12-2013

Face recognition using statistical adapted local binary patterns.

Abdallah Abd-Elghafar Mohamed
University of Louisville

FACE RECOGNITION USING STATISTICAL ADAPTED LOCAL BINARY PATTERNS

By

Abdallah Abd-Elghafar Mohamed
B.Sc., Menoufia University, Egypt 1997
M.Sc., Indiana University Bloomington, IN, USA 2009

Dissertation
Submitted to the Faculty of
Speed School of Engineering of the University of Louisville
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

Department of Computer Engineering and Computer Science
University of Louisville
Louisville, Kentucky

December 2013

FACE RECOGNITION USING STATISTICAL ADAPTED LOCAL BINARY PATTERNS

By

Abdallah Abd-Elghafar Mohamed
B.Sc., Menoufia University, Egypt, 1997
M.Sc., Indiana University Bloomington, IN, USA, 2009

Dissertation Approved on

November 22, 2013

by the following Dissertation Committee:

____
Dr. Roman V. Yampolskiy

____
Dr. Adel S. Elmaghraby

____
Dr. Ibrahim Imam

____
Dr. Dar-jen Chang

____
Dr. C. Tim Hardin

DEDICATION

This dissertation is dedicated to my sons

Hamza and Abdelrahman

and

to my wife Boshra

who has helped me a lot during all my research time.

ACKNOWLEDGMENTS

I would like to take the opportunity to appreciate my supervisor Dr. Roman Yampolskiy, for his invaluable guidance and patience throughout the project. I would also like to thank the other committee members, Dr. Adel Elmaghraby, Dr. Ibrahim Imam, Dr. Dar-jen Chang and Dr. C. Tim Hardin, for their comments and assistance. I would also like to express my thanks to my wife, Boshra, for her understanding and patience during those times when there was no light at the end of anything. She encouraged me and made me stick with it.

Finally, I would like to thank my family members and my friends for all their support and encouragement.

ABSTRACT

FACE RECOGNITION USING STATISTICAL ADAPTED LOCAL BINARY PATTERNS

Abdallah A. Mohamed

November 22, 2013

Biometrics is the study of methods of recognizing humans based on their behavioral and physical characteristics or traits. Face recognition is one of the biometric modalities that received a great amount of attention from many researchers during the past few decades because of its potential applications in a variety of security domains. Face recognition however is not only concerned with recognizing human faces, but also with recognizing faces of non-biological entities or avatars. Fortunately, the need for secure and affordable virtual worlds is attracting the attention of many researchers who seek to find fast, automatic and reliable ways to identify virtual worlds' avatars.

In this work, I propose new techniques for recognizing avatar faces, which also can be applied to recognize human faces. Proposed methods are based mainly on a well-known and efficient local texture descriptor, Local Binary Pattern (LBP). I am applying different versions of LBP such as: Hierarchical Multi-scale Local Binary Patterns and Adaptive Local Binary Pattern with Directional Statistical Features in the wavelet space and discuss the effect of this application on the performance of each LBP version. In addition, I use a new version of LBP called Local Difference Pattern (LDP) with other well-known descriptors and classifiers to differentiate between human and avatar face images.

The original LBP achieves high recognition rate if the tested images are pure but its performance gets worse if these images are corrupted by noise. To deal with this problem I propose a new definition to the original LBP in which the LBP descriptor will not threshold all the neighborhood pixel based on the central pixel value. A weight for each pixel in the neighborhood will be computed, a new value for each pixel will be calculated and then using simple statistical operations will be used to compute the new

threshold, which will change automatically, based on the pixel's values. This threshold can be applied with the original LBP or any other version of LBP and can be extended to work with Local Ternary Pattern (LTP) or any version of LTP to produce different versions of LTP for recognizing noisy avatar and human faces images.

TABLE OF CONTENTS

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

FIGURE PAGE

CHAPTER 1

INTRODUCTION

Biometrics research investigates methods and techniques for recognizing humans based on their behavioral and physical characteristics or traits (Jain, Ross, & Prabhakar, 2004; Mohamed et al., 2011; Mohamed et al., 2012; Mohamed & Yampolskiy, 2012d; Wayman, 2001; Zhenhua, Lei, Zhang, & Xuanqin, 2010). Face recognition is a biomteric trait and it is something that people usually perform effortlessly and routinely in their everyday life and it is the process of identifying individuals from their faces' intrinsic characteristics. Automated face recognition has become one of the main targets of investigation for researchers in biometrics, pattern recognition, computer vision, and machine learning communities. This interest is driven by a wide range of commercial and law enforcement practical applications that require the use of face recognition technologies (Mohamed et al., 2012; Mohamed & Yampolskiy, 2012d). These applications include access control, automated crowd surveillance, face reconstruction, mugshot identification, human-computer interaction and multimedia communication (Haiping, Martin, Bui, Plataniotis, & Hatzinakos, 2009; Mohamed et al., 2012; Mohamed & Yampolskiy, 2012d; Phillips, Martin, Wilson, & Przybocki, 2000; Wayman, 2001).

Face recognition systems have many advantages over traditional security systems: the biometric identification of a person can not be lost, forgotten like complex passwords and PIN codes or easy to be guessed by an illegitimate user like short and simple passwords (Chan, 2008; Li & Jain, 2011).

Face recognition has many advantages over the other biometric traits, such as fingerprint, voice, iris, hand-geometry and signature. Besides being non-intrusive, more natural and easy to use it can also be captured at a distance and in a covert manner (Senior & Bolle, 2011). Since the first automated face recognition system which was developed by Kanade (Kanade, 1973), substantial attention has been given to face recognition. Facial features have the highest suitability among the other six biometric traits (face, finger, hand, voice, eye and signature) considered by Hietmeyer in a machine readable travel documents (MRTD) based on (Haiping et al., 2009; Hietmeyer, 2000; "Machine Readable Travel Documents (MRTD),"): enrollment, renewal, machine requirements and public perception.

Due to the growth of computer power, storage and recent techniques in pattern recognition, face recognition systems can now be applied to solve real life problems and achieve considerable accuracy rates under controlled conditions especially when there is sufficient number of face images in the training database. However, it has turned out to be difficult when face images have been acquired under unconstrained environment where illumination, expression, accessories and so on vary considerably (Li & Jain, 2011; Zhenhua, Lei, Zhang, & Xuanqin, 2010).

1.1 Face Recognition Systems Classification

Face recognition systems can be classified into two types (modes of operation): (i) face verification (or authentication) and (ii) face identification (or recognition) (Chan, 2008; Jain et al., 2004; Li & Jain, 2011; Poli, Arcot, & Charapanamjeri, 2009; Wayman, 2001). A face verification system involves a one to one matching to confirm or deny a person's identity claim. This system compares the captured face image against the person's template(s) stored in the system. If the person presenting himself/herself to the system is the person he/she claims to be then the system will accept that person (client) otherwise the system will reject that person (impostor). There are many applications that require face verification mode, such as mobile or computer log-in, building gate control and E-passport.

On the other hand, a face identification system involves one to many matching. In this system, the captured face image will be compared against all face images stored in the enrollment database to associate the identity of the captured face image to one of those face images stored in the database (Chan, 2008; Jain et al., 2004; Li & Jain, 2011; Poli et al., 2009). So, the system will either make a match and identify the person or fail to make a match and then will not identify that person. In some face identification application systems, the system just tries to find the most similar face image in the database to the captured one. There are many applications that require face identification mode, such as information retrieval (police database), human computer interaction (video games) and video surveillance.

There are many factors that have a direct effect on the performance of the face recognition system. These factors include facial expressions, head pose, lighting conditions (contrast, shadows), age span, hair, occlusions (glasses, make-up) and facial features (beard) (Singh, Vatsa, & Noore, 2008). Based on these factors the face recognition applications can be classified into: (i) cooperative user scenarios and (ii) non-cooperative user scenarios based on the user cooperation with the system (Li & Jain, 2011).

In the cooperative applications, the user of the system has to cooperate with the system by presenting his/her face in a proper way (such as presenting the frontal face pose with natural expressions and open eyes as in the e-passport and physical access control systems) in order to gain access to the system. In the non-cooperative applications the user does not know that he/she is being identified as in street surveillance (Li & Jain, 2011). The most challenging non-cooperative application is the watchlist identification problem.

1.2 Face Recognition System Modules

Face recognition system has four modules: face detection, image normalization, feature extraction and classification (Jain et al., 2004; Mohamed et al., 2012; Mohamed & Yampolskiy, 2012d; Wayman, 2001). Face detection module isolates the face area from the background. The presence of the background with the face image in the same image has an effect on the performance of the recognition system (Jain et al., 2004; Mohamed et al., 2012; Mohamed & Yampolskiy, 2012d; Wayman, 2001).

Varying illumination and pose or expression of the face image can affect the accuracy rate of the recognition system (Jain et al., 2004; Li & Jain, 2011; Wayman, 2001). To reduce the effect of these factors pure face images have to be normalized geometrically and photometrically. The pure face image will be transformed into a standard frame under the effect of the geometrical normalization process while the photometric normalization process normalizes faces based on properties such as illumination.

Face feature extraction module is the module that is responsible for extracting prominent useful information (set of distinguishable features) from the normalized face image. Such information is the key for distinguishing between different faces and the accuracy of any recognition system highly depends on the features extracted from this stage to represent each facial image (Jain et al., 2004; Li & Jain, 2011; Wayman, 2001).

Face classification module is the last stage of any recognition system and it highly depends on the application itself. In the case of identification systems, the features extracted from the input face image have to be matched against features extracted from all face images stored in the database (one to many matcher). The result will be either the identification of the input face image when the highest score match found or the facial image will be unknown if the match score is below the threshold value. In case of the verification applications, the features extracted from the input image have to be matched against the features extracted from one of the enrolled face image (one to one matcher) and the classifier results in either “yes”and accepts this input face image if the match happens or “no” and rejects the input image if there is no match (Jain et al., 2004; Li & Jain, 2011; Wayman, 2001).

1.3 Challenges of Face Recognition

Human visual system can easily identify familiar human faces even if they are observed under challenging viewing conditions such as changing in expression, illumination, occlusion and so on. Automated face recognition systems achieve good results in recognizing facial images captured under constrained conditions. They still have some problems in achieving high performance rates under variations in illumination, occlusion, expression and viewpoint.

Face recognition is not limited only to recognize human faces but it should also work for recognizing faces of non-biological entities such as avatars from virtual worlds. Virtual worlds have millions of avatars which have a strong resemblance to their human owners but how I can differentiate between avatar face images and human face images and between an avatar and another avatar from the same virtual world. Till now, the work done to recognize avatar face images is still very limited and it does not differentiate between human face images and avatar face images.

Consider that there are many techniques that can recognize faces correctly with high recognition rate, what will be the performance of these techniques under the environment of noisy facial images? Many techniques such as local binary patterns (LBP) can not deal with this problem. LBP thresholds all pixels in a specific neighborhood based on the value of the central pixel of that neighborhood to compute a new value for this central pixel. So, if the central pixel is corrupted by noise for any reason the comparison between this corrupted pixel and its neighbors will not be accurate. Also, according to LBP strategy, assigning the value 1 to all pixels greater than or equal to the central pixel value and assigning the value 0 to all pixels less than the central pixel produces inferior. The system may find a pixel with a value which is a little bit less than the central pixel value and there is another pixel which has a value significantly less than the value of the central pixel. Based on the LBP definition both of the two pixels will assign the value 0 and this is undesirable.

1.4 Contributions

The contributions of this thesis to the methodology of face recognition are summarized as follows:

I apply many existing techniques in the wavelet domain and discuss the effect of this supplication on the accuracy rates for recognizing both human and avatar face images. These techniques include traditional local binary pattern (single scale LBP), multi-scale LBP (MLBP), hierarchical multi-scale LBP (HMLBP), adaptive LBP (ALBP), adaptive LBP with directional statistical features (ALBPDSF) and multi-scale adaptive LBP with directional statistical features (MALBPDSF).

To distinguish between human face images and avatar face images for avatar CAPTCHA (Completely

Automated Public Turing test to tell Computers and Humans Apart) challenge, I apply a variety of learning-based recognition approaches to the task of classifying between human and avatar faces. These approaches include, Naïve approaches, which include raw images, summary statistics and grayscale histogram, histograms of oriented gradients (HOG), GIST descriptor, quantized feature descriptors, which include scale-invariant feature transform (SIFT) and speed-up robust features (SURF) and local binary pattern-based features, which include four-patch local binary pattern (FPLBP) and a new developed LBP version called local difference pattern (LDP). For learning and classification for models from all the previous approaches I apply two different types of classifiers: Naïve Bayes, and LibLinear with L2-regularized logistic regression.

To overcome the sensitivity of LBP to noise, I redefine the LBP descriptor with a new definition. In this new definition, all pixels values will change based on its weight in any neighborhood. The new definition of LBP will have an automatically changeable threshold based on the new pixels values and simple statistical operations and not a fixed threshold based on the central pixels value of any neighborhood. I call the new definition of LBP statistical adapted LBP (SALBP). However, since SALBP is single scale version of LBP I go one-step beyond and build the multi-scale version of SALBP, multi-scale SALBP (MSALBP), and the hierarchical multi-scale version of SALBP, hierarchical multi-scale SALBP (HMSALBP). To generalize my work, I build the local ternary pattern version of SALBP and I call it adaptive extended local ternary pattern (AELTP). In addition, I build the multi-scale version of AELTP, multi-scale adaptive extended local ternary pattern (MAELTP) and the hierarchical version of MAELTP, hierarchical multi-scale adaptive extended local ternary pattern (HMAELTP). To evaluate the effect of wavelet domain over all the previous techniques I apply all of them in the wavelet domain and compare between their results in wavelet domain and out of the wavelet domain.

1.5 Overview of Thesis

The outline of the thesis is described below:

Virtual World and Avatars: Chapter 2 includes the virtual worlds definition with examples and pictures from some existing virtual worlds, describing the common purposes for creating and using virtual worlds, defining how individuals are represented in virtual worlds by creating their avatars, relationship between avatars and their creators, some avatar‘s features and how avatars can communicate. Also, explaining why there is an essential need to identify the identity of avatars.

Overview of Human and Avatar face Recognition: some of the most common and widely used techniques in recognizing human faces are listed in Chapter 3 including the type of each technique, the dataset(s) that it recognized its faces on and the accuracy rate that it satisfied. In addition to human face recognition techniques chapter 3 also includes techniques of how to create an avatar dataset, how to detect an avatar face area and how to recognize an avatar face.

Face Recognition and Local Binary Patterns: The structure of discrete wavelet transforms, how to apply them in face recognition and a powerful texture descriptor, called Local Binary Pattern (LBP), and its variants developed for face recognition, have been introduced in Chapter 4 and Chapter 5. However, the original local binary pattern is operating in a single scale space; limit the robustness of the representation to image translation. Therefore, it should be possible to enhance the robustness by extending the representation method to multiresolution by combining the idea of wavelets with multi-scale representation of LBP (MLBP) and Hierarchical Multi-scale LBP (HMLBP), Adaptive LBP with directional statistical features (ALBPDSF) and Multi-scale ALBPDSF. Experiments are carried out on different avatar and human datasets and the results show that proposed techniques outperform other state-of-art contenders.

Avatar CAPTCHA: Avatar CAPTCHA is introduced in Chapter 6 as a challenge presented in ICMLA 2012 conference. This CAPTCHA system presents 12 images in two rows each one has six images, each image either of human or of an avatar, the user’s task is to select all avatar images among these 12 images. In chapter 6 I showed that using machine learning techniques we can achieve significantly higher performance than random guessing and outperform humans. In Chapter 6, a novel LBP representation called Local Difference Pattern (LDP) is proposed and the obtaining results is a proof for its superiority in distinguishing between human and avatar faces.

Statistical Adapted Local Binary Techniques: In chapter 7 two novel representations, called Statistical Adaptive LBP (SALBP) and Adaptive Extended LBP (AELTP), are proposed to treat LBP weaknesses. To extend SALBP and AELTP, I also proposed Multi-scale SALBP (MSALBP), Hierarchical MSALBP (HMSALBP), Multi-scale AELTP (MAELTP) and Hierarchical MAELTP (HMAELTP) to provide tools for multi-resolution analysis of faces. Experiments are carried out and the results show that Statistical adapted techniques outperform other state-of-art traditional contenders.

Conclusions and Future Work: The thesis is drawn to conclusion in Chapter 8 where the directions of future work are also suggested.

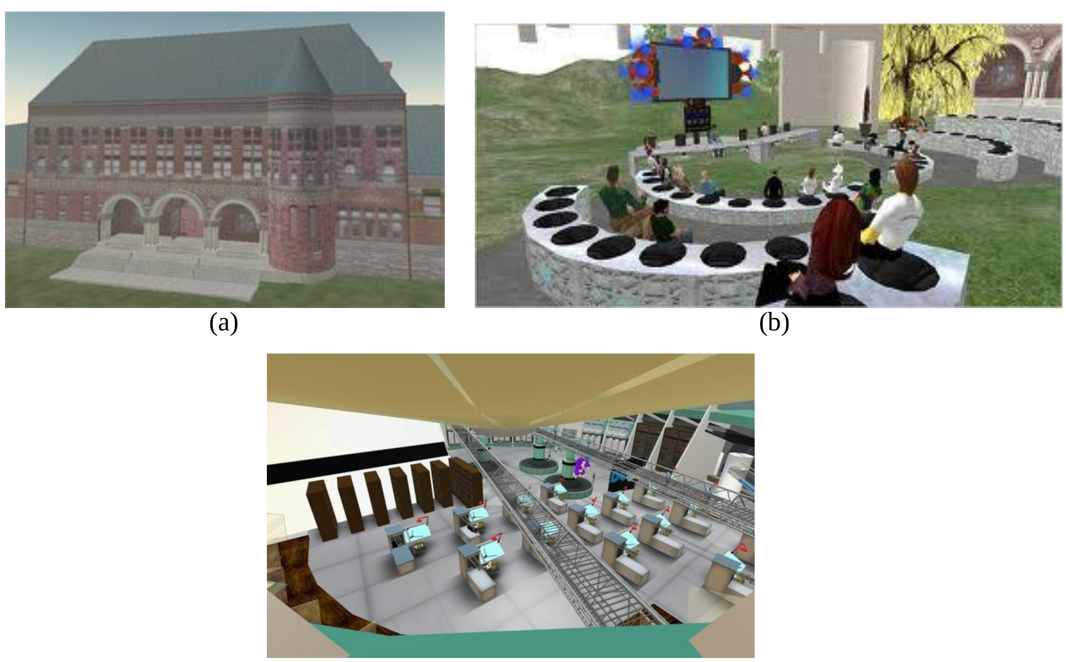
CHAPTER 2

VIRTUAL WORLDS AND AVATARS

2.1 Virtual Worlds

Becoming an indispensable part of today's modern life, the internet has added new contexts for daily activities. Specifically, one of the major breakthroughs of the World Wide Web is that it facilitates the creation of interactive web pages that can be accessed worldwide (Thompson, 2011). The roles these web pages play range from facilitating simple communications (e.g., emails, chat, etc.) to more complex ways of communicating including video conferencing and banking. One of the most recent and fast growing applications of these interactive web pages is what has been called three-dimensional virtual worlds. In these virtual worlds (Virtual Reality), computer graphics are manipulated to render simultaneous, interactive, and three-dimensional environments, which mimics real world environments (Dyck et al., 2008). Designed this way, virtual worlds look realistic to the user to a great extent. This virtual reality thus provides the user with a personal digital space where he or she can perform real world activities. Individuals as well as groups sharing common interests and activities can communicate across the world easily (Trewin, Laff, Hanson, & Cavender, 2009). Accessing these worlds is becoming easier and easier with technology advancement. The presence of virtual worlds and their being easily accessed may lead to transformation of the operation of whole societies. With advancement in building Massively Multiplayer Online Games (MMOG), virtual worlds became even more accessible and popular (Thompson, 2011).

At the present time, there are several well-known virtual world online applications such as Second Life ("Second Life,"), Entropia Universe ("Entropia Universe,"), Sims Online ("Sims Online,") and Active Worlds ("Active Worlds,"). Second Life for instance is a multi-user online three-dimensional virtual world, which includes up to 20 million registered users. It facilitates education, socializing, shopping, starting small businesses and enterprises as well as making money ("Second Life,"). The diversity of interests that can exist in a virtual world is clearly shown in the activities that are facilitated by second life as well as other worlds. Thus, real businesses can exist and actually flourish in virtual worlds. Realizing how popular these sites are becoming, well know companies, TV and radio channels as well as prestigious schools are using them. Reuters for instance has built a virtual headquarters in Second Life so that it would be able to broadcast news not only in the real world but to the virtual one as well. News broadcast sessions have been broadcast by the National Public Radio through Second Life as well. IBM arranged for a gathering of its employees also in Second Life. Universities are building islands in virtual worlds where classes can be offered. For instance, Harvard Law School offers a CyberOne course partly on Berkman Island in Second Life ("Harvard Law Class in Second Life "). Indiana University's Kelly school of business has established a presence also in Second Life virtual world (Boukhris et al., 2011; "Kelly School of Business,"). Companies like Dell, Cisco, Xerox and Nissan have stores within Second Life. Virtual worlds thus host and offer different activities for its residents. They have been used as environments for games and adventure.



(a) (b)

(c)

Figure 1: Second Life images ("Second Life,"): a) Harvard Law School in Second Life
b) A Harvard Law School lecture in Second Life c) DELL in Second Life.

For example in Everquest and World of Craft which are examples of Massively Multiplayer Online Role Playing Games (MMORPGs).

The main activity that the virtual world establishes is the creation of an entertaining virtual world for games. Unlike games, adventure based virtual worlds, offer computer mediated environment so that the residents would interact free of a dictated plot or a specific story or adventure line. Music Television (aka MTV) established a virtual world (i.e., MTV's Virtual Laguna Beach) where users can have access to the MTV Laguna Beach television and can interact live with family and friends. MTV future plans include holding virtual music concerts as well (Bray & Konsynski, 2007).

2.2 Avatars

Originally, the word avatar comes from a religious Hindu expression meaning the appearance or the manifestation of a god in human or super human form ("The Free Dictionary,"). An avatar is simply a digital identity of a user. An avatar is a representation of the user that enables interaction in 3D or in 2D contexts. Users usually prefer to have social presence in these worlds by creating distinct and different avatars. The created avatars sometimes refer to user's own personality or to a made-up identity. Although the avatar is a representation of a user identity, it is still not authentic. Users have the choice of how they would look like as well as how they can express themselves via such chosen appearance. Some users might make decisions to disclose facts about themselves with their choice of the appearance of an avatar. Others might use a popular image as their avatar. The same avatar can be used by a user in different online sessions. Some of the avatars mirror a user's role in virtual world which is reflected by an outfit or a specific appearance. Some users avatars are given a realistic look that resembles a human being. Users who tend to make such realistic choice of avatar appearance believe this would help them create a closer connection with their avatars. Some online websites restrict avatar identities to one per a single user. This requirement would avoid problems of trust, as a user will not be able to use alternative identities. Avatars have different aspects that include animations, emotions, gestures, speech, and voice.

Virtual world service providers require that a user gives up his or her rights of the avatar they created or chose to the providers. Subsequently, this agreement makes ownership of an avatar a debatable issue. Virtual world service providers also have the right to terminate an avatar as well as its user's account (Boberg, Piippo, & Ollila, 2008).

The figures below show examples of avatars from Second Life and Entropia Universe. There is a relationship between how an avatar would look like and how the user would behave within virtual worlds. For instance, users who create attractive avatars usually reveal more information to strangers more than users with unattractive avatars. In addition, tall avatars correspond to a confident user especially during tasks requiring decision-making. Realistic looking avatars show a great deal of positive social interactions.

(a) (b)

Figure 2: Avatar images: a) Second Life avatar ("Second Life,") b) Entropia Universe avatar ("Entropia Universe,").

It has been noticed that users would treat avatars warmly if the avatar looks similar to them (Neustaedter & Fedorovskaya, 2009). Within virtual worlds, an avatar has the ability to move within its 3D or 2D environment to execute a task. Important characteristics of this society is sharing and trading which maintain and increase the unity within avatar groups. Communication is a very essential characteristic of an avatar as it maintains interactions with other users in the virtual world. Communication can take different forms. It could be 1) Verbal, 2) non-Verbal, 3) Asynchronous, 4) Synchronous and 5) direct. Users can communicate using instant messages, message boards, emails, Voice over Internet Protocol (VoIP) as well as text chat.

2.3 The lack of security of Virtual Worlds

Because of their becoming part of our society, determining the identity of avatars is indispensable. Determining the identity of these artificial entities is as important as authenticating human beings. Mostly, an avatar would bear resemblance to its real life owner. There is a high demand for an affordable, fast, reliable means to authenticate avatars (Gavrilova & Yampolskiy, 2010). Toward the establishment of this goal, Yampolskiy and Gavrilova presented the concept of Artimetric. *Artimetrics* is the study of the identification, classification and authentication of virtual entities robots and software agents (Gavrilova & Yampolskiy, 2010).

Terrorist activities as well as cybercrimes are increasing in virtual worlds. For instance, it has been reported that terrorists recruit within virtual communities such as Second Life (Cole). Authorities such as U.S. government's Intelligence Advanced Research Projects Activity (IARPA) believe that they may use virtual worlds for illegal activities. They issue the warning that avatars could be used to recruit new members online, transfer untraceable funds and engage in training exercises useful for real-world terrorist operations (Cole). Several examples of terrorist activities have been reported within Second Life like flying a helicopter into Nissan Building or the bombing of ABC's headquarters. Another example is where armed militants forced their way into an American Apparel store and shot several customers and then planted a bomb outside a store (O'Brien).

Regrettably, these wrong doers cannot be prosecuted for their criminal behavior because these crimes were committed in a virtual world where laws do not exist. Anonymity as well as global access in an online virtual world where there are ease of access banking services that allow for transactions away from the normal routs has made virtual worlds a convenient environment for terrorists (O'Harrow, 2008).

Expressing concern over the consequences of leaving virtual worlds in such as a state, researchers in IARPA note that “The virtual world is the next great frontier and is still a very much a Wild West environment ("Intelligence Advanced Research Projects Activity,"). It provides many opportunities to exchange messages in the clear without drawing unnecessary attention. Additionally, there are many private channels that can be employed to exchange secret messages”. Virtual world has all the activities that the real world has and therefore, possible scenarios of these activities should be thought about (O'Harrow, 2008).

Virtual world environments pose a challenge as communication as well as commercial service between avatars is not recorded. Due to the set-up of the system, companies cannot monitor the creation and use of virtual buildings as well as training centers. Although some of them have been protected by what is described as unbreakable passwords, there have been reports of fraud and other virtual crimes. The situation is getting gloomier as companies in other countries are starting to establish their own virtual worlds. This shows urgency in addressing the issue of the security of virtual worlds. For instance, the founders of the Chinese virtual world HiPiHi ("HiPiHi,") which houses prestigious companies such as IBM and Intel aim to create ways to enable avatars to move freely from their virtual world and other virtual environments such as Second Life or Entropia. This in turn would make it difficult to identify avatar or real users behind avatars. The underground activities associated with real world criminals and terrorists will increase in these environments due to accessibility and secrecy they offer.

CHAPTER 3

OVERVIEW OF HUMAN AND AVATAR FACE RECOGNITION

3.1 Introduction

Face recognition is the process of identifying or verifying persons based on their digital images or videos automatically using computers. One way to satisfy this target is by comparing the facial features for persons. There are many face recognition techniques but generally they are grouped into two main groups: structure-based techniques and appearance based techniques. The methodology of structure-based techniques is based on extracting a group of geometric face features such as nose, eyes and mouth corners. The position of these facial features plays an important role since it forms a feature vector that should be fed to a classifier to identify a specific person. On the other hand, the appearance-based techniques are forming the most recently used face recognition techniques because they are more practical and easy to implement. Their methodology is based on using the appearance of face image as input to the decision making system. These techniques can be divided into three categories: holistic approaches, local feature-based (component-based) approaches and hybrid approaches. Holistic approaches use the whole image region as input to the face recognition system and then their performance affected by changing in pose, illumination and background. In local feature-based approaches, the whole facial image has to be divided into small regions or portions and local features such as eyes, nose and mouth have to be extracted first and their locations and fed to classifiers. Hybrid approaches use both local features and whole face area in recognizing faces, as what LBP does, it divides the whole image into local regions, build the local histogram for each region and then concatenate the local histograms into the whole image histogram. However, what about recognizing avatar faces.

Avatar face recognition can be considered as an extension to human face recognition. To date, very little work in recognizing avatar faces has been reported. Fortunately, the need for secure and affordable virtual worlds is attracting the attention of many researchers who seek to find fast, automatic and reliable ways to identify virtual worlds' avatars. The problem of Avatar Face Recognition (AFR) is not concerned only with finding high quality techniques to recognize a specific avatar among many other avatars but it also related to how to generate avatar face datasets to test and evaluate the developed techniques. In addition, if I generate the required avatar face datasets how I can detect avatar face in each image? Most of the work that was done so far focuses on recognizing the identity of avatars using techniques that are very similar to those applied on human face images datasets.

3.2 Survey of Face Recognition Methods

The following table (table 1) includes a short survey of many appearance-based face recognition techniques including the authors, the tested dataset, classification of the used method, the descriptor used in the experiments and the obtained accuracy rate for each technique.

Table 1: Brief Survey of Face Recognition Methods

	Authors	Dataset	Method Type	Descriptor	Accuracy Rate
1	M. Turk and A. Pentland (Turk & Pentland, 1991)	Over 2500 face images under a wide range of imaging conditions	Holistic	Eigenfaces and Euclidian distance	96% over lighting variations
2	K. Etemad and R. Chellappa (Etemad & Chellappa, 1997)	ORL and FERET	Holistic	LDA and weight-mean absolute square distance	99.2% for face recognition, 95 % for gender classification
3	Yang et al. (Jian, Zhang, Frangi, & Jing-Yu, 2004)	ORL, AR, and Yale	Holistic	2DPCA	96 % (ORL) 89.8% (AR) 84.24 % (Yale)
4	M.S. Barlett et al. (Bartlett, Movellan, & Sejnowski, 2002)	FERET	Holistic	ICA	99.8%
5	M. Yang (M. H. Yang, 2002)	ORL and Yale	Holistic	Kernel PCA	73.99 % (Yale) 97.75 % (ORL)
6	Pentland et al. (Pentland, Moghaddam, & Starner, 1994)	Database of 3000 individual	Hybrid	PCA	98%

7	P.S. Penev and J.J. Atick (Penev & Atick, 1996)	FERET and U.S. Air Force Mini survey database	Local	LFA	82.23% (FERET) 88.34% (Mini survey database)
8	Ahonen et al. (Ahonen, Hadid, & Pietikainen, 2006)	FERET	Hybrid	LBP and KNN	97%
9	Maturana et al. (Maturana, Mery, & Soto, 2009)	ORL, Yale, Georgia Tech. and Ext. Yale	Hybrid	ELBP and NBNN	99.35 % (ORL) 98.18%(Yale) 92.67% (Geor. Tech.) 97.15% (Ext.Yale)
10	Liao et al. (Shu Liao, Fan, Chung, & Yeung2, 2006)	JAFFE	Hybrid	ALBP	94.59%
11	Chan et al. (C. Chan, J. Kittler, & K. Messer, 2007)	FERET, XM2VTS and FRGC2.0	Hybrid	MLBP	97.9% (FERET) 93.2%(XM2VTS) 96.7% (FRGC2.0)
12	Xiaoyang Tan and Bill Triggs (Xiaoyang & Triggs, 2010)	FRGC-104, Ext.Yale, CMU PIE	Hybrid	LTP	86.3% (FRGC) 100% (Ext.Yale) 100% (CMU PIE)
13	Soon lee and Seiichi Ozawa (Lee & Ozawa, 2003)	Japanese face image database	Hybrid	RAN-LTM	99%
14	L. Wiskott et al. (Wiskott, Fellous, Kruger, & Malsburg, 1997)	FERET Bochum	Local	EBGM	84% (FERET) 97% (Bochum)
15	Yufeng Zheng and Adel Elmaghraby (Zheng & Elmaghraby, 2011)	ASUMS	Hybrid	PCA, LDA, EBGM and FPB	100%
16	Rara et al. (Rara, Ali, Elhabian, Starr, & Farag, 2010)	Database of 61 subjects	Hybrid	MAP-MRFAAM	More than 98%
17	Chen et al. (Chen, Liao, Ko, Lin, & Yu, 2000)	Database of 128 subjects	Holistic	LDA	97.34%
18	K.K. Paliwal and A. Sharma (Paliwal & Sharma, 2011)	ORL	Holistic	ALDA	90.00%
19	J. Yang and J.Y. Yang (J. Yang & Yang, 2003)	ORL	Holistic	PCA and LDA	
20	H. Kong, X. Li, J. Wang, E. Teoh and C. Kambhamettu (Kong, Li, Wang, Teoh, & Kambhamettu, 2005)	ORL + Yale + YaleB + CMU PIE + UMIST + CMU AMP + XM2VTS	Holistic	ULDA and MLDA	69.15%
21	W. Jun, J. Kittler, Y. Yu, K. Messer and W. Shitong (Xiao-	ORL and XM2VTS	Holistic	D-LDA	88.5% (ORL) 87.9% (XM2VTS)

	Jun, Kittler, Jing-Yu, Messer, & Shitong, 2004)				
22	B. Zhang, W. Gao, S. Shan and Y. Peng (B. Zhang, Gao, Shan, & Peng, 2004)	ORL and FERET	Holistic	GaborfaceSVM	99.4% (ORL) 85.2% (FERET)
23	L. Zhang, S. Li, Z. Qu and X. Huang (L. Zhang, Li, Qu, & Huang, 2004)	FERET	Holistic	Gabor features and AdaBoost	96.5%
24	H.K. Ekenel and R. Stiefelhagen (Ekenel & Stiefelhagen, 2006)	AR and CMU PIE	Local	DCT and PCA	93.8 % (AR) 91.8% (CMU PIE)
25	A.V. Nefian and M.H. Hayes (Nefian & III, 1998)	ORL	Local	HMM	84%
26	Y. Su, S. Shan, X. Chen and W. Gao (Su, Shan, Chen, & Gao, 2006)	FERET	Local	PGFC	99%
27	G. Zhao and Pietikäinen (Zhao & Pietikainen, 2007)	Dyn Tex and Cohn-Kanade	Hybrid	VLBP and LBP-TOP	95.71 % (Dyn Tex, VLBP) 97.14 % (Dyn Tex, LBP-TOP) 91.18 % (Cohn, VLBP) 94.38 % (Cohn, LBP-TOP)

3.3 Generation of Avatar Face Datasets

Yampolskiy et al., in (Oursler, Price, & Yampolskiy, 2009) applied two different approaches (manual and automated) to generate avatar datasets. Before starting to apply any of the two approaches to generate the required datasets they had to decide from which virtual world they will generate their datasets based on the following criteria:

- Avatar facial features with unstable attributes.
- The ability to see avatar from different angles.
- Generating new avatars with contrasting facial features.
- The simplicity of using the system.

Although the authors considered many virtual worlds such as Entropia Universe and Second Life, they decided to build their dataset based on Second Life virtual world since it is the best fit to the above criteria. Second Life has many advantages in the creation of avatar dataset. It has many physical facial attributes such as right-to-left symmetry and length. Second Life's camera can be easily controlled by changing the camera pan, tilt and zoom, which allow the user to collect several images for the same avatar with different angles. In addition, Second Life allows the user to use scripting language to manipulate the environmental elements.

3.3.1 Avatar Creation Approaches

Manual approach (Oursler et al., 2009): the authors used the Gadwin PrintScreen application to quickly capture and save each avatar face image in the desired directory after it is randomly generated. To capture the profile of avatar, there are 8 steps that should be followed to generate one image set for each avatar. These steps can be limited between one of two categories: either it is for adjusting the appearance menu to be sure that the focus would go to the avatar's face or adjusting the angle of the camera to capture an image for the same avatar each time with a different angle. This approach has two main problems: possible human error and it is a time consuming task. The dataset generated by applying this approach consists of 7 hundreds avatar images collected on 100 subjects (avatars) each has 7 different frontal angle images for the same avatar (see Fig. 3).

Automated approach (Oursler et al., 2009): the authors used programming language, AutoIt, in addition to a scripting language native to Second Life, Linden Scripting Language (LSL), to generate avatar dataset randomly and automatically. 10-step process is followed to generate avatar dataset automatically. This process mainly based on AutoIt and LSL but the user also plays a small role in this process. First, AutoIt was used to build a simulation for the key presses and mouse movements in the windows environment. Second, using the simulated keyboard commands allow the Second Life camera to focus on avatar face. Third, the user has to use the movement control to center avatar's face with horizon. This is required only for the first run and forms the last interaction with the user. Fourth, AutoIt activates LSL and then LSL has to lock Second Life camera's position and rotation. Fifth, AutoIt takes a screen shot for the avatar profile. Finally, AutoIt, rotates the camera to different angles, uses the avatar editing tool to decide the body height, length, leg height, skin, hair and eyes and randomizing each of them and then take screen shot again with each rotation. The dataset generated by this approach consists of 10 different angle avatar images for the same avatar. These images are one upper body image and nine avatar facial images (see Fig. 3).

(a) (b)

Figure 3: Second Life avatar images ("Second Life,"): a) Manually collected b) Automatically collected.

3.4 Avatar Face Detection

Object-class detection is a computer application for finding the locations and sizes of all objects in an image that belongs to a specific class. Face detection can be considered as a special case of object-class detection where the object is the face, the class is the human, and the target is to detect human faces within the image. A complete authentication biometrics system for human consists of two main stages: face detection and face recognition. Since human and avatars are very similar in face components and structure, I can say that a complete biometric authentication system for avatars also consists of two main stages: avatar face detection and avatar face recognition. Available biometric systems are not established to deal with the visual representation and the behavioral nature of the non-biological entities or avatars. There are many challenges emerged in detecting avatar faces. These challenges include, illumination, skin color, pose head rotation, etc. In (Yampolskiy, Klare, et al., 2012) Yampolskiy et al., applied some preprocessing techniques that may be useful to overcome part of these challenges by performing two different types of normalization: geometric and color normalization. To the best of our knowledge, previously there was no available technique specially for detecting avatar faces but there was only one trial to apply human face detection techniques on avatars.

3.4.1 Avatar face Detection using Extended Haar-like Features

In (Darryl D'Souza & Yampolskiy, 2012) D'Souza and Yampolskiy applied an extended version of the Viola & Jones rapid object detection framework, rotated Haar-like features, to detect avatar faces. Thus, rotated Haar-like features are called extended Haar-like features. The object detection framework of the original Viola & Jones contains an efficient set of 45 degree rotated features (see Fig. 4), which expands the learning framework to include an additional domain-knowledge.

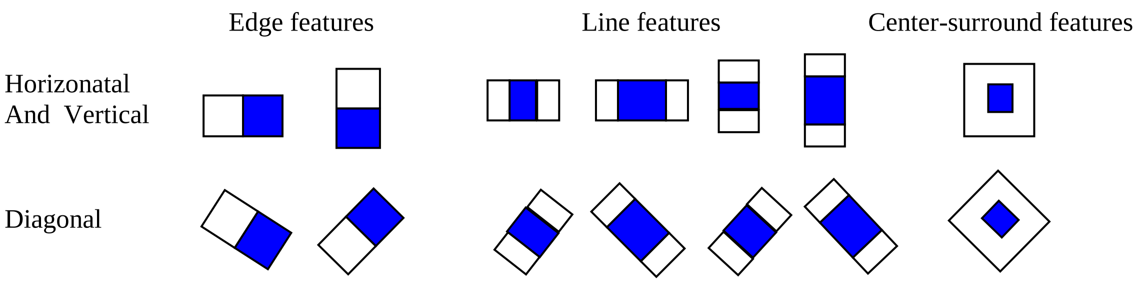


Figure 4: Simple Haar-like features shaded for positive weights and unshaded for negative weights (Darryl D'Souza & Yampolskiy, 2012).

Fig. 4 has 14 different prototypes divided into 3 groups: edge features (4 features), line features (8 features) and center-surrounded features (2 features). The authors used OpenCV to generate the Haar cascade. One of the main characteristics of Intel's OpenCV is that it provides programs to train classifiers for face detection. This feature is called Haartraining. The process of Haartraining has four main steps:

- **Preparing dat**a: this step is related to collecting positive and negative datasets of data. Positive dataset is the dataset that contains the objects of interest. In our case, it will contain the faces that I want to detect. On the other hand, the negative dataset is the dataset that contains other objects that are out of interest such as non-face images and background images.
- **Creating Sampled/Tested data:** the collected dataset was divided into 4 sets:

  Positive images containing the object of interest for training.

  Negative images containing other objects for training.

  Positive images for testing.

  Negative images for testing.

- **Training:** OpenCV applies Adaboost algorithm in Haartraining. As a result for this implementation, the xml cascade file will be generated which in turn will be used in Haardetection tool for object detection.
- **Testing:** the generated haarcascade.xml file resulted from training step will be tested by OpenCV performance.exe tool.

3.4.2 Experimental Results

For the purpose of testing and evaluating the system, the authors (Darryl D'Souza & Yampolskiy, 2012) used two types of datasets: human datasets and avatar datasets. Human datasets consist of two datasets. The first is a dataset from Caltech (California Institute of Technology). It has 450 images for 28 subjects and 3 sketches with complex background and varying illumination. The second human dataset is the FERET dataset. It has 400 images for 52 subjects with 7-8 images for each subject with head rotation from 15 to 67.5 degree, plain background and slightly varying illumination.

Avatar dataset has 450 avatar images collected from two well-known virtual worlds, Second Life (SL) and Entropia Universe (ENT). This dataset is divided into three subsets: 150 female avatar images from ENT with complex backgrounds, 150 male avatar images from SL with plain background and 150 male avatar images from SL with complex backgrounds. After applying the system on these datasets, the obtained accuracy rates in average were 78.8% with error rate of 23.7% for human datasets and 74% with error rate of 26% for avatar datasets.

3.5 Biometric Principles and Avatar Recognition

Yampolskiy and Gavrilova (Gavrilova & Yampolskiy, 2010) showed the need to secure virtual worlds. They introduced some hints about the risks that facing the real world from virtual worlds' entities. Some examples of these risks are (Gavrilova & Yampolskiy, 2010): 1- the risk of terrorism, some terrorism organizations such as Al-Qaeda uses Second Life virtual world for recruiting and communicating with their members and also for training their new members in an environment which is very similar to the real one.

2- Cybercrime risk, related to theft the identity and this happened plenty of times in worlds populated by millions of avatars and it became more dangerous when we know that these virtual worlds operate multibillion economies. 3- Attacking the infrastructure, security experts had reported that computers and networks of the Pentagon and other governmental agencies had been attacked by hackers assisted by hacking software agents. The authors also introduced a survey about non-biological entities. They classified non-biological entities into three categories: Virtual Beings (avatars), intelligent Software Agents (bots), and Hardware Robots. The authors mainly focused in this survey on Virtual Beings. They introduced the definition of the word avatar from both the dictionary and on-line community point of view. The authors also in this study divided avatars based on preferences and the behavior of avatars' creators to: Odd/Shocking, Abstract, Billboard, Lifestyle, Matching, Clan, Animated, Animal, Cartoon, Celebrity, Evil, Real Face, Idiosyncratic, Positional, Power and Seductive avatars.

Figure 5: Facial images of a humanoid robot-model, robot celebrity and a 3D-virtual avatar (Gavrilova & Yampolskiy, 2010; Yampolskiy et al., 2011).

The authors mentioned that using a combination of traditional pattern recognition techniques and biometrics behavioral identifiers to analyze the appearance and the behavioral patterns of avatars could help in identifying these avatars. The authors also summarized these behaviors to be: Mischievous Pranks, Flooding of the server, Blocking, Sleeping, Eavesdropping, Prop Dropping, and Identity Disruption. The authors also stated that there are no available public avatar faces datasets which can be used to test and compare the experimental results achieved by developed systems and they also at the same time provided a way by which I can create avatar faces datasets that are standardized and consistent with real world datasets. The authors studies the two main ways of authentication in virtual worlds: virtual and behavioral and introduced a multi-resolution system to enhance the performance of authenticating non-biological (avatars) entities. The authors also suggested some potential applications and future directions for further research.

3.6 Applying SVM Classification for Avatar Facial Recognition

In another article, Ajina et al., (Ajina, Yampolskiy, & Amara, 2010; Ajina et al., 2011) proposed a biometric recognition system for non-biological entities (avatars) faces. This system is used to recognize avatars from MyWebFace.com. The main goal of this biometric system is to differentiate between different avatars which want to access some virtual world resources. This system has three main stages: collecting the avatar samples, extracting features for characterization and classification. In collecting dataset stage, the authors used an avatar creation web site (MyWebFace). This website is devoted especially to creating avatars.

The authors collected 1800 avatar facial images. These avatars are organized into 100 classes each of which has a series of 18 different images for the same avatar. All images in this dataset are in RGB format taken from frontal position with white homogenous background and under the same lighting conditions. The size for all images is the same 150 x 175 pixels and with the same resolution 90 PPI. This dataset is divided into two independent parts: the first part has 1200 facial image (12 from each class) for training and the second which contains 600 facial images (6 facial images with accessories from each class) are used to evaluate the performance of the system. In extracting features stage, the authors used wavelet transform to extract a set of global characteristics. There are many families of wavelet transforms, to decide which family of wavelets and within this family what is the corresponding level of decomposition that greatly describe the tested dataset, the authors carried out a set of tests on a number of wavelet families. The authors figured out that the best wavelet family for the dataset they have is the Symlet wavelet family and the corresponding level of decomposition is level 5 (See figure 6). In classification stage, the authors used a supervised learning technique called Support Vector Machines (SVM). The results obtained from applying this automatic system were very promising with recognition rate higher than 95% on average.

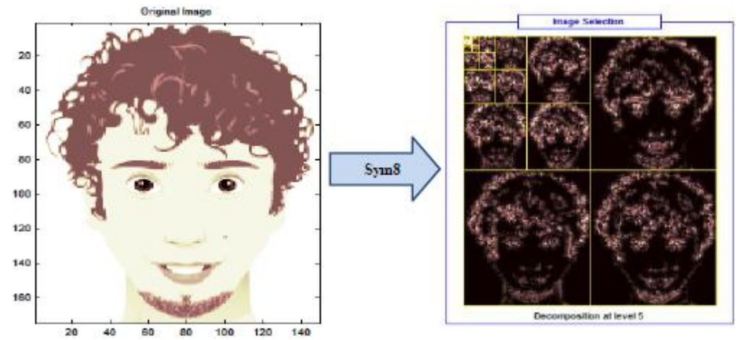


Figure 6: Wavelet decomposition using "Symlet 8" wavelet family with decomposition level 5 (Ajina, Yampolskiy, & Amara, 2011).

3.7 Applying Current Academic and Commercial Software for Human Face Recognition on Avatar Face Datasets

In another article, Yampolskiy et al., (Yampolskiy, Cho, et al., 2012; Yampolskiy et al., 2011) conducted experiments to evaluate the performance of executing current academic and commercial software for recognizing human faces on avatar face datasets. The authors developed the usage of two main techniques to recognize avatar faces. In the first, they combined an academic technique (VeriLook) with two well-known descriptors, Color Structure Descriptor (CSD) and Edge Histogram Descriptor (EHD) while in the second they concluded their experiments by testing and analyzing state-of-the-art commercial software from Google (Picasa).

3.7.1 Using Advanced Face Localization Algorithm

The authors of this article (Yampolskiy, Cho, et al., 2012; Yampolskiy et al., 2011) applied an algorithm (VeriLook) to implement Advanced Face Localization (AFL). VeriLook can perform processing and identification by two different modes, one-to-one (verification) and one-to-many (identification) modes, with a comparison speed of 100000 faces per second. This algorithm is designed to deal with images with variations in roll, pitch and yaw of the head. This variation has different degrees based on the face position. The head roll variation is within ±180 degree, both the head nod and yaw variations are within ±15 degree from the frontal position (see figure 7). The advanced face localization algorithm has the following steps:

1- For all raw images, detect all faces in a particular frame.

2- Build the facial templates (one for each face record) and load them into the computer's RAM (there is no need to store the original face images).

3- Extract the face template.

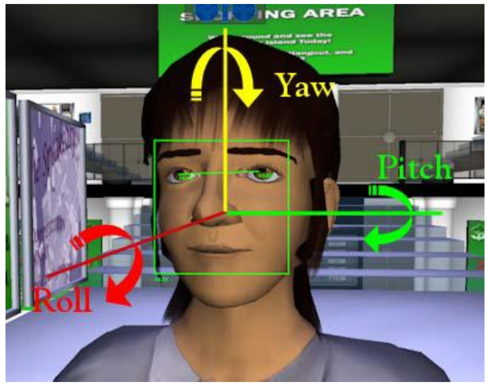


Figure 7: Illustration of different face positions (Yampolskiy, Cho, et al., 2012) (Garcia, Zikos, & Tziritas, 1998; Yampolskiy et al., 2011).

The authors tested their algorithm in a PC with Intel Core 2 processor with 2.66 GHz. In this algorithm the extraction time depends on the defined template size and does not depend on image size. To evaluate the performance of VeriLook algorithm on avatar faces the authors performed two experiments using face templates from Face Recognition Grand Challenge (FRGC) dataset (see figure 8). Experiment 1 is designed to evaluate the performance of the algorithm in recognizing frontal facial images under controlled

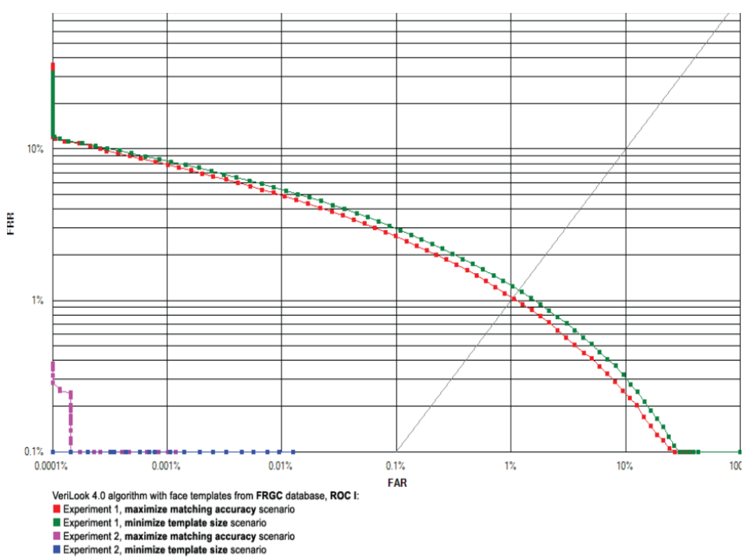


Figure 8: The result of applying VeriLook algorithm on face templates from FRGC dataset (Yampolskiy, Cho, et al., 2012) (Mazloom & Ayat, 2008; Yampolskiy et al., 2011).

illumination and by using a single high-resolution still image from each class. Experiment 2 is used to examine the effect of use multiple still images on performance. In this experiment both query and target datasets has four controlled images from each subject (see figure 8).

3.7.2 Using Color Structure Descriptor

Color Structure Descriptor (CSD) is image-to-image matching algorithm based on color histogram of an image (Yampolskiy, Cho, et al., 2012; Yampolskiy et al., 2011). The authors used pair-wise distance between query image and a set of similar images to overcome a false positive result obtained from their system. CSD is a generalization of the color histogram defined in Hue Min Max difference (HMMD) color space quantizing images using color quantization up to 256 colors. During their experiments, the authors applied a well-known color quantization technique, median-cut algorithm, to quantize avatar facial images (see figure 9). The authors applied CSD only for still images to obtain local color structure of an image by

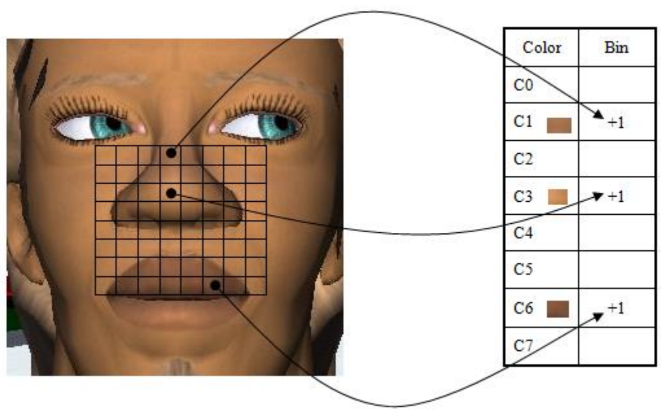


Figure 9: Color structure descriptor (Yampolskiy, Cho, Rosenthal, & Gavrilova, 2012; Yampolskiy, Gyuchoon, Rosenthal, & Gavrilova, 2011).

using structuring element. CSD used a 8 x 8 pixel structure element overlaid on all locations of the image to retrieve colors C related to all pixels contained in the structure element. Therefore, the CSD bins assigned to each color Cm will be increased.

3.7.3 Using Edge Histogram Descriptor

Histogram is one of the most commonly used representations of the global features of an image. Histograms are invariant to both image translation and rotation, and the normalization of histograms leads to scale invariance. Based on the previous properties, they are very useful descriptors in indexing and retrieving of images from their datasets. Edges of an image hold useful information about that image and hence they can be considered as an important feature that can be used to represent the content of these images. One of the most successful ways to represent such important feature is by using histograms. The underlying directionality and brightness in the image can be represented by a descriptor called edge histogram descriptor (EHD). The authors of this article used EHD to find the same avatars' faces based on their characteristics, such as hairstyle, or the shape of their eyebrows. Each facial image has to be divided into 4 x 4 sub-images (see figure 10). All the 16 sub-images have the same dimension regardless of the size of the original image.

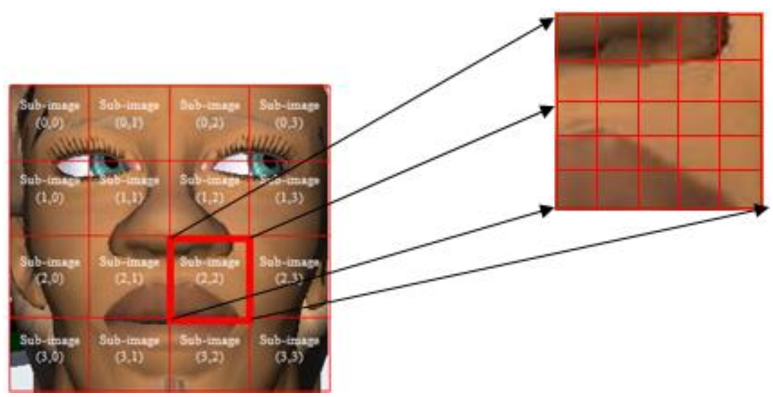


Figure 10: Sub-image grid of input image and image block.

EHD represents the distribution of 5 different types of edges (vertical, horizontal, 45-degree diagonal, 135-dgree diagonal and non-directional) in the area of sub-image. One way to specify the characterizations of a sub-image is by generating the edge distribution histogram for that sub-image. Hence, the EHD histogram of sub-image represents the frequency of occurrences of the five different types of edges in that sub-image. Each local histogram has 5 bins, each one to represent one of the five different types of edges.

Since, each images has to be divided into 16 sub-images (see figure 10), a total of 80 (5 x 16) histogram bins is needed. Each one of the 80-histogram bins has its own location and edge type. For example, the bin for the vertical edge in the sub-image located at (0,3) in Fig. 11 carries the information of the relative population of the vertical edges in the top-right local region of the image.

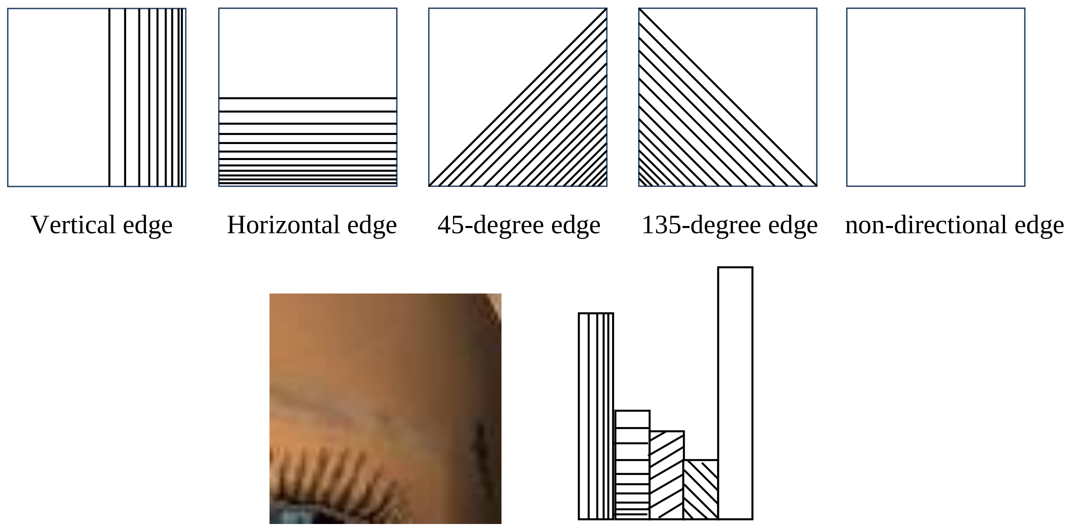


Figure 11: Five types of edge bins for sub-image (top-right) and its histogram (Yampolskiy, Cho, et al., 2012) (Garcia et al., 1998; Yampolskiy et al., 2011).

3.7.4 Experimental Results

To evaluate the performance of the VeriLook algorithm (Yampolskiy, Cho, et al., 2012; Yampolskiy et al., 2011), the authors used avatar dataset of 700 avatar images. These avatar images are used to represent 70 different avatars each in a subject of ten different headshots. The CSD and EHD were combined with VeriLook algorithm to improve the recognition rate. Each time the system returned a head shot as the top match. If this head shot is from the same subject as the query image then the query image is recognized correctly otherwise the query image is recognized incorrectly. Based on this system, 559 out of 700 avatar images were recognized correctly with a percentage of 79.9% and within 1259.33 seconds processing time.

To establish baseline capability in recognizing avatar faces by Picasa, the authors used a dataset of 440 avatar images was collected from Second Life virtual world. There images were organized in 22 subjects (avatars) with 20 pictures of each avatar. This dataset was divided into training or control group and testing group. The control group contains 60% of the dataset (12 images from each subject) while the testing group contains 40% of the dataset (8 images from each subject). Although, Picasa was not able to recognize all avatar facial images in the control group, it can recognize 83.27% of the content of the control group. Out of the recognized images from the control group Picasa can correctly recognize 53.57% of the total images in the dataset.

3.8 Recognizing Avatar Faces using Different Scenarios

In another article, Yampolskiy et al., (Yampolskiy, Klare, et al., 2012) suggested 4 scenarios requiring face recognition algorithms in investigating criminal and terrorist activity in virtual worlds. These scenarios are:

a. **Matching a Human face to an Avatar face**

Generally many users have the tendency to use their real face as their online avatar which helps to represent them well.

b. **Matching one avatar face with another**

This capability helps to continuously track an avatar through cyberspace at different places at different times.

c. **Matching an Avatar's face from one virtual world to the same avatar in a different virtual world**

A recent development within virtual communities is to interconnect different virtual worlds. This will help in uniquely identifying and tracking records of the avatars.

d. **Matching an Avatar sketch to the Avatar face**

Just like the traditional methods of matching the forensic sketch of human faces provided by the description of the victim or witness to their real faces, it is equally important to map this scheme within virtual worlds to match the virtual criminal with its avatar identity.

In this article, the authors also proposed an avatar face recognition framework (Yampolskiy, Klare, et al., 2012). This framework followed the second scenario (avatar to avatar matching) and has the same procedure as standard face recognition systems. Therefore, it has three main steps:

1- Face detection and normalization

2- Face representation

3- Matching

In face detection and normalization step the authors applied Viola and Jones method with the default frontal face Haar cascade packaged with OpenCV to detect avatar faces. Once the avatar face is detected, the authors applied two types of normalizations: geometric normalization to reduce the effects of scale, rotation and translations by estimating the location of the eyes. Appearance normalization is applied by histogram equalization to reduce the effect of changes in illumination. The authors merged the three-color channels for avatar faces into one dimension for normalization in order to reduce the computationally demanding of histogram equalization across three dimensions or color channels.

For face representation, the authors applied two types of representations: structural representation and appearance representation. Structural representation is designed to transfer the face shape and morphology into a digital form. The authors tried many descriptors and finally they were convinced by the effectiveness of the Local Binary Pattern (LBP) descriptor. They applied special case of LBP called Uniform LBP. In appearance representation the color of avatar faces provide discriminative information that can help in determining the identity of avatars. To achieve this goal the authors applied a descriptor called Spatial Appearance Descriptor (SAD).

In matching step, the authors applied a similarity measure, Chi-square, to compute the distance between any two-avatar faces. Chi-square is computed twice between any two images one using structural representation and the second for the appearance representation. These distances are normalized and the final similarity between two faces images is computed based on a distance that is a combination of a distance obtained from the structural representation and the distance obtained from the appearance representation.

To evaluate their framework, the authors used two dataset. The first is the FERET dataset. The authors used a picture to avatar conversion software, AvMaker, to convert each image in FERET dataset to 3D avatar image. By using AvMaker the authors were able to produce 2020 avatar images belongs to 725 subjects. Using FaceVACS face recognition Software Developer Kit on FERET-to-Avatar dataset achieved 99.58% accuracy rates. The authors designed and implemented a technique to collect an avatar dataset automatically from Second Life virtual world. The accuracy rate obtained after applying SAD was about 98%. These results demonstrated the effectiveness of this framework.

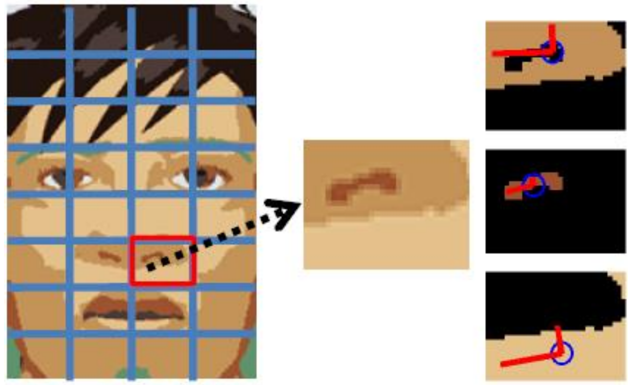


Figure 12: Illustration of the SAD descriptor (Yampolskiy, Klare, & Jain, 2012).

CHAPTER 4

AVATAR FACE RECOGNITION AND HIERARCHICAL MULTI-SCALE LOCAL BINARY PATTERNS (HMLBP)

Recognizing avatars in virtual worlds is a very important issue for law enforcement agencies, terrorism and security experts. In this chapter, a novel face recognition technique based on wavelet transform and Hierarchical Multi-scale Local Binary Pattern (HMLBP) is presented and shown to increase the accuracy of recognition of avatar faces. The proposed technique consists of three stages: preprocessing, feature extraction and recognition. In the preprocessing and feature extraction stages, the wavelet decomposition is used to enhance the common features of the same class of images and the HMLBP is used to extract representative features from each avatar face image without a need for any training. In the recognition stage, the Chi-Square distance is used to achieve a robust decision and to indicate the correct class to which the input image belongs. Experiments conducted on two manually cropped avatar image datasets from two virtual worlds (Second Life and Entropia Universe) show that the proposed technique performs better than traditional (single scale) Local Binary Pattern (LBP), Wavelet Local Binary Pattern (WLBP), Multi-scale Local Binary Pattern (MSLBP) and HMLBP in terms of accuracy.

In the rest of this chapter, I provide an introduction to wavelet transformation, its role and benefits in the field of image processing, describe the LBP operator and its histogram, introduce the meaning of MSLBP and its different versions and introduce the wavelet hierarchical multi-scale LBP (the proposed algorithm). I also present experiments implemented on avatar face datasets, show comparisons between the proposed algorithm and various other methods and finally provide useful conclusions.

4.1 Wavelet Decomposition of an Image

Discrete Wavelet Transform (DWT) is a widely used tool for image compression and texture classification because of its effective ability for multi-resolution decomposition analysis (Mazloom & Ayat, 2008; Mohamed et al., 2011; Mohamed et al., 2012; Mohamed & Yampolskiy, 2012b, 2012d). It was also used to extract the essential features for avatar face recognition. Many articles have discussed its mathematical background and advantages. In the proposed system, DWT is used to decompose images because (Garcia et al., 1998; Mazloom & Ayat, 2008):

- DWT reduces the computational complexity of the system by producing lower resolution images (sub-images) instead of operating on the original images with much higher resolution. For example, applying WT to reduce the resolution of an image from size 128 x 128 to size 32 x 32 will reduce the computational load by a factor of 16.
- DWT decomposes images into sub-images corresponding to different frequency ranges and this can lead to reduction in the computational overhead of the system.
- Using DWT allows obtaining the local information in different domains (space and frequency) while Fourier decomposition concerns only global information in the frequency domain. Thus it supports both spatial and frequency characteristics of an image at the same time.

In case of images I have to apply WT in two directions (row or horizontal direction and column or vertical direction) using four different filters (Mohamed & Yampolskiy, 2012c):

$$\begin{aligned}
\varphi(n_1, n_2) &= \varphi(n_1)\,\varphi(n_2) \\
\psi^H(n_1, n_2) &= \psi(n_1)\,\varphi(n_2) \\
\psi^V(n_1, n_2) &= \varphi(n_1)\,\psi(n_2) \\
\psi^D(n_1, n_2) &= \psi(n_1)\,\psi(n_2)
\end{aligned} \tag{1}$$

where n_1 is the horizontal direction and n_2 is the vertical direction, φ is the scaling function which is essentially a low pass filter, ψ is the wavelet function which is essentially a high pass filter, the product $\varphi(n_1)\ \psi(n_2)$ means applying the low pass filter in the horizontal direction and applying the high pass filter in the vertical direction, by the same way I can understand the meanings of all the four filters. In the second filter there is a super script H since there is a high pass filter applied on the horizontal direction, by the same way I can understand the superscripts V and D (Mohamed & Yampolskiy, 2012c). As a result of applying the four filters an image will be decomposed into four sub-bands LL (low pass filter on the horizontal direction and low filter on the vertical one), HL (high pass filter on the horizontal direction and low pass filter on the vertical one), LH and HH (see Fig. 13). The band LL represents an approximation to the original image while bands LH and HL represent respectively the changes of the image along the vertical and horizontal directions. The band HH records the high frequency component of the image (Mohamed et al., 2012; Mohamed & Yampolskiy, 2011, 2012c, 2012d).

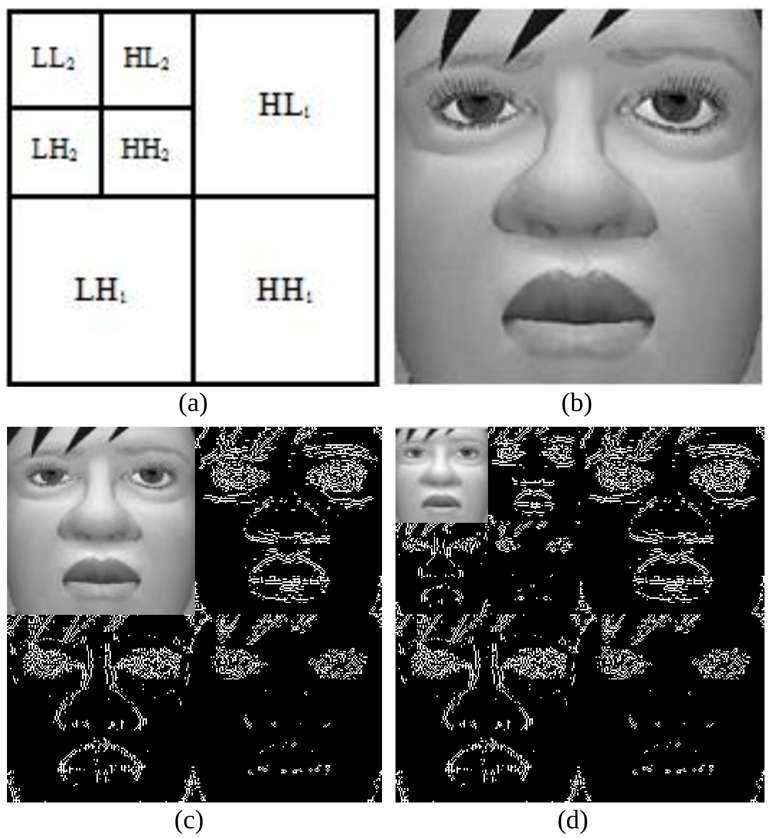


Figure 13: (a) Wavelet coefficient structure (Mohamed, D'Souza, Baili, & Yampolskiy, 2011); (Garcia et al., 1998; Mazloom & Ayat, 2008; Mohamed et al., 2011; Mohamed, Gavrilova, & Yampolskiy, 2012; Mohamed & Yampolskiy, 2012d) (b) A sample image of one of the avatar face images in the dataset (c) One level wavelet decomposition for the avatar face image in b (d) Two levels wavelet decomposition for the avatar face image in b.

To obtain a higher level of decomposition any one of the previous four sub-bands can be analyzed but since images generally are very rich in the low frequency contents, so I have to decompose the LL sub-band of the previous decomposition level using four different filters as I did before. For example, to obtain the second level of decomposition I have to decompose the LL_1 sub-band. The decomposition has to be carried out for the LL_2 to obtain the third level decomposition and so on… . Therefore, I can say that wavelet decomposition of an image provides an approximation image, which is used to obtain the next decomposition level, and three detailed images in horizontal, vertical and diagonal directions (Mohamed & Yampolskiy, 2011, 2012b, 2012d). The two-dimensional wavelet transform, which is required to deal with images, can be obtained by applying a one-dimensional wavelet transform to the rows and columns of the two-dimensional data. Decomposing an image with two scales will give us seven sub-bands: LL_2, HL_2, LH_2, HH_2, HL_1, LH_1 and HH_1 (see Fig. 13).

4.2 Local Binary Patterns

The local binary pattern (LBP) operator, introduced by Ojala et al., (Ahonen et al., 2006; Jun, Yumao, Xiukun, Tsauyoung, & Jianying, 2010; Ojala & Pietikäinen, 1996; Ojala, Pietikainen, & Maenpaa, 2002) is a powerful local descriptor for describing image texture and has been used in many applications such as industrial visual inspection, image retrieval, automatic face recognition and detection. The LBP operator labels the pixels of an image by thresholding the value of the central pixel against its surrounding 8 pixels (for a given size of 3x3 neighborhood of each pixel) and considering the result as a binary value (Jun et al., 2010; Mohamed et al., 2011). The binary value will be converted to the decimal value to get the LBP value. The output value of the LBP operator can be defined as follows (Jun et al., 2010; Wencheng, Faliang, Jianguo, & Zhenxue, 2010):

$$LBP(x_c, y_c) = \sum_{i=0}^{7} 2^i S(g_i - g_c) \qquad (2)$$

where g_c corresponds to the gray value of the central pixel, (x_c, y_c) are its coordinates, g_i ($i = 0,1,2,..,7$) are the gray values of its surrounding 8 pixels and $S(g_i - g_c)$ can be defined as follows:

$$S(g_i - g_c) = \begin{cases} 1, g_i \geq g_c \\ 0, otherwise \end{cases} \qquad (3)$$

Therefore, I can say that LBP is an ordered set of binary comparisons between the central pixel value and the values of its neighborhood pixels (Mohamed et al., 2011; Wencheng et al., 2010). Fig. 14 displays an illustration of the basic LBP operator and how to compute the LBP value.

The LBP operator can be extended to use pixels from neighborhoods of different sizes (Ahonen et al., 2006; Wencheng et al., 2010; Xiaoshan, Minghui, & Lianwen, 2010). Fig. 15 shows us some examples of different LBP operators where *R* is the radius of the neighborhood and *P* is the number of pixels in that neighborhood. The neighborhood can be either in a circular or square order. Using the circular order neighborhood allows any radius and number of the pixels in the neighborhood (NuTao, Lei, & Changping, 2008).

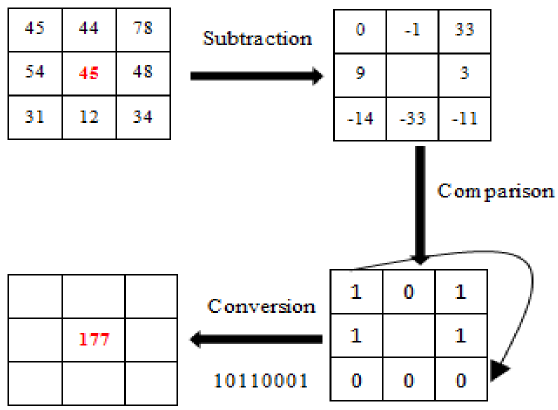
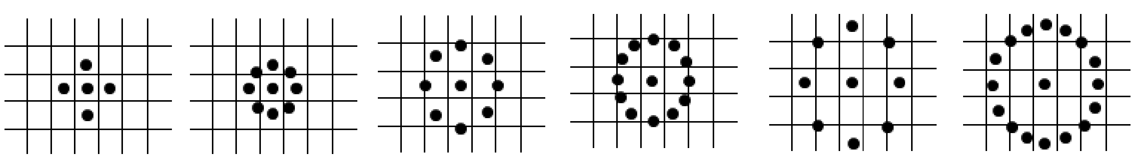


Figure 14: The basic LBP operator.



(*P*=4, *R*=1) (*P*= 8, *R*=1) (*P*=8, *R*=1.5) (*P*=12, *R*=1.5) (*P*=8, *R*=2) (*P*=16, *R*=2)

Figure 15: Different LBP operators.

One of the most important and successful extensions to the basic LBP operator is called uniform LBP (ULBP). An LBP is called uniform when it contains at most two different conversions from 0 to 1 or 1 to 0 when the binary string is viewed as a circular bit string (Ahonen et al., 2006; NuTao et al., 2008). For example, 11111111, 00011000 and 11110011 are uniform patterns. Ojala reported that with *P* = 8 and *R* = 1 neighborhood, uniform patterns account for around 90% of all patterns and with *P* =16 and *R* = 2 neighborhood, uniform patterns account for around 70% of all patterns (Ahonen et al., 2006). After labeling an image using the LBP operator, the histogram of the labeled image can be defined as follows (Wencheng et al., 2010):

$$H_i = \sum_{x,y} I(f(x,y) = i), i = 0,1,..,n-1$$

where '*n*' is the number of different labels produced by the LBP operator, $f(x, y)$ is the labeled image and $I(A)$ is a decision function with value 1 if the event A is true and 0 otherwise. LBP histogram has very useful information about the distribution of the local microstructures, such as spots and edges, over the whole image and so can be used to describe and represent the global characteristics of the image (Wencheng et al., 2010).

4.3 Multi-scale LBP

One of the main weaknesses of the original LBP (single scale LBP) is that it does not provide a complete image representation (S. Liao, Zhu, Lei, Zhang, & Li, 2007). Features obtained by using a local 3x3 neighborhood around a central pixel can only capture small scale structures (microstructures). Hence, the LBP operator is not robust enough against any local changes in the image texture. To overcome this limitation of the original LBP and to capture large scale structures that may have useful features of the faces, new representation of the image, Multi-scale LBP, was presented as a solution.

There are many versions for multi-scale analysis of an image. Mäenpää and Pietikäinen (Mäenpää & Pietikäinen, 2003) propose two novel ways to extend the LBP operator to be able to handle multiple scales. In the first one, the authors use exponentially growing circular neighborhoods with Gaussian low-pass filters to collect information from a large texture area. In this study, both the filters and the sampling positions are planned in a way that makes them able to handle the neighborhood as much as possible and in the meantime be able to reduce repeated information. Additionally, the authors suggest an alternative way to encode arbitrary large neighborhood that has cellular automata. The method was used successfully in compactly encoding even 12-scale LBP operators. Here, a feature vectoring, that is characterized by having marginal distributions of LBP codes and cellular automation rules, was employed as a texture descriptor. However, it is important to note that in these experiments no significant progress could be achieved when performance was compared to the basic multi-scale approach.

Another improvement was performed to the multi-scale LBP operator. It was extended to become a multi-scale block local binary pattern (MB-LBP) (S. Liao et al., 2007). The main point that MB-LBP offers is to enable comparing average pixels values found within small blocks in lieu of comparing pixel values. Here, the operators always use 8 neighbors producing labels from 0 till 255. For example, if the block size is 3x3 pixels, the parallel MB-LBP operator performs a comparison of the average gray value of the center block to the average gray values of the 8 neighboring blocks of the same size. The effective of the operator is 9 x 9 pixels. The MB-LBP was introduced to replace the fixed uniform pattern mapping and to be used with a mapping that is dynamically obtained from a training data. Here, the mapping works as follows: the N recurring MB-LBP patterns take labels 0,…, N-1. The rest of the patterns take a single label. Here, the number of labels and the length of the MB-LBP histogram are parameterized so that the user can set.

Generally, the direct way to analysis an image using multi-scale approach ease to obtain the input image computed at different scales and then concatenating the LBP histogram computed at each scale after resizing each image patch to the same size (see figure 16) (Turtinen & Pietikäinen, 2006). The main problem existing in this approach is the high dimensionality of the final histogram, which contains redundant information.

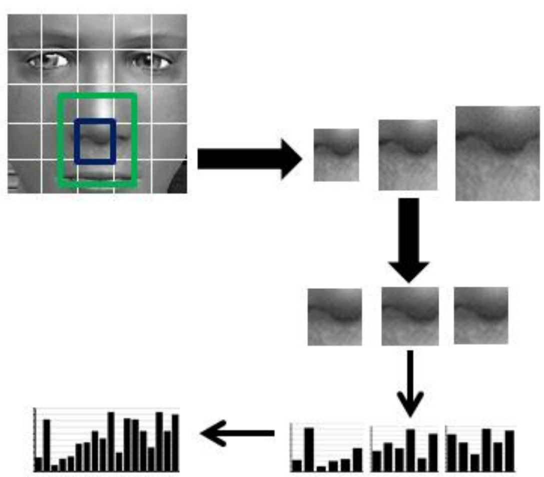


Figure 16: Multi-scale avatar image representation.

To overcome this problem Chan et al., (C. Chan et al., 2007) developed this approach by combining the multi-scale LBP representation with Linear Discriminant Analysis, LDA, and the Principal Component Analysis, PCA. In their approach, they first applied the uniform local binary pattern operators at *R* scales on a face image. Then, they crop the resulting LBP images to the same size and divide these images into non-overlapping sub-regions. The set of histograms computed at different scales for the same sub-region are concatenated into a single histogram. To reduce the dimensionality of the descriptor they applied PCA before LDA. Therefore, to derive discriminative facial features using LDA they applied PCA first to extract statistical independent information.

4.4 Wavelet Hierarchical Multi-scale LBP (WHMLBP)

The proposed algorithm has three steps: preprocessing, feature extraction and recognition or classification.

4.4.1 Preprocessing Face Image

To improve the efficiency of extracting the face features I have to apply a set of preprocessing operations. First, I manually cropped the input images to pure face images by removing the background that is not useful in recognition. Second, these pure face images have to be normalized and then decomposed using the first level of wavelet decomposition to obtain pure facial expression images (See Fig. 17). Detailed images resulted from applying wavelet decomposition contain changes which represent the difference of face images. So considering only the approximation images will enhance the common features of the same class of images and at the same time, the difference will be reduced. For this reason, our experiments were concerned only with the approximation images resulting from the first level of wavelet decomposition and which I used in testing to evaluate the performance of the proposed algorithm

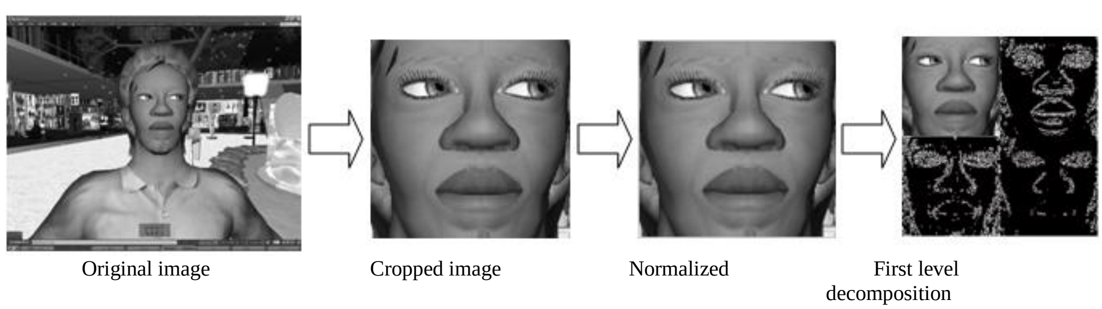


Figure 17: Face image preprocessing.

4.4.2 HMLBP Feature Extraction

The performance of the multi-scale or multi-resolution LBP operator is better than the performance of a single scale LBP operator for many reasons, such as:

a- Multi-scale operator can help to extract more image features under different settings (Mohamed et al., 2011; Zhenhua, Lei, Zhang, & Xuanqin, 2010). Calculating features based on a limited size neighborhood in single scale LBP may lead to inadequate capture of dominant features of an image.

b- As a result of single scale LBP operator "non-uniform" patterns are clustered into one non-uniform pattern. As the radius of the LBP increases, the cluster size of the "non-uniform" patterns increases as well, leading to a substantial loss of information (Mohamed et al., 2011; Zhenhua, Lei, Zhang, & Xuanqin, 2010).

Some work (Mohamed et al., 2011; Zhenhua, Lei, Zhang, & Xuanqin, 2010) was carried out towards extracting more useful features from the image by obtaining information from the "non-uniform" patterns. Such methods are based on a training step to learn the useful patterns and so the training samples have a great effect on the accuracy of recognition (Mohamed et al., 2011; Zhenhua, Lei, Zhang, & Xuanqin, 2010). In HMLBP algorithm the LBPs for the biggest radius are extracted first. The new LBPs of non-uniform patterns have to be extracted further using a smaller radius to extract uniform patterns. This process continues until the smallest radius is processed. This hierarchical scheme does not have a training step and thus it is insensitive to training samples (Mohamed et al., 2011; Zhenhua, Lei, Zhang, & Xuanqin, 2010). Fig. 18 shows an example of the hierarchical multi-scale LBP scheme.

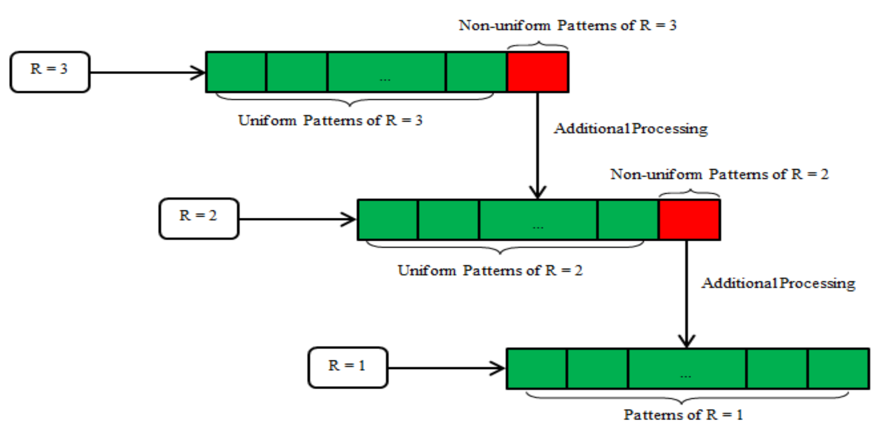


Figure 18: An example of hierarchical multi-scale LBP Scheme (Zhenhua, Lei, Zhang, & Xuanqin, 2010).

The LBP histogram for R=3 is first built. For those “non-uniform” patterns of the R=3 operator, a new histogram is built by the R=2 operator. Then, the “non-uniform” patterns of R=2 lead to the histogram building process for the R=1 operator. Finally, the three histograms are concatenated into one multi-scale histogram to form the feature histogram of an image (Mohamed et al., 2011; Zhenhua, Lei, Zhang, & Xuanqin, 2010).

4.4.3 Similarity Measure

The last stage of our proposed algorithm is to classify each facial image to its class by computing the dissimilarity between training samples and a test (input) sample. To do that I apply Chi-Square distance as follows (Ahonen et al., 2006):

$$D(X,Y) = \sum_{n=1}^{N} \frac{(X_n - Y_n)^2}{X_n + Y_n} \tag{5}$$

where X is the tested image (sample), Y is the training sample(s) or image(s) and N is the sum dimension.

4.5 Experimental Results and Analysis

To ensure the efficiency of the proposed method, two virtual world datasets are used to test the performance of the proposed method. This is the first time given algorithm is used on the gray scale images and consequently there is no baseline results available for direct comparison.

The first dataset, from Second Life virtual world [20], contains 581 (1280 x 1024 pixels) gray scale images of 83 avatars. The second dataset, from Entropia Universe virtual world [21], consists of a total of 490 (407 x 549 pixels) gray scale images representing 98 avatars. I tested these datasets with three well-known algorithms (LBP, WLBP and HMLBP) and compared their result with the results coming from the proposed method.

4.5.1 Experimental Setup

All images in the Second Life dataset are manually cropped to 260x260 pixels while images in Entropia dataset are manually cropped and resized to 180x180 pixels. The resulted 581 Second Life avatar face images dataset is organized into 83 classes each of which has 7 face images of the same avatar with different frontal angles (front, far left, mid left, far right, mid right, top and bottom). Therefore, I can say that the Second Life avatar face images dataset focuses on pose angle and facial expression.

The obtained Entropia avatar face dataset is organized into 98 classes each of which has 5 avatar face images. In one of them the avatar is wearing a mask while in the others the avatar has different facial expressions and eye angles. See Fig. 19 for an example of two classes of avatars (one from each dataset) before and after cropping.

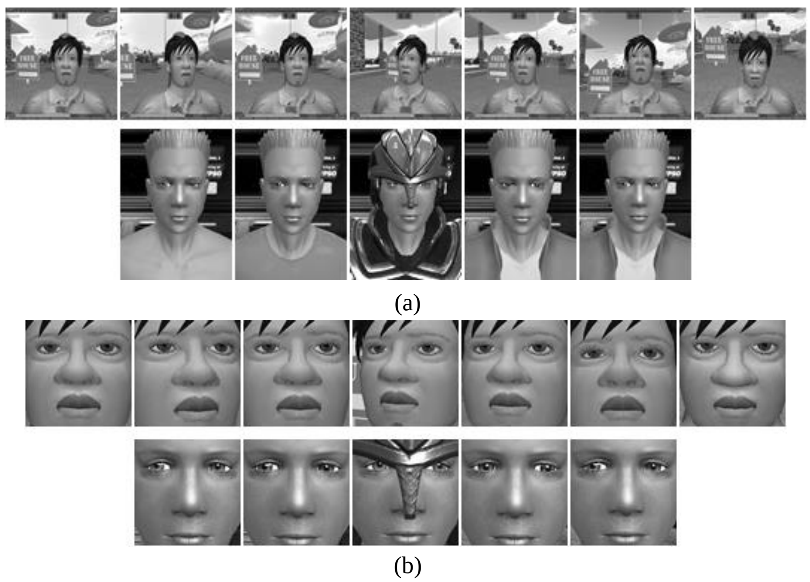


Figure 19: a) Two classes of unprocessed avatar images.
b) The same two classes after cropping the avatar faces.

The resolution of images used in the experiments is changed from 260 x 260 to 130 x 130 pixels (for Second Life dataset) and from 180x180 to 90 x 90 (for Entropia dataset) using the first level of wavelet decomposition. The avatar face images in both datasets are preprocessed and prepared for feature extraction step. HMLBP is used to extract the best descriptive features and then at the end the Chi-Square measure is applied to accomplish classification. The experiments are performed on the condition of a single training image. Each time one image is used as a trainer. The Chi-Square distance computes the similarity between this image and all other images in the dataset. These distances will then be ordered in an ascending order. The 6 images (for Second Life dataset) associated to the least 6 distances in the ascending order will be checked if they are from the same class of the trained image or not. The same will be done but with only 4 images for the Entropia dataset. Based on the number of corrected classified images I can compute the accuracy for each dataset using the following formula: classification accuracy (CA) or recognition rate (*RR*) equation:

$$RR = \frac{number\ of\ corrected\ classified\ images}{total\ number\ of\ samples\ in\ the\ dataset}\ x100\% \qquad (6)$$

4.5.2 Comparing WHMLBP with HMLBP and other Algorithms

In order to gain better understanding on whether using wavelet transform with HMLBP is advantageous or not I compared WHMLBP with HMLBP, WLBP and LBP with several experiments. First I got the performance of WHMLBP with different block size with $R = [3, 2, 1]$ and $P = [16, 16, 16]$ as I can see in Fig. 20.

It is shown from Fig. 20 that that changing the block size affects the result of the recognition rate. The recognition rate is increased as the block size is larger, and the performance is dropped as the block size is larger than 42 x 42 on the two datasets, that is because dense blocks obscure the image features.

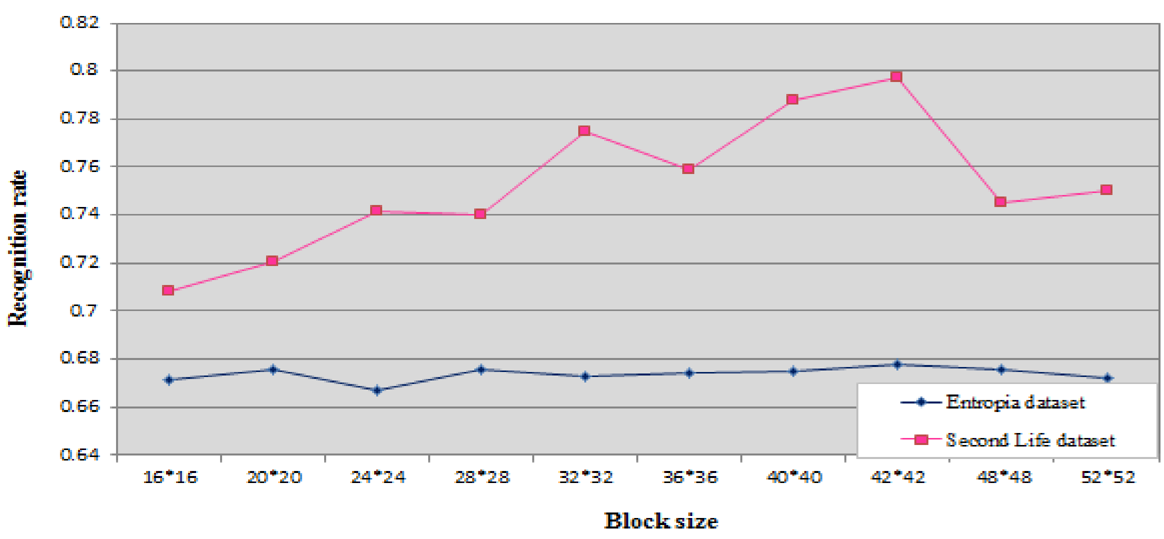
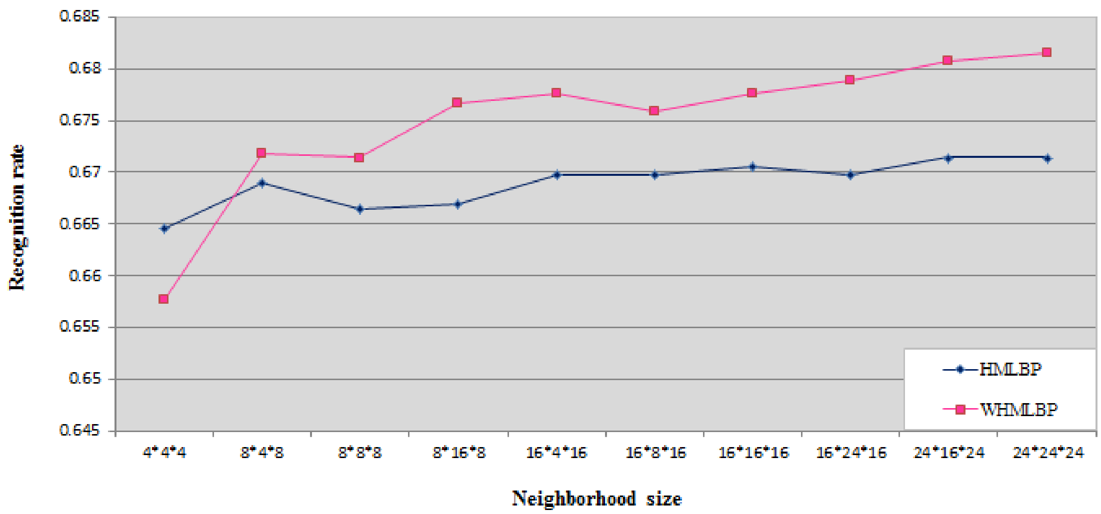
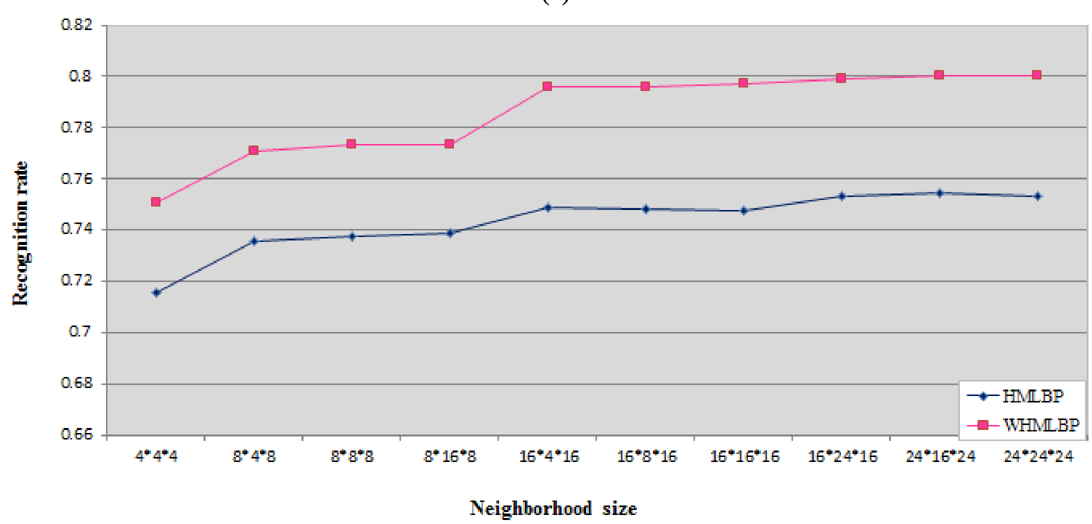


Figure 20: Performance of WHMLBP with different block sizes.

As a result I compared the performance of WHMLBP and HMLBP using 42x42 block size with the same radius R = [3, 2, 1] and different neighborhood sizes for the two datasets as in Fig. 21. The experimental results showed that the recognition rate of WHMLBP increases about 4% to 5% in Second Life dataset and the greatest accuracy is about 80.03% when the neighborhood size is 24*24*24. Moreover, in the Entropia dataset, almost all the cases are better than using HMLBP while the accuracy rate increases about 1%. The average of the recognition rate of the two methods for both datasets using different neighborhood sizes can be seen in table 2.



(a)



(b)

Figure 21: The Recognition rate of WHMLBP and HMLBP on: (a) Entropia dataset (b) Second Life dataset.

To compare the performance of WHMLBP method with other methods, I applied WLBP and LBP methods on the same two datasets. I applied both methods with R = 1, 2, 3 and P = 8, 16, 24 and I got the average of the recognition rate for both datasets as in table 2. The results I obtained demonstrate the effectiveness of our algorithm in comparison to other algorithms.

Table 2: Average Recognition Rate for Different Algorithms Compared to WHMLBP

Dataset	Techniques			
	LBP	WLBP	HMLBP	WHMLBP
Second Life	67.42%	77.27%	74.30%	78.47%
Entropia	66.45%	65.78%	66.87%	67.67%

4.6 Technique Evaluation

To evaluate our method and to be sure that the improvement achieved is statistically significant I performed some statistical tests. Before starting performing the statistical tests I have to be sure that data I have is normally distributed, since these statistical tests can only be performed over normally distributed data. I use Minitab software to plot the distribution of my data as in the following two figures (Fig. 22 and Fig. 23).

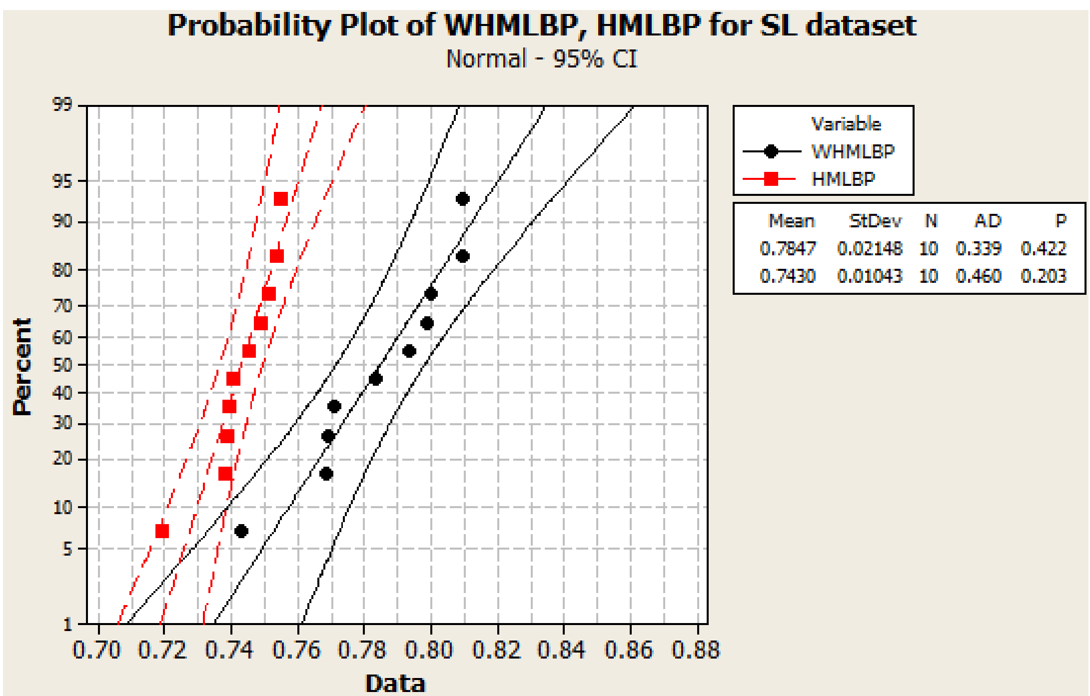


Figure 22: Distribution of WHMLBP and HMLBP data for SL dataset.

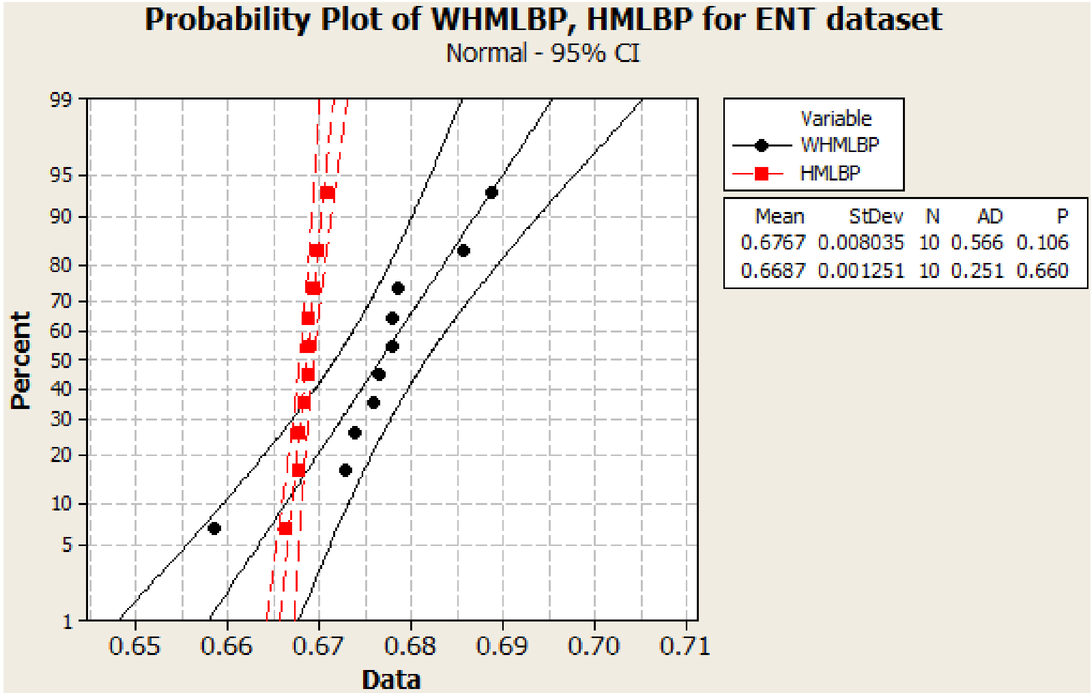


Figure 23: Distribution of WHMLBP and HMLBP data for ENT dataset.

It is clear from Fig. 22 and Fig. 23 that my data is normally distributed for both WHMLBP and HMLBP data since all or nearly all plotted points are falling within the two curves close to the straight line of a normal probability plot. Therefore, I can now check the statistical significance of my results. To satisfy this purpose I use a statistical test, Paired T-Test, and the results are as follows:

<u>Minitab outputs for SL dataset:</u>

Paired T-Test and CI: WHMLBP, HMLBP

Paired T for WHMLBP - HMLBP

	N	Mean	StDev	SE Mean
WHMLBP	10	0.78472	0.02148	0.00679
HMLBP	10	0.74303	0.01043	0.00330
Difference	10	0.04169	0.01294	0.00409

95% CI for mean difference: (0.03243, 0.05095)

T-Test of mean difference = 0 (vs not = 0): T-Value = 10.19 **<u>P-Value = 0.000</u>**

<u>Interpretation of SL result:</u>

The main reason for using Paired T-Test to evaluate the significance of my data and no other tests is that I want to block out some data and evaluate the difference between other data. I obtained these recognition rates after applying different techniques with different neighborhood sizes and I would like to evaluate the difference of recognition rates between the highest two techniques and not the difference of recognition rate between different neighborhood sizes. Therefore, I would like to block out the difference of recognition rates between neighborhood sizes. Paired T-Test can help me to block out the difference of recognition rates between neighborhood sizes and evaluate the difference of recognition rate between the highest two techniques. After applying the Paired T-Test using Minitab on the recognition rates obtained from WHMLBP and HMLBP (the highest two techniques in recognition rates) with assuming that the confidence level is 95.00% I obtained P-value = 0.000 which is less than 0.05. Therefore, I have to reject the Null hypothesis (H_0) of there is no difference in the result obtained from WHMLBP and HMLBP and accept that there is a significant difference in recognition rate (Alternative hypothesis H_1) between WHMLBP and HMLBP.

<u>Minitab outputs for ENT dataset:</u>

Paired T-Test and CI: WHMLBP, HMLBP

```
Paired T for WHMLBP - HMLBP

             N     Mean    StDev  SE Mean
WHMLBP      10  0.67673  0.00804  0.00254
HMLBP       10  0.66872  0.00125  0.00040
Difference  10  0.00801  0.00811  0.00256


95% CI for mean difference: (0.00221, 0.01381)
T-Test of mean difference = 0 (vs not = 0): T-Value = 3.12  P-Value = 0.012
```

<u>Interpretation of ENT result:</u>

The main reason for using Paired T-Test to evaluate the significance of my data and can be seen in the case of SL dataset. After applying the Paired T-Test using Minitab on the recognition rates obtained from WHMLBP and HMLBP (the highest two techniques in recognition rates) with assuming that the confidence level is 95.00% I obtained P-value = 0.012 which is less than 0.05. Therefore, I have to reject the Null hypothesis (H_0) of there is no difference in the result obtained from WHMLBP and HMLBP and accept that there is a significant difference in recognition rate (Alternative hypothesis H_1) between WHMLBP and HMLBP.

4.7 Conclusion

In this chapter, to improve the efficiency of the HMLBP in extracting useful features from an image I applied wavelet transform to the normalized manually cropped images. The effectiveness of this proposed method is shown on two avatar face datasets. Compared with HMLBP method, the proposed method gets more than 4% statistical significance improvement in the first dataset (SL) and about 1% statistical significant improvement in the second one (ENT). Compared with two other well-known methods (LBP and WLBP) the proposed method gets higher recognition rate.

CHAPTER 5

FACE RECOGNITION AND MULTI-SCALE ADAPTIVE LOCAL BINARY PATTERNS WITH DIRECTIONAL STATISTICAL FEATURES (MALBPDSF)

In this chapter, a novel face recognition technique based on wavelet transform and multi-scale adaptive local binary pattern (MALBP) with directional statistical features is proposed. The proposed technique consists of three stages: preprocessing, feature extraction and recognition. In preprocessing and feature extraction stages, wavelet decomposition is used to enhance the common features of the same subject of images and the MALBP is used to extract representative features from each facial image. Finally, the mean and the standard deviation of the local absolute difference between each pixel and its neighbors are used within ALBP and the nearest neighbor classifier to improve the classification accuracy of the LBP. Experiments conducted on ORL dataset and two virtual world avatar face image datasets show that my technique performs better than LBP, PCA, multi-scale local binary pattern, ALBP, ALBP with directional statistical features (ALBPDSF) and MALBPDSF in terms of accuracy and the time required to classify each facial image to its subject.

In this chapter, I propose a new face recognition technique to recognize both human and avatar faces. This technique uses wavelet transform to enhance the common features of the same class of facial images to improve the recognition performance. In addition, it computes the mean and the standard deviation of the local absolute difference between each pixel and its neighbors (in a specific block of pixels) within the adaptive local binary pattern (ALBP) and the nearest neighbor classifier to improve the accuracy rate.

The efficacy of our proposed method is demonstrated by experiments on ORL dataset (Olivetti Research Lab face database) and two avatar datasets from Second Life and Entropia Universe virtual worlds.

5.1 LBP with Directional Statistical Features

Suppose that a given image is of size N x M. Let g_c represents the central pixel for a circular neighborhood of size P and g_p represents its neighbors, where $p = 0,1,\ldots,P-1$. The mean (μ_p) and the standard deviation (σ_p) of the local difference $|g_c - g_p|$ can be computed using (Mohamed et al., 2012; Mohamed & Yampolskiy, 2012d; Zhenhua, Lei, Zhang, & Su, 2010):

$$\mu_p = \sum_{i=1}^{N}\sum_{j=1}^{M}\left|g_c(i,j) - g_p(i,j)\right| /(M * N) \tag{7}$$

$$\sigma_p = \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{M}(\left|g_c(i,j) - g_p(i,j)\right| - \mu_p)^2 /(M * N)} \tag{8}$$

The vector $\vec{\mu} = [\mu_0, \mu_1, ..., \mu_{P-1}]$ refers to the mean vector and $\vec{\sigma} = [\sigma_0, \sigma_1, ..., \sigma_{P-1}]$ refers to the standard deviation (*std*) vector.

The two vectors represent the directional statistical features of the local difference $|g_c - g_p|$ and they carry useful information for image discrimination that can be used to define the weighted LBP dissimilarity. Let $\vec{\mu}_X$ and $\vec{\sigma}_X$ refer to the directional statistical feature vectors for a sample test image X while $\vec{\mu}_Y$ and $\vec{\sigma}_Y$ refer to the two vectors for a class model Y then the normalized distances between $\vec{\mu}_X$ and $\vec{\mu}_Y$, and $\vec{\sigma}_X$ and $\vec{\sigma}_Y$ can be defined as (Mohamed et al., 2012; Mohamed & Yampolskiy, 2012d; Zhenhua, Lei, Zhang, & Su, 2010):

$$d_\mu = \sum_{p=0}^{P-1}\left|\vec{\mu}_X(p) - \vec{\mu}_Y(p)\right| /(P * e_\mu), d_\sigma = \sum_{p=0}^{P-1}\left|\vec{\sigma}_X(p) - \vec{\sigma}_Y(p)\right| /(P * e_\sigma) \tag{9}$$

where e_μ and e_σ are the standard deviations of $\vec{\mu}$ and $\vec{\sigma}$ respectively from training samples images.

So the weighted LBP dissimilarity with statistical features using d_μ and d_σ can be defined as (Mohamed et al., 2012; Mohamed & Yampolskiy, 2012d; Zhenhua, Lei, Zhang, & Su, 2010):

$$D_{LBP}^{F}(X,Y) = D_{LBP}(X,Y) * (1 + c_1 - c_1 * \exp(-d_\mu / c_2)) * (1 + c_1 - c_1 * \exp(-d_\sigma / c_2)) \tag{10}$$

where D_{LBP} (X, Y) is the LBP histogram dissimilarity, c_1 and c_2 are two control parameters for the weights.

5.2 Adaptive Local Binary Pattern (ALBP)

The directional statistical feature vectors can be used to improve the classification performance of an image by minimizing the variations of the mean and the *std* of the directional difference along different orientations. To this end, a new version of the LBP was proposed by Guo et al. (Zhenhua, Lei, Zhang, & Su, 2010), Adaptive LBP (ALBP), to reduce the estimation error of local difference between each pixel and its neighbors. A new parameter called local weight (w_p) is defined in the LBP equation and so the new definition of the LBP equation will have the following form (Mohamed et al., 2012; Mohamed & Yampolskiy, 2012b, 2012d; Zhenhua, Lei, Zhang, & Su, 2010):

$$ALBP_{P,R} = \sum_{p=0}^{P=1} 2^p S(g_p * w_p - g_c) \tag{11}$$

where the objective function to compute the weight w_p is as follows:

$$J = \sum_{i=1}^{N} \sum_{j=1}^{M} (g_c(i,j) - w * g_p(i,j))^2 \tag{12}$$

the target of this equation is to minimize the directional difference $|g_c\text{-}w_p{*}g_p|$ to this end I have to derive equation 12 with respect to w and assign the derivation to zero as follows (Mohamed & Yampolskiy, 2012d):

$$\frac{\partial J}{\partial w} = -2 \sum_{i=1}^{N} \sum_{j=1}^{M} (g_p(i,j)(g_c(i,j) - w * g_p(i,j))) = 0 \tag{13}$$

so I get:

$$w * \sum_{i=1}^{N} \sum_{j=1}^{M} g_p(i,j) g_p(i,j) = \sum_{i=1}^{N} \sum_{j=1}^{M} g_p(i,j) g_c(i,j) \tag{14}$$

$$w = \frac{\sum_{i=1}^{N} \sum_{j=1}^{M} g_p(i,j) g_c(i,j)}{\sum_{i=1}^{N} \sum_{j=1}^{M} g_p(i,j) g_p(i,j)} \tag{15}$$

from equation 15 I get (Mohamed et al., 2012; Mohamed & Yampolskiy, 2012b, 2012d; Zhenhua, Lei, Zhang, & Su, 2010):

$$w_p = \vec{g}_p^{\,T} \vec{g}_c / (\vec{g}_p^{\,T} \vec{g}_p) \tag{16}$$

where g_c = [$g_c(1,1);g_c(1,2);\ldots;g_c(N,M)$] is a column vector that contains all possible values of any pixel $g_c(i,j)$, N x M is the size of an image and g_p = [$g_p(1,1);g_p(1,2);\ldots;g_p(N,M)$] is the corresponding vector for all $g_p(i,j)$ pixels. Let $\vec{w} = [w_0, w_1, \ldots, w_{P-1}]$ refers to the ALBP weight vector. I have to note that each weight w_p is computed along one orientation $2\pi p/P$ for the whole image.

5.2.1 ALBP with Directional Statistical Features

By using the ALBP weight the directional statistics equations (7) and (8) can be changed to (Mohamed et al., 2012; Mohamed, Gavrilova, & Yampolskiy, 2013; Mohamed & Yampolskiy, 2012d; Zhenhua, Lei, Zhang, & Su, 2010):

$$\mu_p = \sum_{i=1}^{N}\sum_{j=1}^{M}\left| g_c(i,j) - g_p(i,j) * w_p \right| /(M*N) \tag{17}$$

$$\sigma_p = \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{M}\left(\left| g_c(i,j) - g_p(i,j) * w_p \right| - \mu_p\right)^2 /(M*N)} \tag{18}$$

Based on the ALBP weight w_p, I have three vectors $\vec{\mu}$, $\vec{\sigma}$ and $\vec{w}$. Similar to the normalized distance between $\vec{\mu}_X$ and $\vec{\mu}_Y$, and $\vec{\sigma}_X$ and $\vec{\sigma}_Y$ I can define the normalized distance between $\vec{w}_X$ and $\vec{w}_Y$ as:

$$d_w = \sum_{p=0}^{P-1}\left| \vec{w}_X(p) - \vec{w}_Y(p) \right| /(P*e_w) \tag{19}$$

where e_w is the standard deviation of $\vec{w}$ from training samples images.

The weighted ALBP dissimilarity with statistical features using d_μ , d_σ and d_w can be defined as (Mohamed et al., 2012, 2013; Mohamed & Yampolskiy, 2012d; Zhenhua, Lei, Zhang, & Su, 2010):

$$\begin{aligned} D_{ALBP}^{F}(X,Y) = D_{ALBP}(X,Y) * (1 + c_1 - c_1 * \exp(-d_\mu / c_2)) \\ * (1 + c_1 - c_1 * \exp(-d_\sigma / c_2)) * (1 + c_1 - c_1 * \exp(-d_w / c_2)) \end{aligned} \tag{20}$$

where D_{ALBP} (X, Y) is the ALBP histogram dissimilarity.

5.3 Wavelet-based Multi-scale ALBP with Directional Statistical Features (WMALBPDSF)

I have presented a general ALBP operator in section 5.1 for extracting the facial images features using a single scale circular symmetric neighbor set of P pixels placed on a circle of radius R with the weight parameter w_p. By altering P and R and combining the resulted images, a multiresolution representation can be obtained. However, the main problem associated with the multiresolution analysis is the high dimensionality of the representation. There are some approaches to overcome this problem. One of these approaches minimizes the redundant information by applying feature selection techniques (Raja & Gong, 2006). Another method reduces the dimensionality of the multiresolution representation by combining the multi-scale local binary pattern representation with linear discriminant analysis (LDA) to extract the features (C.-H. Chan, J. Kittler, & K. Messer, 2007). I propose another method to reduce the dimensionality by decomposing an image into a specific level of decomposition and then using the resulted approximation image for extracting the features.

5.3.1 WMALBPDSF Operator

In this approach, I combine Daubechies wavelet transform with the multi-scale adaptive local binary pattern representation. The first level of decomposition using wavelet transform is first applied to a face image. This decomposition generates three detailed sub-images in three different directions and an approximation sub-image in the fourth one. The three detailed sub-images contain most of the local changes of the facial image while the approximation sub-image contains most of the common features of the same class and so it will be decomposed to the next level of decomposition. To obtain a higher level of decomposition we have to decompose the approximation image of its previous level. We have to repeat this process until we reach to the best level of decomposition describing our data.

The adaptive local binary pattern operators at L scales are then applied to the approximation facial image. This generates a new grey level code for each pixel at every resolution (Mohamed & Yampolskiy, 2012d). The resulting ALBP images are divided into non-overlapping sub-regions, $T_1, T_2, .., T_q$, where q is the number of sub-regions. The set of histograms computed at different scales for the same sub-region provides regional information about that region and they have to be concatenated into a single histogram. This single histogram represents the final multiresolution regional face descriptor for this region.

Concatenating the final multiresolution regional face histogram for each region will form the final multiresolution face histogram for the whole facial image. By using the weighted ALBP dissimilarity with statistical features defined by equation 20 with the nearest neighborhood classifier for the histograms of both training and testing images we can classify each image to its class.

5.4 Experiments

In this section, I verify the performance of the proposed algorithm on two different types of datasets: the first type is real world well-known human faces dataset, ORL database ("The ORL database of faces,"), and the second type is virtual world datasets from Second Life ("Second Life,") and Entropia Universe ("Entropia Universe,") virtual worlds. Fig. 24 shows an example of a subject from each dataset. The proposed method is compared with single scale LBP, multi-scale LBP, ALBP and ALBP with directional statistical features (ALBPDSF).

(a)

(b)

(c)

Figure 24: Samples of one subject of facial images from: a) ORL dataset ("The ORL database of faces,") b) Second Life dataset ("Second Life,") c) Entropia dataset ("Entropia Universe,").

5.4.1 Experimental Setup

To evaluate the proposed technique I have used three facial image datasets. The first one is the ORL dataset. The ORL dataset contains 400 images representing 40 distinct subjects. Each subject has 10 different images. These images were taken at different times, varying the lighting, pose angle, facial expressions (open eyes, closed eyes, smiling, not smiling) and accessories (wearing glasses or no glasses).

The whole dataset was taken against a dark homogeneous background with the subjects in an upright, frontal position and each is grayscale image with a resolution of 92 x 112 pixels. I have used all images in this dataset during our experiments without doing cropping. After applying the first level of wavelet decomposition, the resolution of each image in the ORL dataset was changed from 92 x 112 to 46 x 56.
The second dataset was collected from the Second Life (SL) virtual world. This dataset contains 581 gray scale images with size 1280 x 1024 each to represent 83 different avatars. Each avatar subject has 7 different images for the same avatar with different frontal pose angle (front, far left, mid left, far right, mid right, top and bottom) and facial expression.

The last dataset was collected from Entropia (ENT) Universe virtual world. ENT dataset contains 490 gray scale images with size 407 x 549 pixels. These images were organized into 98 subjects (avatars). Each subject has different 5 images for the same avatar with different frontal angle and occlusions (wearing a mask or no).

The facial part of each image in SL and ENT datasets was manually cropped from the original images based on the location of the two eyes, mouth and the nose. The new size of each facial image in SL dataset is 260 x 260 pixels while in ENT dataset each facial image was resized to the size of 180 x 180 pixels. After applying the first level of wavelet decomposition, the resolution of each facial image in the SL dataset will be reduced to be 130 x 130 and to 90 x 90 for ENT dataset.

The performance of our method is affected by four parameters (Mohamed & Yampolskiy, 2012d). The first one is the wavelet decomposition level. During experiments, I used to apply different families and levels of decomposition of Daubechies wavelet transform on all datasets. I figured out that the performance of my technique differs from one dataset to another and within the same dataset based on the decomposition family and level. Therefore, choose of the best family and level of decomposition is based on the dataset itself. The second parameter is the circular neighborhood size P. Choosing a large size for the neighborhood increases the length of the histogram and then slows down the computation of the dissimilarity measure. Choosing a small size for the neighborhood size may lead to information loss. During my experiments I have chosen a neighborhood of size P = 8, 16. The third parameter is the number of multi-scale operators. Using small number of operators cannot provide sufficient information about the facial images, also using large radius value reduces the size of the corresponding ALBP images. Therefore, in my experiments I have selected L = 10 which means that I have used 10 LBP operators to represent each facial image with P = 8, 16 and R = 1, 2, 3, 4, 5. The fourth parameter is the number of the facial image sub-regions q. Dividing the facial image into a large number of small sub-regions increases the computation time and may reduce the system accuracy while dividing the facial image into a small number of large sub-regions increases the loss of spatial information (C.-H. Chan et al., 2007). In my experiments, each facial image has been divided into q x q non-overlapping rectangle size sub-regions while the best value of q is obtained experimentally.

5.4.2 Experimental Results

In order to gain better understanding on whether using wavelet transform with MALBP with directional statistical features (MALBPDSF) is advantageous or not, I compared WMALBPDSF with ALBPDSF, ALBP, MLBP and LBP. First, I got the average of recognition rate of WMALBPDSF with different dataset using different Daubechies wavelet transform to decide which wavelet family better described my data. After that, I got the average of recognition rate for all datasets with different levels of decomposition within the same decomposition family as in Fig. 25.

It is clear from Fig. 25 that the best family describing each one of SL, ENT and ORL datasets is Db4, Db3 and Db5 respectively. Therefore, I use these wavelet families to build the wavelet form of MALBPDSF to deal with each one of the three datasets.

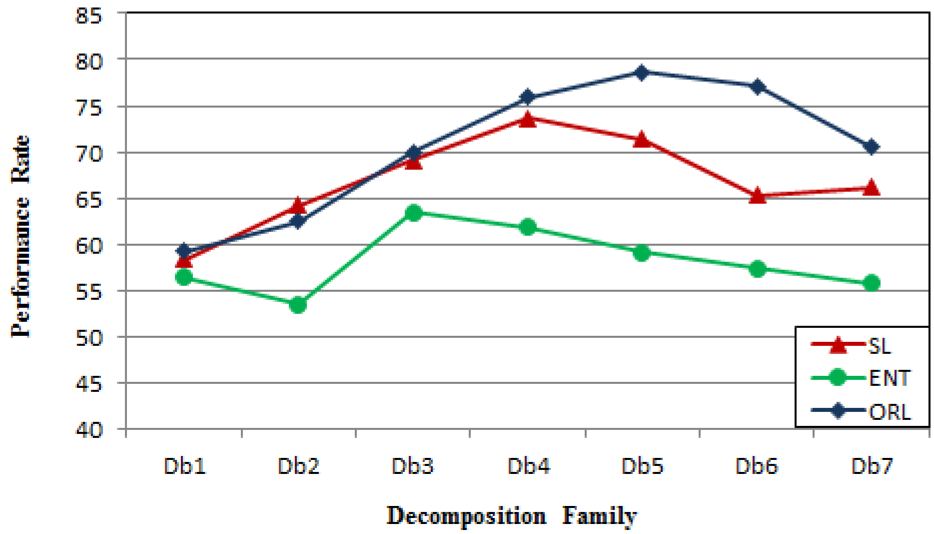


Figure 25: WMALBPDSF average recognition rate for different datasets.

Second, I compared the performance of WMALBPDSF , ALBPDSF, ALBP, MLBP and LBP using the ten different LBP operators and with different number of regions (q) over the SL, ENT and ORL datasets. In this experiment, the training sets were built by randomly selecting one, two and three images from each class for each of the three datasets while the rest is used for testing. The results showed that the average recognition rate of using WMALBPDSF is better than the average recognition rate of using the other methods with almost all values of q and within all datasets. The recognition rate on average using WMALBPDSF is greater than that of its closest competitor, which is MLBP for SL dataset, ALBPDSF for ENT dataset and MLBP for ORL dataset, by about 7%, 3% and 4% respectively as in figures 26, 27 and 28.

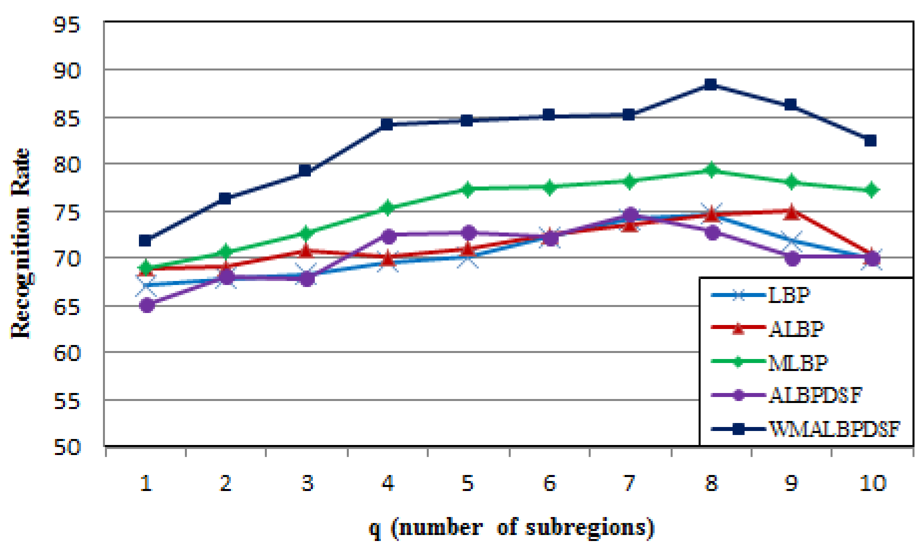


Figure 26: Average recognition rate for SL dataset for different sub-regions.

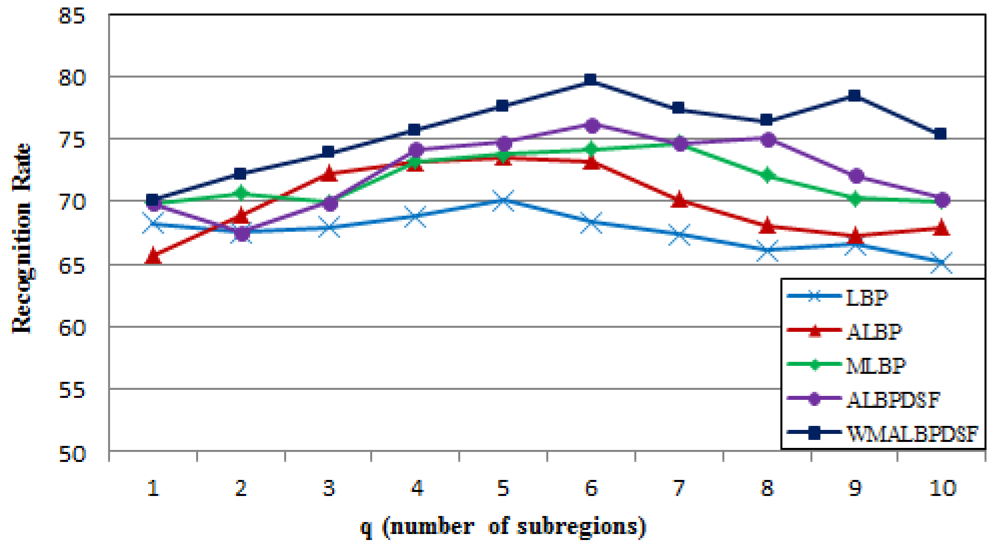


Figure 27: Average recognition rate for ENT dataset for different sub-regions.

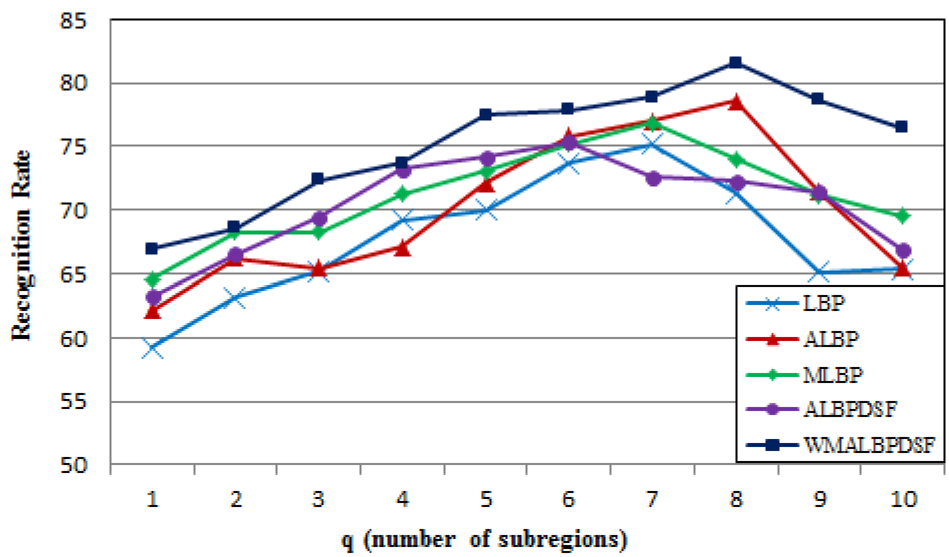


Figure 28: Average recognition rate for ORL dataset for different sub-regions.

Comparing to other methods using wavelet transform with the MALBPDSF improves the recognition rate up to some point and after that the recognition rate starts to be reduced based on the window size. As expected, the recognition rate is reduced with large window (sub-region) size because of the loss of information. Based on the datasets themselves and from figures 26, 27 and 28, it is shown that my technique provides a high and a robust average recognition rate especially when $8 \geq q > 4$ for SL dataset, $9 \geq q \geq 5$ for ENT dataset and $8 \geq q \geq 5$ for ORL dataset.

The results showed also that not only the recognition rate of using WMALBPDSF is better than that of the other methods but also the time that WMALBPDSF requires to classify each input facial image to its class is less than that when compared to other methods (see table 3). This is an expected result since one of the main reasons for using wavelet decomposition in face recognition systems is that it reduces the computational complexity and overhead of the system and so the system can run faster.

Table 3: Average time in seconds required by different algorithms

	Algorithm				
	LBP	ALBP	ALBPDSF	MLBP	WMALBPDSF
Time	17.21	19.71	21.96	19.85	13.81

5.5 System Evaluation

In this section, I evaluate the result obtained from applying WMALBPDSF on different datasets and make sure that the improvements in accuracy rates are statistically significant. Before that, I have to be sure that the data that I have is normally distributed. Therefore, I used Minitab software to plot the probability graph for the result obtained from applying WMALBPDSF and its competitor technique for each one of the datasets (see figures 29, 30 and 31).

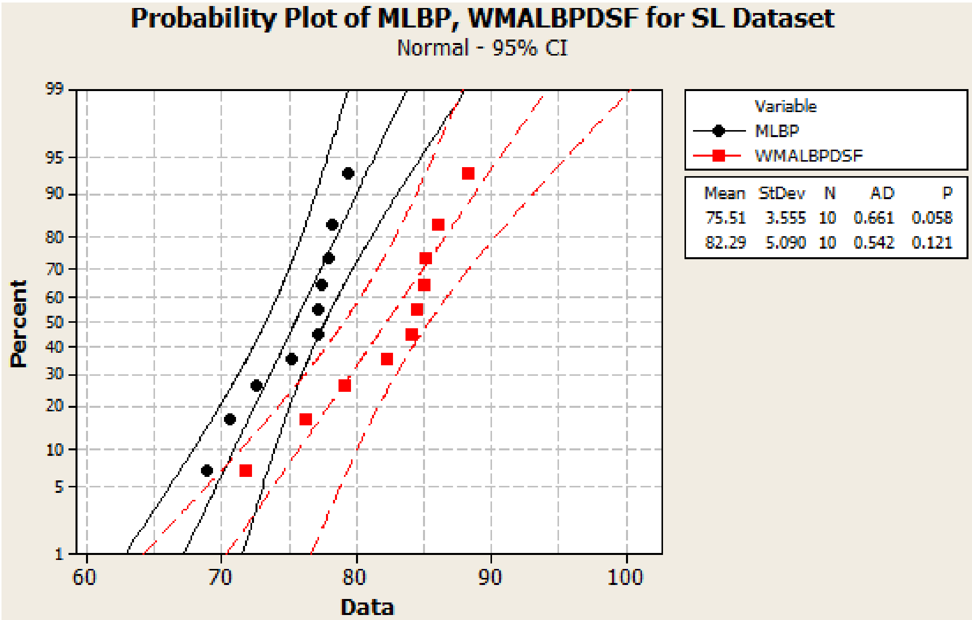


Figure 29: Distribution of MLBP and WMALBPDSF data for SL dataset.

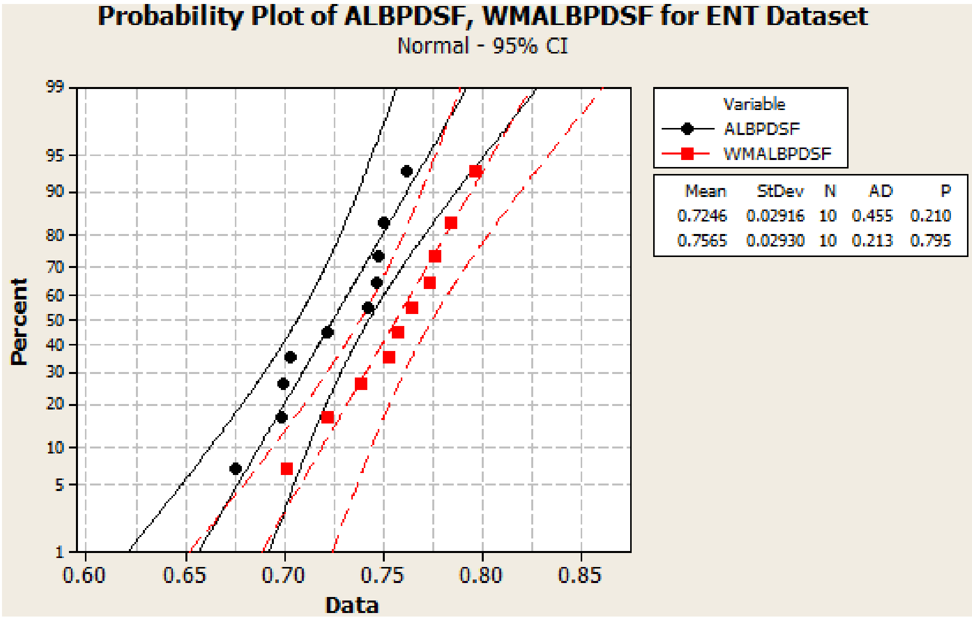


Figure 30: Distribution of ALBPDSF and WMALBPDSF data for ENT dataset.

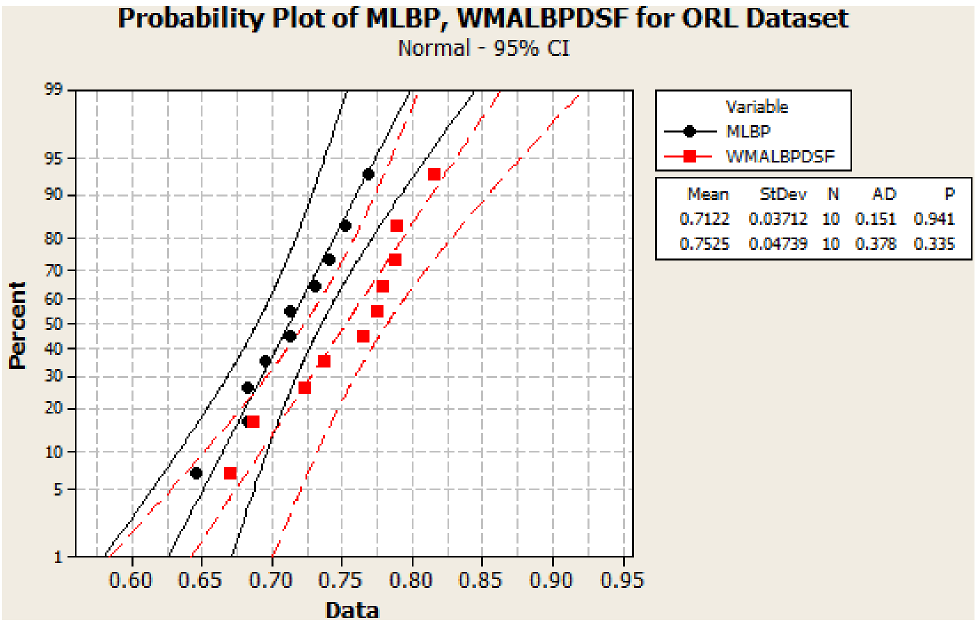


Figure 31: Distribution of MLBP and WMALBPDSF data for ORL dataset.

Based on figures 29, 30 and 31, it is clear that my data is normally distributed and then I can apply a statistical test (Paired T-Test) to study its statistical significance as follows:

<u>Minitab outputs for SL dataset:</u>

Paired T-Test and CI: MLBP, WMALBPDSF

```
Paired T for MLBP - WMALBPDSF

             N    Mean  StDev  SE Mean
MLBP        10   75.51   3.56     1.12
WMALBPDSF   10   82.29   5.09     1.61
Difference  10  -6.782  1.862    0.589


95% CI for mean difference: (-8.114, -5.450)
T-Test of mean difference = 0 (vs not = 0): T-Value = -11.52  P-Value = 0.000
```

Interpretation of SL result:

To evaluate the difference in recognition rates between MLBP and WMALBPDSF, which have the highest two recognition rates for SL dataset, I applied Paired-T Test. In figure 26, recognition rates are obtained after applying techniques, such as LBP and WMALBPDSF, over different number of image sub-regions. I want to evaluate the difference of recognition rates between techniques and not the difference between image sub-regions. Therefore, I want to block out the difference of image sub-regions. Paired T-Test can help me to block out the difference of recognition rates between applying different image sub-regions and evaluate the difference of recognition rate between the highest two techniques. After applying the Paired T-Test using Minitab on the recognition rates obtained from WMALBPDSF and MLBP (the highest two techniques in recognition rates) assuming that the confidence level is 95.00% I obtained P-value = 0.000 which is less than 0.05. Therefore, I have to reject the Null hypothesis (H_0) of there is no difference in the result obtained from WMALBPDSF and MLBP and accept that there is a significant difference in recognition rate (Alternative hypothesis H_1) between WMALBPDSF and MLBP.

Minitab outputs for ENT dataset:

Paired T-Test and CI: ALBPDSF, WMALBPDSF

```
Paired T for ALBPDSF - WMALBPDSF

             N    Mean  StDev  SE Mean
ALBPDSF     10  72.458  2.916    0.922
WMALBPDSF   10  75.649  2.930    0.927
Difference  10  -3.191  1.839    0.581


95% CI for mean difference: (-4.506, -1.876)
T-Test of mean difference = 0 (vs not = 0): T-Value = -5.49  P-Value = 0.000
```

<u>Interpretation of ENT result:</u>

The last row in the results obtained from applying the Paired T-Test states that P-value = 0.000 less than 0.05. With the assumption that the confidence level is 95.00%, I have to reject the Null hypothesis (H_0) of no difference in recognition rate between ALBPDSF and WMALBPDSF. Therefore, I believe there is a significant difference in recognition rate between WMALBPDSF and ALBPDSF.

<u>Minitab outputs for ORL dataset:</u>

Paired T-Test and CI: MLBP, WMALBPDSF

```
Paired T for MLBP - WMALBPDSF

             N    Mean  StDev  SE Mean
MLBP        10   71.22   3.71     1.17
WMALBPDSF   10   75.25   4.74     1.50
Difference  10  -4.030  2.524    0.798


95% CI for mean difference: (-5.836, -2.224)
T-Test of mean difference = 0 (vs not = 0): T-Value = -5.05  P-Value = 0.001
```

<u>Interpretation of ORL result:</u>

The last row in the results obtained from applying the Paired T-Test states that P-value = 0.001 less than 0.05. With the assumption that the confidence level is 95.00%, I have to reject the Null hypothesis (H_0) of no difference in recognition rate between MLBP and WMALBPDSF. Therefore, I believe there is a significant difference in recognition rate between WMALBPDSF and MLBP.

5.6 Conclusions

In this chapter, a novel LBP face recognition approach (WMALBPDSF) is proposed based on wavelet transform and adaptive local binary pattern with directional statistical features. The effectiveness of the proposed method is shown in recognizing faces from both real and virtual worlds. Compared with LBP, ALBP, MLBP and ALBPDSF and with different LBP operators and image sub-regions, my proposed technique improved the recognition rate of the ORL, SL and ENT datasets by about 4%, 7% and 3% respectively. In addition, the time required by my technique to classify each input facial image to its class is less than what is required by other methods.

CHAPTER 6

DISTINGUISHING BETWEEN HUMAN AND AVATAR FACES FOR THE AVATAR CAPTCHA RECOGNITION CHALLENGE

CAPTCHAs are challenge-response tests used in many online systems to prevent attacks by automated bots. Avatar CAPTCHAs are a recently-proposed variant in which users are asked to classify between human faces and computer generated avatar faces, and have been shown to be secure if bots employ random guessing. I test a variety of modern object recognition and machine learning approaches on the problem of avatar versus human face classification. My results show that using these techniques, a bot can successfully solve Avatar CAPTCHAs as often as humans can. These experiments suggest that this high performance is caused more by biases in the facial datasets used by Avatar CAPTCHAs and not by a fundamental flaw in the concept itself, but my results highlight the difficulty in creating CAPTCHA tasks that are immune to automatic solution.

6.1 Introduction

Online activities play an important role in our daily life, allowing us to carry out a wide variety of important day-to-day tasks including communication, commerce, banking, and voting (L. Ahn, Blum, Hopper, & Langford, 2003; Haichang, Dan, Honggang, Xiyang, & Liming, 2010). Unfortunately, undesirable automated programs, or "bots," that abuse services by posing as human beings to (for example) repeatedly vote in a poll, add spam to online message boards, or open thousands of email accounts for various nefarious purposes often misuse these online services. One approach to prevent such misuse has been the introduction of online security systems called CAPTCHAs, or Completely Automated Public Turing tests to tell Computers and Humans Apart (L. Ahn et al., 2003). CAPTCHAs are simple challenge-response tests that are generated and graded by computers, and that are designed to be easily solvable by humans but that are beyond the capabilities of current computer programs (Liming et al., 2010). If a correct solution for a test is received, it is assumed that a human user (and not a bot) is requesting an Internet service.

Three main categories of CAPTCHAs have been introduced (Chandavale, Sapkal, & Jalnekar, 2010). Text-based CAPTCHAs generate distorted images of text which are very hard to be recognized by state-of-the-art optical character recognition (OCR) programs but are easily recognizable by most humans. Sound-based CAPTCHAs require the user to solve a speech recognition task, while others require the user to read out a given sentence to authenticate that he/she is a human. Finally, image-based CAPTCHAs require the user to solve an image recognition task, such as entering a label to describe an image (Haichang et al., 2010). Other work has combined multiple of these categories into multi-modal CAPTCHAs (Almazyad, Ahmad, & Kouchay, 2011), which can increase security while also giving users a choice of the type of CAPTCHA they wish to solve.

The strength of a CAPTCHA system can be measured by how many trials an attacking bot needs on average before solving it correctly (Chandavale et al., 2010). However, there is a tension between developing a task that is as difficult as possible for a bot, but is still easily solvable by human beings. This is complicated by human users who may have sensory or cognitive handicaps that prevent them from solving certain CAPTCHAs. The best CAPTCHA schemes are thus the ones which are easy for almost any human to solve but that are almost impossible for an automated program.

Recently, a novel image-based system was proposed called Avatar CAPTCHA (D. D'Souza, Polina, & Yampolskiy, 2012) in which users are asked to perform a face classification task. In particular, the system presents a set of face images, some of which are actual human faces while others are avatar faces generated by a computer, and the user is required to select the real faces.The designers of the scheme found that humans were able to solve the puzzle (by correctly finding all human faces) about 63% of the time, while a bot that randomly guesses the answers would pass only about 0.02% of the time.

In this chapter, I consider how well a bot could perform against this CAPTCHA if, instead of random guessing, it used computer vision algorithms to try to classify between human and avatar faces. Through experiments conducted on the human and avatar face images released by the authors of (D. D'Souza et al., 2012), I test a variety of modern learning-based recognition algorithms, finding that this task is surprisingly easy, with some algorithms actually outperforming humans on this dataset. While these results indicate that Avatar CAPTCHA is not as secure as the authors had hoped, our results suggest that the problem may not be in the idea of Avatar CAPTCHA, but instead in the way the avatar facial images were generated, allowing the recognition algorithms to learn subtle biases in the data.

6.2 Background and related work

As noted above, text-based CAPTCHAs are currently the most common systems on the web, and have been successfully deployed for almost a decade (L. Ahn et al., 2003). In order to increase the level of security against increasingly sophisticated OCR algorithms, text based CAPTCHAs have had to increase the degree of distortion of the letters or numbers and hence may become so difficult that even humans are unable to recognize all of the text correctly. To address this problem, CAPTCHA systems using image-based labeling tasks have been proposed (L. V. Ahn, Blum, & Langford, 2004; Chandavale et al., 2010; Elson, Douceur, Howell, & Saul, 2007). No distortion is required for many of these tasks because humans can easily identify thousands of objects in images, while even state-of the-art computer vision algorithms cannot perform this task reliably, especially when the set of possible classes is drawn from very large datasets (D. D'Souza et al., 2012). While image-based CAPTCHAs are still never completely secure, they are thought to widen the success rate gap between humans and non-humans.

***Avatar CAPTCHA*:** The authors of (D. D'Souza et al., 2012) proposed Avatar CAPTCHA as a specific type of image-based task. In their approach, the system presents 12 images organized into a two-by-six matrix, with each image either a human face from a face dataset or a synthetic face from a dataset of avatar faces. The relative number of human and avatar faces and their arrangement is chosen randomly by the system. The user's task is to select all (and only) the avatar images among these 12 images by checking a checkbox under each avatar image. The user is authenticated as a human if he/she correctly completes the task, and otherwise is considered a bot. Using brute force attack, a bot has a success rate of 50% for each of the 12 images, since each image is either a human or avatar, so the probability of correctly classifying all 12 images is just $0.5^{12} = 1/4096$. Humans, on the other hand, were found to complete the task correctly about 63% of the time. In this chapter, I show that a bot can achieve significantly higher performance than random guessing, and even outperform humans, using object recognition and machine learning.

6.3 Methods

I apply a variety of learning-based recognition approaches to the task of classifying between human and avatar faces. For data, I used a publicly-available dataset released by the authors of (D. D'Souza et al., 2012) as part of the Face Recognition Challenge held in conjunction with the International Conference on Machine Learning and Applications (ICMLA 2012) conference (Yampolskiy, 2012). This dataset consists of 200 grayscale photos, split evenly between humans and avatars. The human dataset consists of frontal grayscale facial images of 50 males and 50 females with variations in lighting and facial expressions. The avatar dataset consists of 100 frontal grayscale facial images collected from the Entropia Universe and Second Life virtual worlds. All images were resampled to a uniform resolution of 50 x 75. Fig. 32 shows sample images from the dataset. Each of our recognition approaches follows the same basic recipe: I use a particular choice of visual feature, which is used to produce a feature vector from an image, I learn a 2-class (human vs. avatar) classifier using labeled training data, and then apply the classifier on a disjoint set of test images. I now describe the various visual features and classifiers that I employed.

Figure 32: Sample avatar faces (top) and human faces (bottom) used in experiments.

6.3.1 Naïve Approaches

As baselines, I start with three simple approaches using raw pixel values as features.

***Raw images*:** These feature vectors are simply the raw grayscale pixel values of the image, concatenated into a 50 x 75 = 3750 dimensional vector.

***Summary statistics*:** As an even simpler baseline, I use a 1D feature that consists only of the mean grayscale value of the image. A second baseline represents each image as a vector of five dimensions, consisting of the maximum pixel value, the minimum pixel value, the average pixel value, the median pixel value, and the sum of all pixel values.

***Grayscale histograms*:** This feature consists of a simple histogram of the grayscale values in the image. I tested different quantizations of the histogram, in particular testing histograms with 256, 128, 64, 32, 16, 8, 4, and 2 bins.

6.3.2 Histograms of Oriented Gradients (HOG)

Histograms of Oriented Gradients (HOG) features have become very popular in the recognition community for a variety of objects including people (Dalal & Triggs, 2005). Computing these features consists of 5 stages: (1) global image normalization to reduce effect of changing illumination, (2) computing the image gradient at each pixel, (3) dividing the image into small 8x8 pixel cells, and then computing histograms over gradient orientation within each cell, (4) normalization of the histograms within overlapping blocks of cells, and (5) creating a feature vector, by concatenating all normalized histograms for all cells into a single vector. For the images in our dataset, this procedure yields a 2268 dimensional feature vector.

6.3.3 GIST

The GIST descriptor (Oliva & Torralba, 2001) was originally developed for scene recognition but has become popular for other recognition problems as well. This feature applies a series of filters to an image, each of which responds to image characteristics at different scales and orientations. The image is divided into a 4 x 4 grid of regions, and the average response of each filter is calculated within each region. This yields a descriptor that captures the "GIST" of the scene: the orientation and scale properties of major image features at a coarse resolution, yielding a 960 dimensional vector.

6.3.4 Quantized Feature Descriptors

Another popular technique in recognition is to detect a sparse set of highly distinctive feature points in an image, calculate an invariant descriptor for each point, and then represent an image in terms of a histogram of vector-quantized descriptors. The Scale-Invariant Feature Transform (SIFT) (Lowe, 1999) and Speeded-Up Robust Features (SURF) (Bay, Tuytelaars, & Gool, 2006) are two commonly-used descriptors; I use the latter here. I use SURF to detect features points and calculate descriptors for each point, and then use k-means to produce a set of 50 clusters. I then assign each descriptor to the nearest visual word, and represent each image as a histogram over these visual words, yielding a 50 dimensional feature vector. Fig. 33 illustrates some detected SURF features.

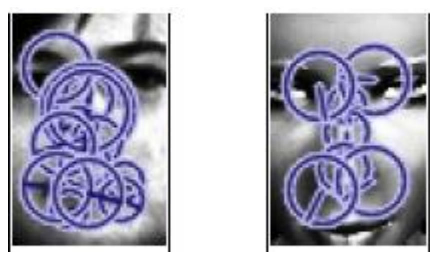


Figure 33: Detected SURF features for a human face (left) and avatar face (right).

6.3.5 Local Binary Pattern-based Features

Four-Patch Local Binary Pattern (FPLBP): The local binary pattern (LBP) descriptor examines each pixel in a small neighborhood of a central pixel, and assigns a binary bit depending on whether the grayscale value is greater than or less than that of the central pixel. The bits that represent the comparison are then concatenated to form an 8-bit decimal number, and a histogram of these values is computed. FPLBP is an extension to the original LBP where for each pixel in the image I consider two rings, an inner ring of radius r_1 and an outer one of radius r_2 (I use 4 and 5, respectively), each centered around a pixel (Wolf, Hassner, & Taigman, 2008). T patches of size s x s (I use s = 3) are spread out evenly on each ring. Since I have T patches along each ring then I have T/2 center symmetric pairs. Two center symmetric patches in the inner ring are compared with two center symmetric patches in the outer ring, each time setting one bit in each pixel's code based on which of the two pairs are more similar, and then calculate a histogram from the resulting decimal values.

Local Difference Pattern Descriptor: I also introduce a simple modification to the above approach which I call Local Difference Pattern. I divide the image into *n* x *n* (3x3) windows and compute a new value for the center of each window based on the values of its neighbors. I compute the new value as the average of the differences between the center and all other pixels in the window (instead of computing the binary window and converting it into its decimal value as in LBP). I tried using both absolute and signed differences. Fig. 34 illustrates this feature. Finally, I compute a histogram for these new values.

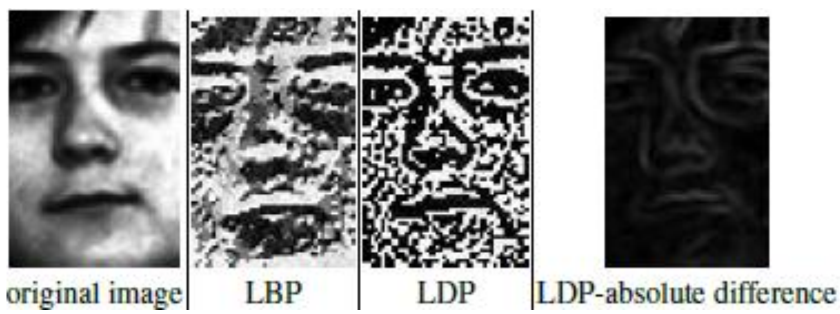


Figure 34: Illustration of LBP and LDP features for a human face.

6.3.6 Classifiers and Feature selection methods

For learning the models from each of the above feature times, I applied two different types of classifiers: Naïve Bayes (Hall et al., 2009; John & Langley, 1995), and LibLinear with L2-regularized logistic regression (Fan, Chang, Hsieh, Wang, & Lin, 2008). I used Correlation-based Feature Selection (CFS) (Hall, 1999) to reduce feature dimensionality.

6.4 Results

Table 4 presents the results on the face-versus-avatar classification task for our simplest features (the Naïve features based on raw pixel values) and our simplest classifier (Naïve Bayes). All results presented here were evaluated using 10-fold cross-validation. The best classification rate obtained in this set of experiments is 93%, when raw grayscale pixel values concatenated into a vector are used as features. Interestingly, even much simpler techniques give results that are significantly better than random guessing (which would yield 50% accuracy). The 128-dimensional grayscale histograms achieve 92% accuracy, but even 4-dimensional histograms achieve almost 70% accuracy. Our simplest method, which encodes an image as a single dimension corresponding to its mean pixel value, gives an accuracy of 56% (M. Korayem, A. A. Mohamed, D. Crandall, & R.V. Yampolskiy, 2012; Mohammed Korayem, Abdallah A. Mohamed, David Crandall, & Roman V. Yampolskiy, 2012).

Table 4: Classification results using Naïve features and Naïve Bayes classifiers

Method	Accuracy	Precision	Recall	F-measure
Pixels-value	**93%**	**93.2%**	**93%**	**93%**
Histograms(256-Bins)	89%	89.8%	89%	88.9%
Histograms(128-Bins)	92%	92.3%	92%	92%
Histograms(64-Bins)	77%	77.3%	77%	76.9%
Histograms(32-Bins)	78%	78.2%	78%	78%
Histograms(16-Bins)	75%	75.1%	75%	75%
Histograms(8-Bins)	77%	77.9%	77%	76.8%
Histograms(4-Bins)	69%	69.1%	69%	69%
Histograms(2-Bins)	52%	52.1%	52%	51.7%
Average-mean-pixel	57%	57.4%	56%	53.8%
Avg Min Max Sum Median	61%	62.9%	61%	59.5%

The fact that such simple recognition tools yield surprisingly high results suggests that there may be some unintended biases in the Avatar CAPTCHA dataset that the classifiers may be learning. These biases could probably be removed relatively easily, by for example applying grayscale intensity and contrast normalization so that the histograms and summary statistics of human and avatar images would be identical. Fig. 35 shows the most informative locations in the raw grayscale pixel features, and suggests

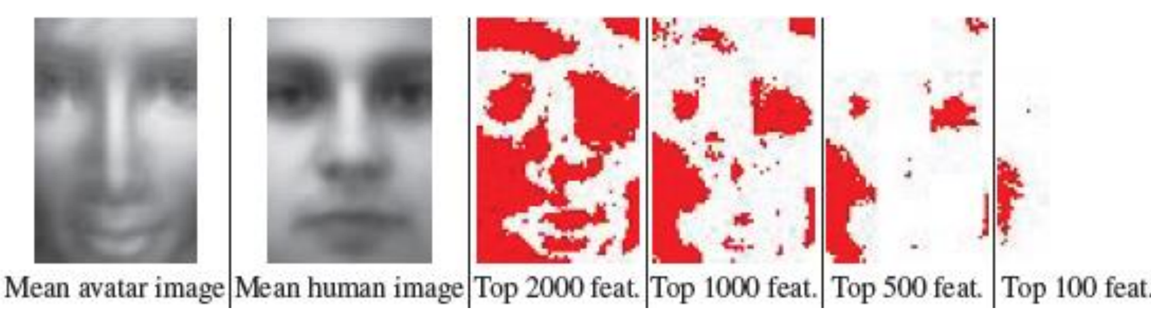


Figure 35: From left: Mean face images, and positions of top features according to information gain.

that the key differences between avatars and humans are in the cheek lines and around the eyes (M. Korayem et al., 2012; Mohammed Korayem et al., 2012).

I next tested more sophisticated techniques that may be much more difficult to guard against. Table 5 shows results for the more sophisticated features and classifiers that I tested. Each row of the table shows a different feature type, while the columns show results for classification using LibLinear, Naïve Bayes (NB), and Naïve Bayes with feature selection (NB+FS). Perfect recognition results (100% accuracy) are achieved by both the LibLinear classifier using raw pixel values, and the local difference pattern (LDP) descriptor using Naïve Bayes with feature selection. HOG features also produced excellent results (99% correct accuracy), while SURF and the local binary pattern variants all yielded accuracies above 95% for at least one of the classifiers. GIST and grayscale histogram features performed relatively poorly at around 90%, but this is still a vast improvement over the random baseline (50%). Fig. 36 presents ROC curves for the different classifiers and features (M. Korayem et al., 2012; Mohammed Korayem et al., 2012).

Table 5: Classification accuracy using different features and classifiers, with feature dimensionality in parentheses

Method	LibLinear	Naïve Bayes	Naïve Bayes + FS
Raw pixels	**100% (3750f)**	93% (3750f)	98% (54f)
Histogram	60% (256f)	**89% (256f)**	82% (24f)
GIST	84% (960f)	88% (960f)	**90% (24f)**
HOG	**99% (2268f)**	94% (2268f)	95% (44f)
FPLBP	94% (240f)	89% (240f)	**95% (26f)**
SURF	**97% (50f)**	96% (50f)	94% (22f)
LDP (absolute differences)	94% (256f)	99% (256f)	**100% (61f)**
LDP (differences)	96% (256f)	98% (256f)	**99% (75f)**
LBP	**98% (256f)**	95% (256f)	98% (31f)

6.5 Discussion and Conclusion

My experimental results indicate that the current Avatar CAPTCHA system is not very secure because relatively straightforward image recognition approaches are able to correctly classify between avatar and human facial images. For example, several of our classifiers achieve 99% accuracy on classifying a single image, which means that they would achieve $(0{:}99)^{12}$ = 88.6% accuracy on the 12-face classification CAPTCHA proposed in (D. D'Souza et al., 2012). This results is actually better than the human performance on this task (63%) reported in (D. D'Souza et al., 2012). Our classifiers work better than baseline even on surprisingly simple features, like summary statistics of an image. These results suggest that there may be substantial bias in the library of face images used in the current system, and that a new dataset without such biases would yield a much more secure system. Our work thus highlights the difficulty of creating image-based CAPTCHA systems that do not suffer from easily-exploitable biases, and how to prevent such biases (and ideally to prove that they do not exist) is a worthwhile direction for future work.

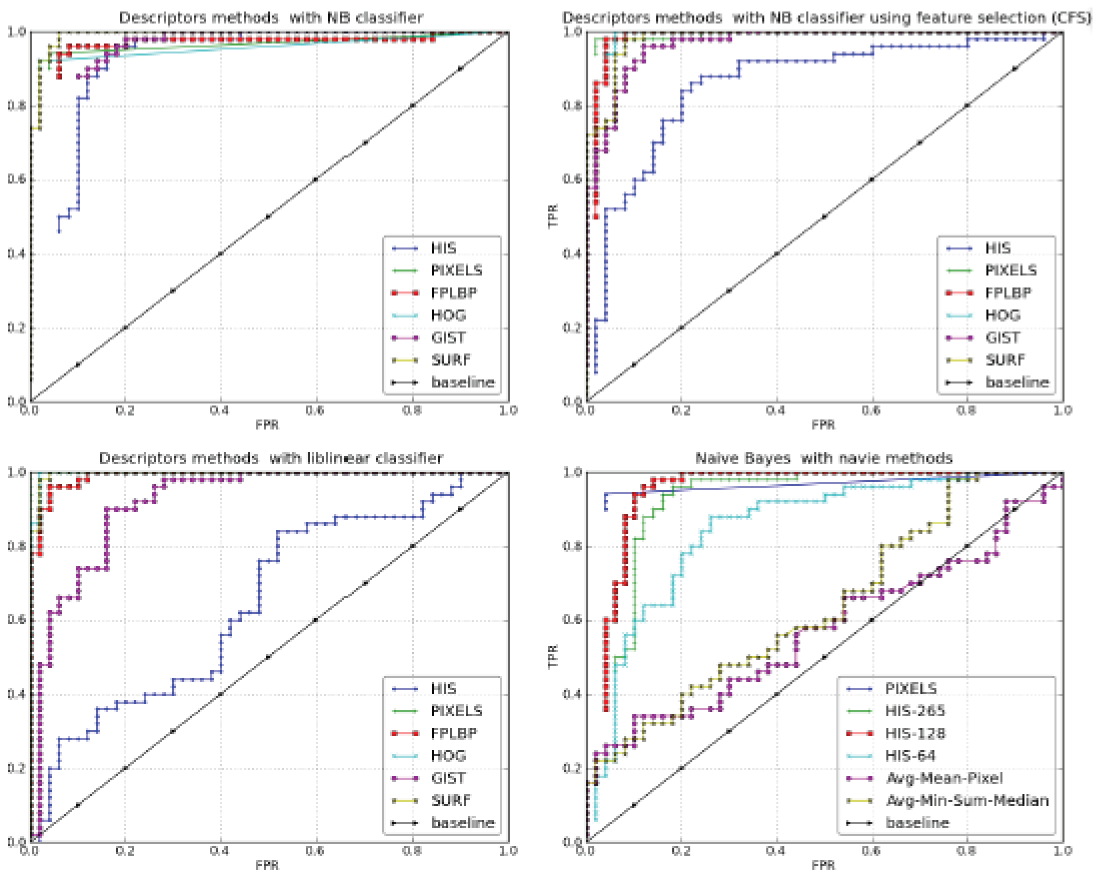


Figure 36: ROC curves for the human versus avatar classification task. Top left: Naïve Bayes classifiers, Top right: feature selection and Naïve Bayes, Bottom row: LibLinear classifiers.

CHAPTER 7

FACE RECOGNITION USING STATISTICAL ADAPTED TECHNIQUES

7.1 Introduction

The local binary pattern method is an effective operator for feature description and it has been applied in many applications. The original LBP system takes a local neighborhood around each pixel, thresholds the pixels in the neighborhood based on the central pixel gray value and uses the resulting binary representation as a local descriptor. Therefore, the original LBP descriptor has the following limitation in its applications: because the LBP methods threshold based on the central pixel value of a certain window around the central pixel, they are sensitive to noise especially in near-uniform regions of an image. The output value of LBP operator can be defined as (Ahonen et al., 2006; Ojala & Pietikäinen, 1996; Ojala et al., 2002):

$$LBP_{P,R} = \sum_{p=1}^{P=1} 2^{p} * S(g_p - g_c) \tag{21}$$

where R is the radius of the neighborhood, P is the number of pixels in that neighborhood, g_p is the value of the pixel p in the neighborhood, g_c is the value of the central pixel and S is the decision function that can be defined as follows (Ahonen et al., 2006; Ojala & Pietikäinen, 1996; Ojala et al., 2002):

$$S(g_p - g_c) = \begin{cases} 1 & g_p - g_c \geq 0 \\ 0 & g_p - g_c < 0 \end{cases} \tag{22}$$

Let us examine an image using a 3x3 window as in Fig 37. Based on the original LBP the corresponding binary value will be 11100101 or 229. If I change the intensity value of the central-up-right pixel from 24 to be 23 the LBP binary value will be 01100101 and hence the corresponding decimal representation will be changed from 229 to be 101 as in figures 37 and 38. Although these two binary bit representations are very similar since their Hamming distance is equal to 1 but their decimal representation is completely different. There are some attempts to change the concept of using the central pixel value as a threshold such as what Heikkila and Pietikainen did in (Heikkila & Pietikainen, 2006) and what Meng et al., followed in (Jun et al., 2010). They modified the threshold scheme presented in the original LBP definition and replaced the term $S(g_p - g_c)$ used in equation 40 to be $S(g_p - g_c + |a|)$ where a is a fixed value during the

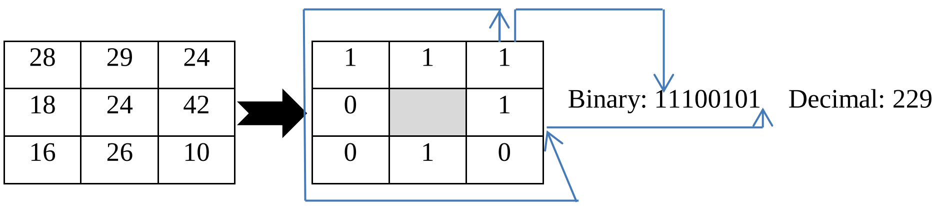


Figure 37: The Original LBP operator.

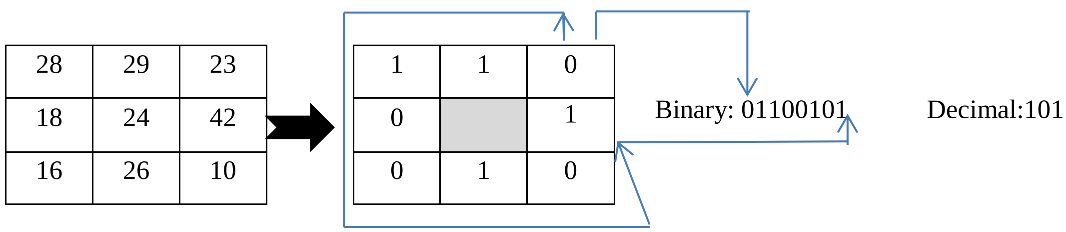


Figure 38: The Original LBP operator after modification.

whole image. If the value chosen for this value is a = 0 then the new definition for this LBP operator will be the same as the original one. The following is a new definition of the decision function used in equation 22 after using a specific threshold value (Heikkila & Pietikainen, 2006; Jun et al., 2010):

$$S(g_p - g_c) = \begin{cases} 1 & g_p - g_c \geq a \\ 0 & g_p - g_c < a \end{cases} \tag{23}$$

Let us examine the 3x3 window defined in Figure 29 using a threshold value a = 5. If the difference between the neighborhood and the central pixel is greater than or equal to 5 then the corresponding bit value is 1 otherwise it will be 0 (Figures 39 and 40 give more explanation for the result of using an example of a threshold with value 5). But what will happen if one pixel changes such as the upper central pixel if it is changed from 29 to be 28? The binary representation for the output will be nearly the same (hamming distance value is 1) while the decimal representation of the output will be different. Also applying one threshold value for the whole dataset is not an ideal way to obtain representative features.

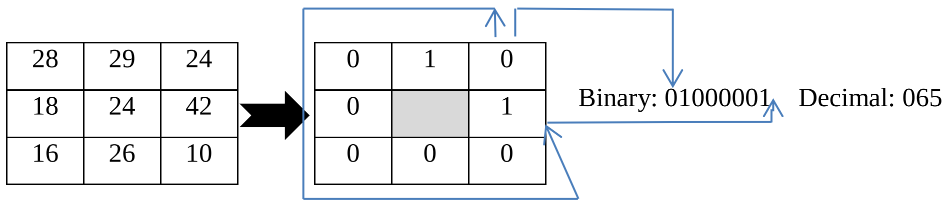


Figure 39: The Original LBP operator with threshold value a = 5.

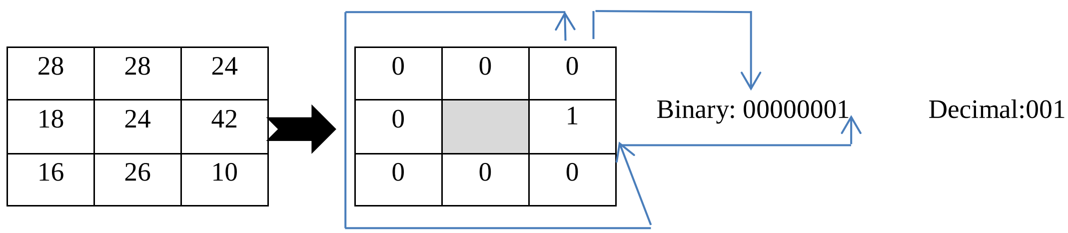


Figure 40: The Original LBP operator with threshold value a = 5 after changing the value of one pixel.

To deal with this problem, I suggested two options: First, by building a new threshold (statistical adaptive threshold). This threshold is not fixed for the whole image but it changes based on a combination of the local statistics of the neighborhood around a certain pixel and the local weight of each pixel in this neighborhood. Different new versions of the LBP descriptors, such as multi-scale statistical adaptive LBP and hierarchal multi-scale statistical adaptive LBP, can be obtained by applying this threshold. Second by extending the LBP to its new extension Local Ternary Pattern (LTP) (Akhloufi & Bendada, 2010; Bendada & Akhloufi, 2010; Shengcai et al., 2010; Xiaoyang & Triggs, 2010) and applying this new threshold on new different versions of LTP operator such as multi-scale statistical adaptive local ternary patterns and Hierarchal multi-scale statistical adaptive local binary patterns (I will explain both options later in this chapter).

Also the LBP and LTP operators suffer from another major problem (especially in case of LTP). Using a base-2 system, as in case of LBP, and a base-3 system as in case of LTP, for representing the feature patterns will increase the feature dimension. For example, the histogram size that is generated by LBP operator (16, 2) is 2^{16} = 65536 and the histogram size that is generated by LTP operator (16, 2) is 3^{16} =

43046721. Both histograms are not suitable for practical implementations, so during our explanation I will suggest how I can reduce or cut down the size of these histograms.

7.2 Statistical Adaptive Local Binary Pattern (SALBP)

Using a fixed threshold or using the central pixel value of the neighborhood of any pixel as a threshold has a negative effect on how the LBP method can deal with noise specially in near uniform or flat area. To avoid the LBP methods from being highly sensitive to noise I propose a novel LBP operator Statistical Adaptive Local Binary Pattern. All parameters in this operator are coming from image pixels themselves. To this end I have to define two parameters std_P and w_p. std_P is the standard deviation of all pixels in the neighborhood around a pixel (central pixel) plus this pixel itself while w_p represents the weight of any pixel p in that neighborhood according to the following equation (objective function) (Mohamed & Yampolskiy, 2012a, 2012e, 2013):

$$J = (g_c(i,j) - \sum_{q=1}^{P} w_q g_q(i,j))^2 \tag{24}$$

where g_c corresponds to the central pixel, g_q coressponds to the surrounding pixels and $W = [w_1, w_2, w_3, \ldots, w_P]$ and $\sum_{q=1}^{P} w_q = 1$. This equation minimizes the overall differences between the central pixel in any neighborhood and all pixels in that neighborhood. By deriving both sides of equation 24 with respect to w_p I get:

$$\frac{\partial J}{\partial w_p} = -2g_p(i,j)(g_c(i,j) - \sum_{q=1}^{P} w_q g_q(i,j)) = 0$$

Then:

$$g_c(i,j) = \sum_{\substack{q=1 \\ q \neq p}}^{P} w_q g_q(i,j) + w_p g_p(i,j) \tag{25}$$

From equation 25 I can obtain the value of w_p using the following equation:

$$w_p = \frac{g_c(i,j) - \sum_{\substack{q=1 \\ q \neq p}}^{P} w_q g_q(i,j)}{g_p(i,j)} \tag{26}$$

For more explanation about how I can compute w_p for different pixels in the neighborhood I have to follow the following steps (Mohamed & Yampolskiy, 2012e, 2013):

1- Initialization $w_p = \frac{1}{P} \ \forall p = 1,2, \ldots, P$

2- Use updated equation 26

Repeat

For $p = 1 : P$

Update w_p with the new value

end

By the end of applying these steps, I will have the weights for all pixels in the neighborhood and then I can define the SALBP operator as follows (Mohamed & Yampolskiy, 2012e):

$$SALBP = \sum_{p=0}^{P-1} s(g_p * w_p - k * std_P) 2^p \tag{27}$$

Therefore, the obtained binary code can be as:

$$s(g_p * w_p - k * std_P) = \begin{cases} 1, & if \quad g_p * w_p \geq k * std_P \\ 0, & otherwise \end{cases} \tag{28}$$

where k is a scaling factor such that o < $k \leq 1$. For more explanation about how I can deal with both the standard deviation and the weights of all pixels in the neighborhood to compute the new value for each pixel by using the definition of SALBP, consider the following example that I discussed previously in figure 37:

p_3=28	p_2=29	p_1=24
p_4=18	C = 24	p_8=42
p_5=16	p_6=26	p_7=10

Figure 41: Window of 8 pixels around the central pixel.

In the beginning I have to compute the standard deviation (*std*) of all pixels in the previous window including the central pixel itself. The *std* for all pixels value in the previous window is 8.595. The first step for computing the updated weight for each pixel in the neighborhood of the central pixel *C=24* is by giving each pixel in the neighborhood an initial weight equal to *1/P*, where *P* is the size of the neighborhood. Therefore, for the previous example each pixel in the neighborhood will be given an initial weight value equal to *1/8*. Using these weights values, I compute the updated weight value for pixel p_1 using equation 26. Multiply the updated new weight value for p_1 by the value p_1 itself and put the result in a new window. Now start computing the updated weight value for p_2 at this point all pixels from 3 to 8 has the same weight *1/P* but pixel p_1 has its updated weight value which will be used to compute the updated weight value for p_2. Keep repeating until I get all updated weight values and then the new pixel value which will be as in the following window:

p_3=3.5	p_2=3.625	p_1=2.875
p_4=2.25		p_8=5.25
p_5=2	p_6=3.25	p_7=1.25

Figure 42: New pixel values after updating.

In figure 42, I have the new value for each pixel in the neighborhood after multiplying the old value for each pixel by its updated weight value. By using the standard deviation of the old pixels values which is 8.595 and a factor k = 0.3 in equation 28 I can get the ASLBP value 11100101 which is the same value by using the original LBP operator. If I apply the same steps on the pixel values in figure 38, I will get the following new values for all pixels in the neighborhood:

p_3=3.5	p_2=3.625	p_1=2.875
p_4=2.25		p_8=5.25
p_5=2	p_6=3.25	p_7=1.25

Figure 43: New pixel values after updating data in figure 38.

By using the same factor k = 0.3 with the standard deviation of the old pixels values, which is 8.602, I get the same ASLBP value 11100101. Suppose I have more than one pixel corrupted by noise for the same image area coming from figure 38 and the new values in this window of data as in figure 44. I followed the same steps that I did in the previous two examples and I obtained the new pixels values that can be seen in figure 45. By using the multiplication of the factor $k = 0.3$ with the standard deviation (which is 8.367) of the old values of these new pixels as in figure 44 as a threshold I get the same ASLBP value 11100101. So with changing the value of pixel p_1 or p_1 and p_2, ASLBP gives the same value which is 11100101 but the value coming from applying the LBP is changing with the change of the values of p_1 and p_2 which means that ASLBP is more robust in nearly flat areas.

28	23	23
18	24	42
16	23	10

Figure 44: New pixel values after corrupting more than one pixel by noise for the data in figure 37.

3.5	2.875	4
2.25		5.25
2	2.875	1.25

Figure 45: New pixel values after using the updated weight values for data in figure 44.

7.3 Multi-scale Statistical Adaptive Local Binary Pattern (MSALBP) Histogram

The original LBP was working based on a single scale of radius R and neighborhood size P of evenly distributed sampling points in a circular shape neighborhood. Based on this I explained the SALBP operator in the previous section. If the sampling points of the neighborhood do not fall exactly on the pixels (center of the pixels) or their coordinates are not integers, bilinear interpolation will be used to express these sampling points (Ahonen et al., 2006; Mäenpää & Pietikäinen, 2003; Ojala & Pietikäinen, 1996; Ojala et al., 2002), therefore, the LBP operator can work with sampling points from different radii. Based on this idea, the SALBP histogram can work for different scales that I can call Multi-scale SALBP histogram. In general the multi-scale representation of an image can be achieved either by varying the size of the neighborhood of the LBP operator or by down-sampling the original image by interpolation or by using the low pass filter and then applying a fixed radius LBP operator (C. Chan et al., 2007; Chan, 2008).

Multi-scale SALBP representation of an image can be obtained by varying the radius of the sampling points and combining the resulted SALBP images. To explain how this can be done, let us suppose I have a single facial image. In the beginning this facial image has to be divided into non overlapping sub-regions say pt_1, pt_2,..pt_n. Apply the definition of SALBP operator on the first sub-region pt_1 with different scales. Applying each scale with SALBP will create a histogram for this region. The set of histograms created by applying the SALBP operators on the first region will provide regional information about that region. These histograms for the same region have to be concatenated into a single histogram to produce multiresolution information about that region (as in Fig. 46).

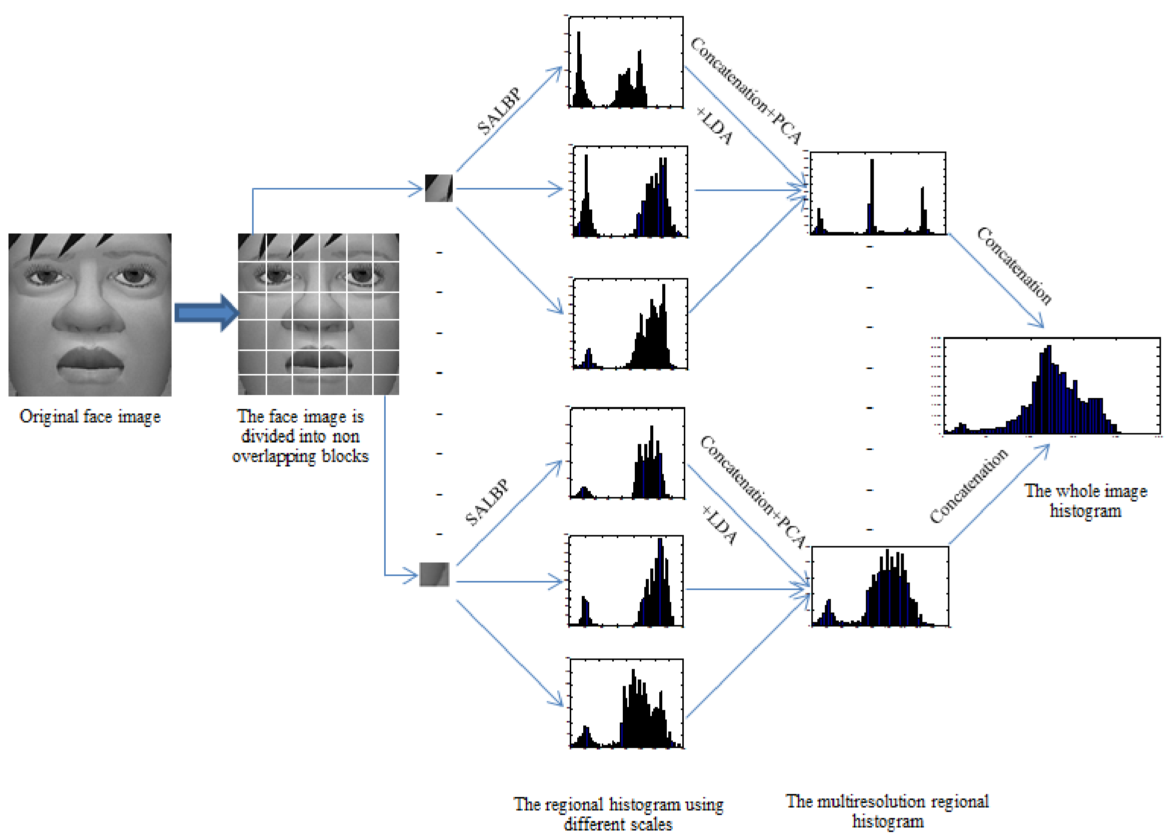


Figure 46: Proposed multi-scale SALBP based avatar facial representation.

However, the main problem associated with this technique is the high dimensionality of that histogram with the small training sample size. One of the best solutions for this problem is by combining Principal Component Analysis (PCA) with Linear Discriminant Analysis (LDA) (Belhumeur, Hespanha, & Kriegman, 1997; Jie, Yu, & Yang, 2001). Applying PCA will extract independent information and reduce the size of that histogram. The result of applying the PCA will be passed to LDA to extract discriminative facial features for each region. Projecting the reduced size histogram for one region on the LDA space will provide the regional discriminative facial descriptor for that region. Concatenating the regional discriminative facial descriptors for all regions will provide the global face description.

7.4 Wavelet Hierarchical Multi-scale Statistical Adaptive Local Binary Pattern (WHMSALBP)

WHMSALBP is another technique to build a multi-scale histogram for the whole image using the SALBP definition (see figure 47). WHMSALBP is the same as the MSALBP that I explained in the previous section except in two main points:

a- In our approach, I am combining the concept of discrete wavelet transform to obtain a new dataset of decomposed images. There are many different families of discrete wavelet transform and each one of them has different levels of decomposition. Practical results guided me to decide which wavelet family and level of decomposition that I have to use to decompose each dataset. For SL I used Db5 with the fourth level of decomposition, for ENT I used Db3 with the third level of decomposition and for ORL I used Db5 with the fourth level of decomposition.

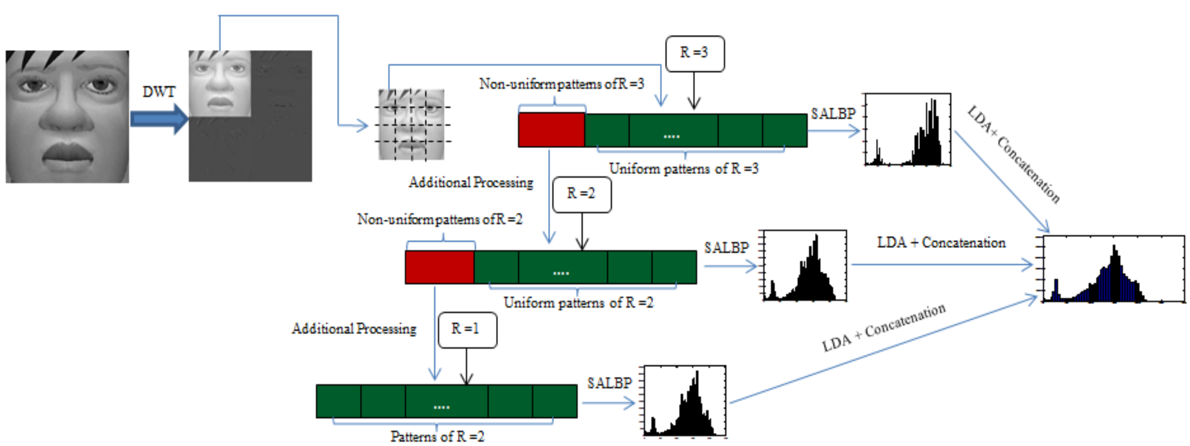


Figure 47: Schematic of the proposed WHMSALBP for one block of an approximation facial avatar image.

b- Instead of just applying the multi-scale definition on every local wavelet image patch, I have to apply this definition in a specific order. I start by applying the biggest neighborhood radius. The resulted non-uniform patterns will be extracted further using smaller radius. This process will continue until the smallest radius is processed (see figure 47).

7.5 Statistical Adapted LBP Techniques Results

To evaluate the performance of the proposed techniques I used two different groups of noisy images datasets. The first group of images belongs to virtual worlds. This group has two different virtual world images datasets. The first dataset belongs to Second Life (SL) virtual world and the second one belongs to Entropia Universe (ENT) virtual world. The second group of images has a noisy real human images dataset, ORL dataset.

7.5.1 SL Dataset Results

SL dataset has 581 images for 83subjects (avatars). Each subject has 7 different images for the same avatar with different frontal angle (front, far left, mid left, far right, mid right, top and bottom) and facial expressions. I corrupted SL dataset with two different types of noise, Gaussian noise and Salt & Pepper noise. Each noisy SL dataset was split into two independent datasets: one is used for training and the second is used for testing. During my experiments, I manually cropped the face area from each image to be 260 x 260 pixels and used three LBP operators, (8, 1), (16, 2) and (24, 3) with different sizes of the training and testing datasets. I started with one image from each subject for training and the rest for testing and continue increasing the number of training images up to 6 images from each subject. All training images are selected randomly.

7.5.1.1 Gaussian Noisy SL Dataset Results

Figures 48 and 49 show us that the performance of different techniques differs based on the LBP operator, each technique uses to perform, and based on the size of training and testing datasets. From figures 48 and 49 we can see that the highest recognition rate is obtained using WHMSALBP technique. This recognition rate is 80.42% and its closest traditional technique is WALBP with 71.71% accuracy rate with (24, 3) LBP operator (see figure 48 (c)).

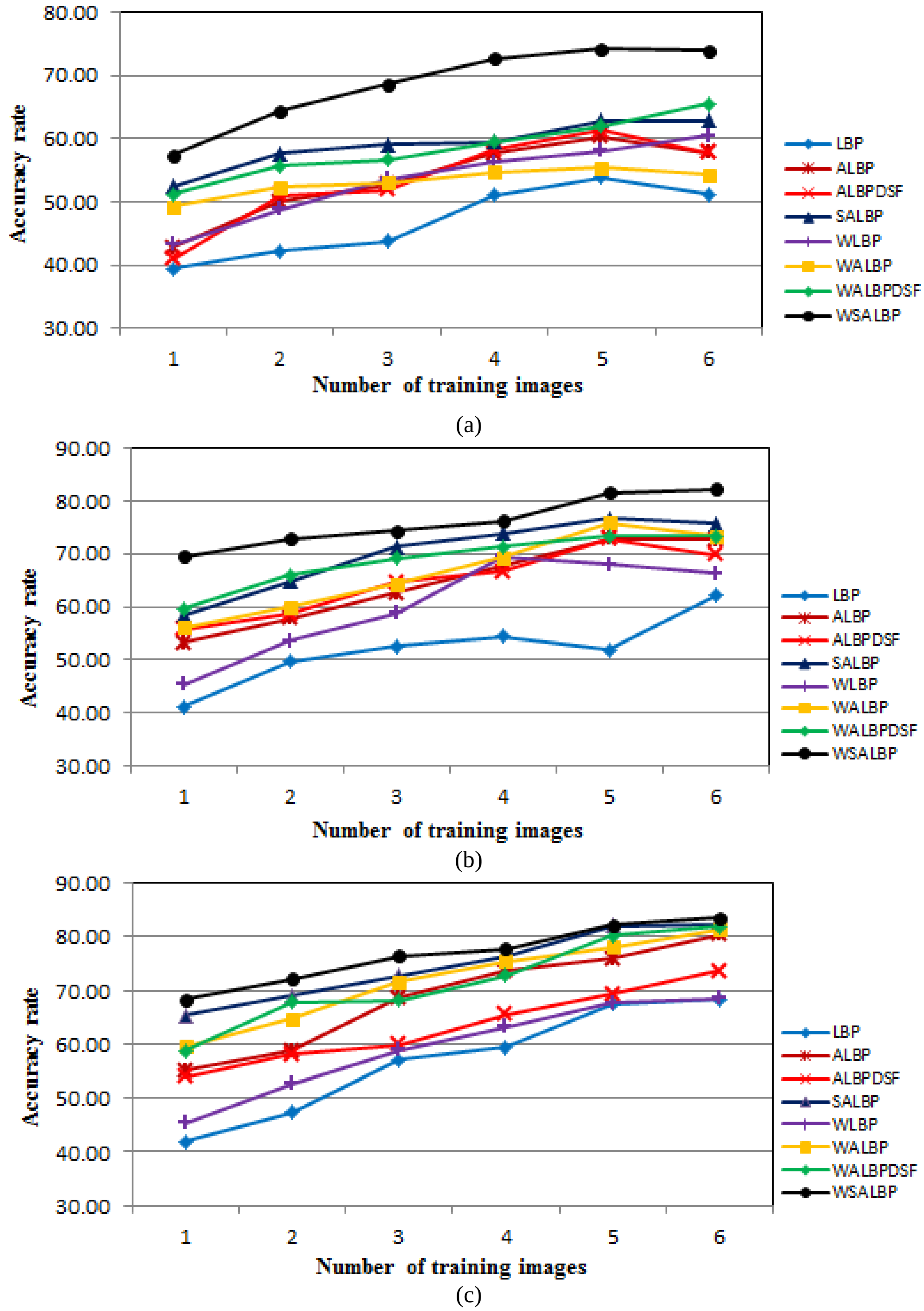


Figure 48: Accuracy rate for Gaussian noisy SL dataset for different LBP techniques with different LBP operators: (a) with (8, 1) operator (b) with (16, 2) operator (c) with (24, 3) operator.

However, in averaging accuracy rates with all LBP operators with all training images the closest traditional technique to WHMSALBP will be WALBPDSF with 66.24% accuracy rate as in Fig. 49.

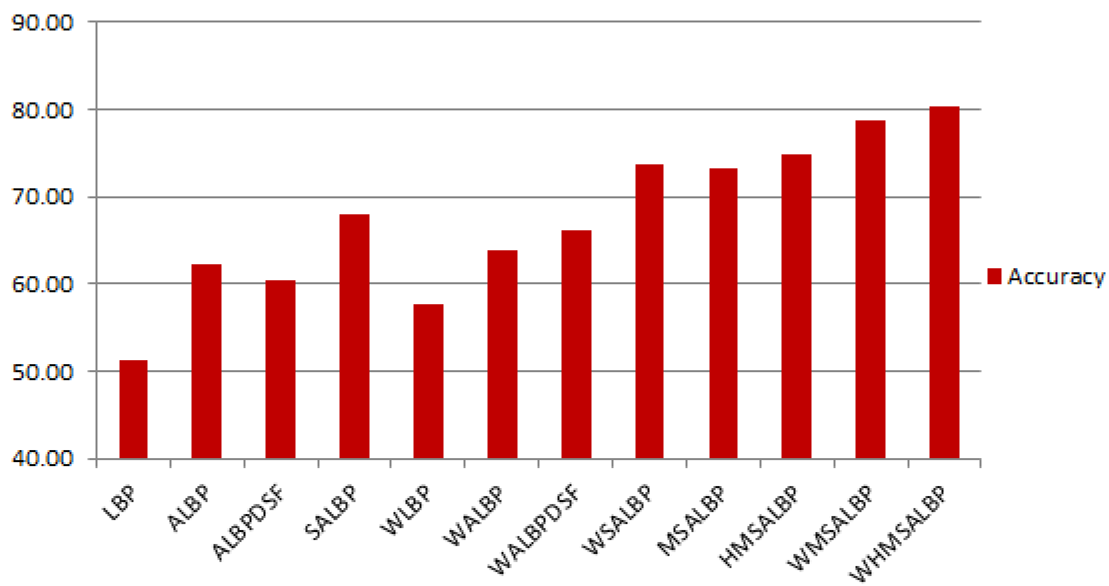


Figure 49: Average of accuracy rates for Gaussian noisy SL dataset with different LBP techniques.

7.5.1.2 Gaussian Noisy SL Dataset Results Evaluation

The highest recognition rate I obtained by WHMSALBP is 80.42% accuracy rate and the closest traditional technique, WLBP, provides 71.71% rate with almost 9% increase in the accuracy rate. However, it is not clear if this increase in accuracy rate is statistically significant or not. I have to follow two steps process: First, I have to check if my data is normally distributed or not. I use Minitab Software to plot my data (see Fig. 50). It is clear from figure 50 that my data is normally distributed. Second, I have to evaluate my data using statistical tests. During my experiments, I used Paired T-Test to evaluate my data. The main reason for using Paired T-Test is that I am concerning only with the difference in accuracy rate between WHMSALBP and WLBP and not concerning with the difference in accuracy rate for different number of training images. The result of applying Paired T-Test on my data as follows:

Paired T-Test and CI: WHMSALBP, WALBP

```
Paired T for WHMSALBP - WALBP

             N      Mean     StDev  SE Mean
WHMSALBP     6    0.8042    0.0717   0.0293
WALBP        6    0.7171    0.0829   0.0338
```

```
Difference  6  0.08703  0.01293  0.00528

95% CI for mean difference: (0.07346, 0.10060)
T-Test of mean difference = 0 (vs not = 0): T-Value = 16.49  P-Value = 0.000
```

The last row in the results obtained from applying the Paired T-Test states that P-value = 0.000 less than 0.05. With the assumption that the confidence level is 95.00%, I have to reject the Null hypothesis (H_0) of no difference in recognition rate between WALBP and WMHMSALBP. Therefore, I believe there is a significant difference in recognition rate between WHMSALBP and WALBP.

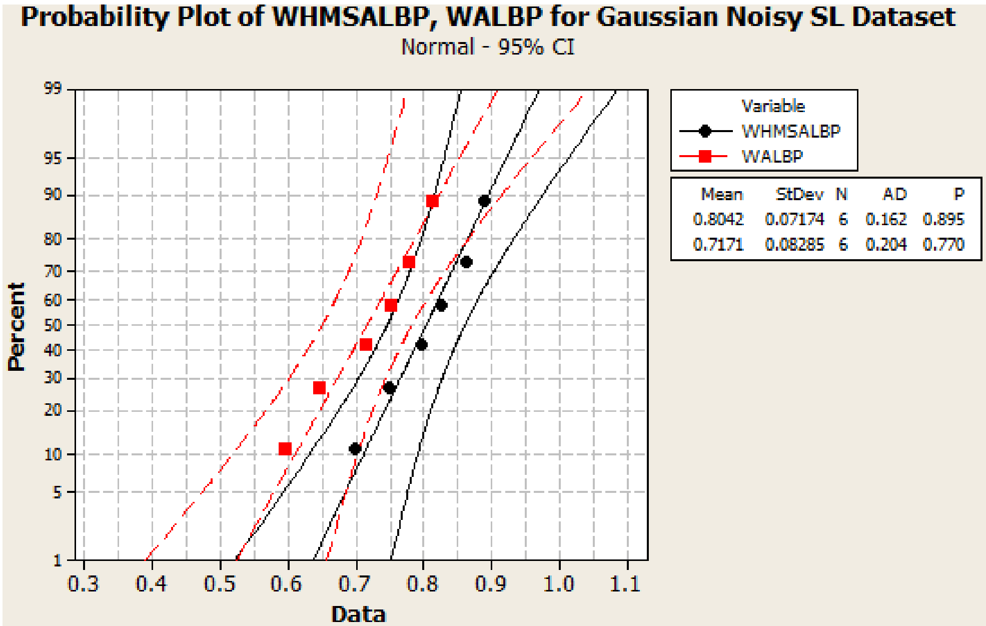
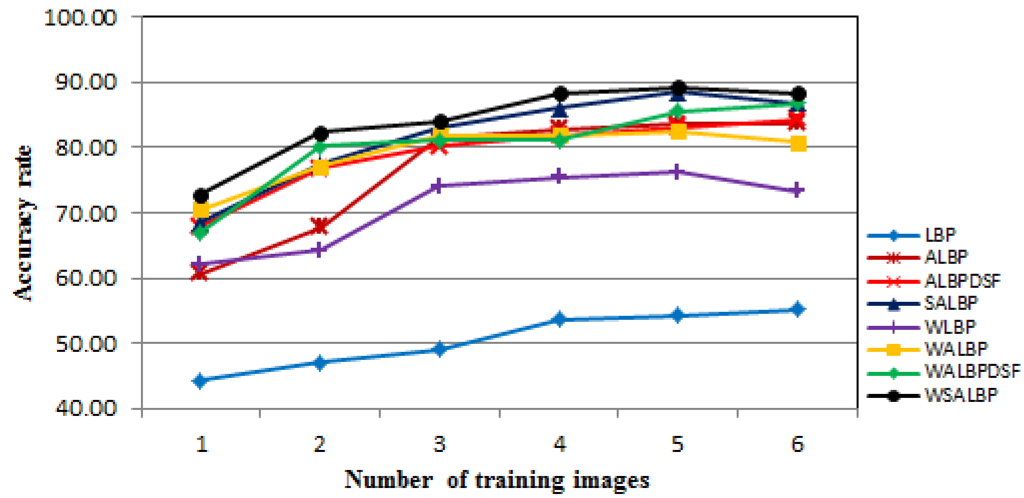


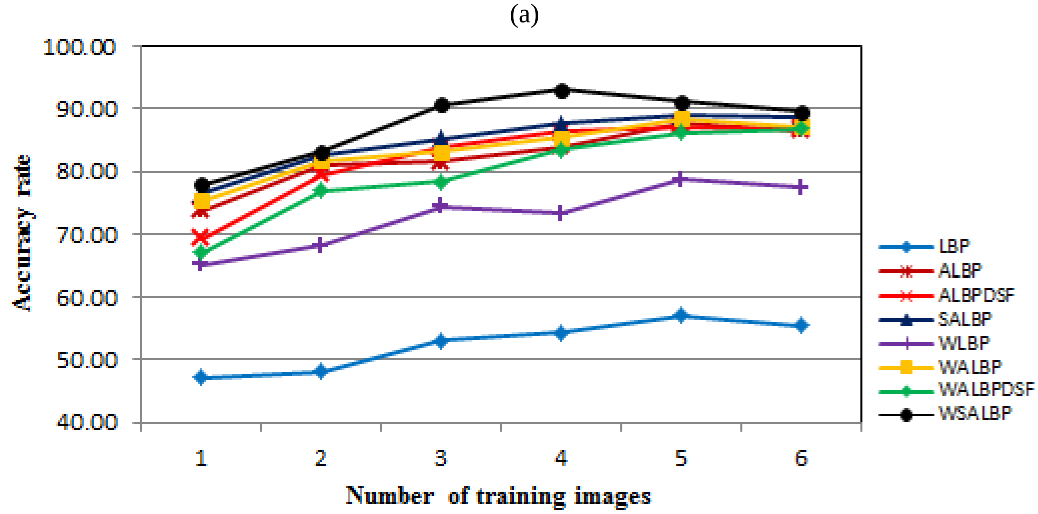
Figure 50: Distribution for WHMSALBP and WALBP data for Gaussian Noisy SL dataset.

7.5.1.3 Salt & Pepper Noisy SL Dataset Results

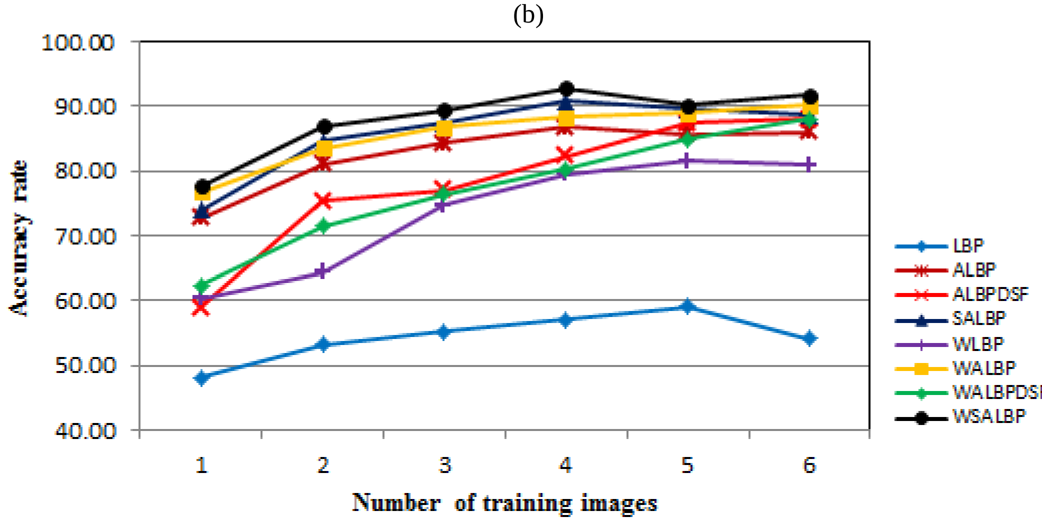
The following two figures, figures 51 and 52, show that the performance rate of WHMSALBP is better than all other techniques. This recognition rate is 89.50% and its closest traditional technique is WALBP with 85.77% accuracy rate with (24, 3) LBP operator (see figure 51 (c)). In addition, in averaging accuracy rates with all LBP operators with all training images, the closest traditional technique to WHMSALBP will be WALBP with 83.24% accuracy rate as in Fig. 52.



(a)



(b)



(c)

Figure 51: Accuracy rate for Salt & Pepper noisy SL dataset with different LBP operators: (a) with (8, 1) operator (b) with (16, 2) operator (c) with (24, 3) operator.

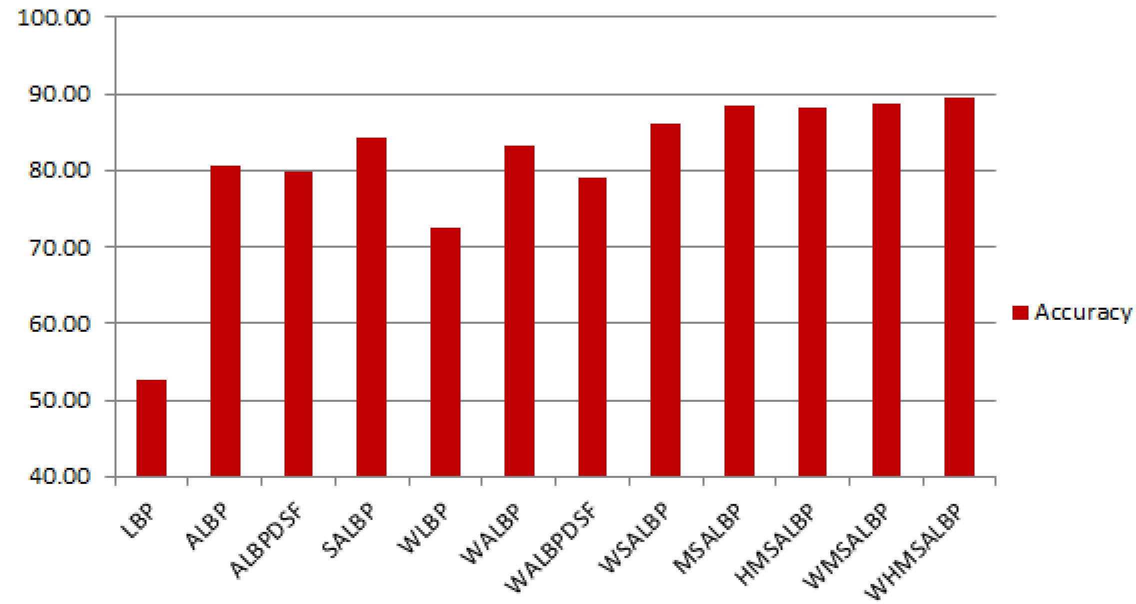


Figure 52: Average of accuracy rates for Salt & Pepper noisy SL dataset with different LBP techniques.

7.5.1.4 Salt & Pepper Noisy SL Dataset Results Evaluation

The difference in recognition rate between WHMSALBP with its closest traditional technique, WALBP, is almost 4%. To evaluate this increase in recognition rate I have to follow two steps process: First, I have to check the distribution of my data since the statistical test that I will use to evaluate my data works only on normally distributed data. Figure 53 shows that my data is nearly normally distributed. Second, I tested my data using a statistical test, Paired T-Test. The reason for using Paired T-Test and no other statistical tests is that I want to block some data and use the other to check the significance of my data. I want to check the difference in recognition rate obtained from applying WHMSALBP and that obtained from applying WALBP. Is it statistically significant or not? At the same time I am not concerning of the difference in recognition rate using different number of training images. The result of applying Paired T-Test on my data is as follows:

Paired T-Test and CI: WHMSALBP, WALBP

```
Paired T for WHMSALBP - WALBP

              N     Mean    StDev  SE Mean
WHMSALBP      6   0.8950   0.0642   0.0262
WALBP         6   0.8577   0.0499   0.0204
Difference    6  0.03730  0.02372  0.00968
```

```
95% CI for mean difference: (0.01241, 0.06219)
T-Test of mean difference = 0 (vs not = 0): T-Value = 3.85  P-Value = 0.012
```

The last row in the results obtained from applying the Paired T-Test states that P-value = 0.012 less than 0.05. With the assumption that the confidence level is 95.00%, I have to reject the Null hypothesis (H_0) of no difference in recognition rate between WALBP and WMHMSALBP. Therefore, I believe there is a significant difference in recognition rate between WHMSALBP and WALBP.

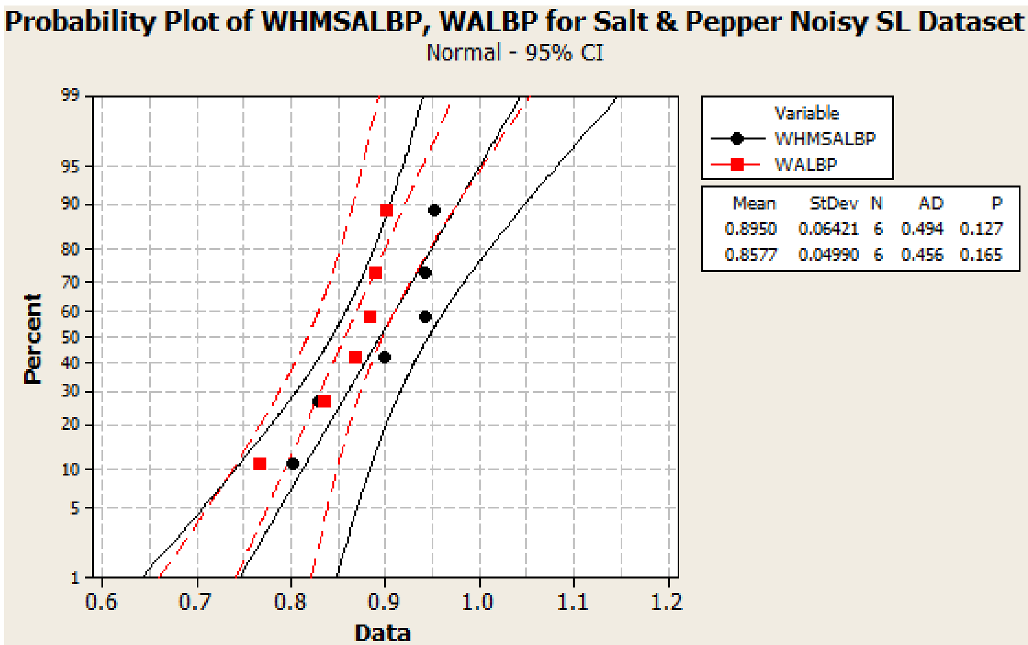


Figure 53: Distribution for WHMSALBP and WALBP data for Salt & Pepper Noisy SL dataset.

7.5.2 ENT Dataset Results

ENT dataset has 490 images for 98 subjects (avatars). Each subject has 5 different images for the same avatar with different frontal angle and facial expressions. I corrupted ENT dataset with two different types of noise, Gaussian noise (see figure 54) and Salt & Pepper noise (see figure 57). Each noisy ENT dataset has split into two independent datasets: one is used for training and the second is used for testing. For my experiments, I manually cropped the face area from each image to be in size 180 x 180 pixels and used three LBP operators, (8, 1), (16, 2) and (24, 3) with different sizes of the training and testing datasets. I started with one image from each subject for training and the rest for testing and continue increasing the number of training images up to 4 images from each subject. All training images are selected randomly.

7.5.2.1 Gaussian Noisy ENT Dataset Results

From figures 54 and 55, I can see that the performance of different techniques changes based on, the LBP operator used to extract an image features and the size of each training and testing dataset.

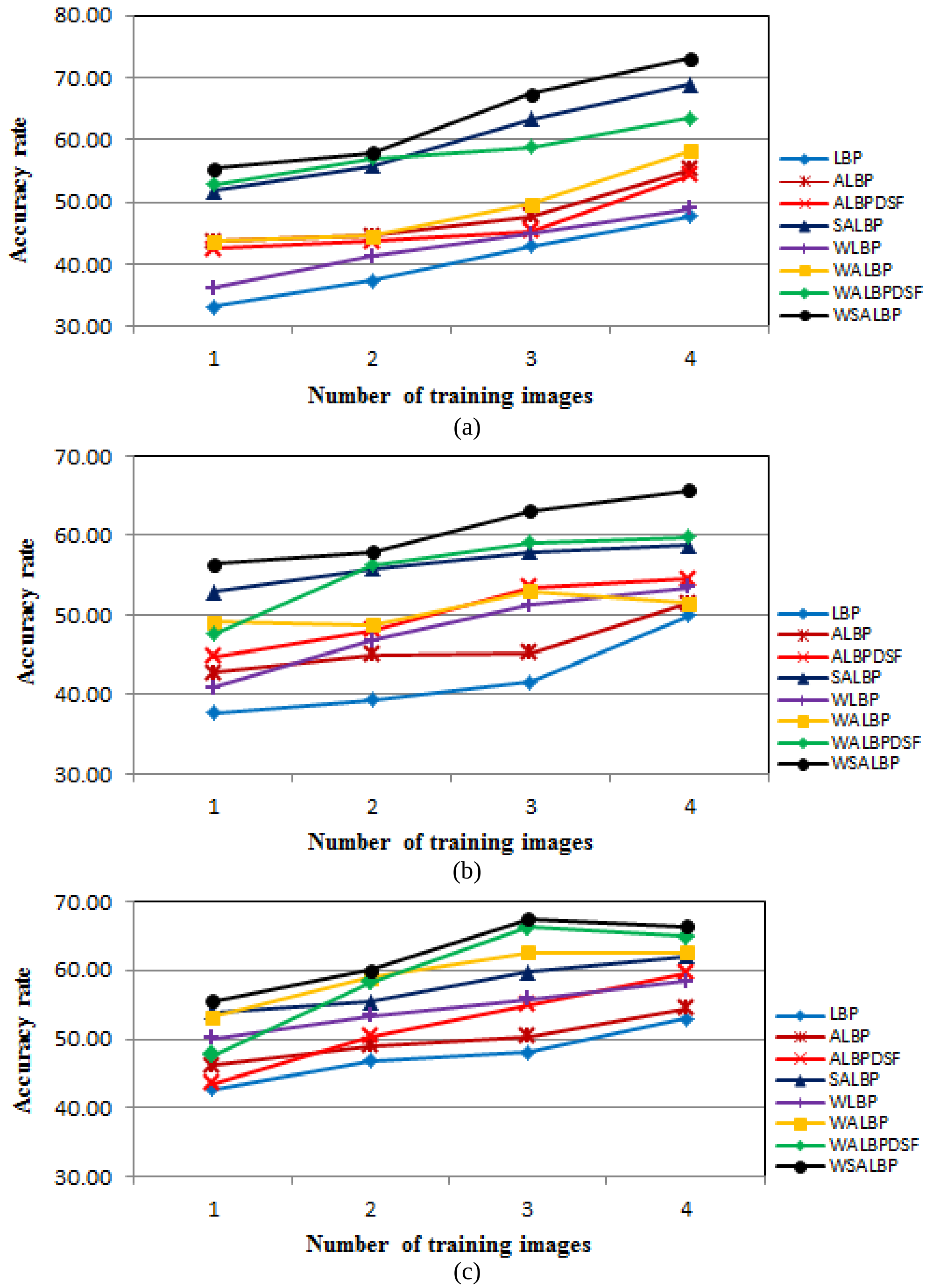


Figure 54: Accuracy rate for Gaussian noisy ENT dataset with different LBP operators:
(a) with (8, 1) operator (b) with (16, 2) operator (c) with (24, 3) operator.

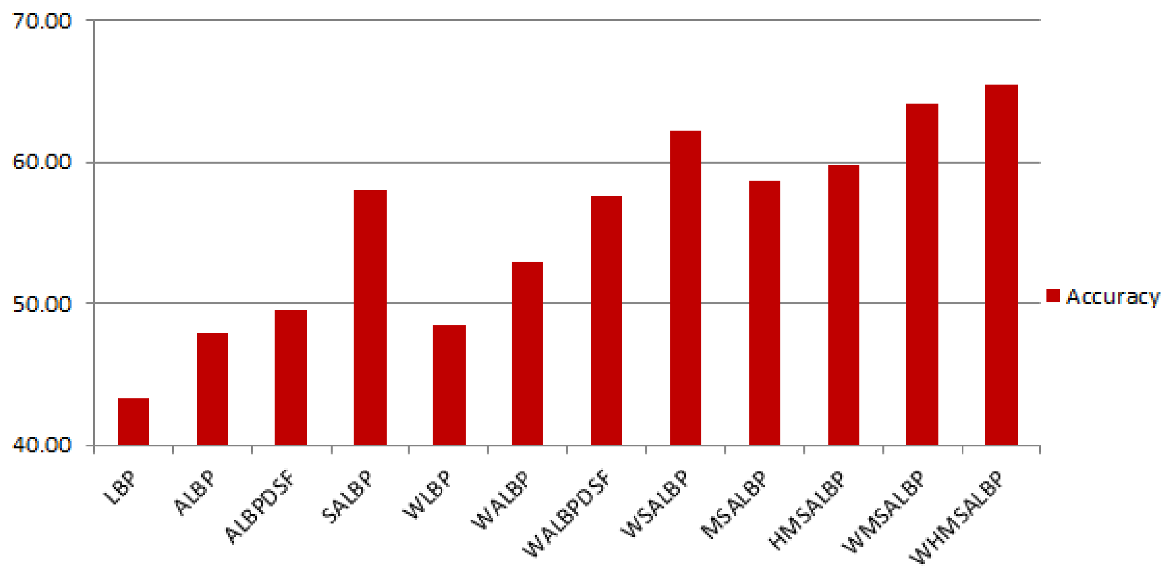


Figure 55: Average of accuracy rates for Gaussian noisy ENT dataset with different LBP techniques.

Increasing the radius of LBP operator and the number of its pixels may lead to increase the accuracy rate. In addition, increasing the number of the training images may lead to increase the accuracy rate.

The highest recognition rate can be obtained after applying WHMSALBP with 65.43% accuracy rate while the highest recognition rate for traditional techniques can be obtained after applying WALBP for (24, 3) LBP operator with 59.37% accuracy rate. However, in averaging accuracy rates with all LBP operators with all training images, the closest traditional technique to WHMSALBP will be WALBPDSF with 57.65% accuracy rate as in Fig. 55.

7.5.2.2 Gaussian Noisy ENT Dataset Results Evaluation

The difference in accuracy rate between the highest technique, WHMSALBP, and the highest traditional technique, WALBP, is almost 6%. To evaluate this difference in accuracy rate I have to follow two steps process: First check the distribution of the data that leads to this difference. To satisfy this target I used Minitab to plot this distribution. Figure 56 shows that the data is nearly normally distributed since almost all data points on or close to the straight line of a normal probability plot. Since my data is normally distributed, I can evaluate the significance of its difference using statistical tests that works on normally distributed data. I checked my data using one statistical test, Paired T-Test and the result as follows:

Paired T-Test and CI: WHMSALBP, WALBP

```
Paired T for WHMSALBP - WALBP

              N      Mean     StDev  SE Mean
WHMSALBP      4    0.6543    0.0399   0.0200
WALBP         4    0.5937    0.0443   0.0221
Difference    4   0.06058   0.00812  0.00406


95% CI for mean difference: (0.04766, 0.07349)
T-Test of mean difference = 0 (vs not = 0): T-Value = 14.92  P-Value = 0.001
```

The last row in the Paired T –Test result shows that P-value = 0.001 less than 0.05. With the assumption that the confidence level is 95.00%, I have to reject the Null hypothesis (H_0) of no difference in recognition rate between WALBP and WMHMSALBP. Therefore, I believe there is a significant difference in recognition rate between WHMSALBP and WALBP.

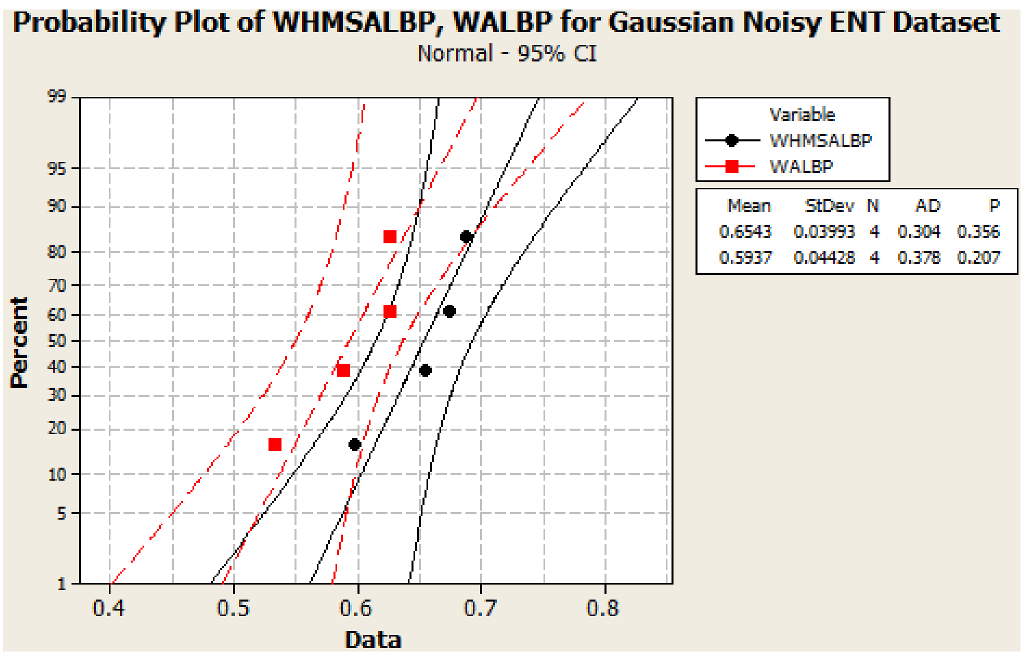


Figure 56: Distribution for WHMSALBP and WALBP data for Gaussian Noisy ENT dataset.

7.5.2.3 Salt & Pepper Noisy ENT Dataset Results

Figure 57 and 58 show that the performance of statistical adapted techniques is better than that of traditional techniques for all LBP operators and with almost all number of training images. The highest recognition rate can be obtained after applying WHMSALBP with 75.51% accuracy rate while the highest

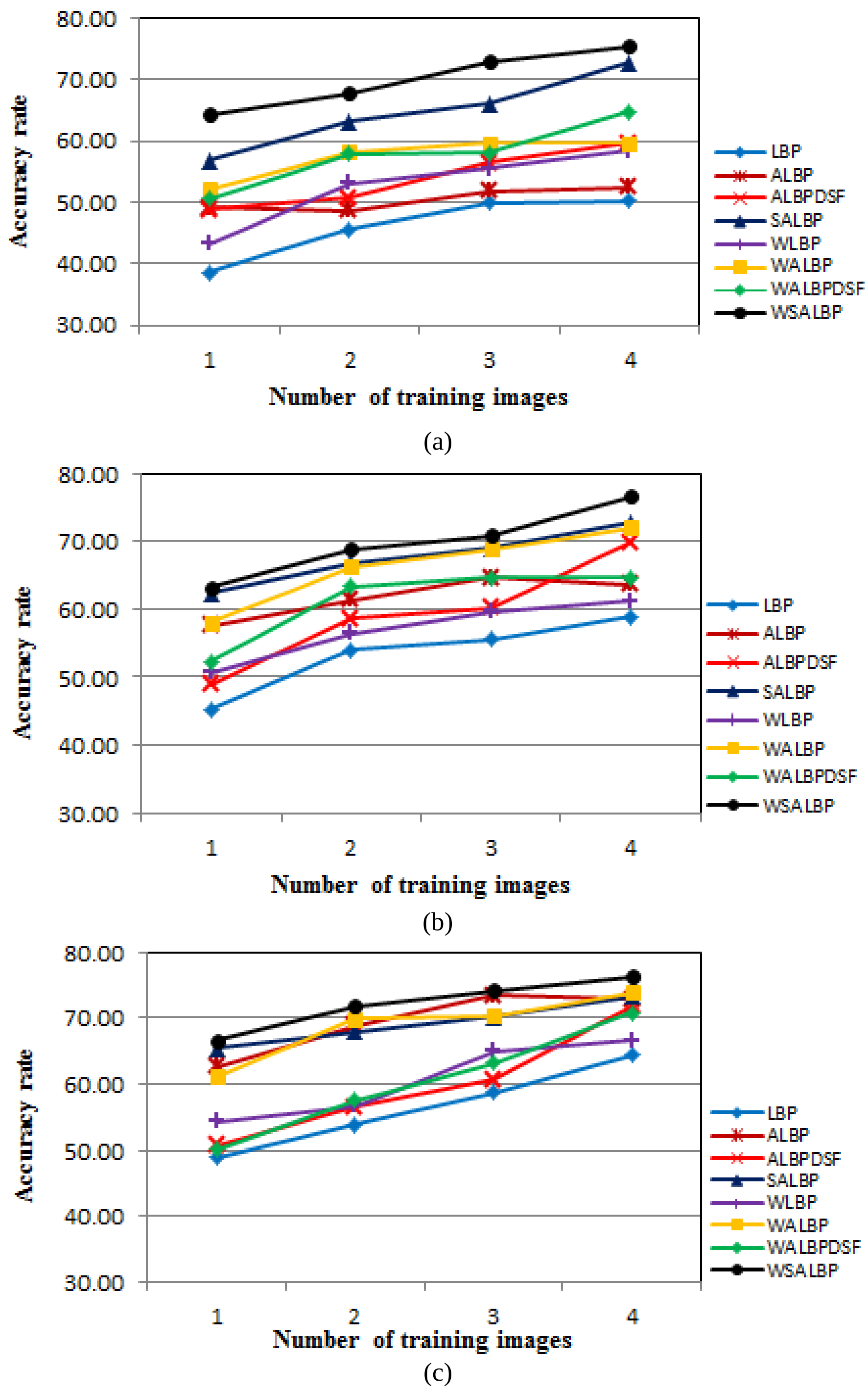


Figure 57: Accuracy rate for Salt & Pepper noisy ENT dataset with different LBP operators: (a) with (8, 1) operator (b) with (16, 2) operator (c) with (24, 3) operator.

recognition rate for traditional techniques can be obtained after applying WALBP for (24, 3) LBP operator with 68.80% accuracy rate. In addition, in averaging accuracy rates with all LBP operators with all training images, the closest traditional technique to WHMSALBP will be WALBP with 64.15% accuracy rate as in Fig. 58.

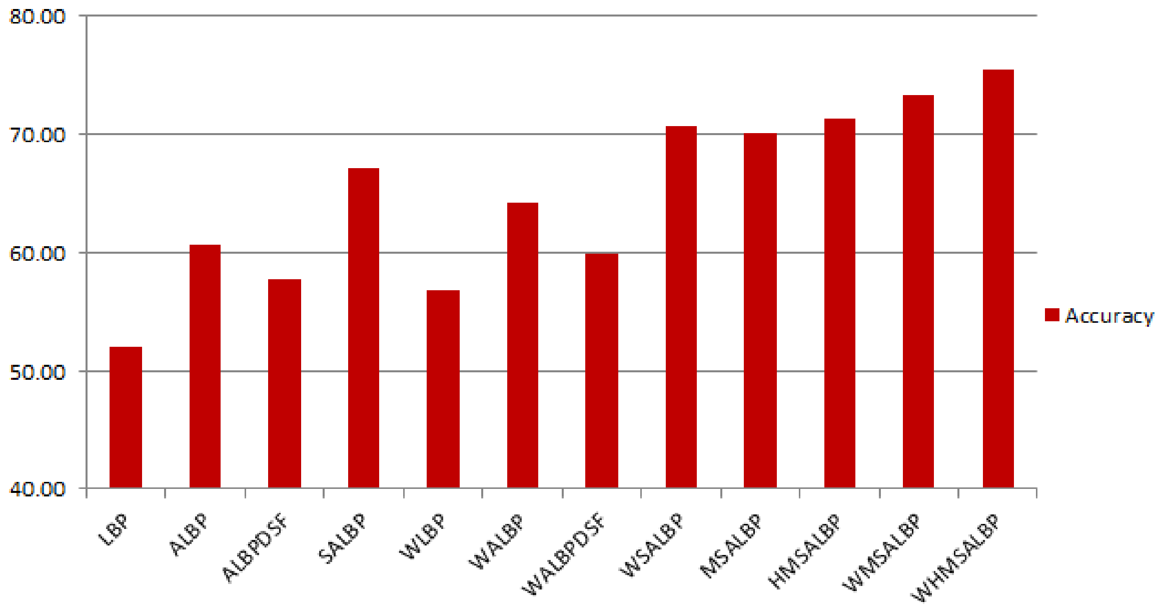


Figure 58: Average of accuracy rates for Salt & Pepper noisy ENT dataset with different LBP techniques.

7.5.2.4 Salt & Pepper Noisy ENT Dataset Results Evaluation

Recognition rate obtained after applying WHMSALBP is higher than that of its traditional competitor technique, WALBP, by about 7%. Figure 59 show how the data obtained from applying both WALBP and WHMSALBP plotted in a probability plot. It is clear from figure 59 that my data is normally distributed and so I can use statistical tests to evaluate the performance of my data. The following results are obtained after applying a statistical test, Paired T-Test, on my data:

Paired T-Test and CI: WHMSALBP, WALBP

```
Paired T for WHMSALBP - WALBP

             N     Mean    StDev  SE Mean
WHMSALBP     4   0.7551   0.0427   0.0214
WALBP        4   0.6880   0.0543   0.0272
Difference   4  0.06710  0.01922  0.00961


95% CI for mean difference: (0.03651, 0.09769)
T-Test of mean difference = 0 (vs not = 0): T-Value = 6.98  P-Value = 0.006
```

The last row in the Paired T –Test result shows that P-value = 0.006 less than 0.05. With the assumption that the confidence level is 95.00%, I have to reject the Null hypothesis (H_0) of no difference in recognition rate between WALBP and WHMSALBP. Therefore, I believe there is a significant difference in recognition rate between WHMSALBP and WALBP.

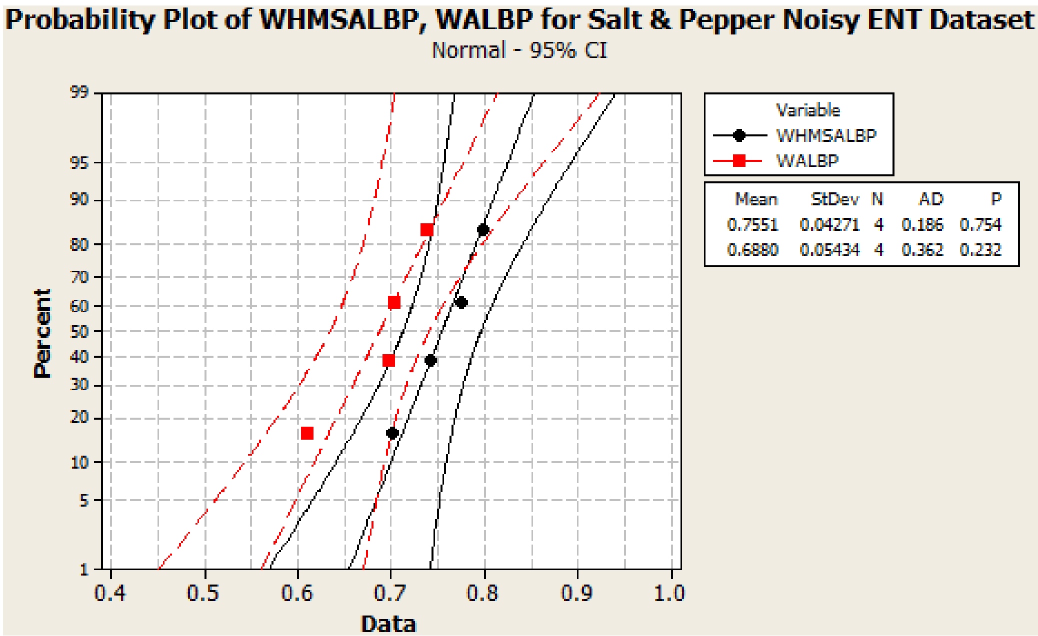


Figure 59: Distribution for WHMSALBP and WALBP data for Salt & Pepper Noisy ENT dataset.

7.5.3 ORL Dataset Results

ORL dataset has 400 images for 40 distinct subjects (humans). Each subject has 10 different images for the same person taken in different conditions. I corrupted SL dataset with two different types of noise, Gaussian noise and Salt & Pepper noise. Each noisy ORL dataset has split into two independent datasets: one is used for training and the second is used for testing. During all my experiments, I used the whole images without cropping the face area from each image. Therefore, each image size is 112 x 92 and I performed my experiments using three LBP operators, (8, 1), (16, 2) and (24, 3) with different sizes of the training and testing datasets. I started with one image from each subject for training and the rest for testing and continue increasing the number of training images up to 9 images from each subject. All training images are selected randomly.

7.5.3.1 Gaussian Noisy ORL dataset Results

Figures 60 and 61 show the results obtained after applying different techniques on ORL dataset. These results show that the performance of statistical techniques, such as WMSALBP and WSALBP, is better than that of traditional techniques with all LBP operators and with almost all sizes of training sets.

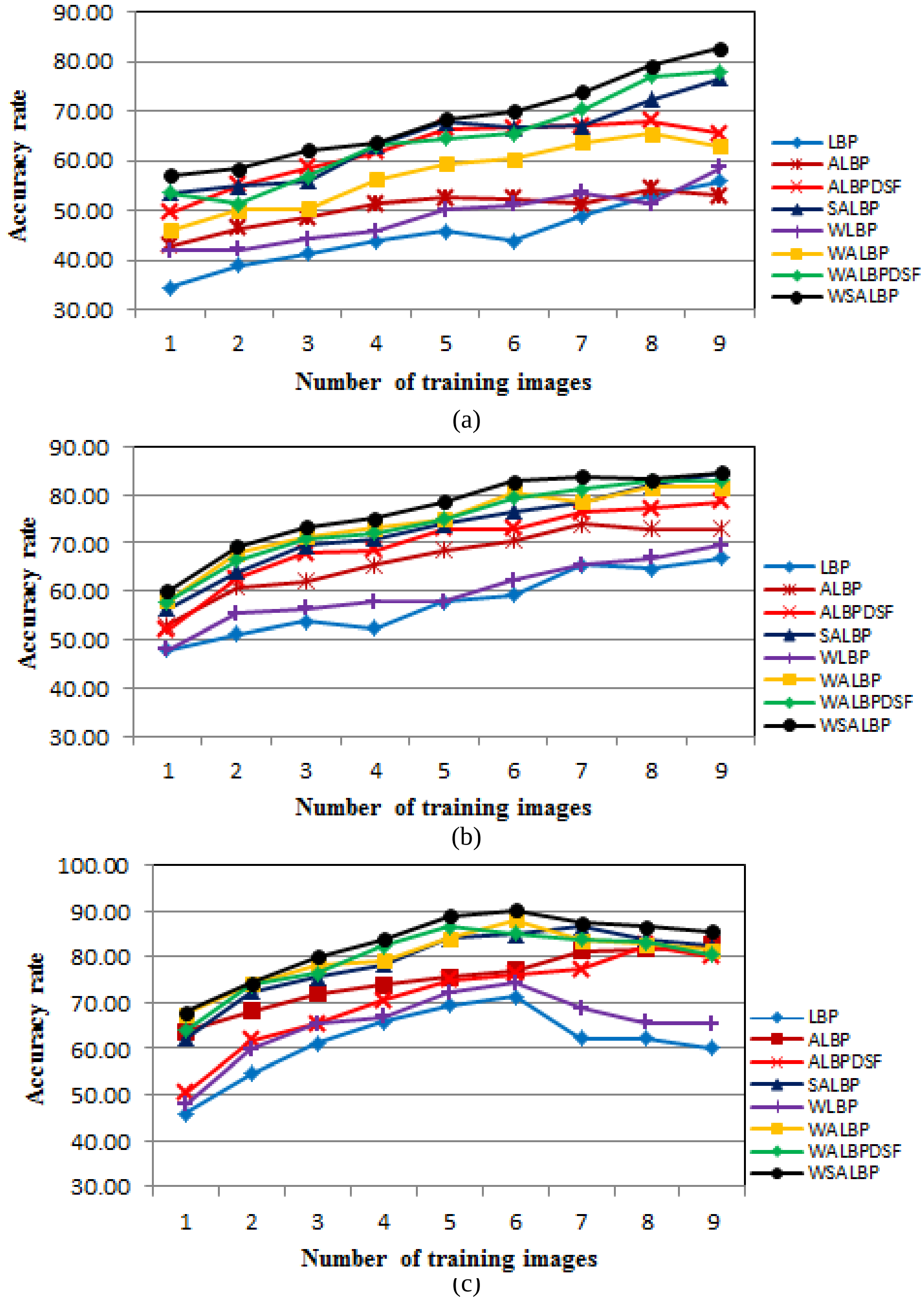


Figure 60: Accuracy rate for Gaussian noisy ORL dataset with different LBP operators:
(a) with (8, 1) operator (b) with (16, 2) operator (c) with (24, 3) operator.

Applying WMSALBP can provide the highest recognition rate, 84.35%, while the best recognition rate can traditional technique provide is 79.89% for WALBP with (24, 3) LBP operator. However, in averaging accuracy rates with all LBP operators with all training images, the closest traditional technique to WMSALBP will be WALBPDSF with 72.77% accuracy rate as in Fig. 60.

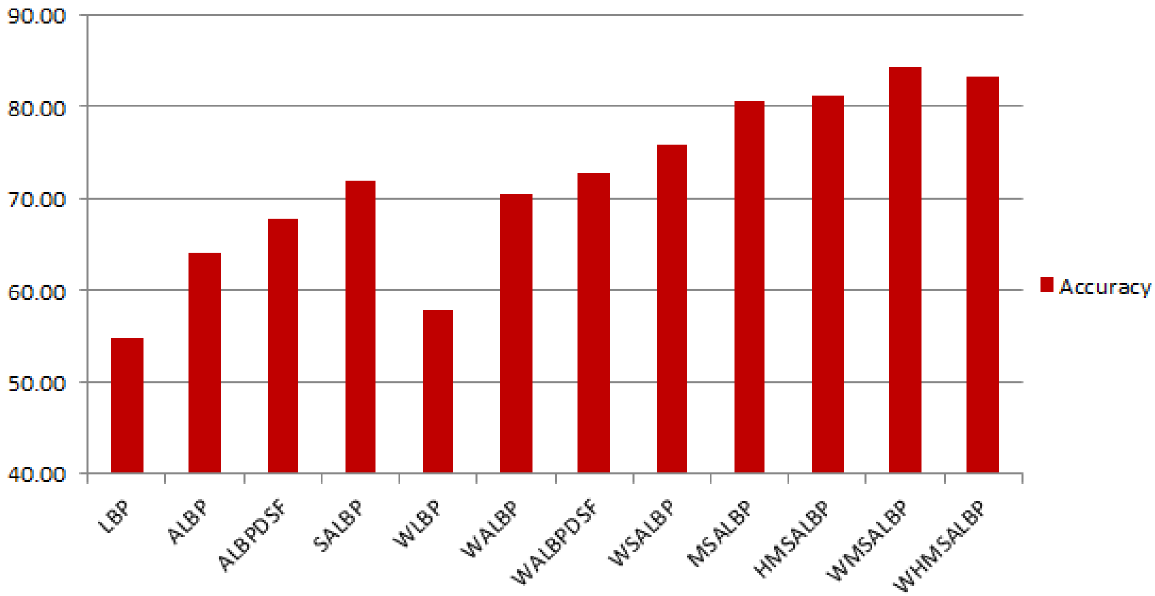


Figure 61: Average of accuracy rates for Gaussian noisy ORL dataset with different LBP techniques.

7.5.3.2 Gaussian Noisy ORL dataset Results Evaluation

To evaluate the performance of WMSALBP against its closest traditional competitor technique, WALBP, I have to test my data using statistical tests such as T-Test and Paired T-Test. First, I have to be sure that my data is normally distributed. Figure 62 shows that my data is normally distributed. Therefore, I can perform statistical tests on my data. The following is the result obtained after applying Paired T-Test on my data:

Paired T-Test and CI: WALBP, WMSALBP

```
Paired T for WALBP - WMSALBP

             N       Mean    StDev  SE Mean
WALBP        9     0.7989   0.0610   0.0203
WMSALBP      9     0.8435   0.0783   0.0261
Difference   9   -0.04461  0.02659  0.00886


95% CI for mean difference: (-0.06505, -0.02417)
T-Test of mean difference = 0 (vs not = 0): T-Value = -5.03  P-Value = 0.001
```

The last row in the Paired T –Test result shows that P-value = 0.001 less than 0.05. With the assumption that the confidence level is 95.00%, I have to reject the Null hypothesis (H_0) of no difference in recognition rate between WALBP and WMSALBP. Therefore, I believe there is a significant difference in recognition rate between WMSALBP and WALBP.

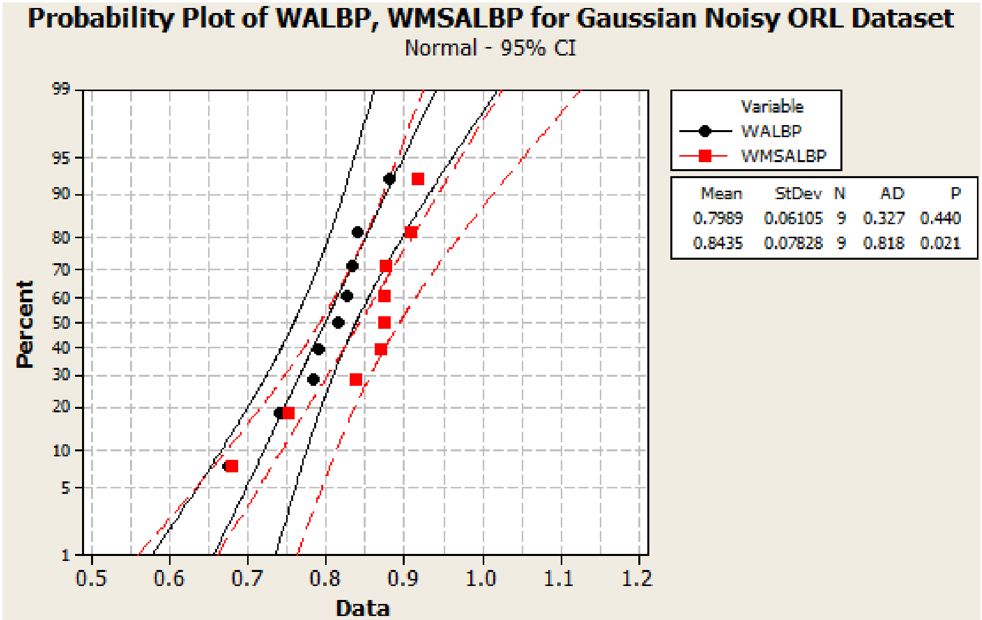


Figure 62: Distribution for WMSALBP and WALBP data for Gaussian Noisy ORL dataset.

7.5.3.3 Salt & Pepper Noisy ORL dataset Results

The following two figures, figures 63 and 64, show the results obtained after applying different LBP versions on Salt & Pepper noisy ORL dataset of face images. These results show that the performance of my proposed statistical adapted versions of LBP is better than that of traditional ones.

The best recognition rate, 86.67%, can be obtained from applying WMSALBP while the highest recognition rate obtained from a traditional technique was 79.21% after applying WALBPDSF with (24, 3) LBP operator. In addition, in averaging accuracy rates with all LBP operators with all training images, the closest traditional technique to WMSALBP will be WALBPDSF with 74.20% accuracy rate as in Fig. 63.

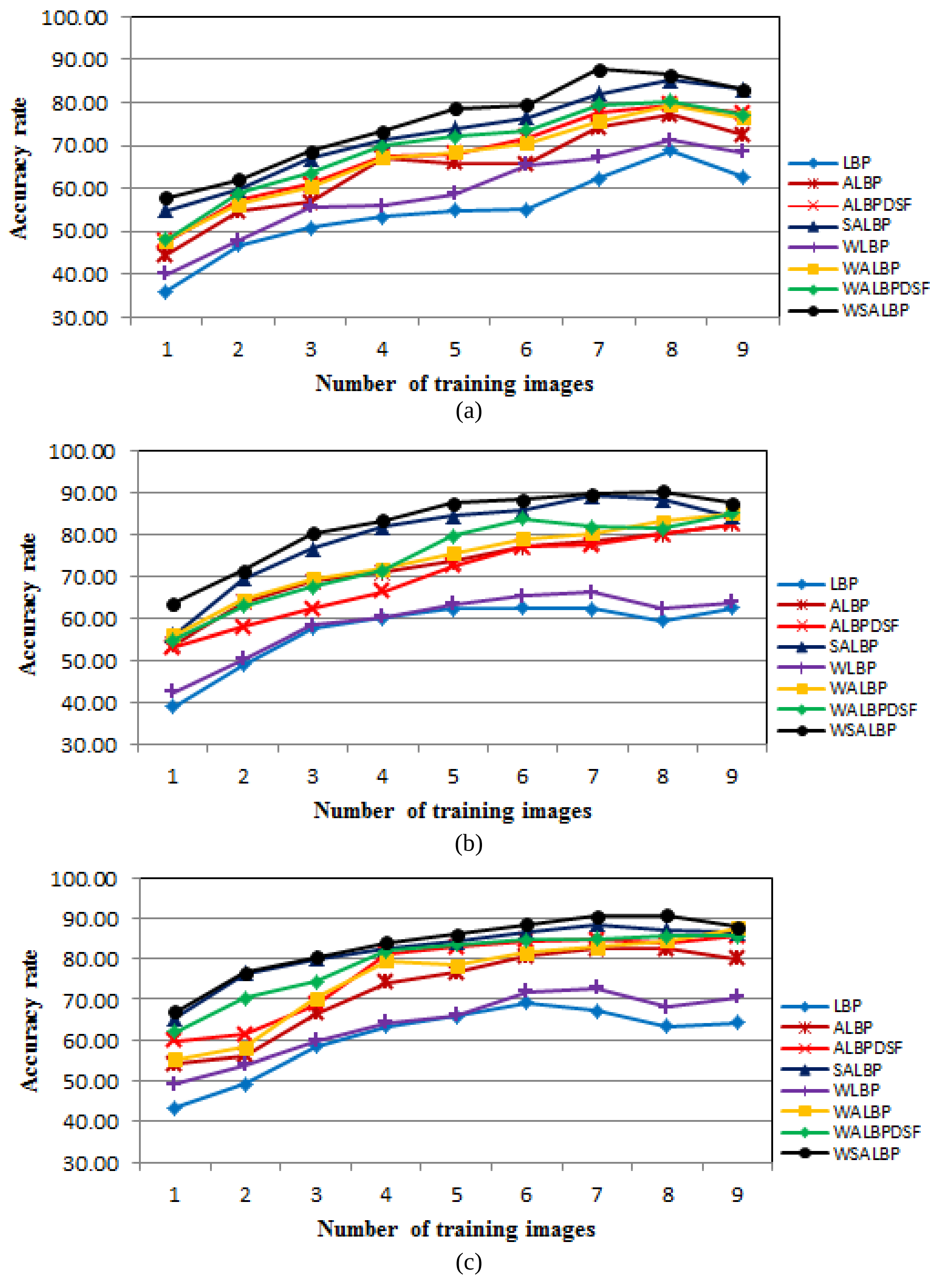


Figure 63: Accuracy rate for Salt & Pepper noisy ORL dataset with different LBP operators:
(a) with (8, 1) operator (b) with (16, 2) operator (c) with (24, 3) operator.

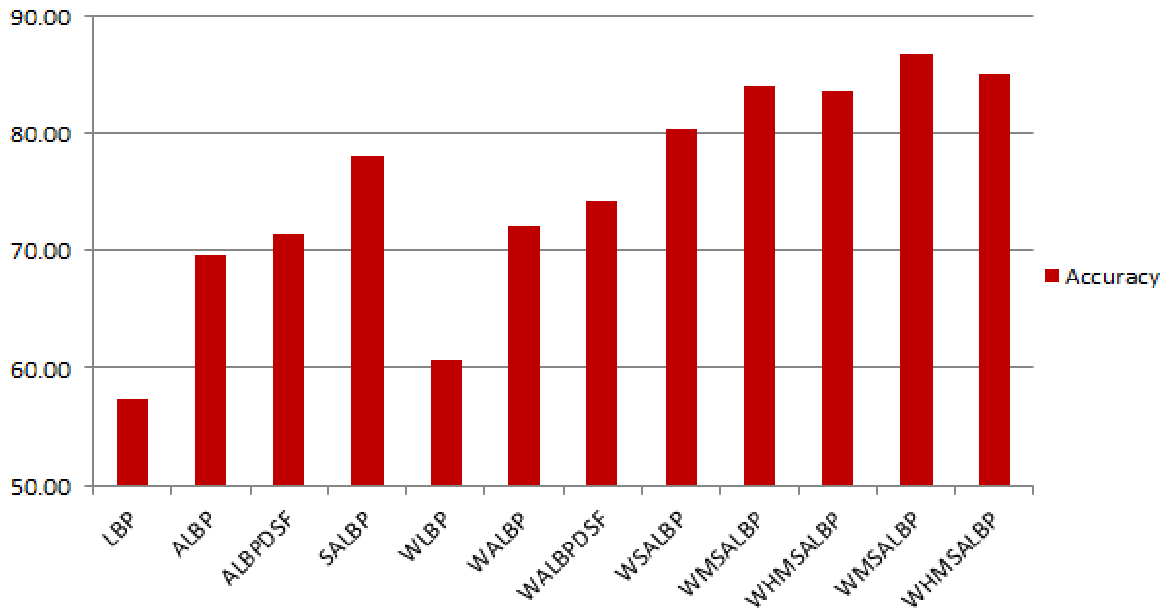


Figure 64: Average of accuracy rates for Salt & Pepper noisy ORL dataset with different LBP techniques.

7.5.3.4 Salt & Pepper Noisy ORL dataset Results Evaluation

The difference in recognition rate between the highest recognition rate obtained from applying WMSALBP and the highest recognition rate obtained from applying a traditional technique, WALBPDSF, is about 7%. To evaluate this difference and see if it is statistically significant or not I followed two steps process: First, check the distribution of the data obtained from applying WMSALBP and WALBPDSF. Figure 65 shows that I can consider my data as a normally distributed data.

Second, using a statistical test, such as Paired T-Test, to evaluate my data. The result obtained from applying Paired T-Test is as follows:

Paired T-Test and CI: WALBPDSF, WMSALBP

```
Paired T for WALBPDSF - WMSALBP

             N      Mean    StDev  SE Mean
WALBPDSF     9    0.7921   0.0845   0.0282
WMSALBP      9    0.8667   0.0854   0.0285
Difference   9  -0.07461  0.01824  0.00608


95% CI for mean difference: (-0.08863, -0.06059)
T-Test of mean difference = 0 (vs not = 0): T-Value = -12.27  P-Value = 0.000
```

The last row in the Paired T –Test result shows that P-value = 0.000 less than 0.05. With the assumption that the confidence level is 95.00%, I have to reject the Null hypothesis (H_0) of no difference in recognition rate between WALBPDSF and WMSALBP. Therefore, I believe there is a significant difference in recognition rate between WMSALBP and WALBPDSF.

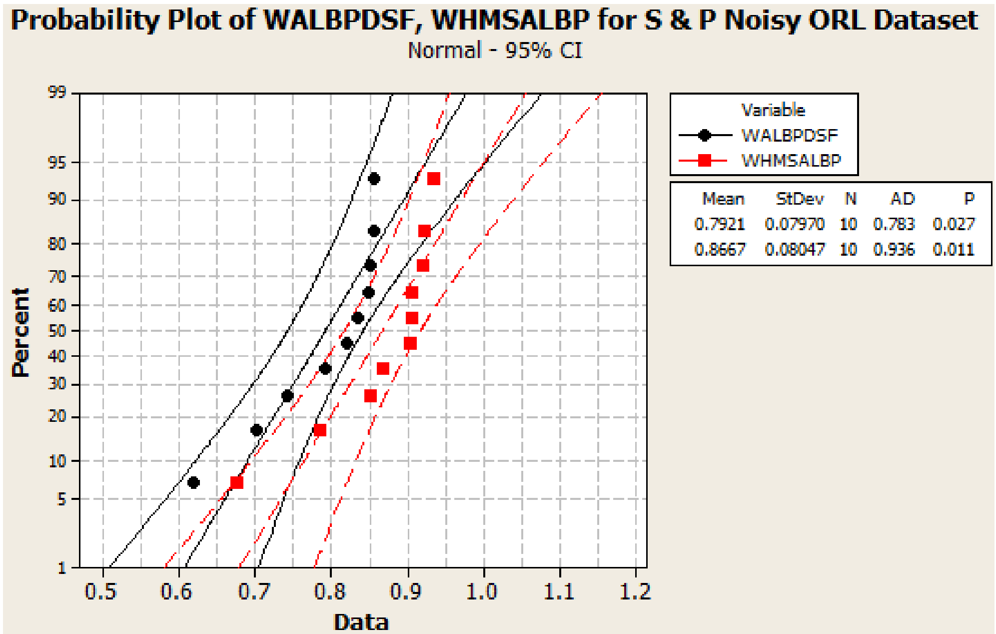


Figure 65: Distribution for WMSALBP and WALBPDSF data for Salt & Pepper Noisy ORL dataset.

7.5.4 Summary of Statistical Adapted LBP Techniques Results

The following three tables summarize the results obtained from applying different LBP variants on the three datasets of facial images that I tested:

Table 6: Percentage average accuracy rates for noisy SL dataset with different LBP variants

Technique	Operator					
	(8, 1)		(16, 2)		(24, 3)	
	Gaussian	S & P	Gaussian	S & P	Gaussian	S & P
LBP	46.90	50.60	50.00	52.61	56.88	54.53
ALBP	53.55	76.64	64.51	82.40	68.71	82.76
ALBPDSF	53.53	79.04	64.75	82.13	63.35	78.20
SALBP	59.02	81.85	70.16	84.91	74.57	85.81
WLBP	53.38	70.87	60.26	72.81	59.33	73.55
WALBP	53.13	80.45	66.48	83.50	71.71	85.77
WALBPDSF	58.41	80.29	68.80	79.78	71.50	77.29
WSALBP	68.54	82.98	76.12	87.57	76.67	88.11
MSALBP	73.18 (Gaussian) 88.49 (S & P)					
HMSALBP	74.87 (Gaussian) 88.22 (S & P)					
WMSALBP	78.73 (Gaussian) 88.79 (S & P)					
WHMSALBP	80.42 (Gaussian) 89.50 (S & P)					

Table 7: Percentage average accuracy rates for noisy ENT dataset with different LBP variants

Technique	Operator					
	(8, 1)		(16, 2)		(24, 3)	
	Gaussian	S & P	Gaussian	S & P	Gaussian	S & P
LBP	40.27	46.13	42.11	53.40	47.57	56.48
ALBP	47.79	50.56	46.10	61.80	49.98	69.43
ALBPDSF	46.42	53.94	50.20	59.34	52.13	59.91
SALBP	59.94	64.66	56.36	67.72	57.80	69.18
WLBP	42.80	52.59	48.08	56.99	54.42	60.65
WALBP	49.00	57.40	50.56	66.25	59.37	68.80
WALBPDSF	58.03	57.83	55.68	61.23	59.24	60.42
WSALBP	63.46	70.15	60.82	69.88	62.36	72.18
MSALBP	58.62 (Gaussian) 70.14 (S & P)					
HMSALBP	59.73 (Gaussian) 71.29 (S & P)					
WMSALBP	64.12 (Gaussian) 73.39 (S & P)					
WHMSALBP	65.43 (Gaussian) 75.51 (S & P)					

Table 8: Percentage average accuracy rates for noisy ORL dataset with different LBP variants

Technique	Operator					
	(8, 1)		(16, 2)		(24, 3)	
	Gaussian	S & P	Gaussian	S & P	Gaussian	S & P
LBP	45.14	54.56	57.66	57.10	61.40	60.51
ALBP	50.33	64.28	66.67	72.02	75.04	72.57
ALBPDSF	62.37	67.55	69.85	69.85	71.00	76.93
SALBP	64.25	72.61	72.95	79.47	78.92	81.92
WLBP	48.79	58.96	59.91	59.03	65.11	64.05
WALBP	57.27	66.95	74.16	73.77	79.89	75.39
WALBPDSF	64.51	69.24	74.29	74.16	79.51	79.21
WSALBP	68.21	75.29	76.77	82.37	82.70	83.51
MSALBP	80.60 (Gaussian) 83.99 (S & P)					
HMSALBP	81.17 (Gaussian) 83.54 (S & P)					
WMSALBP	84.35 (Gaussian) 86.67 (S & P)					
WHMSALBP	83.26 (Gaussian) 85.03 (S & P)					

7.6 Local Ternary Pattern (LTP)

Local binary pattern is a 2-valued (binary) code that is successfully used in many applications such as texture classification and analysis. The LBP operator is based on just two bit values either 1 or 0 which do not allow the LBP operator to discriminate between multiple patterns. The LBP operator has two main points of weaknesses:

1- The LBP operator cannot distinguish between two pixels if the first one is near to the center pixel but a little bit below that pixel and the second one is far below the center pixel value. In this case, the LBP will deal by the same way with both of them and both of them will have the value 0 but this is unfair.

2- In flat image areas, such as in face images, where all pixels nearly have the same gray value if slight amount of noise is added to these areas the LBP operator would give some bits the value 0 and others the value 1. Therefore, the LBP feature will be instable and thus the LBP operator will not be suitable for analyzing these areas.

To solve these problems an extension to LBP, Local Ternary Pattern, was introduced recently (Akhloufi & Bendada, 2010; Bendada & Akhloufi, 2010; Wankou & Changyin, 2011; Xiaoyang & Triggs, 2010). Local ternary pattern (LTP) is a new 3-valued texture operator that can be considered as an extension of local binary pattern. Instead of thresholding based on only the central pixel value of the neighborhood, the user has to define a threshold say t and any pixel value within the interval of $-t$ and $+t$ when compared to the central pixel value has to assign a value 0. Any pixel value above threshold $+t$ when compared to the central pixel value has to assign a value 1 and any pixel value below threshold $-t$ when compared to the central pixel value has to assign a value -1. The following equation shows how to compute the LTP operator (Akhloufi & Bendada, 2010; Bendada & Akhloufi, 2010; Xiaoyang & Triggs, 2010):

$$LTP(i) = \begin{cases} 1 & \text{if } p_i - p_c \geq t \\ 0 & \text{if } |p_i - p_c| < t \\ -1 & \text{if } p_i - p_c \leq -t \end{cases} \qquad (29)$$

where t is a user specified threshold, p_i is a pixel value in the neighborhood and p_c is the central pixel value.

This definition leads to have a texture operator that is less sensitive to noise (since it is no longer mainly based on the value of the central pixel) but no longer strictly invariant to gray-level transformations. The following figure shows an explanation about how the LTP operator works by using the threshold value $t = 5$:

12	34	45
38	35	55
11	65	23

(a)

-1	0	1
0		1
-1	1	-1

(b)

Figure 66: LTP computation: a) The original image window b) the result after applying equation 29.

As we can see from figure 66.b. there are some negative values because of using the threshold $t = 5$ in the previous equation. To solve the problem of getting negative values the LTP is split into two separate LBP descriptors, upper pattern (LTPU) and lower pattern (LTPL). Each one of them has its own histogram. The LTPU is obtained by replacing each negative value in Figure 66.b by 0 and keeping the other values as they are. The LTPL is obtained by following two rules: first replacing each 1 by 0 and second each negative value by 1, both in figure 66.b. Figure 67 shows an example of how to obtain LTPU and LTPL.

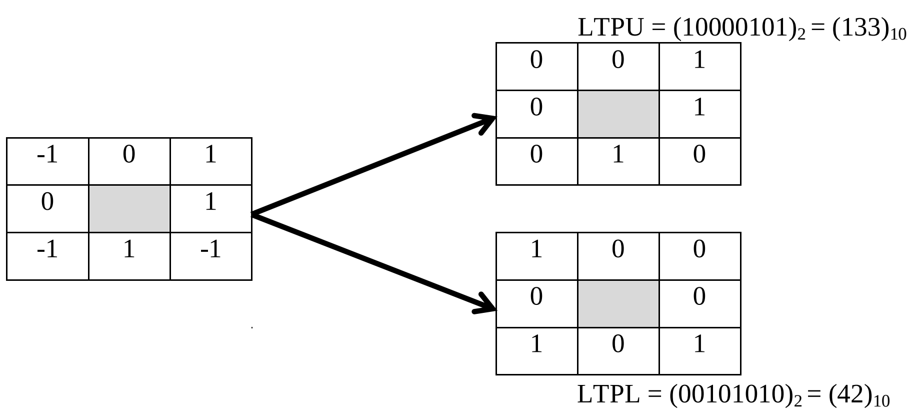


Figure 67: Splitting LTP into two LBP channels (Akhloufi & Bendada, 2010; Bendada & Akhloufi, 2010).

By using this new definition the two 3 x 3 windows of an image in figures 37 and 38 will have the same ternary representation as follows:

28	29	24
18	24	42
16	26	10

OR

28	29	23
18	24	42
16	26	10

0	0	0
-1		1
-1	0	-1

Figure 68: The LTP representation of the two windows in figures 37 and 38 with threshold t = 5.

7.7 Extended Local Ternary Pattern (ELTP)

Liao (Wen-Hung, 2010) figured out that when he applied the original LTP in his experiments the result was worse than the original LBP in the presence of noise. He proposed a new definition to the LTP in (Wen-Hung, 2010; Wen-Hung & Ting-Jung, 2010). Actually, its definition (extended local ternary pattern) is the same as the original one but he did not apply a fixed threshold. He converted image regions to its ELTP representation based on a threshold that employs the local statistics of the neighborhood of a central pixel.

The ELTP representation is the same as the original one except for the definition. In the original one, the threshold value t is fixed while in the ELTP t is not fixed but its value based on the local statistics of the region around the central pixel p_c. Liao used the following equation to compute t:

$$t = \alpha \text{ x } \sigma \quad (0 < \alpha \leq 1) \tag{30}$$

where σ is the standard deviation of the region around the central pixel and α is a scaling factor. So I can rewrite The LTP definition after applying the new threshold to be (Wen-Hung, 2010):

$$ELTP(i) = \begin{cases} 1 & \mathit{if}\ p_i - p_c \geq (\alpha * \sigma) \\ 0 & \mathit{if}\ |p_i - p_c| < (\alpha * \sigma) \\ -1 & \mathit{if}\ p_i - p_c \leq -(\alpha * \sigma) \end{cases} \tag{31}$$

Let us consider that the value of α is 0.5 the following figures 69 and 70 show an explanation and comparison about how LTP and ELTP are working. The result of applying the base 3-system approach is that the feature dimension size increases drastically to be 3^P where P is the number of the sample points in the neighborhood. To reduce the size of the feature dimension Liao (Wen-Hung, 2010) computed the similarity between any two ELTP strings that will transfer the dimension reduction problem to be a graph-partitioning problem.

11	34	30		-1	0	0
25	33	24	➡	0		0
49	32	55		1	0	1

Figure 69: Using LTP with fixed threshold value t =5

11	34	30		-1	0	0
25	33	24	➡	-1		-1
49	32	55		1	0	1

Figure 70: Using ELTP to represent a region of an image with $\alpha = 0.5$.

7.8 Adaptive Extended Local Ternary Pattern (AELTP)

Using the LTP operator allows to overcome some of the weaknesses found in applying the LBP operator. Instead of working with 2-valued codes in LBP, I can work with 3-valued codes in LTP, which increases the available number of patterns. However, in order to obtain good result after using the LTP operator I have to be very careful in choosing the system threshold. Obviously, for face recognition systems to gain good results, there is no fixed threshold and the best performing threshold has to vary depending on the facial dataset. Therefore, the ideal solution is to find ways that compute the threshold automatically based on the available facial dataset. To this end, I will define a weight (based on equation 26 in section 7.2) for each pixel in a neighborhood around a central one. I will use these weights to compute a new value for each pixel in this neighborhood. Local statistics of these new value pixels in that neighborhood will be used to compute the new threshold. This threshold value changes automatically from one patch of pixels to another during the whole image based on the values of the pixels in theses patches. I will use this new threshold in the definition of new LTP descriptor, which I call Adaptive Extended Local Ternary Pattern (AELTP). The following equation shows the proposed AELTP descriptor (Mohamed & Yampolskiy, 2012a):

$$AELTP(i) = \begin{cases} 1 & if\ \ p_i \geq (m + k\sigma) \\ 0 & if\ \ (m - k\sigma) < p_i < (m + k\sigma) \\ -1 & if\ \ \ \ p_i \leq (m - k\sigma) \end{cases} \qquad (32)$$

where σ refers to the standard deviation of the new pixels values in the neighborhood, *m* refers to their median and *k* refers to a constant (see figure 71).

11	34	30
25	33	24
49	32	55

(a)

1.375	4.25	4.25
3.125		3
6.125	4	6.875

(b)

-1	0	0
-1		-1
1	0	1

(c)

-1	0	0
-1		-1
1	0	1

AELTPU = $(00001010)_2 = (10)_{10}$

0	0	0
0		0
1	0	1

AELTPL = $(00110001)_2 = (49)_{10}$

1	0	0
1		1
0	0	0

(d)

Figure 71: AELTP computation (Mohamed & Yampolskiy, 2012a): a) The original image window b) the result after applying equation 26 c) The result of applying equation 32 d) Splitting the result of AELTP into two LBP channels.

I can summarize the steps for applying the AELTP operator on one facial image in the following steps on the basis that it is a single scale operator (Mohamed & Yampolskiy, 2012a):

1- Divide the facial image into non-overlapping sub-regions $J_0, J_1, \ldots J_{t-1}$, where J_0 is the first sub-region for this facial image and *t* is the number of non-overlapping sub-regions that forming this facial image.

2- Decide what the radius is and what is number of pixels in the neighborhood, say *R* and *P*.

3- Starting from the first available patch of pixels, map each sub-region to its corresponding one using AELTP codes (based on equation 32). Divided the result into two separate LBP patterns one for the positive part of AELTP and the second is for the negative part of AELTP. The resulting patterns are concatenated. Using AELTP operator increases the dimension of the features dramatically. Therefore, to reduce the dimensions of the features I use a combination of PCA and LDA.

4- Finally, for classification I use the Chi-Square distance.

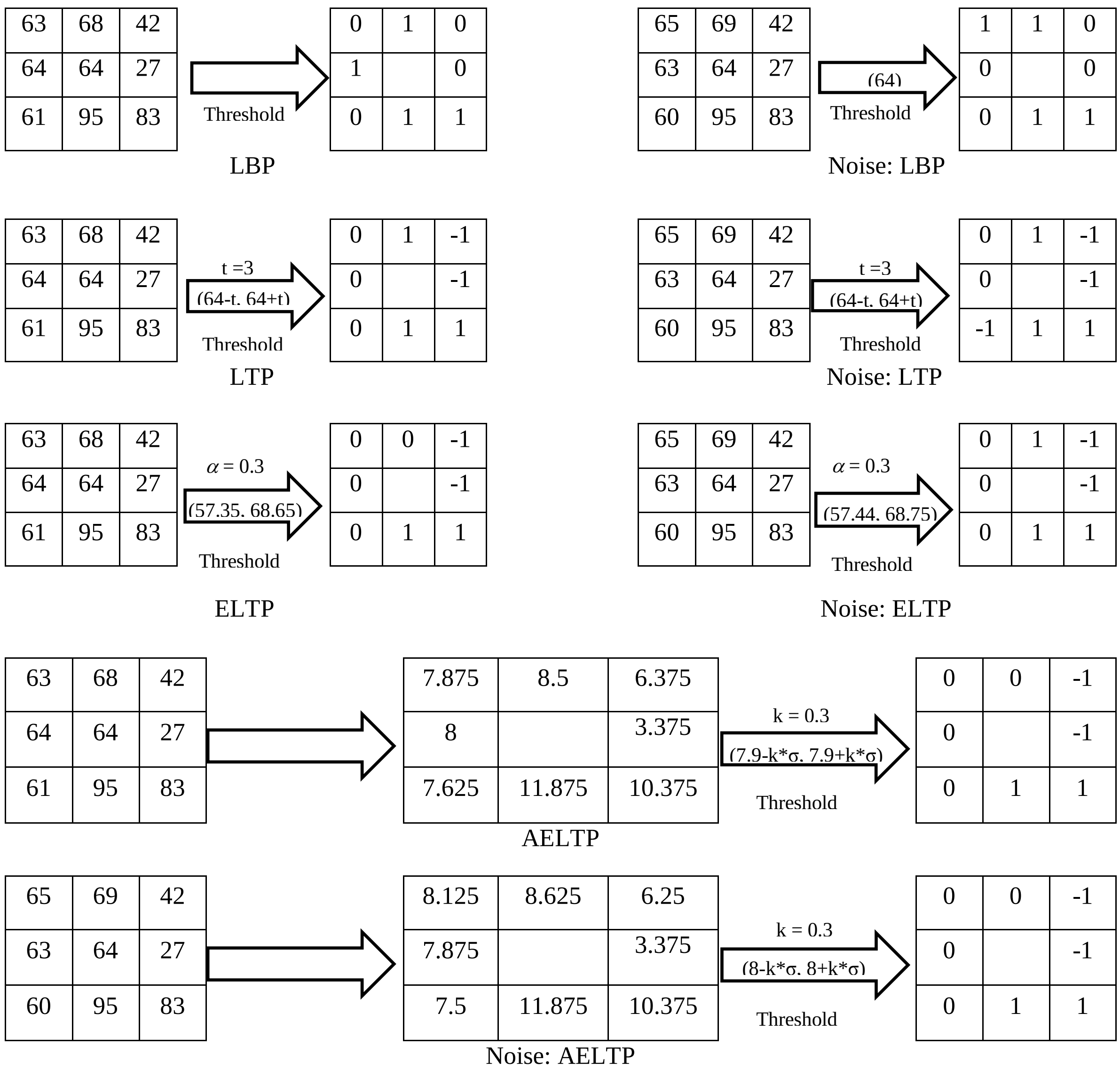


Figure 72: Comparison of LBP, LTP, ELTP and AELTP operators (Mohamed & Yampolskiy, 2012a).

7.9 Multi-scale Adaptive Extended Local Ternary Pattern (MAELTP)

In general, the multi-scale representation of an image had been achieved by different multi-scale LBP versions. To the best of our knowledge until now, there is no multi-scale representation of images using LTP operators. Multi-scale AELTP (MAELTP) representation of an image can be obtained by varying the radius of the sampling points and combining the resulted AELTP images.

To explain how this can be done let us suppose I have a single facial image. In the beginning this facial image has to be divided into non overlapping sub-regions say pt_1, pt_2,..pt_n. Apply the definition of AELTP operator on the first patch of pixels in the first sub-region pt_1 using the first scale. For simplicity, the resulted AELTP code will be separated into two LBP patterns, one for the positive part of the AELTP code and the second pattern for the negative part of the AELTP code. These two patterns have to be concatenated. I have to repeat these steps for all patches in this sub-region to get the histogram of this sub-region using the first scale (see figure 73).

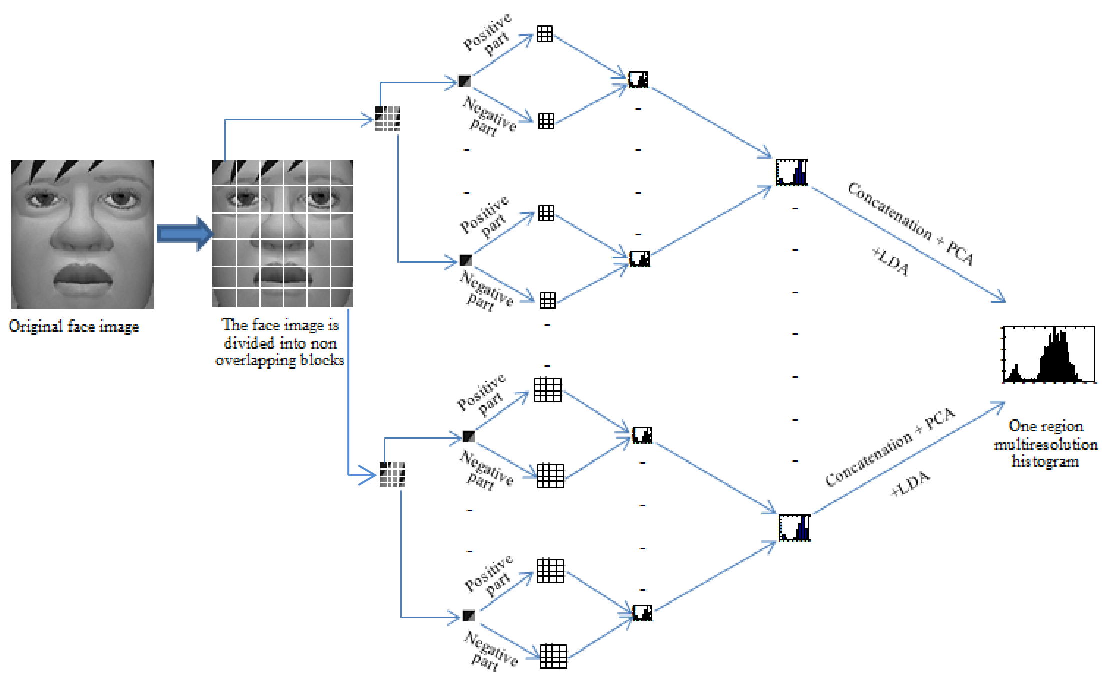


Figure 73: Multiresolution histogram for one region of the facial avatar image using AELTP operators.

I have to repeat these steps with all other scales to produce a histogram for this sub-region after using each scale with the AELTP operator. Applying each scale in the AELTP operator will build a histogram for this region. The set of histograms obtained by applying the AELTP operators on the first region will give regional information about that region. These histograms for the same region have to be concatenated into a single histogram to produce multiresolution information about that region. However, the problem now is the dimension of that histogram which becomes very high and it may contain redundant information. Applying PCA to extract independent information and to reduce the size of that histogram (see Fig. 73). Passing the result of applying PCA to LDA to extract discriminative facial features for each region (Chan, 2008). Projecting the reduced size histogram for one region on the LDA space will provide the regional discriminative facial descriptor for that region. Concatenating the regional discriminative facial descriptors for all regions will provide the global face description for this facial image that I have (see Fig. 73).

7.10 Wavelet Hierarchical Multi-scale Adaptive Extended Local Ternary Pattern (WHMAELTP)

WHMAELTP is another multi-scale definition of images. WHMAELTP is the same as MAELTP that I explained in the previous section except in two points:

a- To reduce the dimensionality of an image and at the same time preserve its representative features I apply different discrete wavelet families on the datasets that I use to evaluate the proposed techniques. Experimental results guided me to decompose SL with Db4 with the fifth level of decomposition, ENT with DB3 with the third level of decomposition and ORL with Db4 with the fourth level of decomposition.

b- The decomposed datasets of facial images are the input to MAELTP descriptor. Starting from the first available local patch, I apply the definition of MAELTP descriptor but in a specific order. I start by the biggest neighborhood radius. The resulted non-uniform patterns have to be extracted again but with the next radius smaller than the biggest one. This process has to continue until the smallest radius is processed. This hierarchical scheme is not sensitive to the number of available training samples since it does not require any training process.

7.11 Statistical Adapted ELTP Techniques Results

To evaluate the performance of the proposed ELTP techniques I tested the same two groups of noisy images that I tested in section 7.5 and the results can be shown as follows:

7.11.1 SL Dataset Results

SL is one of the two virtual worlds datasets that I used to test the performance of the proposed adapted ELTP techniques is the SL that I discussed and tested in section 7.5.1.

7.11.1.1 Gaussian Noisy SL dataset Results

The following two figures, figures 74 and 75, show that the performance of the adapted ELTP is better than the performance of the traditional LTP techniques with regard to the accuracy rates for different LBP operators and with different sizes of training datasets.

The highest obtained recognition rate is 83.23% after applying WHMAELTP and the highest recognition rate obtained from applying traditional technique is 74.11% and it is obtained after applying ELTP with

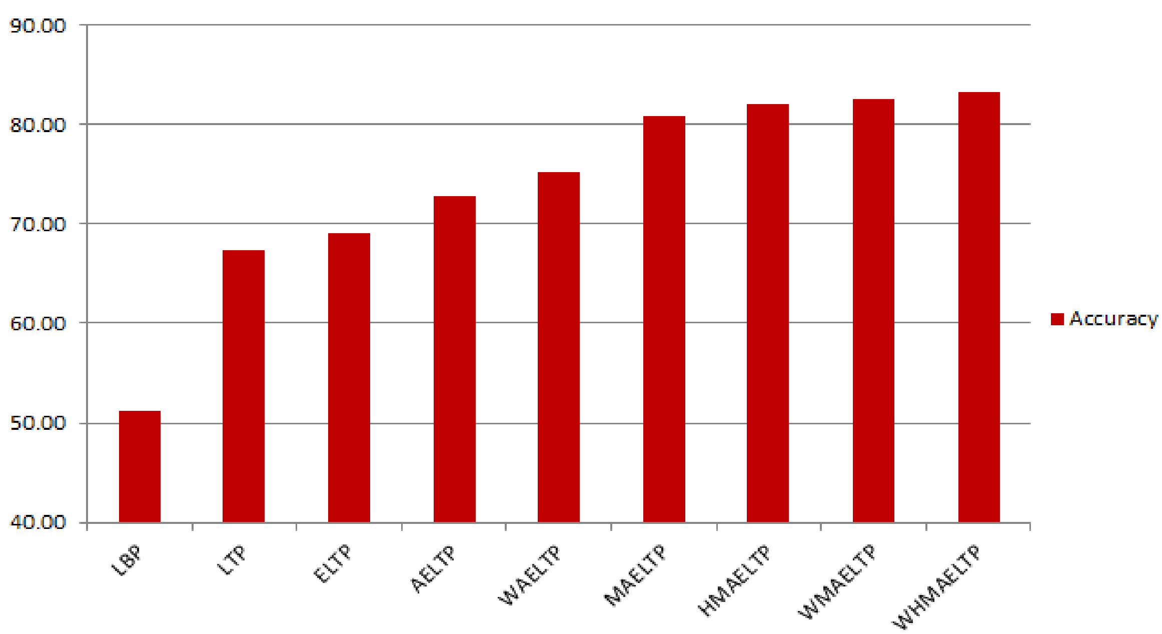


Figure 74: Average of accuracy rates for Gaussian noisy SL dataset with different LTP techniques.

(24, 3) LBP operator. In addition, in averaging accuracy rates of ALL LBP operators with all training images, the closest traditional technique to WHMAELTP is WAELTP with 69.18% accuracy rate as in Fig. 74.

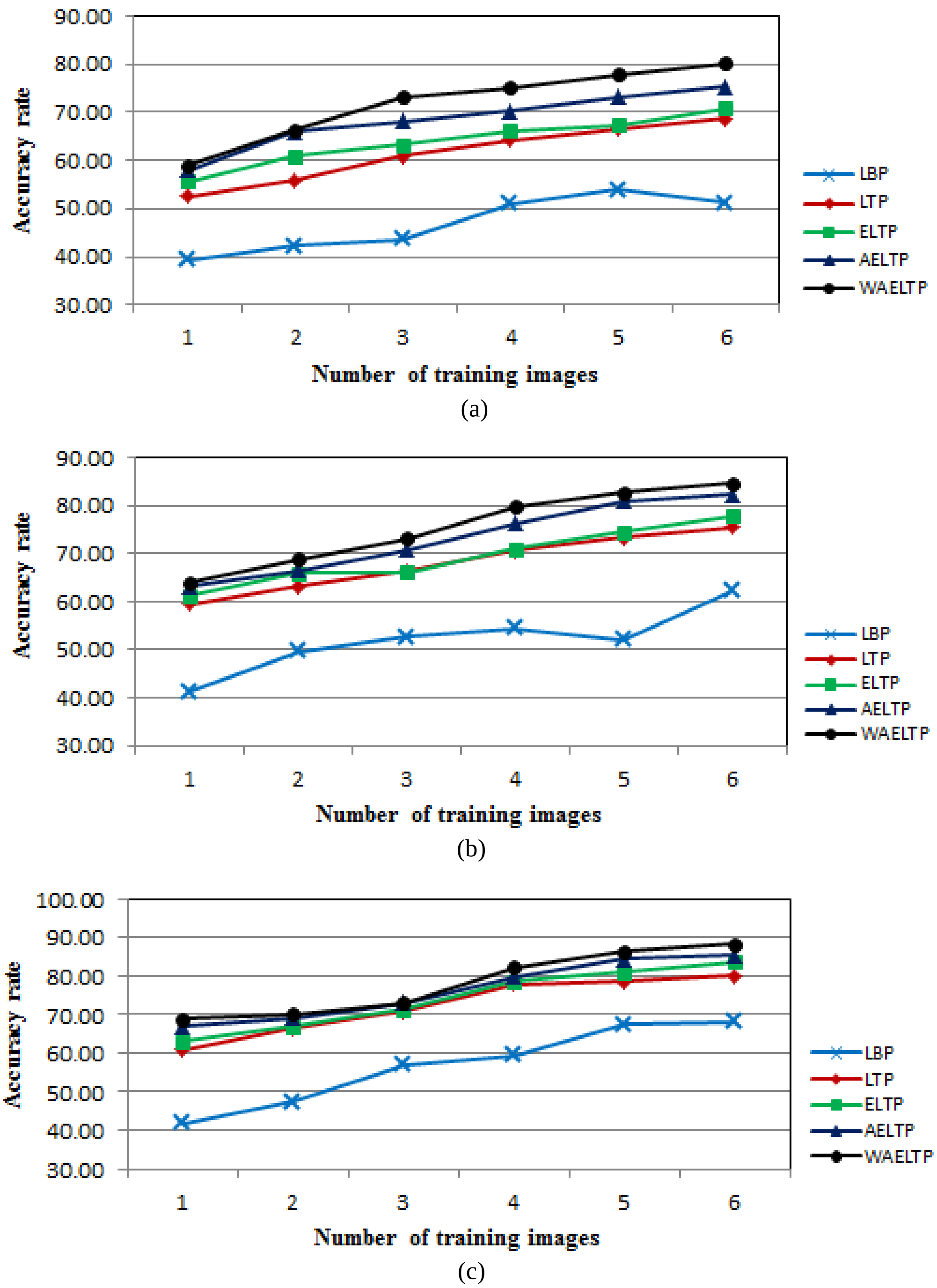


Figure 75: Accuracy rate for Gaussian noisy SL dataset for different LTP techniques with different LBP operators: (a) with (8, 1) operator (b) with (16, 2) operator (c) with (24, 3) operator.

7.11.1.2 Gaussian Noisy SL dataset Results Evaluation

The difference in recognition rate between WHMAELTP (satisfies the highest recognition rate) and ELTP (satisfies the highest recognition rate for traditional techniques) is almost 9%. To evaluate this difference and see either it is statistically significant or not I follow two steps process: First, check the distribution of the data points resulted from applying both WHMAELTP and ELTP (see figure 76).

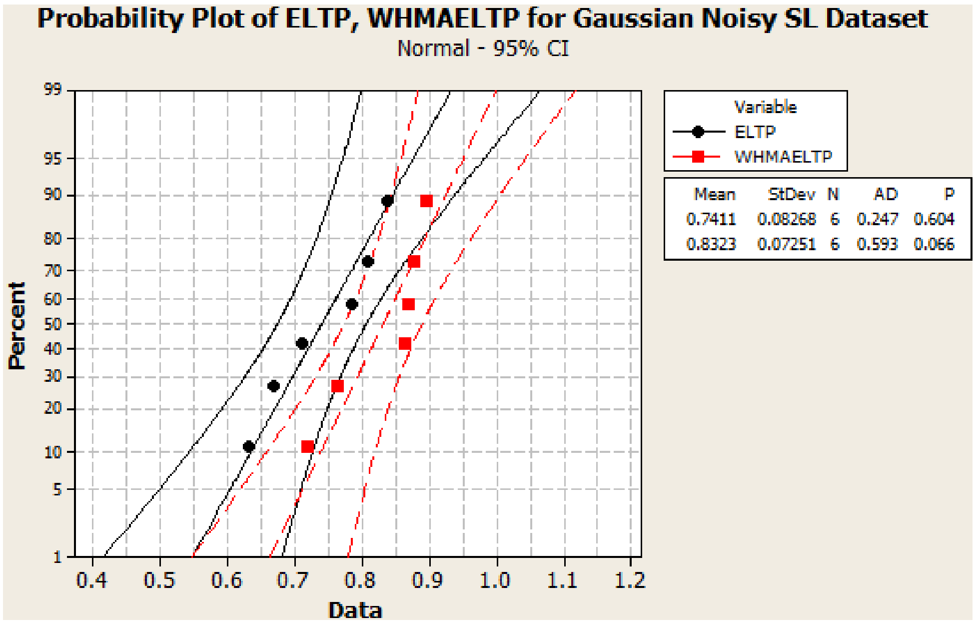


Figure 76: Distribution for WHMAELTP and ELTP data for Gaussian Noisy SL dataset.

From figure 76, it is clear that my data is normally distributed and hence I can evaluate my data using statistical test. Second, I tested my data using one statistical test, Paired T-Test, and the results as follows:

Paired T-Test and CI: ELTP, WHMAELTP

```
Paired T for ELTP - WHMAELTP

              N      Mean   StDev  SE Mean
ELTP          6    0.7411  0.0827   0.0338
WHMAELTP      6    0.8323  0.0725   0.0296
Difference    6   -0.0911  0.0369   0.0151


95% CI for mean difference: (-0.1299, -0.0524)
T-Test of mean difference = 0 (vs not = 0): T-Value = -6.05  P-Value = 0.002
```

The last row of the results show that P-Value is = 0.002 which is less than 0.05. By assuming that the confidence level is 95.00%, I have to reject the Null hypothesis (H_0) of no difference in recognition rate between WHMAELTP and ELTP. Therefore, I believe there is a significant difference in recognition rate between WHMAELTP and ELTP.

7.11.1.3 Salt & Pepper Noisy SL Dataset Results

The result obtained from applying different LTP techniques show that the adaptive extended versions of LTP are higher than the other (traditional) LTP techniques in accuracy rates (see figures 77 and 78).

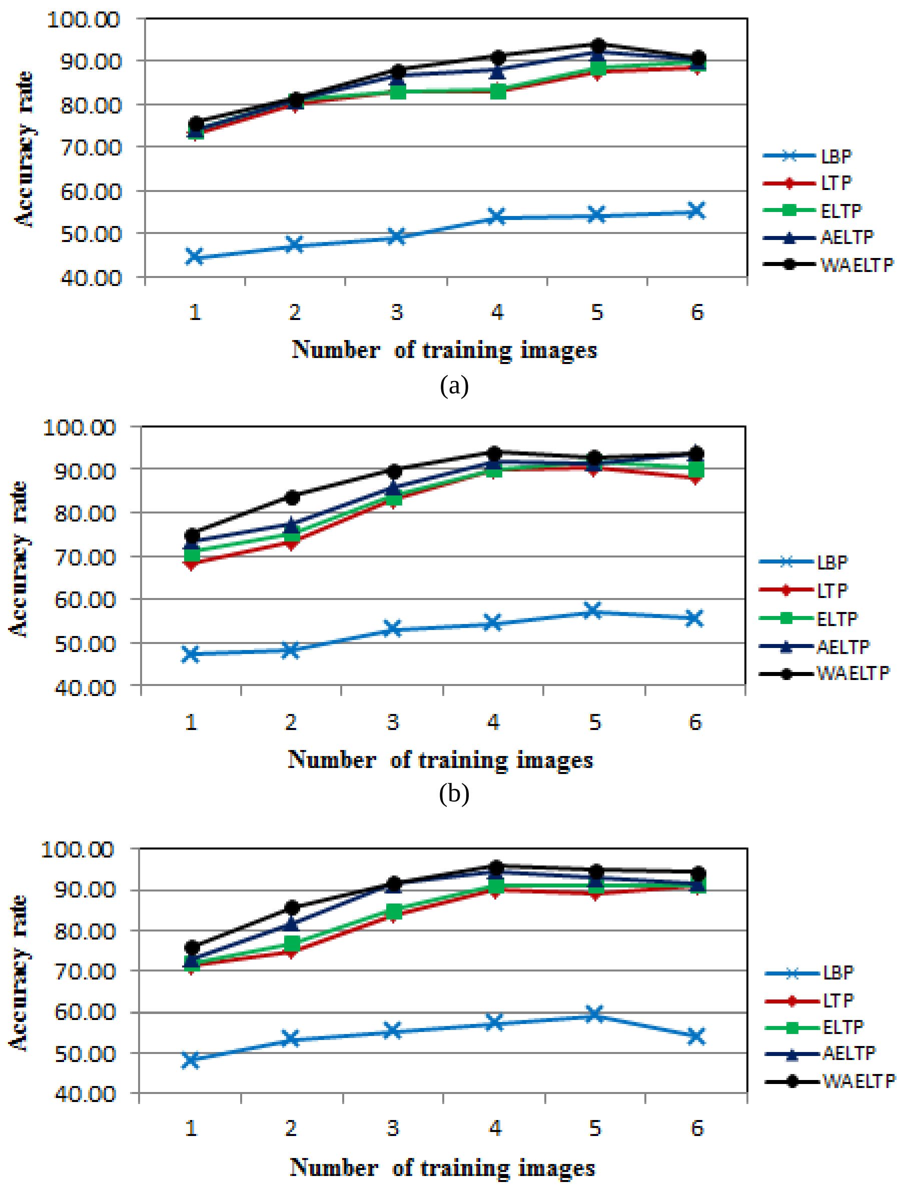


Figure 77: Accuracy rate for Salt & Pepper noisy SL dataset for different LTP techniques with different LBP operators: (a) with (8, 1) operator (b) with (16, 2) operator (c) with (24, 3) operator.

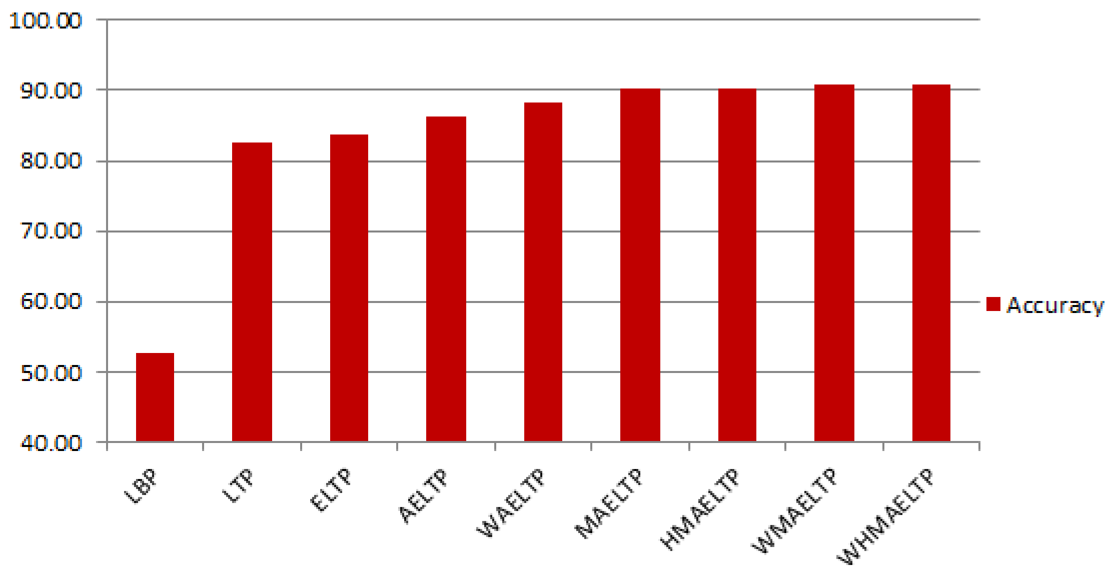


Figure 78: Average of accuracy rates for Salt & Pepper noisy SL dataset with different LTP techniques.

The highest recognition rate is 90.83% obtained from applying WHMAELTP while the highest recognition rate obtained from applying traditional technique is 84.62% obtained from applying ELTP with (24, 3) LBP operator. In addition, in averaging accuracy rates of all LBP operators with all training images, the closest traditional technique to WHMAELTP is ELTP with 83.82% accuracy rate as in Fig. 78.

7.11.1.4 Salt & Pepper Noisy SL Dataset Results Evaluation

WHMAELTP's accuracy rate is higher than that obtained from its closest traditional competitor by about 6%. To check the significance of this increase in accuracy rate I have to follow two steps process: First, check the distribution of the data obtained from WHMAELTP and ELTP (see figure 79) and then test the significance of the difference of accuracy rates obtained from applying WHMAELTP to those obtained from ELTP. Figure79 shows that my data (difference in recognition rate obtained from WHMAELTP to that one obtained from ELTP) is normally distributed and hence I can test its significance using statistical tests. Actually, I tested the significance of my data using Paired T-Test and the obtained results as follows:

Paired T-Test and CI: ELTP, WHMAELTP

```
Paired T for ELTP - WHMAELTP

            N     Mean   StDev  SE Mean
ELTP        6   0.8462  0.0841   0.0343
```

```
WHMAELTP     6    0.9083  0.0707   0.0289
Difference   6   -0.0621  0.0265   0.0108


95% CI for mean difference: (-0.0899, -0.0343)
T-Test of mean difference = 0 (vs not = 0): T-Value = -5.75  P-Value = 0.002
```

Since the P-Value = 0.002 which is less than 0.05 (with the assumption that the confidence level is 95.00%) I have to reject the Null hypothesis (H_0) of no difference in recognition rate between WHMAELTP and ELTP. Therefore, I believe there is a significant difference in recognition rate between WHMAELTP and ELTP.

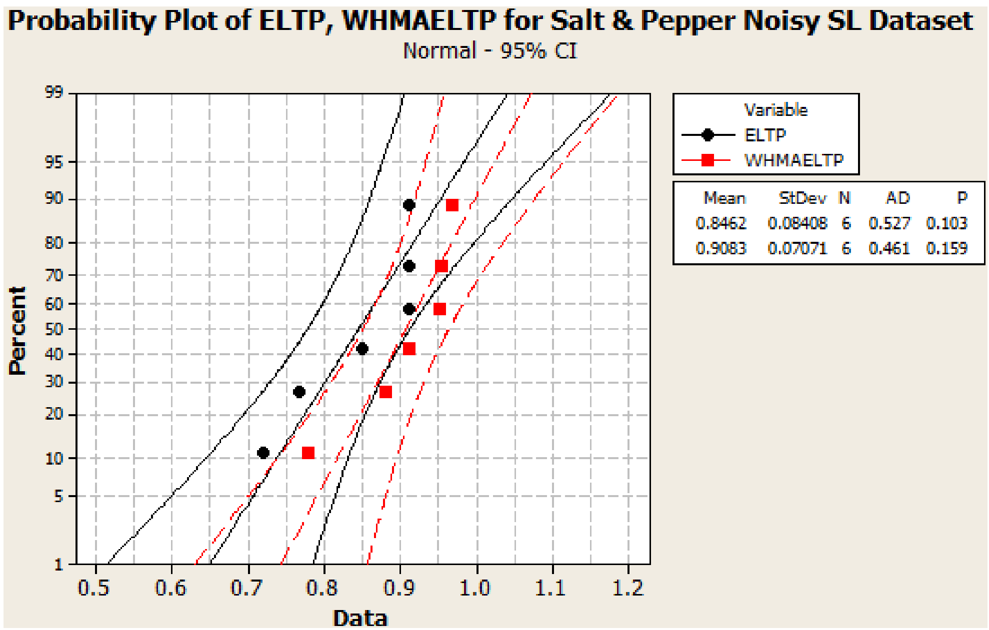


Figure 79: Distribution for WHMAELTP and ELTP data for Salt & Pepper Noisy SL dataset.

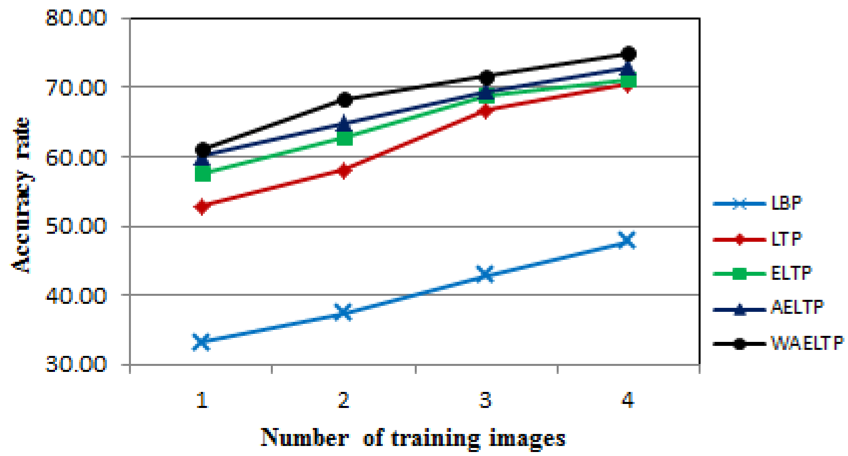
7.11.2 ENT Dataset Results

The second virtual dataset that I used to evaluate the performance of my proposed ELTP techniques is ENT dataset I described in section 7.5.1.

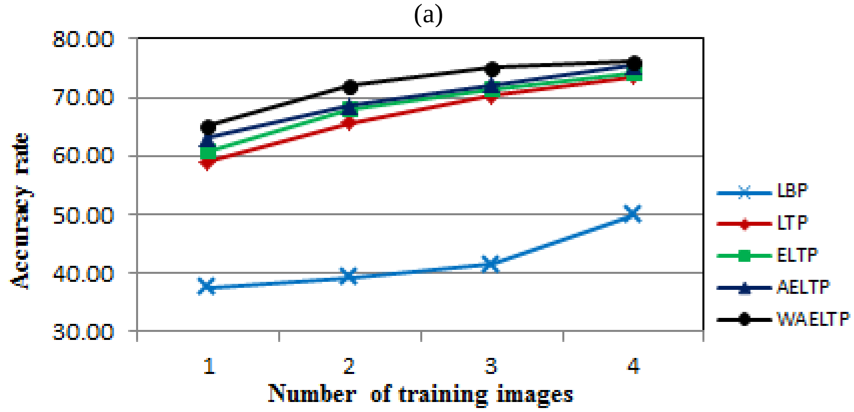
7.11.2.1 Gaussian Noisy ENT Dataset Results

Applying different versions of LTP techniques show that the accuracy rates obtained from applying adaptive extended versions of LTP are higher than those obtained from applying traditional LTP techniques (see figures 80 and 81).

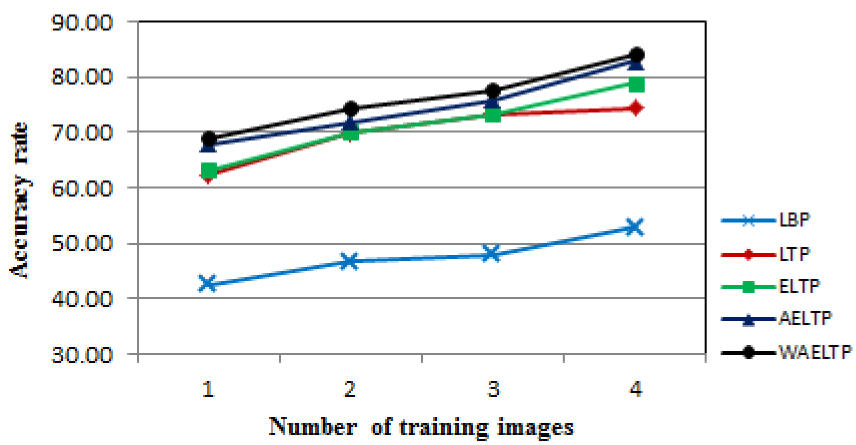
The highest recognition rate is 78.87% which is obtained from applying WHMAELTP while the highest recognition rate obtained from applying traditional technique is 71.31% obtained from applying ELTP with (24, 3) LBP operator. In addition, in averaging accuracy rates of all LBP operators with all training images, the closest traditional technique to WHMAELTP is ELTP with 68.36% accuracy rate as in Fig. 81.



(a)



(b)



(c)

Figure 80: Accuracy rate for Gaussian noisy ENT dataset for different LTP techniques with different LBP operators: (a) with (8, 1) operator (b) with (16, 2) operator (c) with (24, 3) operator.

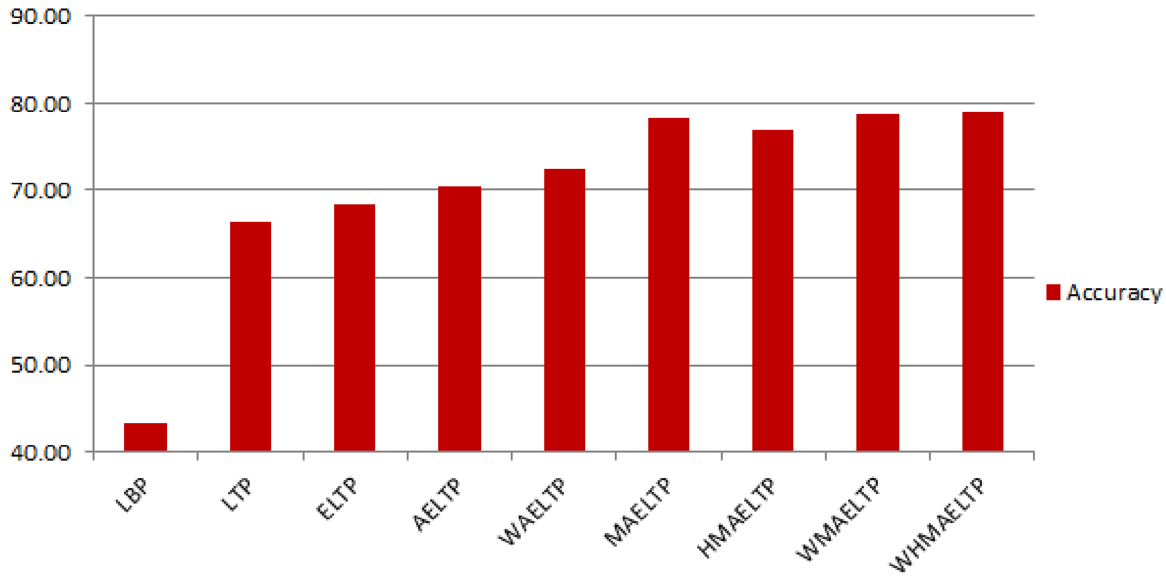


Figure 81: Average of accuracy rates for Gaussian noisy ENT dataset with different LTP techniques.

7.11.2.2 Gaussian Noisy ENT Dataset Results Evaluation

The accuracy rate obtained by applying WHMAELTP is higher than that obtained by applying ELTP by about 7%. To evaluate the significance of this difference in accuracy rate: First I have to check the normality of the result obtained from applying both WHMAELTP and ELTP and then check this difference if it is significance or not. Figure 82 shows the distribution of the resulted points obtained from applying both WHMAELTP and ELTP. It is clear from figure 82 that my data is normally distributed; therefore, I can use statistical tests to check its significance. I tested my data using a statistical test, Paired T-Test, and the result is as follows:

Paired T-Test and CI: ELTP, WHMAELTP

```
Paired T for ELTP - WHMAELTP

              N      Mean    StDev  SE Mean
ELTP          4    0.7131   0.0658   0.0329
WHMAELTP      4    0.7887   0.0654   0.0327
Difference    4  -0.07560  0.01646  0.00823


95% CI for mean difference: (-0.10179, -0.04941)
T-Test of mean difference = 0 (vs not = 0): T-Value = -9.18  P-Value = 0.003
```

Since the P-Value = 0.003 which is less than 0.05 (with the assumption that the confidence level is 95.00%) I have to reject the Null hypothesis (H_0) of no difference in recognition rate between WHMAELTP and ELTP. Therefore, I believe there is a significant difference in recognition rate between WHMAELTP and ELTP.

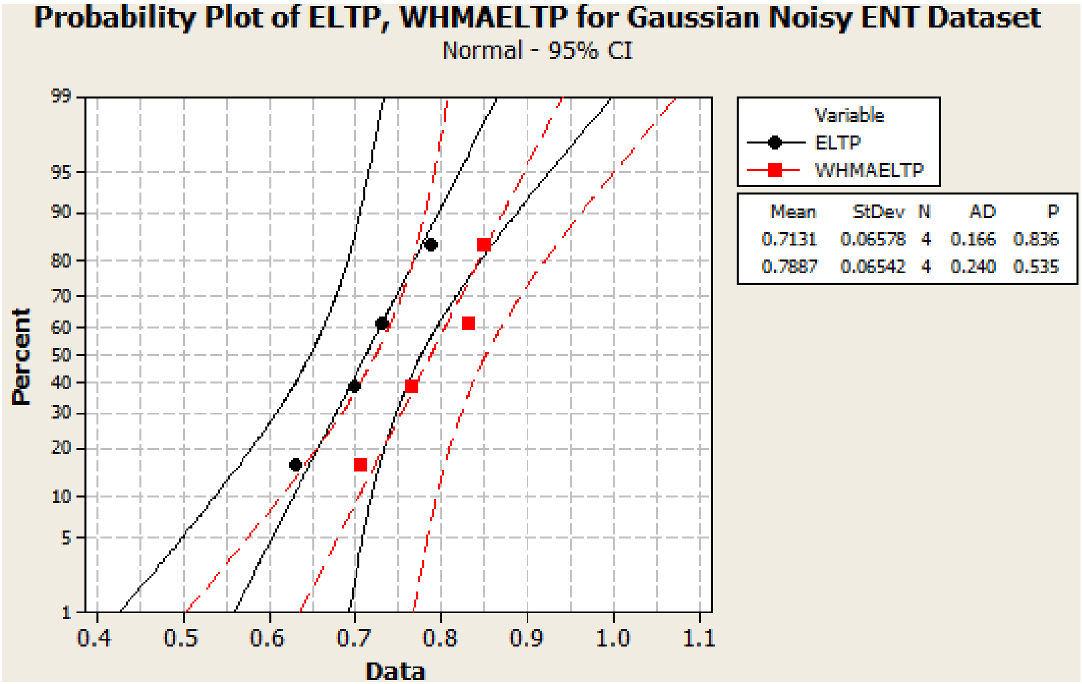
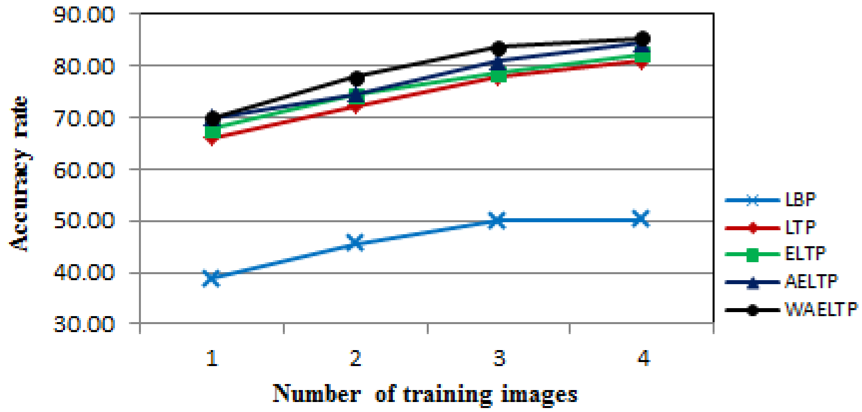


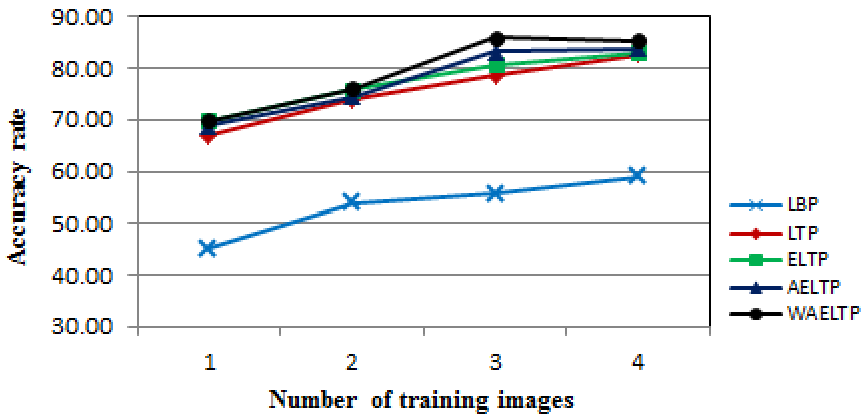
Figure 82: Distribution for WHMAELTP and ELTP data for Gaussian Noisy ENT dataset.

7.11.2.3 Salt & Pepper Noisy ENT Dataset Results

The following two figures, figures 83 and 84, show that the performance of statistical adapted ELTP techniques is better than that for traditional LTP techniques for recognizing avatar faces from a Salt and Pepper noisy avatar dataset. The best recognition rate is 82.77% which satisfied from applying WMAELTP technique while the best recognition rate obtained from a traditional technique is 77.54% obtained from applying ELTP with (24, 3) LBP operator. In addition, in averaging all accuracy rates obtained with all LBP operators with different training images datasets, the closest traditional technique to WMAELTP is ELTP with 76.87% accuracy rate as in Fig. 84.



(a)



(b)

(c)

Figure 83: Accuracy rate for Salt & Pepper noisy ENT dataset for different LTP techniques with different LBP operators: (a) with (8, 1) operator (b) with (16, 2) operator (c) with (24, 3) operator.

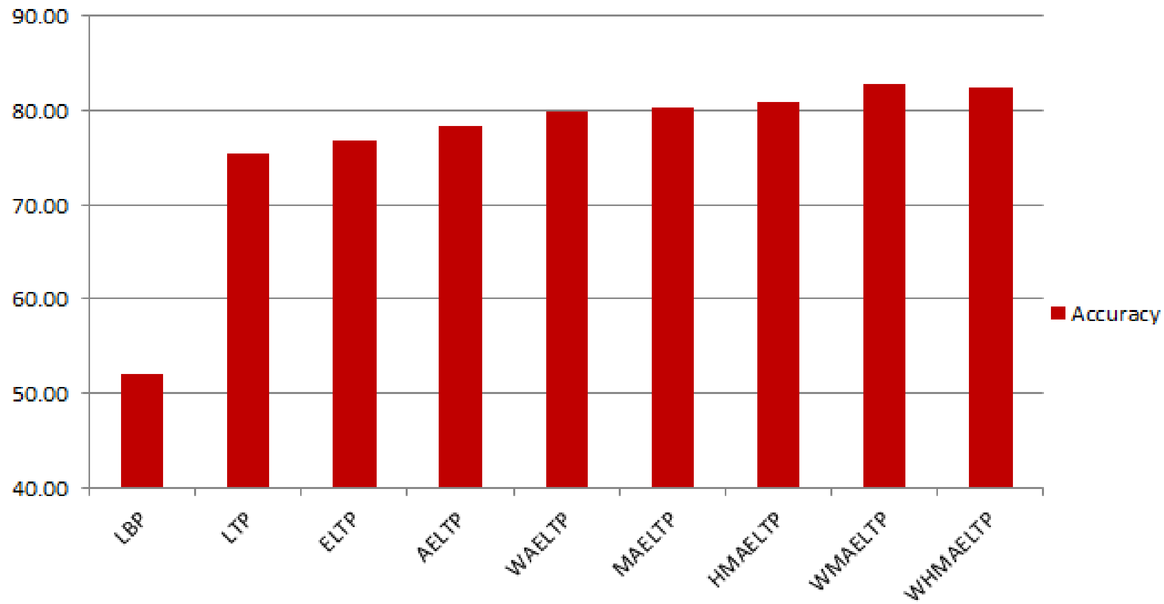


Figure 84: Average of accuracy rates for Salt & Pepper noisy ENT dataset with different LTP techniques.

7.11.2.4 Salt & Pepper Noisy ENT Dataset Results Evaluation

The performance of WMAELTP is better than that for its closest traditional competitor technique, ELTP, by about 5 % increase in the accuracy rate. To evaluate this difference and see how significant it is, first I have to check its normality and then test this difference using statistical tests. Fig. 85 shows the distribution of data obtained from applying both WMAELTP and ELTP on Salt & Pepper noisy ENT dataset.

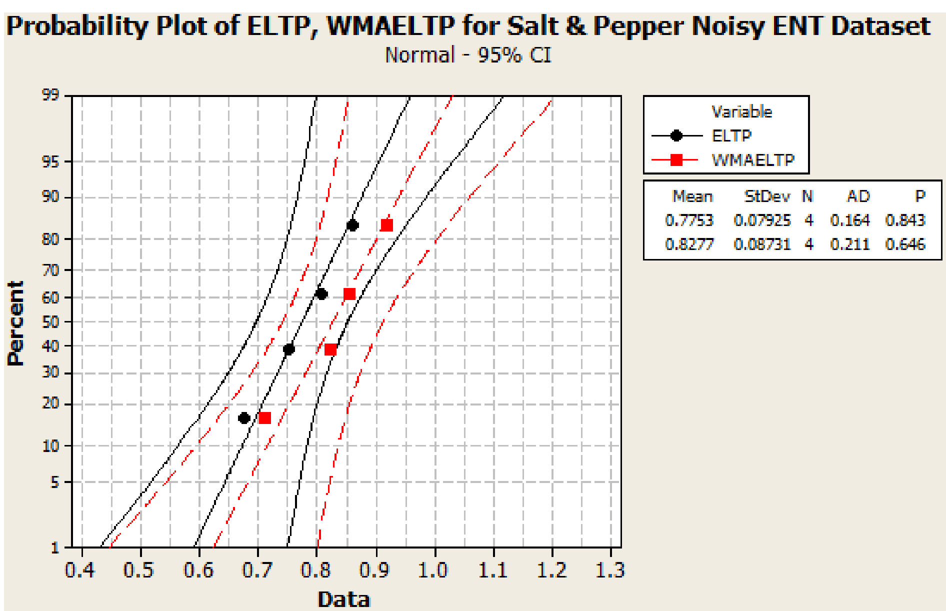


Figure 85: Distribution for WMAELTP and ELTP data for Salt & Pepper Noisy ENT dataset.

Fig. 85 shows that my data is normally distributed and hence I tested my data using a statistical test, Paired T-Test and the result can be shown as follows:

Paired T-Test and CI: ELTP, WMAELTP

```
Paired T for ELTP - WMAELTP

             N      Mean    StDev  SE Mean
ELTP         4    0.7753   0.0793   0.0396
WMAELTP      4    0.8277   0.0873   0.0437
Difference   4  -0.05235  0.01529  0.00764


95% CI for mean difference: (-0.07667, -0.02803)
T-Test of mean difference = 0 (vs not = 0): T-Value = -6.85  P-Value = 0.006
```

The P-Value that obtained because of this test is a great major of how significant is the difference between the accuracy rate obtained from ELTP and WMALTP. Since P-Value = 0.006 which is less than 0.05 (with the assumption that the confidence level is 95.00%) I have to reject the Null hypothesis (H_0) of no difference in recognition rate between WMAELTP and ELTP. Therefore, I believe there is a significant difference in recognition rate between WHMAELTP and ELTP.

7.11.3 ORL Dataset Results

The second type of datasets that I used to evaluate the performance of my proposed ELTP techniques is real human datasets. I tested my proposed techniques using one dataset from this type, ORL dataset, which I described in section 7.5.1.

7.11.3.1 Gaussian Noisy ORL Dataset Results

The following two figures, figures 86 and 87, show that the performance of statistical adapted ELTP techniques is better than that for traditional LTP techniques for recognizing real human faces from a Gaussian noisy ORL dataset. The best recognition rate is 86.17% which satisfied from applying WMAELTP technique while the best recognition rate obtained from a traditional technique is 81.39% obtained from applying ELTP with (24, 3) LBP operator. In addition, in averaging all accuracy rates obtained with all LBP operators with different training images datasets, the closest traditional technique to WMAELTP is ELTP with 77.99% accuracy rate as in Fig. 87.

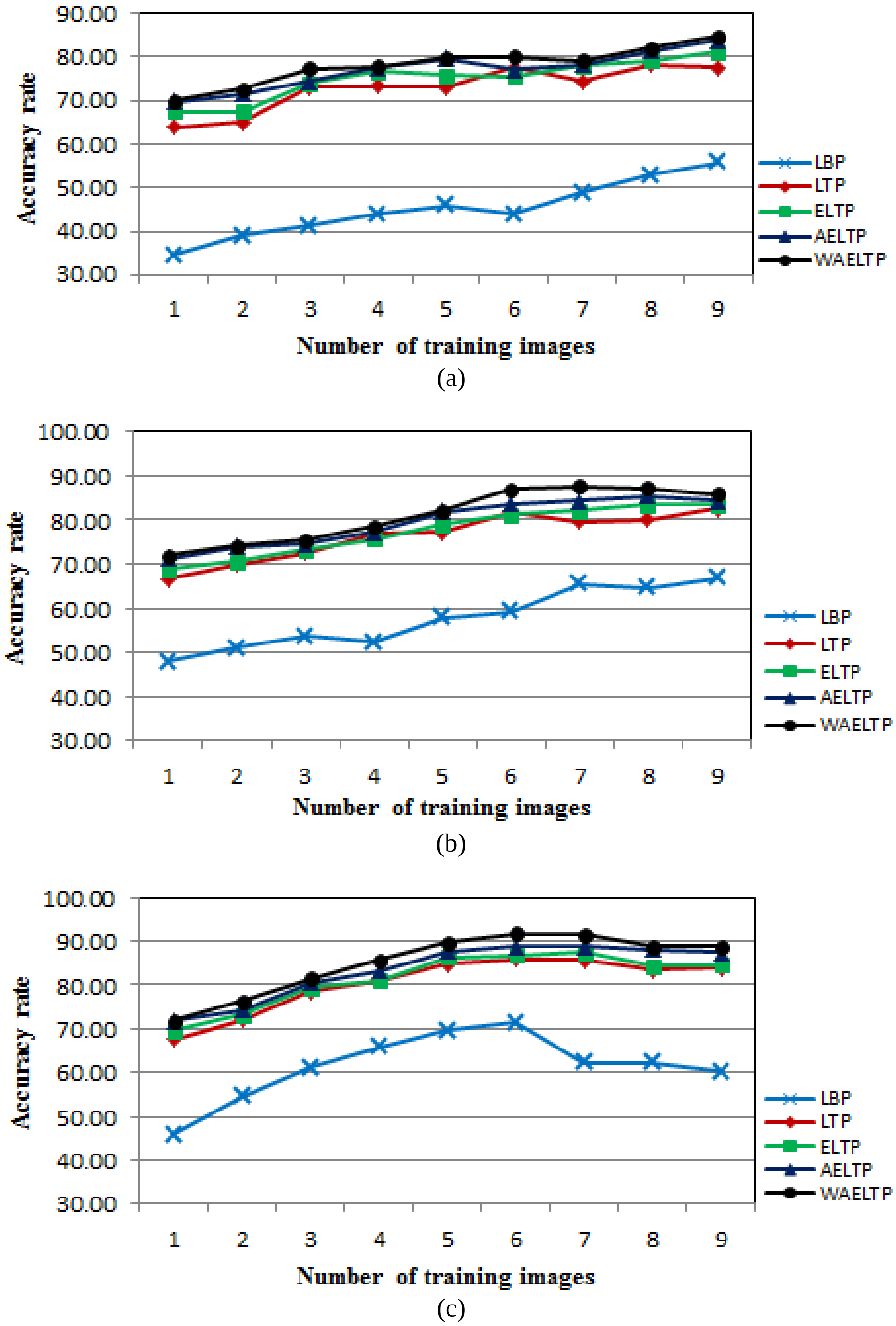


Figure 86: Accuracy rate for Gaussian noisy ORL dataset for different LTP techniques with different LBP operators: (a) with (8, 1) operator (b) with (16, 2) operator (c) with (24, 3) operator.

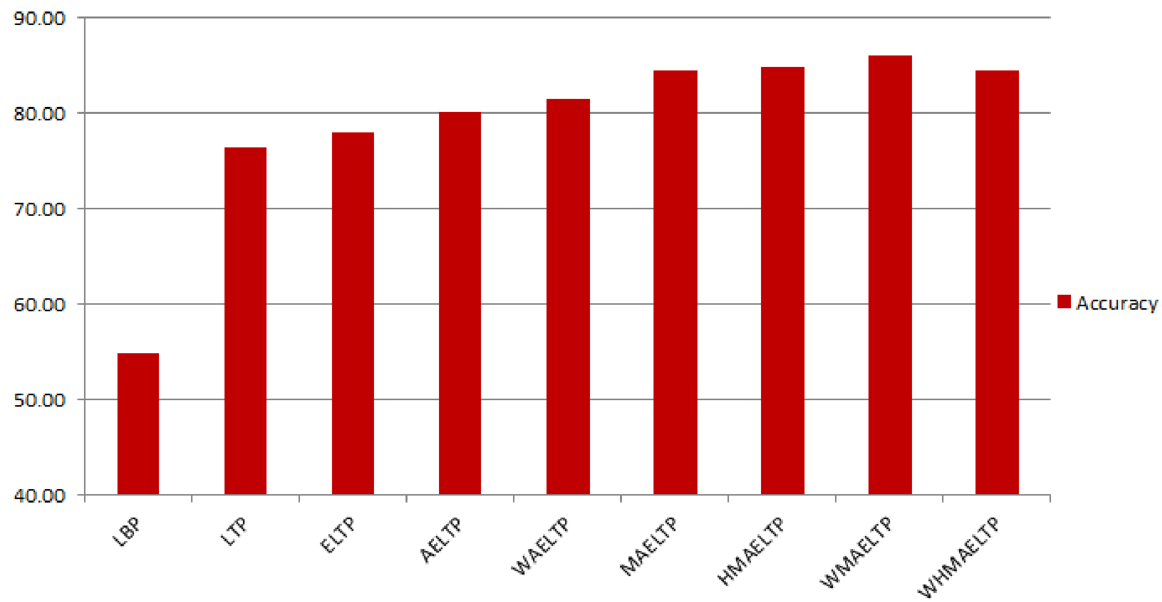


Figure 87: Average of accuracy rates for Gaussian noisy ORL dataset with different LTP techniques.

7.11.3.2 Gaussian Noisy ORL Dataset Results Evaluation

Figures 86 and 87 showed that the performance of adapted statistical ELTP is better than the performance of traditional LTP techniques. The performance of WMAELTP is better than that for its closest traditional competitor technique, ELTP, by about 5 % increase in the accuracy rate. To evaluate this difference and see how significant it is, first I have to check its normality and then test this difference using statistical tests. Fig. 88 shows the distribution of data obtained from applying both WMAELTP and ELTP on Gaussian noisy ORL dataset. Figure 88 shows that my data is nearly normally distributed and hence I can test that data using statistical tests. I test my data using a statistical test, Paired T-Test and the obtained results as follows:

Paired T-Test and CI: ELTP, WMAELTP

```
Paired T for ELTP - WMAELTP

            N      Mean    StDev  SE Mean
ELTP        9    0.8139   0.0633   0.0211
WMAELTP     9    0.8617   0.0599   0.0200
Difference  9  -0.04781  0.01024  0.00341


95% CI for mean difference: (-0.05569, -0.03994)
T-Test of mean difference = 0 (vs not = 0): T-Value = -14.00  P-Value = 0.000
```

The last row of the result obtained from applying the Paired T-Test on my data show that P-Value = 0.000 which is less than 0.05 (with the assumption that the confidence level is 95.00%) I have to reject the Null hypothesis (H_0) of no difference in recognition rate between WMAELTP and ELTP. Therefore, I believe there is a significant difference in recognition rate between WMAELTP and ELTP.

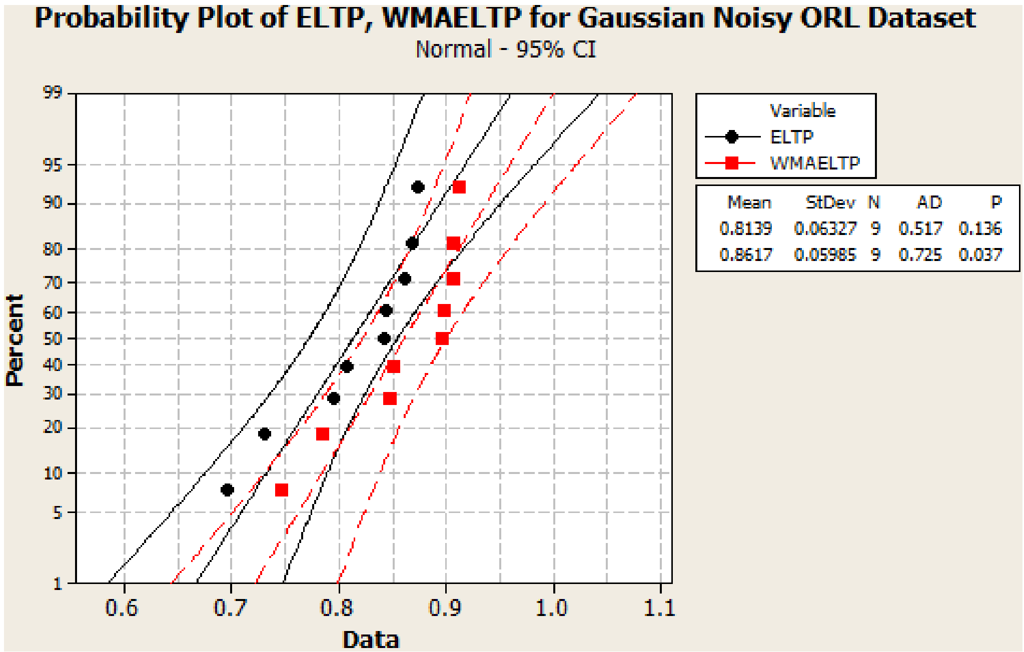
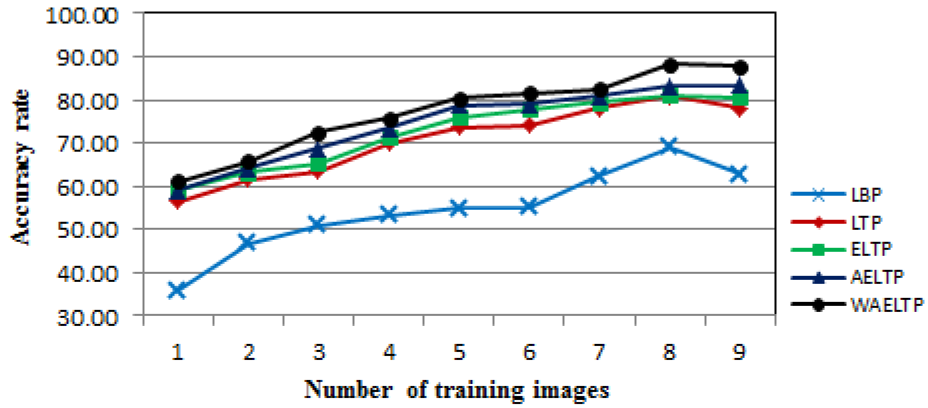


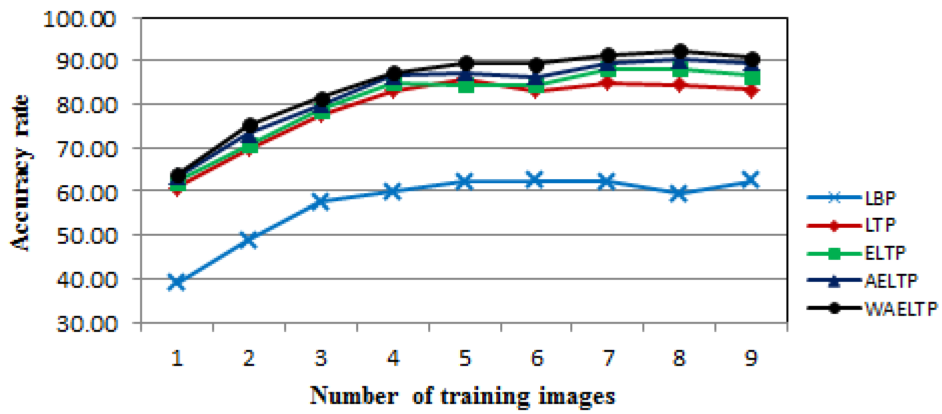
Figure 88: Distribution for WMAELTP and ELTP data for Gaussian Noisy ORL dataset.

7.11.3.3 Salt & Pepper Noisy ORL Dataset Results

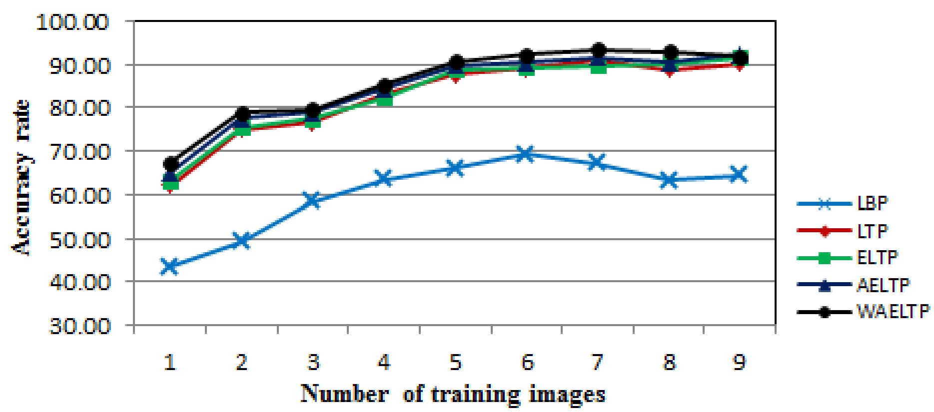
Figures 89 and 90 show that, the performance of statistical adapted LTP techniques in recognizing human faces noised by a Salt & Pepper noise is better than the performance of traditional LTP techniques. The best recognition rate 89.16% is obtained from applying one statistical adapted LTP technique, WHMAELTP, while the best recognition rate that can be obtained from applying a traditional LTP technique is 83.12%, which obtained from applying ELTP with (24, 3) LBP operator. In averaging all recognition rates obtained for the same LBP operator with all training datasets, the closest traditional technique to WHMAELTP is also ELTP with 78.64 % accuracy rate.



(a)



(b)



(c)

Figure 89: Accuracy rate for Salt & Pepper noisy ORL dataset for different LTP techniques with different LBP operators: (a) with (8, 1) operator (b) with (16, 2) (c) with (24, 3).

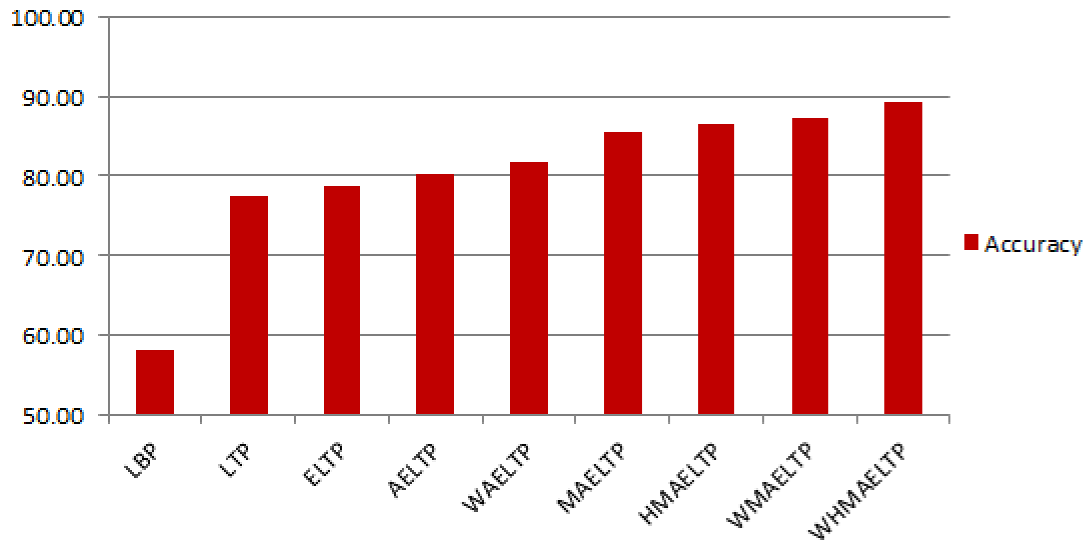


Figure 90: Average of accuracy rates for Salt & Pepper noisy ORL dataset with different LTP techniques.

7.11.3.4 Salt & Pepper Noisy ORL Dataset Results Evaluation

The difference in recognition rate between the best statistical adapted LTP technique, WHMAELTP, in recognizing faces of Salt & Pepper noisy ORL dataset and the best traditional technique, ELTP, in recognizing those faces is about 6% increase in the accuracy rate. To evaluate how significant is the difference, I have to follow two steps process: First, check the distribution of the results obtained from applying both WHMAELTP and ELTP (see figure 91). Second, test the obtained results using statistical tests. Figure 91 shows us that my data is nearly normally distributed and hence I can test it using statistical tests, such as Paired T-Test whose result can be seen as follows:

Paired T-Test and CI: ELTP, WHMAELTP

```
Paired T for ELTP - WHMAELTP

             N       Mean     StDev   SE Mean
ELTP         9     0.8312    0.0948    0.0316
WHMAELTP     9     0.8916    0.0821    0.0274
Difference   9   -0.06043   0.02615   0.00872


95% CI for mean difference: (-0.08053, -0.04034)
T-Test of mean difference = 0 (vs not = 0): T-Value = -6.93  P-Value = 0.000
```

The last row of the result obtained from applying the Paired T-Test on my data show that P-Value = 0.000 which is less than 0.05 (with the assumption that the confidence level is 95.00%) I have to reject the Null hypothesis (H_0) of no difference in recognition rate between WHMAELTP and ELTP. Therefore, I believe there is a significant difference in recognition rate between WHMAELTP and ELTP.

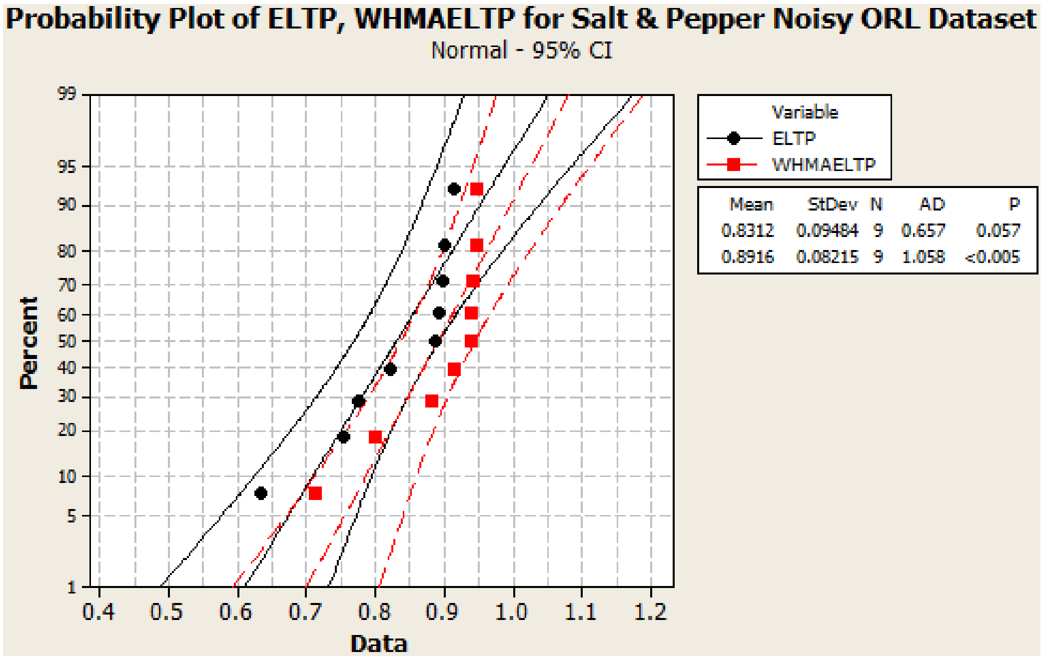


Figure 91: Distribution for WHMAELTP and ELTP data for Salt & Pepper Noisy ORL dataset.

7.11.4 Summary of Statistical Adapted LTP Techniques Results

The following three tables summarize the results obtained from applying different LTP techniques on the tested three datasets of faces:

Table 9: Percentage average accuracy rates for noisy SL dataset with different LTP variants

Technique	Operator					
	(8, 1)		(16, 2)		(24, 3)	
	Gaussian	S & P	Gaussian	S & P	Gaussian	S & P
LBP	46.90	50.60	50.00	52.13	56.88	54.33
LTP	61.44	82.50	68.02	82.13	72.44	83.44
ELTP	64.04	83.14	69.41	83.72	74.11	84.62
AELTP	68.48	85.27	73.29	85.63	76.41	87.64
WAELTP	71.95	86.79	75.50	88.25	78.16	89.80
MAELTP	80.88 (Gaussian) 90.23 (S & P)					
HMAELTP	82.00 (Gaussian) 90.40 (S & P)					
WMAELTP	82.61 (Gaussian) 90.70 (S & P)					
WHMAELTP	83.23 (Gaussian) 90.83 (S & P)					

Table 10: Percentage average accuracy rates for noisy ENT dataset with different LTP variants

Technique	Operator					
	(8, 1)		(16, 2)		(24, 3)	
	Gaussian	S & P	Gaussian	S & P	Gaussian	S & P
LBP	4027	46.13	42.11	53.40	47.57	56.48
LTP	62.01	74.22	67.05	75.54	69.95	76.82
ELTP	65.10	75.73	68.65	77.34	71.31	77.54
AELTP	66.79	77.47	69.82	77.58	74.60	79.86
WAELTP	68.96	79.17	72.11	79.22	76.27	81.29
MAELTP	78.35 (Gaussian) 80.20 (S & P)					
HMAELTP	77.00 (Gaussian) 80.94 (S & P)					
WMAELTP	78.62 (Gaussian) 82.77 (S & P)					
WHMAELTP	78.87 (Gaussian) 82.35 (S & P)					

Table 11: Percentage average accuracy rates for noisy ORL dataset with different LTP variants

Technique	Operator					
	(8, 1)		(16, 2)		(24, 3)	
	Gaussian	S & P	Gaussian	S & P	Gaussian	S & P
LBP	45.14	54.56	57.66	57.76	61.40	62.51
LTP	72.89	70.72	76.36	79.23	80.26	82.56
ELTP	75.07	72.66	77.52	80.13	81.39	83.12
AELTP	77.10	74.52	79.58	81.88	83.40	84.41
WAELTP	78.19	77.30	81.10	81.84	85.10	85.72
MAELTP	84.54 (Gaussian) 85.54 (S & P)					
HMAELTP	84.83 (Gaussian) 86.56 (S & P)					
WMAELTP	86.17 (Gaussian) 87.38 (S & P)					
WHMAELTP	84.52 (Gaussian) 89.16 (S & P)					

7.12 Conclusions

In this chapter, two main novel LBP approaches are presented, SALBP and AELTP. Estimating a suitable threshold for LBP and LTP approaches could be a difficult issue. SALBP and AELTP proposed a solution for this problem by using local statistics to determine the local threshold automatically. Based on the idea of SALBP and AELTP I proposed many other versions of LBP descriptors such as, MSALBP, HMSALBP, MAELTP, HMAELTP and WHMAELTP. The effectiveness of the proposed methods is shown in recognizing faces from both real and virtual worlds. Compared with state of the art and traditional methods and with different LBP operators and different sizes of training datasets, my proposed techniques achieved significant improvement in recognizing human and avatar faces.

CHAPTER 8

CONCLUSIONS AND FUTURE WORK

This research involves designing techniques to recognize human and avatar faces based on a well-known and efficient local texture descriptor, Local Binary Pattern (LBP) descriptor. I combined these different versions of LBP with a very popular tool for image analysis, Discrete Wavelet Transform (DWT). I also compared the performance of the proposed approaches against some popular face recognition techniques such as: LBP, Multi-scale LBP (MLBP) and Adaptive LBP (ALBP). The obtained results after applying the proposed approaches on different avatar and human datasets prove the effectiveness of the proposed techniques regarding both the accuracy rates and the processing time.

Also, to differentiate between human and avatar face images proposed in the ICMLA 2012 CAPTCHA challenge I applied different popular descriptors such as: Histograms of Oriented Gradients (HOG), Scale-Invariant Feature Transform (SIFT), Speeded-Up Robust Features (SURF) and Four-Patch Local Binary Pattern (FPLBP) in addition to my new LBP approach called Local Difference Pattern (LDP). I applied these descriptors with two different classifiers: Naïve Bayes and LibLinear. The obtained results were almost 99% recognition rate for each image and 88.6% accuracy rate for the 12 images. These results mean that the proposed CAPTCHA can easily be broken automatically and the proposed datasets of images need more transformations or ways to change their nature to be difficult to be broken.

I also proposed a new definition to the original LBP to treat one of its major weaknesses, sensitivity to noise. This new definition is based on computing weight for each pixel in the local neighborhood around a central pixel. The computed weight for each pixel can be used to redefine the pixel intensity. The new value for each pixel can be used in simple statistical operations to compute a new threshold. This threshold can be used to build the new definition of LBP. Based on this definition I can define different versions of LBP and Local Ternary Pattern (LTP) that can be used to recognize human or avatar noisy images.

Future work can have many directions, for example instead of using discrete wavelets transforms (DWT) I think combining Curvelet transforms, statistical adapted techniques of LBP and Locality Preserving Projection (LPP) with other versions of LDA like Direct LDA (DLDA) and Approximate LDA (ALDA) may achieve good recognition rates. In addition, I will work to develop my proposed techniques to work not only on gray scale images but also on color human and avatar images. Therefore, I may have descriptors like, SALBP and AELTP for recognizing color face images. I worked in this thesis on LBP and LTP variants but would also be interested in improving Local Quaternary Patterns (LQP) or any other higher level of representations of patterns.

REFERENCES

Active Worlds. from www.activeworlds.com

Ahn, L., Blum, M., Hopper, N., & Langford, J. (2003). *CAPTCHA: Using hard AI problems for security*. Paper presented at the 22nd International Conference on Theory and Applications of Cryptographic Techniques.

Ahn, L. V., Blum, M., & Langford, J. (2004). Telling humans and computers apart automatically. *Communications of the ACM, 47*(2), 56-60.

Ahonen, T., Hadid, A., & Pietikainen, M. (2006). Face Description with Local Binary Patterns: Application to Face Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 28*(12), 2037-2041. doi: 10.1109/TPAMI.2006.244.

Ajina, S., Yampolskiy, R. V., & Amara, N. E. B. (2010). *Avatar facial biometric authentication*. Paper presented at the 2nd International Conference on Image Processing Theory, Tools and Applications, Paris.

Ajina, S., Yampolskiy, R. V., & Amara, N. E. B. (2011). *Evaluation of SVM classification of avatar facial recognition*. Paper presented at the 8th International Conference on Advances in Neural Networks.

Akhloufi, M. A., & Bendada, A. (2010). *Locally adaptive texture features for multispectral face recognition.* Paper presented at the IEEE International Conference on Systems Man and Cybernetics (SMC).

Almazyad, A. S., Ahmad, Y., & Kouchay, S. A. (2011). *Multi-Modal CAPTCHA: A User Verification Scheme.* Paper presented at the International Conference on Information Science and Applications (ICISA),

Bartlett, M. S., Movellan, J. R., & Sejnowski, T. J. (2002). Face Recognition by Independent Component Analysis. *IEEE Transactions on Neural Networks, 13*(6), 1450-1464.

Bay, H., Tuytelaars, T., & Gool, L. V. (2006). *Surf: Speeded up robust features.* Paper presented at the European Conference on Computer Vision.

Belhumeur, P. N., Hespanha, J. P., & Kriegman, D. J. (1997). Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 19*(7), 711-720.

Bendada, A., & Akhloufi, M. A. (2010). *Multispectral Face Recognition in Texture Space.* Paper presented at the Canadian Conference on Computer and Robot Vision.

Boberg, M., Piippo, P., & Ollila, E. (2008). *Designing avatars.* Paper presented at the 3rd International Conference on Digital Interactive Media in Entertainment and Arts, Athens, Greece.

Boukhris, M., Mohamed, A. A., D'Souza, D., Beck, M., Ben Amara, N. E., & Yampolskiy, R. V. (2011). *Artificial human face recognition via Daubechies wavelet transform and SVM.* Paper presented at the 16th International Conference on Computer Games (CGAMES).

Bray, D. A., & Konsynski, B. R. (2007). Virtual worlds: multi-disciplinary research opportunities. *SIGMIS Database 38*(4), 17-25.

Chan, C.-H., Kittler, J., & Messer, K. (2007). *Multi-scale local binary pattern histograms for face recognition*. Paper presented at the International Conference on Biometrics, Seoul, Korea.

Chan, C., Kittler, J., & Messer, K. (2007). *Multi-scale binary pattern histogram for face recognition*. Paper presented at the International Conference on Biometrics. Lecture Notes in Computer Science.

Chan, C. H. (2008). *Multi-scale Local Binary Pattern Histogram for Face Recognition.* (Ph.D), University of Surrey.

Chandavale, A., Sapkal, A., & Jalnekar, R. (2010). A framework to analyze the security of text based captcha. *International Journal of Computer Applications, 1*(27), 127-132.

Chen, L.-F., Liao, H.-Y. M., Ko, M.-T., Lin, J.-C., & Yu, G.-J. (2000). A new LDA-based Face Recognition System Which Can Solve the Small Sample Size Problem. *Pattern Recognition, 33*(10), 1713-1726.

Cole, J. Osama Bin Laden's "Second Life", Salon. from http://www.salon.com/2008/02/25/avatars_2/

D'Souza, D., Polina, P. C., & Yampolskiy, R. V. (2012). *Avatar CAPTCHA: Telling computers and humans apart via face classification.* Paper presented at IEEE the International Conference on Electro/Information Technology (EIT).

D'Souza, D., & Yampolskiy, R. V. (2012). *Baseline Avatar Face Detection using an Extended Set of Haar-like Features*. Paper presented at the Midwest Artificial Intelligence and Cognitive Science Conference, Cincinnati, Ohio.

Dalal, N., & Triggs, B. (2005). *Histograms of oriented gradients for human detection.* Paper presented at the IEEE Computer Society Cnference on Computer Vision and Pattern Recognition, CVPR.

Dyck, M., Winbeck, M., Leiberg, S., Chen, Y., Gur, R. C., & Mathiak, K. (2008). Recognition profile of emotions in natural and virtual faces. *PLoS ONE 3*(11), e3628.

Ekenel, H. K., & Stiefelhagen, R. (2006). *Analysis of Local Appearance-Based Face Recognition: Effects of Feature Selection and Feature Normalization*. Paper presented at the Computer Vision and Pattern Recognition Workshop (CVPRW '06) New York, USA.

Elson, J., Douceur, J., Howell, J., & Saul, J. (2007). *Asirra: a CAPTCHA that exploits interestaligned manual image categorization.* Paper presented at the 14th ACM conference on Computer and communications security, CCS '07.

Entropia Universe. from www.entropiauniverse.com

Etemad, K., & Chellappa, R. (1997). Discriminant analysis for recognition of human face images. *Journal of the Optical Society of America A, 14*(8), 1724-1733.

Fan, R., Chang, K., Hsieh, C., Wang, X., & Lin, C. (2008). Liblinear: A library for large linear classification. *Machine Learning Research, 9*, 1871-1874.

The Free Dictionary. from www.thefreedictionary.com/avatar

Garcia, C., Zikos, G., & Tziritas, G. (1998). *A wavelet-based framework for face recognition.* Paper presented at the 5th European Conference on Computer Vision, Freiburg, Allemagne.

Gavrilova, M. L., & Yampolskiy, R. V. (2010). *Applying Biometric Principles to Avatar Recognition.* Paper presented at the International Conference on Cyberworlds (CW).

Haichang, G., Dan, Y., Honggang, L., Xiyang, L., & Liming, W. (2010). *A Novel Image Based CAPTCHA Using Jigsaw Puzzle.* Paper presented at the IEEE 13th International Conference on Computational Science and Engineering (CSE).

Haiping, L., Martin, K., Bui, F., Plataniotis, K. N., & Hatzinakos, D. (2009). *Face recognition with biometric encryption for privacy-enhancing self-exclusion.* Paper presented at the 16th International Conference on Digital Signal Processing.

Hall, M. (1999). *Correlation-based feature selection for machine learning.* (PhD), The University of Waikato.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. (2009). The WEKA data mining software: an update. *ACM SIGKDD Explorations Newsletter, 11*(1), 10-18.

Harvard Law Class in Second Life. from http://virtuallyblind.com/2007/11/17/harvard-extension-second-life/

Heikkila, M., & Pietikainen, M. (2006). A texture-based method for modeling the background and detecting moving objects. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 28*(4), 657-662. doi: 10.1109/TPAMI.2006.68

Hietmeyer, R. (2000). Biometric identification promises fast and secure processing of airline passengers. *The International Civil Aviation Organization Journal, 55*(9), 10-11.

HiPiHi. from http:// www.hipihi.com/en/

Intelligence Advanced Research Projects Activity. from http://www.iarpa.gov/

Jain, A. K., Ross, A., & Prabhakar, S. (2004). An introduction to biometric recognition. *Circuits and Systems for Video Technology, IEEE Transactions on, 14*(1), 4-20. doi: 10.1109/TCSVT.2003.818349

Jian, Y., Zhang, D., Frangi, A. F., & Jing-Yu, Y. (2004). Two-dimensional PCA: a new approach to appearance-based face representation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 26*(1), 131-137. doi: 10.1109/TPAMI.2004.1261097

Jie, H. Y., Yu, H., & Yang, J. (2001). A direct LDA algorithm for high-dimensional data - with application to face recognition. *Pattern Recognition, 34*, 2067-2070.

John, G., & Langley, P. (1995). *Estimating continuous distributions in bayesian classifiers.* Paper presented at the Uncertainty in Artificial Intelligence.

Jun, M., Yumao, G., Xiukun, W., Tsauyoung, L., & Jianying, Z. (2010). *Face Recognition Based on Local Binary Patterns with Threshold.* Paper presented at the IEEE International Conference on Granular Computing (GrC).

Kanade, T. (1973). *Picture processing system by computer complex and recognition of human faces.* (Ph.D.), Kyoto University.

Kelly School of Business. from http://info.kelley.iu.edu/news/page/normal/8773.html

Kong, H., Li, X., Wang, J.-g., Teoh, E. K., & Kambhamettu, R. (2005). *Discriminant Subspace Analysis for Face Recognition with Small Number of Training Samples.* Paper presented at the British Machine Vision, Oxford, UK.

Korayem, M., Mohamed, A. A., Crandall, D., & Yampolskiy, R. V. (2012). *Learning visual features for the avatar captcha recognition challenge.* Paper presented at the International Conference on Machine Learning and Applications, Boca Raton, FL, USA.

Korayem, M., Mohamed, A. A., Crandall, D., & Yampolskiy, R. V. (2012). *Solving Avatar Captchas Automaticall.* Paper presented at the Advanced Machine Learning Technologies and Applications (AMLTA 2012), Communications in Computer and Information Science series 322, Cairo, Egypt.

Lee, S., & Ozawa, S. (2003). *A Face Recognition System Using Neural Networks with Incremental Learning Ability*. Paper presented at the 8th Australian and New Zealand Conf. on Intelligent Information Systems.

Li, S. Z., & Jain, A. K. (2011). *Hand Book of Face Recognition*: Springer-Verlag.

Liao, S., Fan, W., Chung, A. C. S., & Yeung2, D.-Y. (2006). *Facial Expression Recognition using Advanced Local Binary Patterns,Tsallis Entropies and Global Appearance Features* Paper presented at the IEEE International Conference on Image Processing.

Liao, S., Zhu, X., Lei, Z., Zhang, L., & Li, S. (2007). *Learning multi-scale block local binary patterns for face recognition*. Paper presented at the International Conference on Biometrics. Lecture Notes in Computer Science.

Liming, W., Xiuling, C., Zhongjie, R., Haichang, G., Xiyang, L., & Aickelin, U. (2010). *Against Spyware Using CAPTCHA in Graphical Password Scheme.* Paper presented at the 24th IEEE International Conference on Advanced Information Networking and Applications (AINA).

Lowe, D. (1999). *Object recognition from local scale-invariant features.* Paper presented at the International Conference on Computer Vision.

Machine Readable Travel Documents (MRTD). from www.icao.int/mrtd/overview/overview.cfm

Mäenpää, T., & Pietikäinen, M. (2003). *Multi-scale binary patterns for texture analysis.* Paper presented at the 13th Scandinavian conference on Image analysis.

Maturana, D., Mery, D., & Soto, Á. (2009). *Face Recognition with Local Binary Patterns, Spatial Pyramid Histograms and Naive Bayes nearest Neighbor Classification.* Paper presented at the International Conference of the Chilean Computer Science Society.

Mazloom, M., & Ayat, S. (2008). *Combinational Method for Face Recognition: Wavelet, PCA and ANN.* Paper presented at the Digital Image Computing: Techniques and Applications (DICTA).

Mohamed, A. A., D'Souza, D., Baili, N., & Yampolskiy, R. V. (2011). *Avatar Face Recognition Using Wavelet Transform and Hierarchical Multi-scale LBP.* Paper presented at the 10th International Conference on Machine Learning and Applications and Workshops (ICMLA).

Mohamed, A. A., Gavrilova, M. L., & Yampolskiy, R. V. (2012). *Artificial Face Recognition Using Wavelet Adaptive LBP with Directional Statistical Features.* Paper presented at the International Conference on Cyberworlds (CW).

Mohamed, A. A., Gavrilova, M. L., & Yampolskiy, R. V. (2013). Recognizing Avatar Faces Using Wavelet-Based Adaptive Local Binary Patterns with Directional Statistical Features. *Transactions on Computational Science XVIII, 7848*, 137-154.

Mohamed, A. A., & Yampolskiy, R. V. (2011). *An improved LBP algorithm for avatar face recognition.* Paper presented at the XXIII International Symposium on Information, Communication and Automation Technologies (ICAT).

Mohamed, A. A., & Yampolskiy, R. V. (2012a). *Adaptive Extended Local Ternary Pattern (AELTP) for Recognizing Avatar Faces.* Paper presented at the 11th International Conference on Machine Learning and Applications (ICMLA), Boca Raton, FL, USA.

Mohamed, A. A., & Yampolskiy, R. V. (2012b). *Face recognition based wavelet transform and adaptive local binary pattern.* Paper presented at the 4th International Conference on Digital Forensics & Cyber Crime, Lafayette, Indiana, USA

Mohamed, A. A., & Yampolskiy, R. V. (2012c). *Using discrete wavelet transform and eigenfaces for recognizing avatars faces.* Paper presented at the International Conference on Computer Games (CGAMES), 2012 17th.

Mohamed, A. A., & Yampolskiy, R. V. (2012d). Wavelet-Based Multiscale Adaptive LBP with Directional Statistical Features for Recognizing Artificial Faces. *ISRN Machine Vision, 2012*.

Mohamed, A. A., & Yampolskiy, R. V. (2012e). *Wavelet Based Statistical Adapted Local Binary Patterns for Recognizing Avatar Faces.* Paper presented at the Advanced Machine Learning Technologies and Applications, Cairo, Egypt.

Mohamed, A. A., & Yampolskiy, R. V. (2013). *Recognizing Artificial Faces Using Wavelet Based Adapted Median Binary Patterns.* Paper presented at the The Twenty-Sixth International FLAIRS Conference, FL, USA.

Nefian, A. V., & III, M. H. H. (1998). *Hidden Markov models for face recognition*. Paper presented at the International Conference on Acoustics, Speech and Signal Processing

Neustaedter, C., & Fedorovskaya, E. (2009). *Presenting identity in a virtual world through avatar appearances*. Paper presented at the Graphics Interface, Kelowna, British Colombia, Canada

NuTao, T., Lei, H., & Changping, L. *Face Recognition Based on a Single Image and Local Binary Pattern.* Paper presented at the Second International Symposium on Intelligent Information Technology Application, 2008. IITA '08.

O'Brien, N. Spies watch rise of virtual terrorists. from http://www.news.com.au/top-stories/spies-watch-rise-of-virtual-terrorists/story-e6frfkp9-1111114075761

O'Harrow, R. (2008). Spies Battleground Turns Virtual. from: http://www.washingtonpost.com/wp-dyn/content/article/2008/02/05/AR2008020503144.html

Ojala, T., & Pietikäinen, M. (1996). A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition, 29*(1), 51-59.

Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24*(7), 971-987. doi: 10.1109/TPAMI.2002.1017623

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision, 42*(3), 145-175.

The ORL database of faces. from http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html

Oursler, J. N., Price, M., & Yampolskiy, R. V. (2009). *Parameterized Generation of Avatar Face Dataset.* Paper presented at the 14th International Conference on Computer Games: AI, Animation, Mobile, Interactive Multimedia, Educational & Serious Games, Louisville, KY, USA.

Paliwal, K. K., & Sharma, A. (2011). Aproximate LDA Technique for Dimensionality Reduction in the Small Size Case. *Journal of Pattern Recognition Research, 6*(2), 298-306.

Penev, P. S., & Atick, J. J. (1996). Local Feature Analysis: A General Statistical Theory for Object Representation. *Network-Computation in Neural Systems, 7*(3), 477-500.

Pentland, A., Moghaddam, B., & Starner, T. (1994). *View-Based and Modular Eigenspaces for Face Recognition.* Paper presented at the IEEE Conference on Computer Vision and Pattern Recognition.

Phillips, P. J., Martin, A., Wilson, C. L., & Przybocki, M. (2000). An Introduction to Evaluating Biometric Systems. *Computer, 33*(2), 56-63.

Poli, V., Arcot, N., & Charapanamjeri, J. (2009). Evaluation of Biometrics. *International Journal of Computer Science and Network Security, 9*(9), 261-269.

Raja, Y., & Gong, S. (2006). *Sparse Multiscale Local Binary Patterns.* Paper presented at the 17th British Machine Vision Conference, Edinburgh, United Kingdom.

Rara, H., Ali, A., Elhabian, S., Starr, T., & Farag, A. (2010). *Face Recognition at-a-Distance Using Texture, Dense- and Sparse-Stereo Reconstruction* Paper presented at the International Conference on Pattern Recognition (ICPR).

Second Life. from www.secondlife.com

Senior, A. W., & Bolle, R. M. (2011). Face recognition and its applcations *Achievements and Challenges in Fingerprint Recognition* (pp. 101-115).

Shengcai, L., Guoying, Z., Kellokumpu, V., Pietika, x, inen, M., & Li, S. Z. (2010). *Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes.* Paper presented at the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Sims Online. from www.thesims.ea.com

Singh, R., Vatsa, M., & Noore, A. (2008). Recognizing Face Images with Disguise Variations. In K. Delac, M. Grgic & M. S. Bartlett (Eds.), *Recent Advances in Face Recognition* (pp. 149-160).

Su, Y., Shan, S., Chen, X., & Gao, W. (2006). *Patch-based Gabor Fisher Classifier for Face Recognition.* Paper presented at the International Conference on Pattern Recognition.

Thompson, C. W. (2011). Next-Generation Virtual Worlds: Architecture, Status, and Directions. *Internet Computing, IEEE, 15*(1), 60-65. doi: 10.1109/MIC.2011.15

Trewin, S., Laff, M., Hanson, V., & Cavender, A. (2009). Exploring visual and motor accessibility in navigating virtual world. *ACM Transaction of Accessible Computing 2*(2), 1-35.

Turk, M. A., & Pentland, A. P. (1991). *Face recognition using eigenfaces.* Paper presented at the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Proceedings CVPR '91.,.

Turtinen, M., & Pietikäinen, M. (2006). *Contextual analysis of textured scene images.* Paper presented at the British Machine Vision Conference.

Wankou, Y., & Changyin, S. (2011). *Face recognition using improved local texture patterns.* Paper presented at the 9th World Congress on Intelligent Control and Automation (WCICA).

Wayman, J. L. (2001). Fundamentals of biometric authentication technologies. *International Journal of Image Graphics, 1*(1), 93-113.

Wen-Hung, L. (2010). *Region Description Using Extended Local Ternary Patterns.* Paper presented at the 20th International Conference on Pattern Recognition (ICPR).

Wen-Hung, L., & Ting-Jung, Y. (2010). *Texture Classification Using Uniform Extended Local Ternary Patterns.* Paper presented at the IEEE International Symposium on Multimedia (ISM).

Wencheng, W., Faliang, C., Jianguo, Z., & Zhenxue, C. (2010). *Automatic facial expression recognition using local binary pattern.* Paper presented at the 8th World Congress on Intelligent Control and Automation (WCICA).

Wiskott, L., Fellous, J., Kruger, N., & Malsburg, C. (1997). Face recognition by elastic bunch graph matching *IEEE Transactions on Pattern Analysis and Machine Intelligence, 19*(7), 775-779.

Wolf, L., Hassner, T., & Taigman, Y. (2008). *Descriptor based methods in the wild*. Paper presented at the Faces in Real-Life Images workshop at the European Conference on Computer Vision (ECCV).

Xiao-Jun, W., Kittler, J., Jing-Yu, Y., Messer, K., & Shitong, W. (2004). *A New Direct LDA (D-LDA)Algorithm for Feature Extraction in Face Recognition.* Paper presented at the 17th International Conference on Pattern Recognition.

Xiaoshan, L., Minghui, D., & Lianwen, J. (2010). *Face features extraction based on multi-scale LBP.* Paper presented at the 2nd International Conference on Signal Processing Systems (ICSPS).

Xiaoyang, T., & Triggs, B. (2010). Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions. *Image Processing, IEEE Transactions on, 19*(6), 1635-1650. doi: 10.1109/TIP.2010.2042645.

Yampolskiy, R. V. (2012). ICMLA Face Recognition Challenge. from http://www.icmlaconference.org/icmla12/.

Yampolskiy, R. V., Cho, G., Rosenthal, R., & Gavrilova, M. L. (2012). Experiments in Artimetrics: Avatar Face Recognition. *Transactions on Computational Science, 16*, 77-94.

Yampolskiy, R. V., Gyuchoon, C., Rosenthal, R., & Gavrilova, M. L. (2011). *Evaluation of Face Recognition Algorithms on Avatar Face Datasets.* Paper presented at the International Conference on Cyberworlds (CW).

Yampolskiy, R. V., Klare, B., & Jain, A. K. (2012). *Face recognition in the virtual world: recognizing avatar faces.* Paper presented at the 11th IEEE International Conference on Machine Learning and Applications, Boca Raton, FL, USA.

Yang, J., & Yang, J.-y. (2003). Why can LDA be performed in PCA transformed space? *Pattern Recognition, 36*(2), 563-566.

Yang, M. H. (2002). *Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods.* Paper presented at the IEEE International Conference on Automatic Face and Gesture Recognition

Zhang, B., Gao, W., Shan, S., & Peng, Y. (2004). *Discriminant Gaborfaces and Support Vector Machines Classifier for Face Recognition*. Paper presented at the Asian Conference on Computer Vision, Jeju Island, Korea.

Zhang, L., Li, S. Z., Qu, Z. Y., & Huang, X. (2004). *Boosting Local Feature Based Classifiers for Face Recognition* Paper presented at the Computer Vision and Pattern Recognition Workshop.

Zhao, G., & Pietikainen, M. (2007). Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29*(6), 915-928.

Zheng, Y., & Elmaghraby, A. (2011). *A Brief Survey on Multispectral Face Recognition and Multimodal Score Fusion.* Paper presented at the IEEE International Symposium on Signal Processing and Information Technology (ISSPIT).

Zhenhua, G., Lei, Z., Zhang, D., & Su, Z. (2010). *Rotation invariant texture classification using adaptive LBP with directional statistical features.* Paper presented at the 17th IEEE International Conference on Image Processing (ICIP).

Zhenhua, G., Lei, Z., Zhang, D., & Xuanqin, M. (2010). *Hierarchical multiscale LBP for face and palmprint recognition.* Paper presented at the 17th IEEE International Conference on Image Processing (ICIP).

CURRICULUM VITA

Name: Abdallah Abd-Elghafar Mohamed

Address: Department of Computer Engineering and Computer Science,
Cybersecurity Lab, J.B. Speed School of Engineering,
Duthie Center for Engineering 239,
University of Louisville,
Louisville, Kentucky, USA 40292

Phone: (502) 852-6304

Fax: (502) 852-4713

Email: abdmoham@umail.iu.edu / aamoha04@louisville.edu

DOB: Menoufia, Egypt – August 17, 1974

EDUCATION

1/2010- 12/2013 University of Louisville, Ph.D., Department of Computer Engineering and Computer Science.

8/2007- 12/2009 Indiana University Bloomington, Master of Science in Computer Science, School of Informatics and Computing.

9/1999-12/2002 Menoufia University, Egypt, Master of Science in Computer Science, Mathematics and Computer Science Department, Faculty of Science.

9/1992-5/1997 Menoufia University, Egypt, Bachelor of Science in Pure Mathematics and Computer Science, Mathematics and Computer Science Department, Faculty of Science.

ACADEMIC APPOINTMENTS

Instructor of Computer Science, Department of Computer Engineering and Computer Science J.B. Speed School of Engineering, University of Louisville, Louisville, KY, USA. 2012-2013

Teaching assistant of Computer Science, Department of Computer Engineering and Computer Science J.B. Speed School of Engineering, University of Louisville, Louisville, KY, USA. 2010-2012

Assistant Lecturer of Computer Science, Mathematics and 2002- Present

Computer Science Department, Faculty of Science,
Menoufia University, Egypt.

Teaching Assistant of computer Science, Mathematics and Computer Science Department, Faculty of Science, Menoufia University, Egypt. 1999-2002

Assistant Researcher, Operations and Information Department, National Center for Examinations and Educational Evaluation, Cairo, Egypt. 1998-1999

PUBLICATIONS

JOURNAL PAPERS:

1. **A. A. Mohamed** and R. V. Yampolskiy, "Wavelet-Based Multiscale Adaptive LBP with Directional Statistical Features for Recognizing Artificial Faces," *ISRN Machine Vision,* vol. 2012, 2012.
2. **A. A. Mohamed**, M. L. Gavrilova and R.V. Yampolskiy, "Recognizing Avatar Faces Using Wavelet-Based Adaptive Local Binary Patterns with Directional Statistical," *Transactions on Computational Science XVIII,* 2013, pp. 137-154.
3. R. V. Yampolskiy, N. Ali, D. D'Souza, and **A. A. Mohamed** "*Behavioral Biometrics: Categorization and Review,*" International Journal of Natural Computing Research (IJNCR) 2013, in press.
4. **A. A. Mohamed** and R. V. Yampolskiy, "Face Recognition using Wavelet-based Multi-scale Statistical Adapted Local Binary Patterns," under preparation.
5. **A. A. Mohamed** and R. V. Yampolskiy, "A novel Multi-scale Local Ternary Pattern Representation for Recognizing Faces," under preparation.

BOOK CHAPTERS:

1. **A. A. Mohamed** and R. V. Yampolskiy," Avatar Facial Biometric Authentication Using Wavelet Local Binary Patterns," *Security and Privacy Preserving in Social Networks*, Chapter 10, Lecture Notes in Social Networks. Springer-Verlag Wien, 2013, pp. 317-335.
2. **A. A. Mohamed** and R. V. Yampolskiy,"Face Recognition using A novel Local Binary Patterns Generation" Recognizing Faces" under preparation.

CONFERENCE PAPERS:

1. **A. A. Mohamed** and R. V. Yampolskiy, "An improved LBP Algorithm for Avatar Face Recognition," *XXIII International Symposium on, Information, Communication and Automation Technologies (ICAT),* Sarajevo, Bosnia & Herzegovina, 2011, pp. 1-5.
2. **A. A. Mohamed**, D. D'Souza, N. Baili, and R. V. Yampolskiy, "Avatar Face Recognition Using Wavelet Transform and Hierarchical Multi-scale LBP," *10th International Conference on Machine Learning and Applications (ICMLA),* Honolulu, Hawaii, 2011, pp. 194-199.
3. M. Boukhris, **A. A. Mohamed**, D. D'Souza, M. Beck, N. E. Ben Amara, and R. V. Yampolskiy, "Artificial Human Face Recognition via Daubechies Wavelet Transform and SVM," *16th International Conference on Computer Games (CGAMES),* Louisville, KY, 2011, pp. 18-25.
4. **A. A. Mohamed**, M. L. Gavrilova, and R. V. Yampolskiy, "Artificial Face Recognition Using Wavelet Adaptive LBP with Directional Statistical Features," *International Conference on Cyberworlds (CW),* Darmstadt, Germany, 2012, pp. 23-28.
5. **A. A. Mohamed** and R. V. Yampolskiy, "Face Recognition Based on Wavelet Transform and

Adaptive Local Binary Pattern," *4th International Conference on Digital Forensics & Cyber Crime*, Lafayette, Indiana, USA, 2012.

6. **A. Mohamed** and R. V. Yampolskiy, "Using Discrete Wavelet Transform and Eigenfaces for Recognizing Avatars Faces," *17th International Conference on Computer Games (CGAMES),* Louisville, KY, 2012, pp. 143-147.
7. **A. Mohamed** and R. V. Yampolskiy, “Adaptive Extended Local Ternary Pattern for Recognizing Avatar Faces,” *International Conference on Machine Learning and Applications*, Boca Raton, FL, USA, 2012, pp. 57-62.
8. **A. A. Mohamed** and R. V. Yampolskiy, "Wavelet based Statistical Adapted Local Binary Patterns for Recognizing Avatar Faces,” *Advanced Machine Learning Technologies and Applications Conference (AMLTA 2012), Communications in Computer and Information Science series 322*, Cairo, Egypt, 2012, pp. 92-101.
9. M. Korayem, **A. A. Mohamed**, D. Crandall, and R. V. Yampolskiy, "Learning Visual Features for the Avatar CAPTCHA Recognition Challenge," *International Conference on Machine Learning and Applications*, Boca Raton, FL, USA, 2012, pp. 584-587.
10. M. Korayem, **A. A. Mohamed**, D. Crandall, and R. V. Yampolskiy, "Solving Avatar CAPTCHAs Automatically," *Advanced Machine Learning Technologies and Applications Conference (AMLTA 2012), Communications in Computer and Information Science series 322*, Cairo, Egypt, 2012, pp. 102-110.
11. **A. A. Mohamed** and R. V. Yampolskiy, " Recognizing Artificial Faces using Wavelet Based Adapted Median Binary Patterns,” *26th International Florida Artificial Intelligence Research Society Conference* (FLAIRS 26), *AAAI Press,* St. Pete Beach, Florida, USA, 2013, pp. 286-291.

RESEARCH INTERESTS

Computer Vision, Image Processing, Artificial Intelligence, Biometrics, Machine Learning, Data Mining.

GRANT AND AWARDS

1. Best Excellent undergraduate student in Faculty of Science, Menoufia University, Egypt, 1997
2. Four-year (2007-2011) Ph. D. Scholarship, Ministry of Higher Education, The Arab Republic of Egypt.