ANALYSIS OF HIGH DEFINITION AND STANDARD DEFINITION VIDEO-BASED FACE RECOGNITION FOR MULTIMEDIA APPLICATIONS

BY
WASSEEM N. IBRAHEM AL-OBAYDY

SUPERVISED BY
Dr. HARIN SELLAHEWA

A Thesis
Submitted to the Department of Applied Computing in the University of Buckingham in Partial Fulfillment of the Requirements for Master Degree of Science in Innovative Computing

November 2010

Buckingham, United Kingdom

Dedication

I dedicate this work to my precious parents, brother and sister in appreciation for their love, prayer, unlimited support, and encouragement to complete my higher education. I dedicate this work also to my wife for her continuous love, support and patience in my absence, and to my lovely daughter.

Wasseem

Abstract

Problematic issues including illegal immigration, and the threat to the security of homelands and public areas, as well as the need to control access to buildings, networks and services, have increased the demand for reliable face recognition systems that recognise individuals from a distance. In applications in which the faces are remotely recognised, the spatial resolution of images or video frames has a significant impact on the recognition performance. The assumption is that, increasing the image/video resolution could lead to improvements in recognition accuracy. In recent years, a new digital video standard called high definition (HD) video has been developed to provide high resolution video with a high quality picture. This study evaluates the influence of using high definition (HD) video, in comparison to standard definition (SD) video, to recognise faces from a distance.

A number of requirements and methodologies were needed to ensure an objective comparison. First, a new face video database of 20 subjects was collected at the University of Buckingham using a HD body worn video camera in both indoor and outdoor conditions. Each subject was recorded at varying distances to the camera. Second, three baseline face recognition algorithms, namely Eigenfaces, Fisherfaces and wavelet-based approach were used for the evaluation. The study indicates that using SD face video data captured at a close range as a gallery set result in similar, if not better, recognition accuracy than a gallery set consisting of HD face data of the same range. However, HD video significantly outperforms SD video when the gallery set consists of face images captured from a distance. The latter scenario closely resembles real-life conditions of security (e.g. video surveillance), access control and law-enforcement applications.

Acknowledgements

First and foremost I would like to praise and thank **Allah** glorified and exalted who gave me the energy and ability to carry out my study. I would like to thank the **Ministry of Higher Education and Scientific Research in Iraq** for granting me a fully funded scholarship to obtain the MSc degree in the UK. Special thank to the **Iraqi Embassy/Cultural Department** in London for their efforts and support during my study. I would like to express my deepest gratitude to my supervisor **Dr. Harin Sellahewa** for his effort, support, guidance and his patience through long discussions over the past few months. Moreover, I would like to thank **Prof. Sabah A. Jassim**, Head of the Applied Computing Department, University of Buckingham, for his support and encouragement. Many thanks to **Mr. Hongbo Du**, Deputy of Head, **Dr. Naseer Al-Jawad**, Director of the MSc programme, and **Dr. Ihsan Lami** whose experience and valuable advice I benefited from.

Special thanks to **Mrs. Julie Leach** and **Sharon Taylor**, the two departmental secretaries for their help during my study. I am extremely grateful to the participants who contributed to the face video database in this project. I would like to thank all MSc and PhD students at the Department of Applied Computing, University of Buckingham for their help and support. Finally, thank you all.

Wasseem

Declaration

This work has not previously been submitted towards any qualification, degree or diploma in any university. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

Wasseem N. Al-Obaydy
November, 2010

Table of Contents

List of Figures

List of Tables

List of Acronyms

ATSC	Advanced Television System Committee
CCTV	Closed Circuit Television
DWT	Discrete Wavelet Transform
EBGM	Elastic Bunch Graph Matching
FLD	Fisher's Linear Discriminant
GUI	Graphical User Interface
HD	High Definition
HE	Histogram Equalisation
HH	High-High (used for wavelet subband)
HL	High-Low (used for wavelet subband)
HMM	Hidden Markov Model
LDA	Linear Discriminant Analysis
LH	Low-High (used for wavelet subband)
LL	Low-Low (used for wavelet subband)
MIR	Multimedia Information Retrieval
MOV	QuickTime Movie (movie file extension)
NTSC	National Television System Committee
PAL	Phase Alternating Line
PCA	Principal Component Analysis
PDBNN	Probabilistic Decision-Based Neural Network
PGM	Portable Gray Map (graphic file extension)
PIN	Personal Identification Number
SD	Standard Definition
SR	Super Resolution
SVM	Support Vector Machine
ZN	Z-score Normalisation

Chapter 1

Introduction

Over the last decade of the 20th century, multimedia emerged as a new research area and an arising new industry. Many publications defined this exciting field in several different ways. The most popular definition is the integration of multiple modalities including text, graphics, images, audio, video and animation in a meaningful way that lets the user navigate, interact, create and communicate (Vaughan 2008). Today, multimedia technologies have significantly evolved and have become the foundation of a tremendous number of applications. Movies and video games are examples of rich entertaining applications in audio, video, images and graphics. Financial services, business training and advertising, are also based on using a variety of media types (Hofstetter 1995). Electronic documents (e.g. sheets, books and papers) and distance-learning use multimedia for delivering information and knowledge (International Society for Technology in Education 2010, Georganas 1997). Furthermore, medical imaging (e.g. computerized tomography) is widely used in medical diagnosis, and virtual surgery can be used for medical training (Hofstetter 1995). Video conferencing and video-on-demand are other applications that are used in multimedia communications (Heath 1999). Biometrics-based recognition systems and video surveillance systems are other applications that depend on using multimedia in their operation (Golshani 2008, Cucchiara 2005).

Along with this diversity in multimedia applications, many fundamental challenges face multimedia systems. For example, in communications the major problem is the bandwidth required for transmitting various types of media (Heath 1999). Synchronising multiple media types, especially those coming from distributed sources, is another major difficulty in constructing sophisticated multimedia presentations (Bertino and Ferrari 1998). Another challenge is how to enable or improve multimedia

information retrieval (MIR) depending on media content when text annotations are not available. One of the application areas that belongs to the MIR is the automatic pictorial search in a database of imagery, for example detecting and recognising human face from still images or video sequences stored in a multimedia database (Lew, et al. 2006).

The automated face biometric recognition from an image or video sequence is a fundamental topic in computer vision, robotics and image science. However, it is the most challenging one due to several factors (Kung, Mak and Lin 2005). Section 1.3 highlights these factors. The spatial resolution of an image/video, which is one of these factors, significantly affects the recognition performance especially when the recognition is performed from a distance. The hypothesis that will be tested is that increasing the image/video resolution can improve the performance of face recognition from a distance.

Recently, high definition (HD) video has been introduced as a new video standard that provides high quality video with high resolution. In this thesis, we will investigate the use of high definition (HD) and standard definition (SD) video in face recognition at different distances from the camera. A comparative study of the use of HD vs. SD video data will be conducted using three benchmark face recognition schemes 1) Eigenfaces, 2) Fisherfaces and 3) wavelet-based face recognition. The performance of HD and SD video will be examined for face recognition through analysing and evaluating the experimental results.

The remaining part of this chapter provides a general explanation of basic concepts that are relevant to this work and describes the objective and the outline of this thesis.

1.1 Digital Video

Video is a time-ordered series of still images that are displayed at a rate fast enough to provide the illusion of a continuous motion (Dick 2002). The term “digital video” refers to the acquisition, processing, displaying and storage of the sequential still images in a digital format (Webopedia Online Dictionary 2010). The digital representation of video also contains the audio data that are recorded digitally along with the video. Each still image in video sequence is called a frame, and the number of individual frames that are displayed each second is called frame rate. The higher the frame rate, the smoother the

perceived motion is. Frame rate is often measured in frames per second (fps), and the minimum acceptable frame rate for actual motion is 15 fps (Dick 2002). In addition to frame rate, digital video has a number of other essential characteristics including resolution, size and standards. These features are briefly outlined in the following paragraphs.

Resolution is a concept that describes how finely a digital image or device closes to the continuous (analogue) form using a finite number of pixels (Chapman and Chapman 2004). The higher the resolution, the finer the observed details are. The resolution of digital video is determined by the size of the frame (image), measured in pixels. For example, the size of a PAL video frame is 720 × 576 pixels. It is necessary to distinguish between the image/video resolution and the resolution of display device (i.e. screen resolution). The former shows the amount of detail that is contained in the image or video frame; whereas the latter shows the maximal image/frame size that can be displayed on the screen (Chapman and Chapman 2004).

Since the digital video is represented as a sequence of consecutive digital images, it requires a huge amount of storage capacity. The size per second for uncompressed digital video is calculated by the product of bit colour depth, frame resolution, and frame rate (Dick 2002). For example, a full-colour (i.e. 24-bits) digital video with frame resolution 640 × 480 pixels at a 30 fps would require just over 26 MB for one second, and about 1.6 GB for one minute. Such figures are impractical not only for storage, but also for the transmission of digital video; hence digital video always requires compression techniques (Heath 1999, Chapman and Chapman 2004).

The standardisation of digital video is based on the two major analogue video standards: National Television System Committee (NTSC) and Phase Alternating Line (PAL). Like any other analogue signals, these two types of analogue video signals are translated to the digital form by the sampling process (Chapman and Chapman 2004). In the mid 1990s, the Advanced Television System Committee (ATSC) introduced a new digital video standard called high definition (HD) video (Li and Drew 2004). With the advent of HD video, the term standard definition (SD) video is now used to refer to the NTSC and PAL video formats (Apple 2010). The key features of SD and HD video will be described in more detail in the next sections, and summarised in Table 1.1.

1.1.1 Standard Definition Video

Today, the formats of NTSC, PAL and any video with vertical resolution less than 720 pixels, are classified as standard definition (SD) video formats. Originally, NTSC and PAL are analogue standards, the digital representation of these standards can be obtained by digitising (sampling) the video frames. The NTSC video frame is digitised to 640 × 480 pixels, while a PAL video frame is sampled to 768 × 576 pixels (Chapman and Chapman 2004). Both NTSC and PAL videos have a 4:3 aspect ratio (i.e. the ratio of picture width to height), and follow the interlaced scanning system. The actual frame rate of NTSC video is 29.97 fps but it is often quoted as 30 fps, whereas the frame rate of PAL video is 25 fps (Chapman and Chapman 2004).

1.1.2 High Definition Video

In recent years, the increasing demand to high quality video has accelerated the adoption of high definition (HD) digital video. The term "high definition" points out to the high quality image, video and audio formats. High definition (HD) video is any video that contains 720 or more horizontal lines of the vertical resolution of the video frame. ATSC stated that the frame size of the HD video is either 1280 × 720 or 1920 × 1080 pixels (Browne 2006). All HD video formats support a widescreen aspect ratio of 16:9. Thus, HD video provides high quality picture with high spatial resolution than the SD video. HD video with 720 pixels supports only progressive scanning, and is denoted by (720p), while HD video with 1080 pixels supports both interlaced and progressive scanning, and is denoted by 1080i and 1080p respectively (Apple 2010). For more details about the types of scanning systems, the reader is referred to the "digital multimedia" book complied by Chapman and Chapman 2004. Unlike the SD video, HD video offers a variety of frame rates including 24, 30, and 60 fps (Browne 2006).

Feature	SD	HD
Vertical resolution	< 720	>= 720
Frame rates	25, 30	24, 30, 60
Aspect ratio	4:3	16:9
Scanning system	interlaced	interlaced, progressive

Table 1.1 The key features of SD and HD video

1.2 Biometrics

Over the last two decades, traditional techniques involving keys, passwords, Personal Identification Numbers (PIN), and ID cards have been widely used to determine the identities of individuals. However, many problems have accompanied the use of such techniques, for example misuse, loss, theft, and forgetfulness. Furthermore, such techniques cannot recognise authorised persons from impostors who illegally possess these means of identification. As a result, there was a need to use a new approach based on the unique physical traits of humans to authenticate and recognise them. The traits can be physiological (e.g. face, iris and fingerprint) or behavioural (e.g. voice, handwritten signature and gait). This identification approach is called biometrics (Jain, Bolle and Pankanti 1999).

The term biometrics refers to the unique physiological and behavioural characteristics such as the face, iris, voice, and handwritten signature that can be used to identify and verify individuals (Jain, Bolle and Pankanti 1999). Biometrics technology involves different modalities, for example face recognition, iris recognition, voice verification and signature verification. Multimedia (e.g. images, audio and video) plays a crucial role in computational biometrics, for example human faces can be recognised from still images or video sequences, and gait features can also be extracted from video footage of a walking person (Golshani 2008).

Biometrics can be used in large-scale applications. It can be used together with multimedia surveillance systems to detect remote objects (e.g. persons and vehicles) and extract visual features (e.g. faces and vehicle license plates). Many enterprises can benefit from using multimedia-based biometric technology, for example law-enforcement, financial transactions, government institutions, and border agencies (Golshani 2008).

1.3 Face Recognition

Face recognition is an unobtrusive and reliable biometric identification (authentication) modality which gives the machine the ability to identify humans depending on their facial characteristics (Kung, Mak and Lin 2005, Jain, Bolle and Pankanti 1999). There are a number of reasons why face recognition has become more attractive to many researchers in different areas, such as image processing, pattern recognition, computer

vision, and computer security. The first is that the face is preferable to other types of biometrics for a number of aspects. Firstly, the acquisition of face images is easier even from a certain distance. Secondly, the face conveys, in addition to a person's identity, many other features, for example emotion, age and sex (Park 2009). The second reason is that face recognition has a user-friendly nature, and can be used in a wide range of applications. Such applications vary from the matching of still face images (e.g. photographs in passports and driving licenses) to real-time matching in video sequences. The latter involves surveillance systems (e.g. power grid surveillance, nuclear plant surveillance and Closed Circuit Television (CCTV) control), access control (e.g. border-crossing control, facility access and computer/network access), and security systems in public areas (e.g. airports and stadiums) (Kung, Mak and Lin 2005, Huang, Xiong and Zhang 2005).

However, there are a number of challenges that confront the automated recognition of faces making it a very difficult task for researchers. These include illumination and pose variations, image/video resolution, occlusion, background complexity, facial expression (e.g. blinking, speech and emotion), age variations, facial hair, makeup, and variations of acquisition devices (e.g. noise and distortion) (Kung, Mak and Lin 2005, Park 2009).

1.4 Purpose of the thesis

The aim of this thesis is to investigate whether the HD video will provide better face recognition results than the SD video at different distances from the camera. The motivation to present this study is the growing demand to recognise the human face from a distance, for example face recognition in video surveillance systems and access control applications. The study will also lead to a better understanding of the significance of image/video resolution in face recognition from a distance. In order to achieve an objective comparison, a number of procedures were carefully considered:

- A new face video database was acquired using a HD body worn video camera, and an evaluation protocol was designed to conduct the experiments.
- Three baseline face recognition schemes: Eigenfaces, Fisherfaces and wavelet-based face recognition algorithms were applied.

1.5 Contributions of this work

The main contributions of this thesis include:

- A newly acquired face video database is presented.
- An evaluation of the use of HD video and SD video for face recognition from a distance is introduced.
- W. Al-Obaydy and H. Sellahewa, "Evaluation of High-definition and Standard-definition Video in Face Recognition from a distance", was submitted to SPIE conference for Defense, Security and Sensing, *Biometric Technology for Human Identification VIII* (Conference DS108), Florida USA 2011.
- H. Sellahewa and W. Al-Obaydy, "Performance Evaluation of High-definition Video in Face Recognition from a Distance", was submitted to IEEE Computer Vision and Pattern Recognition (CVPR) 2011 conference, Colorado USA.
- W. Al-Obaydy and H. Sellahewa, "On using High-Definition Body Worn Cameras for Face Recognition from a Distance", was submitted to Biometrics and Identity Management Workshop (BioID 2011), Brandenburg Germany.

1.6 Thesis Outline

The overall structure of this thesis takes the form of six chapters, including this introductory chapter. The chapters have been organised as follows:

- **Chapter one** introduces a general overview about multimedia and fundamental concepts about digital video, high definition and standard definition video, biometrics and face recognition.
- **Chapter two** presents a literature review for the early and recent video-based face recognition approaches and gives a description for a number of the techniques used to improve the low resolution video in automatic face recognition.
- **Chapter three** provides a detailed explanation of the Eigenfaces, Fisherfaces and wavelet-based face recognition approaches.
- **Chapter four** describes the experimental data, evaluation protocol and the software implementation used to conduct the experiments.
- **Chapter five** presents a discussion and evaluation of the experimental results.
- **Chapter six** concludes the findings in this study and discusses the possible areas of the future research.

Chapter 2

Literature Review

This chapter gives a brief description of the components of an automatic face recognition system and lists the benefits and drawbacks of using video in face recognition. The chapter also introduces a review of the major initial and current video-based face recognition approaches that have been published in the literature. The revision concentrates only on 2-dimensional (2D) video-based methods. The advantages of video-based face recognition will be presented in Section 2.6. Finally, this chapter highlights the influential role of video resolution in the automated recognition of faces and outlines some approaches that have been proposed to address the low resolution problem in face recognition.

2.1 Face Recognition System

In general, a face recognition system consists of three major modules: (1) face detection, (2) face normalisation (preprocessing), and (3) face recognition. The face recognition stage involves two subtasks: (a) feature extraction, and (b) face classification (matching). In video-based face recognition systems an additional module called face tracking is needed to track the detected faces in the video sequence (Li and Jain 2005, Zhao, et al. 2003). Most of these systems detect the face in the first frame and track it through the rest of the video sequence. Then, the system performs the face recognition process on the frame that achieves normalised illumination, pose and size (Wang, Wang and Cao 2009).

2.2 Face Recognition in Video

Face recognition based on video has attracted considerable attention for more than two decades. One essential feature of video-based techniques is the abundance of information contained in the video sequence. Such information can be employed to improve the performance of face recognition. In comparison with a still image the video outperforms in two distinct properties. First, it contains redundant frames of the same subject. These manifold frames hold a variety of poses and illumination giving the chance to select the desired frame (e.g. frame containing near-frontal pose) to improve the recognition accuracy. Second, the video sequence provides temporal information besides the spatial information. The temporal information represents the information embedded in the object's motion (e.g. facial motion) in the video sequence (Park 2009). The spatial information and the temporal information can be exploited together (i.e. spatio-temporal information) to solve numerous problems in the field of video processing, for example video retrieval (DeMenthon and Doermann 2003) and video denoising (Zhu, Xue and You 2007). Recently, researchers have developed techniques based on the spatio-temporal information for video-based face recognition (Wang, Wang and Cao 2009). Section 2.3.1 highlights some of these approaches.

On the other hand, there are a number of major drawbacks encountering face recognition in video. First, the quality of video frames is poor in terms of low resolution and extreme variations in illumination, pose, facial expressions, occlusion and distance from camera. Such changes are increased when the video is acquired in an open environment and the human subjects are not cooperative. Second, the size of face image in the video is smaller than that in the still image due to the conditions of video capture, such as the distance between the camera and the subject. The small size of the face image not only increases the complexity of face recognition, but also has a negative influence on the accuracy of face detection and segmentation (Zhao, et al. 2003, Park 2009).

Furthermore, other problems including "out of focus", interlacing, and motion blur also affect the performance of face recognition in video. The "out of focus" problem occurs when the face is out of focus of the lens of acquisition devices. This problem makes the face image blur due to the aberrations of the imaging optics. The "out of focus" problem happens especially in the applications of face recognition from a

distance (Ao, et al. 2009). Interlacing causes motion artifacts that appear in the interlaced video frames when the object (e.g. face) moves quickly to different positions during the acquisition of each individual frame. Finally, the motion blur problem occurs when the camera is shaking or the object moves faster than the exposure of the camera (Ao, et al. 2009).

2.3 Video-based Face Recognition Approaches

Since the video represents a set of consecutive still images, the video-based face recognition can utilize the still image-based approaches. That is, after detecting and segmenting the face image from the video sequence still image-based face recognition algorithms can be applied (Zhao, et al. 2003, Wang, Wang and Cao 2009). Still image-based methods can be classified into two categories: (1) Holistic approaches which use the whole face region in the recognition process and (2) Geometric features-based methods which use the geometric measurements of local facial features such as eyes, nose and mouth to represent the face and recognise it.

The most famous holistic schemes are Eigenfaces approach presented by Turk and Pentland (1991) using Principal Component Analysis (PCA), and Fisherfaces approach developed by Belhumeur et al. (1997) using Fisher's Linear Discriminant Analysis (FLD or LDA). We will use these two approaches in our comparative study, and we will describe them in detail in Chapter 3. Lin et al. (1997) developed a holistic method based on a probabilistic decision-based neural network (PDBNN). This method focuses on the upper facial region (i.e. the eyebrows, eyes and nose excluding the mouth) to recognise faces. The idea behind that is to eliminate the impact of facial variations caused by the motion of the mouth.

Another holistic face recognition scheme that combines wavelet transform, support vector machine (SVM), and a clustering method was introduced by Luo, Zhang and Pan (2005). The wavelet transform was applied on both the probe and gallery images. The authors used the LL subband of the probe face image after 2-level decomposition as an input to the SVM. In SVM, rather than using the sequential search and recognition that took a long time, the authors used fast bisearch technique through applying a two sorts hierarchical-clustering on the LL subbands of gallery images to create a binary tree. Hence, face recognition can be quickly bisearched through the tree.

Sellahewa (2006) proposed a holistic wavelet-based face recognition approach for constrained devices. The wavelet filters that were experimented included Haar, Daubechie 4, and Antonini. The nearest neighbour classification method was used to recognise the probe face image. This approach will be used in our analytical study, and will be elaborated in section 3.3.

Geometric features-based face recognition approaches have also been proposed in the literature. Wiskott et al. (1997) introduced a geometric features-based face recognition method based on an elastic bunch graph matching technique (EBGM). The approach uses Gabor wavelet transform to represent face images as labeled graphs consisting of nodes located at fiducial points such as eyes, nose and mouth. Then, a similarity measure function is used to compare the graphs. Amira and Farrell (2005) developed a features-based face recognition system based on wavelet transform. Two techniques including "error tolerance" and "region tagging" segmentation algorithms were used for extracting facial features. The authors used three wavelet filters including Haar, Gabor and biorthogonal 9/7 filters. The approach involves applying a wavelet decompostion filter on the face image followed by segmenting the facial features from the subbands. After segmentation, the ratios of the features are calculated, for example nose height, mouth width and internal and external distances of the eyes. The recognition stage depends on matching the calculated ratios with the ratios of the stored facial features.

Although it is possible to apply still image-based techniques to recognise the human face in video, video-based approaches that take probe video sequence as input have also been developed. These approaches can be classified into two categories: (1) spatio-temporal information-based approaches, and (2) Statistic model-based approaches. The next two sections present a revision of the methods proposed in each category. Recent surveys of video-based face recognition approaches and techniques can be found in (Zhao, et al. 2003, Wang, Wang and Cao 2009).

2.3.1 Spatio-temporal information based approaches

Most recent approaches exploit both the spatial information and temporal information in video sequences for face recognition. Zhou, Krueger and Chellappa (2002) proposed a time series state space model to track and recognise the face simultaneously in the video

sequence. This probablistic approach uses two parameters: (1) tracking state vector, and (2) identity variable to characterise the kinematics and identity of humans in the probe video. The authors used a condensation algorithm to solve the time series state space, and estimate the joint posterior distribution of the state vector and identity variable. The joint distribution is then marginalised over the state vector to produce a robust estimation of the posterior distribution of the identity variable. To obtain an improved recognition, the posterior distribution of the identity variable is degenerated due to the propagation of identity and dynamics.

2.3.2 Statistic model based approaches

Statistical models like PCA, Linear Discriminant Analysis (LDA), and Hidden Markov Models (HMM) have been used in developing video-based face recognition techniques. Topkaya and Bayazit (2008) introduced a recognition technique based on using a subset of video frames called representative frames (i.e. frames containing faces in frontal poses). The purpose of selecting such frames is that faces in frontal poses provide more information for recognition. The authors used Haar-like features detector to automatically derive the representative frames from the video sequence and localising face regions. On the derived face images, PCA and LDA are applied to reduce the dimension of the data set. The recognition step is performed by applying support vector machine (SVM) algorithm on the compact data.

Another approach based on using adaptive HMM was presented by Liu and Chen (2003). In the training phase, a HMM learns the statistical properties as well as the temporal characteristics of the gallery video sequences of each subject. In the recognition phase, the HMM analyses the temporal properties of the probe video sequence of each subject and generates a probability score. Then, a comparison is performed among the probability scores generated by the HMMs. The identity associated to the highest score is assigned to the probe video sequence. The authors also proposed an adapted version of HMM by using the standard Maximum a Posteriori (MAP) adaptation technique. That is, after recognising the subject in the probe video sequence this sequence is used to update the HMM of that subject. The purpose of adaptation is to enhance the model of the subject and provide better recognition performance over time.

2.6 Video-based vs. Image-based Face Recognition

Face recognition in video outperforms the still image-based recognition in the following major points. First, the temporal information in video sequence can be exploited to improve the recognition performance. Second, efficient facial representations, such as 3D face model or super-resolution images, can be derived from the video frames and used to improve the recognition task. Finally, in a video sequence, the current and past frames can be used to learn and update the subject models over time to improve the recognition performance for the next frames (Liu and Chen 2003).

2.7 Video Resolution in Face Recognition

One of the major factors that significantly affects the performance of video-based face recognition systems is the spatial resolution of video frames. Normally, video signals acquired by digital imaging devices are digitised at resolution levels less than that of still images; hence the quality of a frame extracted from the video is lower than the desired level. In face recognition, when low-resolution face images are used as a probe and/or gallery set, the recognition performance may decrease to an unacceptable level (Arachchige 2008). A well-known example of low-resolution video applications is the video surveillance systems, such as CCTV. In such systems, the resolution of the acquired face images is very low, and their quality is degraded especially when these images are captured in crowded areas and from remote distances. Thus, developing sophisticated face recognition systems in such applications is a very difficult task (Jillela and Ross 2009, Choi, Ro and Plataniotis 2008).

Many techniques have been proposed to address the “low resolution” problem in video-based face recognition. The most popular method is super-resolution (SR) that is used to generate a high-resolution facial image from multiple low-resolution images or video frames. This technique involves three stages: image registration, interpolation, and restoration. The SR technique has been used in many previous occasions to improve the low resolution frames for face recognition. Wheeler, Liu and Tu (2007) proposed an Active Appearance Model (AMM) to provide frame-to-frame face registration that is used by the super-resolution algorithm to create a high-resolution facial image. The registration of face is required to align the face in video frames. The super-resolved face images were tested using a commercial face recognition engine, and the experimental

results showed improved recognition rates. Arachchige (2008) developed a face recognition system based on super-resolution method and principal component analysis (PCA). The author applied two SR techniques based on spatial and frequency domains to improve the resolution of face images. The resulted system can recognise faces in multiple low resolution images/frames.

Wang, Miao and Zhang (2008) suggested another method to extract a high resolution face image from low resolution video sequences. The method consists of two parts: face detection and face hallucination. After detecting the face in the video sequence, an eigentransformation by PCA was applied for face hallucination, resulting in a global high resolution face. The authors also presented a Coupled PCA method to obtain facial local residue. The combination of global- and local-face generates a high resolution face image that can be used for recognition.

Although many techniques have been introduced to optimise the recognition performance in low resolution video facial frames, these techniques may have a number of limitations. For example, the super-resolution method always requires multiple facial frames for the same person acquired in the same scene. Such frames are not always obtainable in some practical applications such as face annotation in the low resolution video-clips on the web (Choi, Ro and Plataniotis 2008). Second, as the spatial resolution of video frames declines, super-resolution tends to be more susceptible to environmental variations, and causes distortion that extensively affects recognition performance (Hennings-Yeomans, Baker and Kumar 2008). Finally, super-resolution routine causes artifacts (i.e. noisy pixels) in the super-resolved output image when there are large changes in facial poses in the low resolution frames. Such changes in the facial pose result from substantial motion occurred over a short time. These changes cause incorrect frame registration resulting in artifacts (Jillela and Ross 2009).

In order to overcome the drawbacks of such techniques, the use of high definititon (HD) video has recently become the most appropriate solution to addresss the low resolution problem in face recognition (Ao, et al. 2009). One of the state-of-art studies that has scrutinised the performance of HD video in face recognition is the research carried out by Thomas et al. (2008). The authors used video data captured by three types of cameras: SD, HD, and webcamera. In these video clips, the subjects were acquired at a short distance from the camera. The authors showed that the use of HD

video in both probe and gallery sets gives the best recognition performance among the other combinations. Thomas et al. also showed that the recognition performance improves as the number of frames per subject increases. In our work, we will evaluate the recognition performance in HD and SD video at four different distance ranges from the camera.

Chapter 3

Eigenfaces, Fisherfaces and Wavelet-based Face Recognition Approaches

This chapter gives an exhaustive description of the face recognition approaches used in this study. The statistical approaches: Eigenfaces and Fisherfaces will be described in Sections 3.1 and 3.2 respectively, while Section 3.3 explains the wavelet-based approach. A summary of the main characteristics of these three approaches is shown in Table 3.1.

3.1 The Eigenfaces Approach

Turk and Pentland (1991) originated a method for face recognition called the Eigenfaces approach, based on a technique proposed by Sirovich and Kirby (1987), for efficient representation of face images using the Principle Component Analysis (PCA). PCA is a statistical analysis tool used to reduce the large dimensionality of data, and to extract features. In this approach, each face image in the high dimensional image space (i.e. training set) can be represented as a linear combination of a set of vectors in the new low dimensional face space. These vectors, calculated by PCA, are the eigenvectors of the covariance matrix of the face images in the training set. Each eigenvector can be displayed as a ghostly face image, hence eigenvectors are called Eigenfaces. When a probe face image is introduced to the system to be recognised, it is projected into the face space and then a nearest neighbour classification method is used to assign the identity to the probe image. The following three sections explain this approach in more detail.

3.1.1 Calculating Eigenfaces

Let an 8-bit face image $f(x,y)$ be a two-dimensional array of size $N \times N$. The face image f may also be represented as a column vector of dimension N^2 or as a point in N^2-dimensional space. For example a face image of size 128 × 128 pixels can be translated into a vector of dimension 16384, or a point in 16384-dimensional space.

Let the set of M face images $f_1, f_2, f_3, \ldots f_M$ be a training set. The average face f_A, of this set can be defined as:

$$f_A = \frac{1}{M} \sum_{n=1}^{M} f_n \tag{1}$$

Figure 3.1 shows an example training set of 16 face images taken from the ORL database which is publicly available online at AT&T Laboratories Cambridge (2002). The average face of these training images is shown in figure 3.2.

Figure 3.1 Training face images from ORL database

Figure 3.2 The average face f_A

The difference between each training face image and the average face is denoted by the vector Φ_n, where:

$$\Phi_n = f_n - f_A \quad (n = 1,2, \ldots, M)$$

This set of very large M vectors is then subjected to the Principal Component Analysis (PCA) which finds a set of M orthonormal vectors, u_n, which best describe the distribution of the training images.

The $N^2 \times N^2$ covariance matrix $\boldsymbol{C}$ is defined as:

$$C = \frac{1}{M}\sum_{n=1}^{M}\Phi_n\Phi_n^T \tag{2}$$

$$= AA^T \tag{3}$$

where the matrix $A = [\Phi_1\ \Phi_2\ \ldots\ \Phi_M]$. The eigenvalues and eigenvectors can be defined as:

$$\lambda_k = \frac{1}{M}\sum_{n=1}^{M}(u_k^T\Phi_n)^2 \tag{4}$$

$$u_l^T u_k = \delta_{lk} = \begin{cases} 1, & \text{if } l = k \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where the vectors u_k and the values λ_k are eigenvectors and eigenvalues of the covariance matrix $\boldsymbol{C}$ in Eq. (2).

However, the size of the covariance matrix $\boldsymbol{C}$ is $N^2 \times N^2$ (e.g. 16384 × 16384), thus calculating the N^2 eigenvectors and eigenvalues of such a very large matrix is a very complex task for typical face image sizes. If the number of face images M in the training set is less than the dimension of the space (N^2), there will be only $M - 1$ useful eigenvectors. Turk and Pentland (1991) found a solution for the above problem by first solving the eigenvalues and eigenvectors of an $M \times M$ matrix rather than an $N^2 \times N^2$

matrix, and then taking a linear combination of the face images Φ_n to find the M eigenfaces.

Consider the eigenvectors v_i of the $M \times M$ matrix $A^T A$ such that:

$$A^T A v_i = \mu_i v_i \tag{6}$$

Premultiplying both sides by A we have:

$$A A^T A v_i = \mu_i A v_i \tag{7}$$

From Eq. (7) we can see that Av_i are the eigenvectors of the covariance matrix $\boldsymbol{C} = AA^T$. Thus, instead of calculating the N^2 eigenvalues and their corresponding eigenvectors of the huge covariance matrix $\boldsymbol{C}$, we can calculate the M eigenvalues and eigenvectors of the matrix $A^T A$. Then, the resulted eigenvectors determine linear combination of the M face images in the training set to form the eigenfaces u_l.

Considering the previous analysis, we can construct the $M \times M$ matrix $L = A^T A$, where:

$$L_{mn} = \Phi_m^T \Phi_n \tag{8}$$

Then, we can calculate the M eigenvectors v_i of the matrix L. The eigenfaces u_l can be defined as:

$$u_l = \sum_{k=1}^{M} v_{lk} \Phi_k \qquad l = 1, \ldots, M \tag{9}$$

Figure 3.3 shows the 16 eigenfaces corresponding to the 16 training face images shown in Figure 3.1. Thus, we can see that the calculations are significantly decreased from the size of face images (N^2) to the number of training face images (M). The eigenfaces can now be used to reconstruct and identify the face images as described in the next two sections.

Figure 3.3 The 16 eigenfaces

3.1.2 Using Eigenfaces for Face Reconstruction

The eigenfaces, calculated as shown above, establish a basis set from which we can reconstruct each face image in the training set. Using all eigenfaces will produce accurate reconstruction of the face image. However, Sirovich and Kirby (1987) argued that only a smaller number of eigenfaces M' is sufficient for a good reconstruction which approximates to the original face image. Only the M' significant eigenvectors with the highest associated eigenvalues are selected to generate the M' eigenfaces. The number of eigenfaces M' is chosen heuristically and can vary from one face image to another.

When a face image f is presented to the system, it is projected into eigenface space by the following operation:

$$w_k = u_k^T (f - f_A) \tag{10}$$

where $k = 1, \ldots, M'$ and w_k represents the contribution (weight) of kth eigenface, u_k, to the face image f. The reconstructed face image f' can be obtained by a set of point-by-point multiplications and summations as follows:

$$f' = f_A + \sum_{k=1}^{M'} w_k u_k \tag{11}$$

Figures 3.4 and 3.5 illustrate two examples of reconstructed face images. The number of eigenfaces used in reconstructing the two images increases from left to right and top to bottom. It can be clearly seen that the number of eigenfaces required for accurate reconstruction of the face image in the first figure is different than that of the face image in the second figure.

Figure 3.4 Reconstruction of the 2nd training face image

Figure 3.5 Reconstruction of the 3rd training face image

3.1.3 Using Eigenfaces for Face Identification

Turk and Pentland (1991) noticed that the M' eigenfaces, calculated as described above, can also be used to recognise new face images. That is, each face image in the training set is projected onto the M'-dimensional eigenface space to obtain its weight [from Eq. (10)]. The resulted weights constitute the vector $\Omega^T = [w_1, w_2, \ldots, w_{M'}]$. When a probe face image is introduced to be identified, it is also projected onto the M'-dimensional eigenface space and its weight vector Ω is calculated in the same manner. A standard pattern recognition algorithm can now be used to compare between the weight vector Ω and each of the weights in Ω^T to find which known face class best describes the

unknown (probe) face image. The easiest way for identifying which face class gives the closest description of the probe face image is to find the face class *k* that gives the minimum Euclidean distance:

$$\varepsilon_k^2 = \left\| \Omega - \Omega_k \right\|^2 \tag{12}$$

where Ω_k is a vector showing the *k*th face class.

3.2 The Fisherfaces Approach

Belhumeur et al. (1997) developed a face recognition approach called Fisherfaces which is insensitive to illumination variations and facial expressions. The authors stated that since the training images are labeled with classes (i.e. individual identities), the class information can be exploited to build a more reliable method to reduce the dimensionality of the feature space. This approach is based on using class specific linear methods for dimensionality reduction and simple classifiers to produce better recognition rates than Eigenfaces method which does not use the class information for dimensionality reduction. Belhumeur et al. used a class specific method called Fisher's Linear Disriminant Analysis (FLD or LDA) to find a set of projecting vectors (i.e. weights) that best discriminate different classes. FLD achieves that objective by maximising the ratio of the *between-class* scatter to that of the *within-class* scatter. This approach will be explained as follows.

Let *M* face images be a training set of *C* individual classes $X_1, X_2, \ldots X_C$. Each class X_i has P_i sample images. The *between-class* scatter matrix is defined as

$$S_B = \sum_{i=1}^{C} P_i (\mu_i - \mu)(\mu_i - \mu)^T$$

and the *within-class* scatter matrix is defined as

$$S_W = \sum_{i=1}^{C} \sum_{x_k \in X_i} (x_k - \mu_i)(x_k - \mu_i)^T$$

where μ_i is the average face image of class X_i, μ is the average face of the set of training images, and P_i is the number of sample images in class X_i.

The optimal projecting matrix W_{opt} is chosen as the matrix with orthonormal columns which maximises the ratio of the determinant of the *between-class* scatter matrix S_B of the projected samples to the determinant of the *within-class* scatter matrix S_W of the projected samples, i.e.

$$W_{opt} = \arg\max_{w} \frac{\left|W^T S_B W\right|}{\left|W^T S_W W\right|}$$

$$= [w_1\ w_2\ \ldots\ w_m]$$

If S_W is nonsingular, the projecting matrix W_{opt} can be generated by finding the generalised eigenvectors and eigenvalues of the matrix $S_W^{-1}.S_B$. The eigenvectors corresponding to the largest *C-1* eigenvalues are chosen to generate the projecting matrix W_{opt} and are called Fisherfaces. However, in face recognition applications, since the number of training images M is much smaller than the number of pixels N^2 in each image, the *within-class* scatter matrix S_W is more likely to be singular. As a result, the projecting matrix W is chosen such that the *within-class* scatter of the projected samples can be made zero. To overcome this problem, Belhumeur et al. proposed to project the high-dimensional training images onto a lower dimensional space so that the *within-class* scatter matrix S_W is nonsingular. The authors achieved that objective by using PCA to reduce the dimension to *M-C,* and then applying FLD on the PCA subspace spanned by the largest *M-C* eigenfaces to reduce the dimension to *C-1*.

3.3 Wavelet-based Face Recognition Approach

Sellahewa (2006) proposed an efficient face recognition approach based on discrete wavelet transform (DWT). The pyramidal scheme was used in this approach due to its computational efficiency over other decomposition schemes (e.g. standard and packet). In the enrolment stage, each face image in the training set is transformed to the wavelet domain to extract its facial feature vector (i.e. subband). The choice of an appropriate subband varies depending on the operational circumstances of the face recognition application. The decomposition level is predetermined based on the efficiency and accuracy requirements and the size of the face image. In the recognition stage, a nearest neighbour classification method was used to classifiy the unknown face images. Figure

3.6 illustrates the stages of this approach. The following paragraphs explain this approach in more detail.

Let the set $F = \{ f_{i,1}, f_{i,2}, f_{i,3}, \ldots, f_{i,m}\}$ be a training set of face images of n subjects, where each subject i has m images. In the enrolment stage, wavelet transform is applied on each training image so that a set $W_k(F)$ of multi-resolution decomposed images result. A new set $LL_k(F)$ of all k-level LL-subbands will be obtained from the transformed face images in the set $W_k(F)$. The new set $LL_k(F)$ forms the set of features for the training images. Thus, the training face image 1 of subject i ($f_{i,1}$) is expressed by its feature vector $LL_{k,i,1}$. The collection of feature vectors $LL_{k,i,1}$, $LL_{k,i,2}$, . . ., $LL_{k,i,m}$ represents the stored template of subject i. In a similar manner, 3 new sets $HL_k(F)$, $LH_k(F)$, $HH_k(F)$ can be created at the same decomposition level k from the kth-level HL-, LH-, HH-subbands respectively.

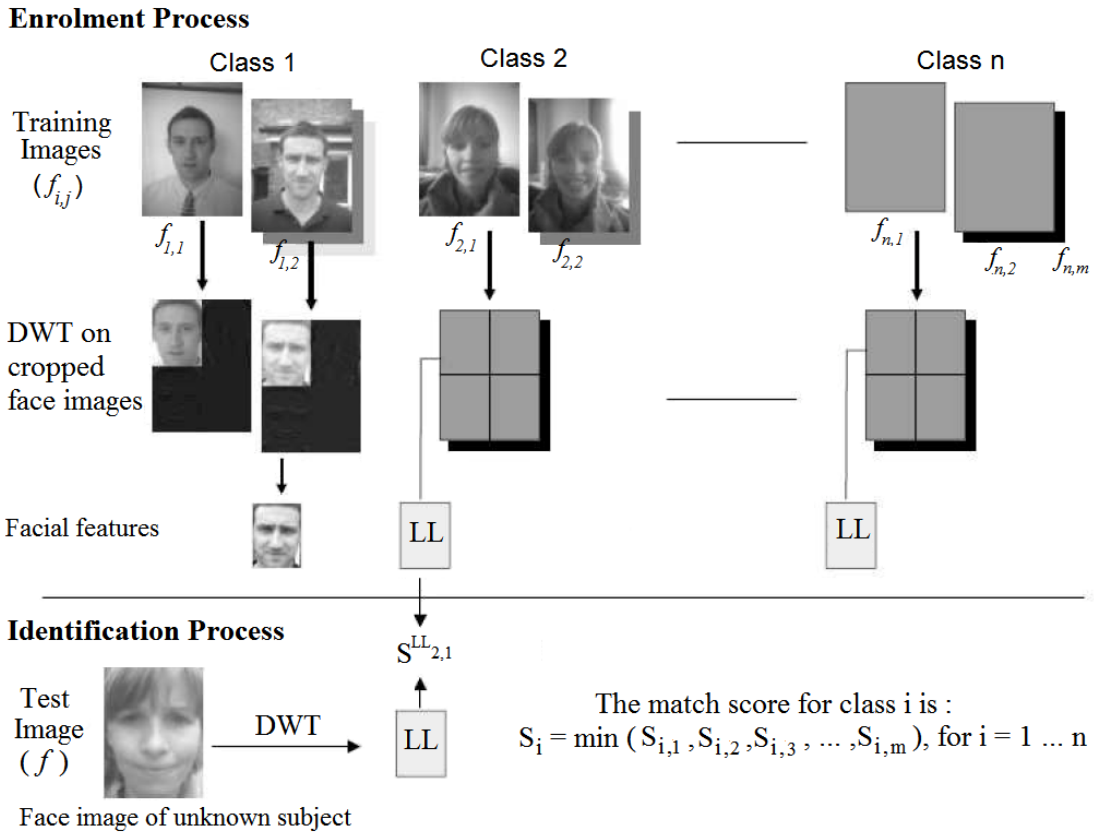


Figure 3.6 Wavelet-based face recognition
Source: (Sellahewa, 2006)

In the recognition phase, a nearest neighbour classifier is used to classify the input face image. When a probe face image is introduced to the system, it is

decomposed by wavelet transform, and a certain subband (e.g. LL_k) is chosen to represent the feature vector of the probe image. A match score $S_{i,j}$ can now be computed between the probe feature vector and each of the feature vectors j of the subject i in the feature set $LL_k(F)$. Then, the identity of the training image which gives the minimum score is assigned to the probe image:

$$S_i = \min (S_{i,j}) \quad (j = 1, \ldots, m)$$

Many similarity measures can be used for the nearest neighbour classifier, for example CityBlock, Euclidean, Cosine or Mahalanobis distance functions. The Haar, Daubechie-4, and Coiflet-1 wavelet filters with 3 decomposition levels per each subband, and CityBlock and Euclidean distance measures were used to implement this approach in our analytical study.

Characteristic	Eigenfaces	Fisherfaces	Wavelet-based
Dimensionality reduction technique	PCA	LDA	Wavelet transform
Use of class information	No	Yes	No
Facial representation	Projected vectors (weights)	Projected vectors (weights)	Subbands
Classification method	Nearest neighbour	Nearest neighbour	Nearest neighbour
No. of required face images per subject	>=1	>1	>=1

Table 3.1 Main characteristics of Eigenfaces, Fisherfaces and wavelet approaches

Chapter 4

Experimental Data, Evaluation Protocol and Software Development

Over the last two decades, a number of face image/video databases have been widely used by the research community for the purpose of training, testing and evaluating various face recognition schemes. The most popular databases include AT&T (formerly Olivetti Research Laboratory (ORL)), Yale, and Biometric Access control for Networked and e-Commerce Applications (BANCA). However, these databases were captured by vague acquisition devices that do not meet the new requirements of our project in terms of high definition video. Therefore, we acquired a new (UBHSD) face video database at the University of Buckingham using a HD body worn video camera. This chapter is devoted to describing the acquisition process and structure of the database as well as the experimental protocol used to evaluate the performance of HD and SD video in face recognition. A summary of the main features of the UBHSD database is shown in Table 4.1. The chapter also describes the software implementation used to conduct the experiments in this study.

4.1 Database Acquisition System

The videos in the UBHSD database were captured using one commercially available HD body worn camera (iOPTEC-P300) that is designed to be used by police officers and security agencies to provide covert/overt video surveillance. This camera can capture both HD and SD video recordings. Since the quality of video is affected by the quality of physical components (e.g. lens and sensors) of the camera, using two different HD and SD cameras will provide inconsistent video data. Therefore, we used the same HD camera to record both HD and SD videos. The SD video was recorded at a frame resolution 848 × 480 pixels, and frame rate 25 fps. The HD video was acquired at

a frame resolution 1920 × 1080p pixels, and frame rate 30 fps. Both the SD and HD videos were recorded in MOV file format.

4.2 Video Database Collection

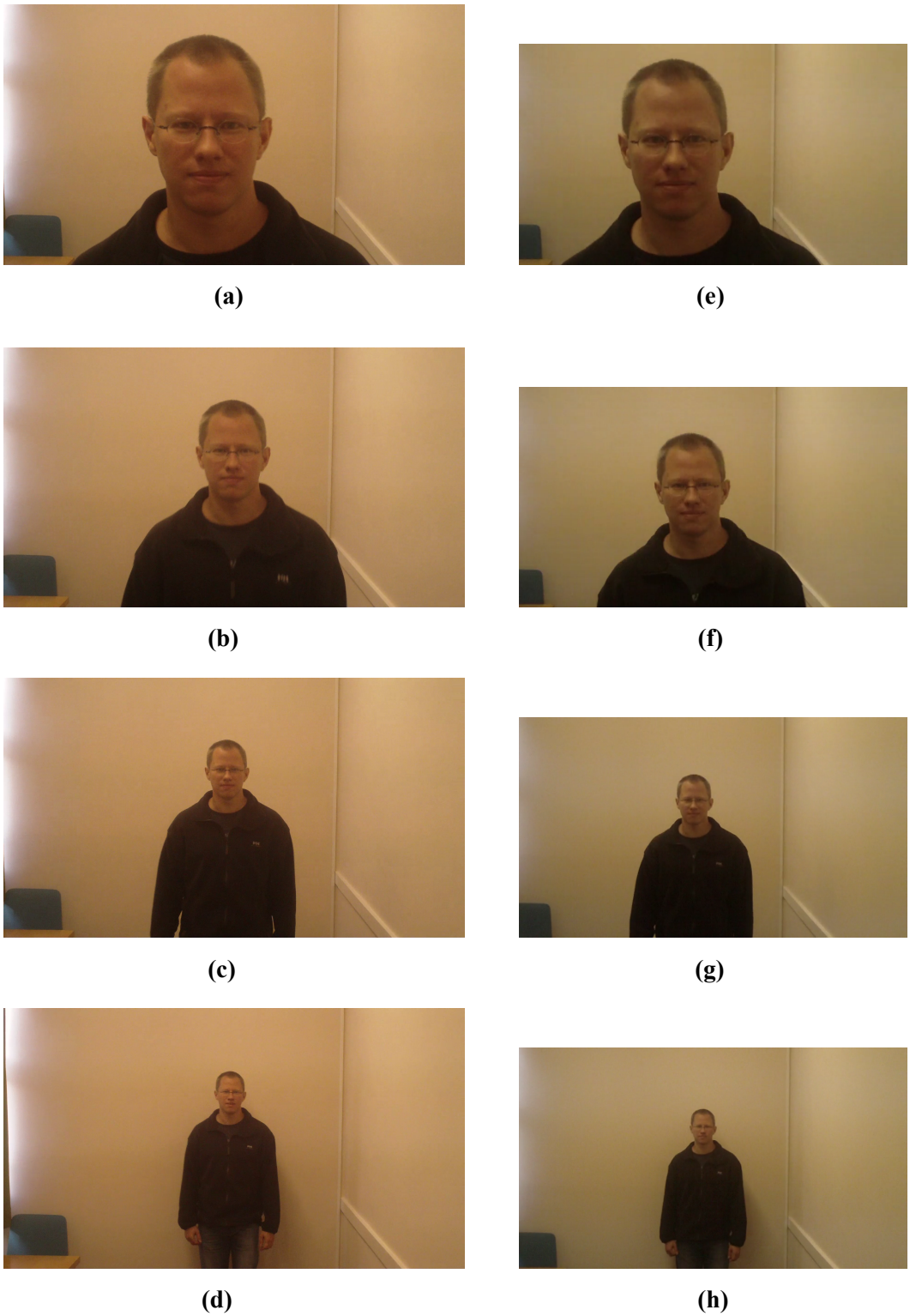
The dataset contains video recordings collected for 20 distinct subjects (17 males and 3 females). The videos of each subject were recorded in two sessions, and each session involved two conditions: indoor and outdoor. The period between the two sessions was at least two days. In each condition, two video recordings (one HD and one SD) of the subject were captured sequentially by the HD camera. Thus, each subject has 8 recordings in total, and the total number of video recordings in the database is 160. The indoor recordings were captured in fixed environmental condition with uniform illumination and background, while the outdoor recordings were acquired in variable lighting condition. These recording conditions represent realistic conditions under which applications of face recognition from a distance can be applied, for example video surveillance and access control applications.

During each recording, the subject was asked to walk a distance of about 4 metres (indoor) and 5 metres (outdoor) in a straight line towards the camera from a start point to a stop point to provide different resolution face data at different distances. During the walk, the subjects are facing the camera, and were free to walk in a natural way which included different head poses. There was an extra metre between the point of camera and the stop point of the subject. Each video recording took between 5 and 10 seconds depending on the speed at which the subject walked.

4.3 Data Preparation

From each video, twelve frames of the subject were extracted. The frames were selected in an automatic way to show the subject at four distance ranges from the camera. Each distance range is represented by 3 frames. The frames in the first range represent the nearest distance from the camera, while the frames in the fourth range represent the farthest distance. Figure 4.1 illustrates the indoor HD and SD video frames representing the four distances, and Figure 4.2 shows the frames captured outdoors. The frames in both figures are rescaled for display purposes.

The division of the total walking distance into 4 ranges is done by dividing the



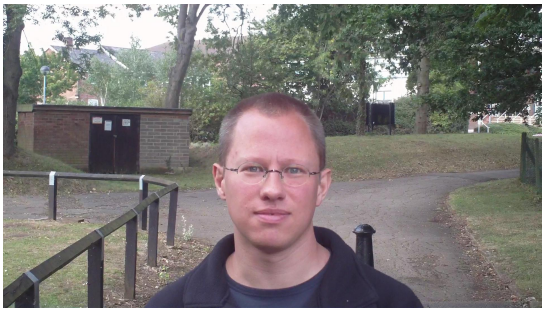
(a) **(e)**

(b) **(f)**

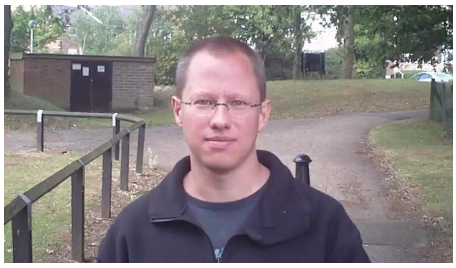
(c) **(g)**

(d) **(h)**

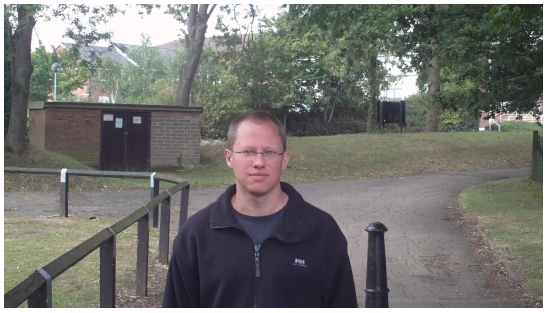
Figure 4.1 Indoor HD and SD video. (a) – (d) HD frames for ranges 1-4 respectively, (e) – (h) corresponding SD frames



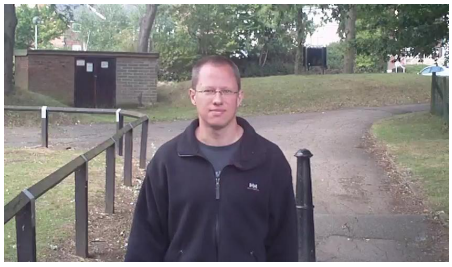
(a)



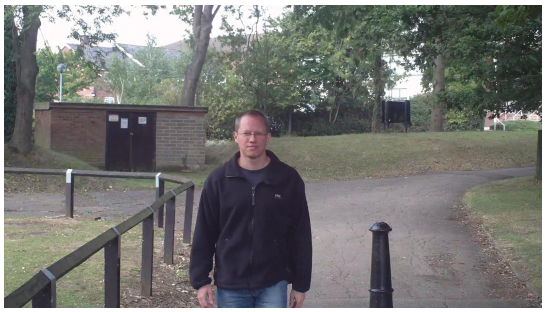
(e)



(b)

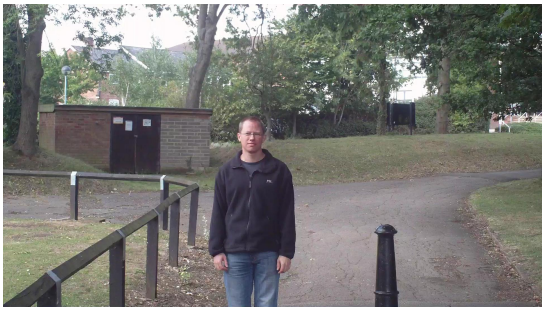


(f)

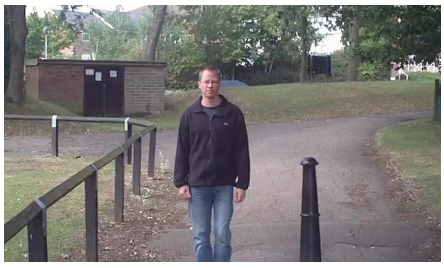


(c)

(g)



(d)



(h)

Figure 4.2 Outdoor HD and SD video. (a) – (d) HD frames for ranges 1-4 respectively, (e) – (h) corresponding SD frames

number of video frames by 4. Then, the mid, mid+5, and mid+10 frames in each frame range were selected and extracted. This process ensures that the subject who appears in HD frames at a certain distance range also appears in their corresponding SD frames at the same distance range from the camera. In some cases, the mid+15 frame was chosen instead of one of the three frames when the latter suffers from motion blur severely. However, the database contains blurred face images with varying poses and facial expressions. The face region in each frame was manually cropped at the top or middle of the forehead, bottom of the chin, and at the base of the ears. Then, all face images were converted to gray scale, rescaled with bilinear interpolation to size of 128 × 128 pixels and saved in "pgm" file format. Figure 4.3 and 4.4 show the cropped and rescaled face images extracted from the respective HD and SD frames in Figure 4.1 and 4.2 respectively.

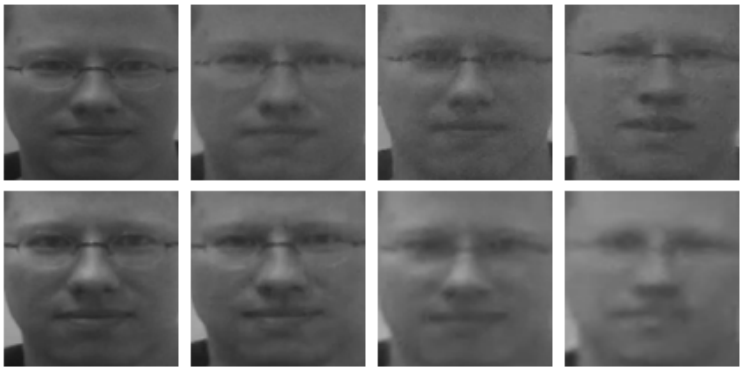


Figure 4.3 Indoor recordings from distance range 1 - 4. Top line HD recordings and bottom line SD recordings

Figure 4.4 Outdoor recordings from distance range 1 - 4. Top line HD recordings and bottom line SD recordings

In total, each subject has 96 face images and the total number of face images in the UBHSD database is 1920. We have not performed any kind of preprocessing (e.g.

illumination normalisation) on the face images in the database. Figure 4.5 and 4.6 show HD and SD face images for another subject in the database. Figure 4.7 shows example outdoor face images for the 20 subjects, where the variations in pose and lighting conditions reflect the real life scenario. Finally, Table 4.1 summarises the main characteristics of the UBHSD database.

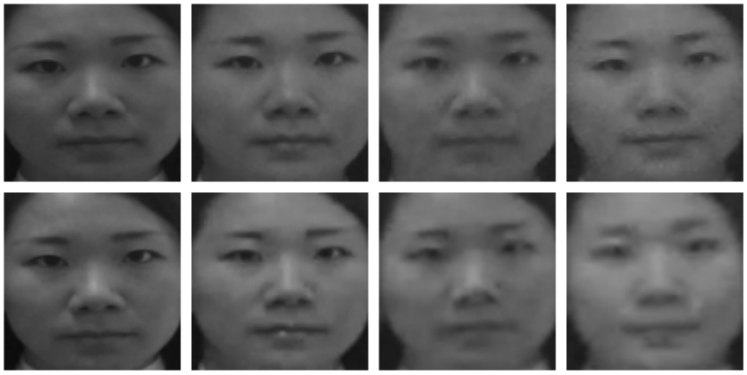


Figure 4.5 Indoor recordings for another subject from distance range 1-4. Top line HD recordings and bottom line SD recordings

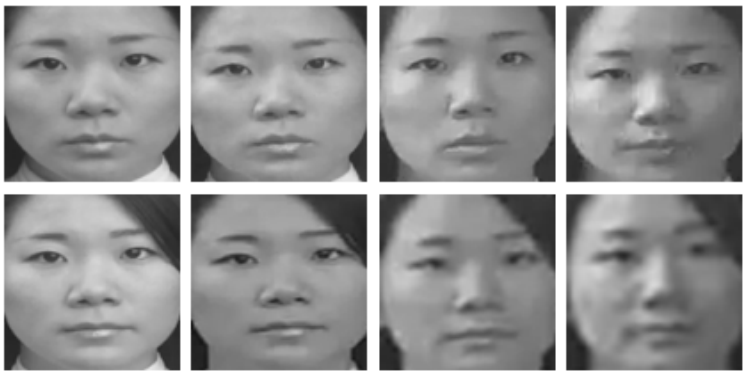


Figure 4.6 Outdoor recordings for another subject from distance range 1-4. Top line HD recordings and bottom line SD recordings

Figure 4.7 Examples of outdoor face images of the 20 subjects in the UBHSD database

Database Property	Value	Database Property		Value	
No. of subjects	20	No. of	males females	17 3	
Recording sessions	2	Video types		HD and SD	
Conditions per session	2	Video resolution		HD SD	1920 × 1080 848 × 480
Videos per subject	8	Frame rate (fps)		HD SD	30 25
Face images per subject	96	Variations in	face size	Yes	
			face pose	Yes	
			facial expressions	Yes	
			eye-glasses	Yes	
			illumination	Yes	
Total no. of videos	160	Video file extension		MOV	
Total no. of face images	1920	Face image file extension		PGM	

Table 4.1 The main features of the UBHSD database

4.3.1 File naming style

We defined a naming style to assign unique names to the video recordings and face images in the UBHSD database. The style, which facilitates the implementation of different evaluation protocols and practical experiments, is defined as follows:

- For video files: (id)_(gender)_(session)_(condition)_(video type).mov
- For face image files: (id)_(gender)_(session)_(condition)_(video type)_ (range number)_(frame number). pgm

where (id) unique identifier of the subject (three digits)
(gender) is either 'm' (male) or 'f' (female)
(session) determines the session 's1' or 's2'
(condition) is either 'i' (indoor) or 'o' (outdoor)
(video type) is either 's1' (standard definition) or 'h1' (high definition)
(range number) is the distance range between the subject and the camera (r1-r4)
(frame number) determines the video frame (4 digits)

4.4 Evaluation Protocol

In order to evaluate the performance of HD and SD video data in face recognition, we defined an experimental protocol to be used to conduct our experiments. This evaluation procedure involves four scenarios P1-P4, where each scenario consists of four tests. Each test was applied twice: one on the HD and one on SD video images by using the three face recognition schemes described in Chapter 3. In each test, recordings from session 1 were used as a training set, while recordings from session 2 were used as a testing set. Thus, no testing data was used as training data.

In test 1 in the first scenario (P1) recordings from session 1, indoor condition, and range 1 were used as a training set, while recordings from session 2, indoor and outdoor conditions, and range 1 were used as a testing set. In the remaining tests in P1, the same configurations of test 1 were used but with different ranges in the testing set, i.e. range 2 in test 2 and so on. Similar configurations of the first scenario were used in the rest scenarios but with different ranges in the training set, i.e. range 2 in P2 and so on. Thus, in each scenario, we evaluate the recognition performance at a certain distance range in the training set with each of the four ranges in the testing set. Table 4.2 shows the configuration of the evaluation protocol.

Scenario	Test	Session 1								Session 2							
		Indoor				*Outdoor*				*Indoor*				*Outdoor*			
		R1	R2	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4
P1	1	Tr								Ts				Ts			
	2	Tr									Ts				Ts		
	3	Tr										Ts				Ts	
	4	Tr											Ts				Ts
P2	1		Tr							Ts				Ts			
	2		Tr								Ts				Ts		
	3		Tr									Ts				Ts	
	4		Tr										Ts				Ts
P3	1			Tr						Ts				Ts			
	2			Tr							Ts				Ts		
	3			Tr								Ts				Ts	
	4			Tr									Ts				Ts
P4	1				Tr					Ts				Ts			
	2				Tr						Ts				Ts		
	3				Tr							Ts				Ts	
	4				Tr								Ts				Ts

Table 4.2 Evaluation protocol. Tr: training set, Ts: testing set

4.5 Software Development

Although in this work we concentrated on providing a comparative study, we also developed a software tool used to conduct the face recognition experiments in a simplified and interactive manner. The remaining sections in this chapter describe the user requirements and the design of the components and algorithm as well as the graphical user interfaces of this software tool.

4.5.1 User Requirements

The user requirements for the software tool used in this project are classified into functional requirements and non-functional requirements. These two types of requirements will be described in detail in the next two sections.

4.5.1.1 Functional Requirements

1- The software tool enables the user to select specific options to determine which session, condition, video type, and distance range are required to prepare the training and testing sets to conduct the experiments.

2- The user shall be able to apply two preprocessing methods: histogram equalisation and z-score normalization for the purpose of normalising the illumination of the training and/or testing face images.
3- The tool gives an option to reload the original sets to the memory without the need to recreate the sets again from the database. This option is useful when the user wants to use the original face images after performing a preprocessing method on the images.
4- The tool provides three face recognition methods: PCA-, LDA- and Wavelet-based face recognition.
5- The tool gives the user a choice to display the probe face image and the recognised face image with a label which indicates that the recognition result is correct or not. In addition, the tool displays a progress bar to show the progress of the processing.
6- The tool provides an option to cancel the test during the run.
7- After conducting each test, the tool automatically saves the experimental results in a tabular form in a text file. In the tests that produce single accuracy, the tool displays the accuracy rate as well as the number of correct recognised faces.

4.5.1.2 Non-functional Requirements

1- A new face video database should be collected and face images should be prepared as described in Section 4.3.
2- The software tool will be implemented using MATLAB language. The reason for using MATLAB is because it provides a rich image processing toolbox.
3- The graphical user interface for the tool will be implemented as a simple window dialogue box.
4- The directory structure of the tool's files should be organised as follows:
    - Directory “Scripts“: contains the MATLAB files of the software tool.
    - Directory “Database”: contains the original face image database, and training and testing sets subdirectories created from the database.

    The “Scripts” and “Database” directories must be put together in the same place in order ensure a proper run of the tool. The training and testing subdirectories can be created automatically by the tool.

4.5.2 System Design

We followed the centralised control style in designing the tool used in this project. The system is composed of a main subsystem called "System_Start" that calls and controls the other three subsystems PCA-, LDA- and Wavelet-based Face Recognition. Figure 4.8 shows an input-process-output model of the system. Each subsystem was designed as a function that contains code organised in subfunctions to run the services, provided by the tool, as well as its graphical user interface. For example, System_Start function contains the following subfunctions: 1) "create" to create the training and testing sets, and 2) "Load_Sets" to load the training and testing sets to the memory.

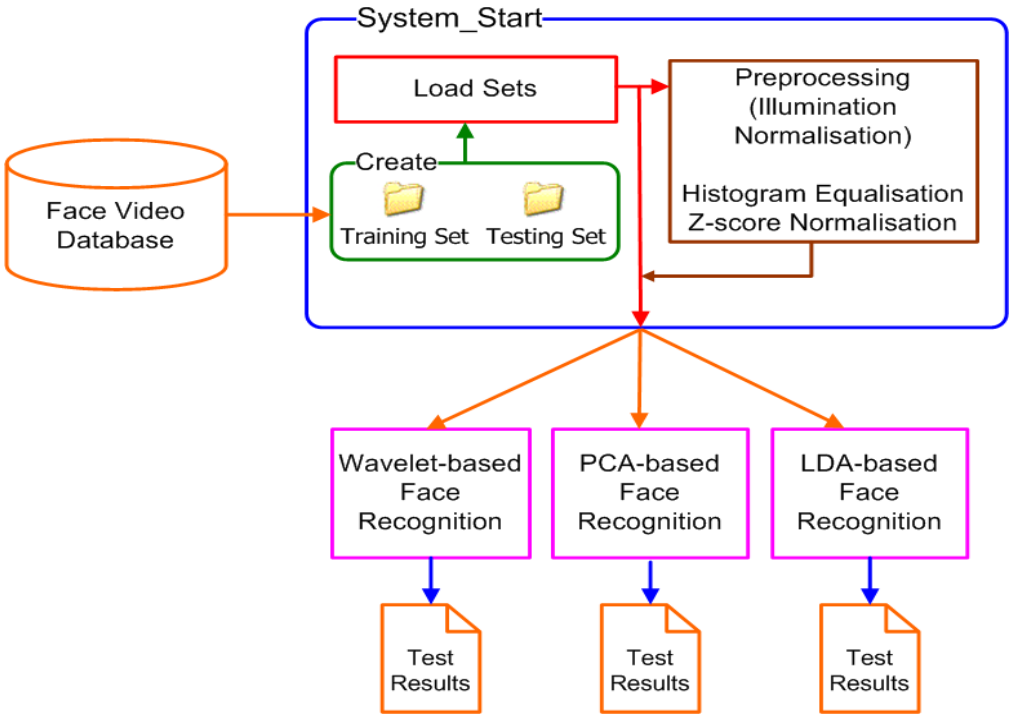


Figure 4.8 An input-process-output model of the software system

As it is clear in the above figure, the System_Start function creates the training and testing sets from the database depending on user selection, and automatically loads these sets into matrices in the memory. Then, it passes these matrices to the other subsystems and subfunctions when they are triggered. For more understanding of the functionality of the software tool, see the system algorithm demonstrated in the next section.

4.5.3 System algorithm design

The flow chart illustrated in Figure 4.9 describes the functionality of the software tool used in this project

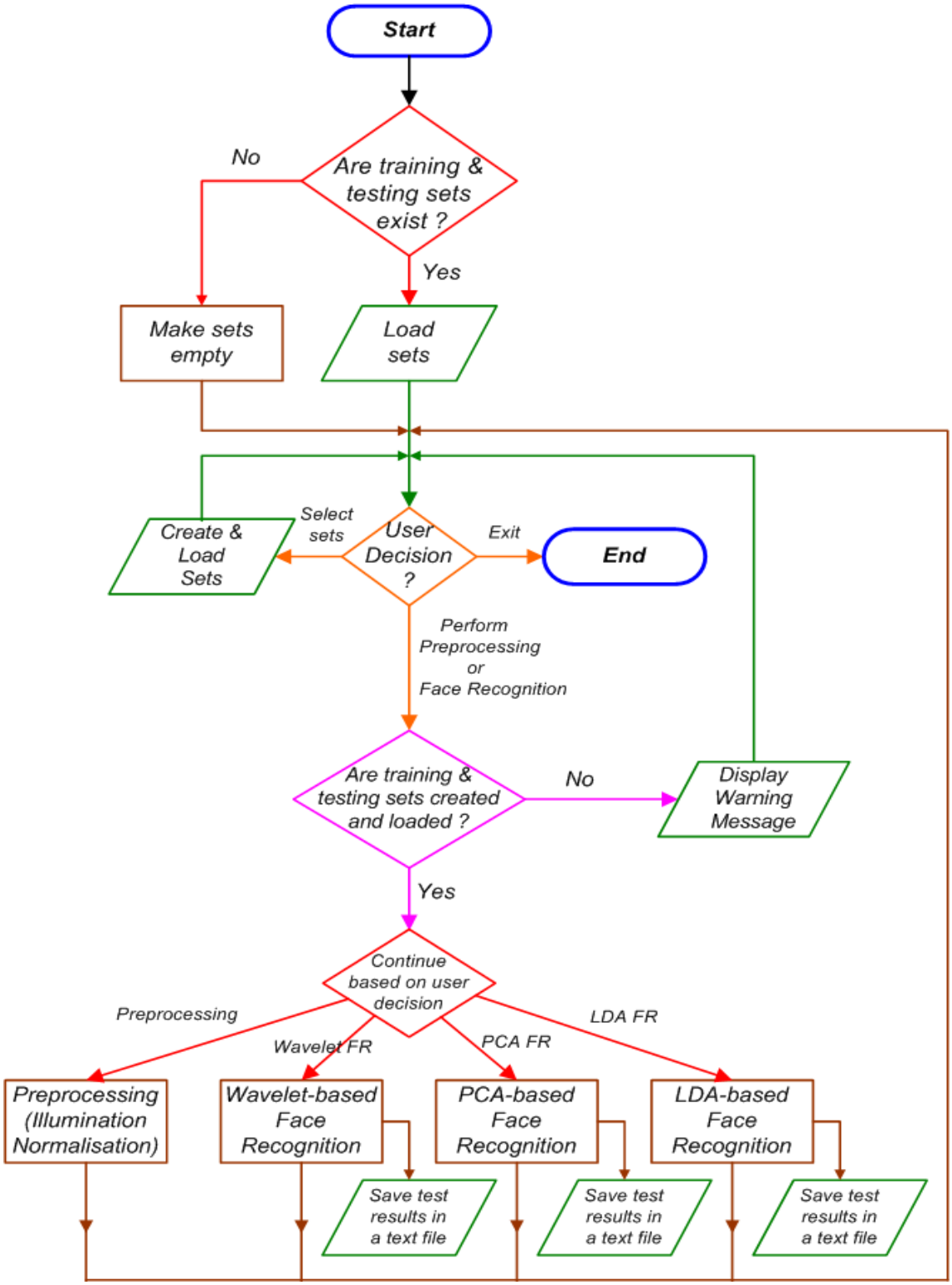


Figure 4.9 Flow chart of the software tool

4.5.4 Graphical User Interfaces

The main graphical user interface (GUI) of the tool is shown in Figure 4.10. The graphical interfaces of the wavelet-based face recognition system are shown in Figures 4.11 and 4.12. Finally, Figures 4.13 and 4.14 illustrate the PCA- and LDA-based face recognition GUIs respectively.

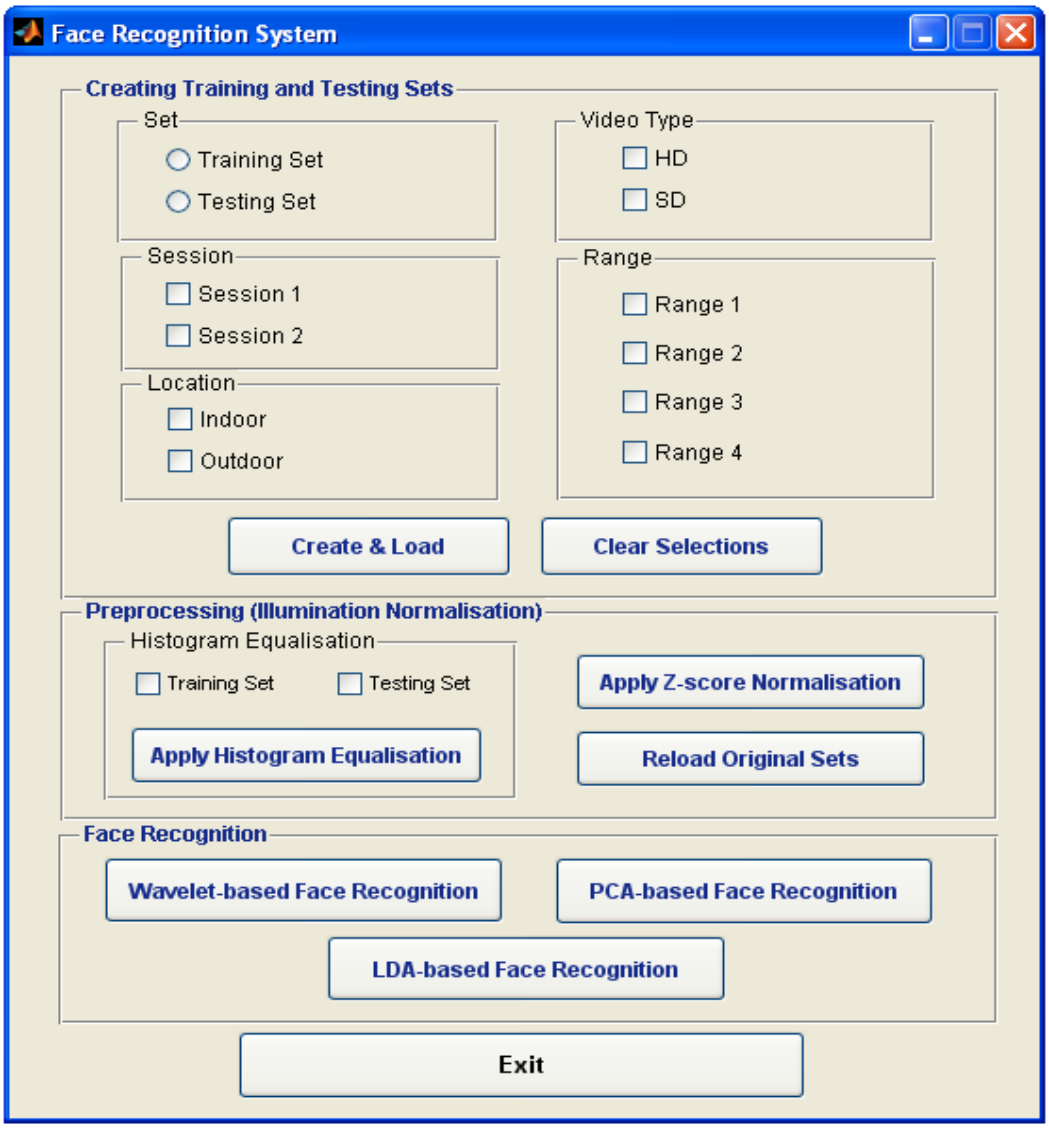


Figure 4.10 Main GUI of the system

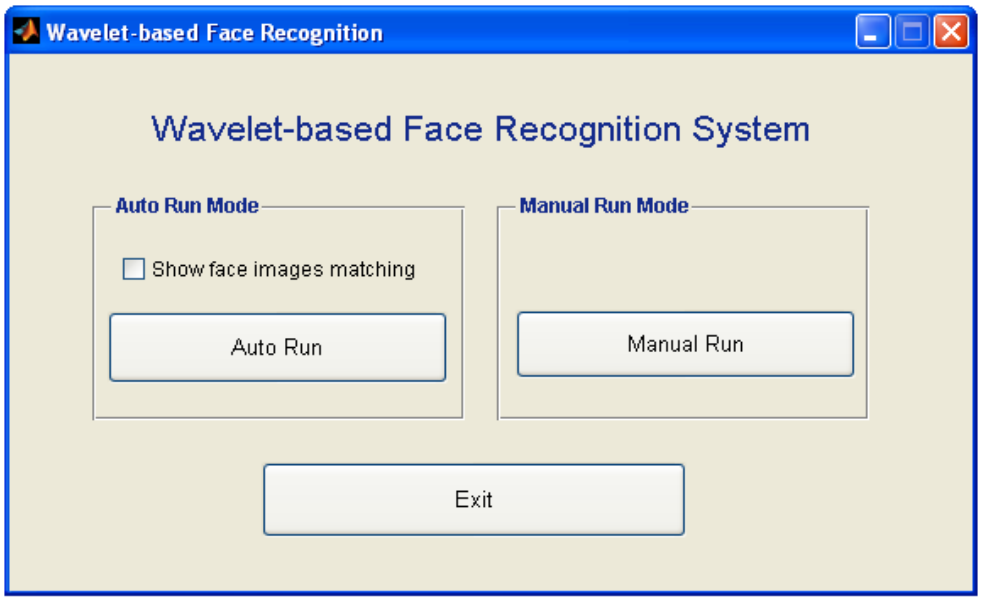


Figure 4.11 Wavelet-based GUI

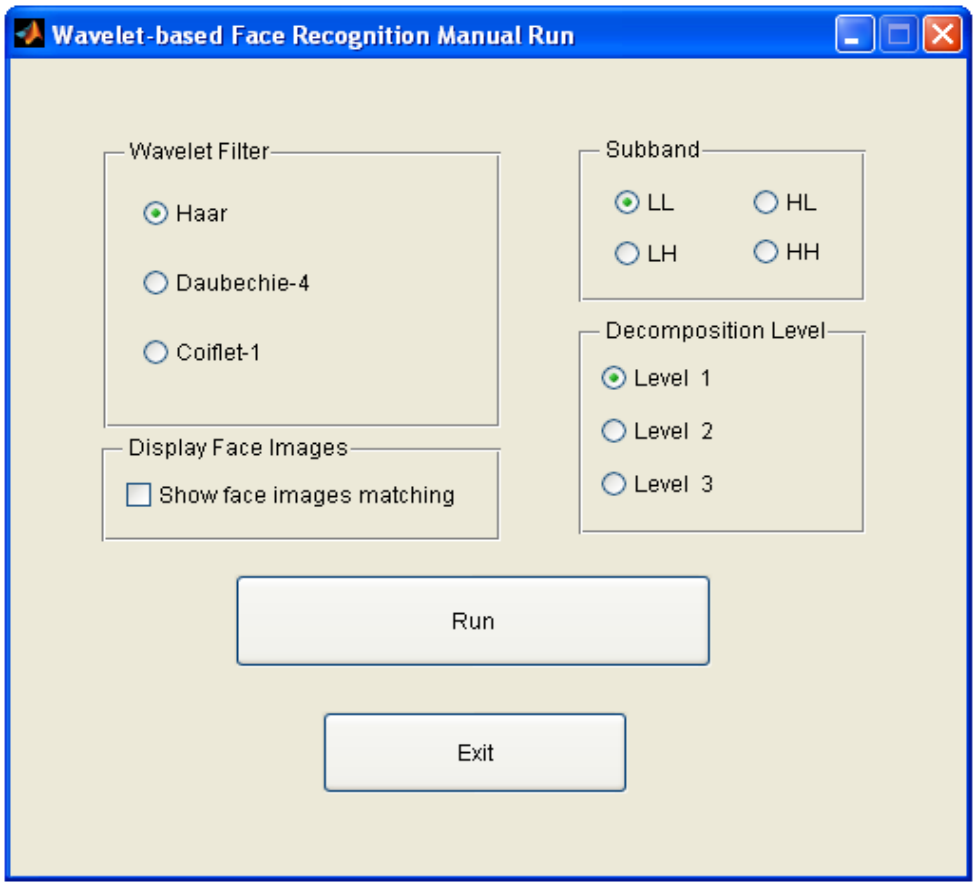


Figure 4.12 Wavelet-based GUI for the manual run mode

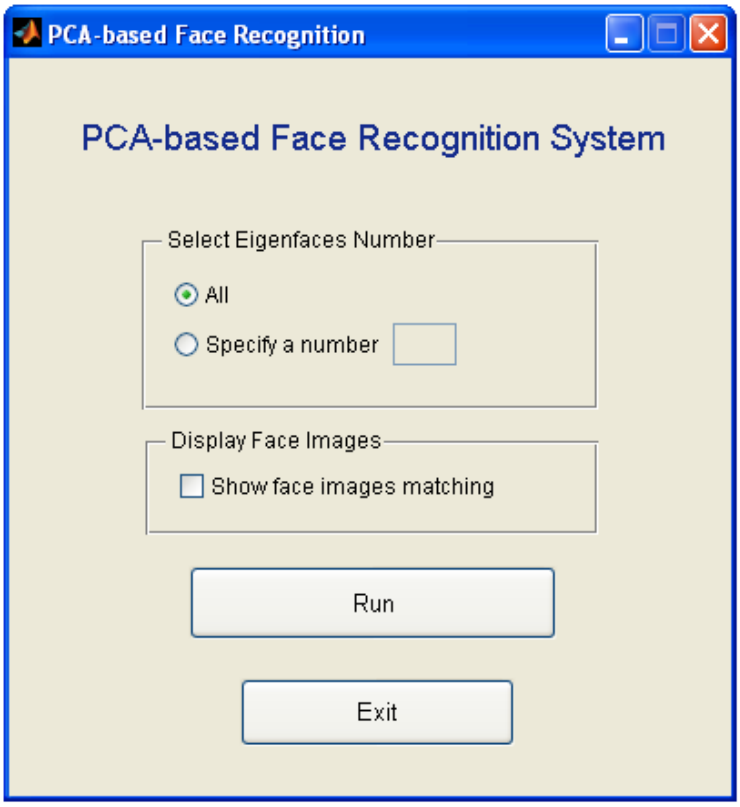


Figure 4.13 PCA-based GUI

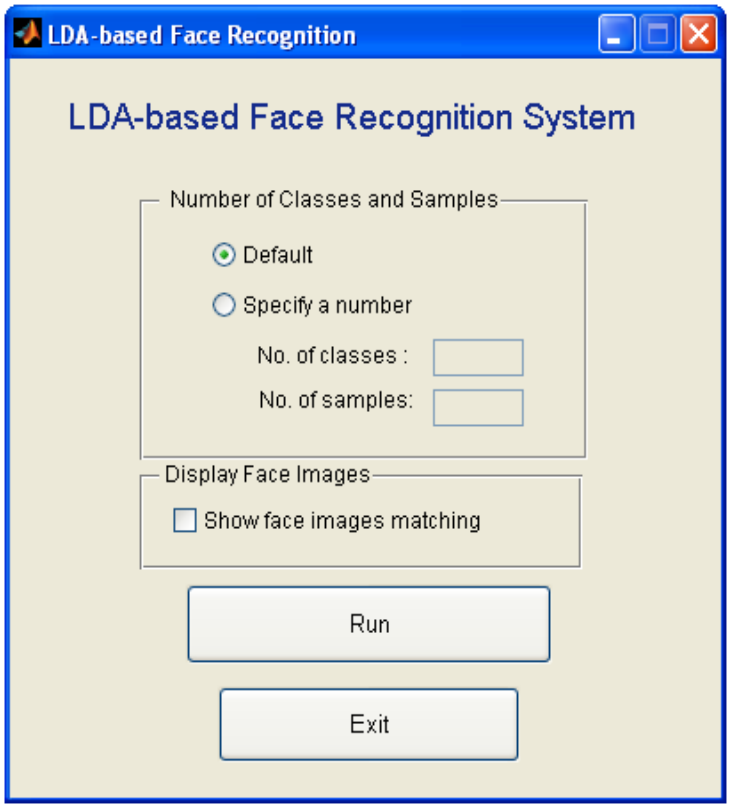


Figure 4.14 LDA-based GUI

4.5.5 Output of Software tool

The visual output of the tool varies depending on user's selection. If the user does not select to display the probe and recognised face images, the tool will display the progress bar window shown in Figure 4.15 during the run of the test. Otherwise it will display the images as shown in Figures 4.16 or 4.17. Figure 4.16 represents an example output of the auto run mode of the wavelet-based face recognition system, where it shows the recognised faces for the four subbands at the same time. The window illustrated in Figure 4.17 is an example of the visual output of PCA- and LDA-based systems as well as the manual run mode of the wavelet-based system. The progress bar window in each figure is to show the processing progress during the run of the test.

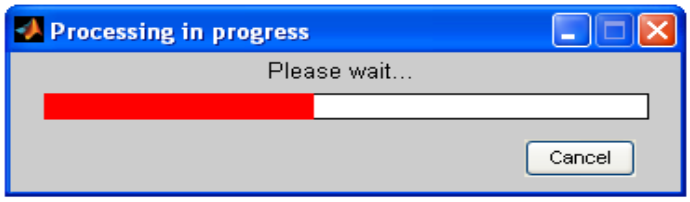


Figure 4.15 Output window without displaying face images

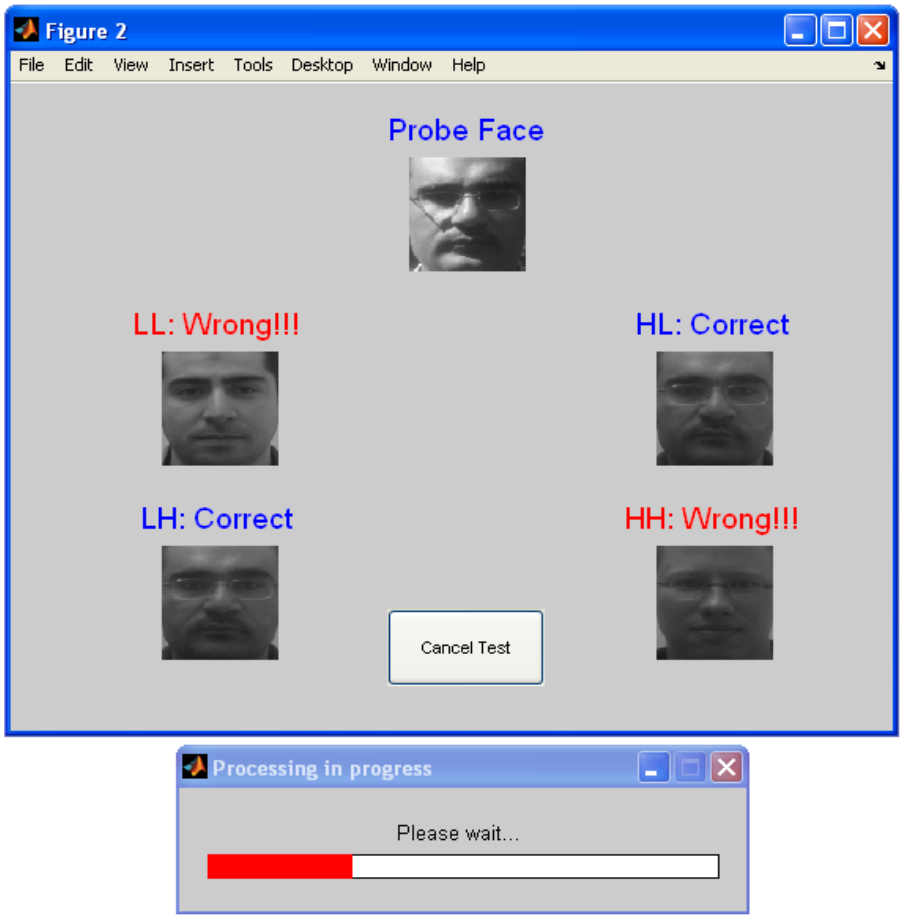


Figure 4.16 Example output of auto run mode of wavelet-based system

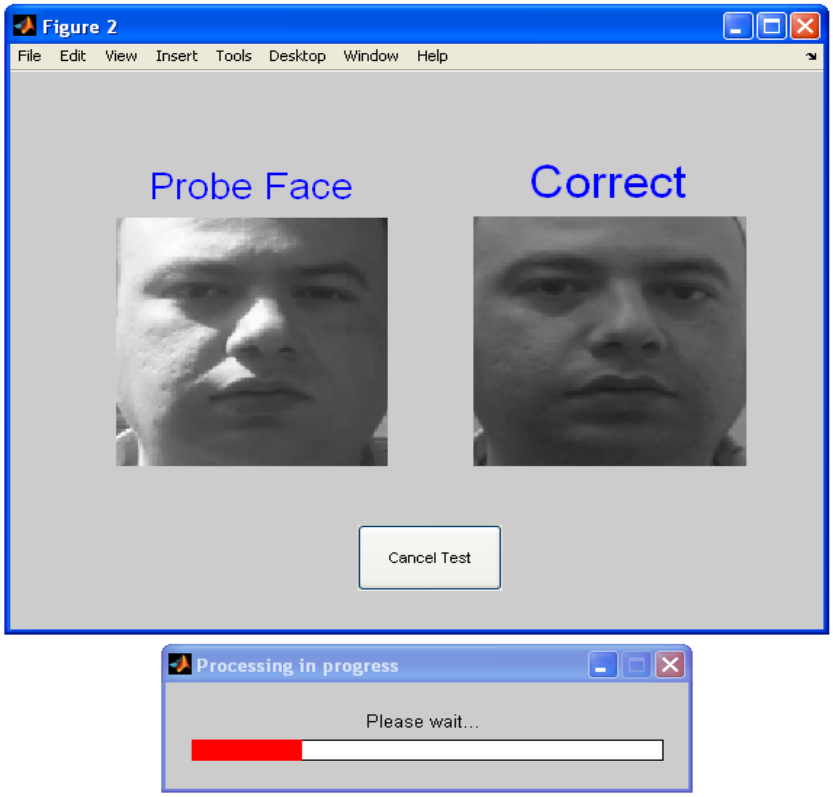


Figure 4.17 Another visual output of software tool

After finishing the test, the tool displays and saves the experimental results in a text file as shown in Figures 4.18 and 4.19 respectively. The window in Figure 4.18 appears in all tests except the auto run mode because this mode produces multiple accuracy rates.

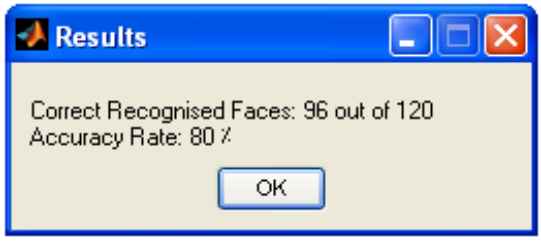


Figure 4.18 Window to display the result

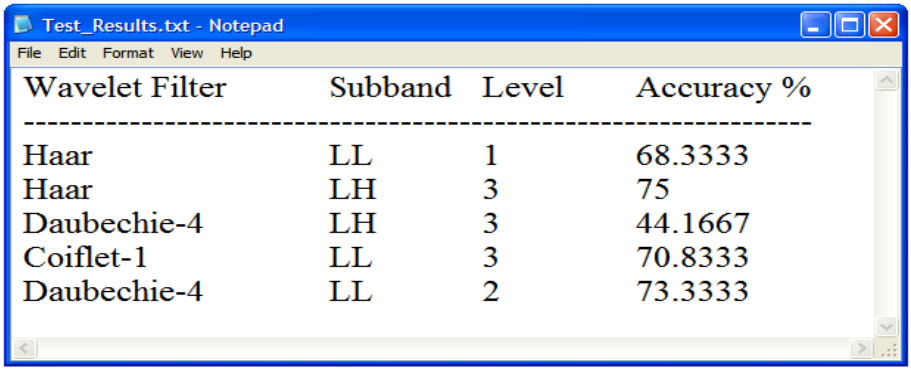


Figure 4.19 An example of text file produced by software tool

Chapter 5

Experiments and Results

In this chapter we will assess the performance of HD and SD video for face recognition at four distance ranges from the camera. We conducted several experiments to test our hypothesis about the impact of video resolution on the identification accuracy. The performance will be examined in two phases: 1) using a single frame to represent each subject in each distance range, and 2) using multiple frames per subject in each range. The purpose of this is to show the benefit of using video rather than still image in face recognition.

Since there are varying lighting conditions between the indoor and outdoor recordings in the UBHSD database, the experiments in each phase were conducted in three stages. These include 1) without illumination normalisation, 2) after normalising by Histogram Equalisation (HE) and 3) after Z-score normalisation (ZN). In each stage, we used two distance measures CityBlock and Euclidean in the three algorithms: Eigenfaces, Fisherfaces, and Wavelets. The Fisherfaces algorithm is not applicable in the first phase because it requires more than one frame per subject. The experimental results using CityBlock measure will be presented and discussed in this chapter and Appendix A, while the results using Euclidean measure are reported in Appendix B.

Throughout this chapter, we will follow a brief notation to simplify the discussion of the experimental results. We will refer to the training set as a gallery set denoted by Tr, and the testing set as a probe set denoted by Ts. The four distance ranges were denoted by R1-R4. We will also refer to the Eigenfaces algorithm by PCA and Fisherfaces by LDA. In the wavelet-based algorithm, the HL and LH subbands at decomposition level 3 (i.e. HL3 and LH3) of the Haar filter were chosen, in addition to PCA and LDA, for the evaluation in this chapter. The reason for selecting these two subbands is because they introduced better recognition rates than the others. The results

of HL3 and LH3 subbands of the Daubechie-4 and Coiflet-1 filters will be presented in Appendix C.

5.1 Performance using single frame per subject in each distance range

In the UBHSD database, since each subject is represented by 3 frames in each distance range, we selected the best frame that contains a frontal or near-frontal pose with minimal motion blur. This shows one of the advantages of using video instead of still image in face recognition (see Section 2.2). The accuracy rates of HD and SD video prior to illumination normalisation will be shown in the next section, while the rates after normalising by HE and ZN will be presented in Appendix A. Besides evaluating the performance of HD and SD video, we will demonstrate in section 5.2 the effect of increasing the number of frames per subject on the recognition accuracy.

5.1.1 Performance prior to illumination normalisation

We followed the evaluation protocol described in Section 4.4 to analyse the recognition performance of HD and SD video. The experimental results of the first scenario are illustrated in Figure 5.1. It can be seen that the performance of SD video is similar, if not better, than that of HD video when the probe set consists of frames in each of the first three ranges. However, when the frames in R4 are used as a probe set, HD video gives better accuracy rates than SD video.

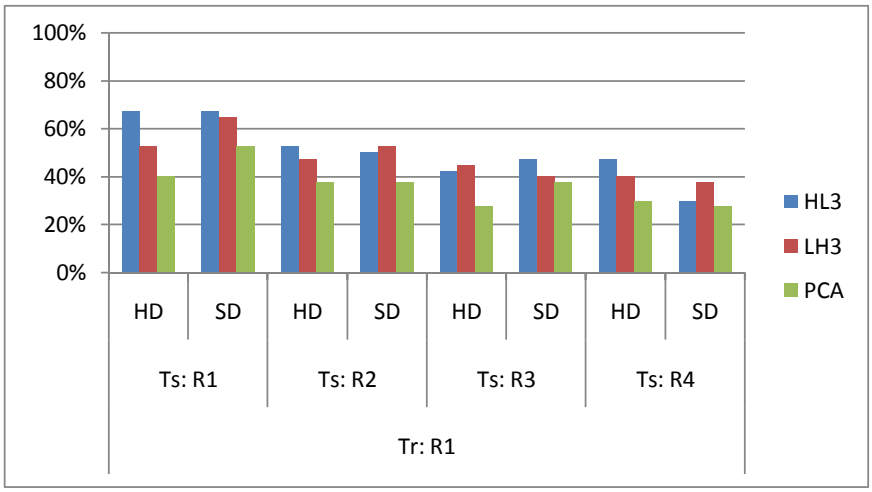


Figure 5.1 Accuracy rates using single frames in range R1 as a gallery set

The superiority of SD video over HD video in R1 could be attributed to the down sampling in the rescaling process. Since the cropped HD face images have spatial

resolution more than double that of cropped SD images, they lose more pixels than SD images when they are rescaled to a size of 128×128. Losing more pixels (i.e. details) from face images can negatively affect the recognition performance.

As is apparent in Figure 5.2, when the frames in R2 are used as a gallery set the HL3 and PCA show that HD video significantly outperforms SD video in each of the four ranges in the probe set. While the LH3 indicates that the SD video performs better than HD video in R1 and R2, and the reverse in R3 and R4.

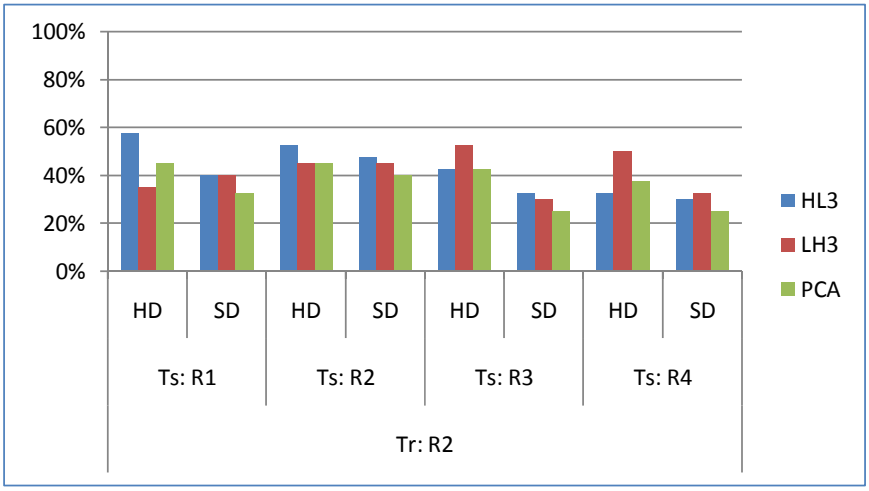


Figure 5.2 Accuracy rates using single frames in range R2 as a gallery set

The results obtained from using frames in R3 as a gallery set are shown in Figure 5.3. The HL3 and LH3 subbands show that HD video gives higher recognition results than SD video when the frames in each of all ranges except R3 are used as a probe set. PCA shows the same scenario but for ranges R2 to R4.

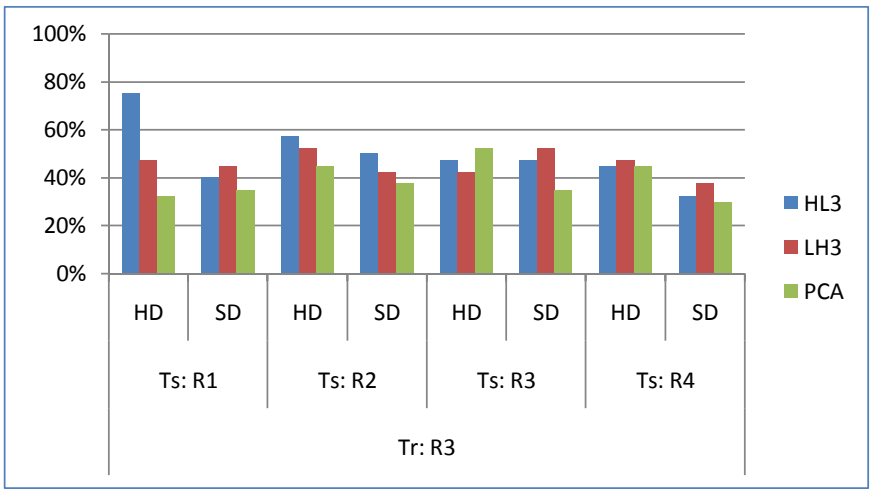


Figure 5.3 Accuracy rates using single frames in range R3 as a gallery set

Finally, the chart in Figure 5.4 demonstrates the fourth experiment in which the gallery set comprises the frames in R4. In this experiment, the performance of HD video approximates to that in the second experiment illustrated in Figure 5.2.

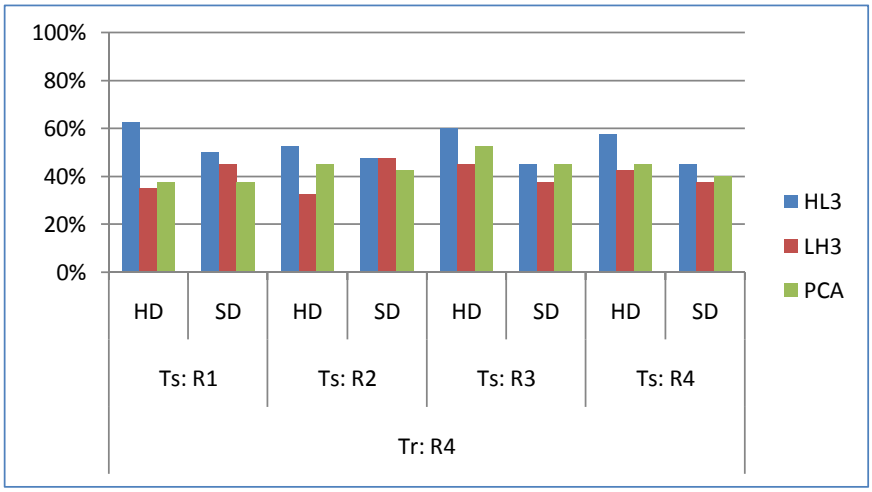


Figure 5.4 Accuracy rates using single frames in range R4 as a gallery set

From the above experiments, it can be concluded that HD video has a superior performance over SD video when frames in each of the ranges R2-R4 are used as a gallery set. Another observation is that HD video perfectly outperforms the SD video when the frames in the farthest range R4 are used as a probe set irrespective of the distance range used in the gallery set. However, the performance may change after applying HE or ZN on the probe and gallery sets (see Appendix A). This will also be illustrated when each subject is represented by 3 frames in the next section.

In general, in this phase, the accuracy rates are low because only one frame is used to represent each subject in each distance range. In the next section we will see the improvements in the recognition accuracy when multiple frames are used per subject.

5.2 Performance using multiple frames per subject in each distance range

We represented each subject by the three frames that we extracted in each distance range. In addition to Eigenfaces and wavelet approaches, we used Fisherfaces algorithm to provide a more reliable evaluation of the recognition performance of HD and SD video. The experimental results, described in the next sections, show a substantial increment in the accuracy rates compared to the rates that resulted from using single frame. This also illustrates another advantage of using video rather than still image in

automated face recognition. The performance of HD and SD video before normalising illumination and after normalising by HE and ZN will be described in the next sections.

5.2.1 Performance prior to illumination normalisation

Once again, we analysed the performance of HD and SD video according to the experimental protocol described in Section 4.4. The recognition rates when the gallery set involves frames in R1 are presented in Figure 5.5. The HL3 subband shows that HD video gives greater rates than SD video when frames in each range are used as a probe set. By contrast, LH3, PCA and LDA show that SD video performs better in R2 and R4.

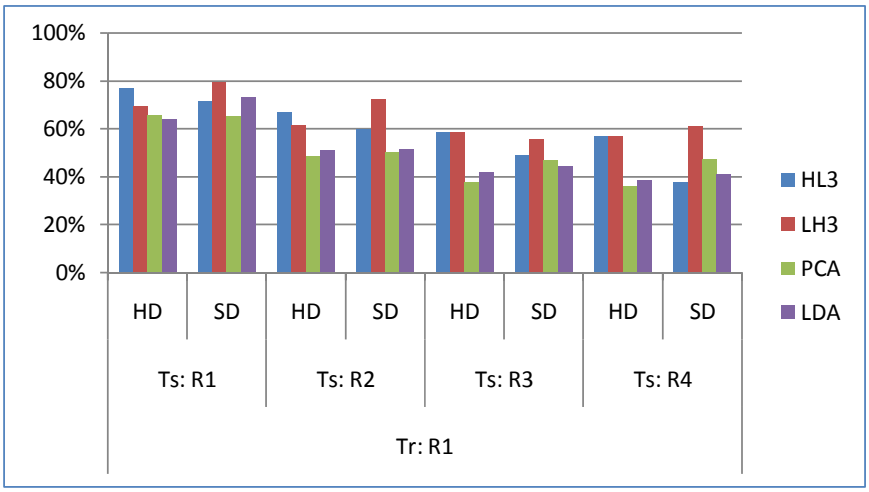


Figure 5.5 Accuracy rates using frames in range R1 as a gallery set

The chart in Figure 5.6 shows the results of the second experiment in which the frames in R2 are used as a gallery set. Only LH3 indicates that SD video gives higher recognition rates than HD video when frames in R1 are used as a probe set. Conversely, HD video perfectly outperforms the SD video in the ranges R2, R3 and R4.

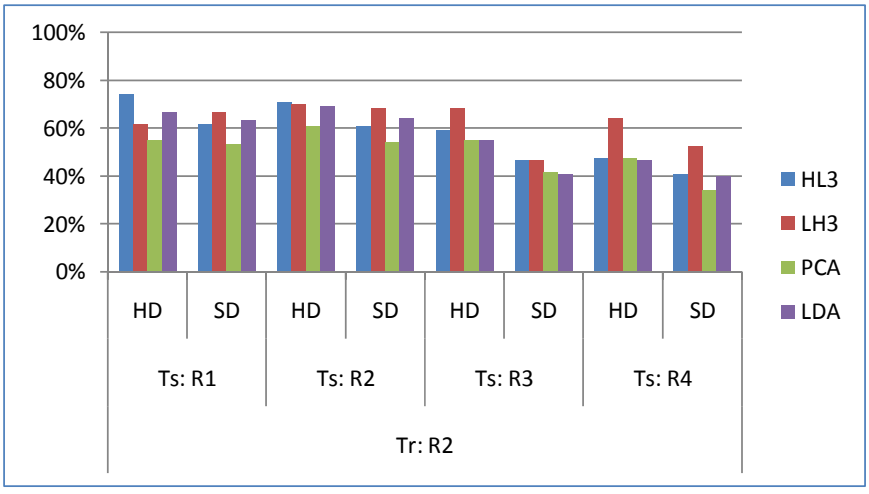


Figure 5.6 Accuracy rates using frames in range R2 as a gallery set

Similarly, when frames in R3 are used as a gallery set, HD video performs better than SD video in the ranges R2, R3, and R4 in the probe set as illustrated in Figure 5.7. The same performance is shown by HL3 when the probe set includes frames in R1.

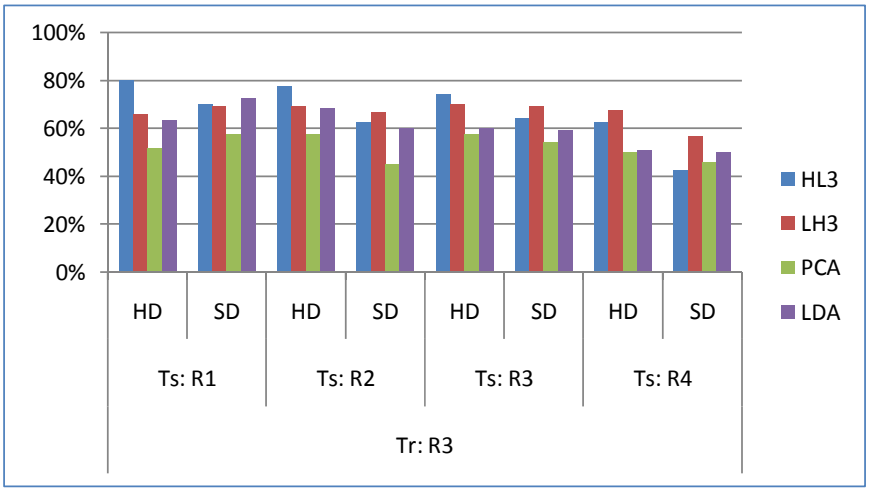


Figure 5.7 Accuracy rates using frames in range R3 as a gallery set

Finally, the bar chart in Figure 5.8 illustrates the fourth experiment in which the frames in R4 are used as gallery set. As can be seen in this figure, the performance of HD video is similar to that in the previous two experiments in each of the ranges R2-R4. The same performance is shown by HL3 and LDA when the probe set consists of frames in R1.

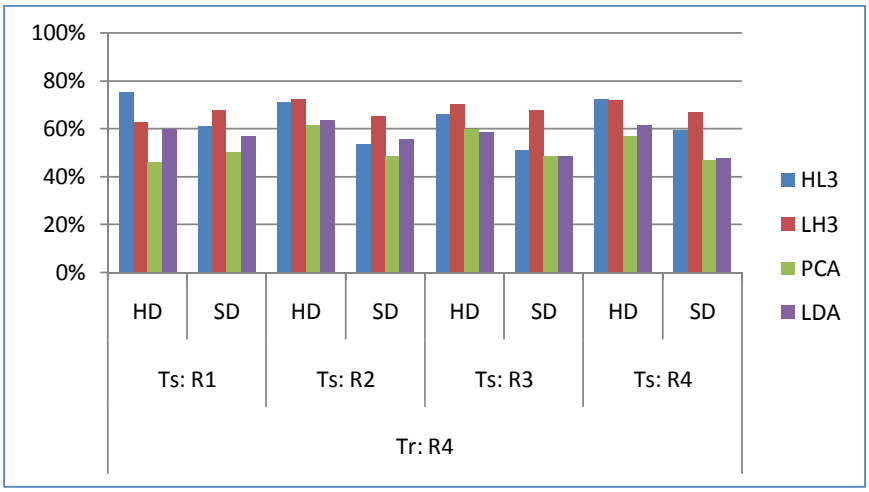


Figure 5.8 Accuracy rates using frames in range R4 as a gallery set

It can be inferred from the previous four experiments that HD video perfectly outperforms SD video when frames in the ranges R2, R3, and R4 are used in both gallery and probe sets. This performance is similar to that when we used a single frame per subject (see Section 5.1.1).

5.2.2 Performance after normalising by HE

We also tested the recognition performance of HD and SD video after applying histogram equalisation on the probe and gallery sets. In general, some of the accuracy rates were increased while others were declined. This irregular change in the rates could be a result of the noise added by HE to the frames. Thus, the performance of HD and SD video are considerably affected. Figure 5.9 demonstrates the recognition rates when the frames in R1 are used in the gallery set. As illustrated in this figure, SD video performs better than HD video in the first two ranges in the probe set. In R3 and R4, PCA shows the same scenario whereas LDA shows the opposite.

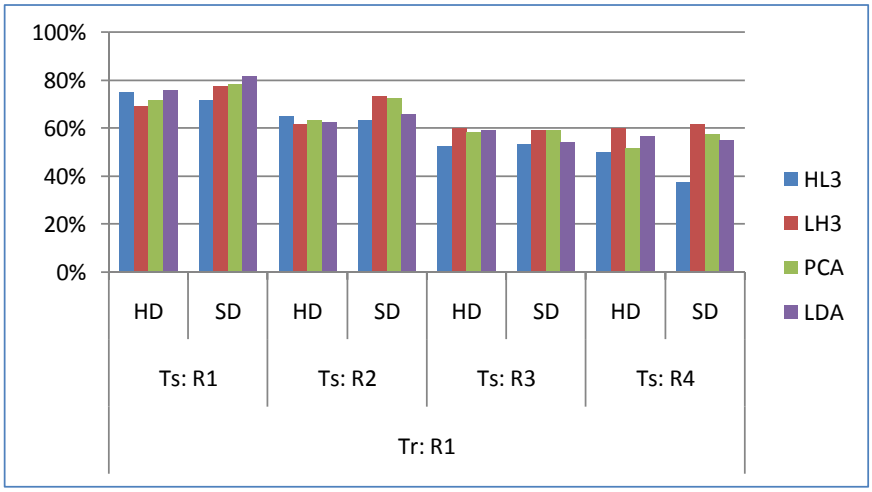


Figure 5.9 Accuracy rates after HE using frames in range R1 as a gallery set

When the frames in R2 are used as a gallery set, the HL3 subband shows the superiority of HD video over SD video in each of the four ranges in the probe set as shown in Figure 5.10. However, PCA and LDA show the reversed case.

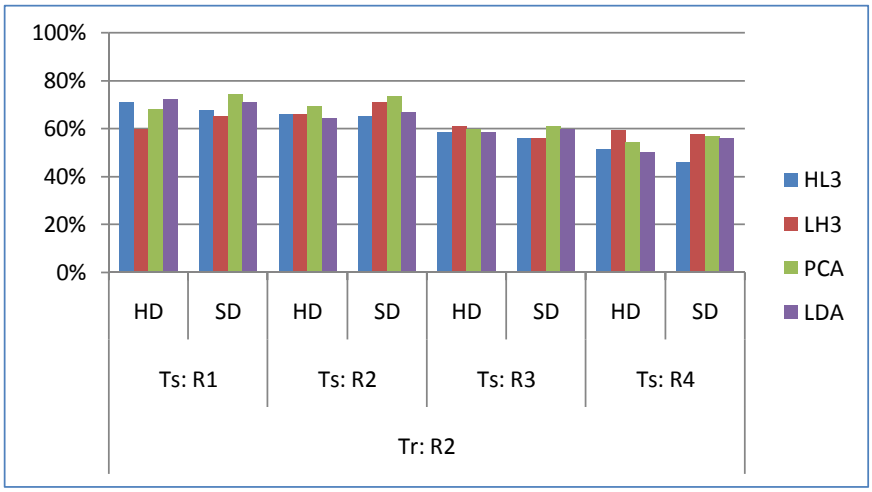


Figure 5.10 Accuracy rates after HE using frames in range R2 as a gallery set

Similarly, as is evident in Figure 5.11 the performance shown by HL3 and PCA, when frames in R3 are used as a gallery set, is close to that in the previous experiment.

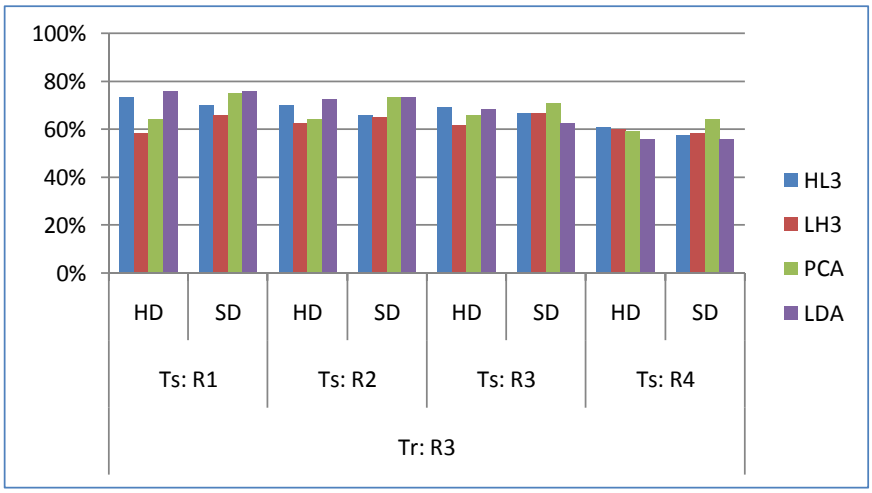


Figure 5.11 Accuracy rates after HE using frames in range R3 as a gallery set

The results of the fourth experiment in which the gallery set is represented by frames in the last range are demonstrated in Figure 5.12. In general, HD video has a superior performance than SD video when the probe set includes frames in each of the ranges R2, R3 and R4.

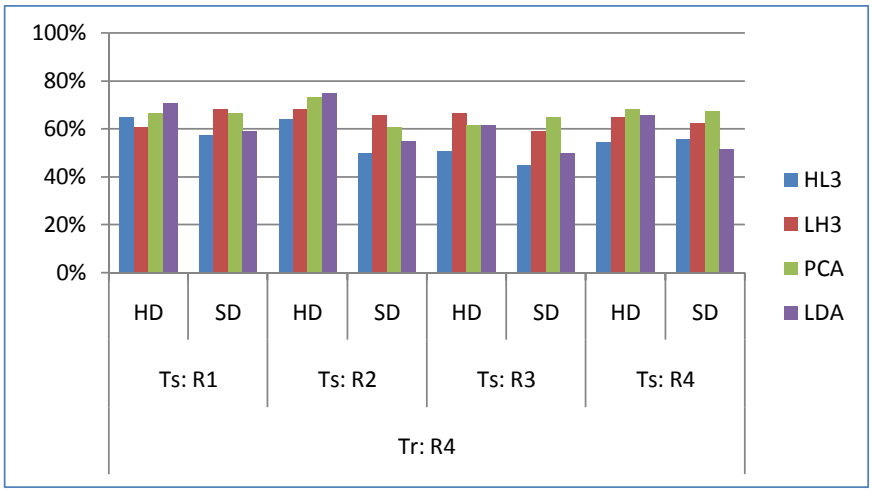


Figure 5.12 Accuracy rates after HE using frames in range R4 as a gallery set

In conclusion, it can be clearly seen that the performance of HD video in the far distances R2-R4 has been considerably affected after applying histogram equalisation. As mentioned earlier, this could be attributed to the noise added by HE to the probe and gallery images.

5.2.3 Performance after normalising by ZN

We also compared the recognition performance of HD and SD video after applying z-score normalisation on the probe and gallery sets. Overall, the accuracy rates were improved compared to the rates obtained before normalisation and after HE. The experimental results when the gallery set consists of frames in R1 are illustrated in Figure 5.13. As shown in this figure, the rates given by LH3, PCA and LDA show that SD video perfectly prevails over HD video when frames in each of the four ranges are used as a probe set. However, HL3 shows the opposite scenario.

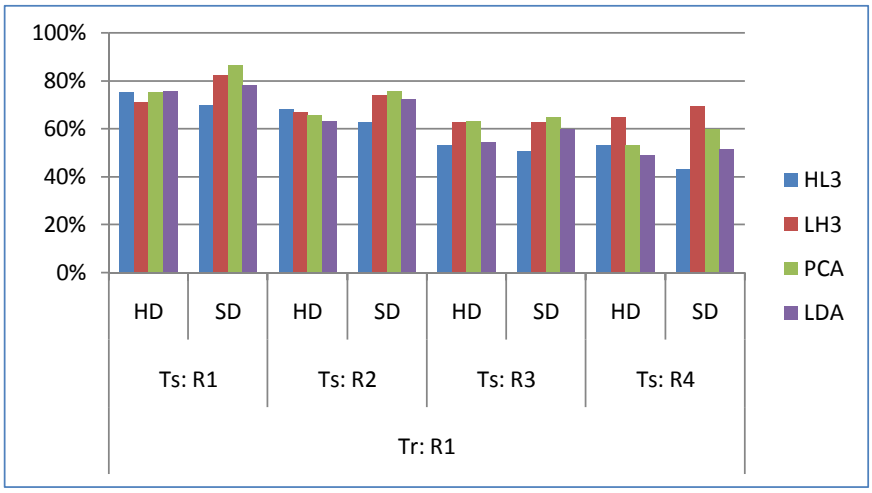


Figure 5.13 Accuracy rates after ZN using frames in range R1 as a gallery set

The chart in Figure 5.14 illustrates the results when the frames in R2 are used in the gallery set. It is clear that HD video overcomes SD video when the frames in each of the ranges R3 and R4 are used in the probe set. The same performance is presented by HL3 and LDA in the first two ranges.

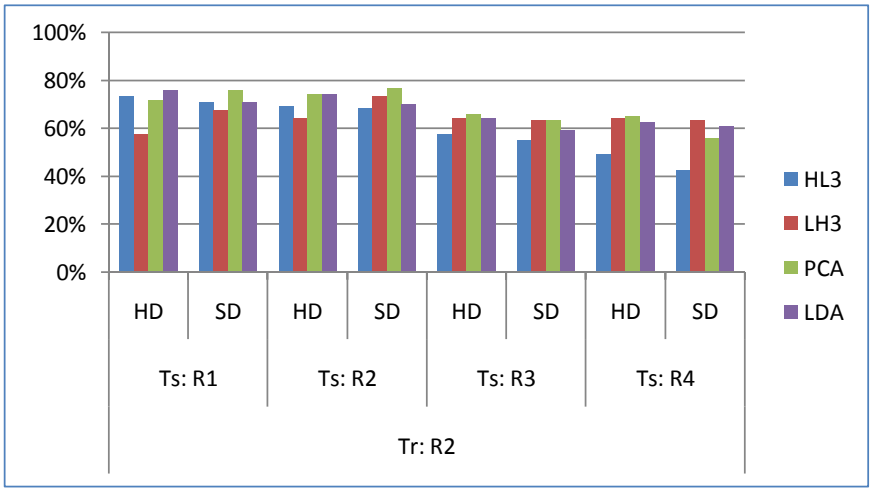


Figure 5.14 Accuracy rates after ZN using frames in range R2 as a gallery set

The results of the third experiment in which the gallery set comprises frames in R3 are shown in Figure 5.15. We can see that HD video gives better recognition rates than SD video when the probe set includes frames in each of the first two ranges. The same performance is given by HL3 and LDA in the ranges R3 and R4.

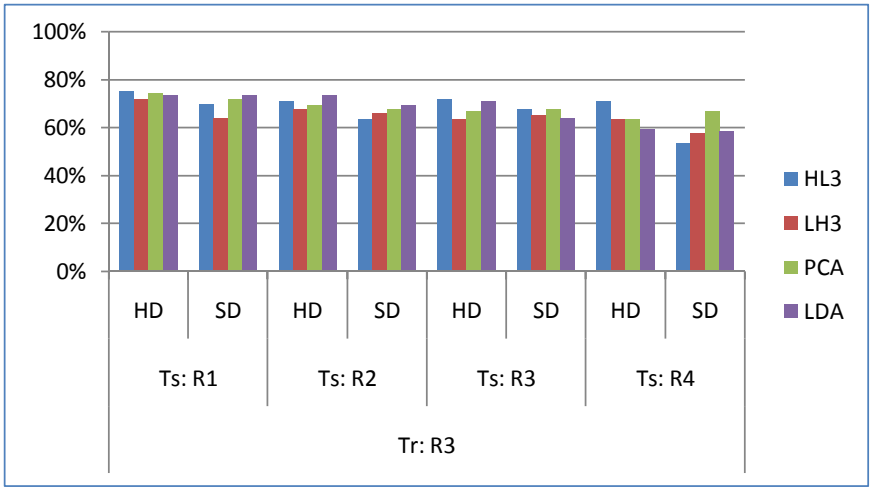


Figure 5.15 Accuracy rates after ZN using frames in range R3 as a gallery set

Finally, the recognition rates when the frames in R4 are used in the gallery set are shown in Figure 5.16. In this figure, the accuracy rates given by HL3, PCA and LDA show that HD video performs better than SD video in the first two ranges. The same performance is shown by HL3 and LDA in R3, and PCA and LDA in R4.

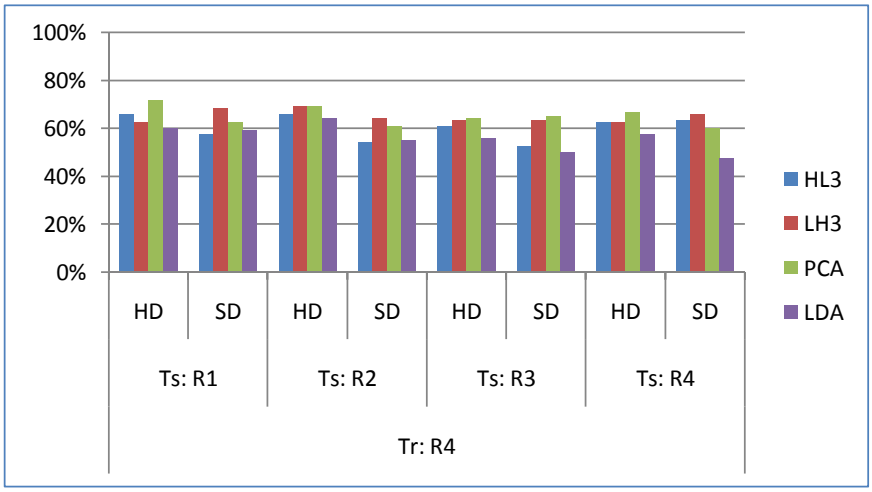


Figure 5.16 Accuracy rates after ZN using frames in range R4 as a gallery set

On this basis, it can be deduced that after applying ZN on the probe and gallery sets, the performance of SD and HD video, compared to that prior to normalisation, has changed when frames in each of the ranges R2-R4 are used in the gallery set. However, HD video still has a superior performance over SD video in these ranges.

Chapter 6

Conclusions and Future Work

This study has presented an investigation of the use of HD and SD video in face recognition at different distances from the camera. We created a new face video database comprising HD and SD video data collected for 20 distinct subjects using a low-cost HD body worn video camera. Three face recognition algorithms, namely Eigenfaces, Fisherfaces and wavelet-based approach were used to evaluate the recognition performance of HD and SD video.

The experiments and results in this study suggest a number of conclusions. First, we have shown that the recognition accuracy has been improved when multiple instances are used to represent each subject in each distance range. Second, what was surprising is that SD video has better recognition performance than HD video at a close range from the camera. This unexpected finding indicates that using very high resolution images at a close range could lead to poor recognition accuracy. Consequently, it can be suggested that SD video is sufficient for applications in which the individuals are closely recognised, for example access to computer/network and facilities (e.g. cash machines). In contrast, HD video substantially outperforms SD video when working with larger ranges. This is useful when the recognition is performed from a distance for example monitoring individuals in video surveillance systems and controlling the access to the buildings. This behaviour of HD video corroborates our hypothesis posed at the beginning of this study.

From these two different behaviours of HD and SD video, we can ask the following question; should we use HD video or SD video for face recognition? According to the results of the experiments, the selection of an appropriate video varies depending on the quality of the gallery set and the probe image. The quality of images is related to the distance from the camera. Therefore, a face recognition system in an

uncontrolled environment (e.g. CCTV with automatic face recognition) should have the ability to automatically select the appropriate resolution that is suited to the distance when identifying a person. Finally, in the case of illumination normalisation, the performance may change due to the effects caused by the normalisation methods on the probe and gallery images. For example, histogram equalisation produces noise to the images, causing, in some cases, a decrease in the recognition rates.

In future investigations it might be possible to use additional face recognition algorithms, such as Independent Discriminant Analysis (ICA) as well as the state-of-art approaches. In addition, we can evaluate the performance of HD and SD video with different image rescaling methods. Another possible area of future work is to compare the recognition performance of HD video and the high resolution video generated by resolution enhancement techniques such as super resolution. In the UBHSD database, we can increase the number of subjects as well as increasing the number of frames per subject. Last but not least, the software tool can be extended to provide additional services, for example summarising and visualising the experimental results, and providing additional illumination normalisation methods as well as face recognition algorithms.

References

Amira, A., and P. Farrell. "An Automatic Face Recognition System Based on Wavelet Transforms." *IEEE International Symposium on Circuits and Systems* , vol. 6 (2005): 6252 - 6255 .

Ao, Meng, Dong Yi, Zhen Lei, and Stan Z. Li. "Face Recognition at a Distance: System Issues." In *Handbook of Remote Biometrics*, by Massimo Tistarelli, Stan Z. Li and Rama Chellappa, 155-167. Springer, 2009.

Apple. "Help Library." 2010. http://documentation.apple.com (accessed July 4, 2010).

Arachchige, Somi R. B. "Face Recognition in Low Resolution Video Sequences using Super Resolution." MSc Thesis, Rochester Institute of Technology, Rochester, NY, 2008.

Belhumeur, Peter N., Joao P. Hespanha, and David J. Kriegman. "Eigenfaces vs. Fisherfaces Recognition using Class Specific Linear Projection." *IEEE Transactions on Pattern Analysis and Machine Intelligence* , vol. 19, no. 7 (1997): 711 - 720.

Bertino, Elisa, and Elena Ferrari. "Temporal Synchronization Models for Multimedia Data." *IEEE Transactions on Knowledge and Data Engineering* , vol. 10, no. 4 (July/August 1998): 612 - 631.

Browne, Steven E. *High Definition Postproduction: Editing and Delivering HD Video.* Oxford: Focal Press, 2006.

Cambridge, AT&T Laboratories. *The Database of Faces.* 2002. http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html (accessed October 10, 2010).

Chapman, Nigel, and Jenny Chapman. *Digital Multimedia.* 2nd edn. West Sussex: Wiley & Sons Ltd, 2004.

Choi, Jae Young, Yong Man Ro, and Konstantinos N. Plataniotis. "Feature Subspace Determination in Video-based Mismatched Face Recognition." *Proceedings of the 8th IEEE International Conference on Automatic Face and Gesture Recognition*, 2008: 1-6.

Cucchiara, Rita. "Multimedia Surveillance Systems." *International Multimedia Conference.* Association for Computing Machinery, 2005. 3-10.

DeMenthon, Daniel, and David Doermann. "Video Retrieval using Spatio-Temporal Descriptors." *ACM Multimedia*, 2003: 508-517.

Dick, David. *PC Multimedia & Web technology and techniques.* 2nd edn. Dumbreck Publishing, 2002.

Georganas, Nicolas D. "Multimedia Applications Development Experiences." *Journal of Multimedia Applications Development Experiences* (Springer) , vol. 4, no. 3 (May 1997): 313–332.

Golshani, Forouzan. "Digital Biometrics." *Encyclopedia of Multimedia*, 2008: 160-165.

Heath, Steve. *Multimedia and Communications Technology.* 2nd edn. Oxford: Reed Educational and Professional Publishing, 1999.

Hennings-Yeomans, Pablo H., Simon Baker, and B.V.K. Vijaya Kumar. "Recognition of Low-Resolution Faces Using Multiple Still Images and Multiple Cameras." *2nd IEEE International Conference on Biometrics: Theory, Applications and Systems*, 2008: 1-6.

Hofstetter, Fred T. *Multimedia Literacy.* NewYork: McGrawHill, 1995.

Huang, Thomas, Ziyou Xiong, and Zhenqiu Zhang. "Face Recognition Applications." In *Handbook of Face Recognition*, by Stan Z. Li and Anil K. Jain, 371-390. Springer, 2005.

International Society for Technology in Education. 2010. http://www.iste.org/content/navigationmenu/research/reports/research_on_technology_in_education_2000_/multimedia/research_on_multimedia_in_education.htm (accessed June 24, 2010).

Jain, Anil, Rund Bolle, and Sharath Pankanti. *Biometrics Personal Identification in Networked Society.* Norwell: Kluwer Academic Publishers, 1999.

Jillela, Raghavender R., and Arun Ross. "Adaptive Frame Selection for Improved Face Recognition in Low-Resolution Videos." *Proceedings of the 8th IEEE International Conference on Automatic Face and Gesture Recognition*, 2009: 1439 - 1445.

Kung, S, Y., M. W. Mak, and S. H. Lin. *Biometric Authentication A machine learning approach.* New Jersey: Prentice Hall, 2005.

Lew, Michael S., Nicu Sebe, Chabane Djeraba, and Ramesh Jain. "Content-Based Multimedia Information Retrieval: State of the Art and Challenges." *ACM Transactions on Multimedia Computing, Communications, and Applications* , vol. 2, no. 1 (February 2006): 1–19.

Li, Stan Z., and Anil K. Jain. "Introduction." In *Handbook of Face Recognition*, by Stan Z. Li and Anil K. Jain, 1-11. Springer , 2005.

Li, Ze-Nian, and Mark S. Drew. *Fundamentals of Multimedia.* USA: Pearson Prentice Hall, 2004.

Lin, Shang-Hung, Sun-Yuan Kung, and Long-Ji Lin. "Face Recognition/Detection by Probabilistic Decision-Based Neural Network." *IEEE Transactions on Neural Networks* , vol. 8, no. 1 (January 1997): 114-132.

Liu, Xiaoming, and Tsuhan Chen. "Video-based face recognition using adaptive hidden markov models." *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* , vol.1 (2003): I-340 - I-345.

Luo, Bing, Yun Zhang, and Yun-Hong Pan. "Face Recognition Based on Wavelet Transform and SVM." *IEEE International Conference on Information Acquisition*, 2005: 373-377.

Park, Unsang. "Face Recognition: face in video, age invariance, and facial marks." PhD dissertation, Michigan State University, 2009.

Sellahewa, Harin. "Wavelet-based Automatic Face Recognition for Constrained Devices." PhD dissertation, The University of Buckingham, Buckingham, 2006.

Sirovich, L., and M. Kirby. "Low-dimensional procedure for the characterization of human faces." *Journal of the Optical Society of America* , vol. 4 (1987): 519 - 524.

Thomas, Deborah, Kevin W. Bowyer, and Patrick J. Flynn. "Strategies for Improving Face Recognition from Video." In *Advances in Biometrics Sensors, Algorithms and Systems*, by Nalini K. Ratha and Venu Govindaraju, 339-361. London: Springer, 2008.

Topkaya, Ibrahim Saygin, and Nilgun Guler Bayazit. "Improving Face Recognition from Video with Preprocessed Representative Faces." *23rd International Symposium on Computer and Information Sciences*, 2008: 1 - 4 .

Turk, Matthew, and Alex Pentland. "Eigenfaces for Recognition." *Journal of Cognitive Neuroscience* , vol. 3, no. 1 (1991): 71-86.

Vaughan, Tay. *Multimedia: Making It Work.* 7th edn. NewYork: McGrawHill, 2008.

Wang, Huafeng, Yunhong Wang, and Yuan Cao. "Video-based Face Recognition: A Survey." (World Academy of Science, Engineering and Technology) , vol. 60 (2009): 293-302.

Wang, Zhifei, Zhenjiang Miao, and Chao Zhang. "Extraction of High-Resolution Face Image From Low-Resolution and Variant Illumination Video Sequences." *Congress on Image and Signal Processing* , vol 4. (2008): 97 - 101.

Webopedia Online Dictionary. 2010. http://www.webopedia.com/TERM/D/digital_video.html (accessed July 1, 2010).

Wheeler, Frederick, Xiaoming Liu, and Peter Tu. "Multi-Frame Super-Resolution for Face Recognition." *First IEEE International Conference on Biometrics: Theory, Applications, and Systems*, 2007: 1 - 6 .

Wiskott, Laurenz, Jean-Marc Fellous, Norbert Krüger, and Christoph von der Malsburg. "Face Recognition by Elastic Bunch Graph Matching." *IEEE Transactions on Pattern Analysis and Machine Intelligence* , vol. 19, no. 7 (July 1997): 775-779.

Zhao, W., R. Chellappa, P. J. Phillips, and A. Rosenfeld. "Face Recognition: A Literature Survey." *ACM Computing Surveys* , vol. 35, no. 4 (December 2003): 399–458.

Zhou, Shaohua, Volker Krueger, and Rama Chellappa. "Face recognition from video: A condensation approach." *Proceedings of the fifth IEEE International Conference on Automatic Face and Gesture Recognition*, 2002: 221 - 226.

Zhu, Hongwei, Caijiao Xue, and Chunyan You. "Video Denoising Using Spatio-temporal Filtering." *ACM*, 2007.

Appendix

A Results using CityBlock measure

A.1 Performance using single frame per subject in each distance range

A.1.1 Performance after HE

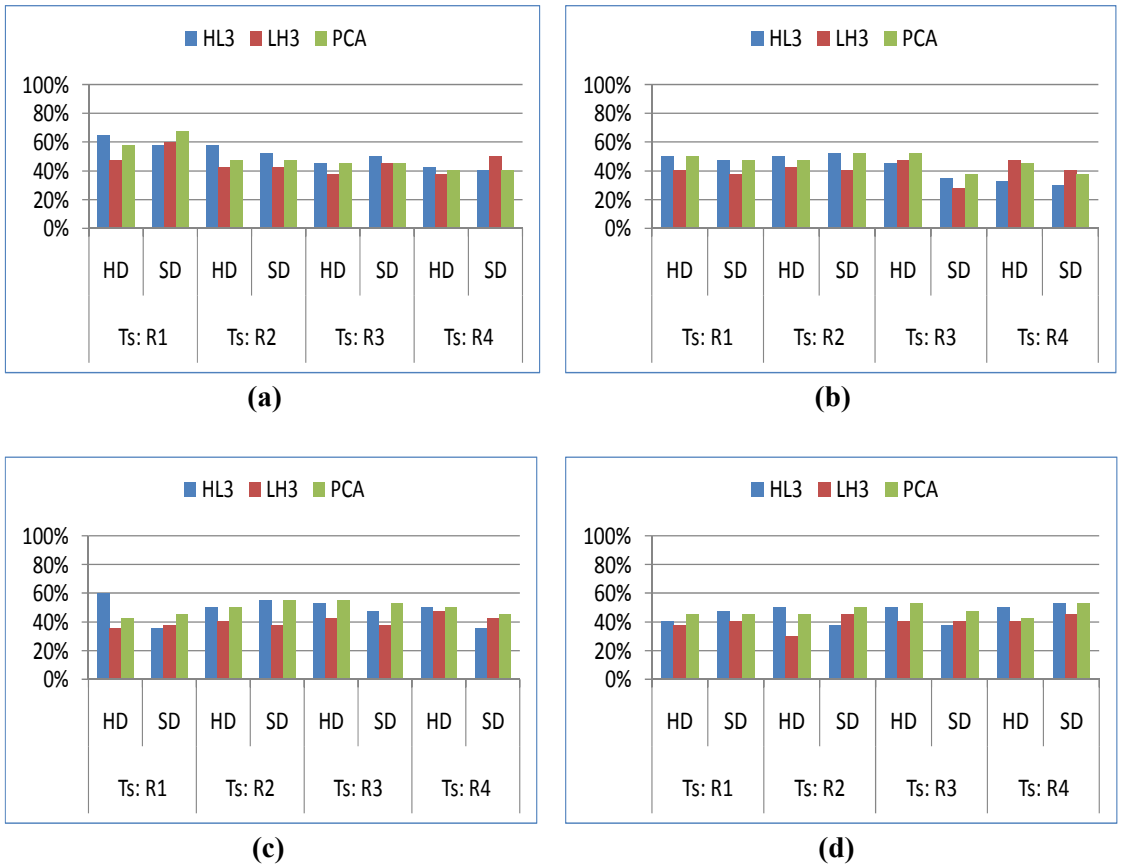
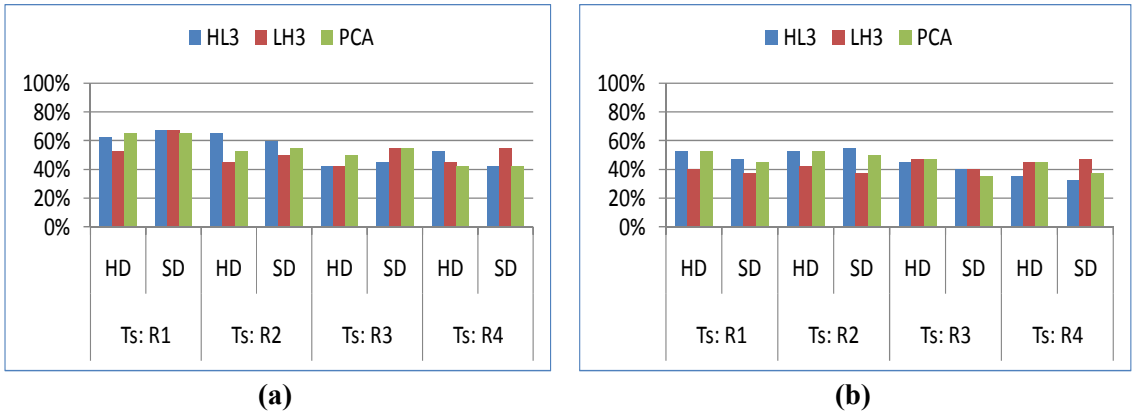
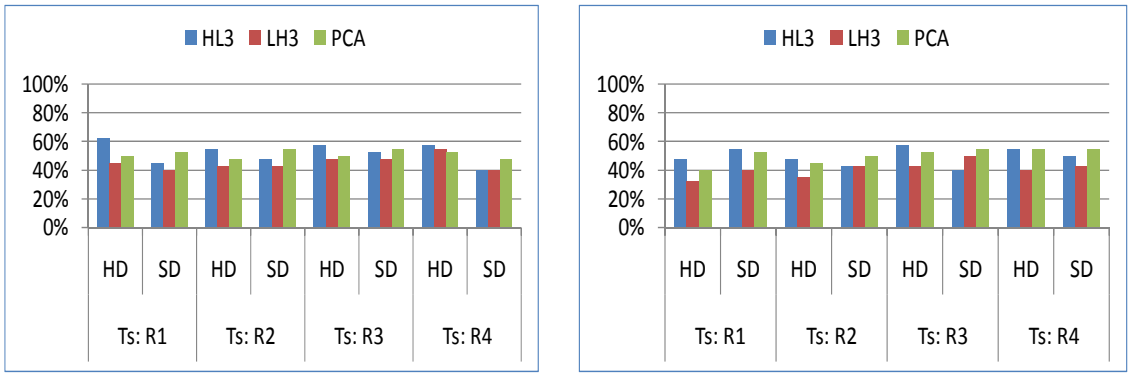


Figure 7.1 Accuracy rates after HE using single frames in range
(a) R1 (b) R2 (c) R3 (d) R4 as a gallery set

A.1.2 Performance after ZN





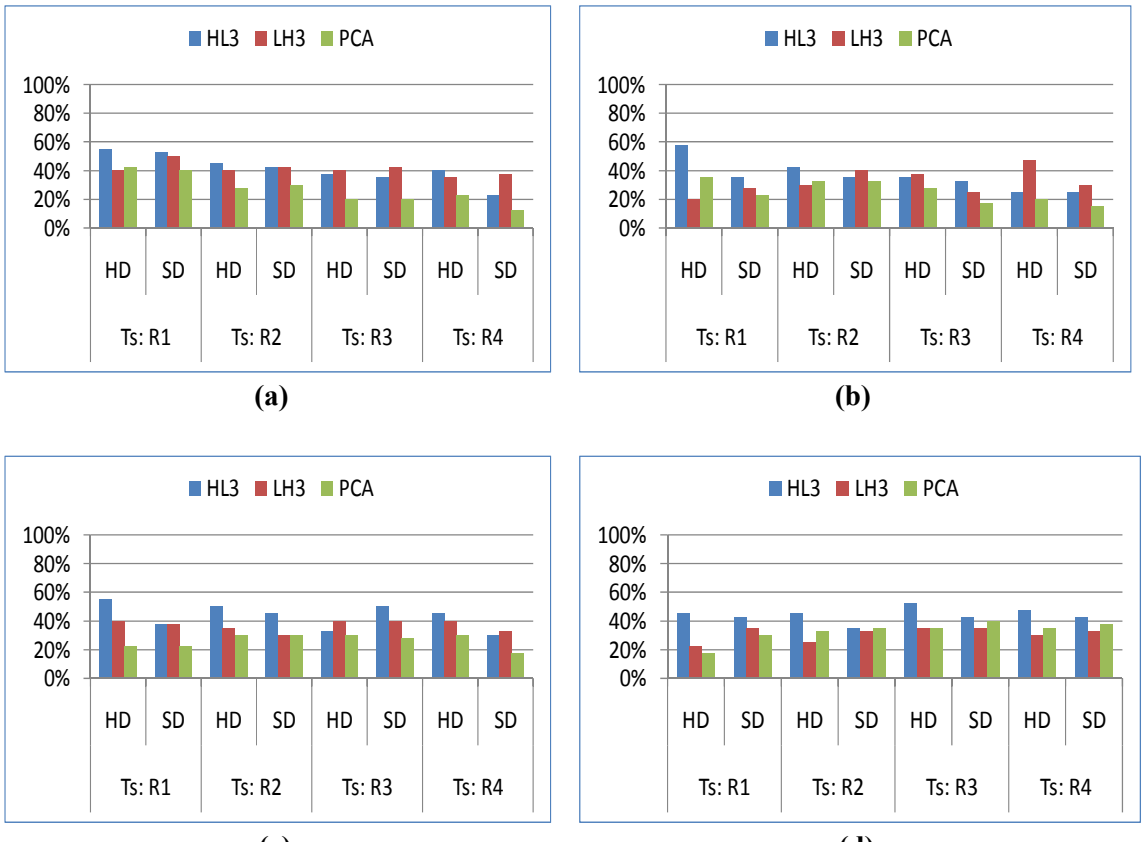
(c) **(d)**

Figure 7.2 Accuracy rates after ZN using single frames in range
(a) R1 (b) R2 (c) R3 (d) R4 as a gallery set

B Results using Euclidean measure

B.1 Performance using single frame per subject in each distance range

B.1.1 Performance prior to illumination normalisation



(a) **(b)**

(c) **(d)**

Figure 7.3 Accuracy rates using single frames in range
(a) R1 (b) R2 (c) R3 (d) R4 as a gallery set

B.1.2 Performance after HE

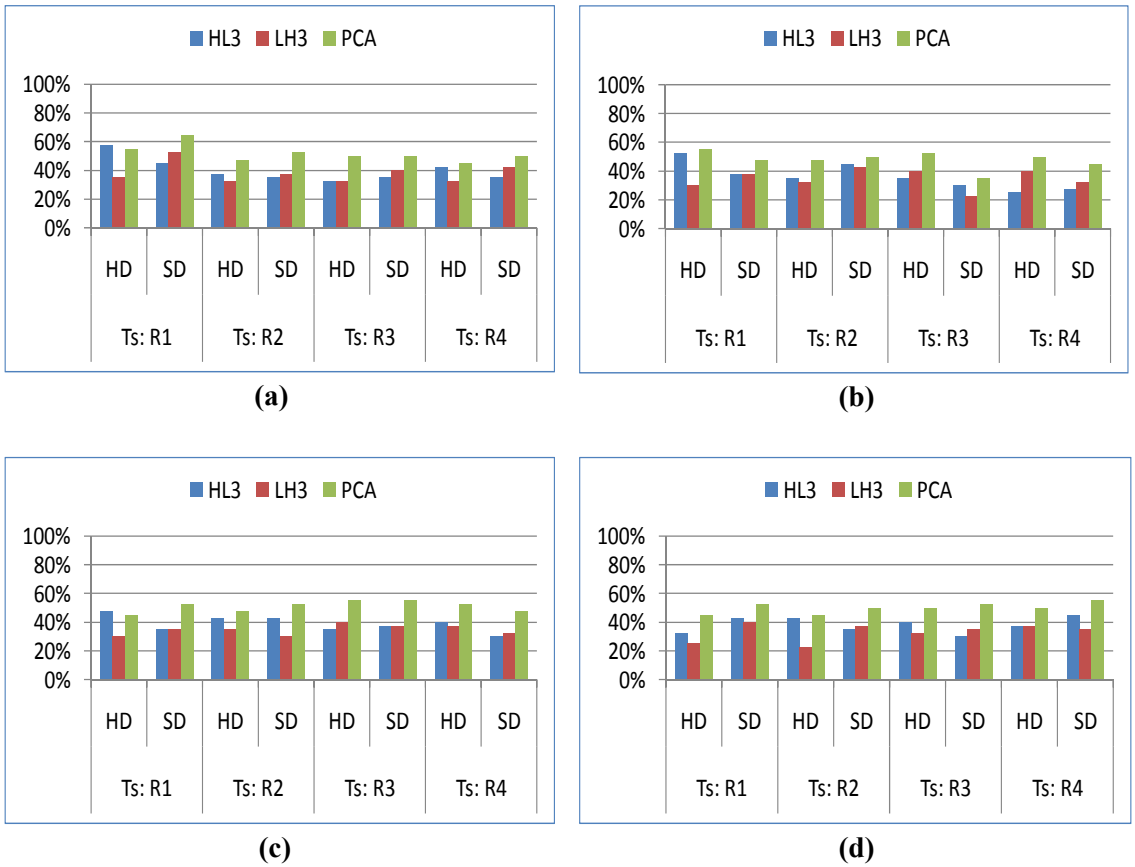
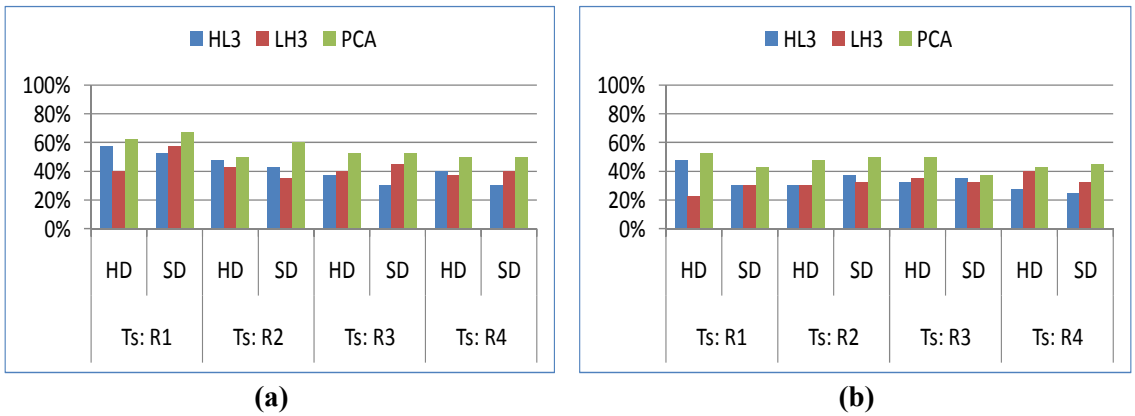
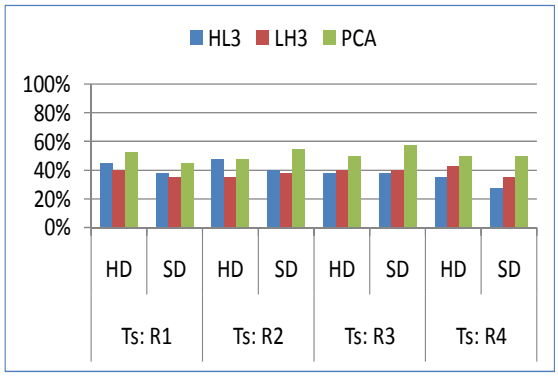


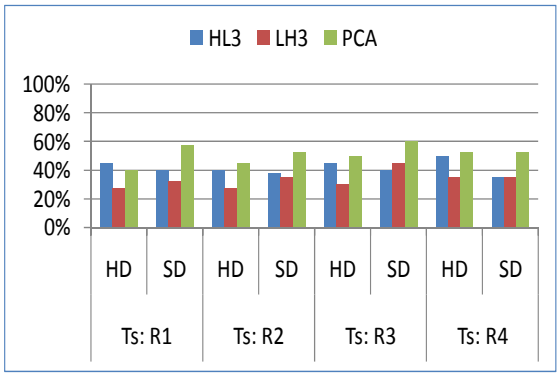
Figure 7.4 Accuracy rates after HE using single frames in range
(a) R1 (b) R2 (c) R3 (d) R4 as a gallery set

B.1.3 Performance after ZN





(c)

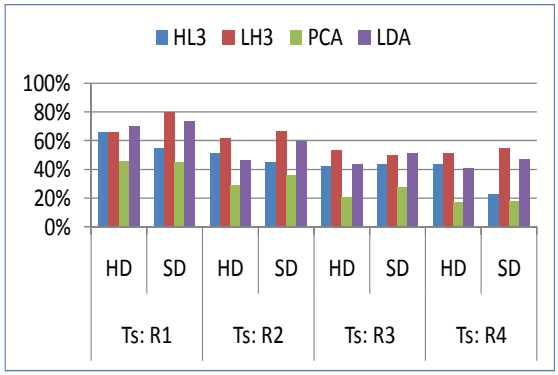


(d)

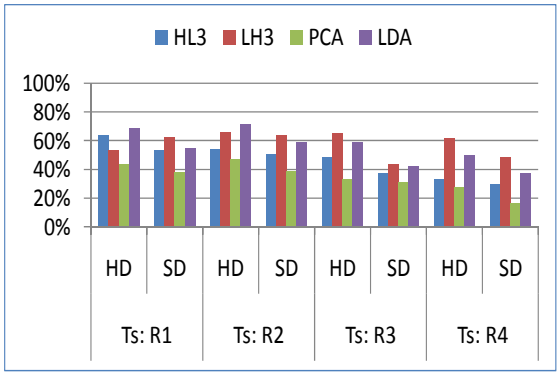
Figure 7.5 Accuracy rates after ZN using single frames in range
(a) R1 (b) R2 (c) R3 (d) R4 as a gallery set

B.2 Performance using three frames per subject in each distance range

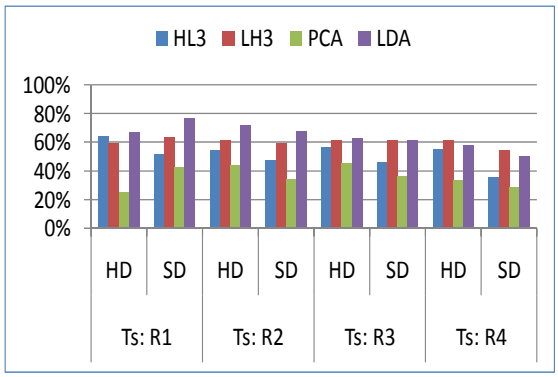
B.2.1 Performance prior to illumination normalisation



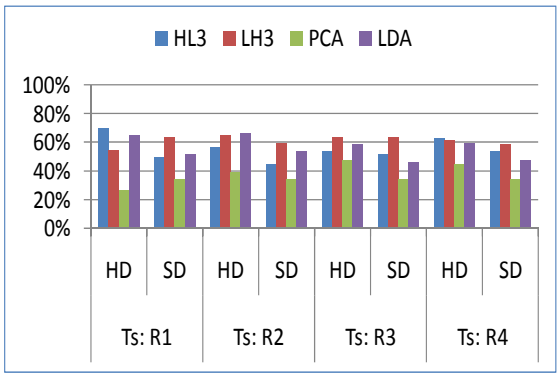
(a)



(b)



(c)



(d)

Figure 7.6 Accuracy rates using frames in range
(a) R1 (b) R2 (c) R3 (d) R4 as a gallery set

B.2.2 Performance after HE

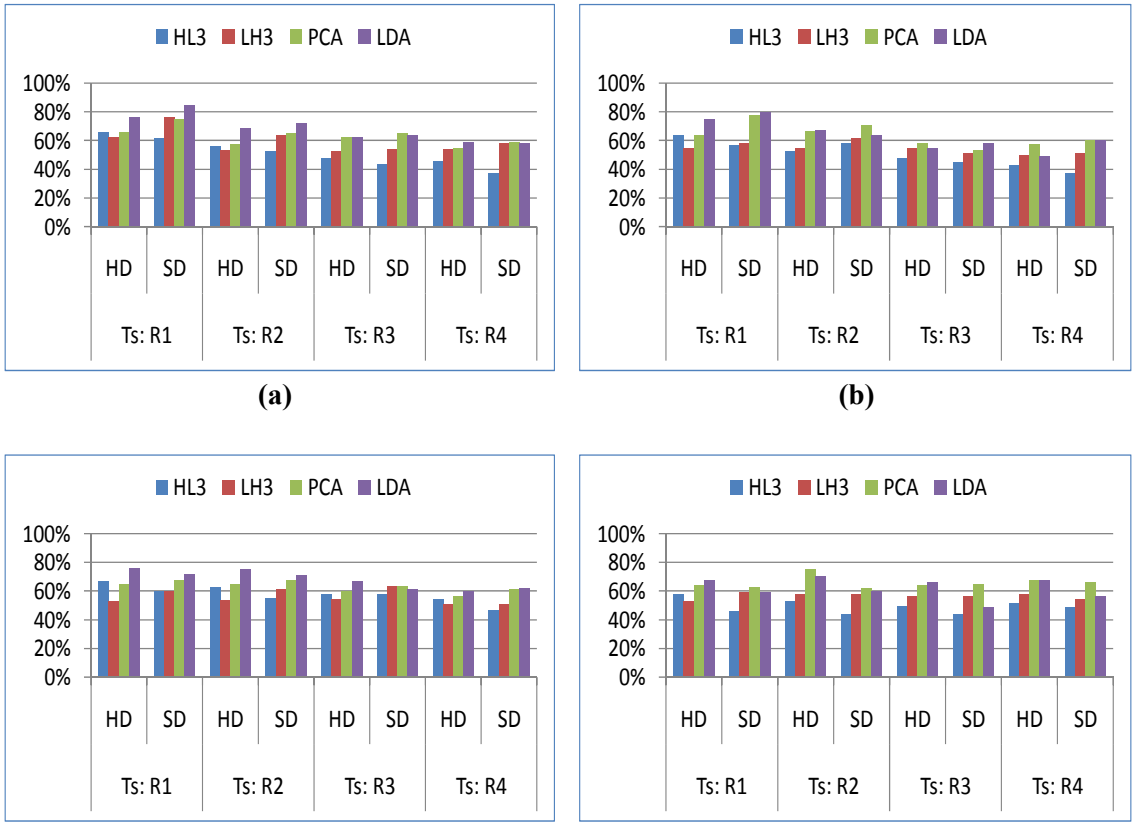
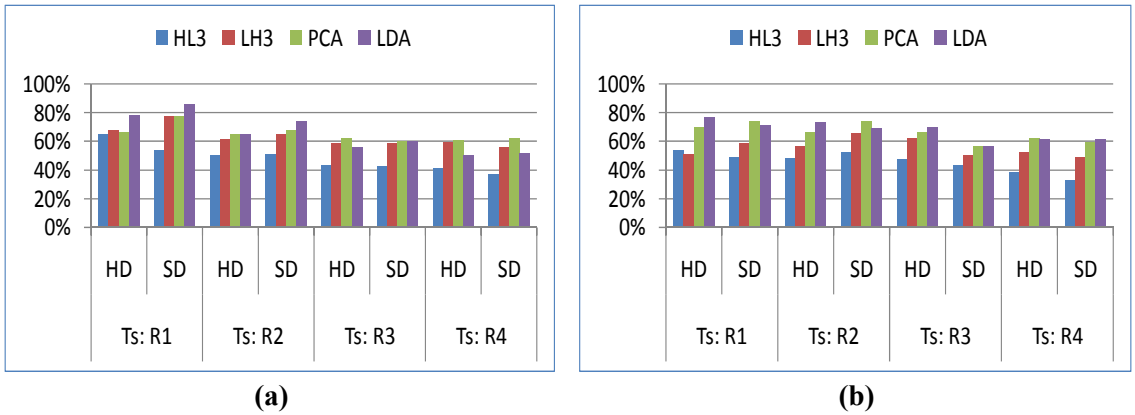
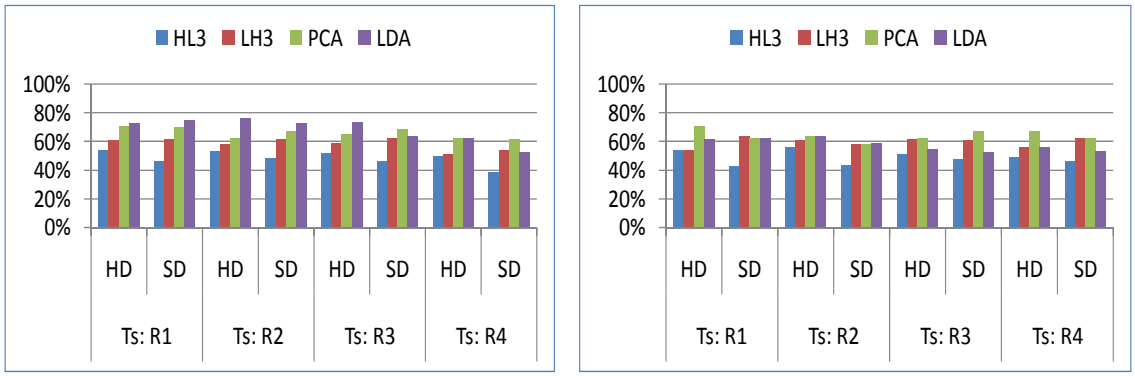


Figure 7.7Accuracy rates after HE using frames in range
(a) R1 (b) R2 (c) R3 (d) R4 as a gallery set

B.2.3 Performance after ZN





(c) **(d)**

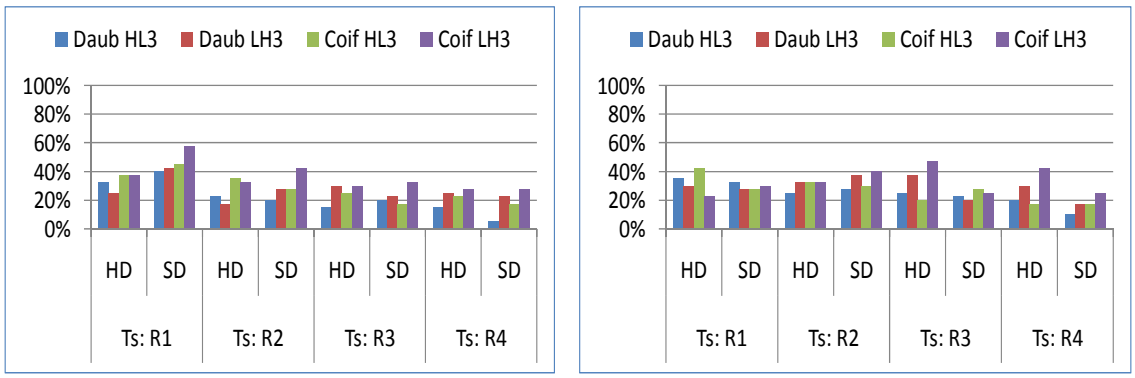
Figure 7.8 Accuracy rates after ZN using frames in range
(a) R1 (b) R2 (c) R3 (d) R4 as a gallery set

C Results using Daubechie-4 and Coiflet-1 filters

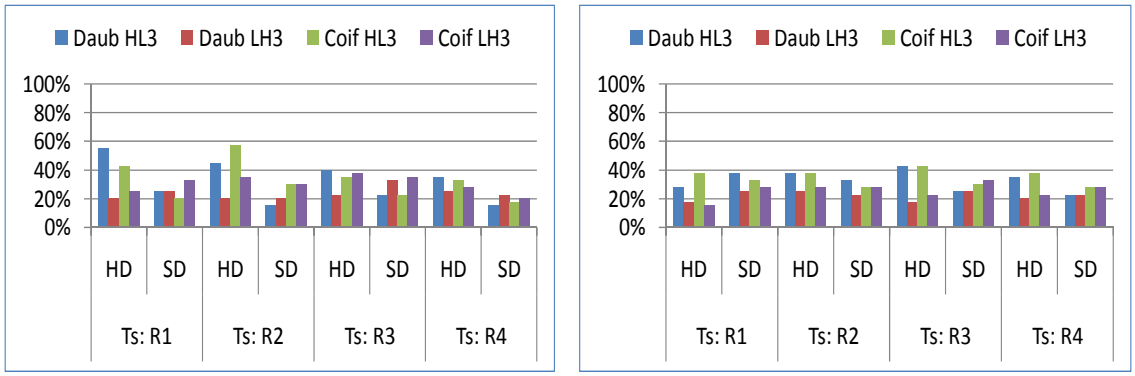
C.1 Using CityBlock measure

C.1.1 Performance using single frame per subject in each distance range

C.1.1.1 Performance prior to illumination normalisation



(a) **(b)**



(c) **(d)**

Figure 7.9 Accuracy rates using single frames in range
(a) R1 (b) R2 (c) R3 (d) R4 as a gallery set

C.1.1.2 Performance after HE

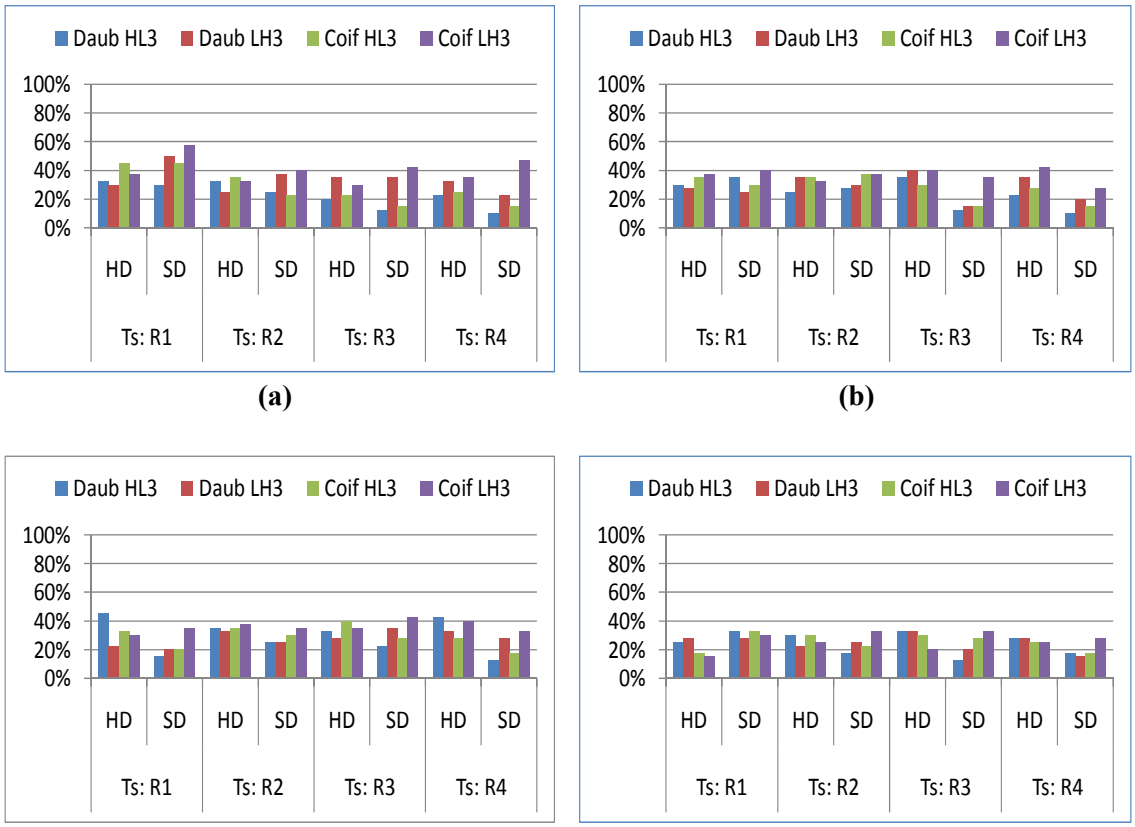
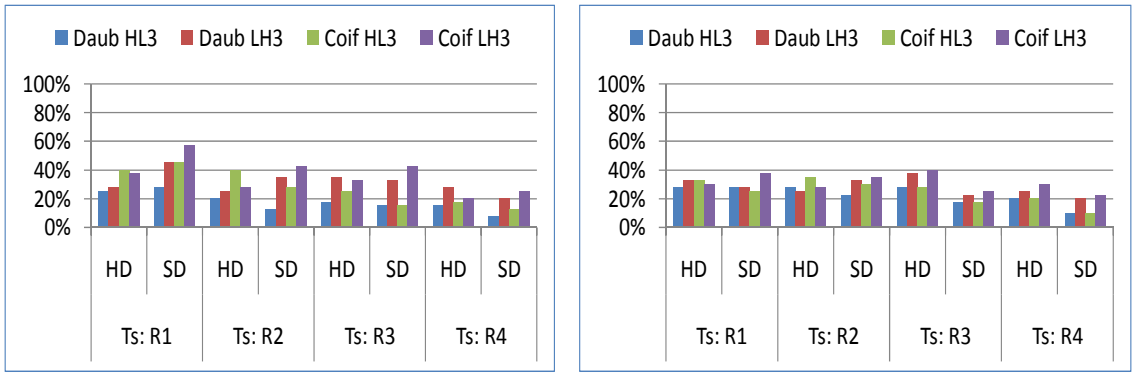
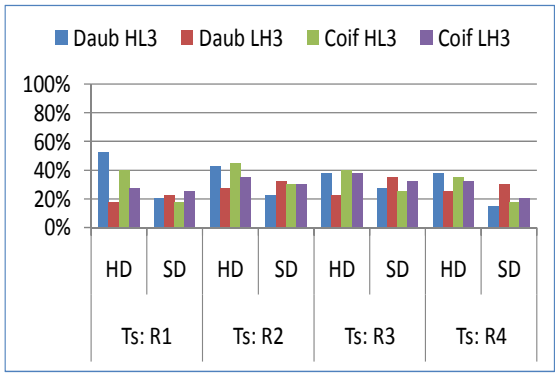


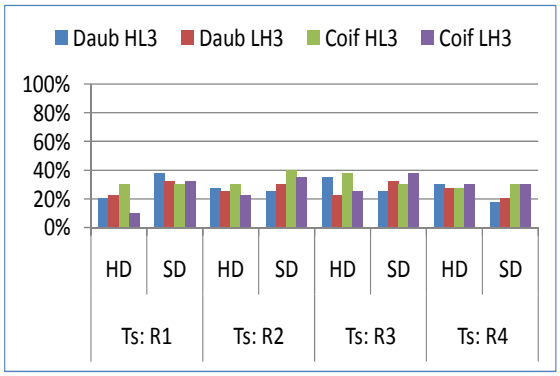
Figure 7.10 Accuracy rates after HE using single frames in range
(a) R1 (b) R2 (c) R3 (d) R4 as a gallery set

C.1.1.3 Performance after ZN





(c)

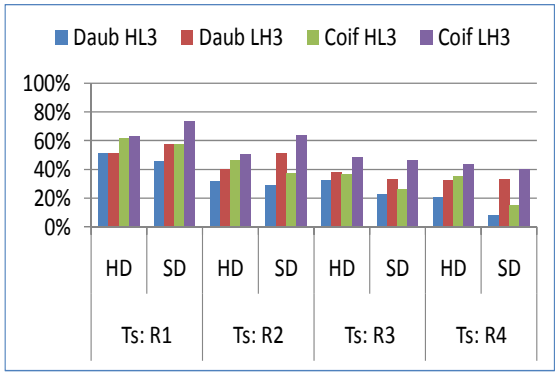


(d)

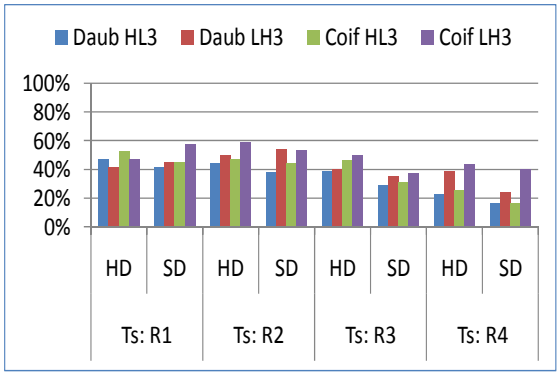
Figure 7.11 Accuracy rates after ZN using single frames in range
(a) R1 (b) R2 (c) R3 (d) R4 as a gallery set

C.1.2 Performance using three frames per subject in each distance range

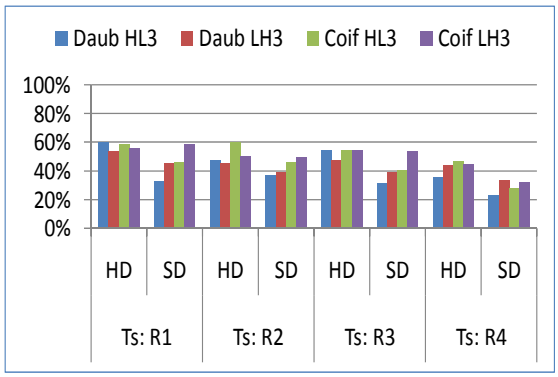
C.1.2.1 Performance prior to illumination normalisation



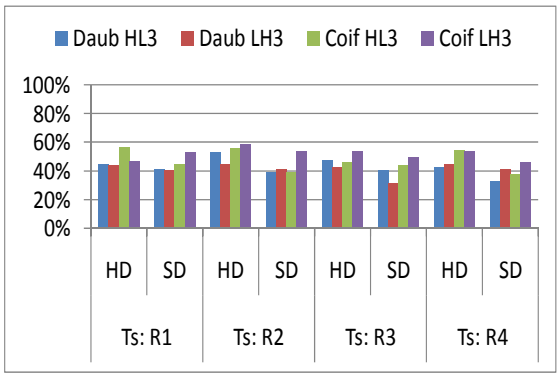
(a)



(b)



(c)



(d)

Figure 7.12 Accuracy rates using frames in range
(a) R1 (b) R2 (c) R3 (d) R4 as a gallery set

C.1.2.2 Performance after HE

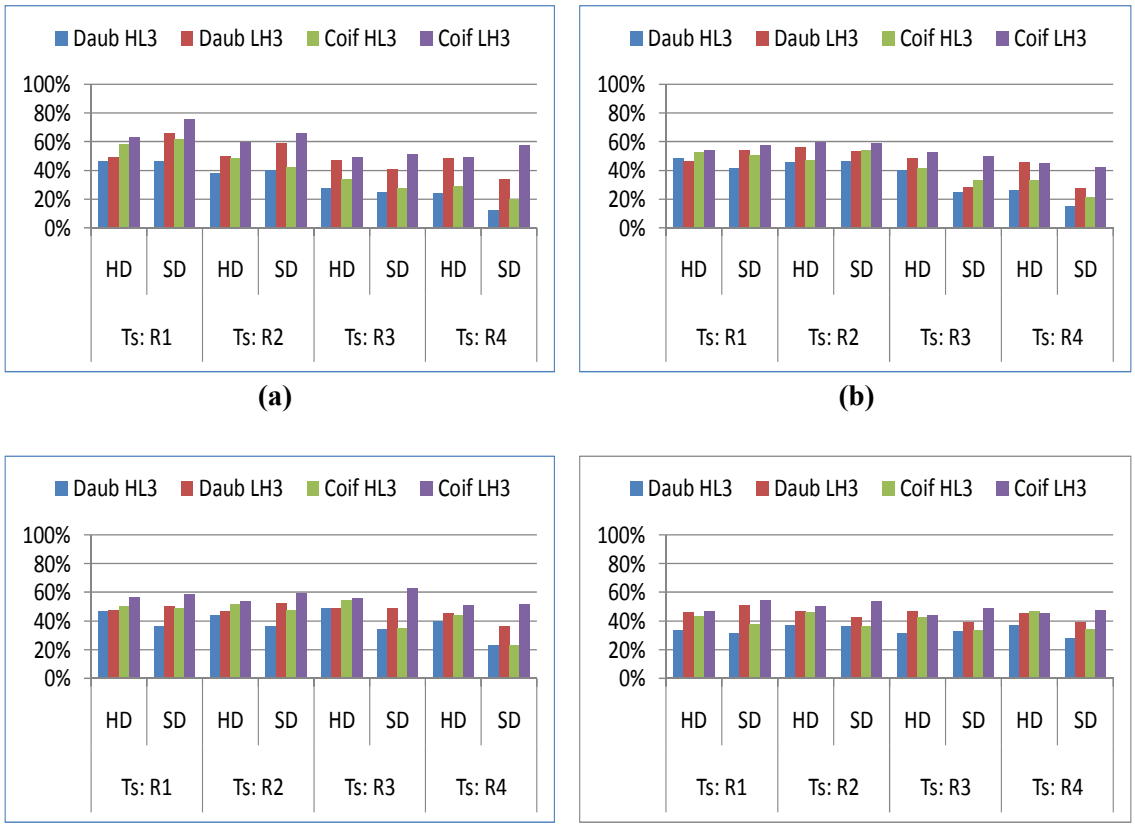
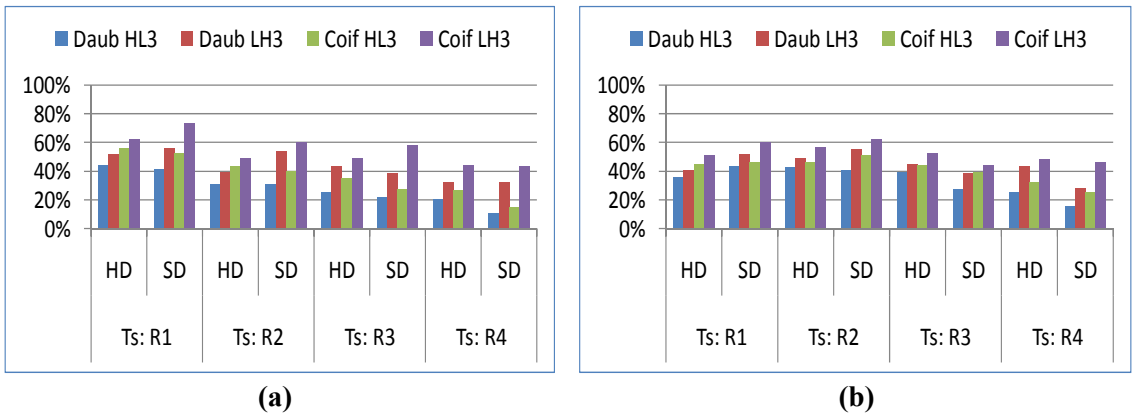
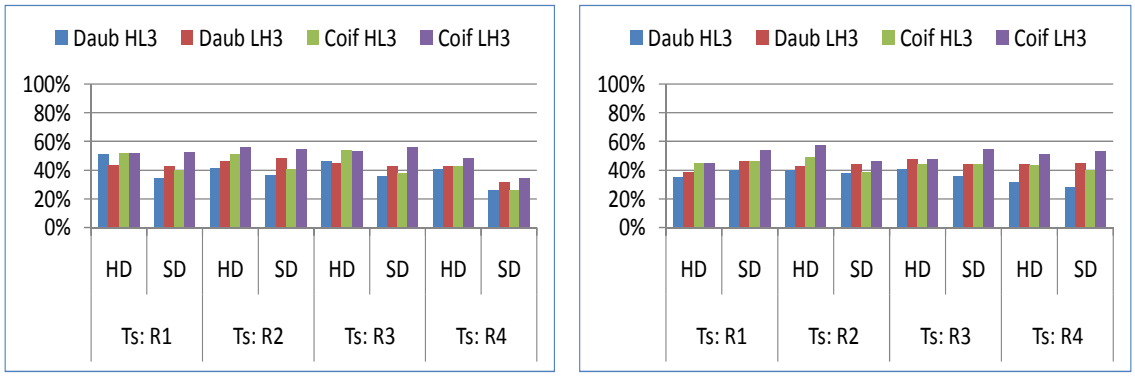


Figure 7.13 Accuracy rates after HE using frames in range
(a) R1 (b) R2 (c) R3 (d) R4 as a gallery set

C.1.2.3 Performance after ZN





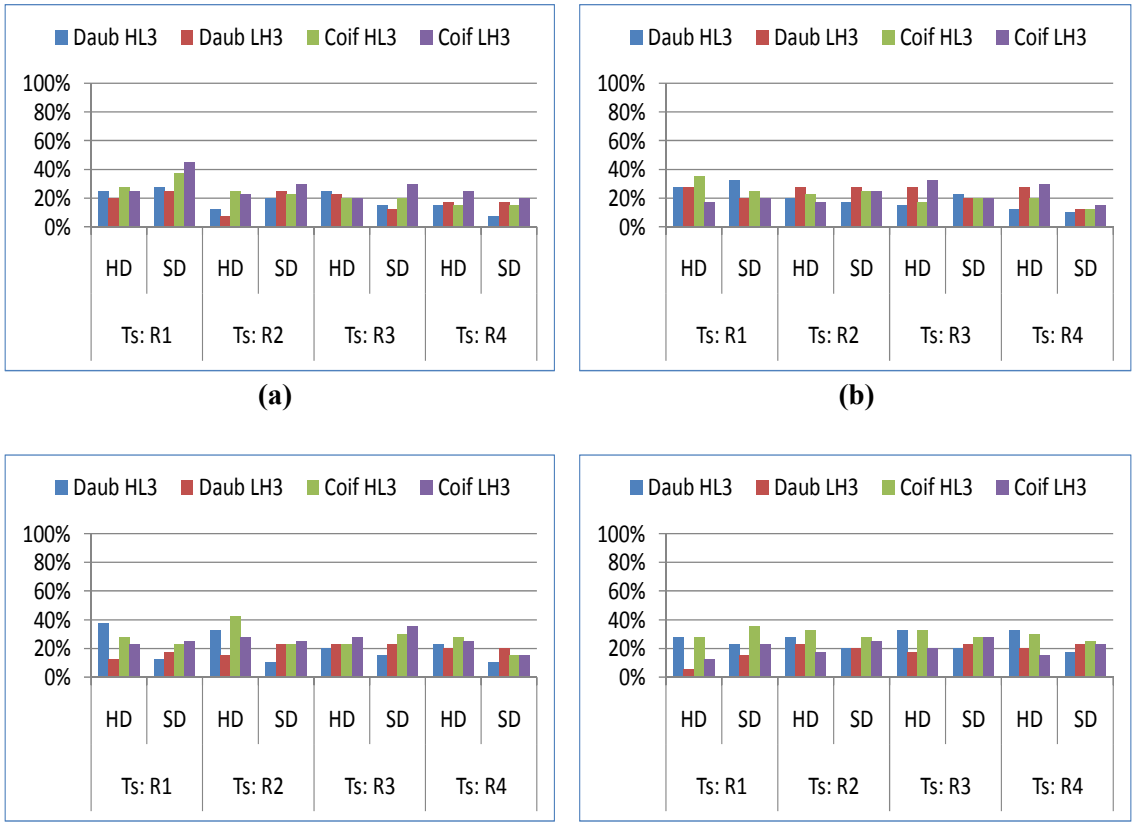
(c) (d)

Figure 7.14 Accuracy rates after ZN using frames in range
(a) R1 (b) R2 (c) R3 (d) R4 as a gallery set

C.2 Using Euclidean measure

C.2.1 Performance using single frame per subject in each distance range

C.2.1.1 Performance prior to illumination normalisation



(a) (b)

(c) (d)

Figure 7.15 Accuracy rates using single frames in range
(a) R1 (b) R2 (c) R3 (d) R4 as a gallery set

C.2.1.2 Performance after HE

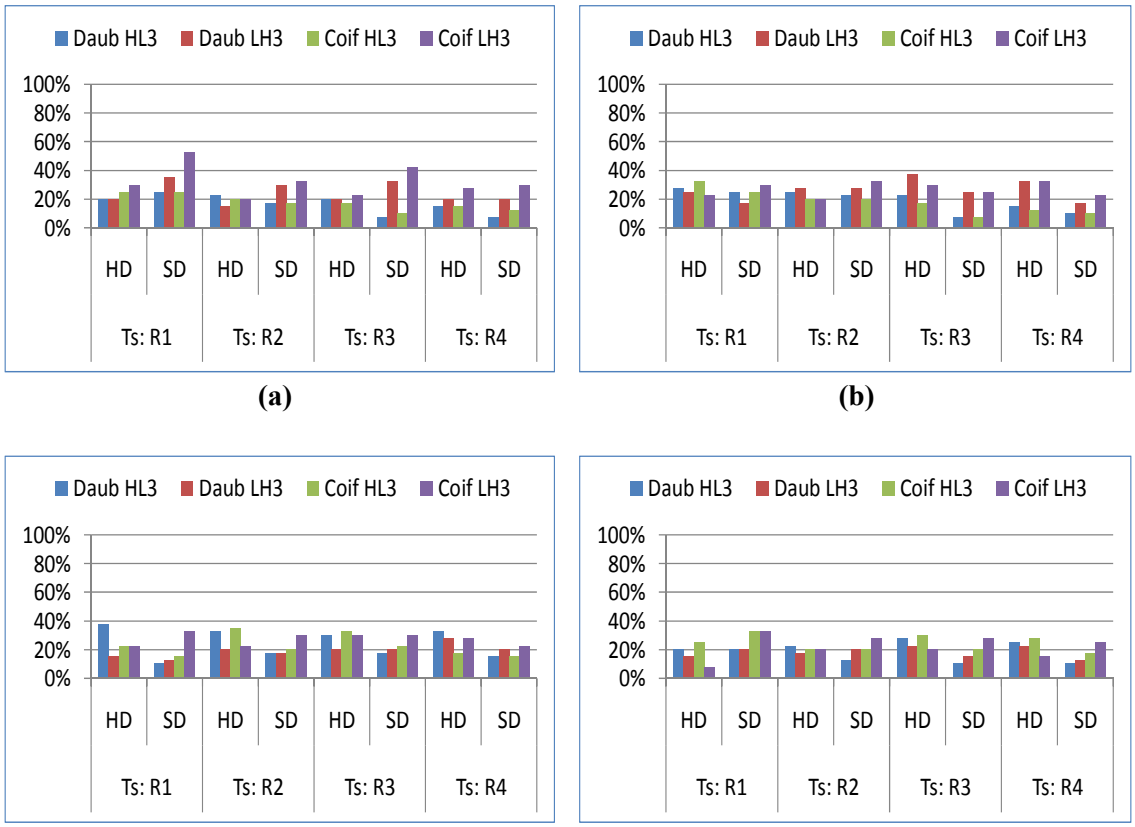
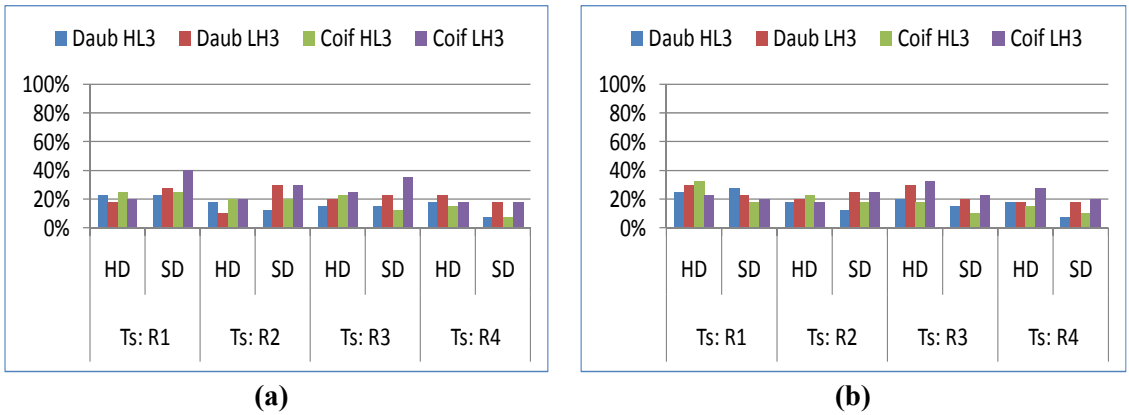
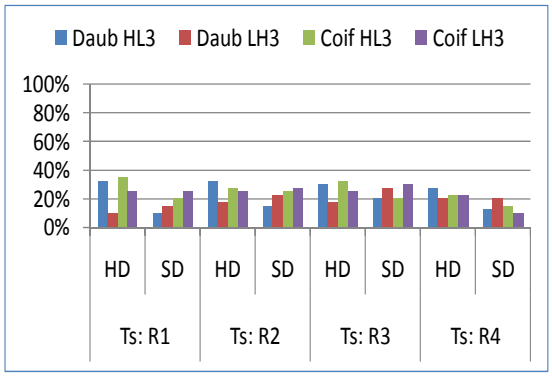


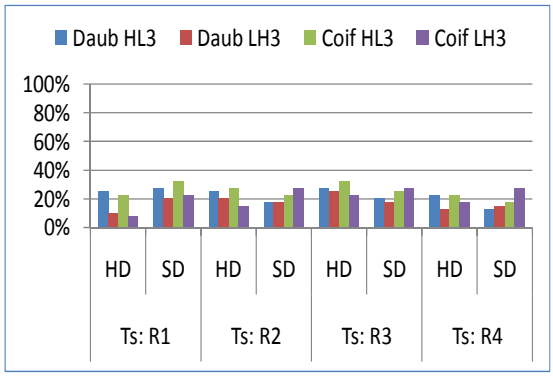
Figure 7.16 Accuracy rates after HE using single frames in range
(a) R1 (b) R2 (c) R3 (d) R4 as a gallery set

C.2.1.3 Performance after ZN





(c)

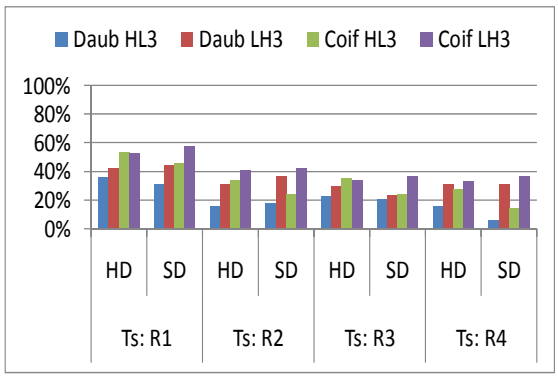


(d)

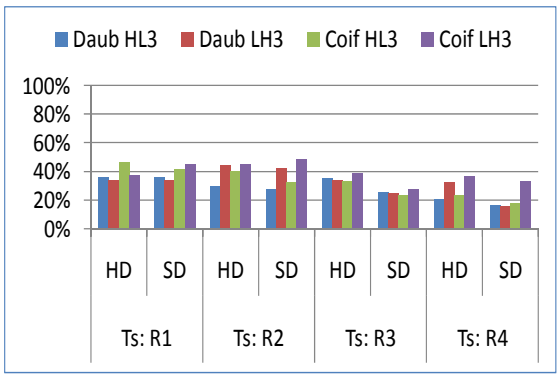
Figure 7.17 Accuracy rates after ZN using single frames in range
(a) R1 (b) R2 (c) R3 (d) R4 as a gallery set

C.2.2 Performance using three frames per subject in each distance range

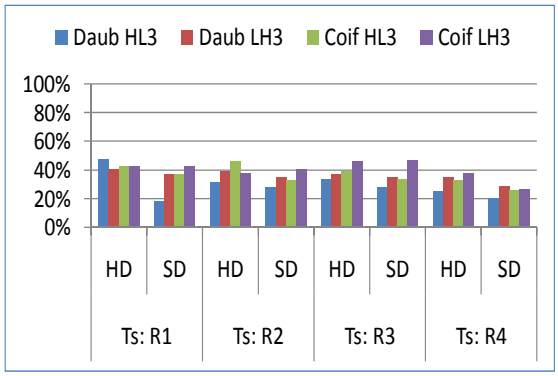
C.2.2.1 Performance prior to illumination normalisation



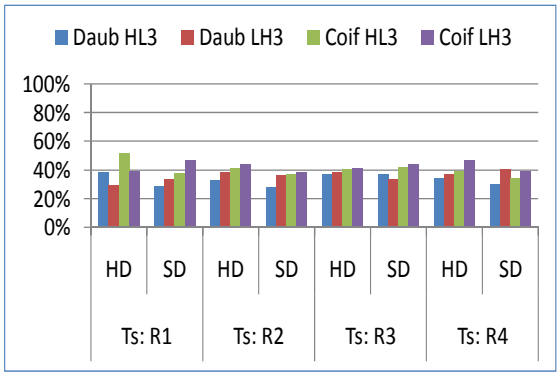
(a)



(b)



(c)



(d)

Figure 7.18 Accuracy rates using frames in range
(a) R1 (b) R2 (c) R3 (d) R4 as a gallery set

C.2.2.2 Performance after HE

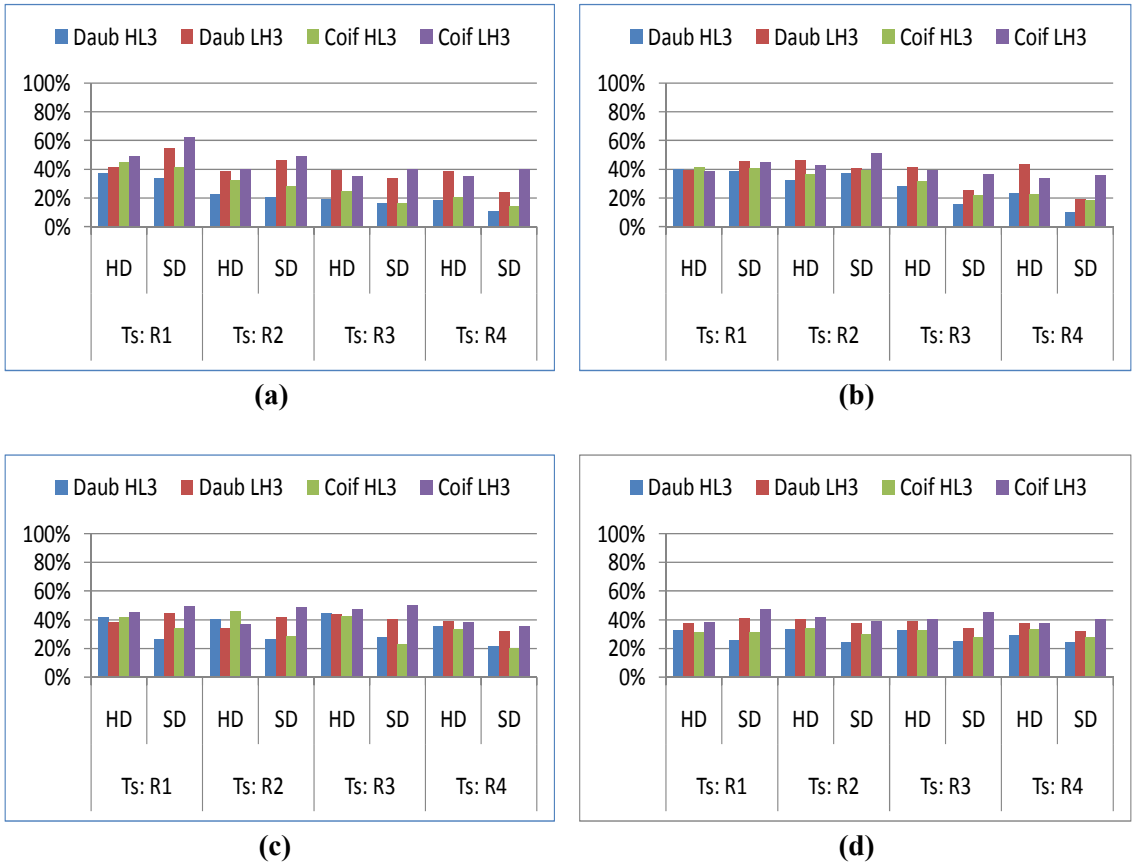
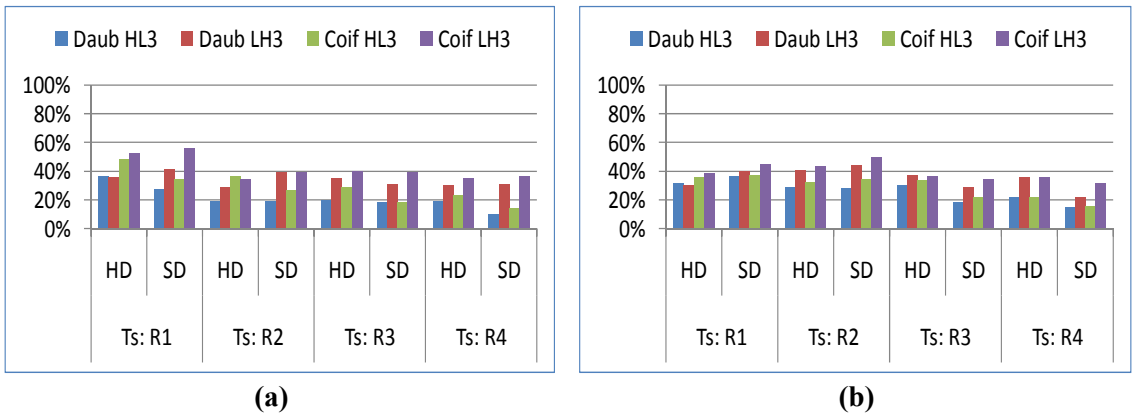
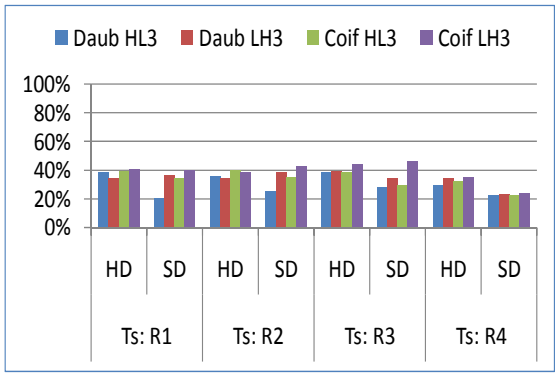


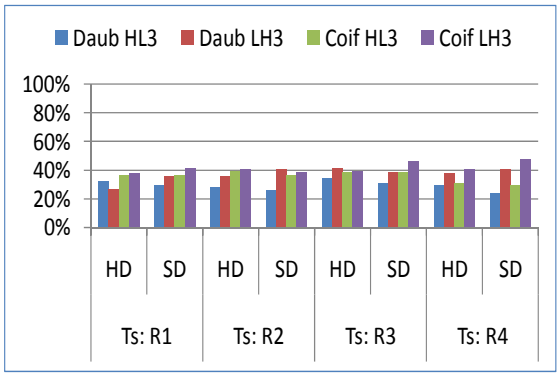
Figure 7.19 Accuracy rates after HE using frames in range
(a) R1 (b) R2 (c) R3 (d) R4 as a gallery set

C.2.2.3 Performance after ZN





(c)



(d)

Figure 7.20 Accuracy rates after ZN using frames in range
(a) R1 (b) R2 (c) R3 (d) R4 as a gallery set