfaces method to obtain several set of Eigenfaces from the input faces to reduce the effect of high variations in human faces. Faruqe and Hasan [18] utilized PCA as features extractor and then Support Vector Machine was applied to classify these features. Chul Kim et al. [19] used Linear Discriminant Analysis (LDA) mixture model method instead of the standard LDA to obtain the transform matrix of each face class. Liu et al. [20]applied Independent Components Analysis (ICA) on the low-frequency coefficients of the 2D Wavelet Transform, and they proposed a criterion to select the useful ICs for recognition. In recognition process they used Fuzzy SVM (FSVM) classifier.

Face recognition systems still encounters many challenges such as pose variation, illumination, expressions, and occlusion. Different solutions have been proposed to reduce the effects of these problems. Zhang and Gao [21]presented detailed survey on face recognition under various poses and they reviewed and discussed the existing techniques. In addition, they classified the techniques into clusters based on its concept. Wei and Lai [22] proposed a new technique to recognize the human faces under different lighting based on matching of relative image gradient. The proposed method was applied for both localization and matching procedure and then matching correlation determines the recognition result. For recognizing the human face with high variation or expressions, Dai and Zhou [23] introduced a new face recognition method based on Support Vector Machine (SVM) classifier and statistical features extracted by Kernel PCA (KPCA) from Gabor Wavelet Transform coefficients of face images. The extracted features were robust against the face variation and it gave high representation of the human faces. Jun Oh et al. [24]proposed solution for face occlusion problem using Selective Local Non-negative Matrix Factorization (S-LNMF) technique. This proposed method consists of the two stages namely: occlusion detection stage and S-LNMF recognition stage.

Generally, all systems are based on face biometrics and consist of three major stages which are face detection or localization from the input image, features extraction stage and finally classification or recognition stage.

2.4 Face Localization

The localization of the Region of Interest (ROI) which contains the face is the first step in any automatic recognition system which an especial case of the face detection. The difference between both problems is: the detection problem is to determine the face existence in the input image then return the location of existed face. While in localization problem there is already existed face in input image and the goal is to determine the location of this face. However, face localization from input image is a challenging task due to variation in location, scale, pose, occlusion, illumination, facial expressions and clutter background.

Despite of various methods have been proposed to detect the faces in input image still there are needs to improve the performance of localization and detection methods. A survey on face recognition and some detection techniques can be found in [25]. While more recent survey mainly on face detection was written by Yang et al. [26], where they classify face detection methods into four main categories as follows:

1. **Knowledge-based methods.** These methods rely on detecting the human faces based on encoding human knowledge of how the face should appear. Normally, it measures the relationship between the face features and then determines the location. These methods are mainly for face localization [27].

2. **Features invariant methods.** The goal behind these methods is to search for features which will be invariant even when there is variation in rotation, pose, illumination or scale and then determine the face

location based on these features. These methods are mainly for face localization [28].

3. **Template matching methods.** In these methods, several face features stored in the system database as templates to represent either the face features or the whole face then calculate the correlation between the input image and stored templates to determine the face location. These methods can be used for localization or detection [29].

4. **Appearance-based method.** In these methods, set of training image will be used to learning the system about the characteristics of the faces then detecting the faces in the input image based on learned characteristics. These methods are mainly used for face detection [30].

Table 2.2 shows the summarized detection approaches and some techniques have been proposed in the literature based on above categories.

Table 2-2**:** Face detection and localization methods categories

Method	Previous Work
Knowledge-based	Rectangular knowledge rule and face structure [27]
Feature invariant method • Facial Features • Texture • Skin Color • Multiple Features	 Geometrical face model [31] Shape Information [32] Gaussian model [33] Combination of skin color and face feature [28]
Template matching • Predefined Face Template • Deformable Template	 Discriminating features analysis [29] 2D deformable model [34]
Appearance-based method • Eigenface • Distribution-based • Neural Network • Support Vector Machine • Naïve Bayes Classifier • Hidden Markov Model (HMM) • Information-Theoretical Approach	 Kernel PCA [35] Calibrated boosted cascade classifier [36] Polynomial neural network [37] SVM [30] Face-nonface classifier [38] Karhunen-Loeve transform with HMM[39] Maximum discrimination [40]

As the result of detection and localization methods' comparison, template matching methods are achieved high accuracy using the simplest algorithms.

2.4.1 Face Localization using Template matching methods

In template matching methods, face samples (mainly frontal face) is predefined and stored in the system database. In the localization stage, the correlation between the input image and stored faces images with different size or make sure the two faces have same size is calculated to determine the face location in the input image. The advantages of these methods are robust to noise, simple to implement and minimum time consuming for localization [41]. However, it's not sufficient for detecting faces in images with high variation in background and illumination due to some image parts appears to be the face location.

1. Predefined Template

The attempt to locate the faces in images has been started at middle of 19^{th} century. One of first attempts introduced by Sakai et al. [42]. They modeled the face based on a set of sub-templates for the eyes, nose, mouth, and the face contour. They matched the sub-template of the face contour with extracted lines from the input image to determine the region of interest (ROI) for the candidate face. Then they computed the correlation between the ROI and the other sub-templates to determine if there is any face. Latter, Craw et al. [43]proposed a new localization method based on shape template of a face. They extracted the face edges by using Sobel filter and then they used these edges to finding the face location in the image based on number of conditions. Govindaraju et al. [44] introduced face detection method relying on establishing face hypotheses and then tested it. In the first step, face model is installed by defining the head edge, and then they used Marr-Hildreth edge operator to capture an edge map of the input image, then a filter is used to remove all parts not containing the face. This proposed method gave 70% localization accuracy based on test 50 newspaper photographs. Tsukamoto et al. [45] presented a new model for face detection, first the sample images are divided into a set of blocks and then extract qualitative features from each block. Then, these features will be used to determine the existence of the face in every position in the input image. The detection result was 66.7%.

2. Deformable Template

One of the well-known problems in template matching methods is the model-based shape matching due to poses varaitions. Many efforts have been introduced in this area, Chin et al.[46] is one of the earliest proposers and they applied a simple transformations such as rotation and scaling to obtain the matched shapes. The template modeled is based on these matched shapes then the correlations were computed to recover the template. During their work, they mainly concentrate on the rigid shape matching.

Lanitis et al. [47] used intensity information and shape to provide face representation technique. A set of the face features based on the shape contours such as eyes and noses were labeled manually from the training images, then a number of points were extracted to represent the features shape. Point distributed model (PDM) was used to classify the features shape of a set of individuals and they used a similar approach to Karhunen-Loeve Procedure to represent the appearance of features shape intensity. Then PDM is used to determine the location of the face in input image and this can be done by using active shape model (ASM). An intensive survey on deformable template matching methods for object detection and recognition has been introduced by Jain [34] to shows the advantages of these methods over predefined template methods; they categorized the various proposed models based on deformable model definition.

Average face is one of the simplest methods of template matching which can be obtained from set of face samples. In localization stage, the rectangular block with high correlation is the candidate to be the face position. From methodology, it can be called filter match method where the input image is convolving with flipped version of the average face as filter. Statistically, filter matching method assumes there is a noise, where this noise is additive white Gaussian noise (AWGN) which is not suaitable for image variation such as clutter back ground, illumination and expressions [41]. To reduce the effect of high face variation problem, Eigenfaces approach is adopted to enhance matched filter performance [48] which makes linear combination for Eigenfaces of the average face and it assume that each face should be closed to this linear combination. However, Eigenfaces approach has problem which it reflects the variation in both the face and the noise [49]. Because of this problem, there is always some localization error which some non-face blocks may appear with high matching correlation to the linear combination of the average face and it's Eigenfaces is more than the face blocks. Therefore, Eigenfaces method can give good detection rate when the noise is white noised clutter. Meng et al. [41] proposed a new method to localize the human faces using linear discriminant from gray scale image. Faces and non-faces are modeled as Gaussian distribution to obtain an optimal Discriminant template. In addition, the result of the proposed method was 92.7% using University of Michigan face database.

One of the methods widely used to calculate the correlation between the average face (template) and the input image blocks is similarity measurements such as Normalized Cross Correlation (NCC) [50] [51]. However, NCC are affected by illumination and clutter background problems, because sometimes there are some non-face blocks have almost the same value of NCC as the average face template matrix. This problem can be solved by using Sum of Absolute Differences algorithm (SAD) [52] which is widely used for image compressing and object tracking but still SAD needs more optimization to give more accurate positions for face in the image. Moreover, SAD can give high face localization rate with high illumination variation, while the variation in the background affects the accuracy rate.

2.4.2 Face localization using segmentation method

One of other ROI detection trend is the idea of segment or group the image pixels into number of clusters or regions based on similarity properties of these groups. It gained more attention in the recognition technologies which rely on grouping the features vector to distinguish between the image regions, then concentrate on particular region which contain the face . One of the earliest surveys in image segmentation has been done by Fu and Mui [53]they classify the techniques into three classes: features shareholding (clustering), edge detection and region extraction. Later, Nikhil and Sankar [54] provided a review on image segmentation techniques but they only concentrate on fuzzy, non-fuzzy and color images techniques. Many efforts have been done in ROI detection which it divided into two types: region growing and region clustering. The difference between the two types is that region clustering search of the clusters with out prior information while region growing need for initial point called seeds to detect these clusters. The main problem in region growing approach is to find the suitable initial points because it's the basic of the clusters where the cluster will grow from the neighbor's pixels of these points based on specified deviation. For seeds selection, Wan and Higgins [55] defined a number of criteria to select insensitive initial points for the sub-class of region growing. To reduce the region growing time, Chang et al. [56] proposed a fast region growing method by using parallel segmentation.

As we mentioned before, region clustering approach search for cluster without prior information. Pappas and Jayant [57] generalized K-Means algorithm to group the image pixels based on spatial constrains and their intensity variations. These constrains include: first the region intensity to be close to the data and second imposes spatial continuity. The algorithm generalization comes from it is adaptive and includes the previous spatial constraints. Like the K-means clustering algorithm, their algorithm is iterative. These spatial constraints are included using Gibbs random field model and they summed that each region is characterized by a slowly varying intensity function. However, still the algorithm not that much accurate because of there are unimportant features. Later, to improve their method, they develop the proposed method by using caricature of the original image to keep only the most significant features and remove unimportant features [58]. Beside the problem of long time of segmentation, there are three basic problems that accure during the clustering process which are center redundancy, dead centers and trapped centers at local minima. Moving K-mean (MKM) proposed in [111] can overcome the three basic problems by minimize center redundancy and center dead as well as reducing indirectly the effect of trapped center at local minima. in spite of that, MKM sensitive to noise, centers are not located in the middle or centroid of a group of data and some members of centers with the largest fitness are enforced to be assigned as a member of a center with the smallest fitness. To reduce the effectcs of the these problems, Isa et al. [59] proposed three modified versions of moving K-means clustering algorithm for image segmentation the fuzzy moving k-means, adaptive moving k-means and adaptive fuzzy moving k-means algorithms. After that Isa and Sulaiman [60] introduced new segmentation method based on new clustering algorithm which is Adaptive Fuzzy k-means clustering Algorithm (AFKM). On ther hand, Fuzzy C-Means (FCM) algorithm was used in image segmentation but still has some drawbacks whih are lack enough robustness to noise and outliers, crucial parameter α selected generally through experience and the time of segmenting an image is dependent on the image size. To overcome the drawbacks of Fuzzy C-Means (FCM) algorithm Weiling and Daoqiang [61] integrated both local spatial and gray information to propose fast and robust Fuzzy C-means algorithm call Fast Generalized Fuzzy C-Means Algorithm (FGFCM) for image segmentation. Moreover, many researchers have provided a definition for some data characteristics where it have a significant impact on the K-means clustering analysis including the scattering of the data, noise, high-dimensionality, the size of the data and outliers in the data, data sets, types of attributes and scales of attributes [105]. However, there is needed for more investigation to understand how data distributions affect on K-means clustering performance. In [106] Hui Xiong et al. they provided a formal and organized study of the effect of skewed data distributions on K-means clustering. For the impact of high-dimensionality on K-means clustering performance the previous research found that the traditional Euclidean notion of proximity is not very useful for K-means clustering on real-world high-dimensional data sets, such as document data sets and gene-expression data sets. To overcome this challenge, researches turn to make use of dimensionality reduction techniques, such as multidimensional scaling [107]. Second, K-means has difficulty in detecting the “natural” clusters with nonspherical shapes [105]. Modify K-means clustering algorithm is a direction to solve this problem. Guha et al. [108] proposed the CURE method which makes use of multiple representative points to obtain the “natural” clusters shape information. For The problem of outliers and noise in the data which can also reduce clustering algorithms performance [109], especially for prototype-based algorithms such as K-means. The direction to solve thiskind of problem, by combine some outlierremoval techniques before conducting K-means clustering. For example, a simple method

[109] of detecting outliers is based on the distance measure. On other hnad, many modified K-means clustering algorithms that work well for smallor medium-size data sets are unable to deal with large data sets. Bradley et al. [110] they introduced a discussion of scaling K-means clustering to large data sets.

However, all prevoius solutions and efforts to increase the performance of K-Means Algorithm still need more investigation because it looking for local minimum. While, if these efforts turn to search for global minimum that will improve the performance of the algorithm. Samir Belhaouari [62] developed a new algorithm K-Means clustering algorithm which maximize the variation between the classes and minimize within the one class. The new algorithm can increase the performance of face localization approach by separating the input image into two classes face and non-face. The face position in the original image can be determined corresponding to the position of face regions which it was determined by algorithm.

After the human face has been located in the input image the features extraction will be the next stage which is used to extract the most appropriate facial features suitable for classification.

2.5 Features Extraction

Recently, appearance-based face recognition approach has gain more attention comparing with the other face recognition approaches which is constructing the classifiers by using a set of facial features extracted from the human faces. Therefore, the selection of invariant facial features especially in the existence of high variations in poses, expressions and illumination is an important step in face recognition systems. It determines the reliability and robustness of the systems. Furthermore, reliable features extraction method will help the classifier to make its operation simpler and less time consuming.

Many research efforts have been proposed to select appropriate facial features for classification. The selection of invariant features from facial image will provide good face representation which can distinguish between the faces. In appearance-based approach, features extraction techniques are based on statistical strategies to extract the facial features which deal with statistical components of the human faces such as PCA, LDA, and ICA. Faruqe et al. [18] used PCA as features extractor with various number of PCs to represent the face. Their accuracy result was 98% using ORL database. In [63] Kanan et al. proposed a new face recognition system to minimize the affection of illumination variation and they projected the human face as an array of Patch PCAs (PPCA) extracted from particular face part which has an information of local regions. AR database is used to evaluate the proposed method with one face as trained sample. The proposed method achieved recognition rate over 70% with various levels of illumination changes. Also Thakur et al. [64] preformed PCA with Redial Basis Function (RBF) to introduce a robust face recognition system. During their experiments, numbers of PCs were extracted from two datasets, where the recognition accuracy was between 80% and 95.50% for ORL dataset and 92.05% and 95.90% for UMIST face dataset. Since PCA is suffering from the illumination changes, new features extraction method have been emerging to represent the facial features vectors in discriminant projection which is Linear Discriminant Analysis (LDA). Tangquan et al. [65] proposed a new algorithm based on Eigenface-Fisher Linear Discriminant as features extractor method to deal with the problem of illumination and reducing the training sample dimensions where LDA is suffering from small sample size problem (SSS). ORL database was used to test the proposed methods and the recognition accuracy was 98.2%. In [66] Chen et al. proposed solution for (SSS) problem, the obtained recognition accuracy rate was 97.54% with 70 projection axes and 128 features. Later, in [67] Juwei introduced a new solution for small sample size problem by using new LDA method based regularized Fisher's separability criterion and they achieved 98% recognition accuracy using FERET face database. Another features extraction method has been proposed to extract the features by using the independent component analysis ICA.[68] Kim et al. proved that ICA is robust against the illumination and the pose changes. During the experimental, they found that extracting of ICA from residual face space is robust to pose and illumination changes and also provide reduction of the features dimensions. Sets of databases were used to evaluate the proposed method namely: ORL, AR, Yale FERET and Bern, while the recognition accuracy was 98% in the case of the illumination and 97% in the case of the pose test. Moreover, there are other features extraction methods that were proposed, Kam et al. [69] proposed new method for features extraction. In this method the features are extracted from the gray level based on descriptive statistic of the face image. The results of the proposed method were compared with traditional Eigenfaces using ORL and Yale databases for evaluation. The recognition accuracy of the method was 92.50% and 80% for ORL and Yale database respectively. In [70] Eigenfaces was used to extract the relevant features for classification and accuracy was 95.6% on small database consist of eight subject with 10 images each.

As its well known the facial images pixels are huge, a number of transformation methods have been proposed to represent the face image in different form by set of transformation coefficients to reduce the image dimensions. Moreover, these methods provide invariant and less sensitive features coefficients for face images variations. In [71] the combination between fractional Fourier transform and fisher-face method was proposed. 2D separability judgment is used to choose the angle value parameter for discrete fractional Fourier transform then the discriminative features extracted by using fisher-face method. ORL and FERET databases are used to test the proposed method and the recognition accuracy was 97.55% and 82.63% respectively. Jadhay et al. [72] applied Discrete Cosine Transform (DCT) on Radon transform frequency components to provide frequency features and reduced the features dimensions. The results showed that the proposed method is robustness against zero mean white noise. The maximum recognition accuracy achieved by the proposed method was 98.93% and 97.38% for FERET and ORL databases respectively. However, these transformation methods only provided sufficient features in frequency domain which are weakness in time domain.

Multi-resolution representations such as Discrete Wavelet Transform (DWT) and Curvelet transformation are effective solutions to overcome the drawbacks of the other transformation techniques. Many efforts have been done on DWT to extract an invariant features from face images for face recognition applications. Hong et al. [73] used the approximation sub-image coefficients of the second level of DWT from ORL database as input to PCA features extractor. During the experiments four different wavelet families have been used and they found wavelet (sym2) gave high recognition rate with 94.2%. Xiang [74] presented face recognition system based 2-dimentional wavelet transform and KPCA as features extractor to reduce the affect of illumination problem. Yale B frontal faces database is used to test the proposed method and the accuracy was 100% for recognition rate. Ki-Chung et al [75] transformed the facial features from Gabor filters response into Eigenspace by using PCA as features extractor. In the results, they found that PCA weakness is against illumination and rotation variation, while they achieved recognition improvement up to 19% and 11% on SAIT and ORL databases respectively. In [76] a new face recognition method was proposed based on DWT and image comparison and it tested on standard face database from ORL database, where the experimental showed that the effects of various parameters on the recognition results. A new technique based on wavelet-curvelet-fractal method has been proposed by Zhong et al. [77] to introduce an invariant features from the input facial images. In this technique, first they performed wavelet decompose and curvelet transform simultaneously on the facial image to extract the facial details then fractal techniques was used as features extractor by using the corresponding fractal dimensions as features for classification process. During the experimental, JAFFE and Faces94 databases were used to test the proposed method and the result were 98.8% and 99.5% respectively.

However, those techniques are either mathematically complicated and time consuming or the inability to extract all facial features due to the face variations such as poses, expressions and illumination. Therefore, new statistical features have been proposed by Cunjian and Jiashu [78] as features extraction methods, where it is ease for extraction and simple to implement. These features called Wavelet Energy Entropy (WEE) and their obtained result was 90% on ORL database. Thus, these new statistical features provide less complexity and increase the classification rate. But the use of energy features is not efficient if there is overlapping between the classes because these entropy features used to capture the features with high information only without considering to the location of these features.

2.6 Classification Process

In any face recognition system, after the features have been extracted from the faces there is an important stage which is matching process. This process determines the matching score between the input candidate face and the stored features faces in the database to confirm the claim identity of the individuals or to identify the individuals. Therefore, the classifier design is an important challenge in face recognition approach due to the need of accurate classification decision for input features comparing with the database. Therefore, there are three conditions that should be provided in any perfect classification process to complete the goal of the recognition systems [79]:

- Highly accurate.
- Less time of classification.
- Small size of training sample.

Many classifiers have been proposed to satisfy these three conditions. K-Nearest Neighbor classifier (K-NN) is one of the popular classifier in face recognition approaches where it widely used to classify unknown faces based on similarity to known templates or samples in the database by calculating the distance between unknown face features to every stored sample in database to select the nearest sample. Then classify the unknown face to the class contain most nearest samples to it. Kam et al. [69] used K-NN classifier for recognition step by using descriptive statistic features as input features for classification. In [80] K-NN has been used to assign the extracted features to particular class in the database by using curvilinear component analysis (CCA) as features extractor. In [81] multi-linear principle components analysis (MPCA)is used as features extractor to construct K-NN classifier to recognize the faces from ORL database. However, K-NN is a simple classifier which consumes a long time in test phase due to the effect of highly data dimensions. Therefore, SVM was the best solution for KNN drawbacks [82] by using geometric concept, where the SVM classifier separate between class margins by search about hyperplane with maximum margins separation of class’s vector. By other mean, to maximize the distance between label vector and hyperplane projected by SVM. Many researchers exploited SVM as features classifier [18, 23, 73] because of its powerful in predicting unseen data and preferable for both linear and non-linear separable databases but it encounters decrease in the predicting power when there are too many classes [82]. In [20] the useful independent components ICs extracted from wavelet coefficients were used as input features for fuzzy support vector machine (FSVM) classifier. Also Neural Networks Classifier (NNC) [70] is one of the most notably classifiers that have been widely exploited in face recognition approach, in multilayer perceptrons (MLPs) and radial basis functions (RBFs). Two Radial Basis Function Network architectures used by Ranganath et al. [83] to design face recognition classifier and K-NN classifier was used in comparison with the proposed classifier on MIT database. Furthermore, the characterization of the statistical properties of a face image has gained more attention recently. Nefian and Hayes [84] used Hidden Markov Model (HMM) to built a new classifier. They assign each of important facial features such as hair, eyes, nose and mouth to stage in one dimensional from left to right HMM. Some researches prefer the combination between HMM and other classifiers to increase the outcome performance of the system in comparison with HMM separate [85]. Later, Qingmiao et al. [86] proposed a mixed classifier by using HMM with K-NN classifiers for face expressions.

2.7 Chapter summary

In the literature, many researches have been done on localization of the human faces in different types of images. The new developments should consider the limitations and drawbacks of the existing algorithms. Furthermore, template matching approaches have achieved significant degree of robustness and effectiveness in localization of faces in various types of images. Future work in template matching approaches should focus on improving the performance of similarity measurements algorithms. Moreover, segmentation techniques mainly region based techniques proved superior on clustering all similar pixels of the image into number of classes which make the localization more easily. Therefore, further work should concern on developing new clustering algorithm. Features extraction and selection is one of important stages in developing face recognition system. Numerous of features for human face characterization have been developed previously. Consequently, more researchers said that there is a necessity to measure the robustness of new features that can produce high classification rate. The selection of invariant features to faces with various expressions, pose, illumination, rotation and complex background needs an intensive study. Since, human face contain huge amount of features which vary in power where some of it not important for classification and makes redundant. Hence, the future work should select invariant features that are less sensitive to variation, useful for classification and small size. A variety of classifiers has been proposed to the features classification problems. KNN classifier is widely used for many applications but it needs a long time for response. NNC and SVM have also been used in classification problems. But still these classifiers have number of limitations. Therefore, future work should focus on developing the performance of one of these classifiers to increase the reliability of recognition rate of whole system.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

The biometrics approach whether physical or behavioral showed its superiority compared with other approaches in the area of authentication. During this work, the study to designed to develop a computer recognition system for human authentication (verification) in frontal image using an invariant biometrics features. The proposed method is applied to provide high reliability of authentication and to solve the limitations of the existing systems [3].

The developed system starts by locating the human face in the captured image using digital camera in the first step. In the second step, multi-resolution wavelet is used to decompose the detected face in a number of coefficients to reduce the dimensions of DWT coefficients. In the third step, significant statistical features for classification are selected from decomposition coefficients. In step four, the candidate face is classified to one of database classes according to selected features and the trusted coefficients are calculated to confirm and quantify this classification decision. In the last stage, the developed system determines whether the claimed face is authorized, not authorized. The next diagram illustrates the general steps of the developed system (see Figure 3.1).

This study consists of three main parts. Part I contains the development of a face localization method. In this part, three Region of Interest (ROI) methods are suggested to localize the face in the input images, two of them are pattern (template) matching methods and the third is clustering method. In part II, the mathematical theory of the statistical selection method from wavelet coefficients is presented to distinguish between the different classes of face. Part III, contains the mathematica l theory of new classifier called Cluster-K-Nearest neighbor (C-K-NN) which is used to evaluate the classification accuracy. In addition, the introduction of new idea to quantify the response of the proposed classifier by using quantification metric to calculate the trusted coefficients of classification is included in this part.

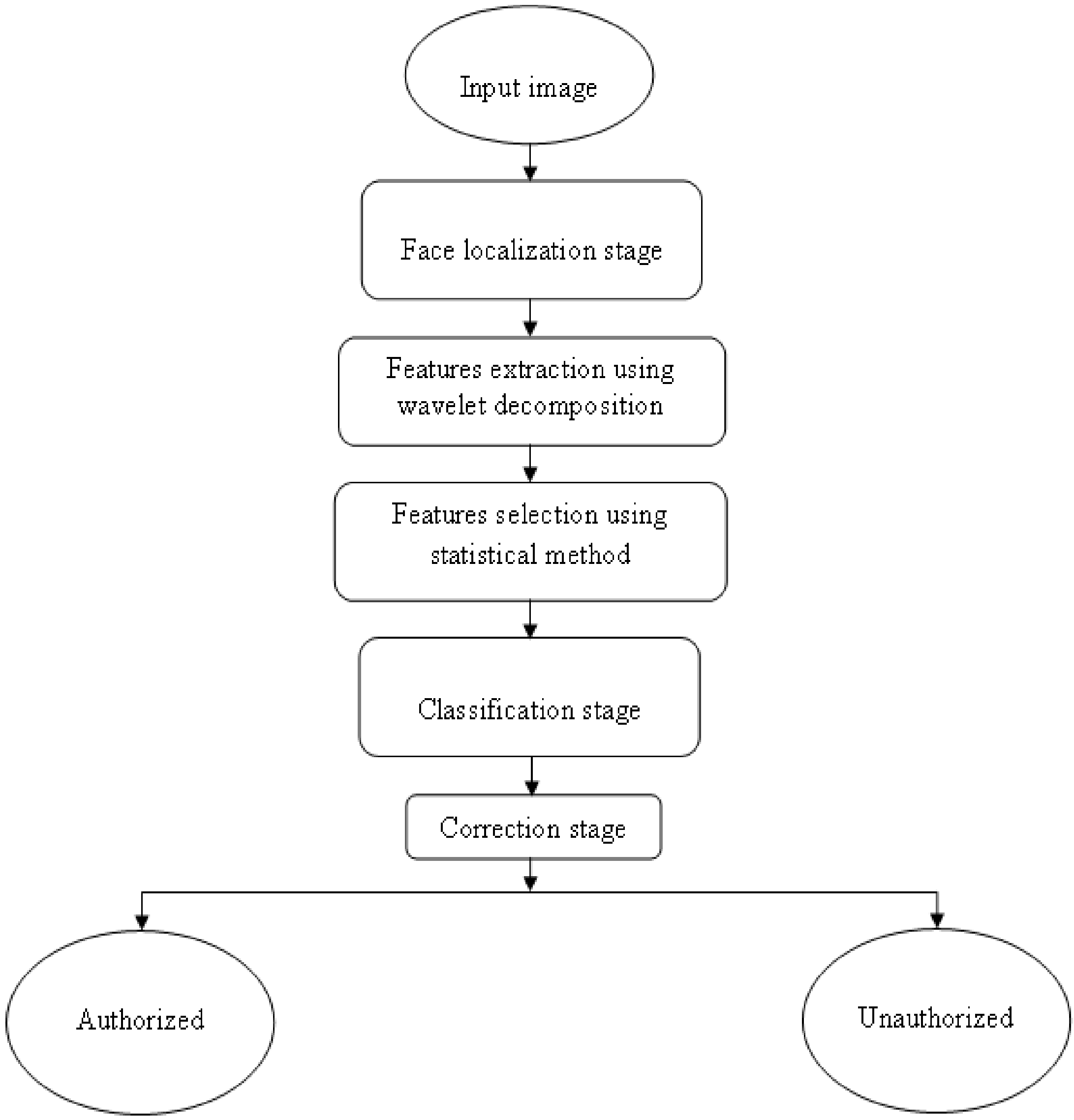


Figure 3-1**:** The general steps of the developed system which contains: face localization, features extraction, features selection, classification and correction

3.2 Part I: Face Localization Based on ROI Methods

This part presents the suggested methods of Face localization. Three ROI methods are proposed to locate the face in the input image, two of them are pattern (template) matching methods which are namely normalized correlation with N-mean Kernel filter (N-mean-Kernel-NC) and the second method is Optimized metrics to calculate the difference between the template image and input image blocks. The third method is clustering method which uses K-Means Modified Algorithm to separate between face and non-face regions.

3.2.1 Face localization using template matching and N-mean Kernel with normalized correlation (N-mean-Kernel-NC)

Since the region of interest (ROI) which includes only the face is almost less than 40% of the original image, where the remaining parts of the image are background comprised a different illumination, clutters conditions and unwanted face parts. Therefore, a cropping operation is applied to the images to remove unwanted parts of the images and extract only the region which contains the face. According to the dataset, a number of ROIs 160 x 120 pixels are cropped from all dataset faces with different cases (expressions, poses, illumination) to construct average template which will be used latter in matching stage. The cropping operation was performed manually, where the given center of the faces area are selected to be the center of ROI. Thus, we will be sure that all background information and unwanted parts are removed. Figure 3.2 shows an example of cropping operation that removes some face parts and illumination background. Once the ROIs are cropped, the average of them will be calculated to obtain an average template which will be used as template to locate faces in the coming input images; Figure 3.3 shows our average template, which is like face without details.

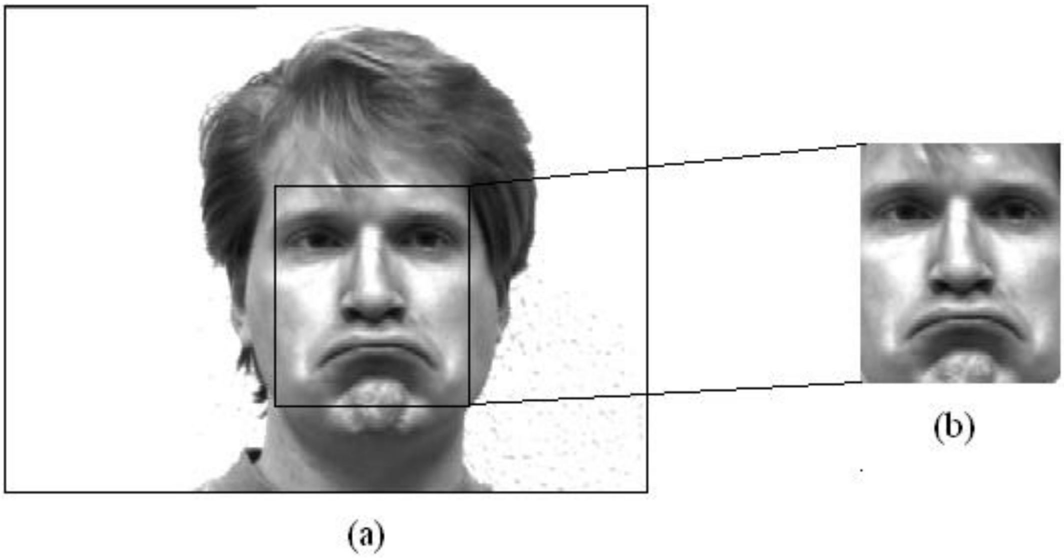


Figure 3-2: (a) Original (320 x 243), (b) Cropped image (100 x 120)

Figure 3-3: Example of the average template

Actually, this method can be used to locate the faces in the images with high illumination variations where the normalized correlation (NC) is affected by these variations. Therefore, there is a need to make preprocessing stage before computing the correlation between the input image blocks and average template to improve the quality of the image. In this preprocessing step n-means Kernel will be used to reduce the effect of the illumination by increasing the image brightness and decreasing the image noise. Thus, we have two stages namely: enhancing and matching stages.

1. **Enhancing Stage**

It seems like low pass filter or more precisely it will be like smoothing the image. Further we will explain the description of this filter and explain how it works.

In general, image filter is used to pass through either the high frequencies for enhancing the image and detecting the edges or to pass through the low frequencies for smoothing the image. There are two domains which we can filter the image namely: frequency domain and spatial domain. On the other hand, we can demonstrate the spatial domain process as to convolve the input image A with filter B. This operation can be formulated by using the equation:

$$C(i,j) = A(i,j) * B(i,j) \tag{3.1}$$

If we have square Kernel with size L×L, the output image can be calculated by the equation:

$$C(i,j) = \sum_{m=1}^{M}\sum_{n=1}^{M} A(m,n)B(i-m, j-n)\,, \tag{3.2}$$

where M is dimension of the filter matrix B.

There are numbers of standard Kernels and it is used to determine the operations characteristics based on the form and the size of them. One of the most important Kernels is mean filter and it is used for smoothing. It's one of the simplest methods of smoothing the image and it is easy to implement. It is used to decrease the variation between the image pixels values and its neighbor pixels. This filter is very useful to reduce the image noise.

The implementation of this filter is by replacing the image pixels value with the mean of neighbors. This process will eliminate the high variation between the pixels and reduce the noise. The following matrix in Figure 3.4 shows the n-mean Kernel.

$1/n^2$	$1/n^2$	$1/n^2$
$1/n^2$	$1/n^2$	$1/n^2$
$1/n^2$	$1/n^2$	$1/n^2$

Figure 3-4: N-Mean Kernel

After the filtering operation, we can see clearly the improvement of the image, Figure 3.5 will demonstrate that:

Figure 3-5: (a) The original image; (b) The filtered image

2. Matching Process

Once the preprocessing stage is finished, a dynamic window will be created on the filtered image by the same size of average face image (template). This dynamic window will be passed through the filtered image with step equal to the small square block then the correlation between the template and the dynamic window will be computed using NC. In general, NC can be computed using the following equation [41]:

$$c = \underline{\mu}^{\mathrm{T}} \frac{\underline{x} - d(\underline{x})}{\left\| \underline{x} - d(\underline{x}) \right\|} = (\underline{\mu} - d(\underline{\mu}))^{\mathrm{T}} \frac{\underline{x}}{\left\| \underline{x} - d(\underline{x}) \right\|} \tag{3.3}$$

where: C gives the maximum probability of a face, $\underline{\mu}$ is a face template, $d\underline{\mu}$ its DC component and which is the mean value of the waveform, and $\underline{x}$ is any square block of the original image. But during our work and for more simplification we will use another equation proposed by [87] to compute the correlation coefficients:

$$R_{ij} = \frac{\sum_m \sum_n (A(m,n) - \overline{A})(B(m,n) - \overline{B})}{\sqrt{\sum_m \sum_n (A(m,n) - \overline{A})^2 * \sum_m \sum_n (B(m,n) - \overline{B})^2}} \tag{3.4}$$

where

$$\overline{A} = \frac{\sum_{m=n}^{N}\sum_{m=n}^{M} A(m,n)}{N.M} \quad , \quad \overline{B} = \frac{\sum_{m=n}^{N}\sum_{m=n}^{M} B(m,n)}{N.M}$$

where A is the average face (template) and B is the dynamic window of the filtered image. The correlation coefficients will be saved in the separated matrix called record matrix, then, the window corresponding to the maximum correlation coefficients value in the record matrix will be extracted as the face region. Figure 3.6 shows the method of correlation coefficients calculation and its recording.

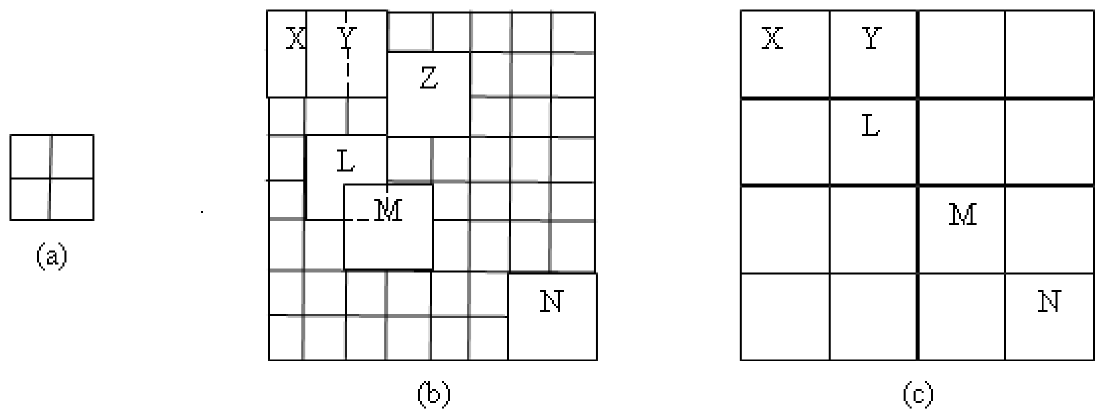


Figure 3-6: (a) The template face, (b) The dynamic windows through the image, (c) The record matrix

Now, based on the position of the face region in filtered image we can locate the corresponding face position in the original image by using windows index. Figure 3.7 shows an example of the face localization.

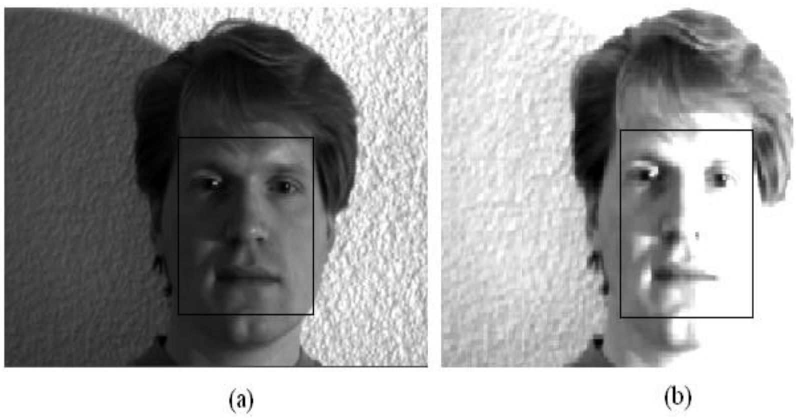


Figure 3-7: (a) Corresponding face location in the original image of the face location in filtered image (b)

Now, we can summarize the method as in the following diagram (see Figure 3.8):

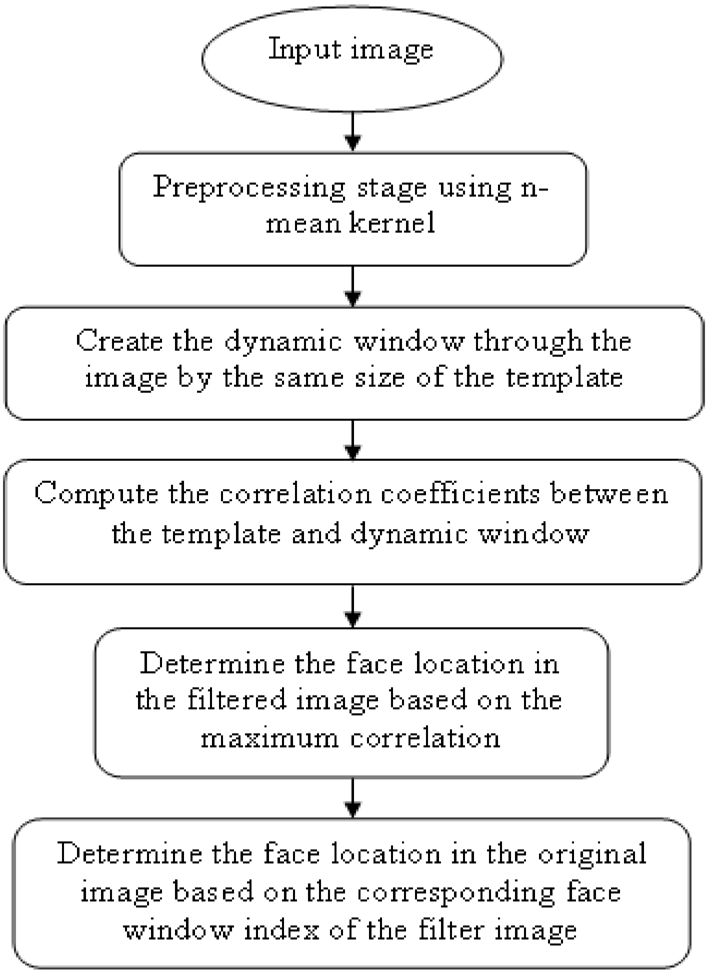


Figure 3-8: Face localization using image enhancing and normalized correlation

3.2.2 Face Localization Using Template Matching and Optimized Metrics

In this method, instead of using n-mean kernel as preprocessing stage, a number of optimized metrics can be used to calculate the difference between the template image and the dynamic window through the input image. Then, the image window correspond to the minimum value in the record matrix will be extracted as face region. To understand the idea of that, first we need to explain the difference meaning in the various spaces, and then explain the optimized metrics.

The difference meaning can be introduced as a number of mathematical definitions for different representation space and the following spaces will present that:

1. Figure 3.9 shows the difference of two points in one dimension and it can be formulated as follow:

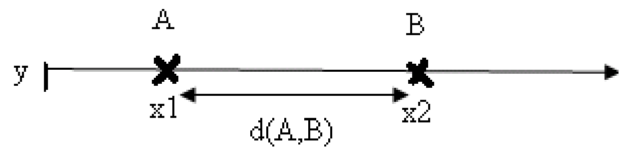


Figure 3-9: The difference of two points in one dimension

$$d(A,B) = |x_1 - x_2| \tag{3.5}$$

2. Figure 3.10 shows the difference of two points in two dimensions and it can be formulated as follow:

$$d(A,B) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{3.6}$$

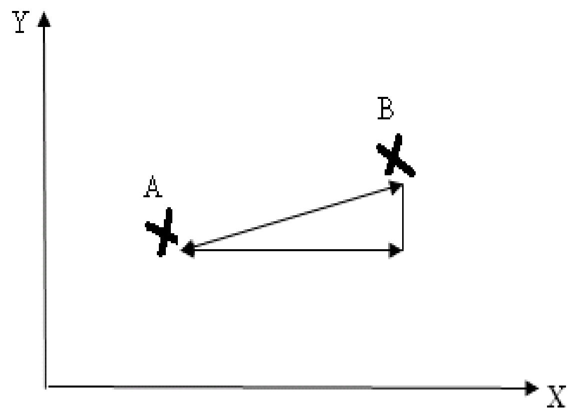


Figure 3-10: The difference of two points in two dimensions

3. Figure 3.11 shows the difference of two functions f(x) and g(x), and it can be formulated as follow:

$$d(f,g) = \int_R |f(x) - g(x)| \mathrm{d}x \ \ or \ d(f,g) = Max_x |f(x) - g(x)| \ or = \frac{\int_R |f(x) - g(x)| \mathrm{d}x}{\int_R |f(x) + g(x)| \mathrm{d}x} \tag{3.7}$$

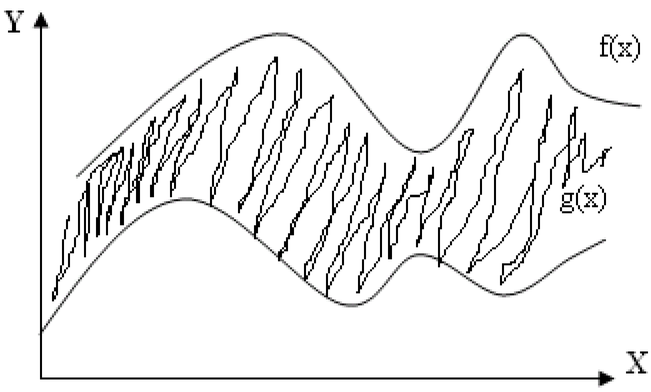


Figure 3-11: The difference of two functions

4. Figure 3.12 shows the difference of two matrices A and B and it can be formulated as follow:

$$d(A,B) = A - B \tag{3.8}$$

$$\underbrace{\begin{array}{|c|c|c|}\hline 1 & 2 & 3 \\ \hline 4 & 5 & 6 \\ \hline 7 & 8 & 9 \\ \hline \end{array}}_{A} - \underbrace{\begin{array}{|c|c|c|}\hline 6 & 3 & 0 \\ \hline 2 & 5 & 1 \\ \hline 8 & 7 & 1 \\ \hline \end{array}}_{B} = \begin{array}{|c|c|c|}\hline -5 & -1 & 3 \\ \hline 2 & 0 & 5 \\ \hline -1 & 1 & 8 \\ \hline \end{array}$$

Figure 3-12: The difference of two matrixes

Now, the difference definitions in different spaces are clear, we can now give a definition for the optimized metrics.

Actually, those metrics are optimizing the Sum of Absolute Difference (SAD) and Sum of Square Difference (SSD) metrics which are widely used in video tracking and image compressing; it's simple, and very easy to implement in order to find a relationship between two images. The idea of those metrics based on calculating the difference between each element in the template image corresponding element in the dynamic window through the input image. Then, the absolute or the square values of the difference will be calculated and gathered together. There are many applications for SAD and SSD such as motion estimation, object recognition and video compression. The result in Figure 3.12 can give an example of SAD and SSD methods after the subtraction of the two matrixes. In the resulting matrix there are some negative values. Therefore we will take the absolute value of all matrix elements and then sum up these elements. The result of this summation gives SAD between the image window and face template image. SAD can be computed by using the equation:

$$d(A,B)=\sum_i\sum_j\left|A(i,j)-B(i,j)\right| \tag{3.9}$$

SAD of the two matrixes =5+1+3+2+0+5+1+1+8=26

While

$$d(A,B)=\sum_i\sum_j\left(A(i,j)-B(i,j)\right)^2 \tag{3.10}$$

SSD of the two matrixes=25+1+9+4+0+25+1+1+64=130

In contrast with the other common correlation based similarity methods namely Normalized Cross Correlation (NCC) and Sum of Hamming Distances (SHD), SAD and SSD are very simple and more accurate and also less sensitive to illumination. However, there are some localization errors it maybe due to two neighbor windows have the face and almost the same SAD or SSD. To improve those metrics we need to optimize equation (3.9) and (3.10) to find the optimum image window that contained exactly the face. The following equations give Optimized (SAD) and Optimized (SSD):

$$OSAD_1(A,B)=\sum_i\sum_j\frac{\left|A(i,j)-B(i,j)\right|}{\max(A(i,j),B(i,j))} \tag{3.11}$$

$$OSAD_2(A,B)=\sum_i\sum_j\frac{\left|A(i,j)-B(i,j)\right|}{\left|(A(i,j)+B(i,j))\right|} \tag{3.12}$$

$$OSAD_3(A,B)=\sum_i\sum_j\frac{\left|A(i,j)-B(i,j)\right|}{\sqrt{A(i,j)+B(i,j)}} \tag{3.13}$$

$$OSSD_1(A,B)=\sum_i\sum_j\frac{\left(A(i,j)-B(i,j)\right)^2}{\left(A(i,j)+B(i,j)\right)} \tag{3.14}$$

$$OSSD_2(A,B) = \sum_i \sum_j \frac{\left(A(i,j) - B(i,j)\right)^2}{\left(A(i,j) + B(i,j)\right)^2} \tag{3.15}$$

In addition to the previous optimized metrics, two optimized error measurement metrics to find the face location in the input image. The first meric is chi-sqaure test which can be used to measure the difference between the expected frequencies and the observed frequencies in one or many cathegories of faces. In general, the chi-square test statisitic is calculate as follows :

$$x^2 = \sum \frac{(Observed - Expected)^2}{Expected} \tag{3.16}$$

If the computed test statistic is high, then the difference between the observed and the expected is significant and the model is poor to fit the data.

In our case, the sample is the set of n human faces, mainly frontal faces, noted

$S = \{A_i, 1 \le i \le n\}.$

The expected face model is the dynamic window within the image that contains a face, noted B. The Chi2 quatify the similarity between the Joint Probability Ditribution (dymamic windows) and the sample (set of human face). The chi2 formula for these data is defined as follows:

$$\chi^2_{u \times v - 1} = \sum_{i=1}^{u} \sum_{j=1}^{v} \frac{\left(\overline{A(i,j)} - B(i,j)\right)^2}{B(i,j)}, \tag{3.17}$$

where u and v are face sizes and the template is defined as the average of a face image sample noted by

$$\overline{A}(i,j) = \frac{\sum_{k=1}^{n} A_k(i,j)}{n}, \quad 1 \le i \le u; 1 \le j \le v.$$

We calculate the statisitic test for each dynamic window in order to find the most likely location of any given face, which is the one corresponding to the minimum chi-square value. To decrease the error of localization it is better to take into consideration the location of lower locals minimum, which are close to the global minimum, as suspected face locations.

The second meric is Sum Square of Student's t-distribution, we use it to measure the difference between the tample and the dynamic windows. Our assymption in this case is to consider all pixels values of image as independent variables, to make the decision if the dynamic windows is or not face, it's enough to test the null Hypothesis, dynamic windows is face, against the alternative hypothesis, dymanic windows is not face, by using the following statistic test:

$$SST = \sum_{i=1}^{n}\sum_{j=1}^{v}\left(\frac{\overline{A}(i,j) - B(i,j)}{S(i,j)/\sqrt{n}}\right)^2, \tag{3.18}$$

where

$$S^2(i,j) = \sum_{i=1}^{n}\left(A_k(i,j) - \overline{A}(i,j)\right)^2 /(n-1).$$

Using Central Limit Theorem and the fact, that t-ditribution square of freedom n-1 is Fisher distribution of freedom 1 and n-1, the Sum Square t-ditribution Normalized (SSTN) convergs in distribution to standard normal distribution:

$$SSTN = \frac{SST - u \times v(n-1)/(n-2)}{\sqrt{u \times v \dfrac{2(n-1)^2(n-2)}{(n-3)^2(n-5)}}} \approx N(0,1). \tag{3.19}$$

We caclculte SSTN for all dynamic windows to extracted all suspected windows to be face corresponding to the location of the lower locals minimum approaching the global minimum.

Now, the proposed method can be summarized as in the following diagram (see Figure 3.13):

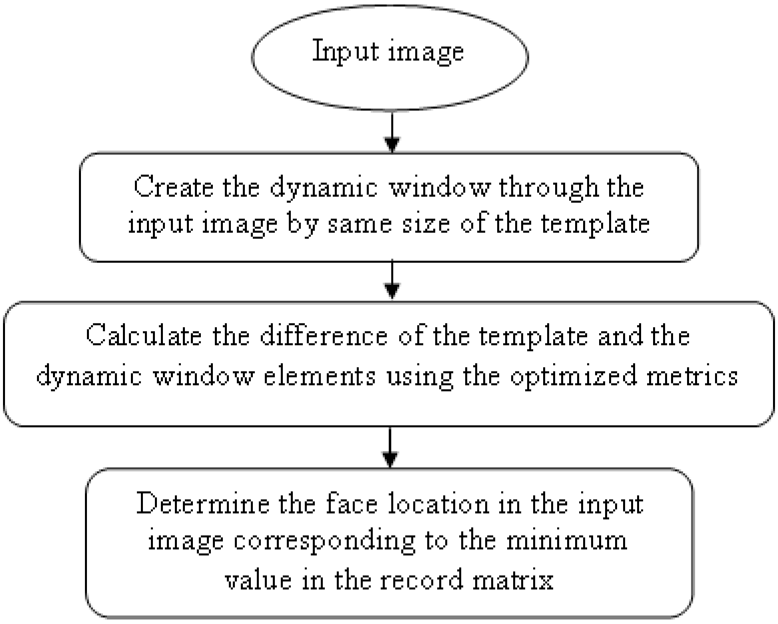


Figure 3-13: face localization using optimized metrics

3.2.3 Face Localization Using K-Means Modified algorithm

In this method, the face location is determined automatically by using K-means modified algorithm. First, the input image will be reshaped into vector of pixels values, followed by applying certain threshold determined by the algorithm to cluster the pixels into two classes. The pixels values belong to the class less than threshold which contains non-face part will be assigned to 0 values. However, some unwanted parts will be clustered to the second class. Therefore, the algorithm will be applied again to obtain the class which contains only the face part. Before we introduce our methodology, we will demonstrate briefly the standard K-means algorithm and modified version which was proposed during this work.

K-means algorithm is a method used to classify the set of observations into number of clusters based on the nearest mean to these observations. Suppose there are set of observations (x_1, x_2, x_3, x_4… x_n), so K-means will partition n observations to k clusters ($k \leq n$), then an initial K means (m_1… m_k) for these clusters will be created. These means will be chosen randomly and the following steps will be sequentially followed [88]:

1. Assignment step:

In this step each observation will be assigned to cluster with the nearest mean. This assignment is based on the result of the Voronoi diagram created by initial means; equation (3.13) shows that:

$$S_i = \{x_j : \left\| x_j - m_i \right\| \le \left\| x_j - m_{i^*} \right\| \quad \text{for all } i^* = 1......k\} \tag{3.20}$$

2. Update step:

In this step, the centroid of each class will be the new mean of the class. Equation (3.21) shows that:

$$m_i = \frac{1}{|s_i|} \sum_{x_j \in s_i} x_j \text{ , where } |s_i| \text{ in the cardinality of set } s_i \tag{3.21}$$

These two steps will be repeated until convergence has been reached. The methodology of the standard K-means algorithm will be shown in following diagram (see Figure 3.14):

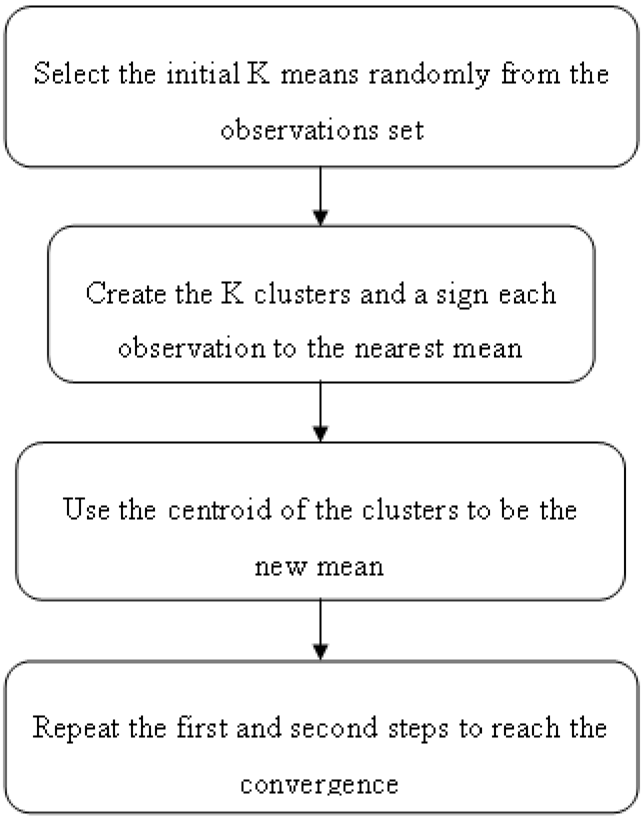


Figure 3-14: The methodology of the standard K-means algorithm

From the previous steps it's clear that the algorithm will not reach to the optimum convergence because it relies on the initial clusters. In addition, the algorithm is vast but it needs to be repeated many times by using different initial means. This problem makes the algorithm slow and not stable to reach the convergence

K-means modified algorithm will solve these limitations of standard version by using differential equations to determine the optimum separation point M. As an example, if we would like to cluster an image into two sub-classes, so we look for M_x and M_y which will separate between the clusters in the two dimensions then the means of the clusters can be found easy, Figure 3.15 shows the separation points M_x and M_y and the mean of each class. To find these points we proposed the following modification in continues and discrete cases:

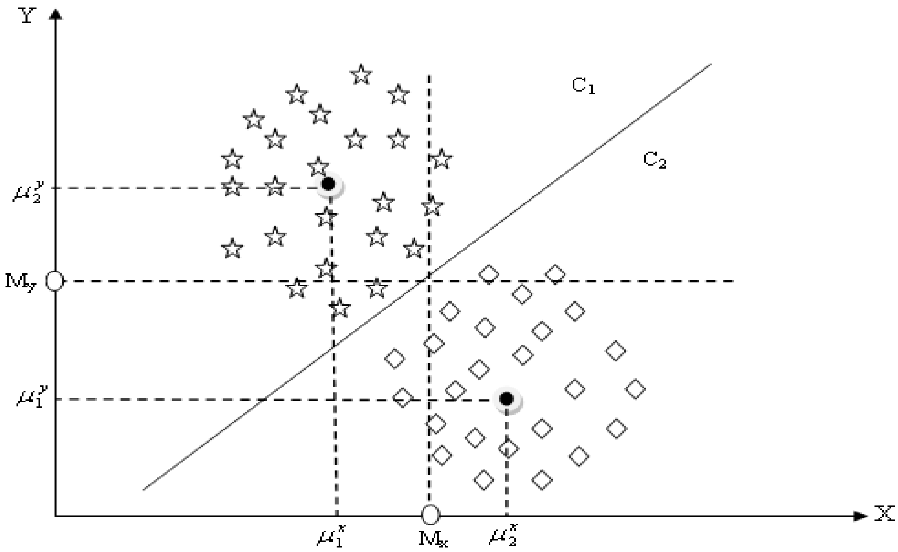


Figure 3-15: The separation points in two dimensions

If we have a number of observations x_j, let f(x) is the probability mass function (PMF) of x.

$$f(x) = \sum_{-\infty}^{\infty} x_i \delta(x - x_i) \ , \quad x_i \in \Re, \tag{3.22}$$

where $\delta(x)$ is Dirac function,

The aim is to get a minimum of

$$Var_T = \sum_{i=1}^{k} \sum_{x_i \in s_i} \left| x_i - \mu_i \right|^2, \tag{3.23}$$

where the μ_i is mean of cluster s_i. Then, it's enough to minimize that equation for dimension 1 without losing the generality

$$\min_{s_i} \sum_{i=1}^{k} \sum_{x_i \in s_i} \left| x_i - \mu_i \right|^2 = \sum_{k=1}^{d} \mathrm{Min}(\sum_{i=1}^{k} \sum_{x_j^k \in s_i} \left| x_j^k - \mu_j^k \right|), \tag{3.24}$$

where d is the dimension, $x_j = (x_j^1, \ldots, x_j^d)$, $\mu_i = (\mu_i^1, \ldots, \mu_i^j)$

a) For continuous case, mean f (x) is like the probability density function (PDF). In case of 2 classes, we need to minimize the following equation:

$$Var_T = \int_{-\infty}^{M} (x-\mu_1)^2 f(x)dx + \int_{M}^{\infty} (x-\mu_2)^2 f(x)dx \tag{3.25}$$

So if $d \geq 2$ we can use the equation (3.24)

Because $\mu_1^1 \perp \mu_1^2$, $M^1 \perp M^2$

$$Min\ \ Var_T = \sum_{k=1}^{d} Min\left(\int_{-\infty}^{M} (x^k - \mu_1^k)^2 f(x^k)dx + \int_{-\infty}^{M} (x^k - \mu_2^k)^2 f(x^k)dx \right) \tag{3.26}$$

Let x be random variable $x \in R^d$ with probability density function $f(x)$, we want to find μ_1 and μ_2 such that minimized Equation 3.23 as follow:

$$(\mu_1^*, \mu_2^*) = \arg_{(\mu_1,\mu_2)} \left(\min_{M,\mu_1,\mu_2} \int_{-\infty}^{M} (x-\mu_1)^2 f(x)dx + \int_{M}^{\infty} (x-\mu_2)^2 f(x)dx \right) \tag{3.27}$$

$$\min_{M,\mu_1,\mu_2} Var_D = \min_M \left[\min_{\mu_1} \int_{-\infty}^{M} (x-\mu_1)^2 f(x)dx + \min_{\mu_2} \int_{M}^{\infty} (x-\mu_2)^2 f(x)dx \right] \tag{3.28}$$

$$\min_{M,\mu_1,\mu_2} Var_D = \min_M \left[\min_{\mu_1} \int_{-\infty}^{M} (x-\mu_1(M))^2 f(x)dx + \min_{\mu_2} \int_{M}^{\infty} (x-\mu_2(M))^2 f(x)dx) \right] \tag{3.29}$$

As we know that:

$Min_x E[(X-x)^2] = E[(X-E(x)^{\cdot})^2] = Var(X)$............(*)

So we can find μ_1 and μ_2 in term of M as following:

$$Var_1 = \int_{-\infty}^{M} (x-\mu_1)^2 f(x)dx \ \ \& \ \ Var_2 = \int_{M}^{\infty} (x-\mu_2)^2 f(x)dx \tag{3.30}$$

$$Var_1 = \int_{-\infty}^{M} x^2 f(x)d(x) + \mu_1^2 \int_{-\infty}^{M} f(x)d(x) - 2\mu_1 \int_{-\infty}^{M} xf(x)d(x) \tag{3.31}$$

And

$$Var_2 = \int_{M}^{\infty} x^2 f(x)d(x) + \mu_2^2 \int_{M}^{\infty} f(x)d(x) - 2\mu_2 \int_{M}^{\infty} xf(x)d(x) \tag{3.32}$$

From $(*)$ we conclude $\begin{cases} \mu_1 = \dfrac{E(X_-)}{P(X_-)} \\ \mu_2 = \dfrac{E(X_+)}{P(X_+)} \end{cases}$

where
$$\begin{aligned} P(X_-) &= \int_{-\infty}^{M} f(x)dx \\ P(X_+) &= 1 - P(X_-) \\ E(X_-) &= \int_{-\infty}^{M} nf(x)dx \\ E(X_+) &= E(X) - E(X_-) \end{aligned}$$

After derivation of μ_i in term of M, we get:

$$\begin{cases} \dfrac{\partial \mu_1}{\partial M} = \left(\dfrac{\int_{-\infty}^{M} xf(x)}{\int_{-\infty}^{M} f(x)} \right) \Rightarrow \dfrac{\partial \mu_1}{\partial M} = \dfrac{Mf(M).P(X_-) - f(M).E(X_-)}{P^2(X_-)} \\ \dfrac{\partial \mu_2}{\partial M} = \left(\dfrac{\int_{M}^{\infty} xf(x)}{\int_{M}^{\infty} f(x)} \right) \Rightarrow \dfrac{\partial \mu_1}{\partial M} = \dfrac{Mf(M).P(X_+) - f(M).E(X_+)}{P^2(X_+)} \end{cases}.$$

Implies that $\begin{cases} \dfrac{\partial \mu_1}{\partial M} = \dfrac{f(M)}{P^2(X_-)}[MP(X_-) - E(X_-)] \\ \dfrac{\partial \mu_2}{\partial M} = \dfrac{f(M)}{P^2(X_+)}[MP(X_+) - E(X_+)] \end{cases}$

$$\frac{\partial Var_1}{\partial M} = M^2 f(M) + 2\mu_1 \mu_1' P(X_-) + \mu_1^2 f(M) - 2\mu_1' E(X_-) + 2\mu_1 Mf(M) \tag{3.33}$$

After simplification, we get

$$\frac{\partial Var_1}{\partial M} = f(M)\left[M^2 + \frac{E^2(X_-)}{P(X_-)^2} - 2M \frac{E(X_-)}{P(X_-)^2} \right] \tag{3.34}$$

And the derivative of Var_2 is:

$$\frac{\partial Var_2}{\partial M} = -M^2 f(M) + 2\mu_2 \mu_2' P(X_+) + \mu_2^2 f(M) - 2\mu_2' E(X_+) - 2\mu_2 Mf(M) \tag{3.35}$$

$$\frac{\partial Var_2}{\partial M} = f(M)\left[-M^2 + \frac{E^2(X_+)}{P(X_+)^2} - 2M \frac{E(X_+)}{P(X_+)^2} \right] \tag{3.36}$$

From (3.35) and (3.36), we have

$$\frac{\partial Var_T}{\partial M} = f(M)\left[\frac{E^2(X_+)}{P(X_+)^2} - \frac{E^2(X_-)}{P(X_-)^2} + 2M\left[\frac{E(X_-)}{P(X_-)} - \frac{E(X_+)}{P(X_+)} \right] \right] \tag{3.37}$$

When $\dfrac{\partial Var_T}{\partial M} \Rightarrow 0$

$$\left[\frac{E^2(X_-)}{P(X_-)^2}-\frac{E^2(X_+)}{P(X_+)^2}+2M\left(\frac{E(X_+)}{P(X_+)}-\frac{E(X_-)}{P(X_-)}\right)\right]=0$$

So

$$\left[-\left(\frac{E(X_+)}{P(X_+)}+\frac{E(X_-)}{P(X_-)}\right)+2M\right]\left(\frac{E(X_+)}{P(X_+)}-\frac{E(X_-)}{P(X_-)}\right)=0$$

$M=\left[\frac{E(X_+)}{P(X_+)^2}+\frac{E(X_-)}{P(X_-)^2}\right].\frac{1}{2}$ give the minimum not the maximum

Then

$$M=\frac{(\mu_1+\mu_2)}{2} \tag{3.38}$$

Therefore, we have to find all M that verify equation (3.39), which it easy than minimizing Var_T directly.

$$\mu_1(M)=\frac{E(X.1_{X<M})}{\mathrm{P(X<M)}}\ \&\ \mu_2(M)=\frac{E(X.1_{X\geq M})}{\mathrm{P(X\geq M)}}$$

We call $X.1_{X<M}=X_-$ and $X.1_{X\geq M}=X_+$.

To find the minimum in terms of M, it's enough to derivate:

$$\left[Var_D(\mu_1(M),\mu_2(M))\right]'=f(M)\left[-\frac{E(X_+)}{P(X<M)}-\frac{E(X_-)}{P(X\geq M)}+2M\right].\left[\frac{E(X_+)}{P(X>M)}+\frac{E(X_-)}{P(X\leq M)}\right]$$

Then

$$M=\left[\frac{E(X.1\ \ X\geq M)}{\mathrm{P(X\geq M)}}+\frac{E(X.1\ \ X<M)}{\mathrm{P(X<M)}}\right]/2$$

But if X is discrete random variables with P_i as mass density function after calculation we find that ΔVar_D is almost equal to the $Var_D(M)$ in the continues case under some conditions of approximation as follows:

Let $x_1<M_1<x_2<M_2<x_3, x_i\in\text{database}$

$$Var_D(M_1) = \sum_{I_1}(x_i - \mu_1)^2 \frac{P_i}{N} + \sum_{I_2}(x_i - \mu_2)^2 \frac{P_i}{N} \tag{3.39}$$

where $I_1 = \{i : x_i : x_1\} \& I_2 = \{i : x_i : x_2\}$

$$Var_D(M_2) = \sum_{I_1}(x_i - \mu_1)^2 \frac{P_i}{N} + \sum_{I_3}(x_i - \mu_2)^2 \frac{P_i}{N} \tag{3.40}$$

where N is data size and $I_3 = \{i : x_i : x_3\}$

$$\Delta Var_D(x_1) = \left[\frac{E^2(X \le x_1)}{P(X \le x_1)} + \frac{E^2(X \ge x_2)}{P(X \ge x_2)}\right] - \left[\frac{E^2(X \le x_2)}{P(X \le x_2)} + \frac{E^2(X \ge x_3)}{P(X \ge x_3)}\right] \tag{3.41}$$

$E(x \le x_1) \equiv E(x.1_{X \le x_1})$ and so on for other cases

$$\Delta Var_D(x_1) = \left[\frac{E^2(X \le x_1)}{P(X \le x_1)} - \frac{E^2(X \le x_2)}{P(X \le x_2)}\right] + \left[\frac{E^2(X \ge x_2)}{P(X \ge x_2)} - \frac{E^2(X \ge x_3)}{P(X \ge x_3)}\right] \tag{3.42}$$

If we consider $N >> P_2 \,\&\, P(X \ge x_1) + \frac{P_2}{N} \approx P(X \ge x_1)$ and $P(X \ge x_3) + \frac{P_2}{N} \approx P(X \ge x_3)$

$$\begin{aligned}\Delta Var_D(x_1) \approx & \left[\frac{E^2(X \le x_1)P_2}{NP^2(X \le x_1)} - \frac{E^2(X \ge x_3)P_2}{NP^2(X \ge x_3)}\right] + \frac{2P_2x_2}{N}\left[\frac{E(X \ge x_3)}{P(X \ge x_3)} - \frac{E(X \le x_3)}{P(X \le x_3)}\right] \\ & + \frac{P_2^2x_2^2}{N^2}\left[\frac{1}{P(X \ge x_3)} - \frac{1}{P(X \le x_3)}\right],\end{aligned} \tag{3.43}$$

when $\frac{P_2^2x_2^2}{N^2}\left[\frac{1}{P(X \ge x_3)} - \frac{1}{P(X \le x_3)}\right] = 0 \quad N \to \infty$

So

$$\Delta Var_D(x_1) \approx \frac{P_2}{N}\left[\frac{E(X \ge x_3)P_2}{P(X \ge x_3)} - \frac{E(X \le x_1)P_2}{P(X \le x_1)}\right]\left[2x_2 - \left[\frac{E(X \ge x_3)}{P(X \ge x_3)} - \frac{E(X \le x_1)}{P(X \le x_1)}\right]\right] \tag{3.44}$$

To find the minimum for the variation total it is enough to find the good separator M between x_i and x_{i+1} such that $\Delta Var_D(x_i) * \Delta Var_D(x_{i+1}) < 0$ and $\Delta Var_D(x_i) < 0$, (see image 3.16)

After finding few minimum locals, we can get minimum global easily.

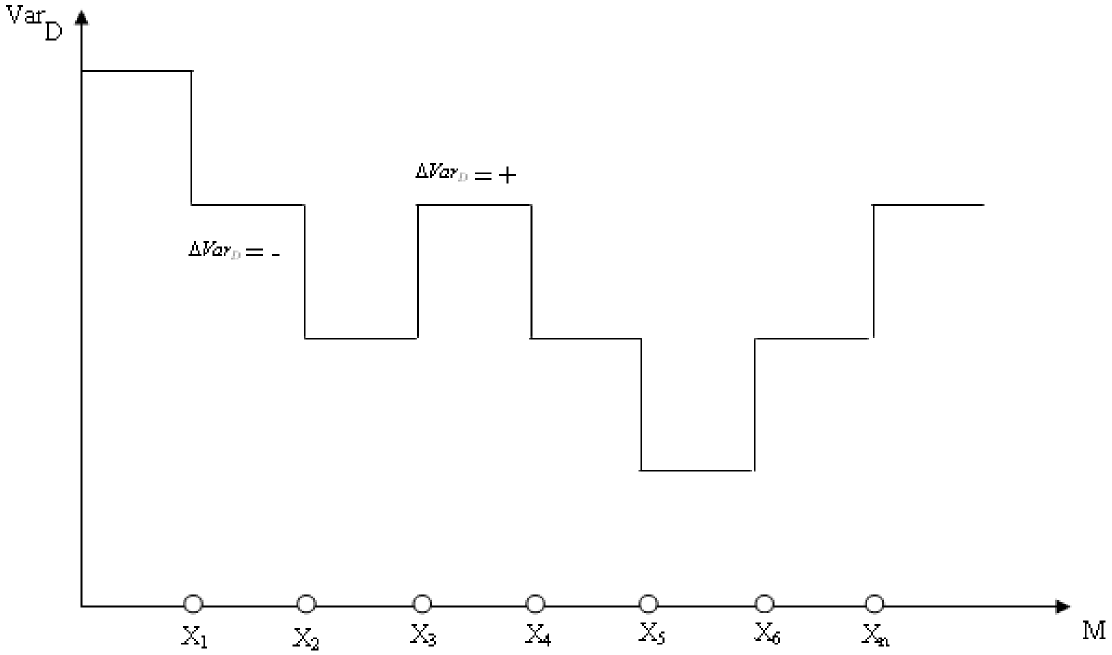


Figure 3-16: Discrete random variables

Now, it becomes a clear the superiority of K-means modified algorithm over the standard algorithm in term of finding global minimum and we can explain our solution based on the modified algorithm. Suppose we have an image with pixels values from 0 to 255. First we reshaped the input image into vector of pixels; Figure 3.17 shows the input image before the operation of K-Means modified algorithm which contains face part, normal background and background with illumination for example. Then we split the image pixels into two classes from 0 to x_1 and from x_1 to 255 by applying a certain threshold determined by the algorithm.

$[0 \quad x_1]$ is class 1 which represents the non-face part.

$[x_1 \quad 255]$ is class 2 which represents the face with the shade.

Figure 3-17: The original image

All pixels values belong to the first class with less than the threshold will be signed to the 0 values and we will keep the pixels above the threshold which contain the face part and background with illumination, Figure 3.18 shows that.

Figure 3-18: The face with shade only

In order to remove the illumination, another threshold will be applied again on the pixels using the algorithm to separate between the face part and illumination part. A new classes will be obtained and let are from x_1 to x_2 and from x_2 to 255. Then, the pixels values with less than the threshold will be signed to 0 values and we keep the pixels from the second class which contains the face part.

[x_1 x_2] is class 1 which represents the illumination part.

[x_2 255] is class 1 which represents the face part.

The removing of the illumination part will be shown in Figure 3.19, but there are some faces pixels are signed to the 0 values due to the affects of the illumination on the pixels values. Therefore, there is a need to return to the original image Figure 3.17 to extract the corresponding image window of the second class with pixels above than the second threshold. Figure 3.20 shows the corresponding window the original image which contain the face part. At the last step before ROI detection, we pass the result image to filter in order to remove the small noise.

Figure 3-19: The face without illumination

Figure 3-20: The corresponding window in the original image

Now, the proposed method can be summarized as shown in the diagram (see Figure 3.21):

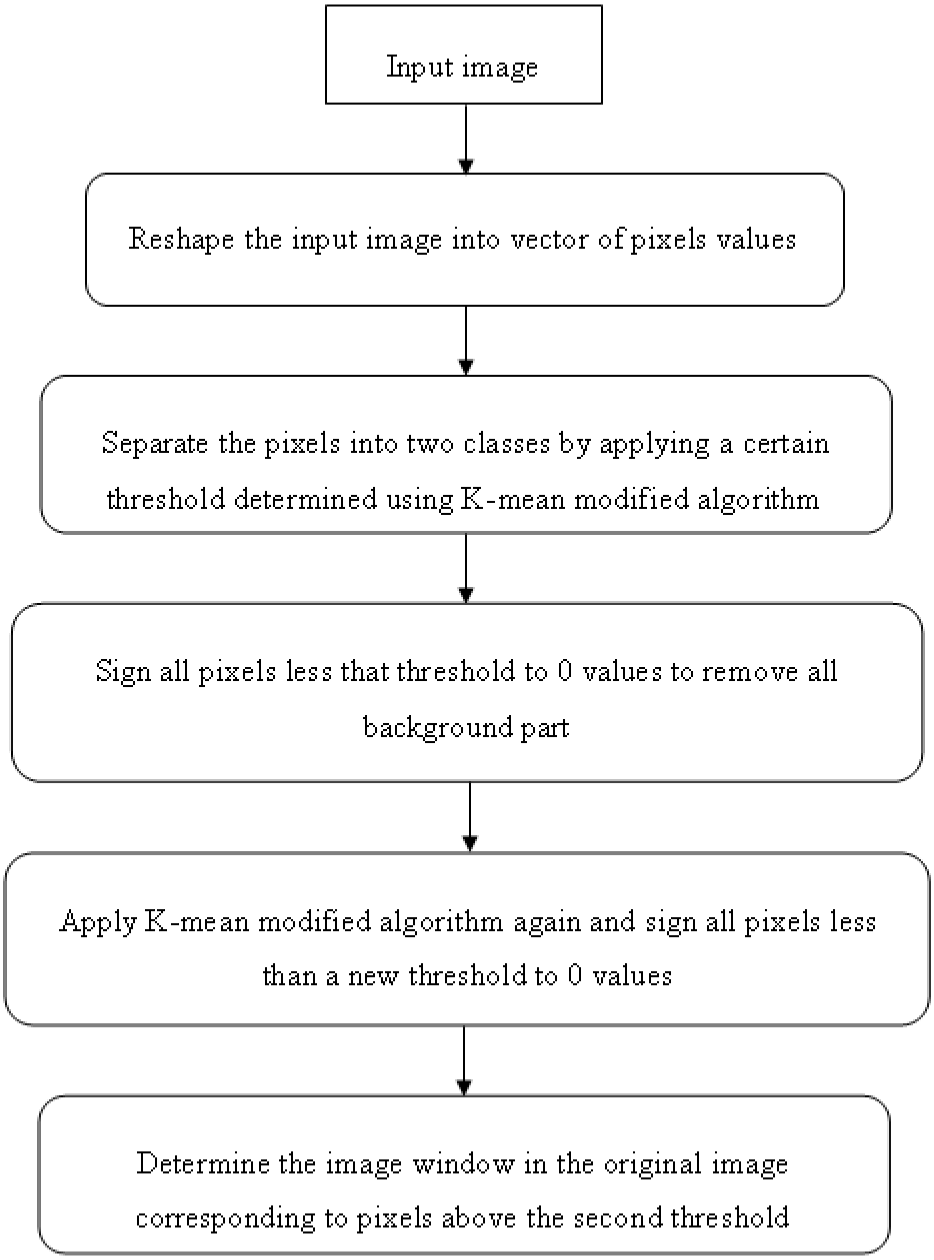


Figure 3-21: Face localization using K-mean modified algorithm

3.3 Part II: Features Extraction

3.3.1 Introduction

Features extraction is the second step in any biometrics authentication system. The extraction and selection of features which can give good face representation and invariant in different variations is a very important issue. Many statistical methods were proposed to extract the invariant facial features such as Principle Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Factor Analysis (FA). Also there is another type of features extraction by using transformations methods and it is widely used to provide robust features against variations such as Fourier Transform (FT) and Discrete Cosine Transform (DCT). Recently, multi-resolution methods such as Discrete Wavelet Transform (DWT) and curvelet have been widely used in many applications due to multi-resolution quality comparing with other transformation methods. In the next sections we will introduce a definition of the wavelet transform in one or two dimension and the comparison with Fourier Transform will be included as well. As DWT gives a huge number of coefficients, we need to introduce statistical solution to reduce the wavelet coefficients redundancy and to select an invariant features. This statistical solution is based on the selection of features with high energy to reduce the dimensions and then select the variance features to represent the face and distinguish between the face classes. The solution is better than using PCA and LDA because it maintains the shape of the face which increases the classification accuracy, less complexity and very fast.

3.3.2 Wavelet Transform

The transform of a signal is considered as a different way of representing the signal. And it saves the information content present in the signal. Wavelets make use of different sets of basis functions to permit the decomposition of continuous and discrete signals. Wavelet Transform offers a time-frequency representation of the signal. It was built up to rise above the short coming of the Short Time Fourier Transform (STFT), which can also be used to analyze non-stationary signals. While

STFT provides a constant resolution at all frequencies, the Wavelet Transform utilizes multi-resolution technique by which different frequencies are analyzed with different resolutions. A wave is a fluctuating function of time or space and is cyclic. Unlike the wavelets are different and localized waves. Also another difference, wavelets have their energy concentrated in time or space and are appropriate to analysis of temporary signals. While Fourier Transform and STFT utilize waves to analyze signals, the Wavelet Transform (WT) uses wavelets of finite energy. The signal which expects to analyze is multiplied with wavelet function and then the transform is calculated for each part created and that is not like the STFT. Furthermore in the wavelet the width of the wavelet function varies with each spectral component. At high frequencies, wavelet transform gives good time resolution and poor frequency resolution, while at low frequencies; gives good frequency resolution and poor time resolution. Figure 3.22 illustrates the differentiation between the wave and wavelet:

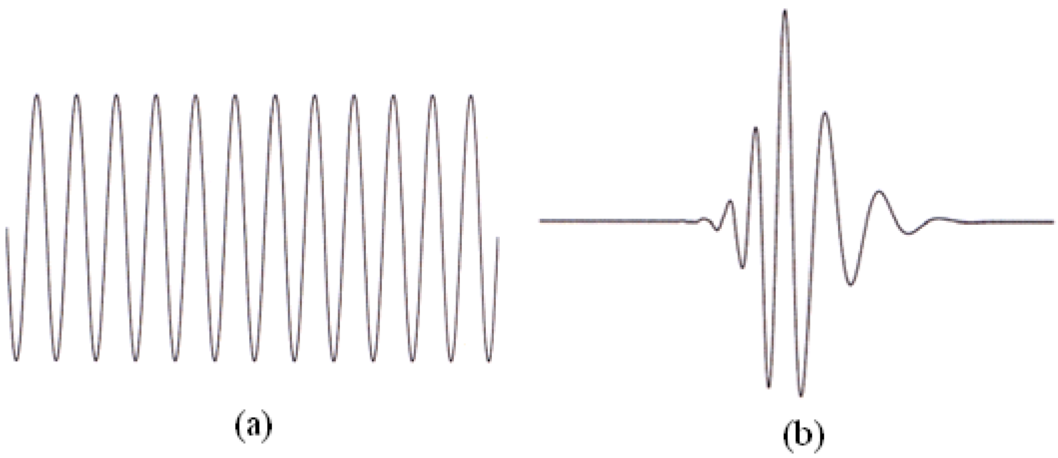


Figure 3-22: Demonstration of (a) The Wave, (b) The Wavelet

From [89], in equation 3.45, the Continuous Wavelet Transform (CWT) is provided where x (t) is the signal to be analyzed. Ψ (t) is the mother wavelet or the basis function. The mother wavelet is considered as source of all the wavelet functions used in the transformation during translation (shifting) and scaling (dilation or compression).

$$X_{WT}(\tau, s) = \frac{1}{\sqrt{|s|}} \int x(t).\Psi^{*}\left(\frac{t-\tau}{s}\right) dt \tag{3.45}$$

All the basic functions are generated from the mother wavelet used and designed based on some most wanted characteristics associated with that function. The two parameters, translation parameter τ and scale parameter s are defined as related to the position of the wavelet function as it is shifted through the signal and corresponds to the time information in the Wavelet Transform. For the parameter τ is defined as |1/frequency| that corresponds to frequency information. The scaling either dilates (expands) or compresses a signal. Large scales expand the signal and gives detailed disappeared information in the signal, whereas small scales compress the signal and give general information about the signal. Another approach, the Wavelet Series is a sampled version of CWT and its calculation may need large amount of time and resources, which is based on the resolution requested. The Discrete Wavelet Transform (DWT), which is based on sub-band coding, is found in [89] and is not like (CWT) its fast computation of Wavelet Transform. Also it is easy to execute and reduce the computation time and resources requested. When we want to compare between CWT and DWT in hand of signal analysis we found the signals are analyzed using a set of basic functions in CWT which relate to each other by simple scaling and translation. But in the case of DWT, we can obtain the time-scale representation of the digital signal using digital filtering techniques. The signal we want to analyze is passed through filters with different cut-off frequencies at different scales.

In signal processing the filters can be considered as one of the most widely used. If we repeat the filters with rescaling we can realize the Wavelets. The amount of the information in the signal is measured by the resolution of signal which is determined by the filter operation; also the scale has been determined by the up-sampling and down-sampling (sub-sampling) operations. The DWT is calculated by in a row low-pass and high-pass filtering of the discrete time-domain signal as shown in Figure 3.23. This is called the Mallat algorithm or Mallat-tree decomposition. Its worth is in the approach where it joins the continuous-time multi-resolution to discrete-time filters. From the Figure 3.23, we can observe that the sequence x[n] represents the signal, where n is an integer. G_0 represents the low pass filter as well as H_0 represent the high pass filters. The detail information is produced by the high pass filter at each level; d[n], while uncouth approximations, a[n] is produced by the low pass filter associated with scaling function.

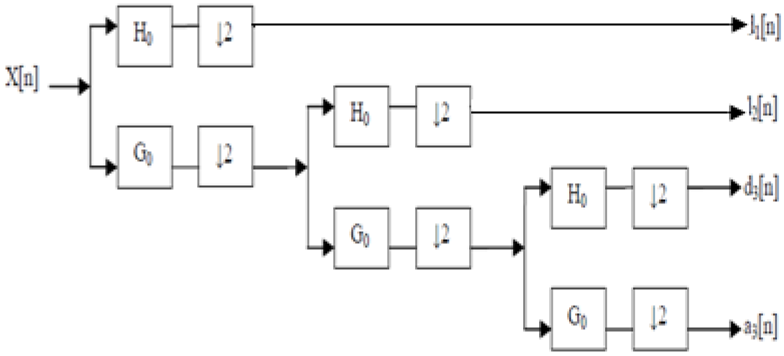


Figure 3-23: Three-level wavelet decomposition tree

For one dimensional wavelet representation, let Z and R denote the set of integers and real number respectively. The multiresolution approximation of one dimensional signal $f(x) \in L^2(R)$ at a resolution 2^j is defined as the orthogonal projection of a signal on subspace V_{2^j} of $L^2(R)$. V_{2^j} can be interpreted as set of all possible approximation at the resolution 2^j of function in $L^2(R)$. The set of vector spaces V_{2^j} is said to be a multiresolution representation of $L^2(R)$ if it satisfies some properties [90].

Let O_{2^j} be the vector space that satisfies that: O_{2^j} is orthogonal to V_{2^j} and $O_{2^j} \oplus V_{2^j} = V_{2^{j+1}}$ i.e. the orthogonal complement of V_{2^j} in $V_{2^{j+1}}$. The approximation $A_{2^{j+1}} f(x)$ at resolution 2^{j+1} contains more information than the approximation $A_{2^j} f(x)$ at resolution 2^j. The details signal of $f(x)$ at resolution 2^j are denoted by $D_{2^j} f(x)$. The details can be defined as different between $A_{2^{j+1}} f(x)$ and $A_{2^j} f(x)$. $D_{2^j} f(x)$ is equivalent to the orthogonal projection of $f(x)$ on the complement O_{2^j} of vector space V_{2^j} in $V_{2^{j+1}}$. According to the theory of multiresolution signal decomposition [90], there exists a unique scaling function $\varphi(x) \in L^2(R)$ and a unique corresponding wavelet function $\psi(x) \in L^2(R)$, where $\varphi_{2^j}(x) = 2^j \varphi(2^j x)$ and $\psi_{2^j}(x) = 2^j \psi(2^j x)$, such that $\{2^{\frac{-j}{2}} \varphi_{2^j}(x - 2^{-j} k)\}_{k \in Z}$ and $\{2^{\frac{-j}{2}} \psi_{2^j}(x - 2^{-j} k)\}_{k \in Z}$ are orthogonal bases of O_{2^j} and V_{2^j} respectively. The approximation is characterized by the sequence of inner products of $f(x)$ with φ_{2^j} and ψ_{2^j} as follows:

$$\{A_{2^j} f(k)\}_{k\in Z} = \{\langle f(0), \varphi_{2^j}(0-2^{-j}k)\rangle\}_{k\in Z} \tag{3.46}$$

$$\{D_{2^j} f(k)\}_{k\in Z} = \{\langle f(0), \psi_{2^j}(0-2^{-j}k)\rangle\}_{k\in Z} \tag{3.47}$$

Let H be a low-pass filter and G be a high-pass filter, where the impulse response of the filter H is $h(k) = \langle \varphi_{-1}(x), \varphi(x-k)\rangle$, and the impulse response of the filter G is $g(k) = \langle \psi_{-1}(x), \psi(x-k)\rangle$. Define $\tilde{H}$ with impulse response $\tilde{h}(k) = h(-k)$ to be the mirror filter of H, and $\tilde{G}$ with impulse response $\tilde{g}(k) = g(-k)$ to be the mirror filter of G. The multiresolution representation of $f(x)$ at any resolution 2^j can be implemented by a pyramidal algorithm as shown in Figure 3.24:

$$A_{2^{j-1}} f(x) = \sum_{k=-\infty}^{\infty} \tilde{h}(2x-k) A_{2^j} f(k) \qquad \text{where j} = 0,\text{-}1,\text{-}2,.. \tag{3.48}$$

$$D_{2^{j-1}} f(x) = \sum_{k=-\infty}^{\infty} \tilde{g}(2x-k) A_{2^j} f(k) \qquad \text{where j} = 0,\text{-}1,\text{-}2,.. \tag{3.49}$$

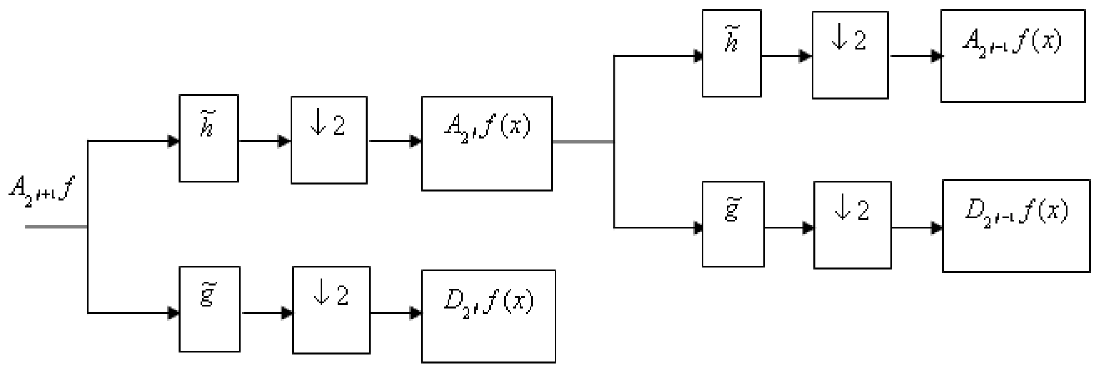


Figure 3-24: One dimensional wavelet decomposition

The wavelet model can be extended to two-dimensional signals by separable multiresolution approximation of $L^2(R^2)$ with scaling function $\varphi(x,y) = \varphi(x)\varphi(y)$. And $\psi(x)$ is the one dimensional wavelet function associated with $\varphi(x)$. There are three associated wavelet functions $\psi^1(x,y) = \varphi(x)\psi(y)$, $\psi^2(x,y) = \psi(x)\varphi(y)$ and $\psi^3(x,y) = \psi(x)\psi(y)$. With this formulation, the wavelet decomposition of a two dimensional signal can be computed with a separable extension of the one-dimensional decomposition algorithm as shown in Figure 3.25.

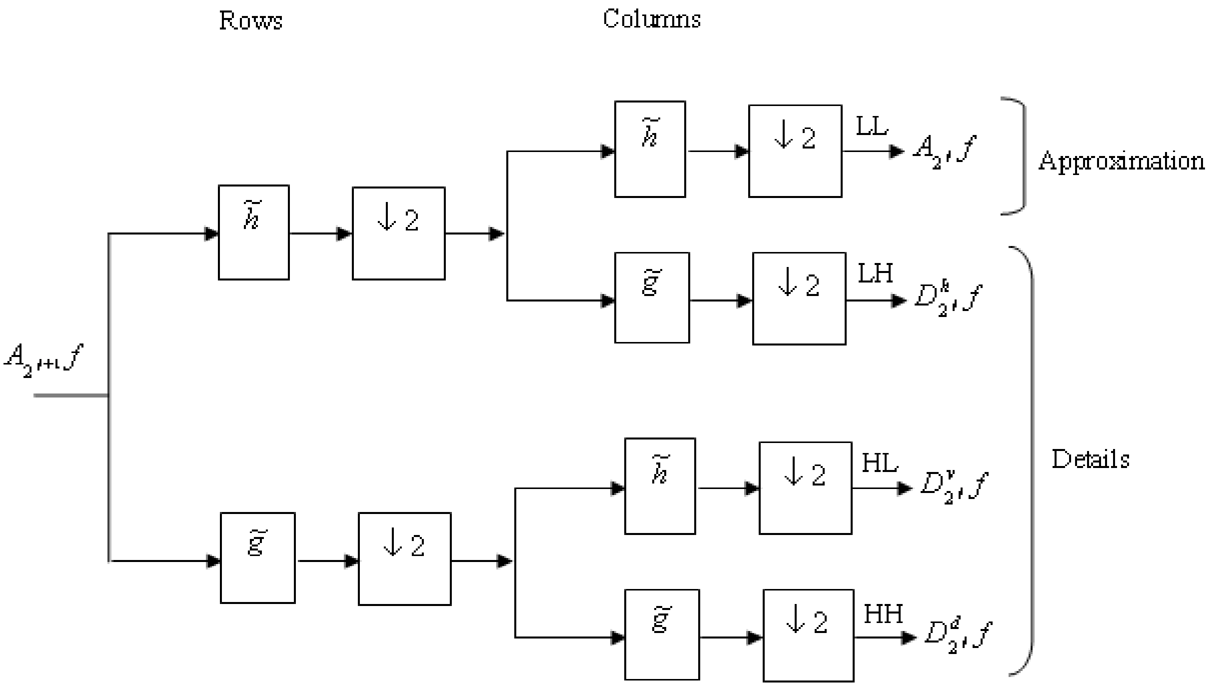


Figure 3-25: Two dimensional wavelet decomposition

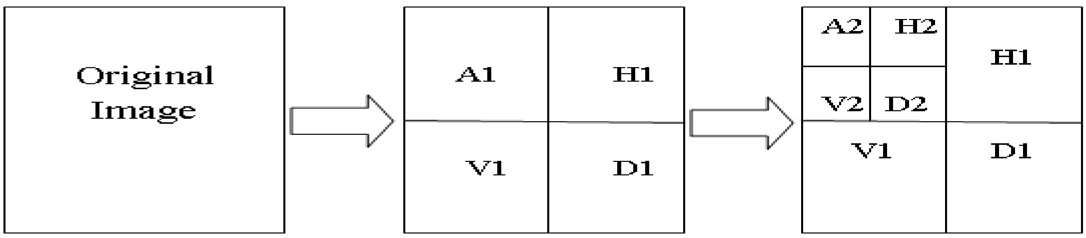


Figure 3-26: Two levels wavelet decomposition

Figure 3.26 shows the decomposition of image $A_{2^{j+1}}f$ into $A_{2^j}f$, $D^h_{2^j}f$, $D^v_{2^j}f$ and $D^d_{2^j}f$ in the frequency domain. The images $A_{2^j}f$, $D^h_{2^j}f$, $D^v_{2^j}f$ and $D^d_{2^j}f$ corresponding to the lowest frequency, the vertical high frequency (horizontal edges), the horizontal high frequency (vertical edges) and high frequency in both directions (diagonal) respectively. i.e., the image $A_{2^{j+1}}f = A_{2^j}f + D^h_{2^j}f + D^v_{2^j}f + D^d_{2^j}f$. This set of images is called an orthogonal wavelet representation in two dimensions [90]. The image $A_{2^j}f$

is the approximation at the resolution 2^j, and the images $D^h_{2^j} f$, $D^v_{2^j} f$ and $D^d_{2^j} f$ give the detail signals for different orientations and resolutions. If the original image has N pixel, then each of the images $D^h_{2^j} f$, $D^v_{2^j} f$ and $D^d_{2^j} f$ will have $2^j N$ pixels $(j < 0)$, so that the total number of pixels in this new representation is equal to the number of pixels of the original image, to keep the volume of data maintained. This process can be summarized as, wavelet decompose an image into orthogonal sub-bands with low-low (LL), low-high (LH), high-low (HL) and high-high (HH) components which correspond to approximation, horizontal, vertical and diagonal respectively. The LL sub-band is further decomposed into another four sub-bands low-low-low-low (LLLL) component, which represents the image approximation at this level and then it is decomposed once again and so on.

3.3.3 Statistical Solution for Features Selection

The reduction of wavelet transform coefficients to represent the face in the image is the most significant problem due to the redundancy of features which are not required to the classification. In the case of face classification, some of these WT coefficients do not have face information that leads to increase the error rate of matching. Therefore, it is important to reduce the coefficients by selecting those coefficients that contains face information and ignoring the remaining. In addition, it is necessary that to create a metric to quantify the classification contribution of each coefficients. For example Figure 3.27, shows three classes in the two dimensions X axis and Y axis.

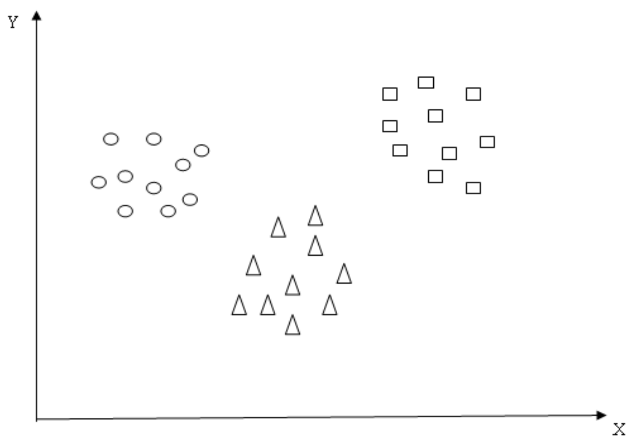


Figure 3-27: Three classes in two dimensions

Now suppose m_1, m_2 and m_3 are the mean of class$_1$, class$_2$, and class$_3$ respectively and m_T is mean of all classes. There are two cases of features with/without overlapping between classes as shown in Figures 3.28 and 3.29:

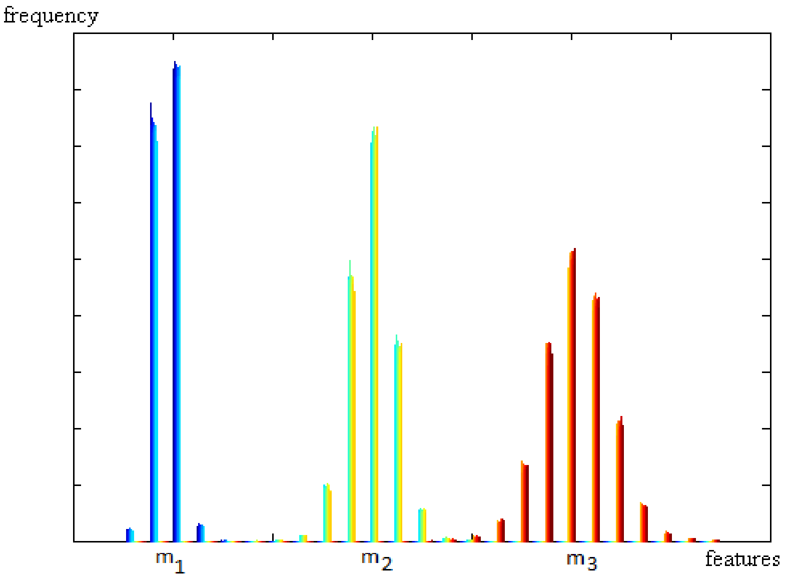


Figure 3-28: Case 1: Histogram of three classes without overlapping, means it good features for classification

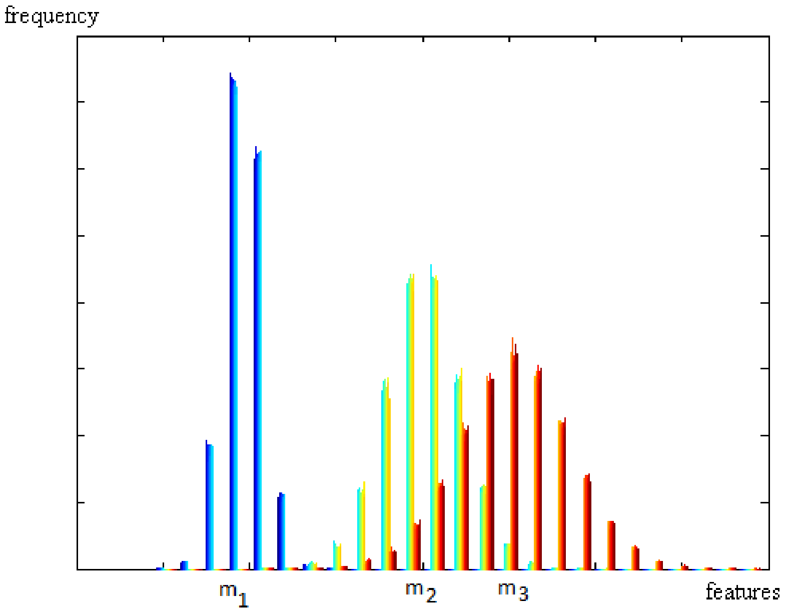


Figure 3-29: Case 2: Histogram of three classes with overlapping, means it bad features for classification

Let $m_T = \frac{\sum_{i=1}^{n} m_i}{n}$ $n = numberofclasses$, from Figures 3.28 and 3.29 show that $\sum (m_i - m_T)^2$ is not sufficient to quantify the classification contribution of the coefficients because its may give same values in the two cases. Therefore, there is a need to introduce another metric to quantify the coefficients contribution. We introduce another metric as follow:

$$Var\_\mathrm{mod} = \frac{1}{n}\left[\sum_{i=1}^{n} \frac{(m_i - m_T)^2}{Var_i}\right], \tag{3.50}$$

where Var_i is variance of the class i, m_i is mean of class i, m_T is mean of all classes and n is the number of classes. Var_i will be calculated using the following formula:

$$Var_i = \frac{\sum_{j=1}^{n_i} \left(x_j^i - m_i\right)^2}{n_i - 1} \qquad \mathrm{i} = 1, 2, 3, \ldots\ldots, \mathrm{n_i} \tag{3.51}$$

where n_i is number of the features in class i

The way to select the desired features coefficients will be as follow:

If variance modified of any feature is less than certain threshold $Var\_\mathrm{mod} \le 1$, we will remove it, else we keep feature. Figure 3.30 shows the original face image and the face image after removing unwanted features and also shows how the face still keep its shape.

Figure 3-30: The original face and reconstruction face by using suitable features

To calculate the error probability of classification we will take the example shown in Figure 3.31:

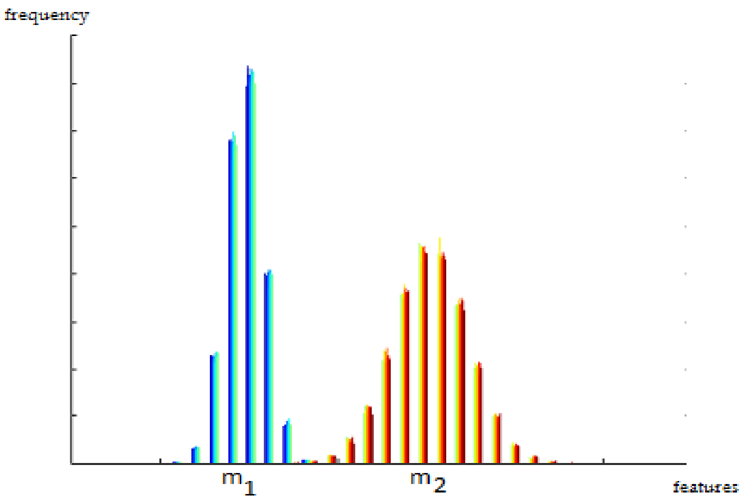


Figure 3-31: Histogram of the features coefficients of two classes'

From Figure 3.31, we can approximate the probability density function (PDF) of features coefficients of each class to be Gaussian density variables,

Let feature extraction of class 1~ Gaussian (m_1, σ_1) and feature extraction of class 2~ Gaussian (m_2, σ_2)

We estimate the expected value by the mean sample

$$m_j = \frac{\sum_{i=1}^{i} x_i^j}{n}, \tag{3.52}$$

where x_i^j is random variable, j^{th} is feature and i class is index.

If we us K-NN as classifier, the probability of error to classify the variable x is:

$$P(error) = \sum_{i=1}^{2} P(x \in class_i).P(error / x \in class_i)\ldots\ldots\ldots\ldots\ldots\ldots(*)$$

let assume $P(x \in class_i) = \frac{1}{2}$ then,

$$P(error / x \in class_1) = P\left(\frac{x - m_1}{\sigma_1} < \frac{m_T - m_1}{\sigma_1}\right) = P\left(N(0,1) < \frac{m_T - m_1}{\sigma_1}\right)$$

$$= P\left(N(0,1) > \frac{m_1 - m_T}{\sigma_1} \right) = \phi\left(\frac{m_1 - m_T}{\sigma_1} \right) \tag{3.53}$$

And we do same step for $P(error / x \in class_2)$

$$P(error / x \in class_2) = P(x < m_T / x \in class_2)$$

$$P(error / x \in class_2) = P\left[N(0,1) > \frac{m_T - m_2}{\sigma_2} \right] = \phi\left[\frac{m_T - m_2}{\sigma_2} \right] \tag{3.54}$$

So replace (3.53) and (3.54) in (*) we get

$$P(error) = \left[\phi\left[\frac{m_1 - m_T}{\sigma_1} \right] + \phi\left[\frac{m_T - m_2}{\sigma_2} \right] \right] / 2 \tag{3.55}$$

If we choose $\left(\frac{m_T - m_2}{\sigma_2} \right) > 1$ and $\left(\frac{m_1 - m_T}{\sigma_1} \right) > 1$

$$P(error) < \phi = 15\%$$

We can also use wavelet coefficients energy since it determine the percentage of information in these coefficients and that to select the relevant coefficients, for that we apply the statistical energy to reduce the dimensions of the overall features by selecting only the features with high statistical energy values, since these features contains high amount of face information more than the other features with low statistical energy values [78]. Therefore, a particular threshold will be applied to the all features, where the statistical energy of each feature X is sum square of the coefficient over all database.

All features that have statistical energy less than certain threshold will be removed and we kept wavelet coefficients with high statistical energy values. The following example will explain our proposed features selection method:

Table 3-1: Example for features selection method

(a)

classes	Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9	Column0
Class A	5	2	3	4	9	5	1	8	7	6
	1	3	4	6	8	7	5	1	9	3
	9	5	1	8	4	3	9	5	6	1
Class B	3	2	6	5	8	9	4	7	3	7
	2	4	8	9	7	6	3	8	4	6
	5	6	7	9	9	8	5	2	4	7
Class C	8	5	2	7	9	1	7	3	6	8
	6	8	2	4	1	7	9	2	3	4
	4	2	1	8	9	7	8	5	6	7

Calculate the energy of each column

(b)

energy	261	187	184	432	518	363	351	245	288	309

Apply for example threshold (270)

(c)

remains				432	518	363	351		288	309

Calculate the mean of each column in all classes of the remaining columns and the mean total of remaining

(d)

Mean A				9	7	5	5		7.3	3.3
Mean B				7.6	8	7.6	4		3.6	6.6
Mean C				6.3	6.3	5	8		5	6.3
Mean T				6.6	7.1	5.8	5.5		5.3	5.4

Calculate the variance of each class and then calculate the variance modify by using equation (3.39)

(e)

variance				2.2	2.6	1.7	1.5		1..2	1.4

Apply for example threshold (1.5)

(f)

remains				2.2	2.6	1.7	1.5			

Before we explain our example we would like to note that each input image with size (M x N) will be reshaped by vector (1 x N). Then we will obtain a new matrix (KxN) where K is the number of the input image and it is 9 in our example distributed among three classes A, B and C as in the first part (a). N is the number of coefficients in each image where it is 10 in above example. The energy of each column is calculated as shown in the second part (b). Let the threshold value applied in this example is (270), the remaining six columns (4, 5, 6, 7, 9, and 10) will be kept to present to the variance modify in the third part (c). The mean of each column in each class and the total mean of each class of remaining columns are calculated as shown in the fourth part (d). The variance modify is calculated as shown in the part fifth (e). Let the threshold value applied in the example is (1.5), the remaining four columns (4, 5, 6 and 7) will be kept to be present to the classifier in the next part (f).

3.4 Classification Process

The matching process is an important step which determines the systems decision. Therefore, the classifier should be robust and more accurate to classify the facial features. We can add other conditions to make perfect classifier like minimum classification time and smaller size of training data. In the next section we will introduce a new classifier that is combined between K-means modified algorithm and K-Near Neighbor (K-NN). This classifier was developed by [79] and it was introduced to classify gases and data for security systems. In addition, the classification performance of C-K-NN classifier is compared with most popular classifiers which are K-NN classifier, Neural Network (NN), Support Vector Machine (SVM). The mathematic equations of this new classifier will be shown in the next sections.

3.4.1 Cluster-K-Near Neighbor Classifier (C-K-NN)

Firstly, we will cluster each class C_i to number of sub-classes $C_{i,j}$, with means $\mu_{i,j}$, with $1 \le j \le m_i$ where m_i is number of sub-classes using K-means modified algorithm. This procedure will minimize the variance within each cluster and maximize the variance between the clusters. As traditional K-means algorithm suffering from determining the number of sub-classes and the initial K-vector, in this classifier, two algorithms were developed to choose the optimum initial K-vector for the minimum variance, namely near to near and near to mean. The next Figure 3.32 will show that each class C_i will be divided into number of sub-classes $C_{i,j}$ represented by the mean $\mu_{i,j}$ of the data.

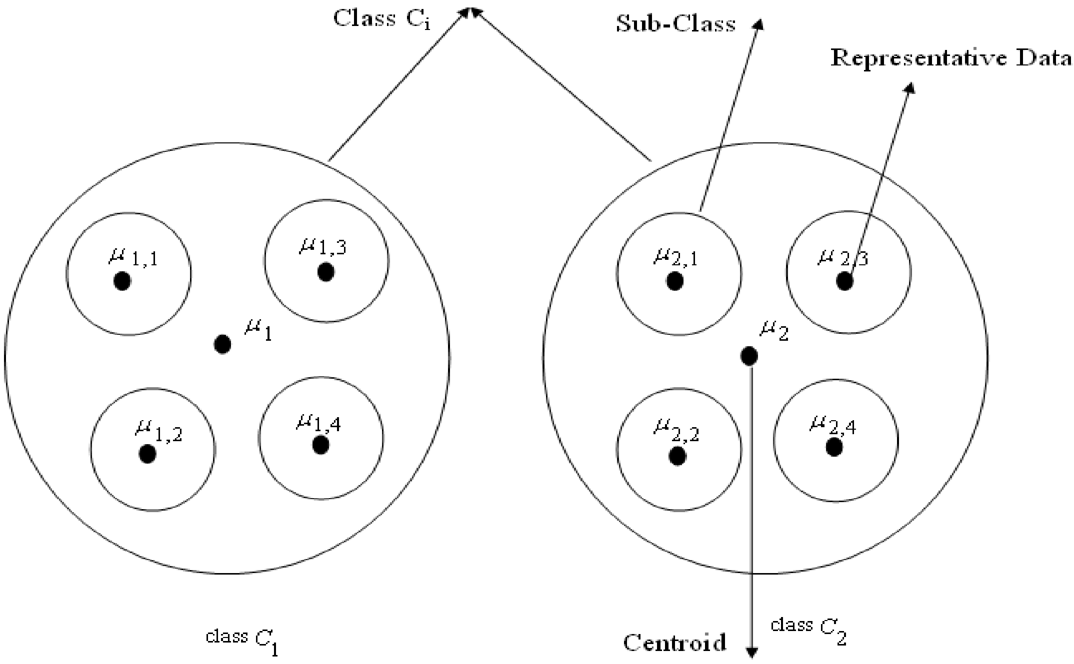


Figure 3-32: Classes, sub-class, representative data

1. Near to Near Algorithm

This algorithm calculates the distance $d(x_{i,n}, x_{i,m})$ for all $x_i \in C_i$, and then starts to cluster each class to $N_i - 1$ sub-classes, where $card(C_i) = N_i$. Then each two closet data will put in the same sub-class $C_{i,1} = \{x_{i,n_0}, x_{i,m_0}\}$, where: $\min_{n \neq m} d(x_{i,n}, x_{i,m}) = (x_{i,n_0}, x_{i,m_0})$

The other data will be gathered at separated sub-class

$C_{i,j} = \{x_{i,j}\}, \forall j \in \{1.......N_i\} - \{n_0, m_0\}$

Now the index n_1 and m_1 be

$$\min_{\substack{n \neq m \\ (n,m) \neq (n_0, m_0)}} d(x_{i,n}, x_{i,m}) = (x_{i,n_1}, x_{i,m_1}) \tag{3.56}$$

The sub-class $C_{i,r}$ will split into two other sub-classes if the x_{i,n_1} and x_{i,m_1} belong to it.

But if they belong to different sub-classes, they will be put in that classes based on the classes *card* . The iteration will stopped after K-subclasses were obtained and the initial K-vector will become the means of each sub-classes.

2. Near to Mean algorithm

It's similar to near to near algorithm, however it deals with the mean of sub-class, therefore, the class C_i will split into two sub-classes:

$$C_{i,1} = \{x_{i,n_0}, x_{i,m_0}\} \tag{3.57}$$

and

$$C_{i,1} = \{x_{i,j} \mid j \notin \{n_0, m_0\}\} \tag{3.58}$$

where, $d(x_{i,no}, x_{i,mo}) = \min\limits_{n \neq m} (x_{i,n}, x_{i,m})$.

Then the class will be updated C_i by replacing $x_{i,no}$ and $x_{i,mo}$ with their average,

$$C_i^1 = \{......x_{i,no-1}, s_0, x_{i,no+1}, ...x_{i,mo-1}, s_0, x_{i,mo+1}, ...\} \tag{3.59}$$

where $so = (x_{i,no} + x_{i,mo}) / 2$.

then $x_{i,n1}$ and $x_{i,m1}$ will be

$$d(x_{i,n1}, x_{i,m1}) = \min \left\{ d(x_{i,n}, x_{i,m}) \mid d(x_{i,n}, x_{i,m}) \neq 0 \right\} \tag{3.60}$$

Then C_i^1 will replace all the data in that are equal to $x_{i,n1}$ or $x_{i,m1}$ by s_1, which is the mean of the union of the two subclasses where $x_{i,n1}$ and $x_{i,m1}$ belong to:

$$s_1 = \frac{C_{n1}x_{i,n1} + C_{m1}x_{i,m1}}{C_{n1} + C_{m1}} \tag{3.61}$$

Where, C_{n1} is the number of iteration of $x_{i,n1}$ inside of C_i^1 and C_{m1} is the number of repetition of $x_{i,m1}$ inside of C_i^1. The algorithm will stop once the number of distinct vectors inside of C_i^r is equal to k. Our classification algorithm does not need to keep all the data, but only the average of each subclass. This is the outstanding feature of this new clustering. To classify a new data or vector x, k-NN algorithm will be used, i.e., we assign x to the class $C_{\hat{i}}$ for which

$$\hat{i} = \arg_i \min_{i,j} d(x, \mu_{i,j}),$$

where $\arg_i \; d(x, \mu_{io,jo}) = i_o$.

Further investigation about the k-NN algorithm is indeed needed to find the closest j- examples in the dataset and select the predominant class. The smallest closest examples could be found in the dataset and select the predominant class which have exactly k examples.

Now, the mathematical derivation of the new classifier is clear.

3.4.2 Quantification Metric

During the training stage of the classifier, the features position matrix J will be constructed beside the classifier library which contained the positions of the selected features by features selection method from part II. This matrix will be updated during the training time until we reach the optimum features positions from the training images. During the testing there is no need to repeat the stage of features selection because we will directly extract the desired features from the input face image based on the features positions in the matrix J. Then, the distance between the extracted features from the input image and classifier library classes will be calculated using Euclidian Distance method to sign the input image to one of these classes. Since, the main responsibility of authentication system is to investigate whether to prevent the users or to permit them into the system; it is necessary for the system to have high predictability trust to be sure that the users can be prevented if they are 100% unauthorized. In the case that the prevention response is not trusted, the correction must be triggered. Using this philosophy, the authentication system is constantly keeping track with the self-correction system. Therefore, the necessity to introduce metric to quantify our classifier response. Two new methods to calculate the metric for prevention predictability trust coefficient, depends on the nearest centroid in the subclasses of database are introduced in the classification algorithms of the proposed authentication system.

We introduce the first metric that the input data and the representatives of each class as exhibited in Figure 3.33. The value of the trust coefficient (TC) can be calculated from equation (3.62):

$$TC = 1 - \left(\frac{\dfrac{1}{\sum_{k=1}^{i} \dfrac{1}{d_i}}}{d_{i+1}} \right) \tag{3.62}$$

where $d_1 < d_2 < d_3 < ... < d_i < d_{i+1}$, and $d_1, d_2, d_3, ..., d_i$ are the smallest distances from the same class, d_{i+1} is the first smallest distance from the other class.

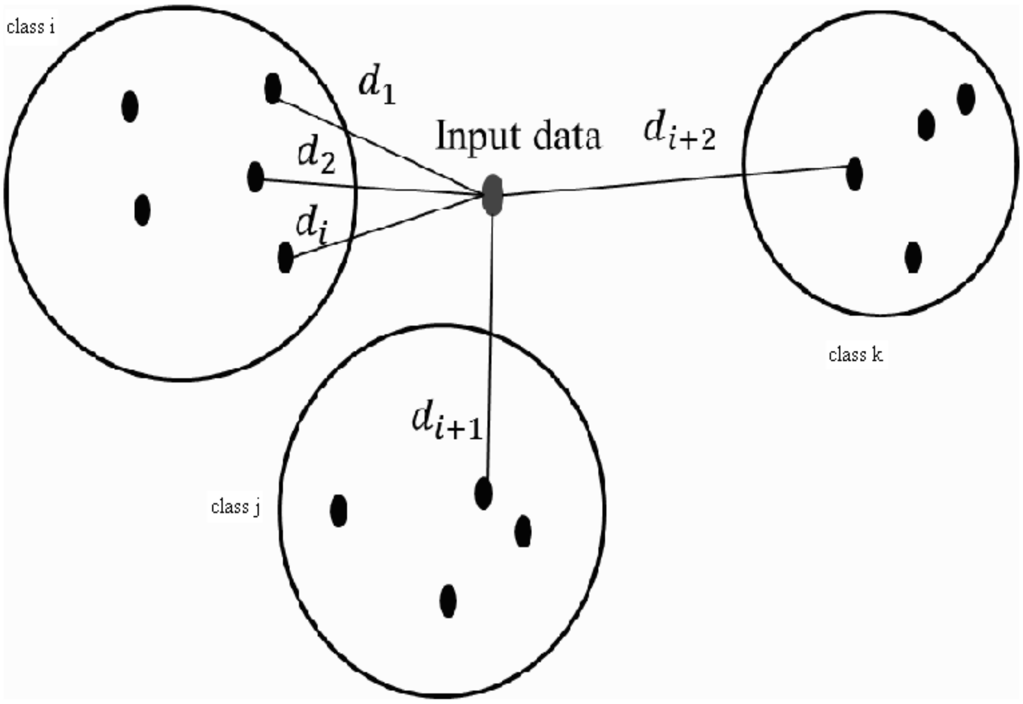


Figure 3-33: First method to calculate the trust coefficient (TC)

The second method depends only on the two nearest representative data to the input data as exhibited in Figure 3.34. The value of the trust coefficient (TC) can be calculated from equation (3.63):

$$TC = 1 - \left(\frac{d_1}{d_{i+1}} \right) \tag{3.63}$$

where $d_1 < d_{i+1}$, d_1 is the smallest distance and d_{i+1} is the first smallest distance from different class other than d_1.

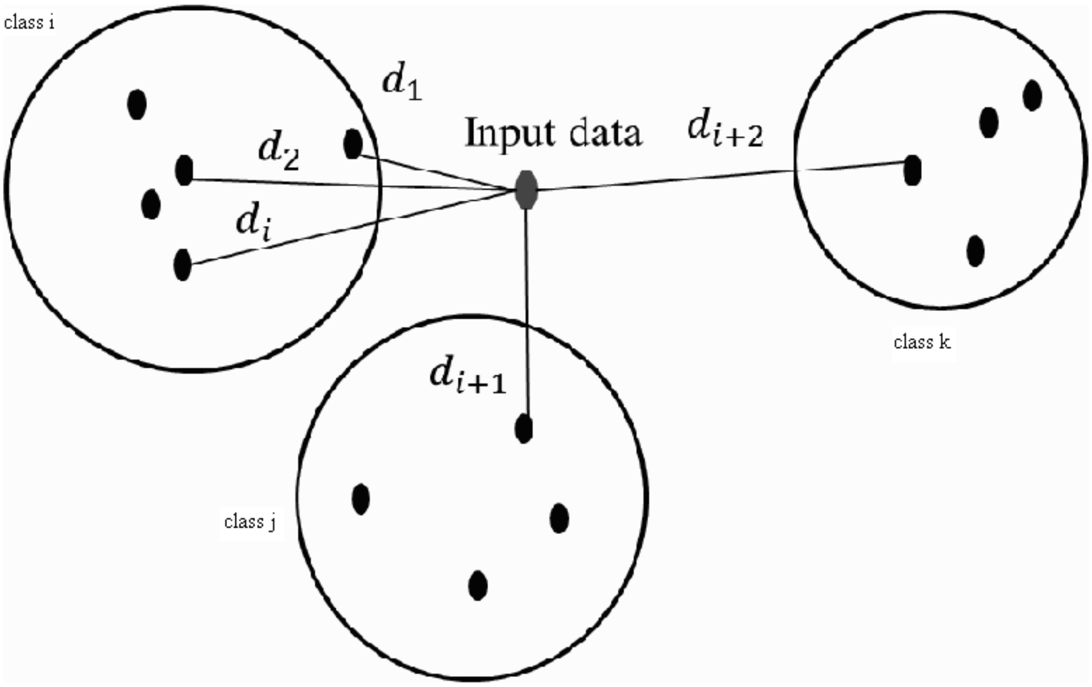
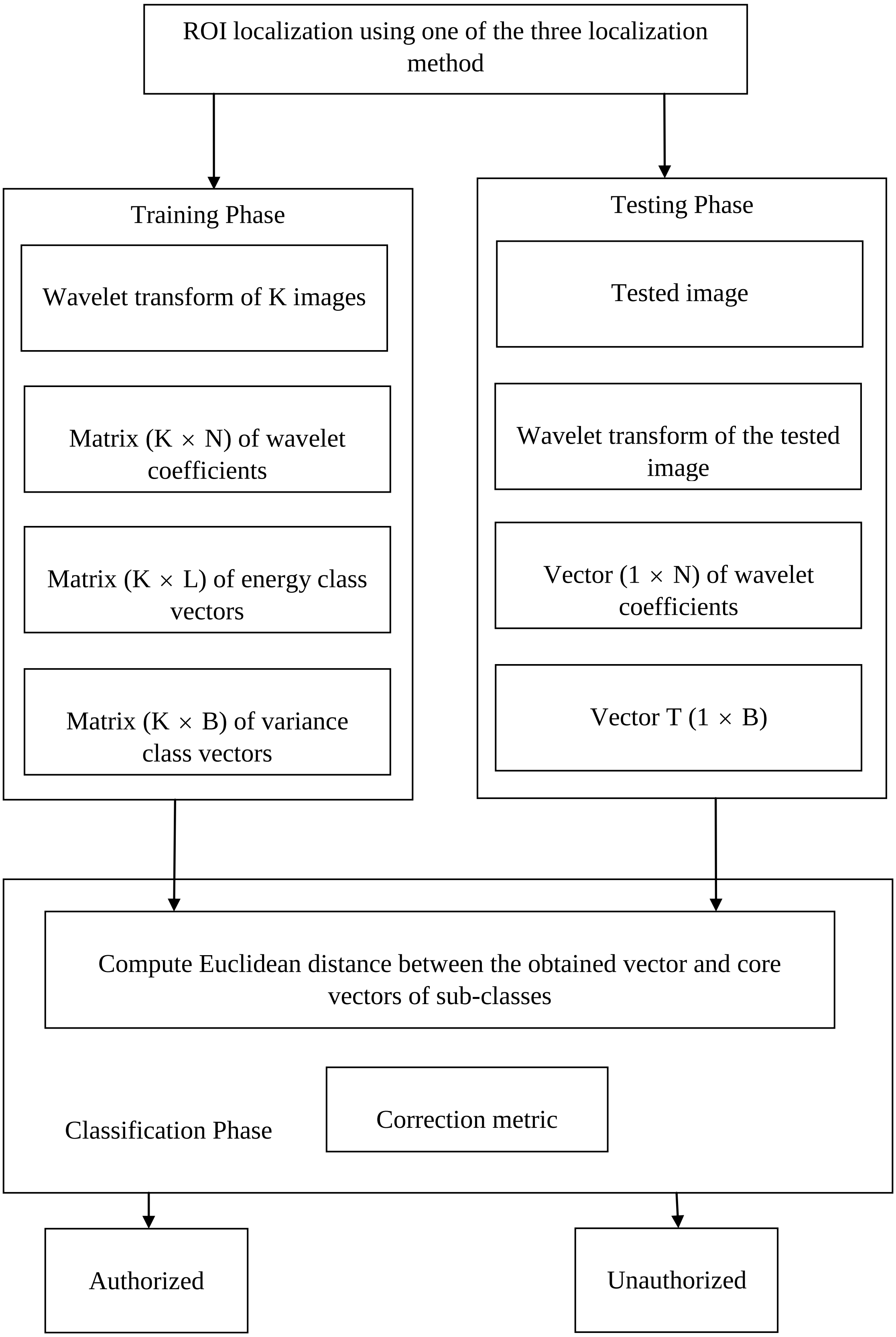


Figure 3-34: Second method to calculate the trust coefficient (TC)

Now the proposed face authentication system is illustrated in the following diagram:


ROI localization using one of the three localization method
Training Phase
Wavelet transform of K images
Matrix (K × N) of wavelet coefficients
Matrix (K × L) of energy class vectors
Matrix (K × B) of variance class vectors
Testing Phase
Tested image
Wavelet transform of the tested image
Vector (1 × N) of wavelet coefficients
Vector T (1 × B)
Compute Euclidean distance between the obtained vector and core vectors of sub-classes
Classification Phase
Correction metric
Authorized
Unauthorized

3.5 Summary

This chapter presents the proposed computer recognition system for human authentication (verification) in frontal image using an invariant biometrics features. The work is divided into three parts. In part I, three ROI methods were proposed to locate the human face in the input image. The first two methods are pattern matching methods and third is clustering method.

In part II, the proposed features extraction and selection method were presented. First multiresolution wavelet decomposition was used to extract invariant features against different variations. Then a statistical method proposed to select the best feature for classification.

In part III, the mathematical derivation of proposed classifier for classification process is introduced. In addition, two methods were proposed to calculate the quantification metric based on trusted features of classification.

In the next chapter, the results of the experiments will be analyzed based on the previous proposed method and algorithms and comparison between the results obtained will be accomplished.

CHAPTER 4

RESULTS AND DISCUSSIONS

4.1 Introduction

In this chapter, the experimental results of the proposed methods will be analyzed and discussed. First, it starts with presenting a description of the used datasets in the work, followed by the results of the first part of the work, which consists of face localization based on ROI using the three face localization methods. Moreover, a comparison between the proposed methods and previous methods will be included in this part. Next, the result of the second part of the work, which start by comparison between the wavelet families in order to select the best wavelet mother function for our system. also consists of threshold versus features reduction and classification accuracy rate, will be analyzed to select the optimum energy and variance thresholds. A comparison between the proposed features extraction and selection method and with the existing methods will be presented. Next, the result of the third part of the work will be discussed based on different parameters to show the superiority of the proposed classifier against number of common existing classifiers. In addition, the results of implementing the response quantification metric will be presented as well. Finally, the proposed system will be evaluated on the five localization database and the statistical features and selection method will be applied and C-K-NN classifier will be used and the result will be included in the end of this chapter.

4.2 Datasets

During the experiment, five of popular databases have been used to test the proposed face localization methods which were Yale database with image size 320×243, MIT-CBCL database with image 115×115, Indian databse with image size 640×480, BioID database with image size 384×286 and caltech database with image size 896×592 and the a number of images from each database were selected based on different conditions such as light variation, pose, background variations , indoor and outdoor. Another two popular Datasets have been used to test the proposed features extraction and selection and the proposed classifier which were ORL Dataset, and Face94 Dataset. The 40 individuals and 10 images for each were selected based on different conditions such as light variation, pose, expressions, glass, and background variations. In addition the faces from Face94 with colour background.

4.2.1 Yale University Dataset

The first database used is Yale database that was established by Yale University [91]. It contains 165 images which is 11 images for each of 15 persons with image size 320×243 pixels. The images are captured with different situations or configurations such as: center-light, with/without glass, happy, sad, left-light, with/without glass, normal, right-light, sad, sleepy, surprised, and wink. Few examples of these images are shown in Figure 4.1.

Figure 4-1: Samples from Yale database for localization

4.2.2 MIT-CBCL Dataset

The second database used to evaluate the proposed localization methods is MIT-CBCL database. It was established by Center for Biological and Computational Learning (CBCL) in Massachusetts Institute of Technology (MIT) [92]. It has 10 persons with 200 images per person and image size 115×115 pixels. The images of each person are with different light conditions, clutter background, scale and different poses. Few examples of these images are shown in Figure 4.2.

Figure 4-2: Samples from MIT-CBCL dataset for localization

4.2.3 Indian Faces Dataset

This database established in February 2002 in the campus of Indian Institute of Technology Kanpur [96]. This database contains 671 images of 40 different subjects with eleven different poses for each of them. All captured images have a bright background and the subjects are in an upright, frontal position. For each individual, the following poses for the face have been included: looking front, looking up, looking down, looking right, looking left, looking up towards right, looking up towards left. Furthermore, images with four emotions - neutral, laughter, sad/disgust, smile- are also included for every individual, (see Figur 4.3).

Figure 4-3: Sample of faces from Indian dataset for localization

4.2.4 BioID Dataset

This dataset [97] consists of 1521 gray level images with a resolution of 384x286 pixels. Each one shows the frontal view of a face of one out of 23 different test persons. The images were taken under different lighting conditions, in a complex background and contain tilted and rotated faces. This dataset is considerate as the most difficult one for eye detection, (see Figure 4.4).

Figure 4-4:Sample of faces from BioID dataset for localization

4.2.5 Caltech database

This database [98] has been collected by Markus Weber at California Institute of Technology. The database contains 450 face images of 27 subjects under different lighting conditions, different face expressions and complexes backgrounds (see Figure 4.5).

Figure 4-5:Sample of faces from Caltech dataset for localization

4.2.6 Olivetti Research Laboratory Face Database (ORL)

The presented statistical model is evaluated on the Olivetti Research Laboratory face database [93]. This database contains 10 different images for each of 40 people with image size 92×112. The images of the same person are taken at different times, under slightly varying lighting conditions and different facial expressions. Some images have been captured with and without glasses. In addition, the individual’s heads is slightly tilted or rotated. Few examples of these images are shown in Figure 4.6.

Figure 4-6: Samples from ORL dataset for classification

4.2.7 Face94 Dataset

This dataset is considered as one of the most popular dataset used for face recognition. It constructs for the purpose of training and testing matching algorithms. It's prepared for computer science research projects at the University of Essex [94]. It contains images of 153 individuals divided into 20 females, 113 males (young) and 20 males' staff (old) with image size 180×200. For the purpose of testing the new classifier in the next section, a total of 400 images were chosen out of the dataset that represent 25 males and 15 females with 10 images each. Figure 4.7 shows samples of Face94 dataset.

Figure 4-7: Samples from Face94 dataset for classification

4.3 Part I: Face localization Face Localization Based on ROI Methods

This part consists of the following methods' results. In the first method, the results of face localization using number of n-means kernels with NC to reduce the effects of the illumination problem on pattern (template) matching method is presented. Instead of n-means kernel, the results of three proposed similarity measurements metrics to calculate the correlation coefficients between the template image and the dynamic window in input image of the second method are presented. In the last method, the results of face localization using K-means modified algorithm is illustrated. The results of this part are presented as follow.

4.3.1 The Results of Face Localization using N-Mean Kernel and Normalized Correlation

N-mean Kernel is used to increase the brightness of the input image by decreasing the image noise. A dynamic window created after the pre-processing step by size similar to the template image and passed through the input image. For the matching process, NC was used to locate the maximum correlation between the template image and the dynamic window in the input image, and then a recorded matrix is constructed to record correlations values. According to the maximum correlation in the recorded matrix, the position of the dynamic window is cropped; Table 4.1 presents the localization accuracy using a number of kernels. From the Table we can observe the decreasing in the localization accuracy when the dimension of the mean kernel is increased and this is due to the increase in brightness of the input image. Therefore 2-mean kernel appears to be the optimum kernel comparing with other. Figure 4.8 shows some of correct and incorrect faces locations on Yale dataset using the proposed method.

Table 4-1**:** : Result of using different N-mean kernels with NC

Kernel	2-mean	3-mean	4-mean	5-mean	6-mean
Result	91%	86%	73%	68%	56%

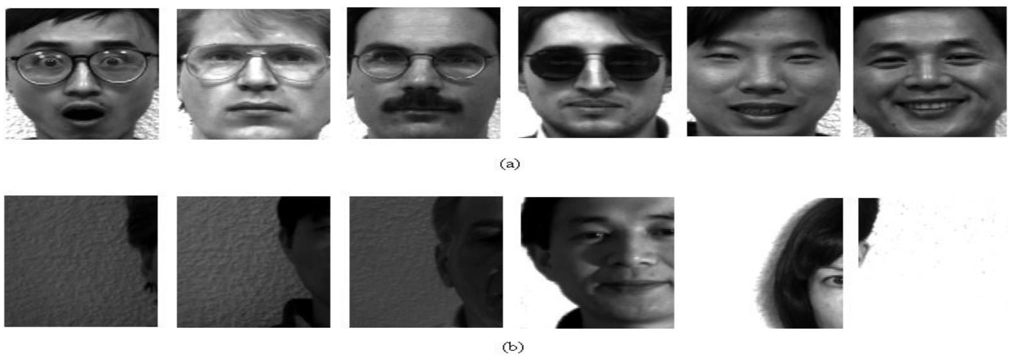


Figure 4-8: Samples of face localization: correct location in the top (a) and incorrect location in the bottom (b)

Table 4.2 shows the comparison on face localization in terms of result and time using the proposed method using NC alone and using LDA proposed by Meng et al. [41] on Yale dataset.

Table 4-2: Comparison between the proposed method and NC, LDA methods on Yale dataset

Method	Result (%)	Time (s)
LDA	92.7	15
NC	80	5
2-mean kernel + NC	91	8

From the table above there is a significant improvement by using N-mean Kernel compared with NC only but in the cost of the time. However, still this processing time is reasonable in comparison with LDA. In addition the result showed the less complexity of the proposed method compared with LDA.

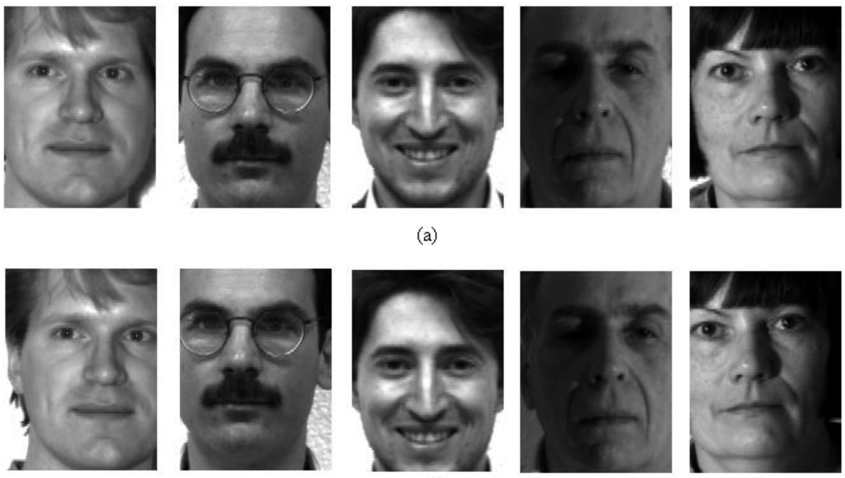
In conclusion, face localization using pre-process stage is preferred than using NCC only for illumination problem, and it gives close accuracy to using the LDA but with less complexity.

4.3.2 The Results of Face Localization Using Optimized Similarity Measurements Metrics

In template matching approach, the correlation between the reference image (template) and the target image (dynamic window) can be calculated by several similarity measurements. Table 4.3, shows the result of face localization established using ten different Statistical measurements. The results demonstrate clearly the increase in accuracy by using our statistical metrics, Chi2, SSTN, and the optimized metrics against the other measurements.

The results in Table 4.3 show clearly the efficiency of the face localization using our statistical metrics, which provides accuracy up to 100%. This result is more important knowing that our algorithm is able to determine precisely the face location whether there is variation in illumination, expressions and shade in the input image or not. In the table, the accuracy of the face localization by using Chi2, SSTN and OSAD (1, 2 and 3) are 100% on Yale dataset. Thus, OSAD overcame the drawback of the NC since there is no effect by the illumination variation. For the SAD, the accuracy is 98% which is acceptable in comparing with the other metrics. This localization error referring to the adjacent windows have almost the same SAD. However, these windows have the face but one of them is more correct than the other. This case will produce an error percentage for locating the face, but it is only a small localization error. Figure 4.9 shows examples of the face localization by using OSAD and SAD on Yale dataset. For the SSD, the accuracy is acceptable but its complexity is higher than OSAD and SAD, thus the error rate is maximized. In addition, if there are two windows with pixels values close to each other SSD is not useful to determine which one is similar to the template. Thus, OSSD (1 and 2) minimized this complexity and increase the accuracy of SSD up to 98%. In case of NCC, there is a significant increase in the error rate and that is referring to the illumination problem. Since, the illumination changed the pixels values of some image parts and this will cause maximum percentage of NCC in a wrong place. In contrast of the SSD error, the error in NCC gives a completely wrong location of the face. SHD gives a poor localization rate because it normally calculates the distance between to strings not between matrices. Therefore, SHD is not useful for face localization or detection but it can be used to calculate the difference between the signals.

The result in table 4.4 showed that the proposed metrics achieved high localization accuracy comparing with the other metrics on MIT-CBCL dataset. Unlike Yale dataset, the images used in this test with variations in the background and the face pose.



(a)

(b)

Figure 4-9: Samples of face localization: OSAD in the top and SAD in the bottom

Table 4-3: Comparison between Similarity Measures and the proposed metrics on YALE Dataset

Similarity Measure	Accuracy (%)
Sum Square T-distribution Normalized (SSTN)	100
Chi-Square distribution (Chi2)	100
Optimized Sum of Absolute Difference ($OSAD_1$ & $OSAD_2$ & $OSAD_3$)	100
Optimized Sum of Squared Differences ($OSSD_1$ & $OSSD_2$)	98
Sum of Absolute Differences (SAD)	98
Zero-mean Sum of Absolute Differences (ZSAD)	98
Locally scaled Sum of Absolute Differences (LSAD)	98
Sum of Squared Differences (SSD)	95
Zero-mean Sum of Squared Differences (ZSSD)	95
Locally scaled Sum of Squared Differences (LSSD)	95
Normalized Cross Correlation (NCC)	80
Zero-mean Normalized Cross Correlation (ZNCC)	80
Sum of Hamming Distances (SHD)	43

Table 4-4: Comparison between Similarity Measures and proposed metrics on MIT-CBCL Dataset

Similarity Measure	Accuracy for Pose (%)	Accuracy for Clutter Background (%)
SSTN	100	100
Chi2	100	100
OSAD1	100	96
OSAD2	100	96
OSAD3	100	96
OSSD1	98	92
OSSD2	98	92
SAD	98	94
ZSAD	98	94
LSAD	98	94
SSD	95	89
ZSSD	95	89
LSSD	95	89
NCC	80	73
ZNCC	80	73
SHD	43	40

Table 4.4 shows the comparison between the proposed metrics and other measurements on the MIT-CBCL Dataset. In this test, the concern is regarding high variation in poses from 30 degrees left to the 30 degrees right and clutter background as well. From Table 4.4 we can see that a template Matching using SSNT and Chi2 will not be affected by the variation in poses. However, OSAD (1,2 and 3) will not be affected by clutter background due to the existence of some objects in the

background. Consequently, some windows in input images will have exactly the same or very close classical metrics values as the template image, and this problem increases the error rate in general then affects the result. This result can be explained by the following. To calculate the error of two couples of values 1 and 11 then 100 and 110. The error is the same for both cases, which is 10 but the relative errors are different. Thus, we introduce the idea to divide the error by the average of the values.

In order to get a fair empirical evaluation of face localization, it is important to test our approaches by using the most comment datasets and compare our results with other techniques. Although several face localization methods have been developed over the past decade, only few of them have been tested on the same dataset. Table 6 shows the reported performance among several face localization methods on five standard datasets described here above

The Table 4.5 shows the performance of our approach Sum Square T-distribution Normalized (SSTN) over the five datasets and the results provide clear evidence that SSTN aproch gives high superriorty performance compare with other techniques.

Table 4-5: Compraison between SSTN and Chi2 with other techniques on different databases

Method	DataSet	Result	Our Result by SSTN	Our Result by Chi2
Orientation Template Matching [99]	BioID database	94.6%	98.7%	97.2%
Gradient Vector Flow [100]	Indian database	97.2%	100%	100%
Features Invariant Method [101]	Yale database	94.73%	100%	100%
Skin Color [102]	Caltech database	93.5%	99.5%	99%
AdaBoost and Artificial Neural Network [103]	MIT-CBCL database	85.34%	100%	100%

4.3.3 The Results of Face Localization Using K-Mean Modified Algorithm

In this method, the input image pixels were reshaped to one vector, and then K-Mean modified Algorithm is applied to cluster the vector pixels from 0 to 255 into two classes. The class one contains all pixels values less than certain threshold determined by K-means algorithm which represent the non-face parts and the second class contains all pixels values more than threshold which represent the face parts with some unwanted parts. Then, all pixel belong to the first class are signed to 0 values. To remove the unwanted parts from the second class, K-Mean Modified Algorithm is applied again to locate the ROI which contains only face part. Two classes are obtained using a new threshold determined by the algorithm as previous step. Then all pixels belong to the first class signed to 0 values as in the first step. Now, the second class contain only the pixels values above the threshold which represent the face part, but perhaps there are some errors in pixels clustering. Therefore, the image block which contains the face is cropped corresponding to the positions of the pixels values above the threshold. The proposed method is evaluated on two dataset namely: Yale dataset where the result was 100% with localization time 4s and the second dataset is MIT-CBCL. Table 4.6 shows the proposed method on the MIT-CBCL, Caltech and BioID databases. In this test, we have focused in the use of images of faces from different angles from 30 degree left the 30 degree right, the variation in the background as well as the cases of indoor and oudoor images. Two sets of images are selected from the MIT-CBCL dataset and two other sets from the other databases to evaluate the proposed method. The first set from MIT-CBCL contains the images with different poses and the second set is images with different backgrounds. Each set contains 150 images. While for the BioID and Caltech sets contains the images with indoor and outdoor status. The results showed the efficiency of the proposed K-mean modified to locating the faces from the image with high background and poses changes. In addition, another parameter was considered in this test which is the localization time. From the results, the proposed method can locate the face in a very small time because of it’s less in the complexity due to using of differential equations. Figure 4.10 shows some examples of face localization on MIT-CBCL, BioID and Caltech datasets.

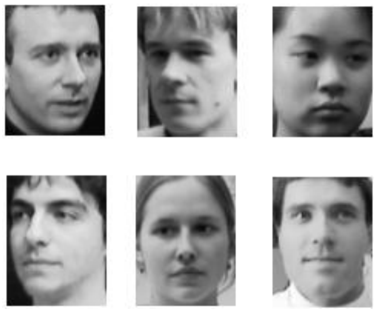


Figure 4-10: Samples of face localization using k-mean modified algorithm

Table 4-6**:** Face localization using K-mean modified algorithm on MIT-CBCL, BioID and Caltech, the sets contains faces with different poses and faces with clutter background as well as indoor/outdoor

Dataset	Image	Accuracy (%)	Time (s)
MIT-Set1	150	100	4
MIT-Set2	150	100	4
BioID	300	100	4
Caltech	300	100	4

4.4 Part II: The Results of Features Extraction and Selection Method

4.4.1 Optimum Wavelet Mother Fuction Selection

According to the main four properties of wavelets functions [104], the experiments results were analyzed based on these properties namely, orthogonality or biorthogonality, existence of associated filters, real or complex wavelets and compact or not compact support. In tables (4.7, 4.8, 4.9, 4.10, 4.11 and 4.12) the seven different wavelets function (Symelt, Daubechig, Coiflets, Mayer Discrete, Biorthogonal, Reverse Biorthogonal, Haar) were examined and the K-NN classifier was used. All coefficients from each level of decomposition and also from all 4 levels were extracted as features vectors. Furthermore, the comparison between the types of families have been done and result shows that for ORL data base the Biothogonal wavelets bior 2.2 gave better accuracy and this is associated to properties like compact support, symmetry, arbitrary number of zero moments, capability to continuous and discrete transformation among which the most important property is the biorthogonal analysis.

Table 4-7: Comparison between number of mother functions from symelt family

Wavelet Mother	Original Features	Classification Accuracy (%)
Sym1	13719	90.30
Sym2	14900	92.12
Sym3	15880	92.12
Sym4	17244	90.91
Sym5	18380	89.70
Sym6	19928	89.09
Sym7	21124	90.91
Sym8	22848	88.48
Sym9	24184	87.88
Sym10	26084	87.27

Table 4-8: Comparison between number of mother functions from Daubechies family

Wavelet Mother	Original Features	Classification Accuracy (%)
Haar	13720	89.09
db2	14900	91.52
db3	15880	91.52
db4	17244	90.91
db5	18380	90.91
db6	19928	90.91
db7	21124	89.09
db8	22848	89.09
db9	24184	89.70
db10	26084	87.88

Table 4-9: Comparison between number of mother functions from Coiflets family

Wavelet Mother	Original Features	Classification Accuracy (%)
Coif1	15880	90.91
Coif1	19928	90.30
Coif1	24184	87.88
Coif1	29716	87.88
Coif1	35391	89.09

Table 4-10: Performance of mother functions from Discrete Meyer family

Wavelet Mother	Original Features	Classification Accuracy (%)
Dmey	161872	87.88

Table 4-11: Comparison between number of mother functions from Biorthogonal family

Wavelet Mother	Original Features	Classification Accuracy (%)
bior1.1	13720	89.09
bior1.3	15880	91.52
bior1.5	18380	90.91
bior2.2	15880	92.73
bior2.4	18380	91.52
bior2.6	21124	90.30
bior2.8	24184	89.70
bior3.1	14900	83.64
bior3.3	17244	92.12
bior3.5	19928	91.52
bior3.7	22848	90.30
bior3.9	26084	89.70

Table 4-12: Comparison between number of mother functions from Reverse Biorthogonal family

Wavelet Mother	Original Features	Classification Accuracy (%)
rbior.1	13720	89.09
rbio1.3	15880	90.30
rbio1.5	18380	88.48
rbio2.2	15880	89.09
rbio2.4	18380	89.09
rbio2.6	21124	87.27
rbior2.8	24184	86.67
rbior3.1	14900	91.52
rbio3.3	17244	89.09
rbio3.5	19928	86.06
rbio3.7	22848	86.67
rio3.9	26084	86.06

To determine the optimum decomposition level, six wavelet functions were selected based on higher classification accuracy which are db2, db3, sym2, sym3, bior2.2 and rbio3.1. Table 4.13 showed that bior2.2 achieved higher classification accuracy compared with other wavelets with 4 levels of decomposition.

Table 4-13: Selection of optimum decomposition level

Decomposition Level	Wavelet Mother					
	db2	db3	sym2	sym3	bior2.2	rbio3.1
1	90.91%	90.91%	90.91%	90.91%	89.70%	90.91%
2	90.91%	90.91%	90.91%	90.91%	90.30%	90.30%
3	90.91%	90.91%	90.91%	90.91%	91.52%	90.30%
4	91.52%	91.52%	92.12%	92.12%	92.73%	91.52%
5	89.70%	89.70%	89.70%	89.70%	90.30%	90.91%
6	89.70%	90.91%	89.70%	90.91%	86.67%	90.91%

4.4.2 Features Extraction Method

In this part, features are extracted as vectors of coefficients from the faces images using wavelet decomposition to construct matrix of coefficients. In the next step, as the coefficients number is very big, the significant coefficients are selected in terms of energy and classification accuracy by applying dynamic energy threshold for each column in constructed matrix. Then all columns have energy values less than certain threshold are removed. To find the optimum threshold, ORL dataset is divided into training and testing data. The training data are used to build the K-NN classifier, while the testing data are used to calculate the classification accuracy rate. The applying of dynamic threshold is repeated until reaching best classification rate with the minimum number of coefficients. Once the optimum energy threshold is found, the proposed variance modified metric is calculated for the remaining columns. Then, another dynamic threshold is applied to select the most significant features by removing all columns less than the threshold in order to increase the classification accuracy.

The selection of the invariant and important features can be achieved by using the previous statistical method. Firstly, we apply the wavelet decomposition to each image of dataset, followed by the proposed statistical energy reduction method. Then, K-NN classifier is constructed using the training data and the classification accuracy rate is calculated using the testing data. Figure 4.11 shows the performance of the K-NN classifier after the dynamic removing of the columns with minimum energy values corresponding to the number of extracted feature with different threshold values. It shows that the maximum accuracy obtained is 92% with 437 features. It can be noted that there are several points which reached the highest performance but we choose the point where the number of the features is the minimum and the classification accuracy rate is the maximum.

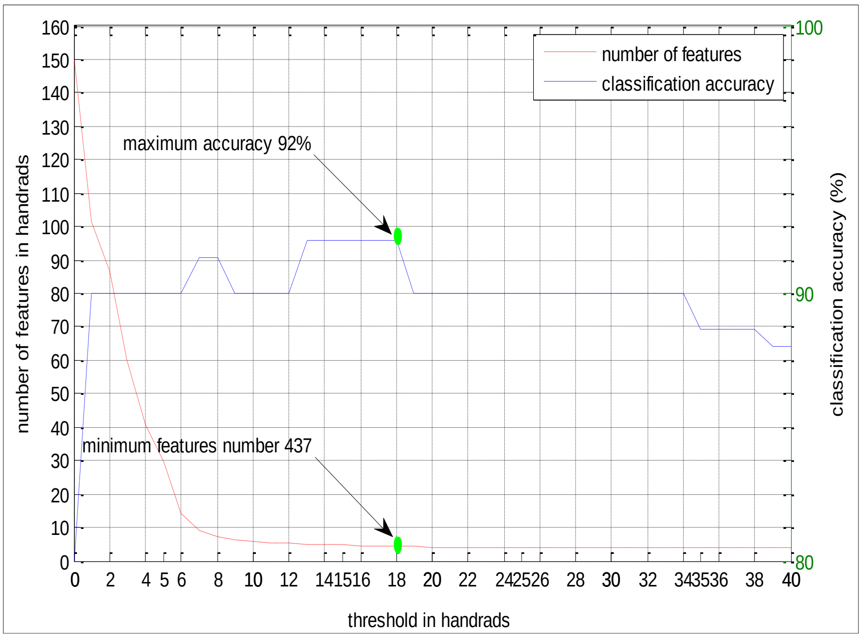


Figure 4-11: The performance of the K-NN classifier after applying the dynamic threshold to remove all columns with minimum energy values.

Subsequently, after the columns with minimum energy are removed, the proposed variance modified metric is applied on the remaining columns to select the most important coefficients for classification. Then, K-NN classifier is constructed again based on the new coefficients and the classification accuracy rate is calculated. Figure 4.12 shows the performance of the K-NN classifier after the dynamic removing of the columns with minimum variance values corresponding to the number of extracted feature with different threshold values. It shows that the maximum accuracy obtained is 95.35% with 171 features. It can be noted that there are several points which reached the highest performance but we choose the point where the number of features is the minimum and the classification accuracy rate is the maximum. Moreover, a significant increase in the classification accuracy rate 3.35% is obtained by using proposed variance modified metric which can distinguish between the dataset classes.

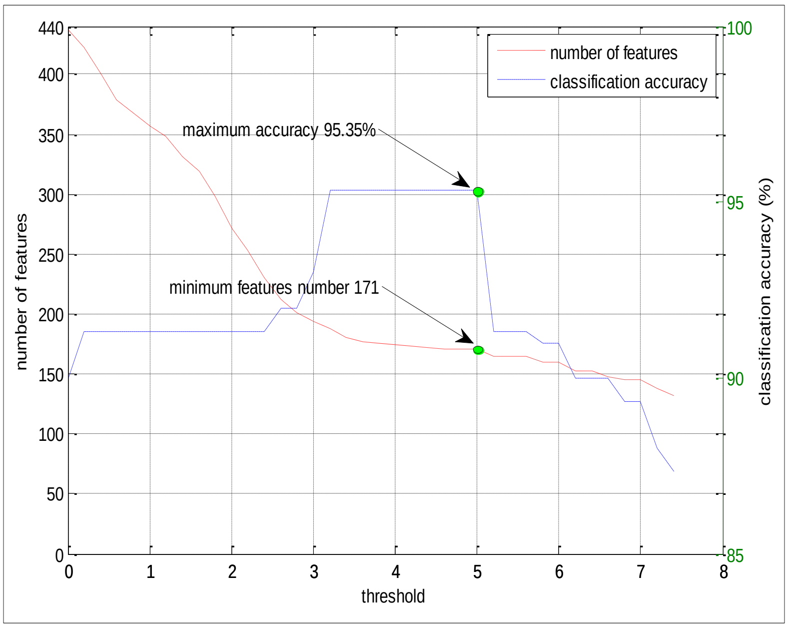


Figure 4-12: The performance of the K-NN classifier after applying the dynamic threshold to remove all columns with minimum variance values.

The result from Table 4.14 shows the comparison between the proposed features extraction and selection method with the number of existed methods namely ICA, LDA, and PCA. In addition, the method proposed by Cunjian and Jiashu [78] to extract only the features with high information is included in this comparison. In this test, ORL dataset with different pose, face expressions and illumination changes is divided into training set and test data. The training set is used to construct K-NN classifier, while the testing data is used to classify the features. During the test, not all ICs are extracted where only 40 ICs of rank 3 are extracted from the low frequency coefficients of the wavelet transform referring to the method in [20]. For the PCA, the total population scatter matrix of the wavelet transform low frequency coefficients in level two is calculated, then the eiganvalues and the corresponding eigenvectors of the matrix are calculated as in [73]. For the LDA, the method proposed by [95] is used to extract 2DLDA features from wavelet transform low frequency coefficients in level one. As for the method proposed by Cunjian and Jiashu [78], the energy-entropy features are extracted from the wavelet transform low frequency coefficients in level two. In spite of all previous methods extraction the features from the low frequency coefficients referring to it is less sensitivity to the facial expressions changes, the proposed features selection method considered high frequency coefficients where it also contain facial information. Latter, K-NN classifier is used to evaluate the performance of the proposed methods against these existed methods. Table 4.14 shows the increase of recognition performance by using the proposed features extraction selection method with classification accuracy rate 95.35% and minimum consuming time 7s.

Table 4-14**:** Proposed statistical method against other methods

Method	Accuracy (%)	Time (s)
Principle Component Analysis (PCA)	87.2	29
Linear Discriminant Analysis (LDA	89.67	29
Wavelet Energy Entropy	90.5	19
Independent Component Analysis (ICA)	92.5	32
Wavelet Energy & Variance Modify	95.35	7

4.5 Part III: The Results of Classification Process

4.5.1 C-K-NN Classifier Results

In This part, first ORL and Face94 datasets are divided into training and testing data, followed by selection of the features vectors from the training data using the proposed method to construct C-K-NN classifier library. Then, K-means modified algorithm is applied to cluster the training classes into set of sub-classes with representative data μ for each sub-class. This cluster process will increase the performance of the classification by using its representative data instead of the centroid of each class as in the traditional K-NN classifier. During the training, there is another vector created beside the library (storing as data classes) called position vector which contains the positions of the selected coefficients. The benefit behind this vector is to reduce the testing time, where there is no need to apply the features extraction and selection method again. In the testing phase, the coefficients are extracted from the image using wavelet decomposition, followed by coefficients selection based on the position vector, Then distances between the input feature coefficient and the mean μ of the all sub-classes are calculated using Euclidian Distance metric (ED) to classify the features to the sub-class with the minimum distance value. A comparison between the proposed classifier and a number of existed classifiers [70, 82]showed in table 4.15 and table 4.16 on ORL and Face94 datasets respectively. During the comparison, three parameters are considered to evaluate the performance of the proposed classifier which are training images per-person, classification time and classification accuracy.

Table 4-15**:** Comparison between C-K-NN Classifier and the other Classifies on ORL dataset

Parameter	Neural Networks (NN)	K-Nearest Neighbor (K-NN)	Support Vector Machine (SVM)	Cluster-K-Nearest Neighbor (C-K-NN)
Training images	9/person	6/person	8/person	6/person
Classification time	10s	7s	17s	4s
Classification accuracy	96.7%	95.35%	97.93%	99.39%

Table 4-16: Comparison between C-K-NN Classifier and the other Classifies on Face94 dataset

Parameter	Neural Networks (NN)	K-Nearest Neighbor (K-NN)	Support Vector Machine (SVM)	Cluster-K-Nearest Neighbor (C-K-NN)
Training images	9/person	6/person	8/person	6/person
Classification time	10s	7s	17s	4s
Classification accuracy	98.35%	97%	99%	100%

The results of above tables show that neural networks classifier need a big number of training images to achieve high classification accuracy while K-NN and C-K-NN classifiers used a few number of training images. In terms of classification time, the uses of position vector in C-K-NN classifier to extract the features from the test images decreased the testing time while the other classifiers apply the features extraction and selection method again on the tested images. Moreover, SVM classifier needs more time as compared with other classifier, where it is basically designed to classify between two classes. In addition, C-K-NN classifier achieved the highest classification accuracy by comparing the other classifiers where are 99.35% and 100% for ORL and Face94 respectively. The reason behind that, the clustering process minimized the probability of the misclassification.

4.5.2 Quantification Metric Results

In this part, the trusted coefficients of the ORL dataset are calculated to quantify the classification accuracy by using the correction metric as shown in Figure 4.13. From the figure, the trust coefficients are arranged from 0 up 1 where the 0 represent 0% and 1 represent 100%. Suppose 0.2 is selected as threshold to separate between the confident and unconfident images classifications. Then, all classification results with trust coefficient above the threshold are confirmed as confident results. While all classification results below threshold are considered as unconfident results. However, there are incorrect classifications with trust coefficients above the threshold as in the image 159 those need to reclassify again for correction. On the other hand, there are incorrect classifications below the threshold are considered as confident classification errors as in the image 75. Thus, there are 151 images with trust coefficients above the threshold which consist of 150 confident classifications and one incorrect classification. While, there are 9 images with trust coefficients below the threshold where is consists of 8 unconfident classifications and one incorrect classification.

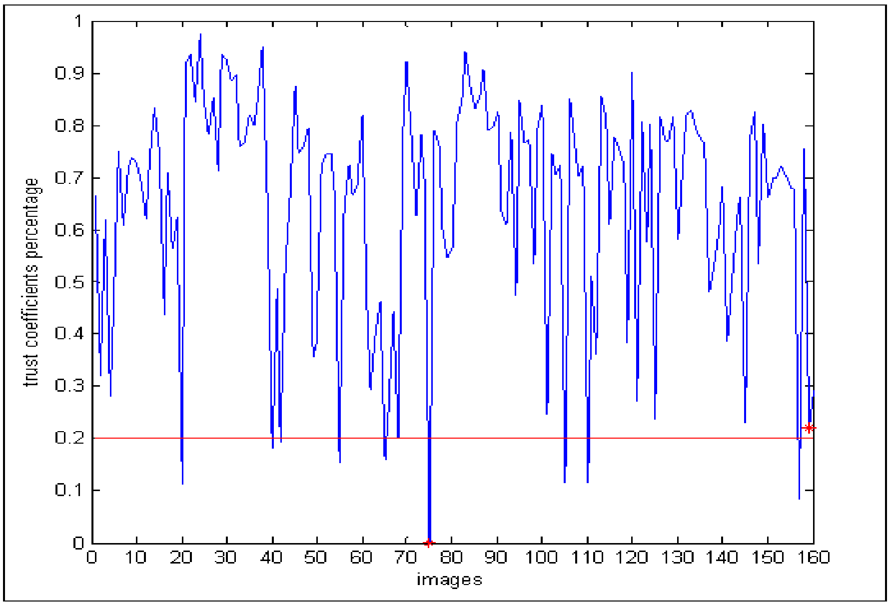


Figure 4-13: Trust coefficients response of our classifier, * is classification error

Our metric is relevant when the accuracy is weak, Yale dataset [91] was used without the localization stage to show the importance of the metric. The relation between the trust coefficients and the image classification can be shown in the Figure 4.14, where the red points represent the error positions of the classification.

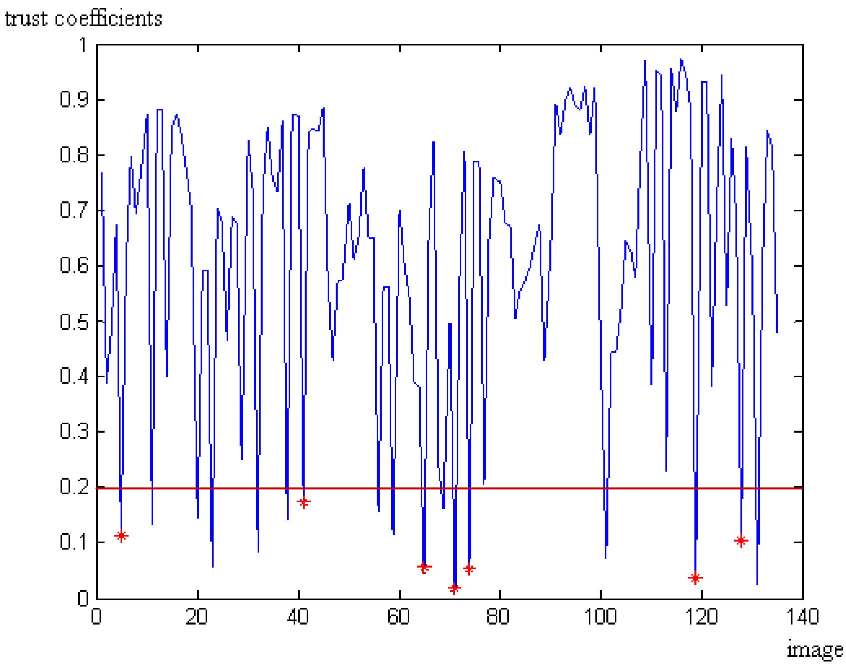


Figure 4-14: The relation between the trust coefficients and the image classification, * is classification error

The relation between the trust coefficients and the probability of error is shown by the follows graph and table (see Figure 4.15 and Table 4.17). In the table, T.C is trust coefficients and P. error is the propbability of error.

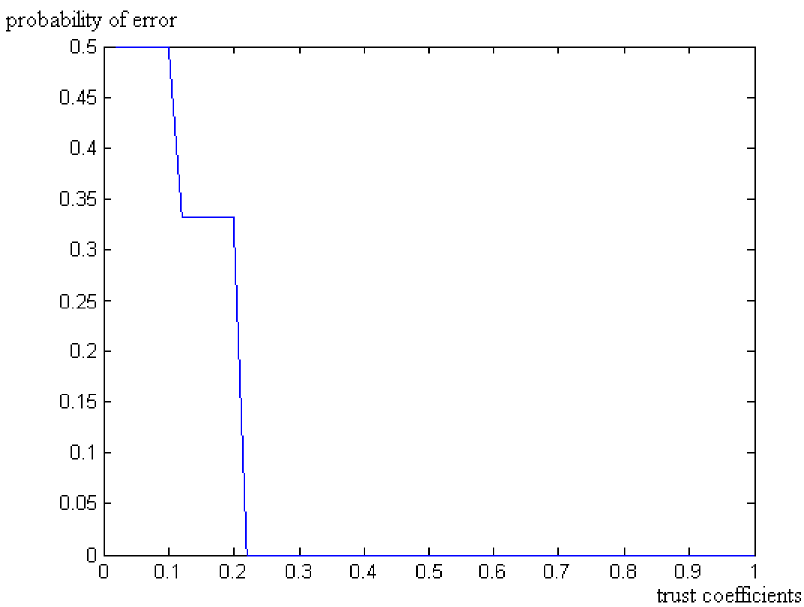


Figure 4-15:The relation between the trust coefficients and the probability of error

Table 4-17:The relation between the trust coefficients and the probability of error

Range	ORL dataset		Yale dataset		Total-T.C	Total-P. Error
	T.C	P. Error	T.C	P. Error		
0-0.1	2	1	8	4	10	5
0.1-0.2	3	1	9	3	12	4
0.2-0.3	12	0	4	0	16	0
0.3-0.4	6	0	6	0	12	0
0.4-0.5	9	0	9	0	18	0
0.5-0.6	12	0	18	0	30	0
0.6-0.7	21	0	21	0	42	0
0.7-0.8	42	0	19	0	61	0
0.8-0.9	40	0	29	0	69	0
0.9-1	13	0	12	0	25	0
Total	160	2	135	7	295	9

4.6 The Proposed System Results

In This part, first previous five face localization datasets are divided into training and testing data. Then, the proposed face localization methods were applied to locate the face in the input image. Followed by extraction and selection of the features vectors from the training data using the proposed method to construct C-K-NN classifier library. The performance of the proposed system showed in table 4.11 on localization datasets.

Table 4-18: The performance of the proposed system

Dataset	Localization Method		
	n-Mean Kernel and NC	Optimized Metrics	K-Means Algorithm
Yale Dataset	91%	100%	100%
MIT-CBCL Dataset	85.34%	100%	100%
Indian Dataset	97.2%	100%	100%
BioID Dataset	83%	95%	94.7%
Caltech Dataset	83%	96%	95%

4.7 Summary

This chapter presents the results obtained by applying the proposed methods in the three parts of study. In part I, the ROI which contain the face is located using two patter matching methods and clustering method. In the first pattern matching method, numbers of ROIs are cropped manually, and then the average face (template) is created from those ROIs. To locate the face in the input image, N-mean kernel is applied to increase the brightness of the image. Then, dynamic window with same size of template is created and passed through the input image. NC is used to calculate the correlation coefficients between the template and the dynamic window and the correlation values saved in the record matrix. Then, the window corresponding to the maximum correlation in the record matrix is extracted to be the face location. A number of N-mean kernels are applied on the Yale dataset. The results showed that 2-mean kernel achieved highest accuracy as compared with other kernels. In addition, a comparison between the proposed method and to other methods which are NC and LDA is presented. The results showed that the proposed method achieved highest localization accuracy with reasonable time. In the second pattern matching method, instead of using N-mean kernel, numbers of optimized metrics are used to calculate the correlation coefficients between the template image and the dynamic window. Then, the window corresponding to the minimum correlation in the record matrix is extracted to be the face location. A comparison study between several similarity measurements metrics is presented. The results showed that the proposed metrics increased the performance of the SAD and SSD algorithms. Also SSTN and Chi2 achieved high localization accuracy for indoor and outdoor images. In the last localization method, K-Means modified algorithm applied twice after reshape the input image in vector to cluster the input image into two classes. Then, the class contained the background and unwanted parts removed. Then, the first pixel from the top, bottom, left and right in the clustered image is determined and the corresponding window is extracted from the input image. Yale and Face94 dataset are used to evaluate the proposed method. The results showed that the proposed method achieved localization accuracy of 100% in the image with various poses, clutter background and illumination changes.

In part II, the invariant features for classification are selected using statistical method. First, wavelet transform is applied to decompose the input images then a matrix of wavelet coefficients is constructed. A dynamic threshold is applied on the coefficients matrix to reduce the coefficients dimensions by removing all columns with energy value less than the threshold. Finally, another threshold applied on the remaining columns to select the most features with high variance from the total mean of the classes. A comparison study between the proposed method and number of the exited feature selection methods is presented using ORL dataset and NN classifier. The results showed that the proposed method differentiated better than the other method and has minimum selection time as well.

In part III, First, ORL dataset is divided into training and testing data. Then followed by features selection method to construct the library of C-K-NN and position vector. Then, K-means modified algorithm is applied by cluster training data into set of sub-classes and determined the representative data of each sub-class. In the testing phase, the features are extracted based on the position vector and then the distance between the input feature and the representative data is calculated using ED. A comparison between the proposed classifier and number of exited classifier is presented using ORL and Face94 datasets. The results showed that the C-K-NN classifier overcame the drawback of the K-NN classifier and achieve highest classification accuracy 99.35% and 100% for ORL and Face94 respectively. In addition, C-K-NN need small amount of Training data with minimum of classification time as compared with other classifiers. A new metric is used to quantify the classification accuracy. This metric calculated the trust coefficients of image classification and determined whether the classification is confident or unconfident based on particular threshold. Moreover, the metric introduced some error correction which increased the classification accuracy rate.

Finally, the proposed biometrics system has been tested on the five localization databases and the result showed that the higher classification accuracy in term of illumination and variation in poses and clutter background (indoor-outdoor).

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Introduction

The work in this thesis concerns with the development of a face classification for face biometrics authentication trend. We have studied several parts of face classification in image with various face situations and background changes. In particular, we proposed three ROI detection methods to locate the faces even under high illumination and background variety. To improve the classification we proposed a new feature extraction and selection method from the multiresolution representation of wavelet decomposition. We applied a new classifier; it is a combination of K-mean modified algorithm and K-NN classifier. In this chapter all research works presented in this thesis will be summarized in section 5.2. The contributions of this work are presented in section 5.3. Some suggestions for future works are highlighted in section 5.4.

5.2 Conclusions of the Work

One of the main points in this thesis is proposing an invariant features to distinguish between the classes for face biometrics authentication systems based on wavelet transform. This thesis consists of three parts. In part I, we focus our attention on developing the RIO detection by suggesting some methods for face localization. Two pattern matching methods and one clustering method were proposed to locate the faces. In the first two methods, the average face (template) has been constructed from a number of ROIs extracted manually from a training set of images. In the first pattern method, n-mean kernel was used to increase the brightness of the input image and decrease the noise. Later, for locating the face in the input image, a dynamic window has the same size of the template image was created and passed through the input image. The NC was used to calculate the correlation coefficients between the template image and the dynamic window. The window position corresponding to the maximum value was extracted to be the face location. A comparison between a different n-mean kernels shows that 2-mean kernel is the optimum kernel with localization accuracy 91% on Yale dataset with different images situations which provide reasonable increasing in the image brightness. Also another comparison between 2-mean kernel with NC and LDA and NC alone shows that n-mean-kernel-NC increase the accuracy around 11% comparing with the NC alone and less localization time 8s than LDA. In the second pattern matching method, instead of using n-mean kernel three optimized metrics were proposed to calculate the difference between the template image and the dynamic window. Then the corresponding window to the minimum value in the record matrix was extracted to be the face location. A comparison between the optimized metrics and other similarity measurement metrics shows that optimized metrics increase the accuracy of SAD and SSD from 98% and 95% respectively up to 100% on Yale dataset and from 90% and 87% up to 100% on MIT-CBCL dataset. While the other metrics achieved high localization results only against expressions and poses and poor results against illumination and small background changes. In addition, the proposed error measurement metrics achieved high localization accuracy up to 95% and 96% on BioID and Caltech databases respectively as indoor and outdoor localization databases. In the last method of localization, the input image was reshaped to the vector then K-means modified algorithm was applied twice to cluster the image pixels into classes. Then block window corresponding to the class of pixels values above the threshold which contain the face only is extracted from the input image. Two set of images with high illumination and background changes were established, the result shows that the proposed method achieved 100% localization accuracy with only 4s.

In part II, ORL dataset was divided into training and testing data. Firstly, a number of wavelet mother function were tested in order to select the optimum mother wavelet and the result showed that Biorthogonal wavelet appeared to be best choise in between the other wavelet. After that for the selection method, wavelet transform was used to decompose the training set into set of coefficients. The motivation behind using wavelet decomposition is to extract the invariant features against illumination and expressions. Coefficients were ranked into matrix as rows of vector for each image. The statistical energy of each column is calculated, a dynamic threshold is applied over the score of the ranked features. The columns with high statistical energy are kept according to the applied threshold value. The obtained features are then presented to K-NN classifier in order to validate our reduction. To optimize the number of features with the maximum classification accuracy rate, the threshold value is changed and the classification is performed by using the remaining features. This process is repeated until reaching the maximum accuracy with the minimum number of coefficients. The proposed method makes the K-NN classifier achieved 92% average classification accuracy rate with only 437 coefficients out of 15020 coefficients. To keep coefficients with high energy is not sufficient for classification; therefore we need to measure the contribution of each coefficient to classification. Once the minimum number of coefficients is obtained, the mean of each column is calculated followed by the mean total of all columns to calculate the variance of each remaining column. Then, the variance modified for each column is calculated. A dynamic threshold is applied over the score of the ranked features; the most significant features are kept according to the applied threshold value. The obtained features are then presented to K-NN classifier to validate our classification. To optimize the number of features with the maximum classification accuracy rate, the threshold value is changed and the classification is performed by using the remaining features. This process is repeated until reaching the maximum performance with the minimum number of coefficients. The proposed method makes the K-NN classifier achieved 95.35% average classification accuracy rate with only 171 coefficients out of 437 coefficients. The reason behind using modify variance is to select the features that separate between the classes and decrease the overlapping. A comparison study between the proposed methods of features selection and some of existed selection methods is accomplished. The statistical analysis shows that the proposed method achieved highest classification accuracy with fewer amounts of coefficients. Moreover, the proposed method performs the significant decreasing in classification time compared with the PCA.

In part III, ORL and Face94 datasets are used to evaluate the performance of proposed classifier C-K-NN. Once the features are selected from training data, the classifier library and position vector of selected features are created. In the testing, based on the values in position vector, the features is selected from the input image, then the distance between the input features and the representative data of each sub-class is calculated using ED to sign the features to the sub-class with the minimum distance. A comparison between C-K-NN classifier and a number of existed classifier shows that achieved highest accuracy 99.39% and 100% for ORL and Face94 respectively. The uses of position vector perform minimum classification time 4s as compared with other classifier. A new metrics are proposed to quantify the classification response. Moreover, some classification errors are detected and corrected by using these metrics.

Finally, to evaluate the proposed system, the five localization databases were used in order to test the system from locating the faces passing through extracting and selecting the invariant features then classifying these based on the classifier library. The results showed that our proposed system achieved up to 100% classification accuracy.

5.3 Contributions of the work

In this work we have verified for a number of methods that are necessary for any recognition system. We have made improvements and enhancements to existing methods in some cases and new methods are proposed. The specific contributions of this work are summarized as follows:

1. Three ROI methods for face localization from facial images with various pose, illumination, expressions and clutter background are presented. Two of these methods using template matching approach based on similarity measurements metrics which are as follows:

    - Template matching based preprocessing and normalized correlation to measure the similarity.

    - Template matching based optimized metrics to measure the difference.

In the third method, a new face localization method based on clustering concept is proposed and by using K-Means modified algorithm to distinguish between face and non-face regions.

2. New method for features selection from multi-resolution representation of discrete wavelet transform using the integration between statistical energy and variance modify metric of the wavelet coefficients is proposed to reduce the coefficient redundancy and to select only the significant features.

3. A new robust and efficient classifier for facial features is investigated.

4. A new metrics to quantify the exactness of classification response is introduced with eventual correction of some classification errors.

5.4 Future work

The following approaches are recommended for the future work:

1) Allocating a separate template for each class might increase the effectiveness of the optimized metrics in order to locate the faces in images with very high background changes.

2) Applying K-means modified algorithm to separate between more than two classes might likely increase the accuracy of the unwanted parts removing.

3) For faster and accurate classification, selection of the features extraction i.e. reducing number of features can be implemented by using the following metric:

$$Var_{total} = \min_{i,j}\left\{\sum_{i \neq j} \frac{(m_i - m_j)^2}{\sqrt{\mathrm{var}_i + \mathrm{var}_j}}\right\} \tag{8.1}$$

where the m_i is the mean of class i, m_j is the mean of all classes and var_i is variance of class i

4) To make the probability density function of classification error more highest, we need to investigate other metric as follow:

$$PPC = 1 - \left(\frac{\dfrac{1}{\sum_{k=1}^{i} \dfrac{1}{\dfrac{d_i}{n_i}}}}{\dfrac{d_{i+1}}{n_{i+1}}} \right) \tag{8.2}$$

where n is the cardinality of subclass i.

REFERENCES

[1] S. Mitra, "Efficient biometric authentication based on statistical models," pp. 227, 2005.

[2] R. de Luis-García, C. Alberola-López, O. Aghzout and J. Ruiz-Alzola, "Biometric identification systems," *Signal Process,* vol. 83, pp. 2539-2557, 12, 2003.

[3] Qinghan Xiao, "Technology review - Biometrics-Technology, Application, Challenge, and Computational Intelligence Solutions," *Computational Intelligence Magazine, IEEE,* vol. 2, pp. 5-25, 2007.

[4] K. P. Khushk and A. A. Iqbal, "An overview of leading biometrics technologies used for human identity," in *Engineering Sciences and Technology, 2005. SCONEST 2005. Student Conference on,* 2005, pp. 1-4.

[5] D. Voth, "Face recognition technology," *Intelligent Systems, IEEE,* vol. 18, pp. 4-7, 2003.

[6] A. K. Jain, A. Ross and S. Prabhakar, "An introduction to biometric recognition," *Circuits and Systems for Video Technology, IEEE Transactions on,* vol. 14, pp. 4-20, 2004.

[7] A. K. Jain, A. Ross and K. Nandakumar, "An introduction to biometrics," in *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on,* 2008, pp. 1-1.

[8] W. Zhao, R. Chellappa, A. Rosenfeld and P. J. Phillips, "Face Recognition: A Literature Survey," vol. 35, pp. 399-458, December, 2003.

[9] F. Ahmad and D. Mohamad, "A review on fingerprint classification techniques," in *Computer Technology and Development, 2009. ICCTD '09. International Conference on,* 2009, pp. 411-415.

[10] L. Collins, "Earmarked [biometrics]," *IEE Review,* vol. 51, pp. 38-40, 2005.

[11] A. K. Qin, P. N. Suganthan, C. H. Tay and H. S. Pa, "Personal identification system based on multiple palmprint features," in *Control, Automation, Robotics and Vision, 2006. ICARCV '06. 9th International Conference on,* 2006, pp. 1-6.

[12] W. Jia, D. Huang and D. Zhang, "Palmprint verification based on robust line orientation code," *Pattern Recognit,* vol. 41, pp. 1504-1513, 5, 2008.

[13] Leong Lai Fong and Woo Chaw Seng, "A comparison study on hand recognition approaches," in *Soft Computing and Pattern Recognition, 2009. SOCPAR '09. International Conference of,* 2009, pp. 364-368.

[14] M. S. Nixon and J. N. Carter, "Automatic Recognition by Gait," *Proceedings of the IEEE,* vol. 94, pp. 2013-2024, 2006.

[15] T. Barbu, "Comparing various voice recognition techniques," in *Speech Technology and Human-Computer Dialogue, 2009. SpeD '09. Proceedings of the 5-Th Conference on,* 2009, pp. 1-6.

[16] A. K. Jain, R. P. W. Duin and Jianchang Mao, "Statistical pattern recognition: a review," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 22, pp. 4-37, 2000.

[17] H. Kim, D. Kim and S. Yang Bang, "Face recognition using the mixture-of-eigenfaces method," *Pattern Recog. Lett.,* vol. 23, pp. 1549-1558, 11, 2002.

[18] M. O. Faruqe and M. Al Mehedi Hasan, "Face recognition using PCA and SVM," in *Anti-Counterfeiting, Security, and Identification in Communication, 2009. ASID 2009. 3rd International Conference on,* 2009, pp. 97-101.

[19] H. Kim, D. Kim and S. Y. Bang, "Face recognition using LDA mixture model," *Pattern Recog. Lett.,* vol. 24, pp. 2815-2821, 11, 2003.

[20] Yongguo Liu, Gang Chen, Jiwen Lu and Wanjun Chen, "Face recognition based on independent component analysis and fuzzy support vector machine," in *Intelligent Control and Automation, 2006. WCICA 2006. the Sixth World Congress on,* 2006, pp. 9889-9892.

[21] X. Zhang and Y. Gao, "Face recognition across pose: A review," *Pattern Recognit,* vol. 42, pp. 2876-2896, 11, 2009.

[22] Shou-Der Wei and Shang-Hong Lai, "Robust face recognition under lighting variations," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on,* 2004, pp. 354-357 Vol.1.

[23] Guang Dai and Changle Zhou, "Face recognition using support vector machines with the robust feature," in *Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003. the 12th IEEE International Workshop on,* 2003, pp. 49-53.

[24] H. J. Oh, K. M. Lee and S. U. Lee, "Occlusion invariant face recognition using selective local non-negative matrix factorization basis images," *Image Vision Comput.,* vol. 26, pp. 1515-1523, 11/1, 2008.

[25] R. Chellappa, C. L. Wilson and S. Sirohey, "Human and machine recognition of faces: a survey," *Proceedings of the IEEE,* vol. 83, pp. 705-741, 1995.

[26] Ming-Hsuan Yang, D. J. Kriegman and N. Ahuja, "Detecting faces in images: a survey," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 24, pp. 34-58, 2002.

[27] Zhao Fei and Qiao Qiang, "Face detection based on rectangular knowledge rule and face structure," in *Information Science and Engineering (ICISE), 2009 1st International Conference on,* 2009, pp. 1235-1239.

[28] W. Widjojo and Kin Choong Yow, "A color and feature-based approach to human face detection," in *Control, Automation, Robotics and Vision, 2002. ICARCV 2002. 7th International Conference on,* 2002, pp. 508-513 vol.1.

[29] P. Shih and C. Liu, "Face detection using discriminating feature analysis and Support Vector Machine," *Pattern Recognit,* vol. 39, pp. 260-276, 2, 2006.

[30] C. Shavers, R. Li and G. Lebby, "An SVM-based approach to face detection," in *System Theory, 2006. SSST '06. Proceeding of the Thirty-Eighth Southeastern Symposium on,* 2006, pp. 362-366.

[31] S. Jeng, H. Y. M. Liao, C. C. Han, M. Y. Chern and Y. T. Liu, "Facial feature detection using geometrical face model: An efficient approach," *Pattern Recognit,* vol. 31, pp. 273-282, 3, 1998.

[32] J. Wang and T. Tan, "A new face detection method based on shape information," *Pattern Recog. Lett.*, vol. 21, pp. 463-471, 6, 2000.

[33] L. M. Bergasa, M. Mazo, A. Gardel, M. A. Sotelo and L. Boquete, "Unsupervised and adaptive Gaussian skin-color model," *Image Vision Comput.*, vol. 18, pp. 987-1003, 9, 2000.

[34] A. K. Jain, Y. Zhong and M. Dubuisson-Jolly, "Deformable template models: A review," *Signal Process*, vol. 71, pp. 109-129, 12/15, 1998.

[35] M. -. Yang, N. Ahuja and D. Kriegman, "Face recognition using kernel eigenfaces," in *Image Processing, 2000. Proceedings. 2000 International Conference on*, 2000, pp. 37-40 vol.1.

[36] H. Takatsuka, M. Tanaka and M. Okutomi, "Distribution-based face detection using calibrated boosted cascade classifier," in *Image Analysis and Processing, 2007. ICIAP 2007. 14th International Conference on*, 2007, pp. 351-356.

[37] L. Huang, A. Shimizu, Y. Hagihara and H. Kobatake, "Face detection from cluttered images using a polynomial neural network," *Neurocomputing*, vol. 51, pp. 197-211, 4, 2003.

[38] S. L. Phung, A. Bouzerdoum, D. Chai and A. Watson, "Naive bayes face-nonface classifier: A study of preprocessing and feature extraction techniques," in *Image Processing, 2004. ICIP '04. 2004 International Conference on*, 2004, pp. 1385-1388 Vol.2.

[39] A. V. Nefian and M. H. Hayes III, "Face detection and recognition using hidden markov models," in *Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on*, 1998, pp. 141-145 vol.1.

[40] A. J. Colmenarez and T. S. Huang, "Face detection with information-based maximum discrimination," in *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, 1997, pp. 782-787.

[41] Lingmin Meng and T. Q. Nguyen, "Frontal face localization using linear discriminant," in *Signals, Systems, and Computers, 1999. Conference Record of the Thirty-Third Asilomar Conference on*, 1999, pp. 745-749 vol.1.

[42] T. Sakai, M. Nagao and S. Fujibayashi, "Line extraction and pattern detection in a photograph," *Pattern Recognit,* vol. 1, pp. 233-236, IN9-IN12, 237-248, 3, 1969.

[43] I. Craw, H. Ellis and J. R. Lishman, "Automatic extraction of face-features," *Pattern Recog. Lett.,* vol. 5, pp. 183-187, 2, 1987.

[44] V. Govindaraju, D. B. Sher, R. K. Srihari and S. N. Srihari, "Locating human faces in newspaper photographs," in *Computer Vision and Pattern Recognition, 1989. Proceedings CVPR '89., IEEE Computer Society Conference on,* 1989, pp. 549-554.

[45] A. Tsukamoto, Chil-Woo Lee and S. Tsuji, "Detection and pose estimation of human face with synthesized image models," in *Pattern Recognition, 1994. Vol. 1 - Conference A: Computer Vision & Image Processing., Proceedings of the 12th IAPR International Conference on,* 1994, pp. 754-757 vol.1.

[46] Roland T. Chin and Charles R. Dyer, "Model-based recognition in robot vision ," *ACM Comput. Surv.,* vol. 18, pp. 67-40, 1986.

[47] A. Lanitis, C. J. Taylor and T. F. Cootes, "Automatic interpretation and coding of face images using flexible models," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 19, pp. 743-756, 1997.

[48] A. Pentland, B. Moghaddam and T. Starner, "View-based and modular eigenspaces for face recognition," in *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on,* 1994, pp. 84-91.

[49] Lingmin.Meng and Truong.Q.Nguyen, "Frontal face detection using multi-segment and wavelet," in 1999, .

[50] D. Tsai and C. Lin, "Fast normalized cross correlation for defect detection," *Pattern Recog. Lett.,* vol. 24, pp. 2625-2631, 11, 2003.

[51] Shou-Der Wei and Shang-Hong Lai, "Fast Template Matching Based on Normalized Cross Correlation With Adaptive Multilevel Winner Update," *Image Processing, IEEE Transactions on,* vol. 17, pp. 2227-2235, 2008.

[52] M. J. Atallah, "Faster image template matching in the sum of the absolute value of differences measure," *Image Processing, IEEE Transactions on,* vol. 10, pp. 659-663, 2001.

[53] K. S. Fu and J. K. Mui, "A survey on image segmentation," *Pattern Recognit,* vol. 13, pp. 3-16, 1981.

[54] N. R. Pal and S. K. Pal, "A review on image segmentation techniques," *Pattern Recognit,* vol. 26, pp. 1277-1294, 9, 1993.

[55] S. -. Wan and W. E. Higgins, "Symmetric region growing," *Image Processing, IEEE Transactions on,* vol. 12, pp. 1007-1015, 2003.

[56] Y. Chang and X. Li, "Fast image region growing," *Image Vision Comput.,* vol. 13, pp. 559-571, 9, 1995.

[57] T. N. Pappas and N. S. Jayant, "An adaptive clustering algorithm for image segmentation," in *Acoustics, Speech, and Signal Processing, 1989. ICASSP-89., 1989 International Conference on,* 1989, pp. 1667-1670 vol.3.

[58] T. N. Pappas, "An adaptive clustering algorithm for image segmentation," *Signal Processing, IEEE Transactions on,* vol. 40, pp. 901-914, 1992.

[59] N. A. M. Isa, S. A. Salamah and U. K. Ngah, "Adaptive fuzzy moving K-means clustering algorithm for image segmentation," *Consumer Electronics, IEEE Transactions on,* vol. 55, pp. 2145-2153, 2009.

[60] S. N. Sulaiman and N. A. M. Isa, "Adaptive fuzzy-K-means clustering algorithm for image segmentation," *Consumer Electronics, IEEE Transactions on,* vol. 56, pp. 2661-2668, 2010.

[61] W. Cai, S. Chen and D. Zhang, "Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation," *Pattern Recognit,* vol. 40, pp. 825-838, 3, 2007.

[62] Brahim Belhaouari Samir, "Modified k-means cluster," 2008.

[63] H. R. Kanan and M. S. Moin, "Face recognition using entropy weighted patch PCA array under variation of lighting conditions from a single sample image per person," in *Information, Communications and Signal Processing, 2009. ICICS 2009. 7th International Conference on,* 2009, pp. 1-5.

[64] S. Thakur, J. K. Sing, D. K. Basu, M. Nasipuri and M. Kundu, "Face recognition using principal component analysis and RBF neural networks," in *Emerging Trends in Engineering and Technology, 2008. ICETET '08. First International Conference on,* 2008, pp. 695-700.

[65] Tangquan Qi, Huiwen Deng and Weiping Hu, "Face recognition using eigenfaces-fisher linear discriminant and dynamic fuzzy neural network," in *Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on,* 2010, pp. 166-170.

[66] L. Chen, H. M. Liao, M. Ko, J. Lin and G. Yu, "A new LDA-based face recognition system which can solve the small sample size problem," *Pattern Recognit,* vol. 33, pp. 1713-1726, 10, 2000.

[67] J. Lu, K. N. Plataniotis and A. N. Venetsanopoulos, "Regularization studies of linear discriminant analysis in small sample size scenarios with application to face recognition," *Pattern Recog. Lett.,* vol. 26, pp. 181-191, 1/15, 2005.

[68] T. Kim, H. Kim, W. Hwang and J. Kittler, "Independent component analysis in a local facial residue space for face recognition," *Pattern Recognit,* vol. 37, pp. 1873-1885, 9, 2004.

[69] R. Kam-art, T. Raicharoen and V. Khera, "Face recognition using feature extraction based on descriptive statistics of a face image," in *Machine Learning and Cybernetics, 2009 International Conference on,* 2009, pp. 193-197.

[70] N. Jamil, S. Lqbal and N. Iqbal, "Face recognition using neural networks," in *Multi Topic Conference, 2001. IEEE INMIC 2001. Technology for the 21st Century. Proceedings. IEEE International,* 2001, pp. 277-281.

[71] X. Jing, H. Wong and D. Zhang, "Face recognition based on discriminant fractional Fourier feature extraction," *Pattern Recog. Lett.,* vol. 27, pp. 1465-1471, 10/1, 2006.

[72] D. V. Jadhav and R. S. Holambe, "Radon and discrete cosine transforms based feature extraction and dimensionality reduction approach for face recognition," *Signal Process,* vol. 88, pp. 2604-2609, 10, 2008.

[73] Hong Wang, Su Yang and Wei Liao, "An improved PCA face recognition algorithm based on the discrete wavelet transform and the support vector machines," in *Computational Intelligence and Security Workshops, 2007. CISW 2007. International Conference on,* 2007, pp. 308-311.

[74] Xiang-fei Nie, "Face recognition using wavelet transform and kernel principal component analysis," in *Future Information Technology and Management Engineering (FITME), 2010 International Conference on*, 2010, pp. 186-189.

[75] Ki-Chung Chung, Seok Cheol Kee and Sang Ryong Kim, "Face recognition using principal component analysis of gabor filter responses," in *Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, 1999. Proceedings. International Workshop on*, 1999, pp. 53-57.

[76] Zheng Dezhong and C. Fayi, "Face recognition based on wavelet transform and image comparison," in *Computational Intelligence and Design, 2008. ISCID '08. International Symposium on*, 2008, pp. 24-29.

[77] Z. Zhang, G. Wang, X. Lin and Q. Wu, "Face recognition based on wavelet-curvelet-fractal technique," in Anonymous Springer Berlin / Heidelberg, 2009, pp. 532-540.

[78] Cunjian Chen and Jiashu Zhang, "Wavelet energy entropy as a new feature extractor for face recognition," in *Image and Graphics, 2007. ICIG 2007. Fourth International Conference on*, 2007, pp. 616-619.

[79] Waset,"http://www.waset.org/journals/waset/v49/v49-182.pdfvol,", 2009.

[80] R. Lotlikar and R. Kothari, "Face recognition using curvilinear component analysis," in *Neural Networks Proceedings, 1998. IEEE World Congress on Computational Intelligence. the 1998 IEEE International Joint Conference on*, 1998, pp. 1778-1783 vol.3.

[81] C. Chen, Zhang Shi-qing and Chen Yue-fen, "Face recognition based on MPCA," in *Industrial Mechatronics and Automation (ICIMA), 2010 2nd International Conference on*, 2010, pp. 322-325.

[82] Pallabi Parveen and Bhavani Thuraisingham, "Face recognition using multiple classifiers," in *Tools with Artificial Intelligence, 2006. ICTAI '06. 18th IEEE International Conference on,* 2006, pp. 179-186.

[83] S. Ranganath and K. Arun, "Face recognition using transform features and neural networks," *Pattern Recognit,* vol. 30, pp. 1615-1622, 10, 1997.

[84] A. V. Nefian and M. H. Hayes III, "Face detection and recognition using hidden markov models," in *Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on,* 1998, pp. 141-145 vol.1.

[85] Ti-QiongXu, Bi-Cheng Li and Bo Wang, "Face detection and recognition using neural network and hidden markov models," in *Neural Networks and Signal Processing, 2003. Proceedings of the 2003 International Conference on,* 2003, pp. 228-231 Vol.1.

[86] Qingmiao Wang and Shiguang Ju, "A mixed classifier based on combination of HMM and KNN," in *Natural Computation, 2008. ICNC '08. Fourth International Conference on,* 2008, pp. 38-42.

[87] Rafael C. Gonzalez, Richard E. Woods and Steven L. Eddins, *Digital Image Processing.* Prentice Hall, 2002.

[88] David John Cameron MacKay, "An example inference task: Clustering," in *Information Theory, Inference and Learning Algorithms*Anonymous United Kingdom: Cambridge University Press, 2003, pp. 284-8.

[89] Gilbert G. Walter and Xiaoping Shen, "An introduction to orthogonal wavelet theory " in *Wavelets and Other Orthogonal Systems*Anonymous United State of America: CRC Press, 2001, pp. 28-20.

[90] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 11, pp. 674-693, 1989.

[91] Yale Face Database, "http://cvc.yale.edu/projects/yalefaces/yalefaces.html," vol. 2010, .

[92] CBCL Database, "http://cbcl.mit.edu/software-datasets/heisele/facerecognition-database.html," vol. 2010, .

[93] ORL, "http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html, " vol. 2010, .

[94] Face94 dataset, "http://cswww.essex.ac.uk/mv/allfaces/index.html," vol. 2010, .

[95] Jun-Ying Gan and Si-Bin He, "Face recognition based on 2DLDA and support vector machine," in *Wavelet Analysis and Pattern Recognition, 2009. ICWAPR 2009. International Conference on*, 2009, pp. 211-214.

[96] http://vis-www.cs.umass.edu/~vidit/IndianFaceDatabase/.

[97] "BioID : BioID Support : Downloads : BioID Face Database " vol. 2011, .

[98] http://www.vision.caltech.edu/html-files/archive.html.

[99] B. Froba and C. Kublbeck, "Orientation template matching for face localization in complex visual scenes," in Image Processing, 2000. Proceedings. 2000 International Conference on, 2000, pp. 251-254 vol.2.

[100] M. Vatsa, R. Singh and P. Gupta, "Face detection using gradient vector flow," in Machine Learning and Cybernetics, 2003 International Conference on, 2003, pp. 3259-3263 Vol.5.

[101] Jiatao Song, Jilin Liu, Zheru Chi and Wei Wang, "Locatization of human eyes based on a series of binary images," in Multimedia and Expo, 2004. ICME '04. 2004 IEEE International Conference on, 2004, pp. 1199-1202 Vol.2.

[102] M. H. Khan, N. A. Khan and A. Zuberi, "Fast and accurate localization of human faces," in Communications and Information Technology, 2004. ISCIT 2004. IEEE International Symposium on, 2004, pp. 1232-1237 vol.2.

[103] Thai Hoang Le and Len Tien Bui, "A hybrid approach of AdaBoost and artificial neural network for detecting human faces," in Research, Innovation and Vision for the Future, 2008. RIVF 2008. IEEE International Conference on, 2008, pp. 79-85.

[104] M.Misiti, Y. Misiti, G. Oppenheim and J. M. Poggi,"Wavelets and their Applications," Digital Signal and Image Processing Series, 2006.

[105] Pang-Ning Tan, Michael Steinbach and Vipin Kumar, *Introduction to Data Mining.* Pearson Addison-Wesley, 2006.

[106] Hui Xiong, Junjie Wu and Jian Chen, "K-Means Clustering Versus Validation Measures: A Data-Distribution Perspective," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on,* vol. 39, pp. 318-331, 2009.

[107] I, borg and patrick J.F. Groenen, *Modern Multidimensional Scaling—Theory and Applications.* Springer science+business media, 2005.

[108] Sudipto Guha, Rajeev Rastogi and Kyuseok Shim, "CURE: an efficient clustering algorithm for large databases," vol. 27, pp. 73-84, 1998.

[109] E. Knorr, R. Ng and V. Tucakov, "Distance-based outliers: algorithms and applications," vol. 8, pp. 237-253, 2000.

[110] P.S. Bradley, Usama Fayyad and Cory Reina, "Scaling clustering algorithms to large databases," in *4th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining,* 1998, pp. 9-15.

[111] Mohd Yusoff Mashor, "Hybrid Training Algorithm for RBF Network," vol. 8, pp. 50-65, 2000.

PUBLICATIONS

This study has been the basis for the following peer-reviewed journals papers and conferences talk.

- Peer-reviewed journals papers:

1. **Nadir Nourain Dawoud**, Brahim Belhaouari Samir,” Statistical Model and Wavelet Function for Face Recognition”, *Journal of Applied Science*, Vol.11(7), PP.1213-1218, March 2011.

2. **Nadir Nourain Dawoud,** Brahim Belhaouari Samir, Josefina Janier,” Fast Template Matching Method Based Optimized Sum of Absolute Difference Algorithm for Face Localization”, *International Journal of Computer Applications*, Vol. 18(8), PP. 30-34, March 2011.

3. **Nadir Nourain Dawoud,** Brahim Belhaouari Samir, Josefina Janier,” N-Mean Kernel Filter and Normalized Correlation for Face Localization”, *published in IEEE trans of the 7th International Colloquium on Signal Processing & Its Applications*, pp.416-419, May 2011.

4. **Nadir Nourain Dawoud**, Brahim Belhaouari Samir, Josefina Janier,” Fast Template Matching Method Based on Optimized Metrics for Face Localization”, accepted for publishing in IAENG International Journal of Applied Mathematics of International MultiConference of Engineers and Computer Scientists (IMECS).

- Submitted Journal Papers

5. **Nadir Nourain Dawoud,** Brahim Belhaouari Samir,” Cluster-k-Nearest-Neighbor (k-NN) Classifier for Face Classification and Recognition,”.

6. **Nadir Nourain Dawoud,** Brahim Belhaouari Samir,” Cluster-k-Nearest-Neighbor (C-k-NN) Classifier and its Classification Response Metric for Face Classification,”.

7. **Nadir Nourain Dawoud,** Brahim Belhaouari Samir,” Face Localization Using Template Matching Method Based on New Optimized Metrics ,”.

8. **Nadir Nourain Dawoud,** Brahim Belhaouari Samir,” A New Features Extraction and Selection Method based on Wavelet Transform and Statistical Method for Face Classification ,”.

9. **Nadir Nourain Dawoud,** Brahim Belhaouari Samir,” A New K-Means Modified Algorithm for Image Segimentation,”.

- Refereed Conferences:

10. **Nadir Nourain Dawoud**, Brahim Belhaouari Samir,” Statistical Model and Wavelet Function for Face Recognition”, *Proceeding of International Conference on Fundamental and Applied Science*, 15 – 17 June 2010, Kuala Lumpur, Malaysia.

11. **Nadir Nourain Dawoud,** Brahim Belhaouari Samir, Josefina Janier,” N-Mean Kernel Filter and Normalized Correlation for Face Localization”, *Proceeding of 2011 the 7th International Colloquium on Signal Processing & Its Applications*, 4-6 March 2011, Penang, Malaysia.

12. **Nadir Nourain Dawoud,** Brahim Belhaouari Samir, Josefina Janier,” Fast Template Matching Method Based on Optimized Metrics for Face Localization”, *proceeding of International MultiConference of Engineers and Computer Scientists (IMECS),* 16-18 March 2011, Hong Kong.