UNIVERSITY OF SOUTHAMPTON

Wavelet Based Approaches for Detection and Recognition in Ear Biometrics

by

Mina Ibrahim Samaan Ibrahim

A thesis submitted in partial fulfillment for the

degree of Doctor of Philosophy

in the

Faculty of Physical and Applied Sciences

School of Electronics and Computer Science

June 2012

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF PHYSICAL AND APPLIED SCIENCES
SCHOOL OF ELECTRONICS AND COMPUTER SCIENCE

Doctor of Philosophy

by Mina Ibrahim Samaan Ibrahim

One of the most recent trends in biometrics is recognition by ear appearance in head profile images. Ear localization to determine the region of interest containing ears is an important step in an ear biometric system. To this end, we propose a robust, simple and effective method for ear detection from profile images by employing a bank of curved and stretched Gabor wavelets, known as banana wavelets. Our analysis shows that the banana wavelets demonstrate better performance than Gabor wavelets technique for ear localization. This indicates that the curved wavelets are advantageous for the detection of curved structures such as ears. This ear detection technique is fully automated, has encouraging performance and appears to be robust to degradation by noise. Addition of a preprocessing stage, based on skin detection using colour and texture, can improve the detection results even further.

For recognition, we convolve the banana wavelets with an ear image and then apply local binary pattern (LBP) for texture analysis to the convolved image. The LBP histograms of the produced image are then used as features to describe an ear. A histogram intersection technique is then applied on the LBP histograms of two ears to measure their similarity for recognition. Analysis of variance is also exploited here to select features to identify the best banana filters for the recognition process. We show that the new banana wavelets, in combination with other analysis, can be used to achieve recognition by the ear, with practical advantages. The analyses focus particularly in simulating addition of noise and occlusion to a standard database, and their evaluation on a newer and much more demanding ear database.

We also present an experimental study to investigate the effect of time difference between image acquisition for gallery and probe on the performance of ear recognition. This experimental research is the first study on the effect of time on ear biometrics and show that the recognition rate remains unchanged over time, confirming another advantage of deploying the human ear as a biometric.

Contents

1 Introduction………………………………………………………………………………………… 1

1.1 Ear Biometrics……………………………………………………………………………… 1

1.2 Research Context and Contributions…………………………………………………… 4

1.3 Thesis Outline……………………………………………………………………………… 5

1.4 List of Publications………………………………………………………………………… 6

2 Previous Work ………………………………………………………………………………… 7

2.1 Ear Detection……………………………………………………………………………… 8

2.2 2D Ear Recognition………………………………………………………………………… 16

2.3 3D Ear Recognition………………………………………………………………………… 26

2.4 Multimodal Biometrics…………………………………………………………………… 27

3 Ear Detection ………………………………………………………………………………… 31

3.1 Banana Wavelets…………………………………………………………………………… 31

3.2 Ear Detection……………………………………………………………………………… 34

3.3 Effects on Ear Detection…………………………………………………………………… 45

3.3.1 Occlusion…………………………………………………………………………… 45

3.3.2 Noise………………………………………………………………………………… 47

3.4 Preprocessing Algorithm…………………………………………………………………… 48

3.4.1 Skin Detection……………………………………………………………………… 48

3.4.1.1 Skin Texture…………………………………………………………………… 49

3.4.1.2 Skin Colour…………………………………………………………………… 50

3.4.2 Adaptive Filter………………………………………………………………………… 51

3.5 Conclusions……………………………………………………………………………… 52

4 Experimental Results………………………………………………………………… 53

4.1 Database Description………………………………………………………………… 53

4.1.1 XM2VTS Database……………………………………………………………… 53

4.1.2 SOTON Database………………………………………………………………… 54

4.2 Ear Detection Results………………………………………………………………… 55

4.3 Occlusion Analysis…………………………………………………………………… 62

4.4 Noise Analysis…………………………………………………………………………… 68

4.5 Gabor Filter Analysis………………………………………………………………… 68

4.6 Parameter Sensitivity………………………………………………………………… 71

4.7 Preprocessing…………………………………………………………………………… 74

4.8 Conclusions……………………………………………………………………………… 75

5 Ear Recognition ………………………………………………………………………… 79

5.1 Basic Techniques……………………………………………………………………… 79

5.1.1 Analysis of Variance (ANOVA)……………………………………………… 79

5.1.2 Local Binary Pattern (LBP)…………………………………………………… 81

5.2 The Proposed Recognition Technique…………………………………………… 87

5.2.1 Ear Description…………………………………………………………………… 87

5.2.2 Ear Comparison…………………………………………………………………… 89

5.3 Filter Selection………………………………………………………………………… 90

5.4 Ear Recognition………………………………………………………………………… 90

5.4.1 LBP and ULBP……………………………………………………………………… 93

5.4.2 Similarity Measure………… 97

5.4.3 Banana and Gabor Filters………… 97

5.4.4 Occlusion………… 102

5.4.5 Noise………… 103

5.5 Verification Performance………… 104

5.6 Parameters Sensitivity………… 108

5.7 Conclusions………… 110

6 The Effect of Time on Ear Biometrics………… 111

6.1 Introduction………… 111

6.2 Database Description………… 112

6.3 The Effect of Time on Recognition………… 113

6.4 The Effect of Time on Verification………… 115

6.5 Conclusions………… 118

7 Conclusions and Future Work ………… 119

7.1 Overall Conclusions………… 119

7.2 Future Work………… 120

Appendix A Adaptive Filter ………… 123

References………… 127

List of Figures

Figure 1.1: Invasive (top) and non-invasive (bottom) biometric methods....................3

Figure 1.2: The anatomy of the ear 4

Figure 2.1: The locations of the measurements used in the “Iannarelli System” 8

Figure 2.2: Example of the steps taken to achieve successful ear enrolment. 12

Figure 2.3: The features used in training the AdaBoost 13

Figure 2.4: The accumulator for image (a) is in (c). 15

Figure 2.5: Stages in building the ear biometric graph model 17

Figure 2.6: Convergence field 19

Figure 2.7: The SIFT keypoints for an ear image...23

Figure 2.8: From the left: the accumulator gathering votes for the putative centre; the detected centre; sampling the image along radial lines; the input template for the wavelet analysis. 24

Figure 2.9: A gallery ear image and its associated mask 25

Figure 3.1: Banana Wavelets (top) and Gabor Wavelets (bottom) 33

Figure 3.2: (a)-(h) 8 filters used in this work...34

Figure 3.3: Our ear detection technique 35

Figure 3.4: (a) Input image, and (b)-(i) after convolution with 8 banana filters. 36

Figure 3.5: Convolution of an input image with 8 banana filters 38

Figure 3.6: Regions of interest 39

Figure 3.7: An ear (neighbourhood) and its partition into blocks 40

Figure 3.8: Ear template and the spatial arrangement of the positions of the filters’ responses 41

Figure 3.9: Examples of partitioned ears into blocks 42

Figure 3.10: Many overlapping neighbourhood windows..43

Figure 3.11: Bounded window around an ear 43

Figure 3.12: (a)-(d) The computation for parameters of filter 4 45

Figure 3.13: An occluded ear for top (10% to 60%)..46

Figure 3.14: An occluded ear for left (10% to 60%)..46

Figure 3.15: (a)-(d) The effect of Gaussian noise various η values, and (e)-(h) their histograms...47

Figure 3.16: Ear detection technique with applying preprocessing algorithm 48

Figure 3.17: Skin detection technique 49

Figure 3.18: The co-occurrence matrices for skin and non-skin regions....................50

Figure 3.19: Examples of applying the adaptive filter 52

Figure 4.1: Samples from XM2VTS face profile database..54

Figure 4.2: Placement of cameras and break-beam sensors in system 54

Figure 4.3: Samples from SOTON dataset 1: (a)-(e) show the groups according to Table 4.1...56

Figure 4.4: Samples of ear detection using our technique...57

Figure 4.5: Samples for detection of blurred ears using our technique......................58

Figure 4.6: Samples for detection of rotated ears using our technique 58

Figure 4.7: Samples for detection of occluded ears by earrings using our technique 59

Figure 4.8: Samples for detection in presence of glasses using our technique 59

Figure 4.9: Samples of correct ear detection from group A (SOTON dataset 1).........61

Figure 4.10: Samples of correct ear detection from group C (SOTON dataset 1).......61

Figure 4.11: Detection results for four ears with occlusion from top 63

Figure 4.12: Detection results for four ears with occlusion from left 64

Figure 4.13: Detection rate for ears with occlusion from top and left........................65

Figure 4.14: Samples of correct ear detection from group B (SOTON dataset 1).......66

Figure 4.15: Samples of correct ear detection from group D (SOTON dataset 1).......66

Figure 4.16: Samples of correct ear detection from group E (SOTON dataset 1) 67

Figure 4.17: Samples of incorrect ear detection from group B, group D and group E (SOTON dataset 1) 67

Figure 4.18: Samples for the results in presence of noise..........69

Figure 4.19: Detection rate for banana wavelets technique and Gabor wavelets technique in the presence of noise 70

Figure 4.20: The effect of changing theta parameter on the detection rate 72

Figure 4.21: The effect of changing frequency parameter on the detection rate 73

Figure 4.22: Example of applying our ear detection technique with the preprocessing step..........76

Figure 5.1: The basic LBP operator 82

Figure 5.2: Example for the LBP operator..........82

Figure 5.3: Three different circular neighbourhoods 83

Figure 5.4: The 36 unique rotation invariant binary patterns 84

Figure 5.5: Uniform Local Binary Pattern Histograms 86

Figure 5.6: Our ear description algorithm 88

Figure 5.7: Filter selection algorithm..........91

Figure 5.8: The selected banana filters..........92

Figure 5.9: Samples from SOTON dataset 2..........93

Figure 5.10: Recognition rate versus number of filters applied to XM2VTS database and SOTON dataset 2 using LBP and ULBP 95

Figure 5.11: Confusion matrix of ear recognition for XM2VTS database and SOTON dataset 2 using LBP and ULBP..........96

Figure 5.12: Recognition rate versus number of filters applied to XM2VTS database and SOTON dataset 2 using Intersection and difference 98

Figure 5.13:Confusion matrix of ear recognition for XM2VTS database and SOTON dataset 2 using intersection and difference 99

Figure 5.14: Recognition rate versus number of filters applied to XM2VTS database and SOTON dataset 2 using banana and Gabor filters 100

Figure 5.15:Confusion matrix of ear recognition for XM2VTS database and SOTON dataset 2 using banana and Gabor filters 101

Figure 5.16: Recognition rate for ears with occlusion from top and left for XM2VTS database 102

Figure 5.17: Recognition rate for banana filters and Gabor filters in the presence of noise for XM2VTS database 103

Figure 5.18: Intra-class and inter-class distributions for different databases. 105

Figure 5.19: Decidability measure versus number of filters applied for different databases. 106

Figure 5.20: The ROC curve for XM2VTS and SOTON databases 109

Figure 6.1: Morphological change in configuration of male and female ears during first twelve months. 112

Figure 6.2: Samples from SOTON dataset 3 113

Figure 6.3: Recognition rate over time 114

Figure 6.4: Examples of highly and slightly rotated ears 115

Figure 6.5: The ROC curve SOTON dataset 3 with error bar 116

Figure 6.6: The ROC curve for SOTON dataset 3 117

Figure A.1: The adaptive filter structure 124

List of Tables

Table 1.1: Examples of recognition rates for different biometrics systems.................2

Table 2.1:Detection rates for different ear biometrics systems 16

Table 2.2:Recognition rates for different ear biometrics systems 28

Table 3.1: Examples of (block) positions of filters’ responses for the ears in Figure 3.9 42

Table 3.2: Parameter Settings for the Banana Wavelets...44

Table 4.1: Classification of SOTON dataset 1 into five groups according to head appearance and occlusion 55

Table 4.2: Ear detection rates for SOTON dataset 1 by applying banana and Gabor filters 60

Table 4.3: Parameters sensitivity for ear detection and the change in each parameter 74

Table 4.4: Comparing the correct detection rates for SOTON dataset 1 77

Table 5.1: Examples of uniform and non uniform binary patterns 85

Table 5.2: Comparison of different recognition techniques and our recognition technique (based on XM2VTS database)..92

Table 5.3: Comparison of recognition rate for different techniques 94

Table 5.4: The measured decidability for different databases.................................107

Table 5.5: How changes in parameter values affect the correct recognition rate (CCR) 110

Table 6.1: Number of Subjects and Images in each Session....................................113

Table 6.2: The Combinations of Probe and Gallery 114

Table 6.3: The Decidability and EER values over time 116

Acknowledgements

First of all, I thank my God for achieving this work and giving me the ability to finish it in that distinct and satisfactory form. I would like to express my appreciation to my supervisors Prof. Mark Nixon and Dr. Sasan Mahmoodi for their continuous support and encouragement during the research study in this thesis. They really influenced my way of thinking and developing the research ideas adopted in this thesis.

I must express my gratitude to Eman, my wife, for her continued support and encouragement to finish this work. I also need to thank my sons (Martin and Pola), mother, father, sisters for their support.

I would also like to extend my thanks to father Botros El Moharaky for his prayers. Thanks are also to Mina and Fady Awad for their help. Thank you all …

To my dear wife and sons

Chapter 1

Introduction

1.1 Ear Biometrics

Biometrics recognition systems are one of the most essential security aspects in the modern world. Biometric recognition occupies a vital place in security methods. Biometric recognition refers to the automatic recognition of individuals based on a feature vector derived from their physiological and/or behavioural characteristic.

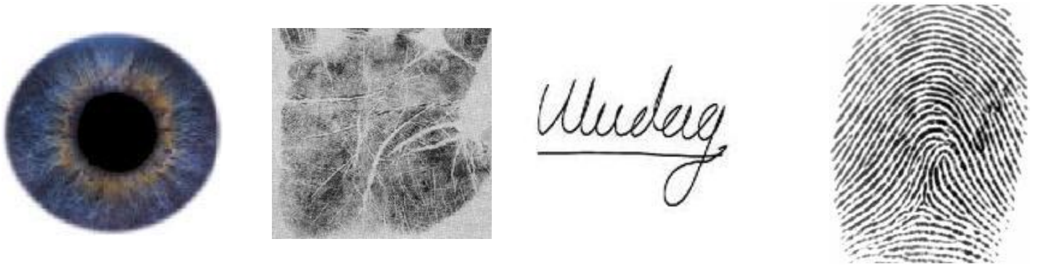
Personal identification aims to associate a particular individual with his/her identity and plays an important role in our society. Any human physiological or behavioural characteristic can be used as indicator to make a personal identification as long as it satisfies the following requirements and issues [1]: universality, uniqueness, permanence, collectability, and acceptability. For a practical biometric system, we must also consider the recognition accuracy. A highly accurate biometric recognition algorithm is required in many applications, such as in security and by the police to identify potential criminals. Table 1.1 shows some examples of recognition performance for different biometrics approaches. As illustrated in the table, it is not unusual for the recognition rate for a biometric system to be above 97%. The techniques have primarily been selected from the most recent edition of the

International Joint Conference on Biometrics (IJCB), which is one of the top conferences in this field. The process of detecting and isolating the area of interest in an automatic biometric system is detection which is largely synonymous with enrolment or registration, when a person's (biometrics) data is entered into a database.

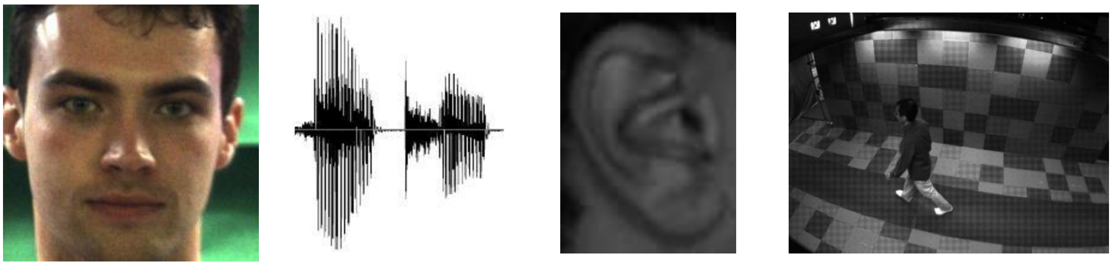
Biometrics		Recognition Rate
Ear	Hurley et al. [2]	99.2%
	Cadavid et al. [3]	97.6%
Face	Juefei-Xu et al. [4]	100%
	Smeets et al. [5]	98.6%
Iris	Daugman [6]	100%
	Ko et al. [7]	98.2%
Fingerprint	Shen and Kot [8]	97.6%
	Win and Sein [9]	99%

Table 1.1: Examples of recognition rates for different biometrics systems

Biometric methods can be divided into invasive and non-invasive methods (Figure 1.1). Invasive methods are those that need the assistance from the individual in order to obtain data required to compare his/her biometric features to the ones stored in a database. Non-invasive biometrics does not require the assistance of the individuals; actually data capture may be done even without their knowledge. Invasive biometrics, for example, fingerprint and iris are unsuitable for some applications such as surveillance since the resolution at which data is acquired is too low. In addition to that, good quality scanners or cameras are required to produce good results for fingerprint and iris based methods. Alternatively, ear and face recognition methods are non-invasive biometric techniques, and they do not need high quality cameras or optics as required for fingerprint or iris biometrics.



(a) iris (b) palmprint (c) signature (d) fingerprint



(e) face (f) voice (g) ear (h) gait

Figure 1.1: Invasive (top) and non-invasive (bottom) biometric methods

Most of the current research in biometrics has focused on face, fingerprint, gait, iris, signature, hand geometry or palmprint. There has been much interest in ear biometrics in recent years. Ear recognition is a physical biometric, which is characterized by the shape of the outer ear, lobes and bone structure.

The shape of the ear includes the outer rim or helix, the ridge or antihelix that runs inside and parallel to the helix, the lobe, and the distinctive u-shaped notch known as the intertragic notch between the ear hole (meatus) and the lobe [10,11]. The antihelix splits into two branches at the upper which forms triangular fossa. The lower of the two branches and the inner helix form the top and left side of the concha. The shape of the ear also includes the crus of helix where the helix intersects with the lower branch of the antihelix. Figure 1.2 shows the anatomy of the ear.

Ear identification can be preferred over other biometric technologies for various reasons. The ear contains a large amount of specific and unique features that allow for human identification. It has a rich and stable structure that does not change significantly with time [12]. The ear is unlike the fingerprint or the iris, it can be easily

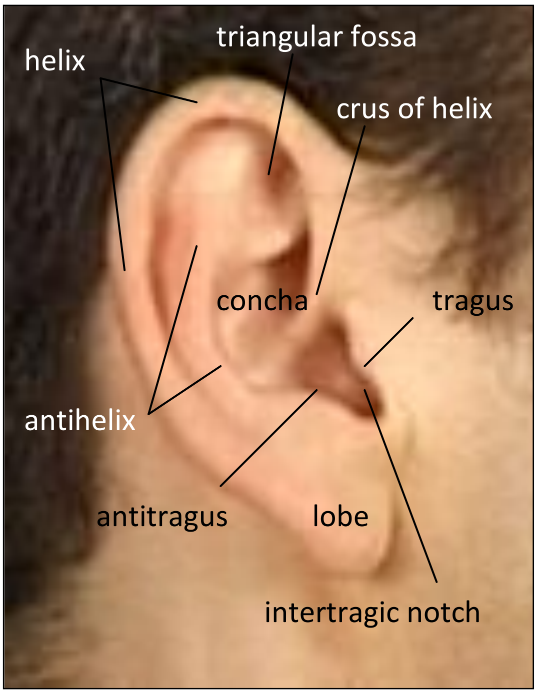


Figure 1.2: The anatomy of the ear

captured from a distance without knowledge of the person under examination. Also, an ear is unlike a face, it does not suffer from changes in facial expression. The ear is fixed in the middle of the both sides of the head so that the immediate background is predictable. The colour distribution is more uniform for the ear than for the human face, iris or retina. These properties make ears a very attractive biometric identifier. As a result the ear biometrics is suitable for security, surveillance, access control and monitoring applications.

1.2 Research Context and Contributions

We contend that it is prudent to continue investigating approaches which consider the ear as a planar surface. This will allow for application in access control and surveillance, and for acquisition from documents. The ear plane is not congruent to that of the head, but it is chosen as such we can consider the ear to be on a flat surface. For enrolment and recognition, we shall need to locate the ear in a profile image automatically with an algorithm which is robust in the presence of noise. It therefore appears appropriate to investigate a technique which depends on the general structure of the ear.

There is a natural interest in frequency domain image analysis as it can be used to derive an alternative representation which is convenient for analysis. Wavelets allow for localised frequency domain analysis. In order to derive an approach which appears more suited to ear biometrics, we employ a bank of banana wavelets, which are Gabor wavelets generalized to extract curvilinear structures. In addition to the frequency and orientation, banana wavelets are also characterized by properties associated with the bending and curvature of the filter. Such properties are essential for ear detection and recognition and such filters therefore produce little response in other parts of the image. These features then appear well matched to the general structure of the ear which has many curvilinear structures, particularly in the region of the helix (the uppermost part of the ear) and the tragus (which are the lower parts).

1.3 Thesis Outline

This thesis is organized as follows: Chapter 2 reviews the main approaches to ear biometrics. Chapter 3 describes our automatic ear detection technique which enrols ears from 2D face profile images using the banana wavelet filters [A]. A preprocessing stage to our ear detection technique is also presented in Chapter 3.Chapter 4 represents the ear detection results for our detection technique. Chapter 5 describes the proposed technique for ear recognition which combines the banana wavelet filters and the local binary pattern operator to recognize the correct subject [B]. In Chapter 6, the first experimental study to investigate the effect of time on ear biometrics is proposed [C]. Finally in Chapter 7,overall conclusions are reviewed and possible directions for further work are discussed. Appendix A describes the adaptive filter.

1.4 List of Publications

[A] M. I. Ibrahim, M. S. Nixon, and S. Mahmoodi, “Shaped Wavelets for Curvilinear Structures for Ear Biometrics,” in *Proc. ISVC’10*, Las Vegas, USA, Nov. 2010.

[B] M. I. Ibrahim, M. S. Nixon, and S. Mahmoodi, “Ear Detection and Recognition By Banana Wavelets,”*Pattern Recognition journal*, submitted

[C] M. I. Ibrahim, M. S. Nixon, and S. Mahmoodi, “The Effect of Time on Ear Biometrics,” in *Proc. IJCB*, Washington DC, USA, Oct. 2011.

Chapter 2

Previous Work

There exist some techniques developed for ear detection and recognition using 2D and 3D images. These systems have been examined and validated on different databases in the literature and its potential for human identification was noted first in the late nineteenth century.

The potential for using the ear’s appearance as a means of personal identification was recognised and advocated by the French criminologist Alphonse Bertillon in 1890 as part of his new scientific method of criminal identification [13] which suggested identification by anthropometry, ear and iris. Alfred Iannarelli [12] performed two studies in 1989 to give a support evidence for recognition by ear. The first study Iannarelli compared over 10,000 ears drawn from a randomly selected sample, and the second study examined fraternal and identical twins. These two studies were based on measuring the distances between specific points of the ear. Iannarelli developed a method where 12 measurements are used as features to distinguish individuals (as illustrated in Figure 2.1). The results from these studies have proved that the ear contains unique features. Even identical twins had similar but not identical ears. This work then awaited automation. There are existing surveys on 2D and 3D ear biometrics described different detection and recognition techniques for ear biometric systems [14, 15, 16, 17]. We shall first consider techniques for finding the ear: ear detection. Then we shall distinguish between approaches which consider the ear as a planar structure, 2D approaches, and those which consider the 3D information.

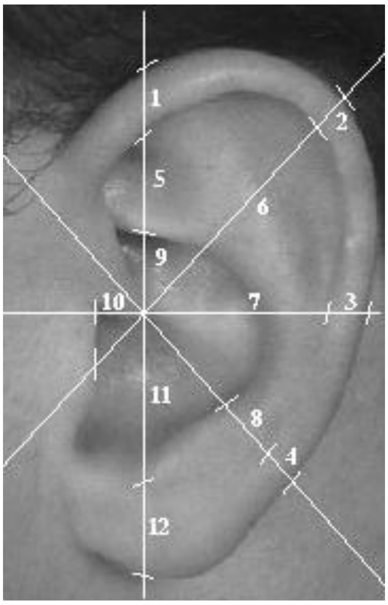


Figure 2.1: The locations of the measurements used in the “Iannarelli System”

2.1 Ear Detection

Ear detection is the most important step in an ear recognition system, and the quality of detection will directly affect the performance of the whole recognition system. Burge and Burger [18] used a relatively simple approach based on applying the deformable contour technique on a Gaussian pyramid representation of the image gradient. A grey scale image was obtained from the subject's head in profile. Then the location of the ear in the image was localised by deformable contours. The deformable contour algorithm is based on the minimization of the energy of a generalized active contour model (g-snake).

Abdel-Mottaleband Zhou [19] detected the ear automatically from the profile image of the face. The detection of the ear was achieved in three steps: skin-tone region detection, edge detection and template matching. Detecting the skin-tone region used a skin filter which relied on colour and texture information. The output skin region was extended to encompass the whole ear; this was achieved by applying a morphological dilation operation to the output of the skin filter. Canny edge detection was applied only to the region where skin was detected. The result was filtered to remove short and isolated edges that appeared due to noise. Finally the ear region was detected using template matching. The Hausdorff distance was used to search for an ear template in the edge map of the skin region. This technique was a part of an ear recognition system which recognized the ear after its detection. This system will be explained in the next section.

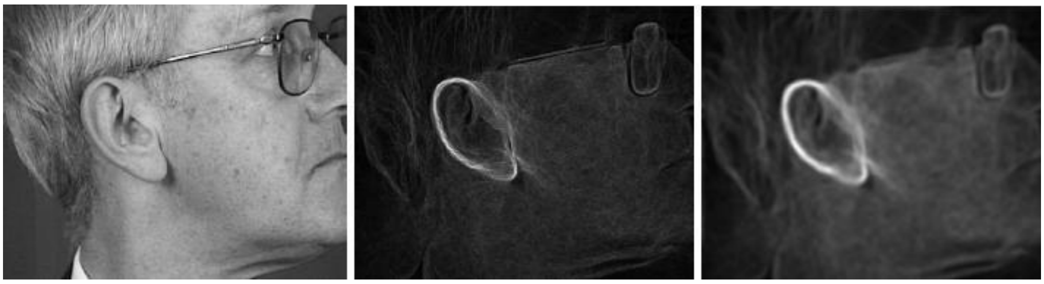
Yuan and Mu [20] combined skin colour model and intensity contour information to localise and track the human ear in a sequence of video frames. A two-stage ear tracking method was proposed. In the first stage, The Continuously Adaptive Mean Shift (CAMSHIFT) approach was used to track the face profile in the video frame sequences. In the second stage, the contour-based fitting algorithm was used for accurate ear detection. The aim for profile tracking was to fast and locate the main skin colour region as the region of interest (ROI) which contained the ear. The mean shift algorithm was operated on probability distribution, to track coloured objects in a video sequence. The colour image data was initially represented as colour distribution. The probability distribution was calculated by histogram back-projection that associated the pixel values in the image with the value of the corresponding histogram bin of the hue (colour). The histogram bin values were scaled to be in the discrete pixel range of the 2D probability distribution image. Then, the ROI was extracted by implementing the CAMSHIFT algorithm. The ROI was divided equally into four rectangular parts. In most cases, the ear appeared in the left-top rectangle part of the ROI, so a new ROI was defined by that part. Canny edge detection and contour fitting were applied on this new ROI. Since the shape of the ear is similar to an ellipse, ellipse fitting was used to locate the ear by finding a set of parameters that minimized the distance between the data points and the ellipse. Experimental results showed that the ear tracking was applicable in a practical scenario. The experiment used a common USB camera and a subject walked perpendicular to the direction of the camera.

Chen and Bhanu [21] introduced a simple and effective method to extract ears from side face range images which had two phases: offline model template building and on-line detection. Ears were extracted manually from a set of training side face range images. Then, a histogram of shape index image was computed for each detected ear. The shape index captured the geometric information of the ear wherein calculation depended on principal curvatures. After calculating the histograms for all the extracted ears, the histograms were averaged. The averaged histogram was used as the model template. Step edge detection was calculated as the maximum distance in depth between the centre point and its neighbours in a small window. The step edge image was thresholded to obtain a binary image. There were some holes in the thresholded binary image so the images were dilated to determine the potential regions containing ears. Then smaller regions, whose area was less than a chosen threshold, were removed. For every region, the grown rectangular box with the minimum dissimilarity was chosen as the candidate ear region. After that, the dissimilarity was calculated between the template model and each region. Finally over all of the candidate regions, the one with the minimum dissimilarity was selected as the detected region. There were 30 subjects in their database and every subject had two side face range images. They reported a 91.5% correction rate.

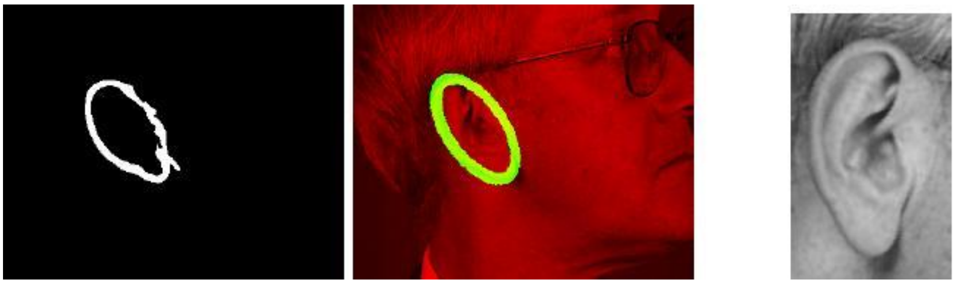
Ansari and Gupta [22] presented an ear detection approach relied on the edges of outer ear helices. They used the parallelism between the ear outer helix curves to detect the ear. Ear edges were obtained by using Canny edge detector. These edges were thinned and junctions between the edges were removed. After removing junctions, each edge was treated separately as a chain code. The shape of helix curve was convex so edges were segmented into those which were convex and those which were concave. For segmentation, the edges were divided into groups where the angles of tangents were either monotonically decreasing or increasing. A set of curves was considered to be possible outer ear curves, from those curves which were parallel and the average distance between two curves and the standard deviation of their distance were less than a certain threshold value. To select the curve which belonged to outer ear helix, one necessary condition was that angle produced by the lines joining the points in the path and the one end of the path was either monotonically decreasing or increasing. Among the remaining curves, the longest was selected as the ear helix curve. The helix curve of ear was used to determine the ear boundary by joining two end points of the ear helix curve with straight line. The outer region to the boundary was marked as background. The experiment was performed on IITK database which had images of 700 individuals. The accuracy of the proposed localisation was 93.3%.

Prakash et al. [23, 24] presented an efficient distance transform and template based technique for automatic ear localization from a side face image. The technique first segmented skin and non-skin regions in the image, based on a colour model. This model transformed a colour image into a grey scale image such that the grey value at each pixel showed the likelihood of the pixel belonging to the skin. Then the grey scale image was transformed to a binary image with an appropriate threshold. After skin segmentation, the edge map of the skin regions was obtained using the Canny edge detector. Then edge length and curvature based criteria were employed to eliminate spurious edges from the edge map. Cross-correlation based searching was applied between distance transforms of the edge maps of the face and an ear template. All the points having normalized cross-correlation coefficient value greater than a threshold were accepted as probable locations of the ear. The ear template was created by averaging the intensities of a set of ear images. Resizing of the ear template was done in accordance with the size of each input image and the resized template was used for ear localization and ear verification. Finally, a similarity between two sets of Zernike moments (one for the template and the other for the detected ear) was estimated to determine whether the detected ear was actually an ear or not. The proposed technique was tested on IIT Kanpur ear database containing side face images of 150 individuals and achieved 95.2% accuracy rate.

Cummings et al. [25] presented a novel ear enrolment technique using the image ray transform, based upon an analogy to light rays. The transform is able to highlight tubular structures such as the helix of the ear and, by exploiting the elliptical shape of the helix, can be employed as the basis of a method for enrolment for ear biometrics. The initial step was to apply the image ray transform to the original ear image (Figure 2.2-a). This produced an image in which the helix of the ear was highlighted (Figure 2.2-b) in addition to other tubular features such as spectacles. Then a Gaussian smoothing was applied to the result (Figure 2.2-c) to reduce noise. The image was then thresholded (Figure 2.2-d) to produce an image with a strong helix as its focus. Histogram equalisation was used to allow a consistent threshold across all images to be used. Simple template matching was then used with an elliptical template across a range of rotations and scales (Figure 2.2-e) and then the matched section was then normalised and extracted (Figure 2.2-f). They achieved 99.6% success at enrolment across 252 images of the XM2VTS database.



(a) Original (b) Ray Transform (c) Smoothed



(d) Threshold (e) Detection (f) Extraction

Figure 2.2: Example of the steps taken to achieve successful ear enrolment.

Islam et al. [26] modified the cascaded AdaBoost approach to detect the ear from 2D profile face images. The approach was divided into three steps: preprocessing, training and detection. In the preprocessing step, input images were collected from different sources with varying size and intensity values. Therefore, all the input images were scale normalized to a chosen input pattern size. The intensity values of images were also normalized to minimize the effect of lighting. In the training step, weak classifiers were built up based on some rectangular features (Figure 2.3). These features were derived from Haar wavelets which are a natural set of basis functions computed the difference of intensity in neighbouring regions. Eight types of features were proposed for building the weak classifiers. All the input images and the sub-windows were represented as integral images to efficiently calculate the feature value. Then a strong classifier was created by taking a weighted combination of a set of selected weak classifiers using the AdaBoost (Adaptive Boost) algorithm. The algorithm used supervised learning to determine the weight of each weak classifier. Once the training was completed, the trained classifiers of all the stages were used in a cascaded manner to build an ear detector. A classifier in the cascade was only used when a sub-window in the test image was detected as positive (ear) by all previous classifiers. The sub-window was accepted as an ear, only when it passed through all of the classifiers. Ear detection based on the cascaded AdaBoost algorithm was fast, accurate and robust, while the training phase was time consuming caused by the large number of samples and the features.

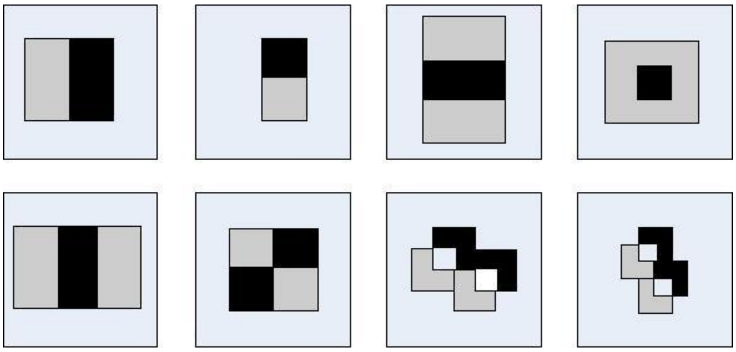
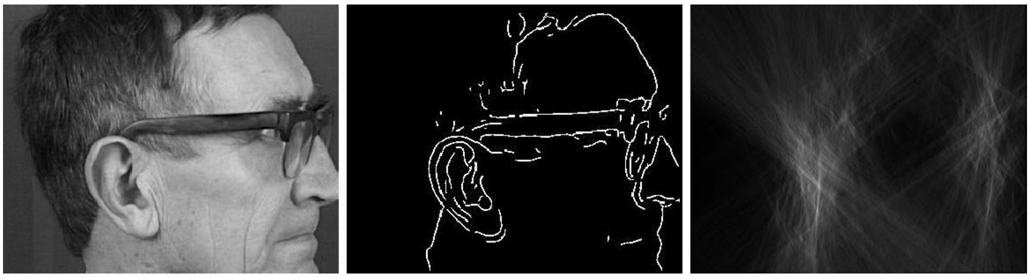


Figure 2.3: The features used in training the AdaBoost

The positive training set was built with 5000 ear images collected from the profile face images of various databases covering a wide range of races, sexes, appearances, orientations and illuminations. This set contained images from UND, MID, XM2VTS, USTB, MIT-CBL, and UMIST databases. It also contained around 3000 images synthesized by rotating -15 to +15 degrees of some images from USTB, UND and XM2VTS databases. The negative training set included 10,000 images of non-ear images. These images were mostly collected from profile face images, excluding the ear area. Training with a large variety of ear images from various ear databases made the detector robust to slight rotation, occlusion and also to the degradation of image quality to a significant extent. The detection performance of the cascaded detector was tested against 203 profile images from the UND database; all the ears were accurately extracted by the proposed detector with a very low false positive rate. In the presence of partial occlusions involving hair and earrings, Out of 104 selected occluded images from the XM2VTS database, 54 images were correctly detected.

Arbab-Zavar and Nixon [27] proposed an ear detection technique based on locating the elliptical shape of the ear using a Hough Transform (HT) accruing tolerance to occlusion and noise. Head detection was the first step in the algorithm (Figure 2.4-a). Diffuse lighting or skin detection in colour was employed to eliminate the head from background relying on which database they used. Then, Canny edge detection was applied to detect the edges (Figure 2.4-b). The HT generated an accumulator with a number of peaks corresponding to the position of ellipses' centres (Figure 2.4-c). To detect the peak corresponding to the ear, a number of refinements were applied to remove some of the votes which did not adhere the properties of a prospective ear rim ellipse. The peaks corresponding to non-vertical ellipses were eliminated which the ear usually appeared as a vertical ellipse with a vertical axis approximately twice as large as the horizontal axis. After that, the edge direction information was employed to determine whether or not a pair of points can possibly be part of an ellipse. Selecting the pairs more carefully reduced the noise level in the accumulator and the peak corresponding to the ear became clearer. Then, one third of the front of the faces was cropped to eliminate the effect of spectacles which also appeared as elliptical shapes.



(a) (b) (c)

Figure 2.4: The accumulator for image (a) is in (c).

After these refinements, an averaging template was applied to the accumulator array to remove the single peaks. Finally, the centre of the widest peak was found which represented the ear. The proposed enrolment algorithm was tested on two databases. The first one was the XM2VTS database which consisted of 252 images from 63 individuals. The second database was the UND ear biometric database which consisted of 942 2D images from 302 subjects. The enrolment process successfully detected the ear region in all the 252 profile images of the XM2VTS database. 83% of the ears were enrolled successfully in the case of 40% occlusion from the top which was studied as it closely follows occlusion by hair. For the UND database, the algorithm offered a 91% enrolment success rate.

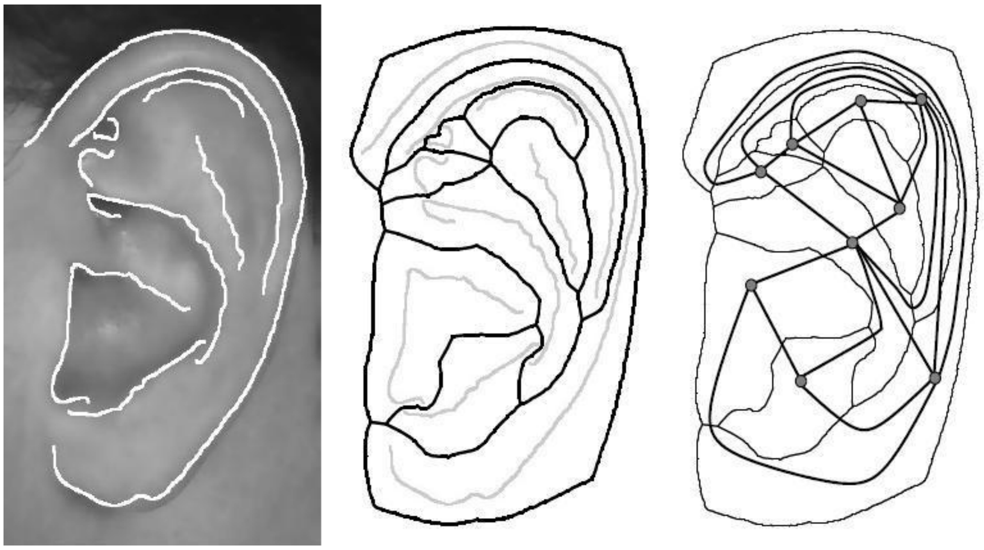
These are two most sophisticated approaches yet for enrolment [26, 27]. Islam's technique depended on statistical approach. It reported good results on a database of large size, and the results depended on different databases in the training stage. But it did not work well in presence of occlusion, also if the ear was rotated or an ear was different in appearance from the ears in the training set, the technique could not detect the ear because it was not included in the training set. The advantages of Arbab-Zavar's technique were that it used the shape of the ear and reported a good result with occlusion but there were some limitations in which it depended on some conditions and certain database.Table 2.1 summarizes all of the above ear detection techniques.

Technique	Database	Detection Rate
Deformable contour Burge and Burger [18]	N/A	N/A
Template matching Abdel-Mottaleb and Zhou [19]	103 subjects	N/A
Skin colour and intensity contour Yuan and Mu [20]	USB camera	N/A
Histogram of shape index Chen and Bhanu [21]	30 subjects	91.5%
Edges of outer ear helices Ansari and Gupta [22]	IITK, 700 individuals	93.3%
Distance transform and template based Prakash et al. [23, 24]	IIT Kanpur, 150 individuals	95.2%
Image ray transform Cummings et al. [25]	XM2VTS, 252 profile images	99.6%
Haar based and Cascaded AdaBoost Islam et al. [26]	UND, 203 profile images XM2VTS, 104 occluded images	100% 52%
Arbab-Zavar and Nixon [27]	XM2VTS, 252 profile images UND, 942 profile images	100% 91%

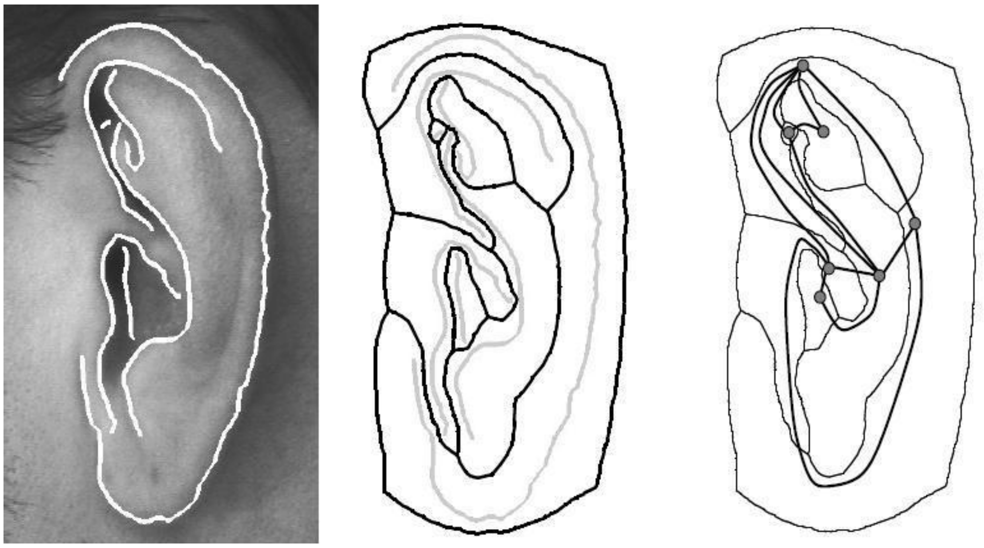
Table 2.1:Detection rates for different ear biometrics systems

2.2 2D Ear Recognition

Burge and Burger [18] were the first to consider the human ear as a biometric in the field of machine vision. They introduced a graph matching based algorithm for authentication in which they took into account the erroneous curve segments which can occur because of changes in the ear image. After ear localization, edges were



(a) Ear print (b) Voronoi diagram (c) Neighbour graph



(d) Ear print (e) Voronoi diagram (f) Neighbour graph

Figure 2.5: Stages in building the ear biometric graph model

computed using the Canny operator. Edge relaxation was used to form larger curve segments, which the remaining small curve segments (length less than threshold) were eliminated. Differences in lighting and orientation caused erroneous curves. These false curves were eliminated by first segmenting the inner cavity and then eliminating small, high curvature, and closed curves occurring within it. A Voronoi neighbourhood graph of the curves was constructed to achieve invariance under affine transformations (Figure 2.5). They identified the problem of occlusion by hair as a main obstacle and proposed the use of a thermogram image to overcome this obstacle. A thermogram image is one where the surface heat of the subject is employed to form an image. Though they introduced the ear as a potential biometric, the study did not include an evaluation of recognition potential.

Moreno et al. [28] investigated the use of outer ear images for human identification. In this system ears were manually extracted from the profile images. For recognition, they combined the results obtained by three neural classifiers. The first classifier was based on a biometric vector which contained the location of considered feature points from the outer ear image. This vector was normalized to be invariant with respect to translation and scale. The second classifier was based on a morphology vector which captured the shape and wrinkles structure that were specific for each individual ear. The third classifier used neural networks in two stages.

A first network, called a compression network, was trained auto associatively on the original ear image to extract the statistically salient properties of the image data. These properties formed the compression vector. This vector constituted the input to a single perceptron that performed the identification task. The outputs of any of the three classifiers corresponded to each one of the individuals to be identified. The classification results were then combined by methods such as voting, Bayesian combination, linear weighting, or neural network techniques. For experiments, a set of 168 images were acquired in different sessions from 28 individuals. The experimental results have demonstrated that the best identification results, without considering rejection thresholds, have been obtained using a compression network (93%). The combination techniques have increased the identification rates by taking into account rejection thresholds (84% achieved by the Weighting Bayesian method versus 79% of compression network).

Hurley et al. [2] used force field transformation to reduce the dimensionality of the original pattern space. The algorithm was divided into two steps: image to force field transformation, and force field feature extraction. The entire image was transformed into a force field by pretending that each pixel exerted a force on all the other pixels, which was proportional to the pixel's intensity and inversely proportional to the square of the distance to each of the other pixels. Force field feature extraction was performed by computing convergence maps (Figure 2.6). The convergence function mapped the force field to a scalar field, considering the force as input and returning the additive inverse of the divergence of the force direction. A Fourier based cross-correlation technique was employed to perform template matching on thresholded convergence maps. The force field technique resulted in a correct classification rate of 99.2% on a subset of the XM2VTS face profile database. The subset consisted of 252 images from 63 subjects with relatively unoccluded ears. This method proved to yield a much better performance than Principal Components Analysis (PCA) when the images were poorly registered. Force field was derived directly from the entire image without any preprocessing and the ear did not need to be explicitly extracted from the background. The approach was also robust to noise.

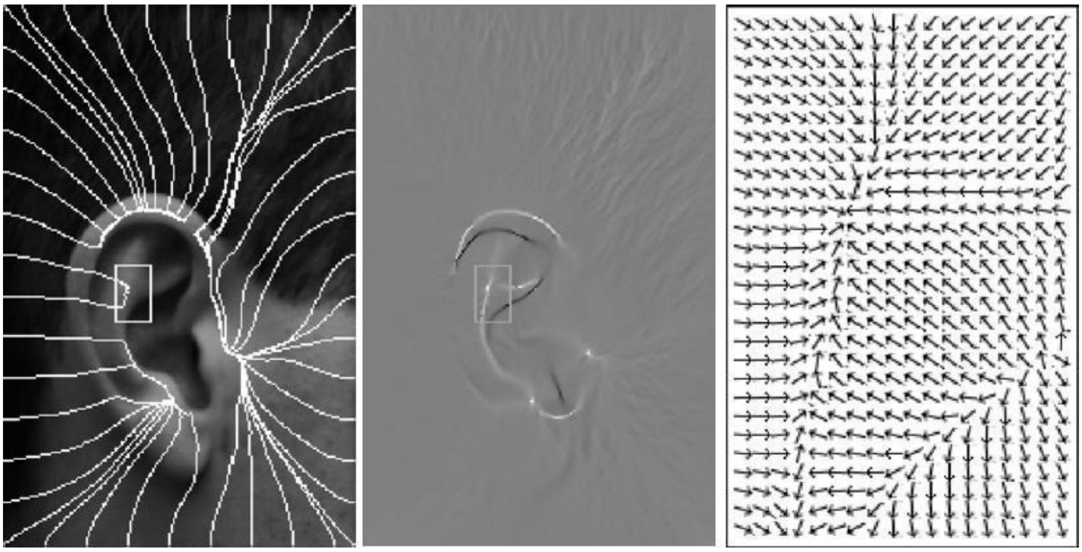

(a) Field line features (b) Convergence map (c) Magnified insert force field

Figure 2.6: Convergence field

Abdel-Mottaleband Zhou [19] presented a technique for ear identification from profile images of the face. The ear was automatically extracted from the profile image of the face. A force field transform was applied to the extracted ear region.

From the transformed image, principal curvatures k1 and k2 were calculated to detect the points lying in ridge-valley line as the feature points. These points correspond to the extreme ridge points on the considered surface (an extreme ridge point is a point where k1 has a local positive maximum). During recognition stage, the feature points were aligned and compared with those in the database using Hausdorff distance. For the experiments, a database of profile face images was collected for 103 subjects. For each person, one image was used for training, where the ear region was extracted and the detected ear feature points were stored in the database. The proposed ear recognition method was applied to 58 images for 29 subjects, two images from each subject. Out of the 58 images, 51 obtained the correct match.

The most popular methods used to extract the feature vectors of ear images in ear biometrics research is Principal Components Analysis (PCA) which is widely used in facial recognition. PCA is a dimensionality reduction technique in which variation in the dataset is preserved. The classification is achieved in a lower dimension space called eigenspace.

Victor et al. [29] applied PCA to make a comparison between face and ear recognition and concluded that the face recognition performed better than the ear recognition. The process consisted of three steps: preprocessing, normalization, and identification. In the preprocessing step the images were cropped to a specific size. Coordinates of two distinct points were supplied to the normalization routine. The two points picked in a face image were the centre points of the two eyes, and the two points in an ear image were Triangular Fossa and the Antitragus. The normalization step included geometric normalization, masking and photometric normalization. The identification step consisted of two phases: training and testing. There were three experiments performed to test this comparison: probes of same day with different expressions (other ear), probes of different days with similar expression (same ear), and probes of different days with different expressions (other ear).

However, Chang et al. [30] made another comparison between face and ear recognition by using PCA and found that there is no considerable difference between the ear and the face in terms of recognition. The process was same as that in the research of Victor et al. [29]. Subject's images were dropped from the study if either the face or the ear was largely obscured by hair, if the subject wore an earring or analogous face jewellery or if either image had technical problems. There were three experiments: different days experiment, lighting condition variation experiment, and pose variation experiment (with ± 22.5° rotation). The final result was that the recognition rate for ears was 71.6 % and for face 70.5 %. The difference is not statistically significant. The different conclusion obtained by the two comparisons may be due to the quality of their data [30]. The database in the study by Victor et al. had less control over the variations such as earrings, hair over ears and lighting.

Zhang et al. [31] proposed a hybrid system for classifying ear images. In their system they combined Independent Component Analysis (ICA) and a Radial Basis Function (RBF) network. The method was divided into three stages: preprocessing, feature extraction, and identification. In the preprocessing stage, the grey-scale ear images were transformed into 64×40 pixel resolution. Training and probe data were filtered by using a Wiener filter which achieved the best results. Finally histogram equalization was applied to the filtered image. The ICA algorithm was used in the feature extraction stage. The ear image was decomposed into linear combinations of several independent basic images. The corresponding coefficients of these combinations were the inputs to a RBF network. In the identification stage, the RBF network was used for classification. RBF was a special three-layered feed-forward neural network and the activation function was the Gaussian Kernel Function. The experiments were performed on two ear databases. The first database was Carreira-Perpinan ear database which consisted of 102 images from 17 individuals. The second database was their own database which consisted of 180 images from 60 individuals. They showed that better performance can be achieved by using ICA instead of PCA, where the recognition rate by using ICA was 88.3% but for PCA, it was 85%. However, both PCA and ICA offer almost no invariance and therefore need very accurate registration in order to achieve consistently good results [10].

A few studies have focused on geometrical approaches [32, 33]. Mu et al. [32] presented an ear identification system based on geometrical feature extraction. The system was achieved in three steps: edge detection of the ear image, feature extraction and identification. The ear edge detection was obtained by applying median filtering, grey level stretch, Sobel operator, binarization, and noise removal. After ear edge detection step, the shape feature vector of the outer ear and the structural feature vector of the inner ear were extracted to form the feature vectors which were invariant to ear image’s translation, scale and rotation. Finally identification was performed by applying template matching between the database templates and the feature vector of the tested ear image. They reported an 85% correct recognition rate using their database which consisted 308 images from 77 subjects.

Choras [33] proposed invariant geometrical method in order to extract features needed to classification. The method was divided into four steps: contour detection, coordinates normalization, feature extraction and classification. Contour detection consisted of edge detection and binarization techniques. Coordinate normalization was performed to the detected contour, such that the centroid became the centre of the image. In the feature extraction step, different circles centred in the centroid were created. Feature vector was divided into two vectors. The first vector contained radius length, the number of intersection points for each radius, and the sum of all the distances between the intersection points for the considered radius. The second vector contained the number of endings, bifurcations and intersection points, and all the coordinates of those points for all the extracted and traced contours. The classification step relied on, the simple comparison between the input image feature vectors, and all the vectors from the database. For such easy images from their database, they obtained error-free recognition.

Arbab-Zavar et al. [34] proposed the first model based approach, capitalizing on clear structure with the advantages of being robust in occlusion and noise. An automatic ear detection process based on Hough transform for ellipses was applied. The Scale Invariant Feature Transform (SIFT) was employed to detect the features (keypoints) within the extracted ear images (Figure 2.7). The keypoints determined by SIFT have assigned locations, scales and orientations. The model was learned off-line using an unsupervised learning algorithm over a detected training set of 63 ear images. In this algorithm, the SIFT keypoints of each ear image were repeatedly presented to the algorithm. In each iteration, the matching keypoints between the ear and the model were determined, and the model was updated by changing the matched keypoints using a cumulative average and also by adding the unmatched keypoints of the ear image to the model. A derivative based measure was employed to quantify the model alteration in each iteration. After learning, the model was a group of generalized ear parts. A simple nearest neighbour approach was used for recognition which obtained 91.5% correct recognition. A subset of the XM2VTS

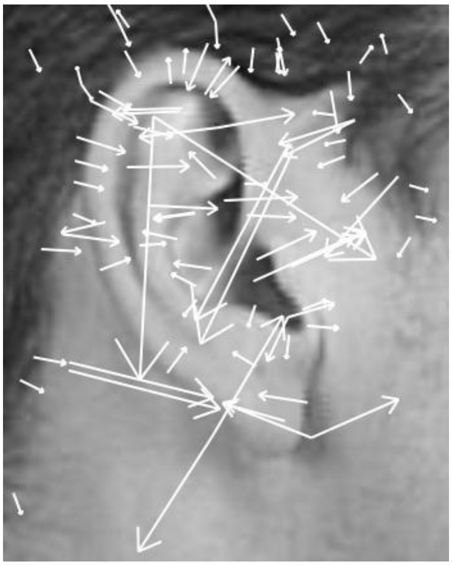


Figure 2.7: The SIFT keypoints for an ear image

database was used for testing the algorithm which consisted of 252 images from 63 individuals.

Arbab-Zavar and Nixon [35, 36] improved their model based method by applying a one dimensional log-Gabor filter. This filter was applied to obtain the frequency content of the fluctuating surface which provided the structures of the Helix and the Anti-helix in an ear. The boundaries and stretched curves were not well captured by the SIFT descriptors, so a Log-Gabor filter was employed (Figure 2.8). Before applying the Log-Gabor filter, aligned templates of image data were prepared to facilitate the localisation of helix-related features. The template was the sampled image in a semi-circular region which contained the Helix. The model parts were employed to vote for the position of the centre of this semicircle. After the centre was determined, the template was formed by sampling the image intensities along radial lines which were mostly normal to the Helix curve. Finally, a one dimensional Log-Gabor filter on the columns of the templates was applied. A robust matching algorithm was implemented in recognition stage. Log-Gabor filter reported an 88.4% recognition rate. Log-Gabor filter and the model were combined, using a simple decision fusion technique of sum of the normalized scores. The hybrid classifier showed a significant improvement with a 97.4% recognition rate.

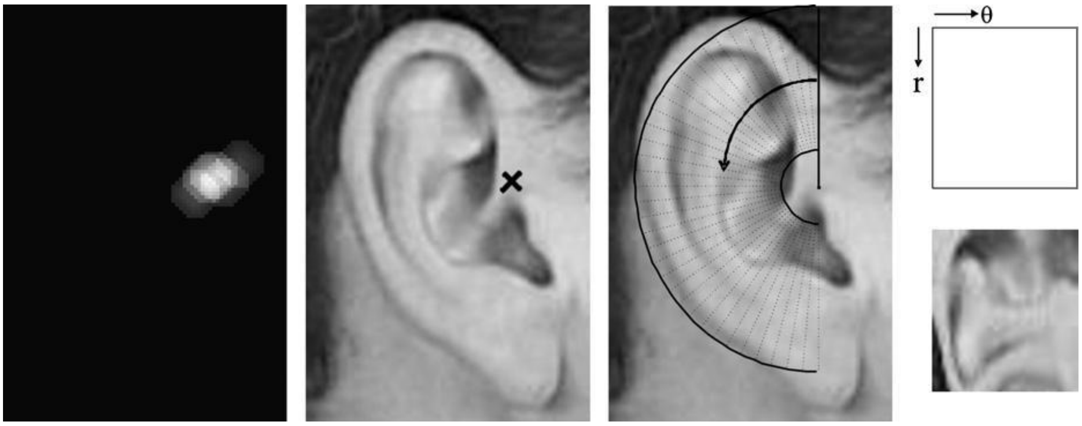
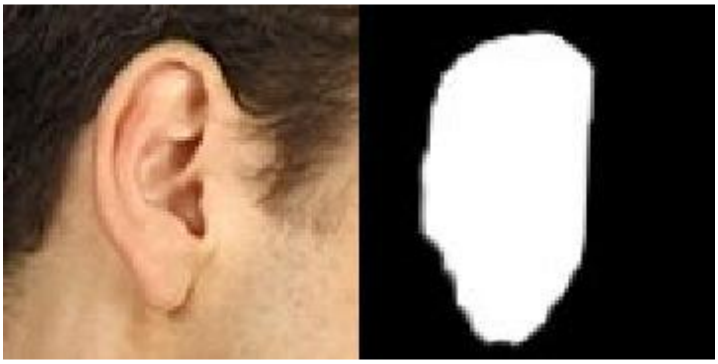


Figure 2.8: From the left: the accumulator gathering votes for the putative centre; the detected centre; sampling the image along radial lines; the input template for the wavelet analysis.

Bustard and Nixon [37, 38] proposed a technique which improved the robustness of ear detection and recognition, addressing issues of pose variation, background clutter and occlusion. The technique was achieved in two main stages: creation of masks for gallery ear images which were created off-line, and registration and recognition for a probe which were executed on-line. To create the masks, SIFT was used for the detection of feature points. After that a seed ear (Figure 2.9-a) was selected and a mask manually created for that ear (Figure 2.9-b). Then the rest of the gallery of ears were matched against the seed to find four matched points which were used to calculate the homography matrix. The calculated homography matrix was used to align the mask of the seed ear image to compute the mask of the matched ear image. This process was repeated until there were no more matches. If there were any gallery images remaining, a new mask was constructed manually and that image was added to the seed. This was repeated until all gallery images had masks. The first step, in registration and recognition, was to determine feature points in the probe. For each of these points the gallery was searched to find correspondences. If four points were matched between the probe and the gallery, they were used to compute a homography matrix that registered the probe. Once the two images were aligned, the distance between the images was computed. Finally, the nearest gallery image identified the person. They reported a 96% recognition rate on the same dataset used by Hurley et al. [2]and Arbab-Zavar et al. [27].



(a) Seed Ear (b) Mask

Figure 2.9: A gallery ear image and its associated mask

2.3 3D Ear Recognition

Although acquiring 3D ear data needs more elaborate equipment compared with that for 2D data, some researchers have focused on 3D recognition. The process of ear extraction and detection in a side view image becomes easier when the 3D data of the image is available in which the auricle has a rich and deep 3D structure. The Iterative Closest Point (ICP) algorithm has been widely used in 3D ear images alignment due to its simplicity and accuracy.

Chen and Bhanu [39] introduced a two-stage ICP technique for matching 3D ears. In the first stage, the ear helix in 3D images was detected. The helix of the ear was computed as the maximum distance in depth between the centre point and its neighbours in a small window. Edge thinning and linking, and removal of small edge segments were implemented to detect the helix correctly. Then the ICP algorithm was run to determine the initial rigid transformation to align a model ear helix with the test ear helix. In the second stage, the initial transformation was applied to selected locations of model ears. The ICP algorithm iteratively refined the transformation by minimizing the distance between the control points of the model ear and their closest points of the test ear to bring model ears and test ear into best alignment. Finally, the root mean square (RMS) registration error was calculated as the matching error criterion and the model ear with the minimum RMS error was declared as the recognized ear. There were 30 subjects in their database and every subject had two side face range images. They achieved a 93.3% recognition rate.

Yan et al. [40] presented approaches to speed up the ICP algorithm and improve the performance of ICP registration to recognise 3D ear images. Two approaches were considered to make the ICP algorithm faster. One was to control the number of iterations, and the other was to use appropriate data structures to shrink the running time for each iteration by using *k-d* tree instead of Octree. The time for traversing a tree is the most expensive stage of the ICP algorithm. The smaller the number of cells, the faster the traversal time. The number of cells in the *k-d* tree were smaller than the number of cells in Octree. Therefore, the ICP algorithm with k-d tree was much faster than with Octree. The approaches were considered to improve the performance included noise removal, outlier elimination and using point-to-triangle error metric. They achieved a 98.8% recognition rate on their database which consisted 302 subjects, one image for each subject.

Cadavid and Abdel-Mottaleb [41] proposed an approach for 3D ear recognition from video sequences. A series of frames were extracted from a video clip and the region of interest in each frame was independently reconstructed in 3D using Shape from Shading. The resulting 3D models were then registered using the ICP algorithm. Iteratively each model in the series was considered as a reference and the similarity between the reference model and every model in the series was calculated using a similarity cost function. The model that demonstrated the greatest overall similarity was determined to be the most stable 3D model and was subsequently enrolled in the database. A probe model was globally aligned to a 3D model in the gallery using ICP. Then, the root mean square distance (RMSD) between the two models was computed. The smallest RMSD value was declared the identity of the probe ear model. Experiments were conducted using a gallery set of 462 video clips, collected by WVU, where in each clip the camera moved in a circular motion around the subject’s face. They reported a 95% recognition rate.

3D can achieve good results, but needs specialized equipment but it cannot be applied to images derived from documents. Table 2.2 summarizes all of the above ear recognition systems.

2.4 Multimodal Biometrics

Given a natural propensity for biometric fusion, to handle scenarios where one biometric is obscured but when another is available, it is no surprise that there are a number of multimodal biometrics approaches to ear recognition exists. Rahman et al. [42] presented multimodal biometric systems using PCA to extract the features on

Technique	Database	Recognition Rate
A graph matching based Burge and Burger [18]	N/A	N/A
Three neural classifiers Moreno et al. [28]	28 subjects (168 images)	93%
Force field transformation Hurley et al. [2]	XM2VTS, 63 subjects (252 images)	99.2%
Force field transformation Abdel-Mottaleb and Zhou [19]	29 subjects (58 images)	87.9%
Principal Components Analysis Victor et al. [29]	72 subjects (808 images)	40%
Principal Components Analysis Chang et al. [30]	197 subjects	71.6%
Independent Component Analysis Zhang et al. [31]	17 subjects (102 images) 60 subjects (180 images)	88.3%
Geometrical feature extraction Mu et al. [32]	77 subjects (308 images)	85%
Geometrical feature extraction Choras [33]	Easy images from their database	100%
Model based Arbab-Zavar et al. [34]	XM2VTS, 63 subjects (252 images)	91.5%
Log-Gabor filter and model based Arbab-Zavar and Nixon [35]	XM2VTS, 63 subjects (252 images)	97.4%
Scale Invariant Feature Transform (SIFT) Bustard and Nixon [37]	XM2VTS, 63 subjects (252 images)	96%
Iterative Closest Point (ICP) Chen and Bhanu [39]	30 subjects (60 images)	93.3%
ICP Yan et al. [40]	302 subjects	98.8%
Shape from Shading and ICP Cadavid and Abdel-Mottaleb [41]	462 video clips	95%

Table 2.2:Recognition rates for different ear biometrics systems

both face and ear.Iwano et al. [43] combined speech and ear images using a composite posterior probability. Yan et al. [44] proposed a combination of 2D and 3D ear data. They used PCA for 2D ear images. PCA, an edge-based Hausdorff distance method, and ICP registration were used for 3D ear images. Katiyar and Pathak [45] proposed a simple multimodal biometric identification system including ear, face and gait using Kernel Principal Component Analysis (KPCA) method. Nixon et al. [46] described how ear and gain biometrics could be combined for use in forensic identification. They used point based approaches to ear and gait recognition. Kisku et al. [47] combined ear and face biometrics using support vector machines. All of these studies achieved an increase in performance when employing multimodal biometrics instead of individual biometrics.

Chapter 3

Ear Detection

Using local structure appears an interesting approach to ear detection especially when combined with frequency domain analysis. One of the main motivations of the methods based on local structure is that the ear image can be represented in a common structure for all ears. A local structure-based method depends on finding points or local areas in an ear by using generalised Gabor wavelets, known as banana wavelets. We aim to choose a selection of local features which are generic to ear enrolment and recognition.

In this chapter, a novel method is proposed based on using banana wavelet analysis to determine local structure points. In the following sections, the details about the banana wavelet transform will be presented. Then the proposed algorithm will be explained.

3.1 Banana Wavelets

Kruger et al. [48, 49, 50] introduced an object recognition system in which objects were represented as a spatially organized set of local line segments. The line segments corresponded to binarized Gabor wavelets or banana wavelets. In their technique, representations of object classes were learned autonomously. In order to avoid the necessity of manual intervention for the generation of ground truth, the system was equipped with a mechanism which produced controlled training data. This was achieved by moving an object with a robot arm and following the object by fixating the robot hand. The robot produced training data on which a certain view of an object was shown with varying background and illumination but with corresponding landmarks having the same pixel position in the image.

Significant features per instance were extracted by positioning a rectangular grid on a roughly segmented object in front of homogeneous background. After extracting the significant features per instance in different pictures, invariant local features were extracted for a class of objects. The learning algorithm was applied to this data to extract an object representation. A similarity function between a graph labelled with the learned banana wavelets and a certain position in the image was applied for matching process.

Many approaches have mentioned the existence and relationship of their own work to banana wavelets [51, 52, 53] but do not appear directly to have exploited this new technique.

Gabor filters are now being used extensively in various computer vision applications. Banana wavelets are a generalization of Gabor Wavelets and are localized filters derived from a mother wavelet [48], particularly suited to curvilinear structures (Figure 3.1). Each filter of this family of banana wavelets models the receptive field structure of an object.

A banana wavelet $B^{\mathbf{b}}$ is parameterized by a vector $\mathbf{b}$ of four variables, i.e. $\mathbf{b} = (f, \theta, c, s)$ where f, θ, c and s are the frequency, orientation, curvature, and size, respectively. This filter consists of two components i) a rotated and curved complex wave function $F^{\mathbf{b}}(x, y)$ and ii) a Gaussian $G^{\mathbf{b}}(x, y)$ function rotated and curved in the same way as $F^{\mathbf{b}}(x, y)$ [48]:

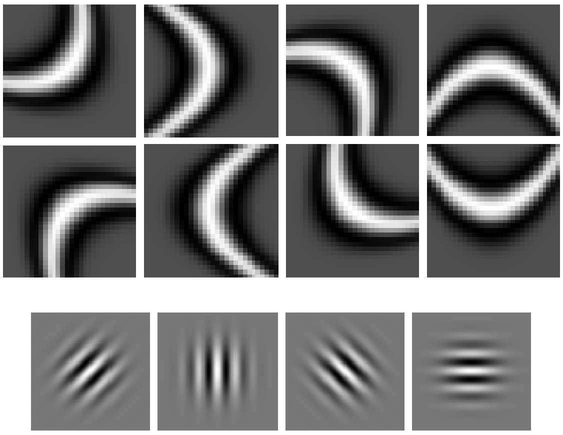


Figure 3.1: Banana Wavelets (top) and Gabor Wavelets (bottom)

$$B^{\mathbf{b}}(x,y) = \gamma \cdot G^{\mathbf{b}}(x,y) \cdot (F^{\mathbf{b}}(x,y) - DC^{\mathbf{b}}) \tag{3.1}$$

where

$$G^{\mathbf{b}}(x,y) = \exp\left(-\frac{f^2}{2} \cdot \left(\frac{\left(x_c + c \cdot x_s^2\right)^2}{\sigma_x^2} + \frac{x_s^2}{\sigma_y^2 \cdot s^2} \right) \right) \tag{3.2}$$

$$F^{\mathbf{b}}(x,y) = \exp\left(i \cdot f \cdot \left(x_c + c \cdot x_s^2\right)\right) \tag{3.3}$$

$$x_c = x \cdot \cos\theta + y \cdot \sin\theta \tag{3.4}$$

$$x_s = -x \cdot \sin\theta + y \cdot \cos\theta \tag{3.5}$$

$$DC^{\mathbf{b}} = \frac{\int G^{\mathbf{b}}(x,y) \cdot F^{\mathbf{b}}(x,y)}{\int G^{\mathbf{b}}(x,y)} = e^{-\frac{\sigma_x}{2}} \tag{3.6}$$

γ is a constant, chosen empirically, σ_x and σ_y are the scales of the Gaussian filter in the x and y directions respectively, and $DC^{\mathbf{b}}$ is the bias of the banana wavelets.

Features can be extracted from any image by the banana wavelet transform to describe both spatial frequency and curvilinear structure present in the image. The convolution of an image with complex banana filters with various frequencies, orientations, curvatures, and sizes, captures the local structure of the image.

3.2 Ear Detection

We argue that any ear contains curvilinear structures such as the helix, anti-helix and inter-tragic notch. Our ear detection technique is shown in Figure 3.3. The essence of our ear detection technique is to initially calculate the magnitude of the filter responses $AI(\mathbf{x},\mathbf{b})$ by convolving a banana wavelet $B^{\mathbf{b}}$ with an image I and then to find the positions where this magnitude has local maxima, i.e.:

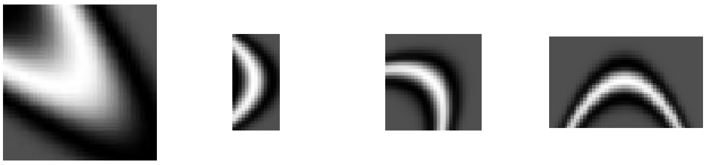
$$\mathbf{x}_0 = \arg\left(\max_{\mathbf{x}}\left(\left|AI(\mathbf{x},\mathbf{b})\right|\right)\right) \tag{3.7}$$

where

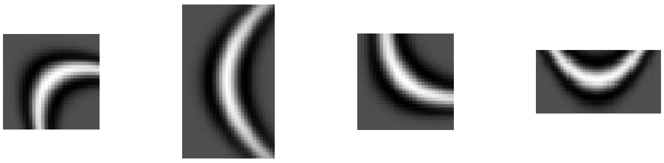
$$AI(\mathbf{x},\mathbf{b}) = (B^{\mathbf{b}} * I)(\mathbf{x}) \tag{3.8}$$

and $\mathbf{x}$ is the position of a pixel in an input image I .

A banana wavelet $B^{\mathbf{b}}$ produces a strong response at pixel position $\mathbf{x}_0$ when the local structure of the image at that pixel position is similar to $B^{\mathbf{b}}$. An input ear image and the response magnitudes, which are calculated by convolving the input image with the filters depicted in Figure 3.2, are shown in Figure 3.4. In this figure, white pixels represent large values in the response magnitudes. Therefore there are local maxima



(a) Filter1 (b) Filter2 (c) Filter3 (d) Filter4



(e) Filter5 (f) Filter6 (g) Filter7 (h) Filter8

Figure 3.2: (a)-(h) 8 filters used in this work

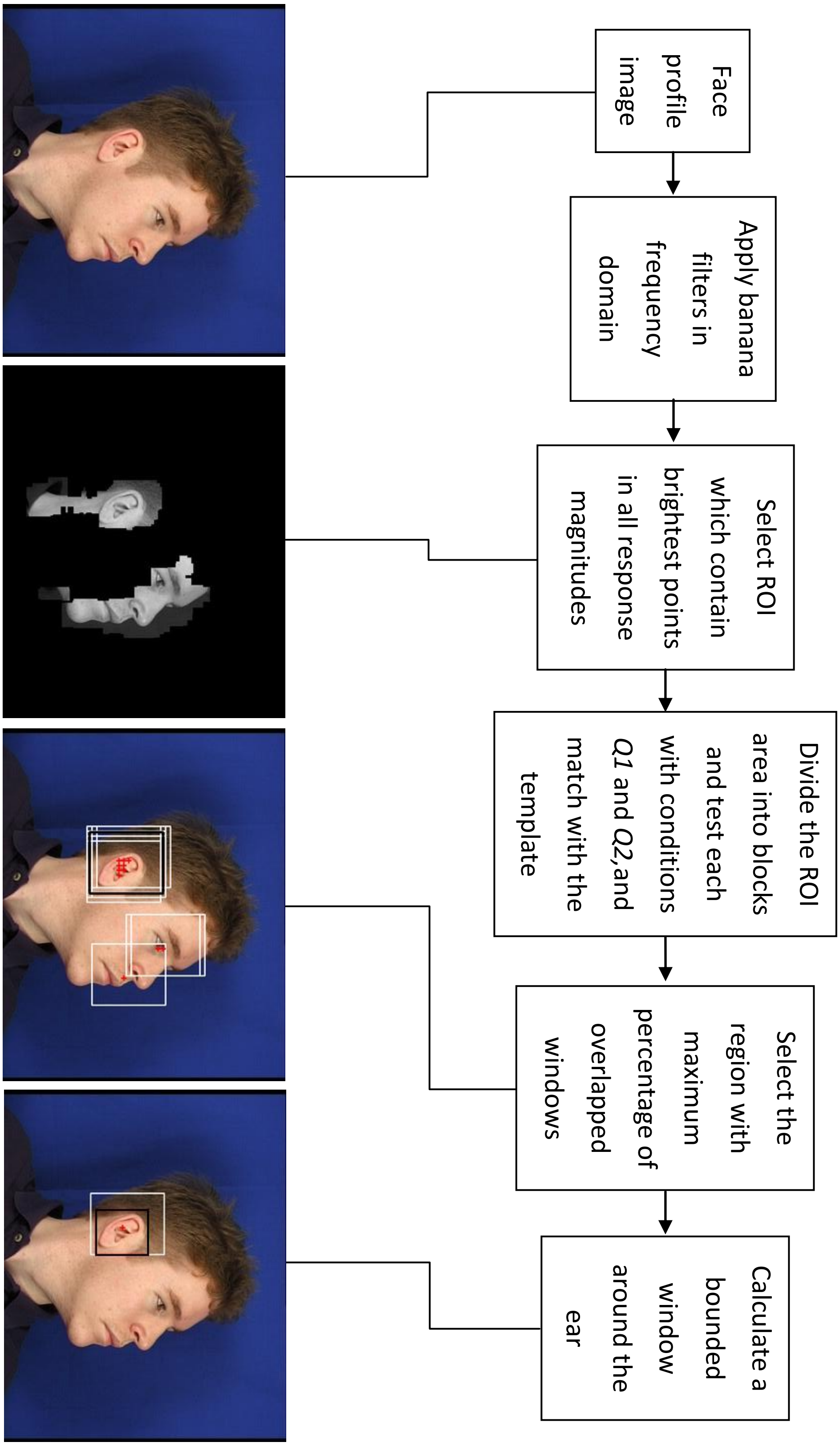
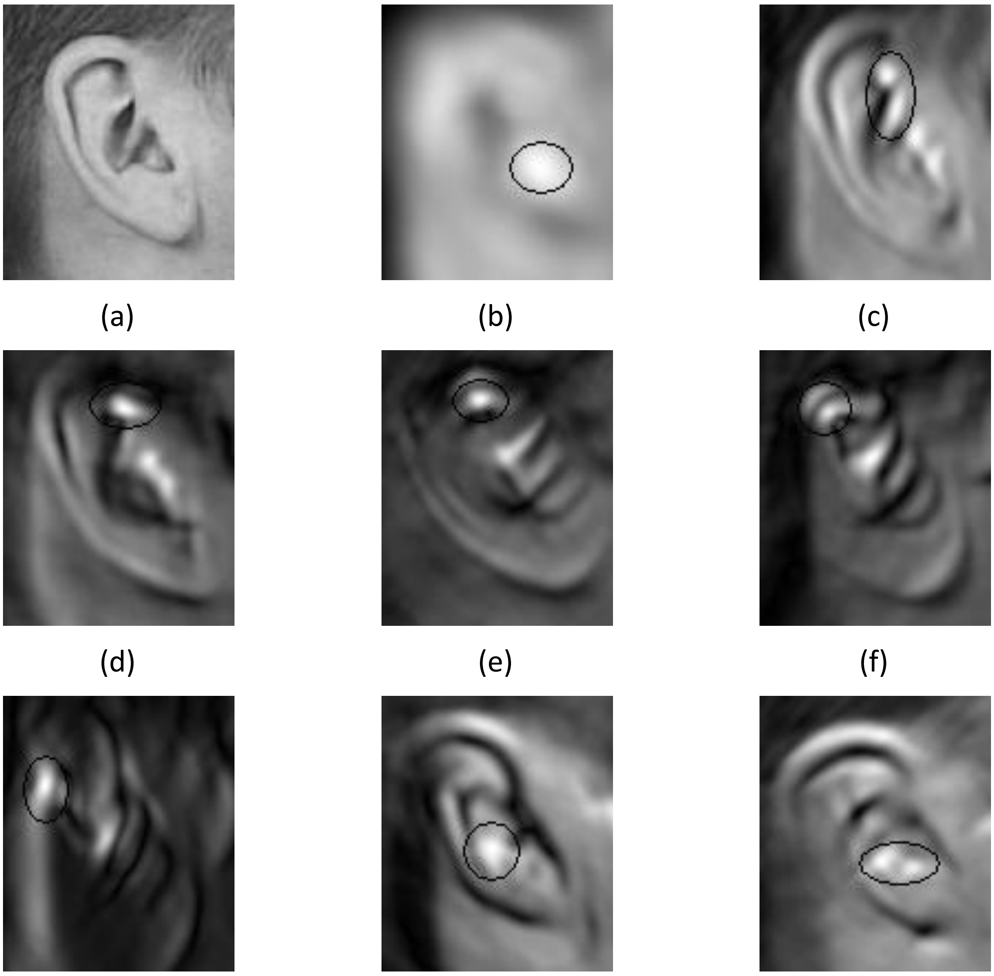


Figure 3.3: Our ear detection technique



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 3.4: (a) Input image, and (b)-(i) after convolution with 8 banana filters.

at those positions where the ear has similar frequency, curvature, orientation, and size to those of the corresponding banana wavelet.

To reduce the computational burden, the ear detection process starts with a coarse search by applying 8 banana filters to the whole image to extract the regions of interest containing curved lines in order to perform a finer search for the ear within these regions:

$$BI_i = B^{\mathbf{b}_i} * I \qquad 1 \le i \le 8 \qquad (3.9)$$

where i is the filter number. The brightness points represent strong values in the locations of BI_i where the local structures of image I have similar properties to those of the corresponding banana filters. Each BI_i is converted to a binary image M_BI_i according to the brightness points. The region (a 121×81 pixel window) surrounding the top 10 local maxima in each BI_i is set to 1. The regions of interest are chosen as regions which contain brightness points in five or more banana filtered images, i.e.:

$$ROI = \left\{ (x, y) \middle| (x, y) \in \Omega \ \wedge \sum_{i=1}^{8} M\_BI_i (x, y) \geq 5 \right\} \tag{3.10}$$

where Ω is the domain of image I. These regions of interest are much smaller than the whole image and therefore the fine search performed by applying the banana filters to these regions, requires less computational demand.

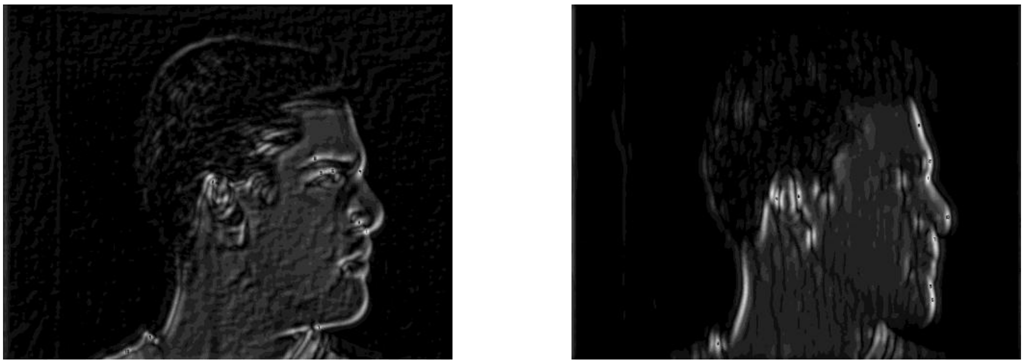
Figure 3.5 illustrates the convolution of an input image with the filters shown in Figure 3.2. In Figure 3.5, the brightest points represent strong values in the response magnitudes which have similar properties to those of the corresponding banana wavelet.

As illustrated in Figure 3.6, the regions of interest are chosen as regions which contain brightest points in all response magnitudes of the convolved image with banana filters. These regions are much smaller than the whole image and therefore the fine search performed by applying the banana wavelets to these regions, requiring less computational.

In the fine search, the regions of interest are divided into a group of smaller neighbourhoods and convolved with a bank of banana filters (8 filters are chosen in this thesis as shown in Figure 3.2) to calculate positions corresponding to local maxima in each neighbourhood. A position of interest is selected by considering two

Filter 1 Filter 2

Filter 3 Filter 4



Filter 5 Filter 6

Filter 7 Filter 8

Figure 3.5: Convolution of an input image with 8 banana filters

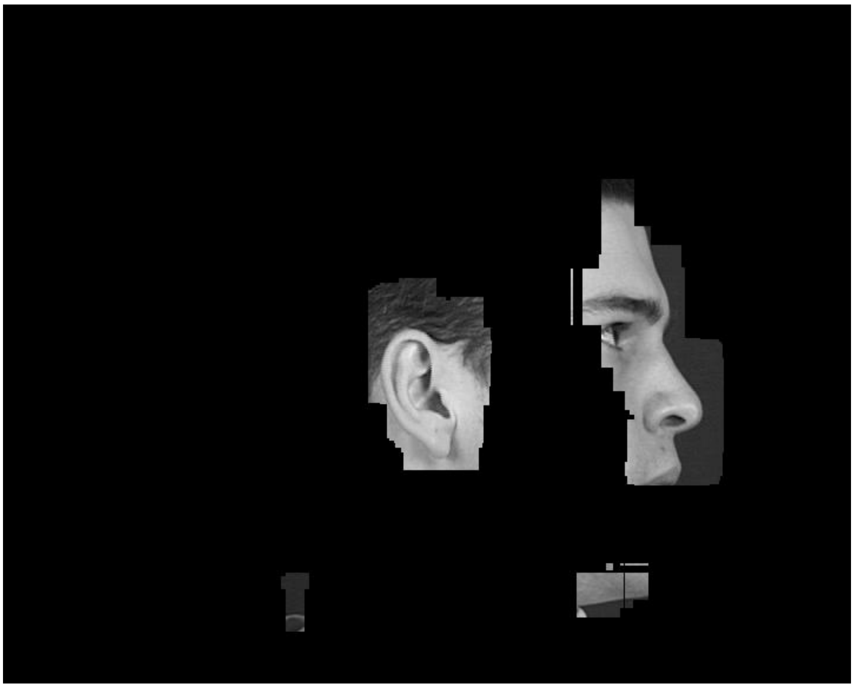


Figure 3.6: Regions of interest

conditions: i) the response magnitude has to represent a local maximum($Q1$) and ii) its value should be greater than a certain threshold ($Q2$):

$$Q1: AI(\mathbf{x}_0,\mathbf{b}) \geq AI(\mathbf{x}_i,\mathbf{b}) \qquad \mathbf{x}_i \in N(\mathbf{x}_0) \tag{3.11}$$

$$Q2: AI(\mathbf{x}_0,\mathbf{b}) > \lambda \cdot T(\mathbf{x}_0) \tag{3.12}$$

where

$$T(\mathbf{x}_0) = 0.25 \cdot (E(I) + E(I,\mathbf{x}_0)) \tag{3.13}$$

$$E(I) = \sum_{\mathbf{x}\in I} \sum_{\mathbf{b}\in B} AI(\mathbf{x},\mathbf{b}) \tag{3.14}$$

$$E(I,\mathbf{x}_0) = \sum_{\mathbf{x}\in N(\mathbf{x}_0)} \sum_{\mathbf{b}\in B} AI(\mathbf{x},\mathbf{b}) \tag{3.15}$$

λ is a constant and $N(\mathbf{x}_0)$ represents a square window with centre $\mathbf{x}_0$ and width w.

In addition to the conditions $Q1$ and $Q2$, the spatial arrangement of the positions of local maxima should match a template representing the ear structure. The position of any local maxima is determined by block number (block number 1 to 30) by partitioning the neighbourhood into 30 blocks (as shown in Figure 3.7). The spatial arrangement of the positions of local maxima (as illustrated in Figure 3.8) should be as following:

- Filter 1 should be on the right of filter 8 and below filter 2, and its position is in block number 19, 20, 24 or 25.
- Filter 2 should be above filter 1 and below filter 3, and its position is in block number 8, 9, 13 or 14.
- Filter 3 should be on the right of filter 4 and above filter 2, and its position is in block number 3, 4, 8 or 9.
- Filter 4 should be on the right of filter 5 and on the left of filter 3, and its position is in block number 2, 3, 7 or 8.
- Filter 5 should be on the left of filter 4 and above filter 6, and its position is in block number 1, 2, 6 or 7.
- Filter 6 should be above filter 7 and below filter 5, and its position is in block number 6, 7, 11 or 12.
- Filter 7 should be on the left of filter 8 and below filter 6, and its position is in block number 18, 19, 23 or 24.
- Filter 8 should be on the left of filter 1 and on the right of filter 7, and its position is in block number 18, 19, 23 or 24.

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25
26	27	28	29	30

Figure 3.7: An ear (neighbourhood) and its partition into blocks

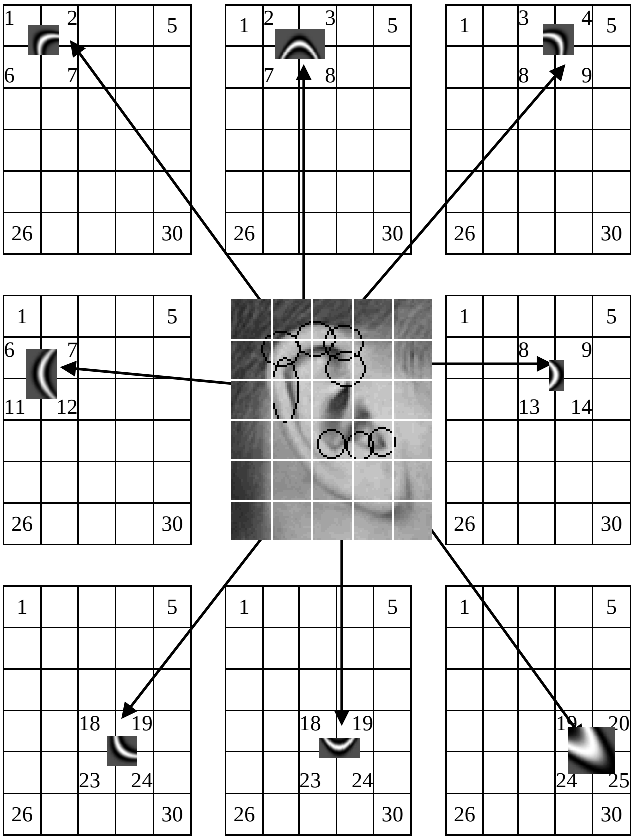


Figure 3.8: Ear template and the spatial arrangement of the positions of the filters' responses

Locations of the local structure points in this template are general for any ears. For example, the convolution of Filter 4 with any ear, produces a strong response at the top middle part (block number 2, 3, 7 or 8) as illustrated in Figure 3.4-e and Figure 3.8. Also when convolving Filter 3 with any ear, it produces a strong response at the top right part (block number 3, 4, 8 or 9) for any ear (Figure 3.4-d and Figure 3.8). In addition to that, when convolving the other filters, they produce strong responses nearly in the same positions for all ears. So the template is not particular for certain ear or database but it is general for any database, which the positions of the produced responses are general for any ear. Examples of various ears (with variations in rotation and scale) are shown in Figure 3.9, and the positions of the local maxima of each filter for these ears are recorded in Table 3.1. As illustrated in Figure 3.9 and Table 3.1, the spatial arrangement of the positions of the filters' responses matches the ear template and are determined in the correct places.

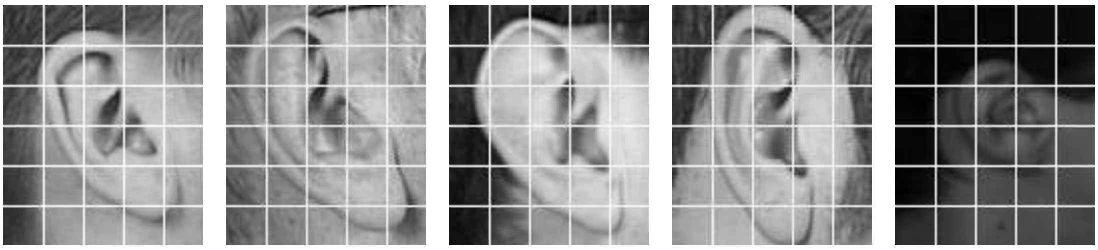


Figure 3.9: Examples of partitioned ears into blocks

	Ear 1	Ear 2	Ear 3	Ear 4	Ear 5
Filter 1	19	19	24	24	19
Filter 2	8	8	8	9	14
Filter 3	8	3	3	4	9
Filter 4	8	2	2	3	8
Filter 5	7	1	7	2	7
Filter 6	12	6	11	12	12
Filter 7	18	18	18	18	18
Filter 8	19	19	18	18	19

Table 3.1: Examples of (block) positions of filters' responses for the ears in Figure 3.9

The neighbourhood with maximum number of positions matching the ear template and meeting conditions $Q1$ and $Q2$ is considered as the neighbourhood containing the ear. In the case that many overlapping neighbourhood windows are detected (Figure 3.10), only one region is selected which contains maximum percentage of overlapped windows.

The window size, which contains the ear, is initially fixed for all ears. So, a bounded window around the ear is calculated. Boundary refinement can be achieved by convolving the top part of the fixed window with a bank of banana filters. Then, for each filter response magnitude, a position of local maxima is located. After that, from those local maxima positions, an accumulator space is calculated to choose the maximum accumulator position as a position of top point at the ear helix. The same process is applied to calculate left point at the helix ear and bottom point at the lobe, but with a different bank of banana filters. Finally, those points (the top and left points at the ear helix, and the bottom point at the lobe) are used to representthe boundaries of the bounded window around the ear (Figure 3.11). The calculated window is variable in the size corresponding to each ear.


(a) (b)

Figure 3.10: Many overlapping neighbourhood windows


(a) (b)

Figure 3.11: Bounded window around an ear

The parameters of the 8 filters are shown in Table 3.2 (R and C in the table are the number of rows and columns of the banana wavelet filters, respectively, which are used in the convolution process between banana wavelet filters and the image).

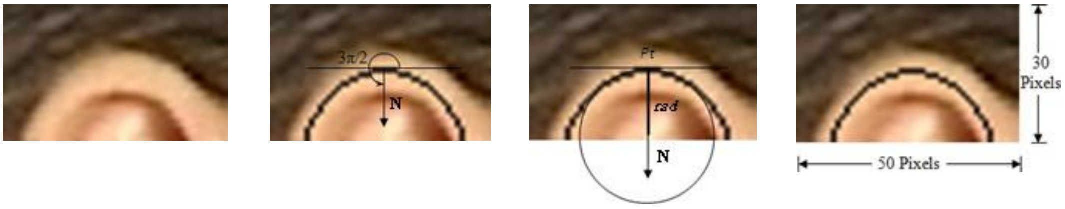
	f	θ	c	s	R	C
Filter 1	0.05	$\pi/4$	0.1	1	50	50
Filter 2	0.28	$\pi/2$	0.05	1	30	15
Filter 3	0.28	$3\pi/4$	0.05	1	30	30
Filter 4	0.28	π	0.05	1	30	50
Filter 5	0.28	$5\pi/4$	0.05	1	30	30
Filter 6	0.28	$3\pi/2$	0.02	1	50	30
Filter 7	0.28	$7\pi/4$	0.03	1	30	30
Filter 8	0.28	2π	0.05	1	20	40

Table 3.2: Parameter Settings for the Banana Wavelets.

These parameters are chosen experimentally by averaging the computed parameters from 63 images (one image from each of 63 subjects selected from XM2VTS database).Firstly, the orientation of the curve is estimated by computing the angle of the normal to the ear curve and subtracting $\pi/2$ from this angle. As illustrated in Figure 3.12-b, the orientation of the curve equals π, so θ in Filter 4 is set to equal π.

Secondly, the curvature of the ear curve is computed by finding the best fitting circle at a point on the curve. If Pt is a point on the curve, then the best fitting circle passing through the point Pt will have the same curvature as that of the helix (the upper part). The curvature of a circle is a constant $1/rad$, where rad is the radius of the circle. The centre of the circle will be on the line containing the normal vector **N** to the circle. As illustrated in Figure 3.12-c, rad is small which means that the curvature of the curve is large, so c in Filter 4 is chosen to be large, here c is chosen to equal 0.05.

Thirdly, the average number of rows and columns of the top part of the ear are approximately 30 and 50 (Figure 3.12-d), respectively, so these values are used to be the values of R and C parameters of Filter 4.



(a) (b) (c) (d)

Figure 3.12: (a)-(d) The computation for parameters of filter 4

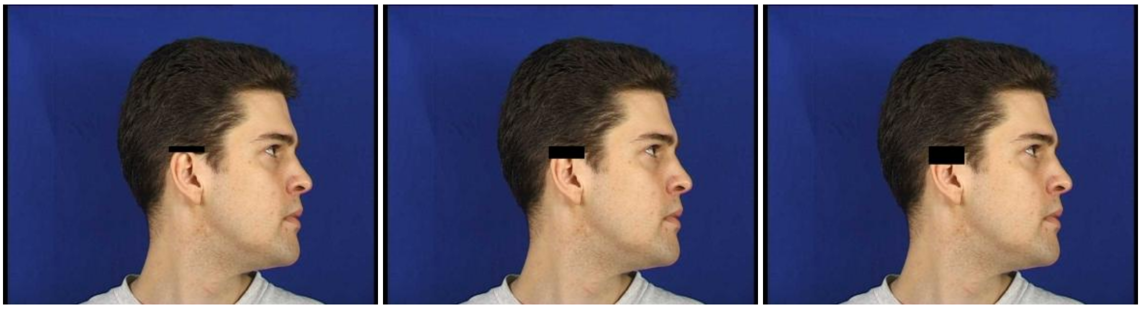
Finally, parameters f and s of the filter are set to 5 and 1, respectively, producing the best results. The same procedure is performed to compute the parameters of the other filters.

3.3 Effects on Ear Detection

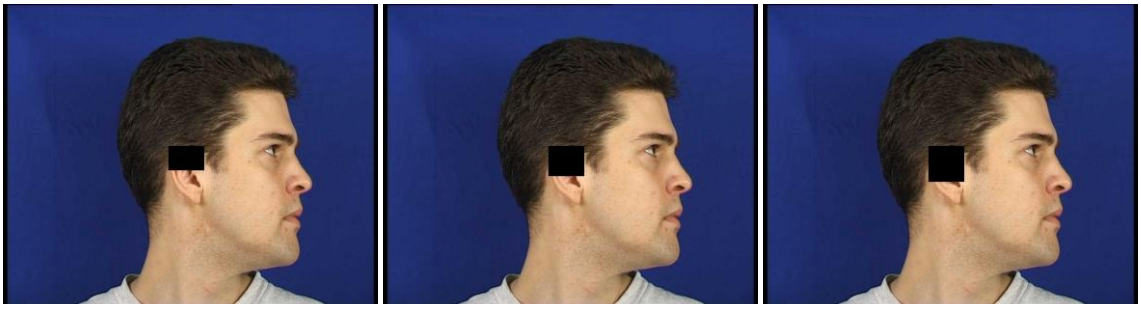
There exist many possible side effects to the ear enrolment process, for example: occlusion, noise, rotation and lighting. Occlusion and noise are applied to our database images to show the strength of our technique in presence of occlusion or noise. For the other effects, our databases have many variations in rotation and lighting, so these are not simulated.

3.3.1 Occlusion

The main obstacle preventing the deployment of ear biometrics is the occlusion. Hair and earrings are the most probable objects to occlude an ear image. It is not easy to measure the degree of occlusion in ear images, since the ear images are occluded in different ways. So, occlusion is synthetically applied to the ear region of the face-profile images with different level of occlusion from top (see Figure 3.13) and left (see Figure 3.14). Occlusion from the top is chosen as this is the most likely scenario for occlusion by hair. The ear image is occluded by 10% to 60% from the ear size (when the ear image is occluded synthetically more than 60%, only a small part from the ear appears and this part does not contain information about the ear).

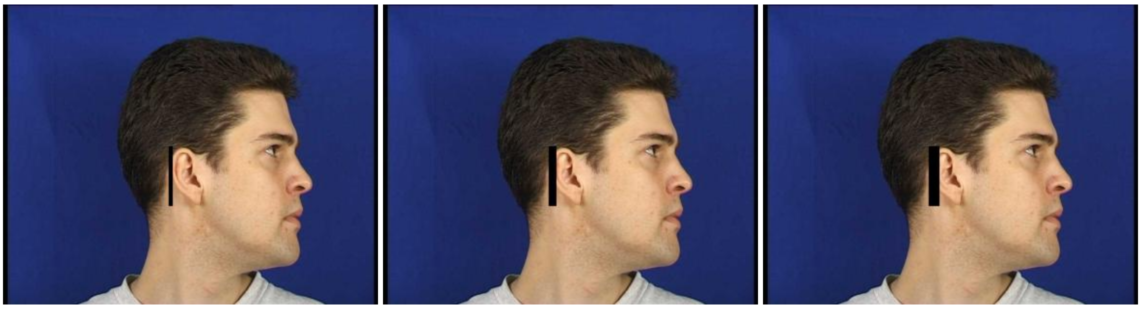


(a) 10% (b) 20% (c) 30%

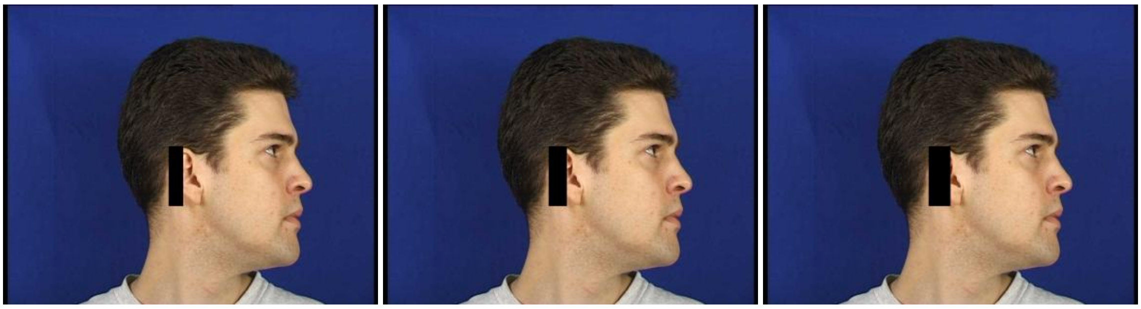


(d) 40% (e) 50% (f) 60%

Figure 3.13: An occluded ear for top (10% to 60%)



(a) 10% (b) 20% (c) 30%



(d) 40% (e) 50% (f) 60%

Figure 3.14: An occluded ear for left (10% to 60%)

3.3.2 Noise

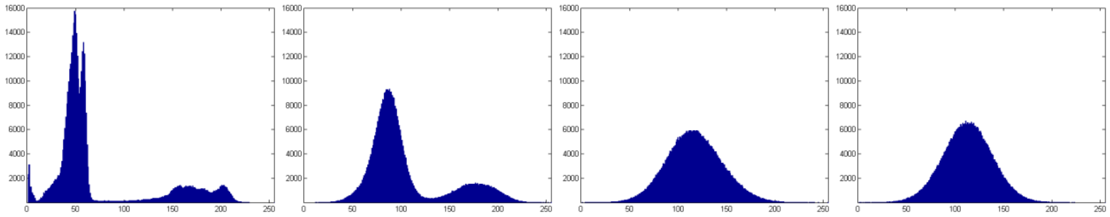
By the Central Limit Theorem, a common assumption is that images are contaminated with an additive zero mean Gaussian noise $\psi(i,j)$ [54]. In Gaussian additive noise models, the noise density follows a Gaussian normal distribution $G(0,\eta^2)$ defined by a standard deviation η. It can be expressed by adding the normal distribution G to image I :

$$IN(i,j) = I(i,j) + \eta\,\psi(i,j) \qquad \forall i \in [1..Row]\ and\ j \in [1..Col] \qquad (3.16)$$

where Row and Col are the width and height of image I , respectively. Figure 3.15 shows an image with the Gaussian additive noise applied, showing a range of η values and their histograms. After adding the Gaussian additive noise, the image is normalised from 0 to 255.

Gaussian noise is added to our database images to test our algorithm performance in the presence of noise with various standard deviations.

(a) η = 0 (b) η = 20 (c) η = 100 (d) η = 500



(e) histogram of image in (a) (f) histogram of image in (b) (g) histogram of image in (c) (h) histogram of image in (d)

Figure 3.15: (a)-(d) The effect of Gaussian noise various η values, and (e)-(h) their histograms

3.4 Preprocessing Algorithm

The purpose of preprocessing is to minimize the variation in the acquired image, while keeping the characteristic features of the subject. We propose that superior preprocessing methods are necessary to improve the performance of the ear detection.

This section briefly discusses skin detection and adaptive filter techniques required for the preprocessing step before applying our proposed ear detection algorithm. The preprocessing algorithm is shown in Figure 3.16. The skin detection technique is applied to the face profile image to produce a skin mask (skin regions are white and non-skin regions are black). Also the adaptive filter is applied to the input image to produce enhanced edges. Then, the skin mask multiplied by the enhanced image to produce skin image with enhanced edges. Finally our ear detection technique is applied to the processed image.

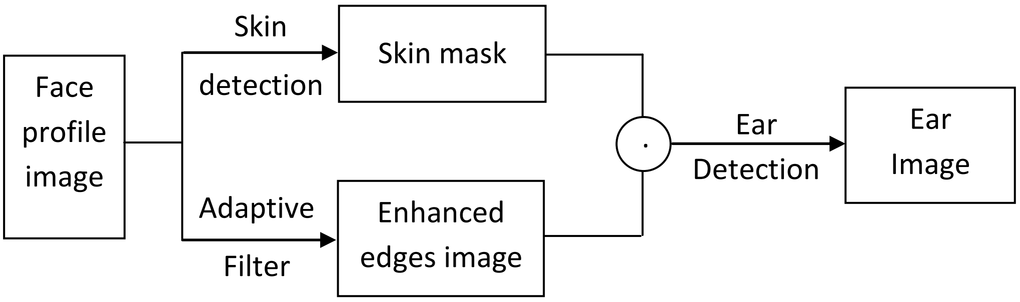


Figure 3.16: Ear detection technique with applying preprocessing algorithm

3.4.1 Skin Detection

The first step of the preprocessing algorithm is the segmentation and detection of skin regions in the image. Normally, ears are in the skin region, therefore the skin detection technique is applied to isolate the face and ear region from the hair and clothes as much as possible. Our skin detection technique is based on skin texture and skin colour analysis (as shown in Figure 3.17). The main problems of using colour only to detect the skin region are the following: i) the appearance of skin colour can be affected by lighting and ii) the skin colour can change from person to person (even with people of same ethnicity).

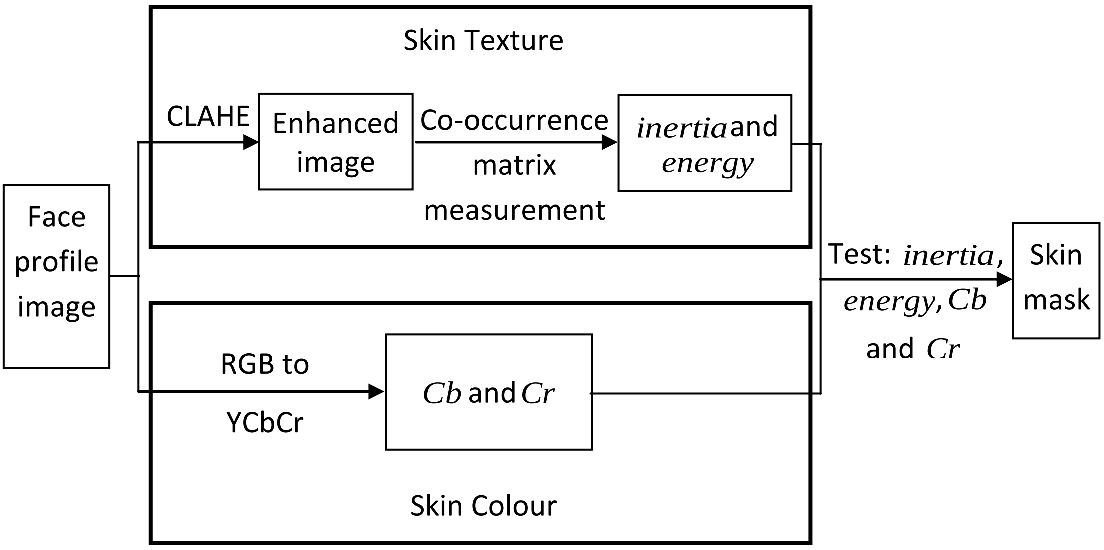


Figure 3.17: Skin detection technique

3.4.1.1 Skin Texture

The co-occurrence matrix [54] is used for skin texture detection. The co-occurrence matrix contains elements that are counts of the number of pixel pairs for specific brightness levels, at a particular distance and along a specified direction. For brightness levels $b1$ and $b2$ the co-occurrence matrix Co is [54]:

$$Co(b1,b2)=\sum_{x=1}^{row}\sum_{y=1}^{col}\left(I(x,y)=b1\right)\wedge\left(I(x',y')=b2\right) \quad (3.17)$$

where $\wedge$ denotes the logical AND operation and the x coordinate x' is the offset given by the specified distance d and direction ϕ by:

$$x'=x+d\cdot\cos(\phi) \quad \forall \quad (d\in[1..\max(d)])\wedge(\phi\in[0..2\pi]) \quad (3.18)$$

and the y coordinate y' is:

$$y'=y+d\cdot\sin(\phi) \quad \forall \quad (d\in[1..\max(d)])\wedge(\phi\in[0..2\pi]) \quad (3.19)$$

At first, an input image I is converted to greyscale and then a Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm [55, 56] is applied to enhance the contrast of the greyscale image. Next, Eq. (3.17) is applied to the enhanced image to obtain a square, symmetric, matrix whose dimensions equal the number of grey levels N in the image. The co-occurrence matrices for skin and non-skin regions are shown in Figure 3.18. In the co-occurrence matrix generation, the maximum distance d was 5 pixels and the directions ϕ were set to select eight neighbours of each point. The co-occurrence matrix can then be described by the $inertia$ [54]:

$$inertia = \sum_{b1=1}^{N} \sum_{b2=1}^{N} (b1 - b2)^2 \cdot Co(b1, b2) \qquad (3.20)$$

which emphasises components which have a large separation. An alternative measure is the $energy$ [54]:

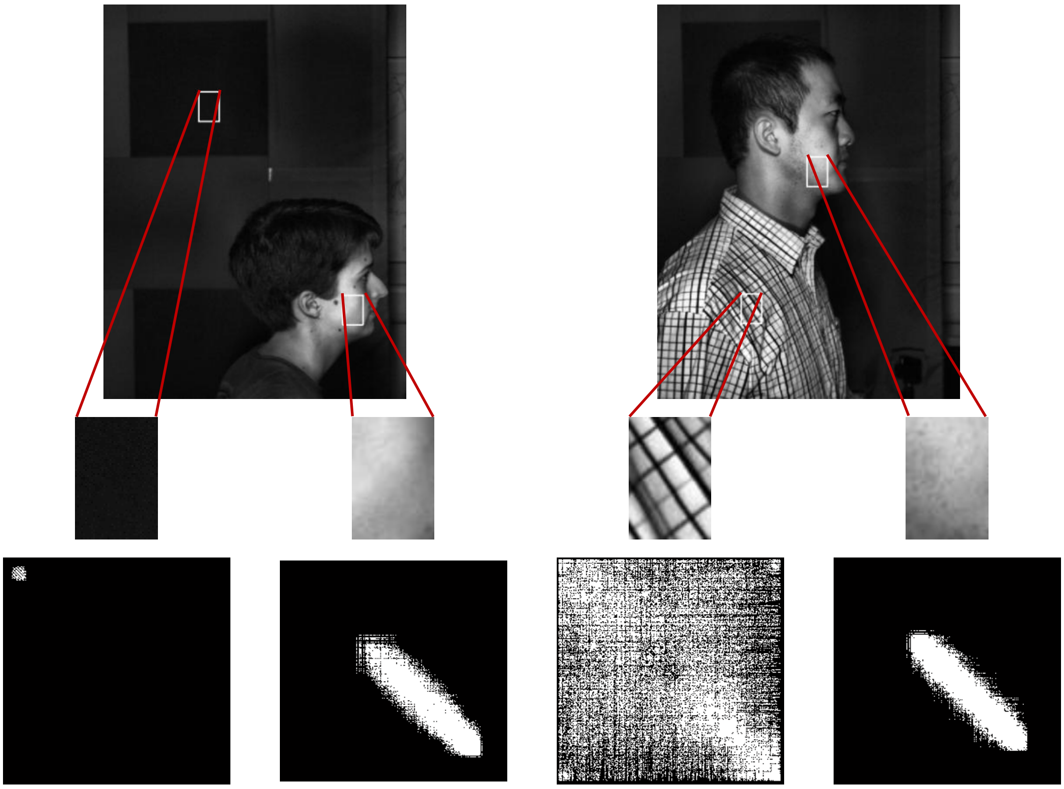


Figure 3.18: The co-occurrence matrices for skin and non-skin regions

$$energy = \sum_{b1=1}^{N}\sum_{b2=1}^{N} Co(b1,b2)^2 \tag{3.21}$$

which gives priority to larger items. Each measure ($inertia$ and $energy$) describes a different facet of the underlying data.

3.4.1.2 Skin Colour

Skin texture detection sometimes detects non-skin region as skin region, so skin colour is combined with skin texture to detect the skin region only. Shin et al. [57] pointed out that among 9 colour spaces, YCbCr is the best 2D colour space for skin detection. The RGB components are separated into luminance (*Y*), chrominance blue (Cb) and chrominance red (Cr). In the YCbCr colour space, a human skin colour model can be considered practically independent of the luminance and the skin colour is concentrated in a small region of the Cb - Cr plane [58]:

$$Cb = 128 - 37.797 \cdot R - 74.203 \cdot G + 112 \cdot B \tag{3.22}$$

$$Cr = 128 + 112 \cdot R - 93.786 \cdot G - 18.214 \cdot B \tag{3.23}$$

3.4.2 Adaptive Filter

The adaptive filter algorithm [59] is applied to the face profile image to enhance the edges. The algorithm consists of two adaptive filters with different cutoff frequencies and two different DC (zero frequency) gains (the gain differ by one). The two adaptive filters are connected in parallel-like structure to perform image edge enhancement (see Appendix A for further details). Figure 3.19 shows examples of applying the adaptive filter to face profile images. As illustrated in the figure, the edges of each image specially those of ears are enhanced.

Figure 3.19: Examples of applying the adaptive filter

3.5 Conclusions

This chapter shows how banana wavelets can be used to find the ear from head profile images for biometric purposes. The proposed technique detects ear automatically and does not require any user interaction. Therefore, it is appropriate to be used in an automatic ear based biometric system. A preprocessing algorithm is also applied to the face profile image before applying our ear detection technique. The preprocessing algorithm consists of two techniques: skin detection and adaptive filter. The skin regions are segmented to isolate the face and ear regions from the hair, clothing and background. Skin texture and colour are used to detect the skin region in face profile images. The adaptive filter is applied to enhance the edges of the image.

Chapter 4

Experimental Results

The analysis is performed using Matlab on Windows operating system on a Core2Duo CPU 2.4 Ghz with 2 GB of RAM. The aim of the experimental results section is to illustrate the results of enrolment technique used in this thesis.

4.1 Database Description

The efficiency of banana wavelet technique is tested using databases of 2D images.There are two databases are used in our experiments:the XM2VTS face profile database and the SOTON database.

4.1.1 XM2VTS Database

The first ear database is selected from the XM2VTS face profile database [60]. The selected dataset consists of 252 images from 63 individuals with four images per person collected during four different sessions over a period of five months to ensure the natural variation between the images of the same person. The images selected are those where the whole ear is visible in a 720×576 24-bit image. The ears in the dataset are not occluded by hair but there are few images with some occlusion by earrings. This is the same dataset used by Hurley et al. [2] and Arbab-Zavar et al. [27] and thus offers an avenue for comparison of performance with previous approaches. Figure 4.1 illustrates samples from XM2VTS face profile database.

Figure 4.1: Samples from XM2VTS face profile database

4.1.2 SOTON Database

The second ear database (SOTON dataset 1) is selected from SOTON database [61].The database is acquired as subjects walk past a camera triggered by a light beam signal through a tunnel, and other biometrics are acquired at the same time (face and gait biometrics). As shown in Figure 4.2, the tunnel has a central region that participants walk along, with the face and ear cameras placed at the end of the walkway and the gait video cameras positioned around the upper perimeter of the tunnel.

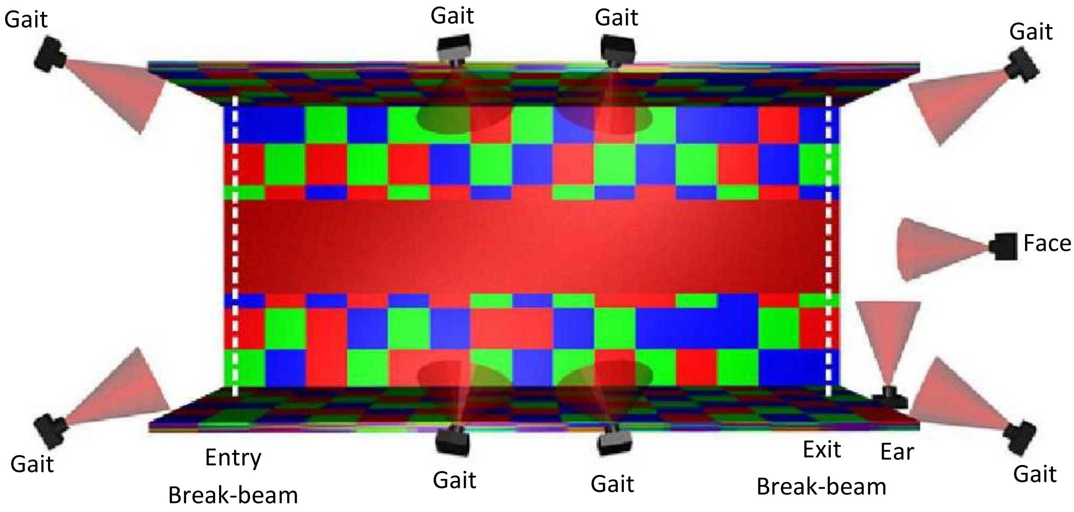


Figure 4.2: Placement of cameras and break-beam sensors in system

The advantage of SOTON dataset 1 is that, it has a lot of variations of ear orientation, size, colour skin, and lighting condition, allowing investigation of the performance of our technique on a data acquired in a more realistic scenario. In addition to that, this dataset has a large number of images which contains 2875 images. As the acquisition is largely uncontrolled, subjects sometimes present the whole head without occlusion and other combination with partial or large occlusion, and partial or sometimes the head is even absent. Clearly, this is a more demanding database than those used previously. This is the first use of the new database for evaluation, so there is no comparative analysis available.

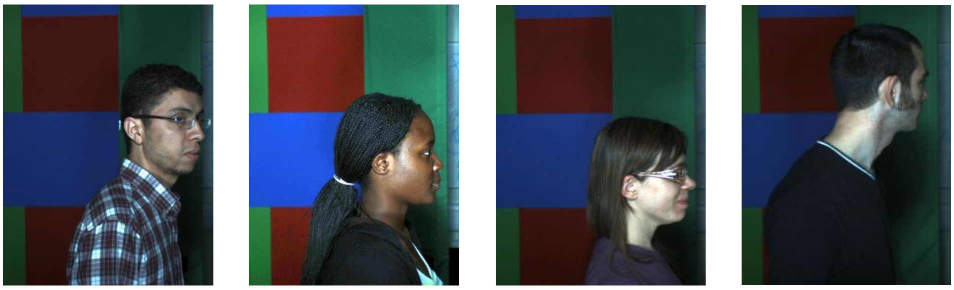
We have classified the database into five groups according to two criteria (Table 4.1). The first criterion is according to no occlusion, small occlusion and large occlusion. The second criterion is according to partial head and full head. Samples from our classified database are shown in Figure 4.3.

	Number of samples
Group A: whole head, no occlusion	1100
Group B: whole head, small occlusion	323
Group C: partial head, no occlusion	798
Group D: partial head, small occlusion	238
Group E: large occlusion	416

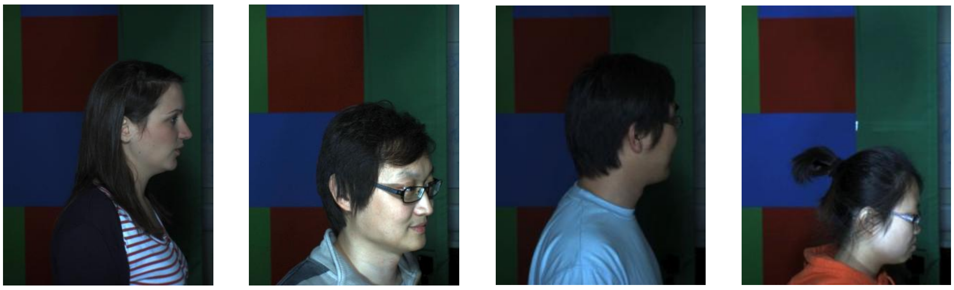
Table 4.1: Classification of SOTON dataset 1 into five groups according to head appearance and occlusion

4.2 Ear Detection Results

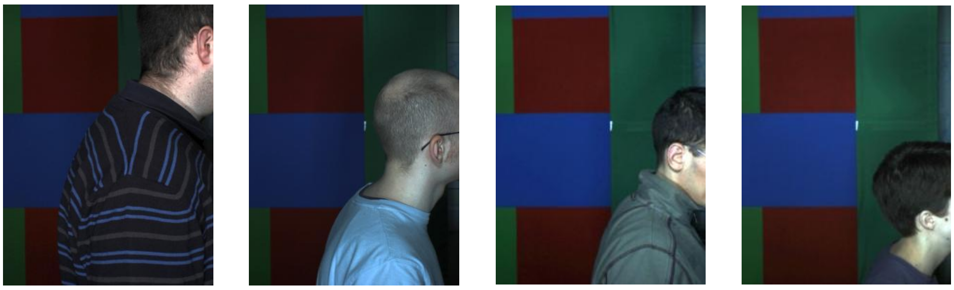
The proposed technique correctly detects all the ears in the images in the database (the detection rate was 100%). In addition to that, the technique detects an accurate bounded window around the ear according to the size of each ear. Some results of detection using banana wavelets are shown in Figure 4.4. The system is fully automatic and it does not require any manual interference for ear detection. As such



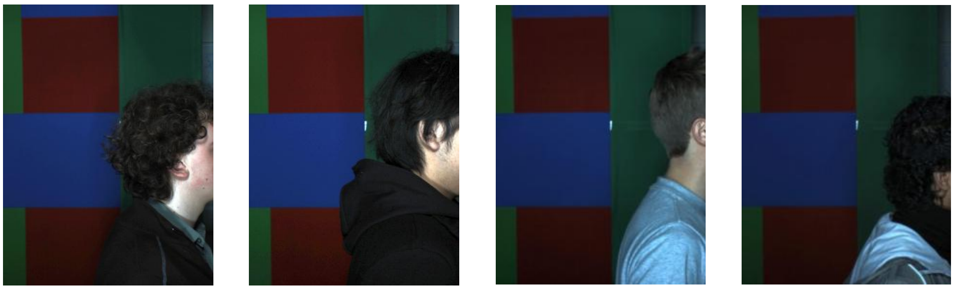
(a) Group A



(b) Group B



(c) Group C



(d) Group D

(e) Group E

Figure 4.3: Samples from SOTON dataset 1: (a)-(e) show the groups according to Table 4.1

the approach appears suitable for real time biometric applications. The parameters used in these results are $\sigma_x = 1$, $\sigma_y = 2\pi$, $\gamma = 0.08$, $\lambda = 1.8$, and $w = 7$.

The technique appears robust to degradation of images such as the motion blur, shown in Figure 4.5. It is also accurate and robust to some rotations (note the large subject rotation in Figure 4.6). In addition, the technique is robust in the presence of earrings (Figure 4.7) and glasses (Figure 4.8).

Arbab-Zavar and Nixon [27] applied a Hough Transform (HT) to locate the elliptical shape of the ear. They reported 100% detection rate on the same XM2VTS dataset (the same detection rate as here). They depended on some conditions such as cropping the third of the front of the faces to eliminate the effect of spectacles which also appeared as elliptical shape. Hurley et al. [2] also used the same XM2VTS dataset. They proposed a recognition algorithm without need to detect the ear from the face profile.

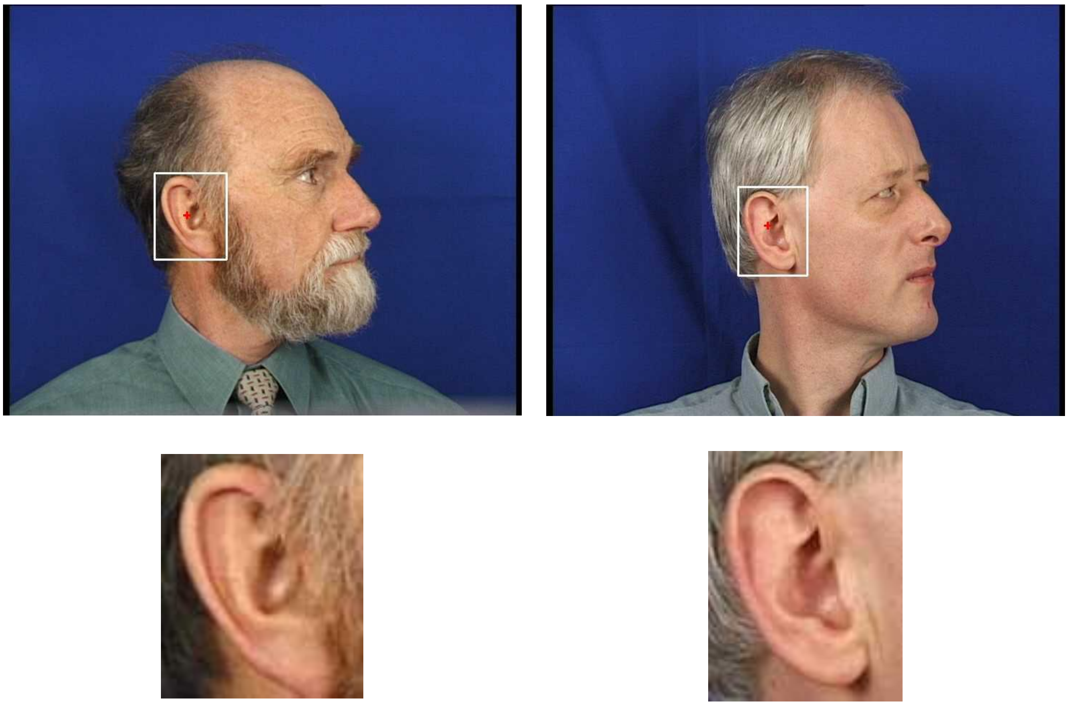


Figure 4.4: Samples of ear detection using our technique

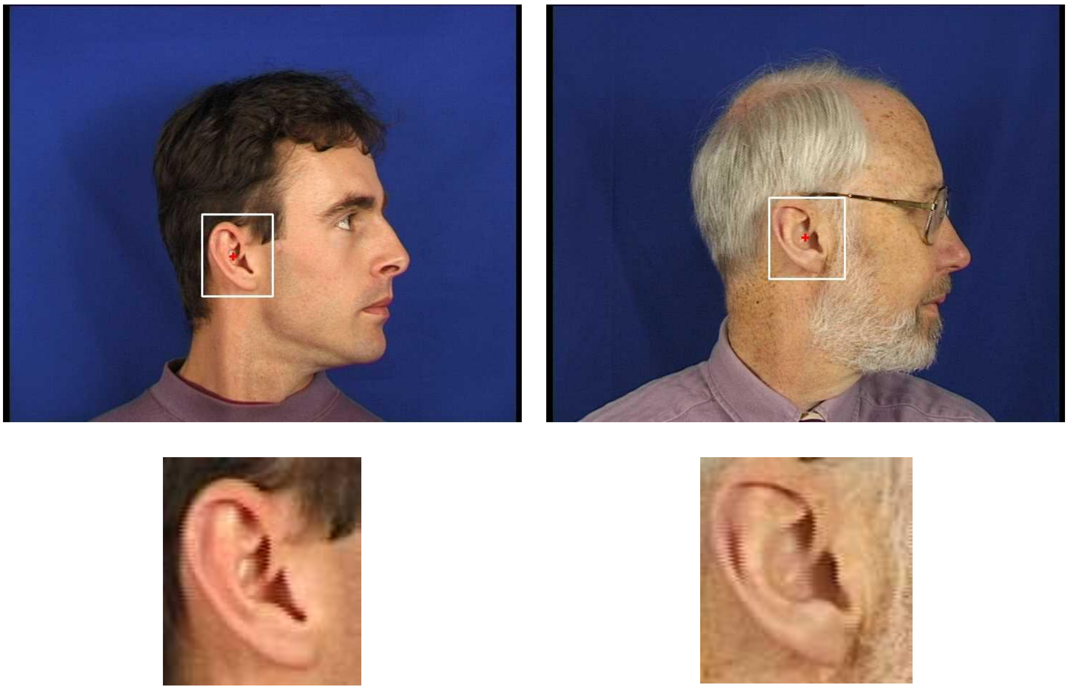


Figure 4.5: Samples for detection of blurred ears using our technique

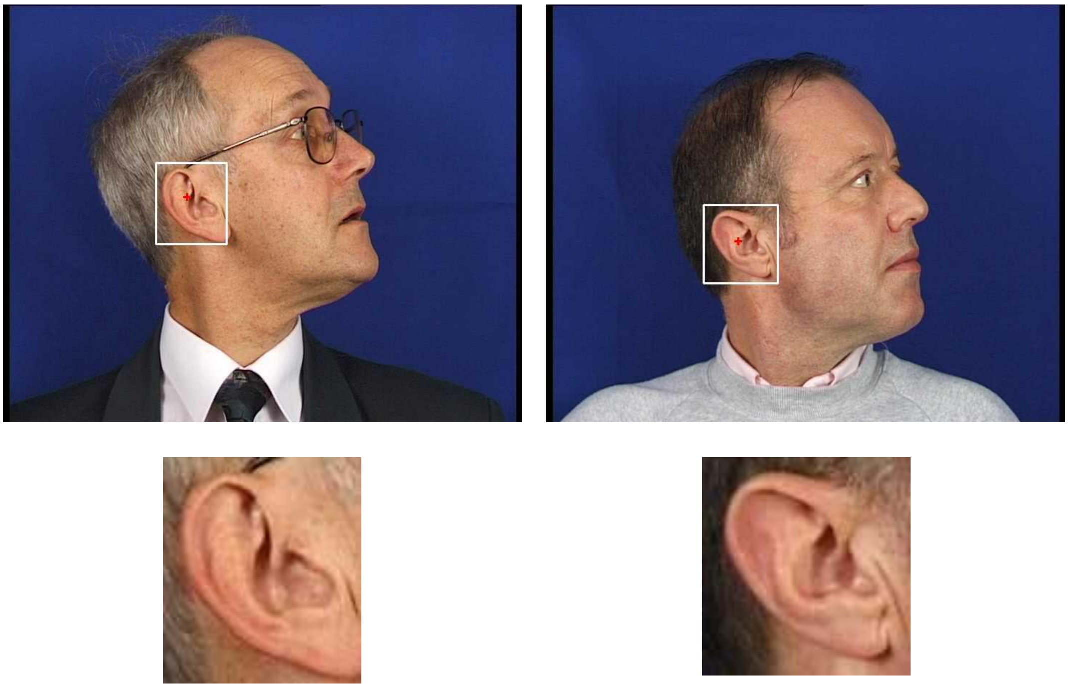


Figure 4.6: Samples for detection of rotated ears using our technique

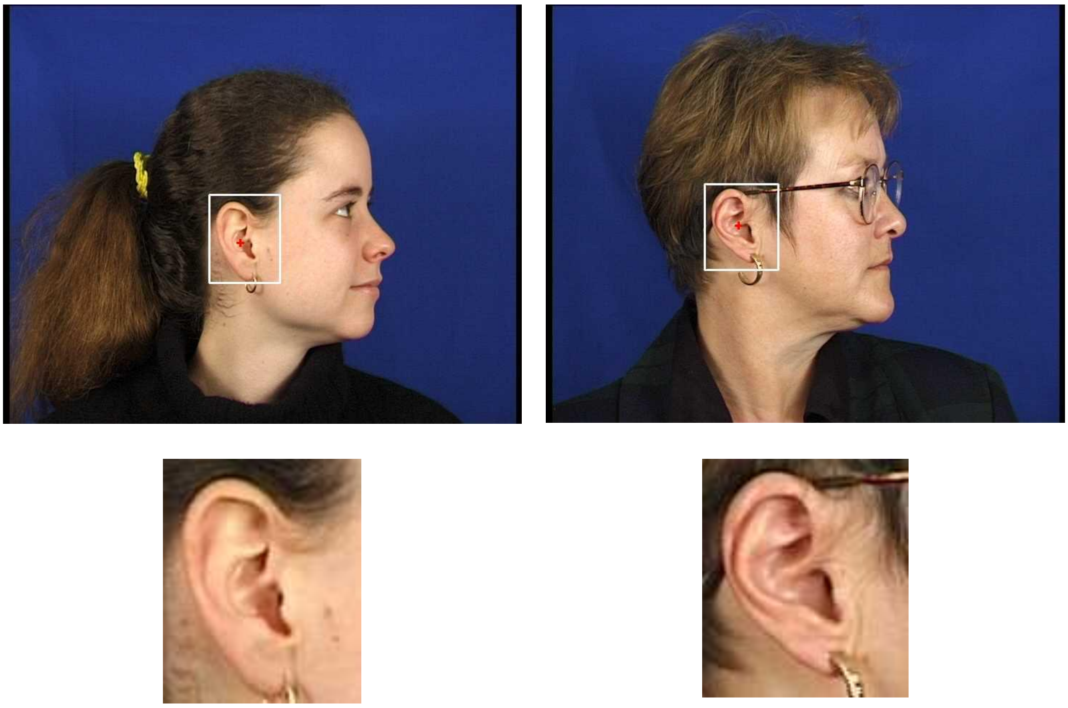


Figure 4.7: Samples for detection of occluded ears by earrings using our technique

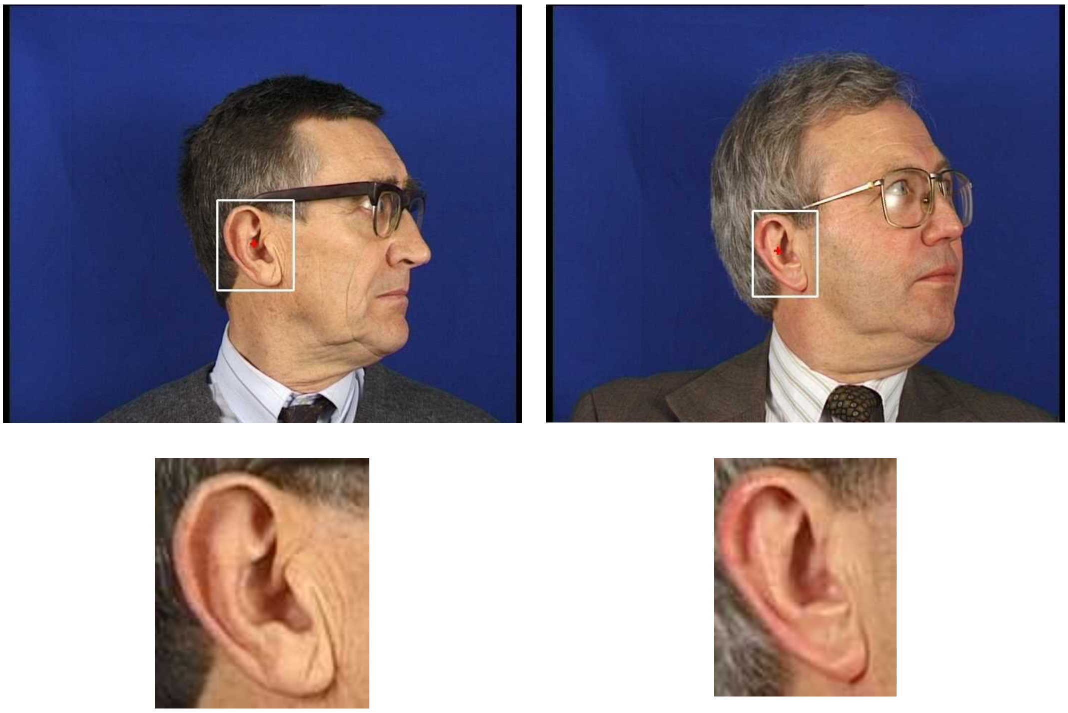


Figure 4.8: Samples for detection in presence of glasses using our technique

We also apply the banana wavelet based technique to detect the ears from SOTON dataset 1. Table 4.2 shows the results of applying banana and Gabor wavelet techniques to the SOTON dataset 1. The detection rate is lower than that for the XM2VTS database since the images are considerably more challenging, as consistent with a less controlled and hence more realistic scenario. The Gabor approach is of consistently lower performance than the Banana wavelet approach.

		Detection rate
Group A	Banana filters	944/1100 ~ 85.8%
	Gabor filters	695/1100 ~ 63.2%
Group B	Banana filters	213/323 ~ 65.9%
	Gabor filters	127/323 ~ 39.3%
Group C	Banana filters	639/798 ~ 80.1%
	Gabor filters	443/798 ~ 55.5%
Group D	Banana filters	105/238 ~ 44.1%
	Gabor filters	65/238 ~ 27.3%
Group E	Banana filters	81/416 ~ 19.5%
	Gabor filters	41/416 ~ 9.9%

Table 4.2: Ear detection rates for SOTON dataset 1 by applying banana and Gabor filters

The same parameters used in the analysis of the XM2VTS database are employed in this analysis. As such, the approach is not tuned for the SOTON database (and scenario) and it is likely that these results could be improved further. Figure 4.9 shows some results of applying banana wavelet filters to group A (whole head, no occlusion).The results of the correct detected ears in group C (partial head, no occlusion) are shown in Figure 4.10.

Figure 4.9: Samples of correct ear detection from group A (SOTON dataset 1)

Figure 4.10: Samples of correct ear detection from group C (SOTON dataset 1)

4.3 Occlusion Analysis

Detecting occluded ear images is more challenging, due to the missing parts from the ear image. To analyze the ear images with occlusion, we synthetically occlude the ear region of the face-profile images in the XM2VTS database from top and left by 10% to 60% of the ear image. Our technique shows promising results when used to enrol partially occluded ears.Figure 4.11 and Figure 4.12 show enrolled ear samples after occluding them from top and left; respectively. The results of the occlusion analysis are provided in Figure 4.13, which illustrates the detection success rate of the XM2VTS database against increasing synthetic occlusion. As shown in Figure 4.13-a, the detection is still above 99% at 30% occlusion from top, and around 90% of the ears are enrolled successfully in case of 40% occlusion. The detection rate reduces to 33% for 60% occlusion from top. Also as illustrated in Figure 4.13-b, for the occlusion from left, above 99% of the ears are enrolled successfully in case of 40% occlusion, and it drops to 39% at 60% occlusion from the left.

For SOTON dataset 1, we do not need to occlude the ear region synthetically because there are many occluded ears in this dataset. Table 4.2 shows the results of our ear detection technique for group B (whole head, small occlusion), group D (partial head, small occlusion), and group E (large occlusion). As expected the correct ear detection rate of group E is low since this group contains occlusion of the ear region more than 50%, and in many cases ear region is occluded more than 80% by hair. The correct detection rates of occluded ears of SOTON dataset 1 are close to those of XM2VTS database, which indicates that, our synthetically occluded ears look natural and realistic enough to have similar results with naturally occluded SOTON dataset 1.

Figure 4.14, Figure 4.15 and Figure 4.16 illustrate samples of correct ear detection from group B, group D and group E, respectively. Figure 4.17 shows some samples of incorrect ear detection from group B, group D and group E. As illustrated in the figure our ear detection technique fails to detect the ear correctly which there is large occlusion (nearly 90% for the third subject).

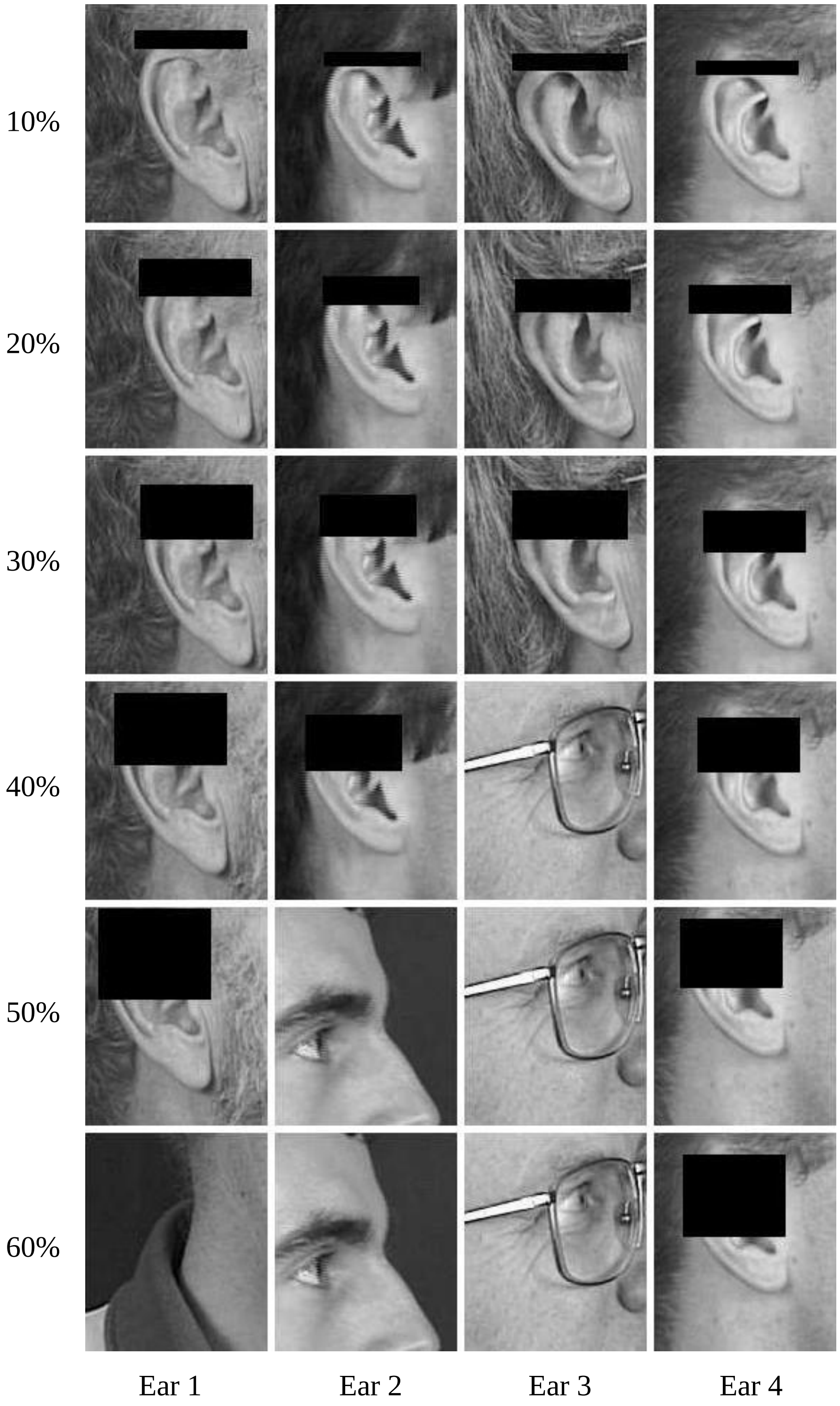


Figure 4.11: Detection results for four ears with occlusion from top

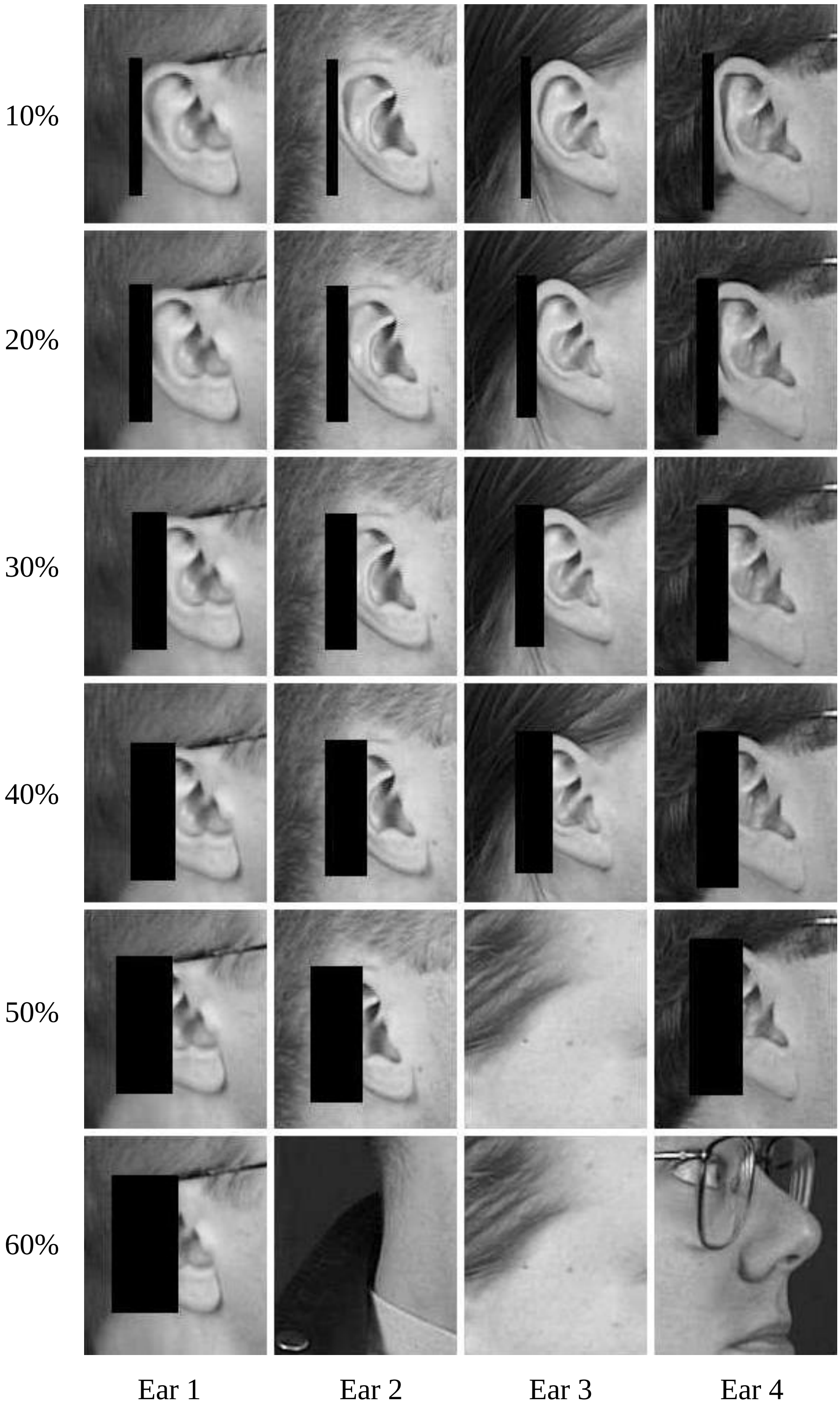
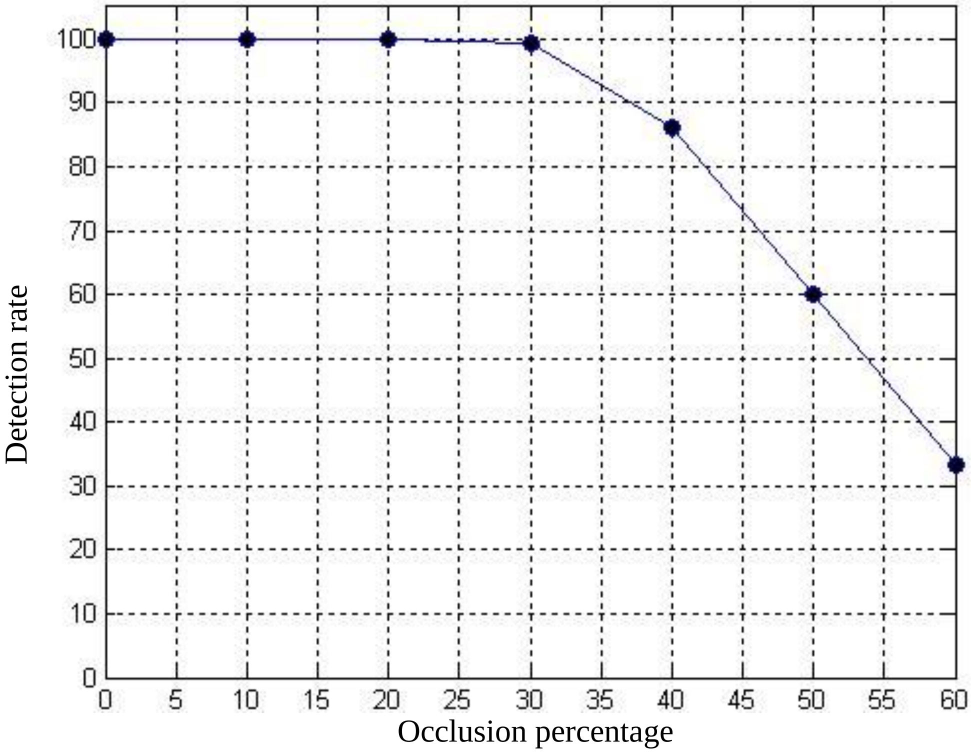
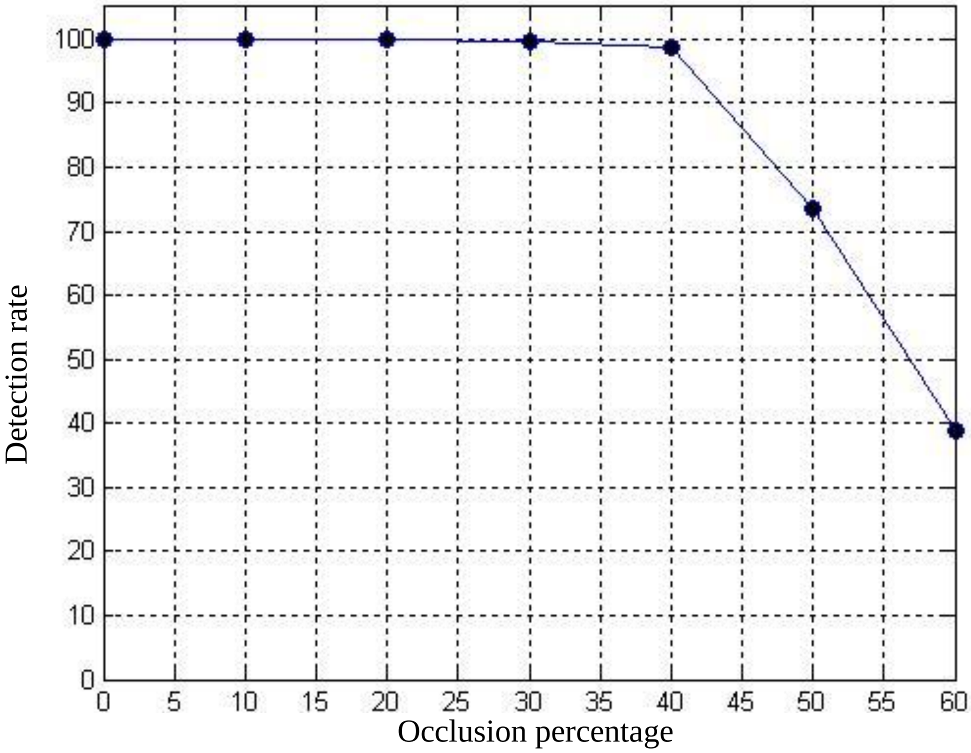


Figure 4.12: Detection results for four ears with occlusion from left



(a) Occlusion from top



(b) Occlusion from left

Figure 4.13: Detection rate for ears with occlusion from top and left

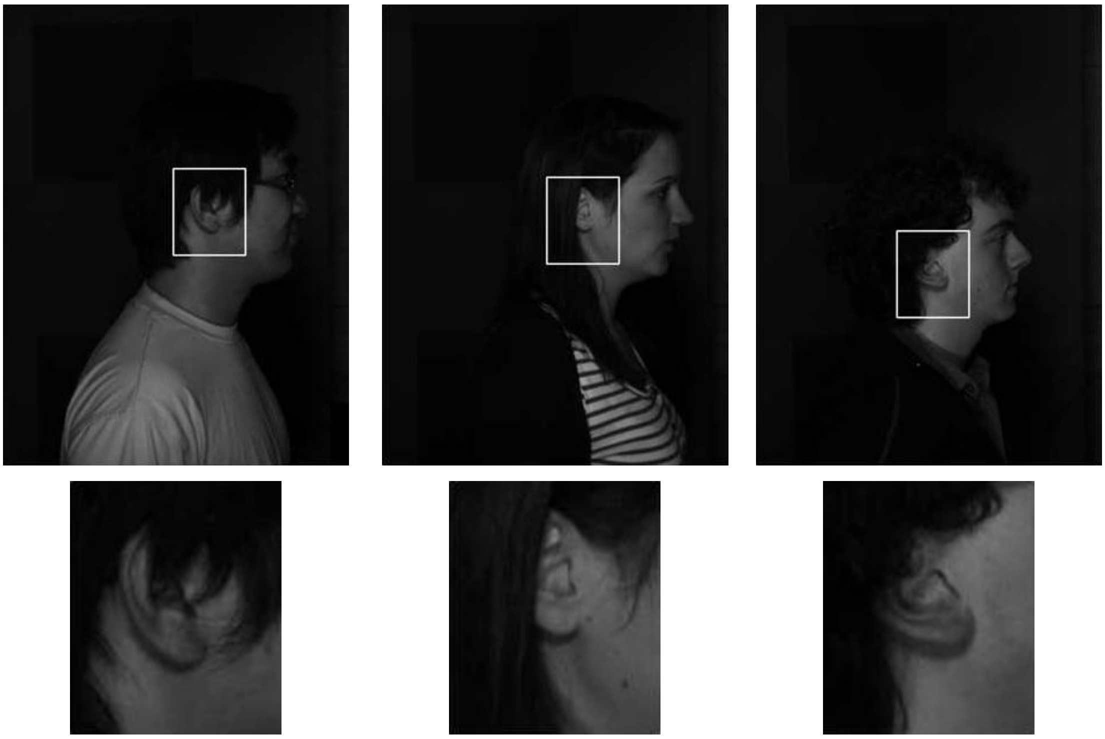


Figure 4.14: Samples of correct ear detection from group B (SOTON dataset 1)

Figure 4.15: Samples of correct ear detection from group D (SOTON dataset 1)

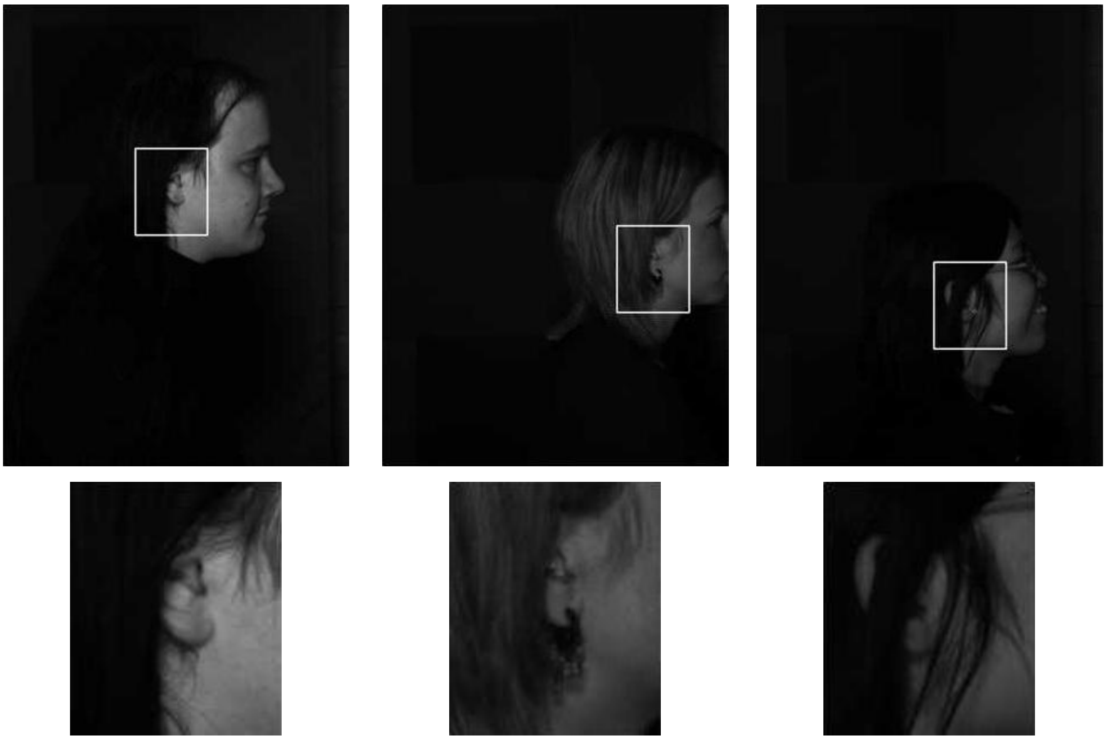


Figure 4.16: Samples of correct ear detection from group E (SOTON dataset 1)

Figure 4.17: Samples of incorrect ear detection from group B, group D and group E (SOTON dataset 1)

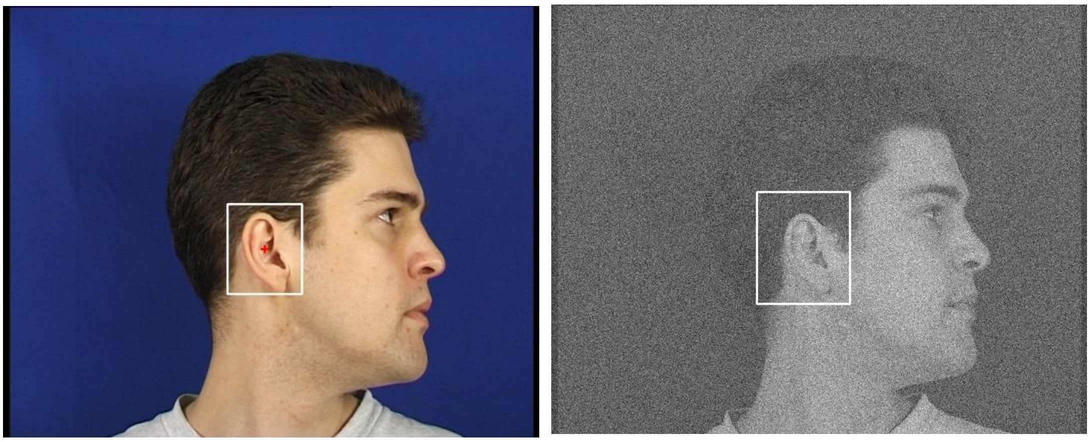
4.4 Noise Analysis

The proposed technique also appears robust to noise. The accuracy of detection in the presence of noise is more than 90% when the noise standard deviation η is quite high. These results are illustrated in Figure 4.18. Here, images are contaminated by additive zero mean Gaussian noise with various noise standard deviations. Naturally, this noise could be filtered by application of a smoothing operator, but the purpose here is to investigate performance basis, rather than in tuned system. The technique proposed here is successful until η = 500 in which case a region containing the eye is erroneously selected (see Figure 4.18-f).

The results of the noise analysis are provided in Figure 4.19. We note that for noise-free images the enrolment rate is 100%. As expected, the enrolment rate drops with increasing noise. The detection rate for η = 0 to η = 500 is shown in Figure 4.19-a where at η = 100 the detection rate is around 90% but this drops to under 43% when η = 200. This is actually quite a severe level of noise, as shown in Figure 4.18-d. In much worse cases, such as the one shown in Figure 4.18-f, the detection drops to about 5% which is to be expected given the severity of the effects of noise. Figure 4.19-b illustrates more detail for the detection rate from η = 0 to η = 100. The detection rate drop to 97% at η = 20.Then, it decreases steadily to slightly above 95.5% until η = 70. After that, the detection drops to about 90% at η = 100.

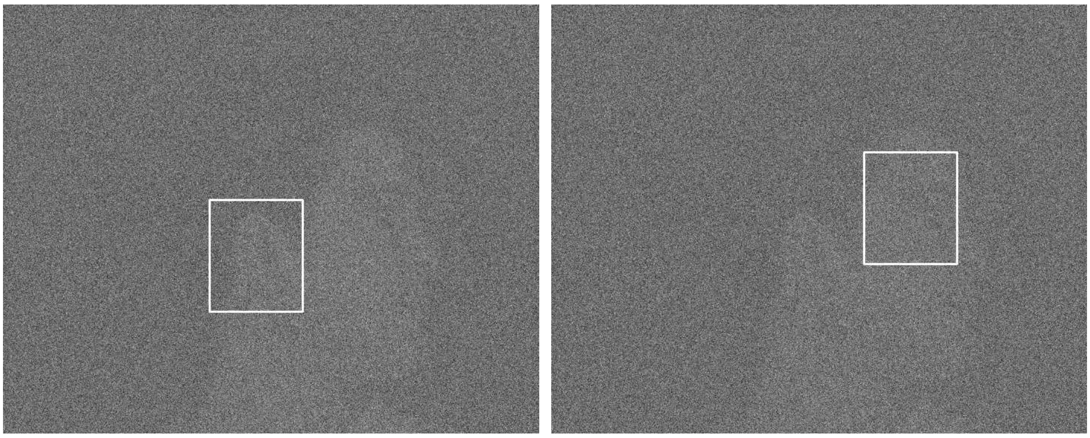
4.5 Gabor Filter Analysis

Banana wavelet filters can capture the curved structures better than Gabor wavelet filters. To show this, Gabor wavelet technique is applied to the same subset of the XM2VTS face profile database with the same filter sizes, orientations and frequencies as those of banana wavelet filters (the same parameters in Table 3.2, except that curvature c was set to zero). The detection rate obtained by Gabor wavelets is 97.2%.



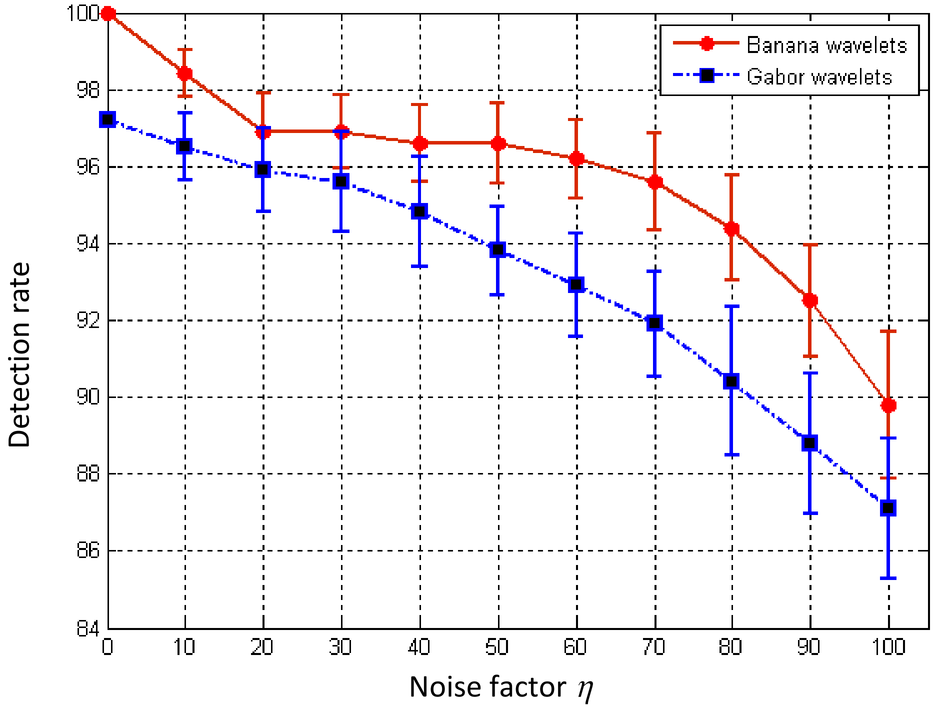
(a) 0% noise (b) η = 50

(c) η = 100 (d) η = 200



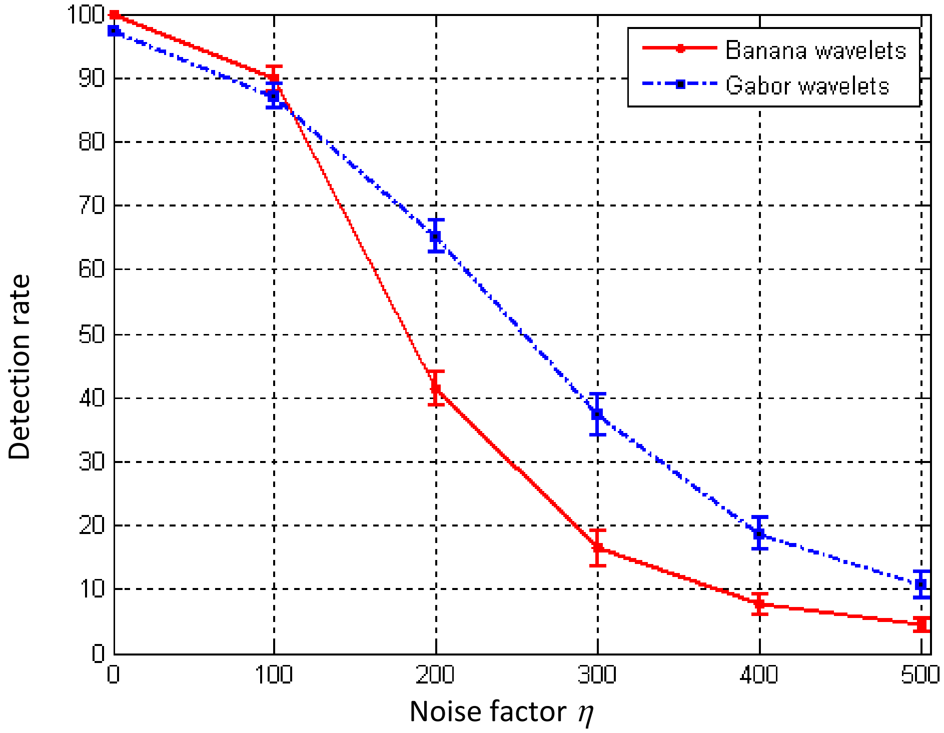
(e) η = 400 (f) η = 500

Figure 4.18: Samples for the results in presence of noise



SNR (19.12) (13.10) (9.58) (7.08) (5.14) (3.56) (2.22) (1.06) (0.04) (-0.88) db

(a) Detection rate for η = 0 to η = 100



SNR (-0.88) (-6.90) (-10.42) (-12.92) (-14.86) db

(b) Detection rate for η = 0 to η = 500

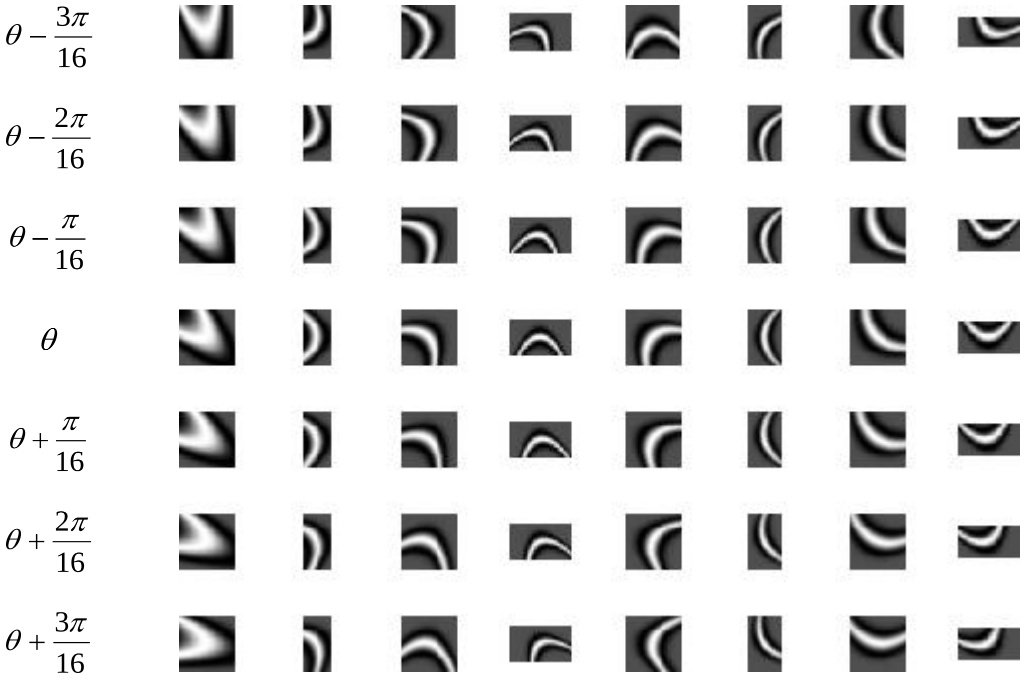
Figure 4.19: Detection rate for banana wavelets technique and Gabor wavelets technique in the presence of noise

The Gabor wavelet technique has also been applied to SOTON dataset 1. Table 4.2 shows the detection rates of applying Gabor filters to SOTON dataset 1. As illustrated in the table, for all of the five groups, the results of applying banana filters are better than those of applying Gabor filters.

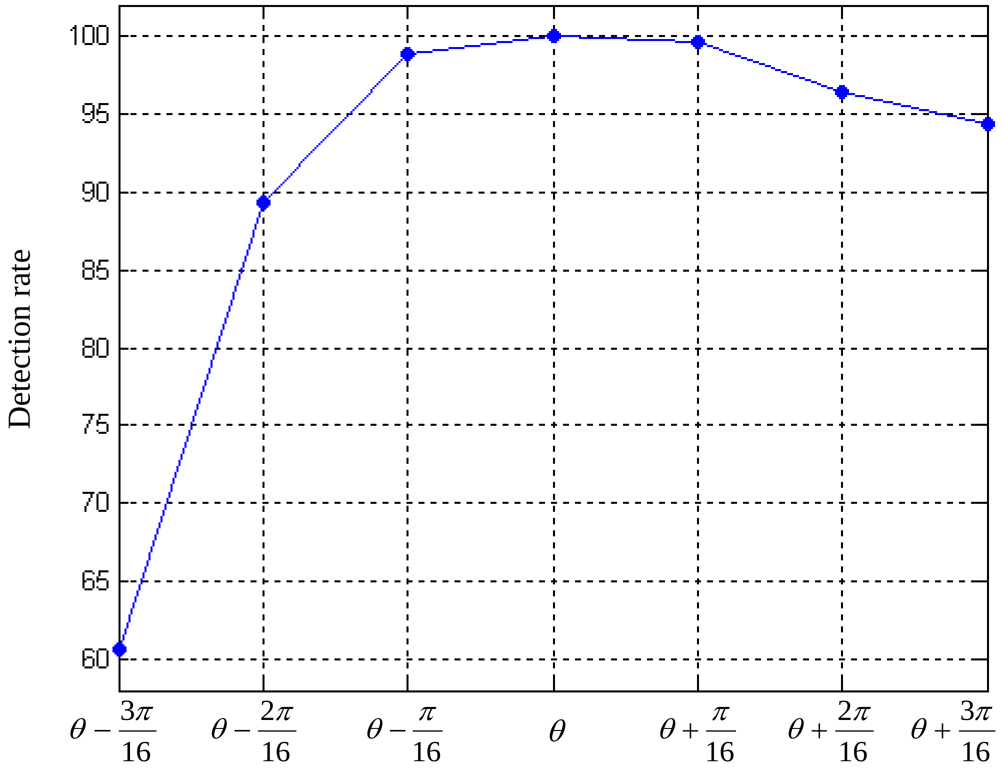
Figure 4.19 shows a comparison between the performance of the Gabor wavelet and banana wavelet in the presence of noise. For lower values of noise (Figure 4.19-a) the banana wavelet is more successful than the Gabor wavelet. As the noise increases (Figure 4.19-b), the advantages associated with using curvature become masked by the noise and in cases of severe noise the Gabor wavelet is actually more successful than the banana wavelets. However, Figure 4.18 suggests that such noise corruption is highly unlikely in a realistic scenario.

4.6 Parameter Sensitivity

The parameters used to detect the ear region are chosen by experiment. Generally, the purpose of parameters sensitivity is to determine which banana filter parameter has the most influence on the ear detection rate. Parameters sensitivity is determined by varying each parameter while keeping the other parameters unchanged. Experiments show that the two parameters with the most important influence on the detection rate are orientation θ and frequency f of filters. Figure 4.20shows the effect of changing θ on the detection rate. As shown in Figure 4.20-a $\theta + n$ perseveres the main structure of the ear but with slight rotation, whereas $\theta + n$ completely changes the structure of the ear. So the detection rate drops slightly to 94.4% when θ is changed by $+3\pi/16$ but drops to around 60% when θ parameter is varied by $-3\pi/16$ (as shown in Figure 4.20-b). It is clear from Figure 4.21-a that $f + n$ is very similar to f without change whereas $f - n$ is very coarse. Therefore the detection rate slightly drops to 98% when f parameter is varied by +0.18 but drops to just above 53% when f is changed by -0.18 (as shown in Figure 4.21-b).

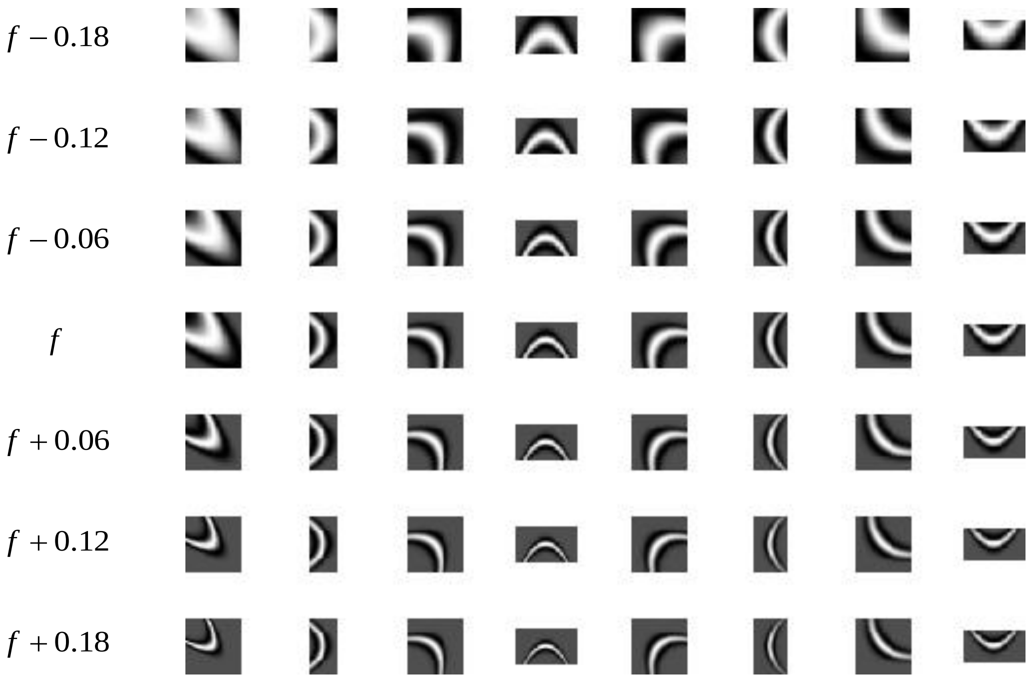


(a) chnage in theta parameter

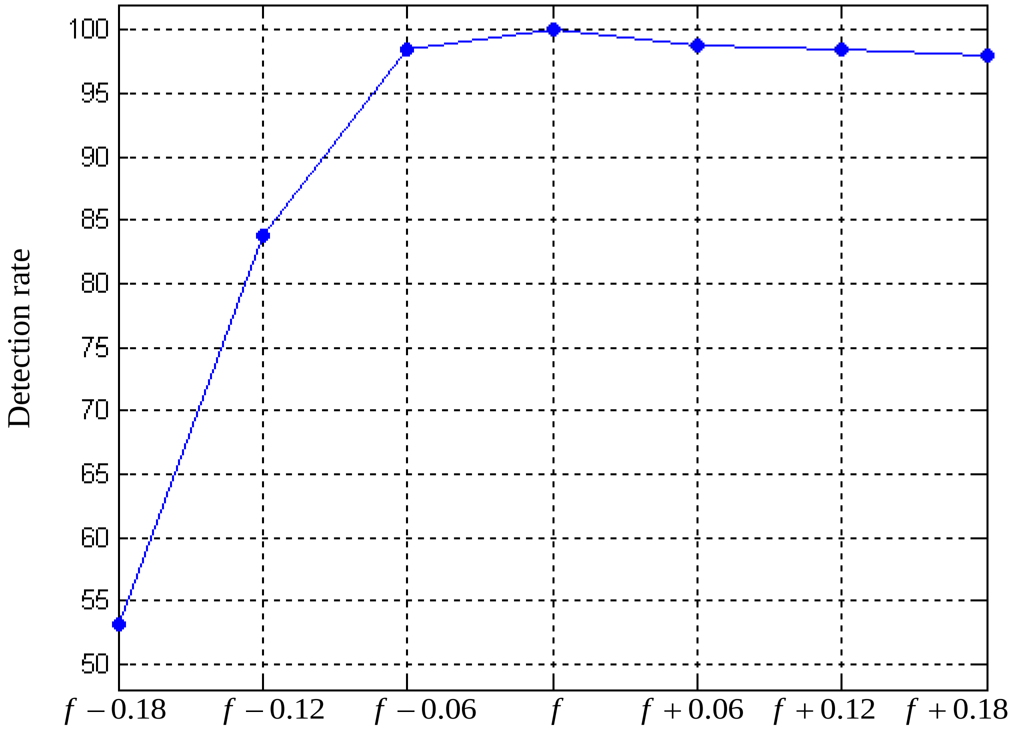


(b) detection results versus change in theta parameter

Figure 4.20: The effect of changing theta parameter on the detection rate



(a) change in frequency parameter



(b) detection results versus change in frequency parameter

Figure 4.21: The effect of changing frequency parameter on the detection rate

θ		f		c		s		R,C	
change	detection rate	change	detection rate	change	detection rate	change	detection rate	change	detection rate
$-\frac{3\pi}{16}$	60.7%	-0.18	53.2%	-0.04	98.0%	$\frac{s}{4}$	91.3%	-20	97.2%
$-\frac{2\pi}{16}$	89.3%	-0.12	83.7%	-0.02	100%	$\frac{s}{2}$	99.6%	-10	96.0%
$-\frac{\pi}{16}$	98.8%	-0.06	98.4%	0	100%	0	100%	0	100%
0	100%	0	100%	+0.02	98.8%	$2s$	100%	+10	99.2%
$+\frac{\pi}{16}$	99.6%	+0.06	98.8%	+0.04	97.2%	$4s$	100%	+20	96.4%
$+\frac{2\pi}{16}$	96.4%	+0.12	98.4%						
$+\frac{3\pi}{16}$	94.4%	+0.18	98.0%						

Table 4.3: Parameters sensitivity for ear detection and the change in each parameter

The detection results are much less affected by the other parameters (c, s, R and C). Table 4.3 illustrates the effect of change in each parameter on the detection rate.

4.7 Preprocessing

The effectiveness of the preprocessing step is tested using SOTON dataset 1. The skin detection technique and the adaptive filter are applied to the five groups in SOTON dataset 1.To determine the threshold values of the skin detection technique, skin regions are segmented manually from SOTON dataset 1 and analysed. The regions are classified as skin if their $inertia, energy, Cb$ and Cr values are between the following thresholds:

$$20 \leq inertia \leq 2200, \tag{4.1}$$

$$0 \leq energy \leq 0.01, \tag{4.2}$$

$$87 \leq Cb \leq 130 \text{, and} \tag{4.3}$$

$$129 \leq Cr \leq 176 \tag{4.4}$$

Table 4.4 compares the ear detection rate based on the following methods:

i) applying our ear detection technique without the preprocessing step,
ii) applying the adaptive filter before our ear detection technique,
iii) applying the skin detection before our ear detection technique, and
iv) applying adaptive filter and skin detection before our ear detection technique.

As illustrated in the table, the results show the effect of the preprocessing step. Here, the ear detection rate of Group A (whole head, no occlusion) increases from 85.8% to 97.4%. The ear detection rates of the other groups also show an improvement, ranging from 16.1% increase in Group C (partial head, no occlusion) to 34.9% in Group D (partial head, small occlusion). Table 4.4 also compares the detection rate of applying banana and Gabor filters. As shown in the table, the results of applying banana filters to all of the five groups (based on the different methods described above) are better than those of applying Gabor filters. An example of applying our ear detection technique with the preprocessing step is shown in Figure 4.22.

4.8 Conclusions

In this chapter the proposed technique is tested on the XM2VTS face profile database and the SOTON database. The ear detection technique correctly detects all the test images selected from the XM2VTS database. Our technique demonstrates encouraging results when used to enrol partially occluded ears or in the presence of noise. When the proposed technique is applied to SOTON dataset 1, which the acquisition of this database is largely uncontrolled, the technique does not assume any prior normalization of the ear region and the detection result is found to be good enough to varying light conditions, skin colour, occlusion and different head rotations. As expected, there is an improvement in the ear detection rate when applying our ear detection technique with the preprocessing step.

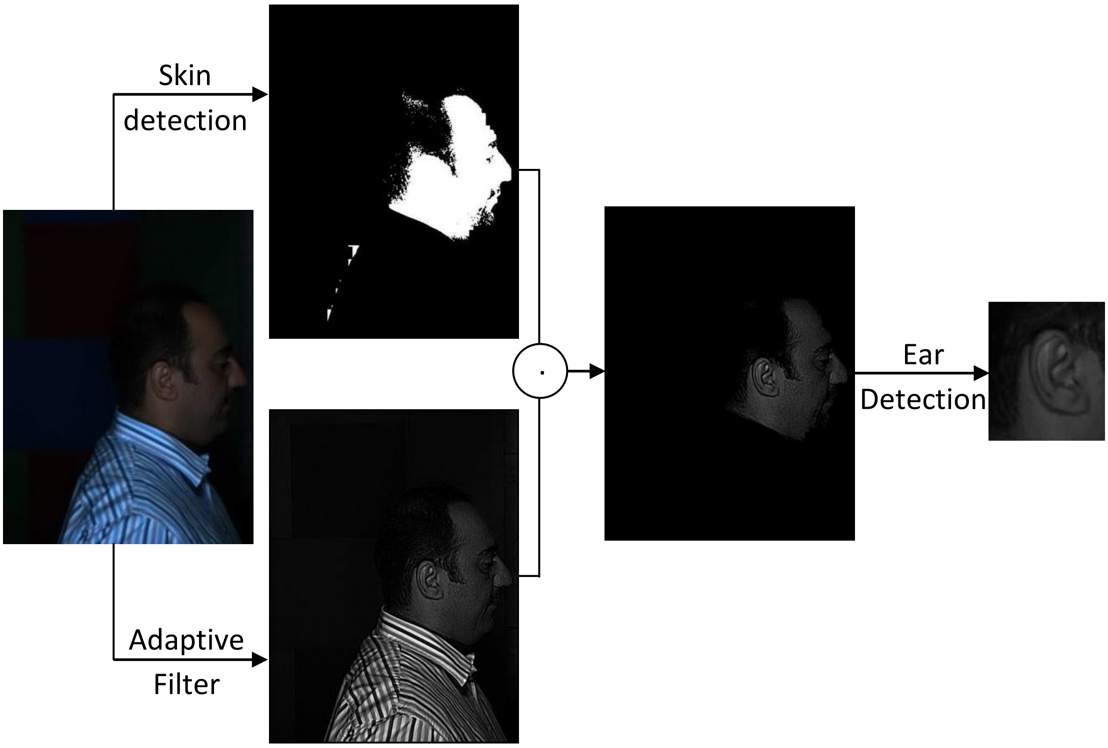


Figure 4.22: Example of applying our ear detection technique with the preprocessing step

		No Preprocessing	Adaptive Filter	Skin Detection	Adaptive Filter + Skin Detection
Group A	Banana filters	944/1100 85.8%	1030/1100 93.6%	1038/1100 94.4%	1071/1100 97.4%
	Gabor filters	695/1100 63.2%	805/1100 73.2%	896/1100 81.5%	922/1100 83.8%
Group B	Banana filters	213/323 65.9%	241/323 74.6%	256/323 79.3%	279/323 86.4%
	Gabor filters	127/323 39.3%	138/323 42.7%	165/323 51.1%	172/323 53.3%
Group C	Banana filters	639/798 80.1%	702/798 88.0%	755/798 94.6%	768/798 96.2%
	Gabor filters	443/798 55.5%	543/798 68.0%	680/798 85.2%	715/798 89.6%
Group D	Banana filters	105/238 44.1%	119/238 50.0%	161/238 67.6%	188/238 79.0%
	Gabor filters	65/238 27.3%	67/238 28.2%	140/238 58.8%	136/238 57.1%
Group E	Banana filters	81/416 19.5%	108/416 26.0%	136/416 32.7%	171/416 41.1%
	Gabor filters	41/416 9.9%	52/416 12.5%	119/416 28.6%	120/416 28.8%

Table 4.4: Comparing the correct detection rates for SOTON dataset 1

Chapter 5

Ear Recognition

Biometrics concerns the recognition of individuals based on a feature vector extracted from their anatomical and/or behavioural characteristic, and plays a vital role in security and surveillance systems. Finding efficient descriptors to represent these characteristics is essential for many pattern recognition tasks.

5.1 Basic Techniques

This section briefly discusses analysis of variance (ANOVA) and local binary pattern (LBP) techniques required for the development of our proposed ear recognition algorithm.

5.1.1 Analysis of Variance (ANOVA)

ANOVA is a well established statistical method to provide a statistical analysis of data [62]. The purpose of ANOVA is to test the significant differences between the mean of two or more classes, and this is done by partitioning the total sum of squares into between-class SS_b and within-class SS_w. One way ANOVA is the simplest case, in which, all classes are compared with each other simultaneously instead of individually.

If $\chi_{i,j}$ is the sample for i^{th} class ($i = 1,2,\ldots,k$) and j^{th} data point ($j = 1,2,\ldots,ns_i$) then the grand total GT is:

$$GT = \sum_{i}^{k} T_i \tag{5.1}$$

where $T_i = \sum_{j}^{ns_i} \chi_{i,j}$ is the total of group i, $NS = \sum_{i}^{k} ns_i$ is the total number of samples, k is the number of classes, and ns_i is the number of samples in i^{th} class.

The sum of squares for the between class SS_b is defined as:

$$SS_b = \sum_{i}^{k}\left(\frac{T_i^2}{ns_i}\right) - \frac{GT^2}{NS} \tag{5.2}$$

with degrees of freedom:

$$df_b = k - 1 \tag{5.3}$$

Then, the mean squares for the between class MS_b is defined as:

$$MS_b = \frac{SS_b}{df_b} \tag{5.4}$$

The sum of squares for the within class SS_w is defined as:

$$SS_w = \sum_{i}^{k}\sum_{j}^{ns_i}\left(\chi_{i,j}\right) - \sum_{i}^{k}\left(\frac{T_i^2}{ns_i}\right) \tag{5.5}$$

with degrees of freedom:

$$df_w = NS - k \tag{5.6}$$

Then, the mean squares for the within class MS_w is defined as:

$$MS_w = \frac{SS_w}{df_w} \qquad (5.7)$$

Finally, the $\mathbf{F\text{-}ratio}$ is calculated by comparing between classes' mean squares MS_b with within classes' mean squares MS_w, thus reflecting discrimination capacity:

$$\mathbf{F\text{-}ratio} = \frac{MS_b}{MS_w} \qquad (5.8)$$

These calculations are used to analyse the null hypothesis. The null hypothesis states that there are no differences between the means of different classes. The null hypothesis is rejected if $\mathbf{F\text{-}ratio} \geq \mathbf{F}_{(df_b, df_w, \alpha)}$ which means there are differences between the means of different classes.

For classification, ANOVA is a fast and efficient way to evaluate the importance of each extracted feature [63]. The relative importance of a feature can be determined by its $\mathbf{F\text{-}ratio}$ in a descending order.

5.1.2 Local Binary Pattern (LBP)

The LBP texture operator is currently one of the most popular techniques for texture description and it has become a popular approach in various applications. The original LBP [64] operator labels every pixel of an image by thresholding the 3 x 3 neighbourhood of each pixel $px_i\,(i = 0,1,\ldots,7)$ with the value of the centre pixel px_c (as shown in Figure 5.1) and considering the result as a binary number:

$$LB(px_i - px_c) = \begin{cases} 1, & px_i \geq px_c \\ 0, & px_i < px_c \end{cases} \qquad (5.9)$$

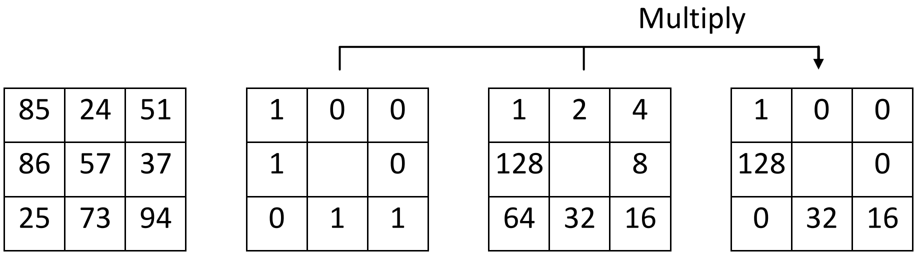
px_0	px_1	px_2
px_7	px_c	px_3
px_6	px_5	px_4

Figure 5.1: The basic LBP operator

Then, the values of the pixels in the thresholded neighbourhood are multiplied by the weights given to the corresponding pixels. Finally, the values of the eight pixels are summed to obtain the LBP number for the centre pixel:

$$LBP = \sum_{i=0}^{7} LB(px_i - px_c) \cdot 2^i \qquad (5.10)$$

The histogram of these 2^8 = 256 different labels can then be used as a texture descriptor. Figure 5.2 shows an example for the LBP operator. The advantage of using LBP texture operator is that it is invariant with illumination and shifting.



85	24	51
86	57	37
25	73	94

(a)

1	0	0
1		0
0	1	1

(b)

1	2	4
128		8
64	32	16

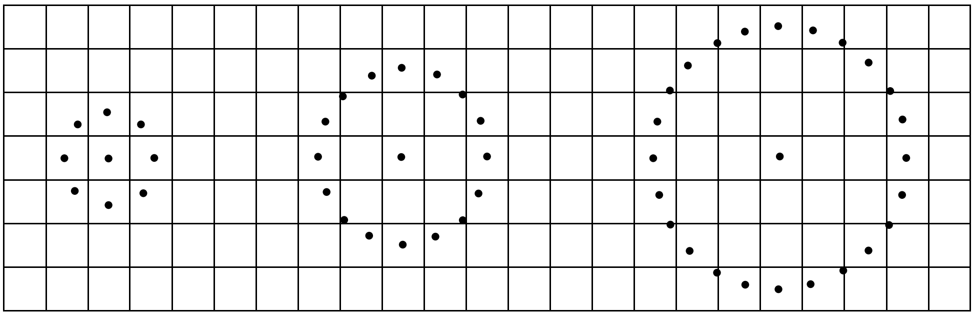
(c)

1	0	0
128		0
0	32	16

(d)

Figure 5.2: Example for the LBP operator

Ojala et al. [65] modified the original LBP into uniform local binary pattern. The LBP operator has been extended to consider different neighbourhood sizes with circular form. In general, the operator $LBP_{P,Rad}$ refers to a neighbourhood size of P equally spaced pixels on a circle of radius Rad that form a circularly symmetric neighbour set. For example, the operator $LBP_{8,1}$ uses only 8 neighbours on a circle of radius 1 while $LBP_{16,2}$ uses 16 neighbours on a circle of radius 2. Figure 5.3 illustrates three circularly symmetric neighbour sets for different values of P and Rad.



(a) $LBP_{8,1}$ (b) $LBP_{16,2}$ (c) $LBP_{24,3}$

Figure 5.3: Three different circular neighbourhoods

If g_c corresponds to the grey value of the centre pixel of a local neighbourhood and $g_p (p = 0,1,...,P-1)$ correspond to the grey values of P circular pixels on a circle of radius Rad (Rad >0), then the coordinates $\left(x_{g_p}, y_{g_p}\right)$ of the neighbours g_p are given by:

$$\begin{pmatrix} x_{g_p} \\ y_{g_p} \end{pmatrix} = \begin{pmatrix} x_c + Rad \cdot \cos\left(\frac{2\pi}{P} p\right) \\ y_c - Rad \cdot \sin\left(\frac{2\pi}{P} p\right) \end{pmatrix} \tag{5.11}$$

where $\left(x_c, y_c\right)$ are the coordinates of the centre pixel g_c.

The basic LBP is not rotation invariant but the uniform LBP solves the problem of rotation by rotating back each $LBP_{P,Rad}$ code to a reference position to remove the effect of rotation. This transformation can be defined as follows [66]:

$$LBP_{P,Rad}^{ri} = \min\left\{ROR\left(LBP_{P,Rad}, i\right) \mid i = 0,1,...,P-1\right\} \tag{5.12}$$

where

$$LBP_{P,Rad} = \sum_{p=0}^{P-1} LB(g_p - g_c) \cdot 2^p \tag{5.13}$$

$$ROR\left(LBP_{P,Rad}, i\right) = \begin{cases} \sum_{k-i}^{P-1} LB\left(g_k - g_c\right).2^k + \sum_{k-0}^{i-1} LB\left(g_k - g_c\right).2^{P-i+k} & i > 0 \\ LBP_{P,Rad} & i = 0 \end{cases} \tag{5.14}$$

and $LB(x)$ is the binary number as defined in Eq.(5.9), and $ROR(x,i)$ is the circular rotation operator which circularly shifts the P bit binary number x i times to the right. For rotation invariant $LBP_{8,1}^{ri}$, number of codes is 36 individual patterns (as shown in Figure 5.4). It was observed that certain patterns seem to be fundamental properties of texture, providing the vast majority of patterns, sometimes over 90% [65]. These patterns are called uniform. A local binary pattern is called uniform if the binary pattern contains at most two bitwise transitions from zero to one or one to zero when the bit pattern is considered circular. Table 5.1 shows examples of uniform and non uniform binary patterns. In Figure 5.4, patterns 0 to 8 correspond to uniform binary patterns. Pattern 0 represents a bright spot and pattern 8 represents a dark spot. Patterns 1 to 7 represent lines of varying degrees of curvature. The remaining patterns in Figure 5.4, which have no label, are non uniform binary patterns.

To detect whether the patterns are uniform or not, a uniformity measure U is introduced which counts the number of transitions of one to zero (or vice versa):

$$U(LBP_{P,Rad}) = |LB(g_{P-1} - g_c) - LB(g_0 - g_c)| + \sum_{p=1}^{P-1} |LB(g_p - g_c) - LB(g_{p-1} - g_c)| \quad (5.15)$$

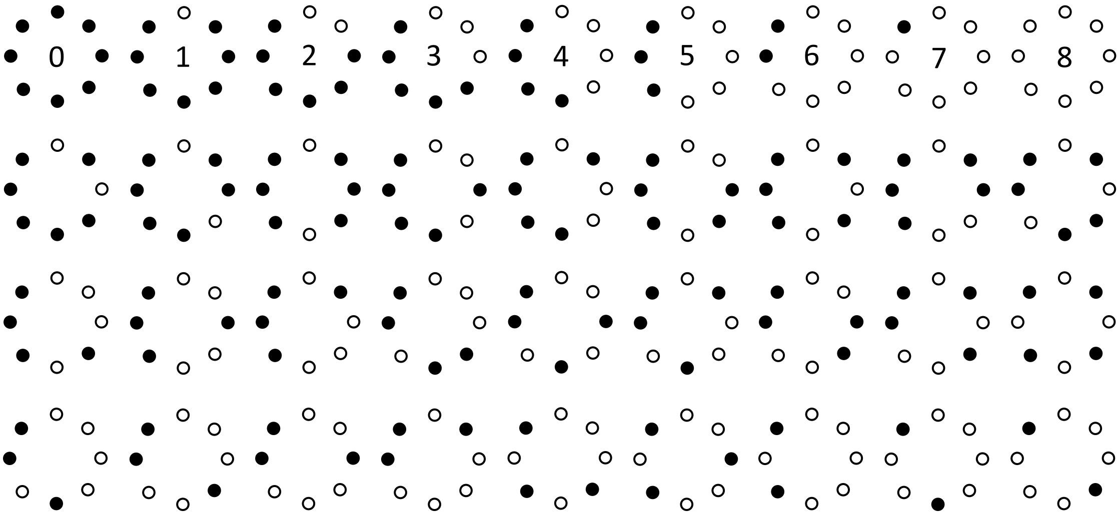


Figure 5.4: The 36 unique rotation invariant binary patterns

The Pattern	Uniform or Non Uniform	Number of Transitions
00000000	Uniform	0
11111111	Uniform	0
00111110	Uniform	2
01000000	Uniform	2
11000001	Uniform	2
01101100	Non Uniform	4
00110101	Non Uniform	6
10101010	Non Uniform	8

Table 5.1: Examples of uniform and non uniform binary patterns

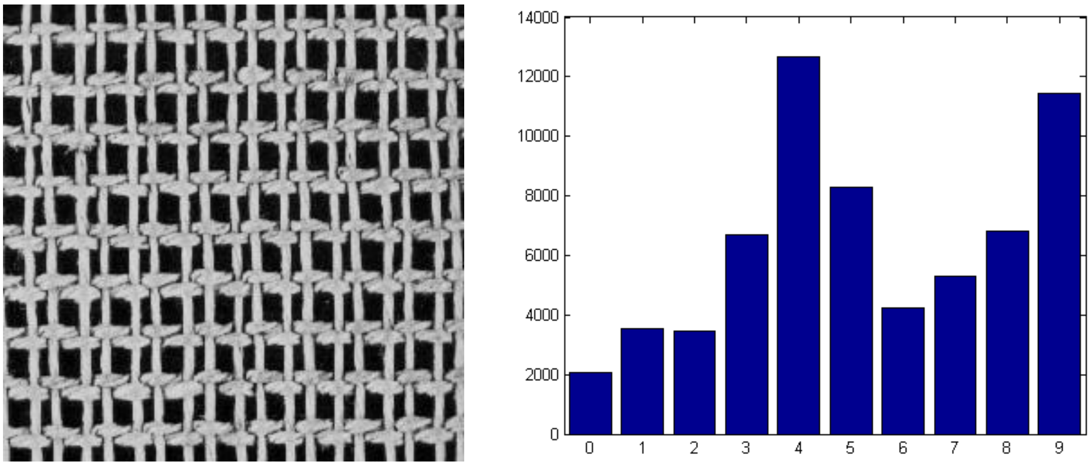
When $U(LBP_{P,Rad}) \leq 2$, the total number of patterns is $P^2 - P + 2$, but the total number of patterns becomes $P+1$ when uniform codes are rotated to their minimum values (using Eq.(5.12)).

The rotation invariant uniform local binary pattern ($LBP_{P,Rad}^{riu2}$) code for any uniform pattern is calculated by counting ones in the binary number, and the non uniform patterns are all set to the same code value:

$$LBP_{P,Rad}^{riu2} = \begin{cases} \sum_{p=0}^{P-1} LB(g_p - g_c) & U(LBP_{P,Rad}) \leq 2 \\ P+1 & otherwise \end{cases} \tag{5.16}$$

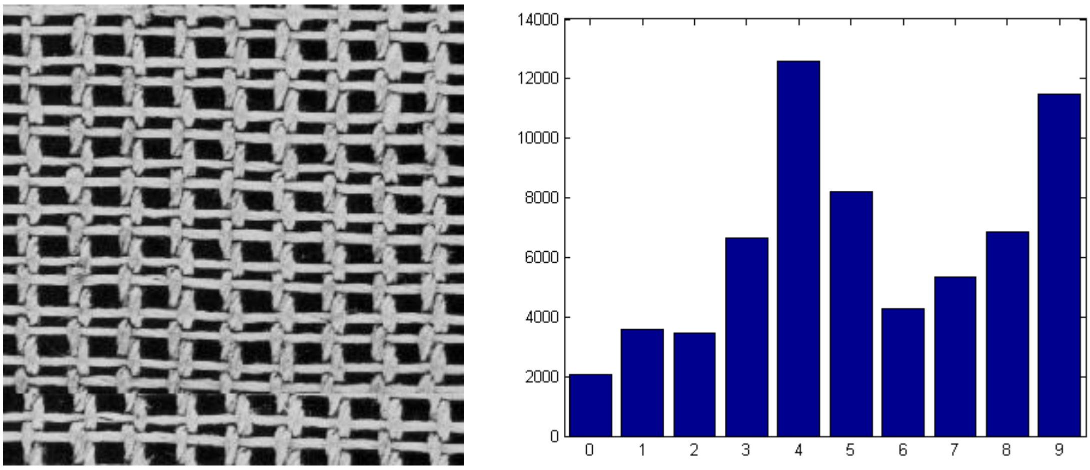
The uniform LBP approach then determines a histogram of the codes derived for an entire image and this histogram describes the texture.

To evaluate the effect of the uniform LBP operator on rotated texture, it is applied to french canvas texture and its shifted and rotated version [54]. The results for a shifted and rotated version of french canvas texture are shown in Figure 5.5. The description of the original texture is given in Figure 5.5-b and its shifted and rotated version is in Figure 5.5-d. Visually there is no difference between these histograms but there is actually some slight difference, of less than 100 count values (as shown in Figure 5.5-e), which is considerably smaller than the count values (14000) exposed by the uniform LBP technique.



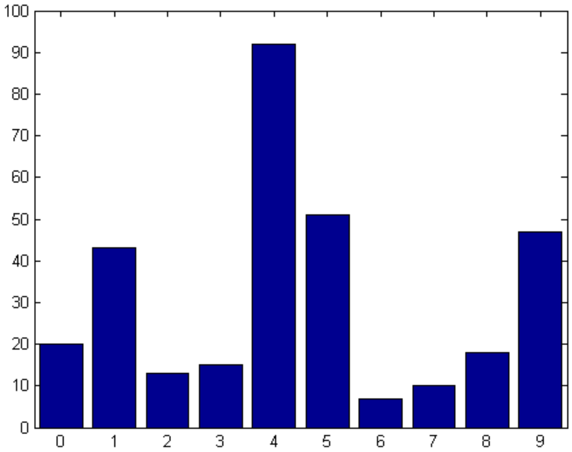
(a) french canvas

(b) histogram of uniform LBP codes for (a)



(c) shifted and rotated french canvas

(d) histogram of uniform LBP codes for (c)



(e) the difference between the two histograms

Figure 5.5: Uniform Local Binary Pattern Histograms

5.2 The Proposed Recognition Technique

In this section, the proposed ear recognition technique is discussed in detail. There are two main steps for our technique: the first step is to describe the ear as a feature vector, and the second step is to compare between these feature vectors.

5.2.1 Ear Description

To describe the ear with a feature vector we apply the representation method proposed in [67], but instead of using Gabor wavelet filters, we use banana wavelet filters. The structure of the ear mainly contains features similar to those of banana wavelets. Any ear image can therefore be described as a histogram sequence by the following procedure (Figure 5.6):

- The ear is extracted manually from a head profile image. The size of extracted ear images is different. This chapter is largely concerned with ear recognition rather than ear detection. The ear is therefore extracted manually in order to minimize the effect of detection on the recognition performance. Also the ear is extracted automatically in a separate experiment to obtain the result for cases, where the whole process is considered automatic.

- The magnitude of the banana filter responses $M^{\mathbf{b}}$ is calculated by taking the magnitude of convolving banana wavelet filters $B^{\mathbf{b}}$ for all possible $\mathbf{b}$ (8 filters are selected by using ANOVA, this will be discussed in section 3) with the extracted ear image I:

  $$M^{\mathbf{b}} = \left|\left(B^{\mathbf{b}} * I\right)\left(x_0\right)\right| \qquad (5.17)$$

  where x_0 and $*$ are the location of a pixel in I, and the convolution operator respectively.

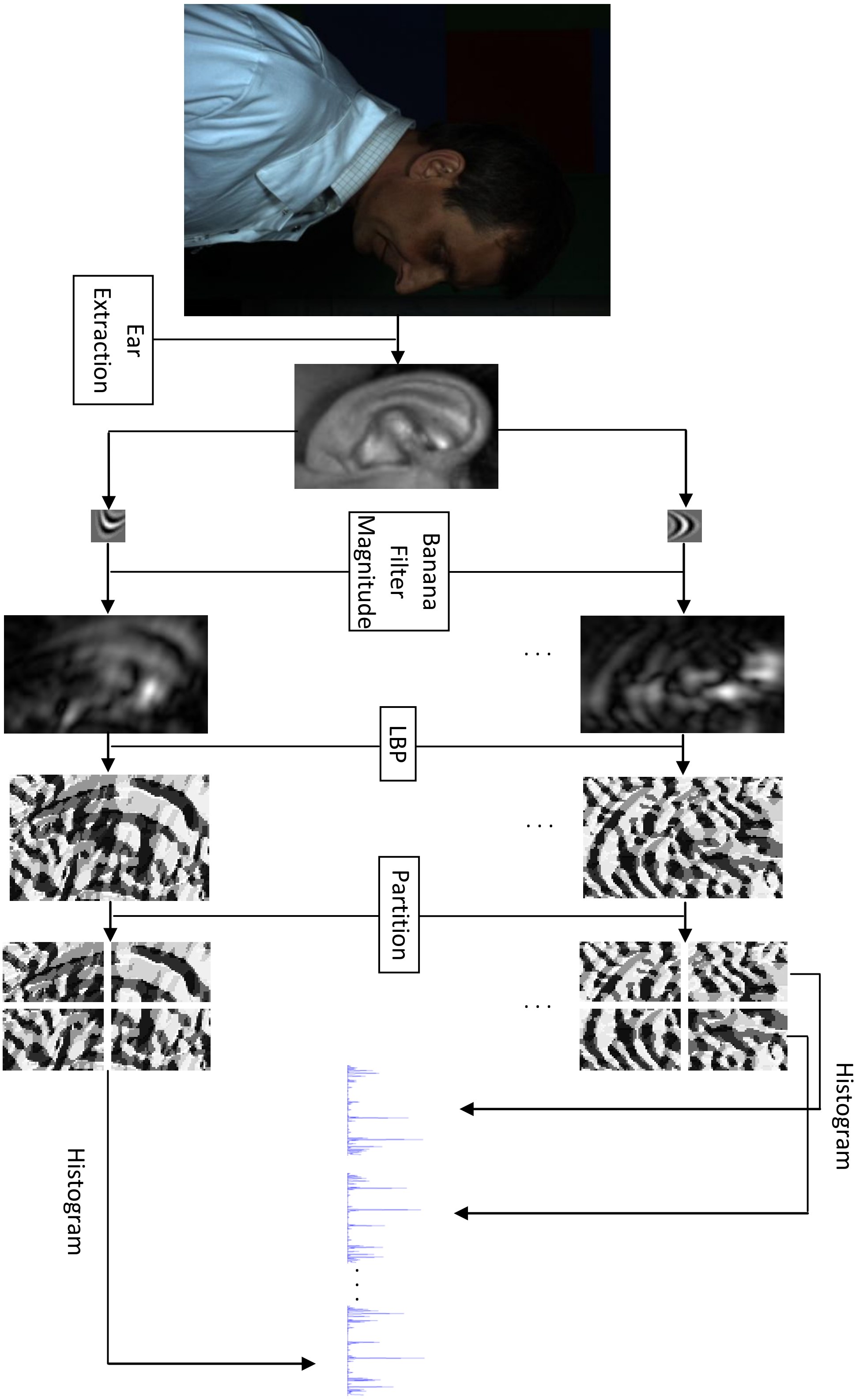


Figure 5.6: Our ear description algorithm

- Local binary pattern $LBP^{\mathbf{b}}$ is calculated for each $M^{\mathbf{b}}$ which has invariant characteristics.
- Each $LBP^{\mathbf{b}}$ is divided into four non-overlapping regions. These quadrants (regions) expose basic structures in the ear.
- Histogram $H_{r,n}$ calculated by applying filter n (n is from 1 to 8) on region r (r is from 1 to 4) is used to produce the ear features.
- All of the calculated histograms are concatenated into a single histogram sequence HS to describe the given ear image as a feature vector:

$$HS = \left(H_{1,1}, H_{2,1}, H_{3,1}, H_{4,1}, H_{1,2}, \ldots, H_{4,8}\right) \tag{5.18}$$

5.2.2 Ear Comparison

Histogram intersection HI [68] is used as a similarity measure between two histograms. We normalize each histogram to produce a probability distribution function. The histogram intersection can then be calculated as:

$$HI\left(H^1, H^2\right) = \sum_{i=0}^{L} \min\left(\frac{h_i^1}{\sum_{j=0}^{L} h_j^1}, \frac{h_i^2}{\sum_{j=0}^{L} h_j^2}\right) \tag{5.19}$$

where h^1 and h^2 are two histograms, and L is the number of bins in the histogram (in this case L equals 255). Once the ear is described as histogram sequence, the similarity S between two histogram sequences is measured by calculating HI between them:

$$S\left(HS^1, HS^2\right) = \sum_{r=1}^{4} \sum_{n=1}^{8} HI\left(H_{r,n}^1, H_{r,n}^2\right) \tag{5.20}$$

To select the matched ear with the test one, the similarities between the histogram sequence of the test ear and all of the other histogram sequences are measured, and the ear with maximum value for S is selected as the correct one.

5.3 Filter Selection

The ANOVA based feature selection algorithm is applied to select the best filters from a bank of banana filters used in the recognition process. To select the best filters, first a bank of banana filters (720 filters) is convolved with XM2VTS database and the magnitudes of the filters' responses are computed(XM2VTS database is used for filter selection as it is constrained and has fewer variations in lighting, rotation and etc.). Then, LBP algorithm is applied on each magnitude image to produce a feature image which is then divided into four parts (quadrants). The histogram is calculated for each part and the four histograms are concatenated for each filter. After that, ANOVA algorithm is applied to compute the $\mathbf{F\text{-}ratio}$ for each filter [63]. Then, the filters are ranked in descending order according to their $\mathbf{F\text{-}ratio}$ and the first 160 filters are selected as $\mathbf{F\text{-}ratio} \geq \mathbf{F}_{(df_b=62, df_w=189, \alpha=0.05)}$ (63 subjects with 252 images). For classification, large values of the $\mathbf{F\text{-}ratio}$ do not necessarily indicate useful features. They only indicate a well spread feature space. Therefore, the best 8 filters are selected from the chosen filters according to the algorithm in Figure 5.7. The selected filters are shown in Figure 5.8.

5.4 Ear Recognition

Having selected the best filters, the recognition technique proposed here is then applied to XM2VTS face profile database to obtain a recognition rate of 100% in the case of extracting the ear manually, and 99.6% in the case of extracting the ear automatically (which indicates that our automatic ear detection technique is accurate and similar to manual ear enrolment). Table 5.2 compares between the

```
 1: set $All\_f = \{BF^{(1)}, BF^{(2)}, ..., BF^{(i)}, ..., BF^{(160)}\}$
 2: for each $BF^{(i)} \in All\_f$
 3:     set $SF^{(i)} = BF^{(i)}$
 4:     repeat
 5:         set $Remain\_f = All\_f - SF^{(i)}$
 6:         for each $BF^{(k)} \in Remain\_f$
 7:             compute $CRR(SF^{(i)} \cup BF^{(k)})$
 8:             if $CRR(SF^{(i)} \cup BF^{(k)}) \geq CRR(SF^{(i)})$
 9:                 set $SF^{(i)} = SF^{(i)} \cup BF^{(k)}$
10:                 goto step 13
11:             end if
12:         end for
13:     until $size(SF^{(i)}) = 8 \; or \; CRR(SF^{(i)}) = 100$
14: end for
15: compute $max\_CRR$
16: let $max\_SF$ be the set containing all of $SF^{(i)}$ $where \; CRR(SF^{(i)}) = max\_CRR$
17: if $size(max\_SF) = 1$
18:     choose this set as the set of the selected filters
19: else
20:     for each $SF^{(m)} \in max\_SF$
21:         compute $d'^{(m)}$ for $SF^{(m)}$ (using Eq. (5.21))
22:     end for
23:     choose the set with maximum $d'$ as the set of the selected filters
24: end if
```

Figure 5.7: Filter selection algorithm. Here, All_f is the set of all 160 banana filters, BF is the banana filter, $Remain\_f$ is the subset of the filters that are not selected, CRR is the correct recognition rate, SF is the subset of the selected filters, and d' is the decidability measure.

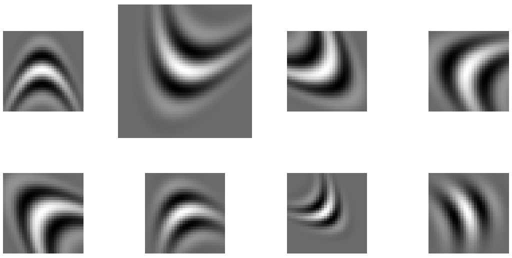


Figure 5.8: The selected banana filters

performance of our recognition technique and other recognition techniques (applied to the same XM2VTS dataset). As illustrated in the table, our recognition technique offers the highest recognition rate.

A recognition rate of 99.5% is also achieved by applying the algorithm proposed here to the SOTON ear dataset 2. SOTON dataset 2 is selected from SOTON database [61] and contains 548 face profile images from 137 subjects with four images per person. Figure 5.9 shows two subjects (with four images per subject) from SOTON dataset 2. As illustrated in the figure there are a lot of variations of ear orientation, size, and lighting condition.

Technique	Recognition Rate
Force field transformation Hurley et al. [2]	99.2%
Model based Arbab-Zavar et al. [34]	91.5%
Log-Gabor filter and model based Arbab-Zavar and Nixon [35]	97.4%
Scale Invariant Feature Transform (SIFT) Bustard and Nixon [37]	96%
Our recognition technique (manual enrolment)	100%
Our recognition technique (automatic enrolment)	99.6%

Table 5.2: Comparison of different recognition techniques and our recognition technique (based on XM2VTS database)

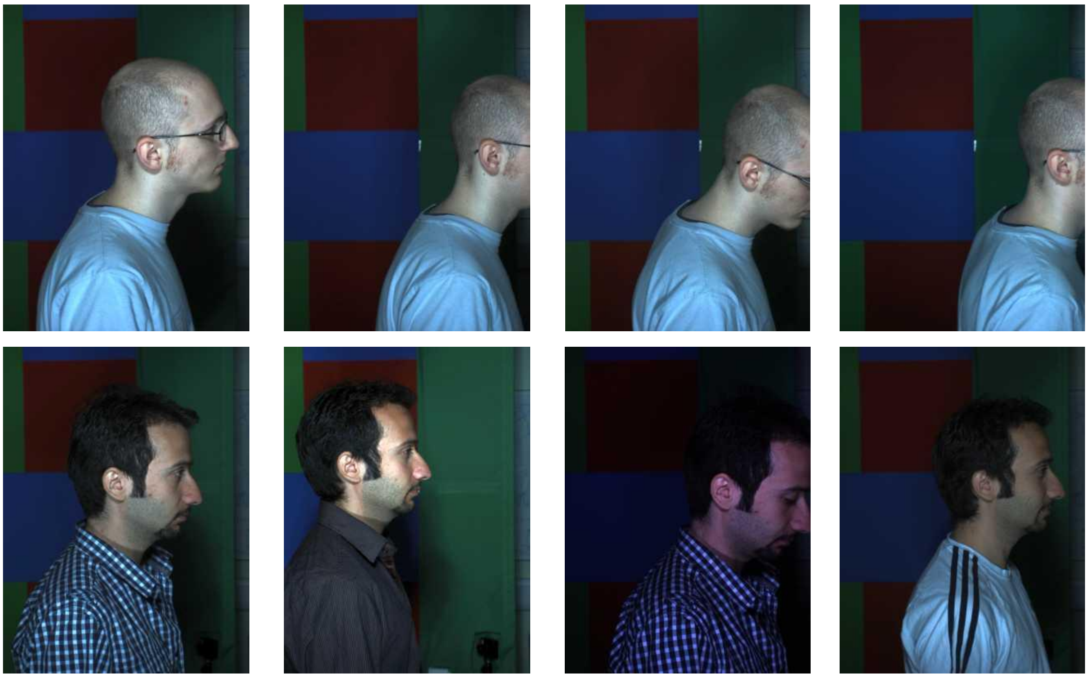


Figure 5.9: Samples from SOTON dataset 2

In these experimental results we also show the difference in recognition rate in the following cases:

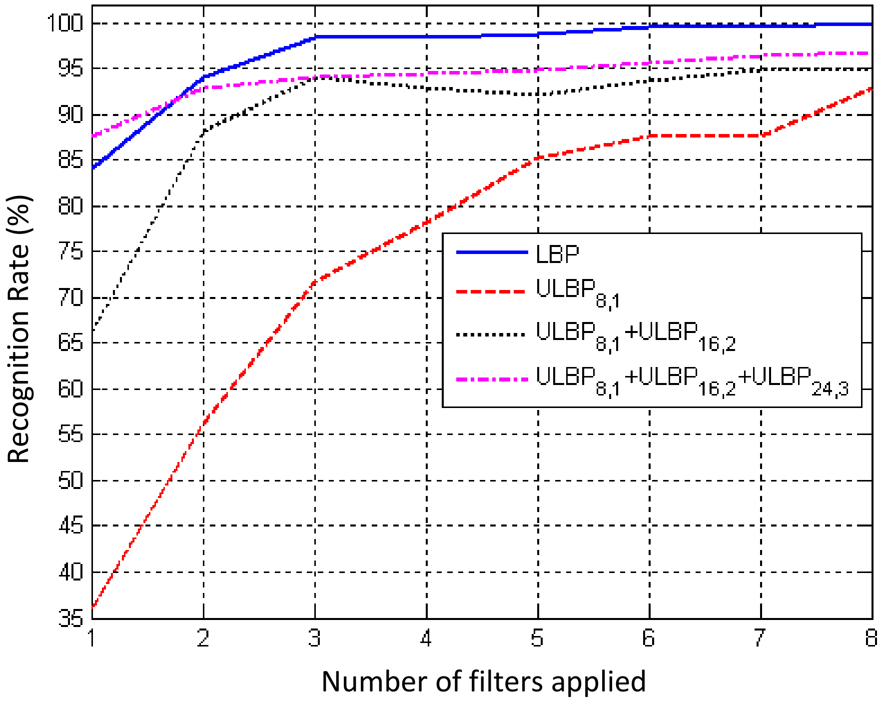
i) between the LBP and multi-resolution ULBP algorithms,
ii) between two different similarity measures(histogram intersection and absolute difference),
iii) between banana and Gabor filters,
iv) between top and left occlusions and
v) between banana and Gabor filters in the presence of Gaussian noise.

5.4.1 LBP and ULBP

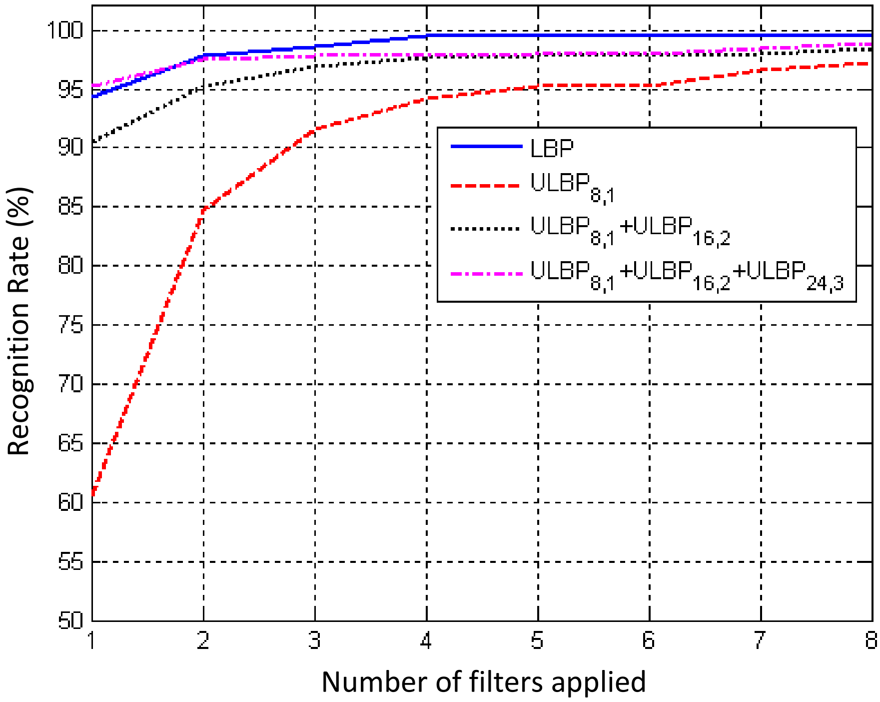
We have applied the uniform LBP for ear recognition. However recognition rate was not as high as that for basic LBP.Table 5.3 shows the difference in recognition rate between the LBP and multi-resolution ULBP ($ULBP_{8,1}$+$ULBP_{16,2}$+$ULBP_{24,3}$) to XM2VTS database and SOTON dataset 2. There is an additional comparison between the LBP and ULBP illustrated in Figure 5.10, which shows the recognition rate versus the number of banana filters applied to different databases when applying LBP and multi-resolution ULBP techniques. As shown in Figure 5.10, in all the stages of applying the filters the recognition rate of applying LBP is better than the recognition rate obtained by the multi-resolution ULBP technique. A confusion matrix is computed to assess the accuracy of the ear recognition (as shown in Figure 5.11). For a classifier to have good accuracy, ideally most of the matched samples would be represented along the diagonal of the confusion matrix.

Database	Recognition rate		
		Intersection	Difference
XM2VTS database (manual enrolment)	LBP (Banana filters)	100%	99.2%
	LBP (Gabor filters)	98.8%	98.8%
	$ULBP_{8,1}$ (Banana filters)	92.9%	88.9%
	$ULBP_{8,1}$+$ULBP_{16,2}$ (Banana filters)	94.8%	94.4%
	$ULBP_{8,1}$+$ULBP_{16,2}$+$ULBP_{24,3}$ (Banana filters)	96.8%	94.4%
XM2VTS database (automatic enrolment)	LBP (Banana filters)	99.6%	98.0%
	LBP (Gabor filters)	96.4%	96.0%
	$ULBP_{8,1}$ (Banana filters)	89.3%	82.9%
	$ULBP_{8,1}$+$ULBP_{16,2}$ (Banana filters)	93.3%	91.3%
	$ULBP_{8,1}$+$ULBP_{16,2}$+$ULBP_{24,3}$ (Banana filters)	95.2%	91.7%
SOTON dataset 2	LBP (Banana filters)	99.5%	99.3%
	LBP (Gabor filters)	98.9%	99.1%
	$ULBP_{8,1}$ (Banana filters)	97.3%	96.2%
	$ULBP_{8,1}$+$ULBP_{16,2}$ (Banana filters)	98.4%	98.2%
	$ULBP_{8,1}$+$ULBP_{16,2}$+$ULBP_{24,3}$ (Banana filters)	98.9%	98.4%

Table 5.3: Comparison of recognition rate for different techniques

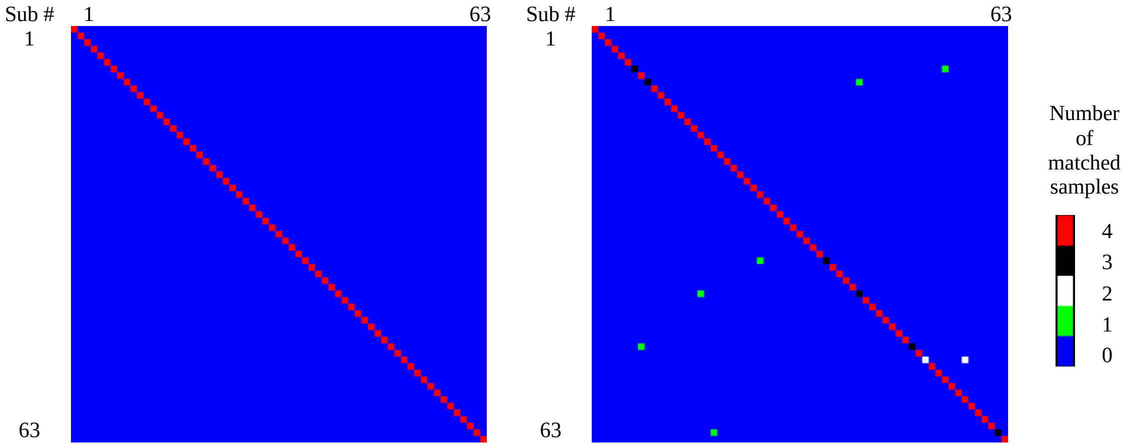


(a) XM2VTS database (manual enrolment)



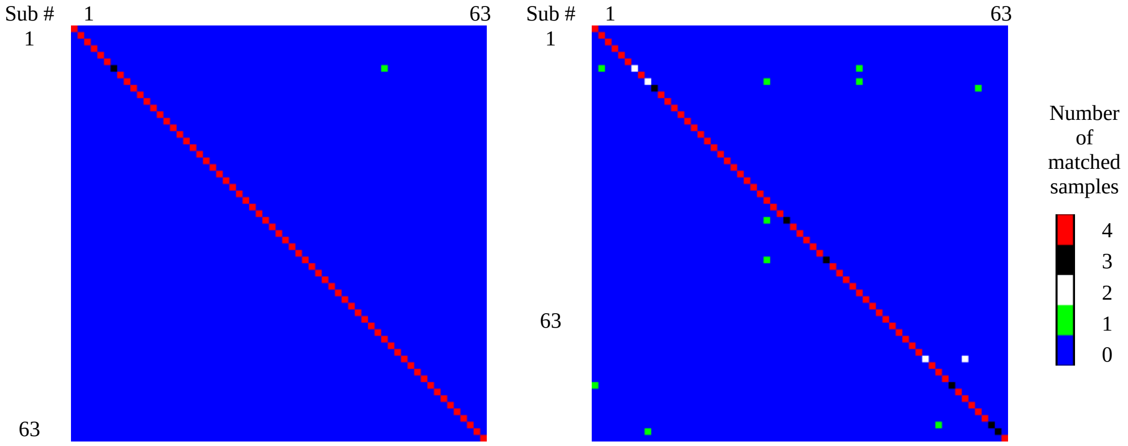
(b) SOTON dataset 2

Figure 5.10: Recognition rate versus number of filters applied to XM2VTS database and SOTON dataset 2 using LBP and ULBP



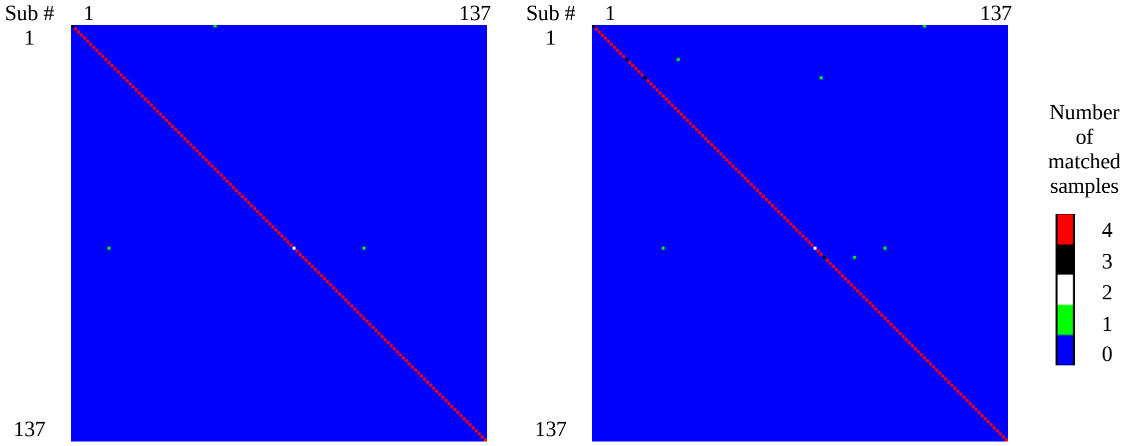
(a) XM2VTS database with manual enrolment (LBP)

(b) XM2VTS database with manual enrolment (ULBP)



(c) XM2VTS database with automatic enrolment (LBP)

(d) XM2VTS database with automatic enrolment (ULBP)



(e) SOTON dataset 2 (LBP)

(f) SOTON dataset 2 (ULBP)

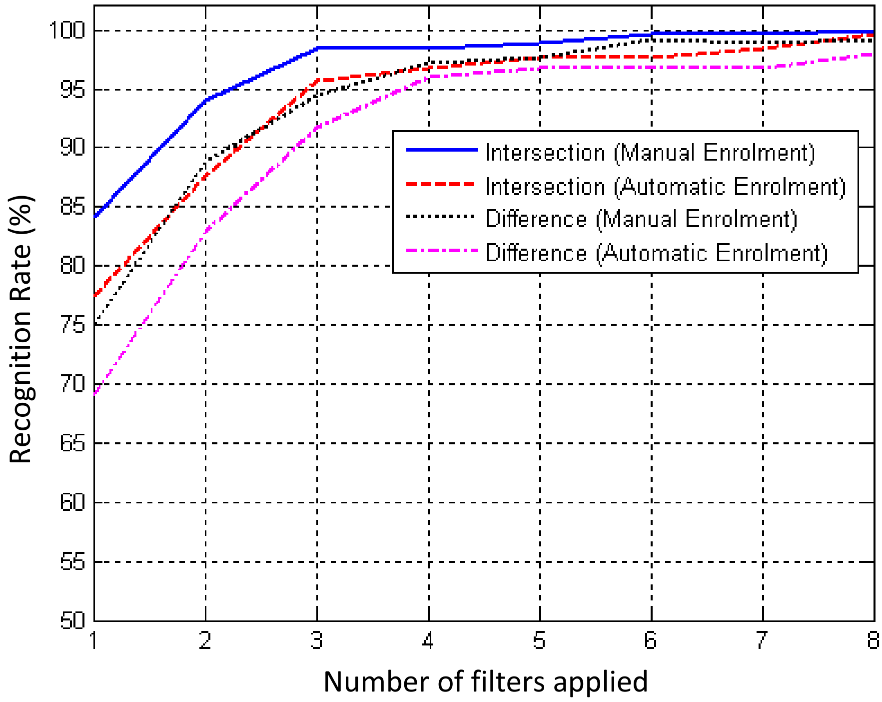
Figure 5.11: Confusion matrix of ear recognition for XM2VTS database and SOTON dataset 2 using LBP and ULBP

5.4.2 Similarity Measure

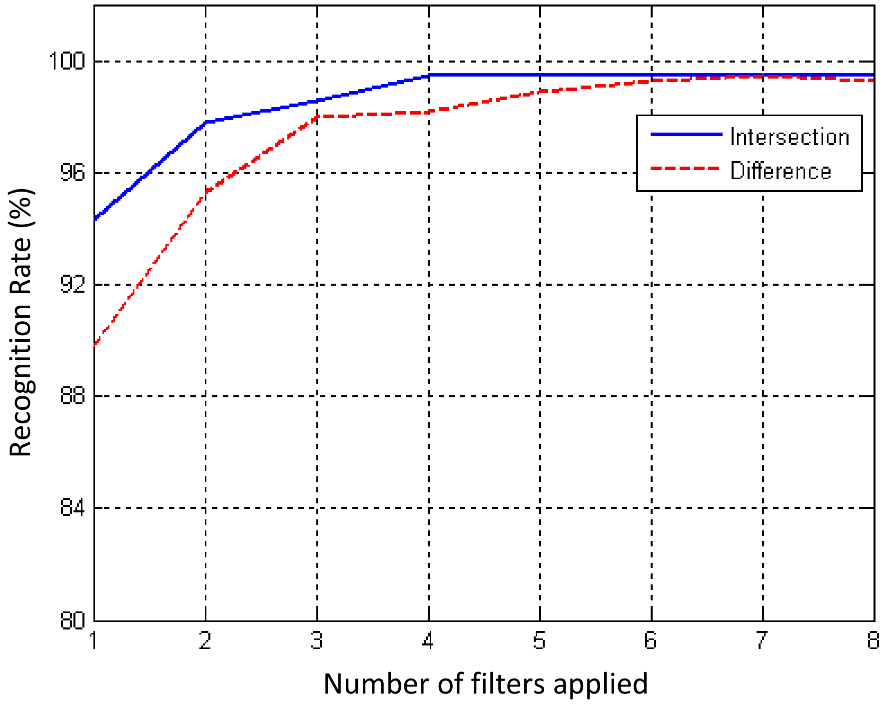
In Table 5.3, there is also a comparison between using histogram intersection and difference as a similarity measure between two histograms. As shown from the table, histogram intersection results in better recognition rate than histogram difference similarity measure. Figure 5.12 shows the performance of our recognition technique versus the number of banana filters applied to different databases when using histogram intersection or difference as a similarity measure. As illustrated in Figure 5.12, when applying 3 banana filters to XM2VTS database (manual enrolment and histogram intersection similarity measure), the recognition rate increases slightly from 98.4% to 100% (when applying 8 banana filters). However, when applying 4 banana filters to SOTON dataset 2 the recognition rate is 99.5% which is the same result when applying 8 banana filters.Figure 5.13 shows the confusion matrix of ear recognition for XM2VTS database and SOTON dataset 2 using intersection and difference.

5.4.3 Banana and Gabor Filters

Our hypothesis in this thesis is that the banana filters can describe the curved structures better than Gabor filters. Numerical results presented in this chapter confirm our hypothesis. Banana and Gabor filters are applied to ear images to the XM2VTS face profile database and SOTON dataset 2 with the same size, orientation and frequency (except that curvature parameter was set to zero for Gabor filters). The recognition rate for XM2VTS face profile database obtained by Gabor filters is 98.8% with manual ear detection, and 96.4% with automated ear detection. A recognition rate of 98.9% is also achieved by applying the Gabor filters to SOTON ear dataset 2.Figure 5.14 shows the recognition rate versus the number of banana and Gabor filters applied to different databases. As expected the recognition rate rises by increasing the number of banana filters or Gabor filters.Figure 5.15 illustrates the accuracy of the ear recognition by creating the confusion matrix of the recognition for different databases using banana and Gabor filters.

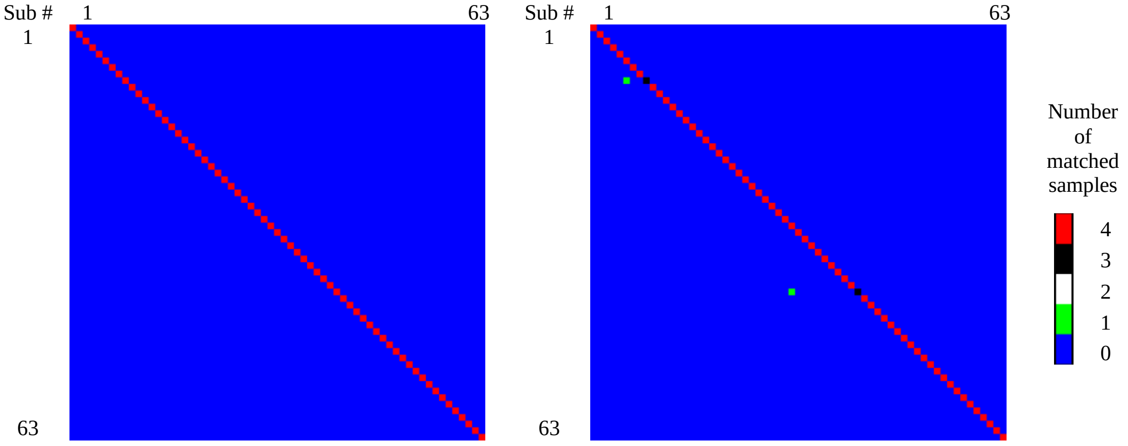


(a) XM2VTS database



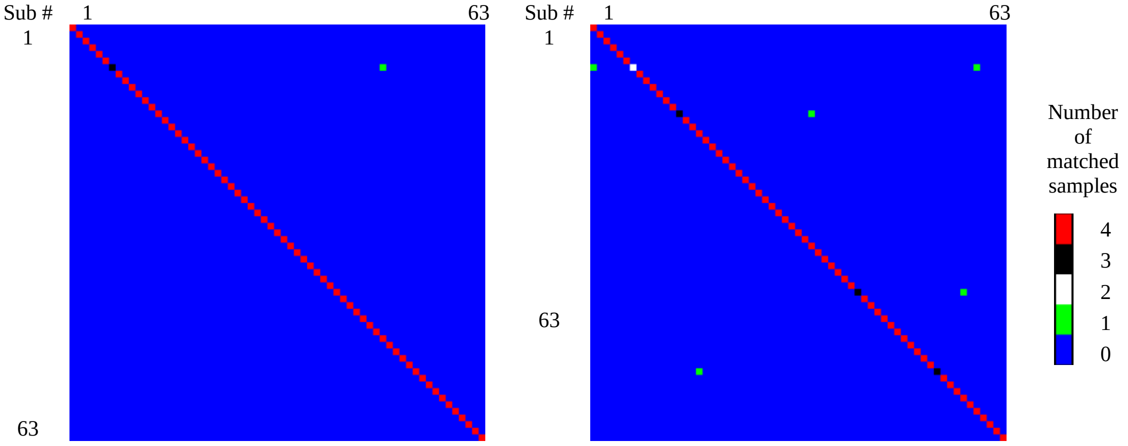
(b) SOTON dataset 2

Figure 5.12: Recognition rate versus number of filters applied to XM2VTS database and SOTON dataset 2 using Intersection and difference



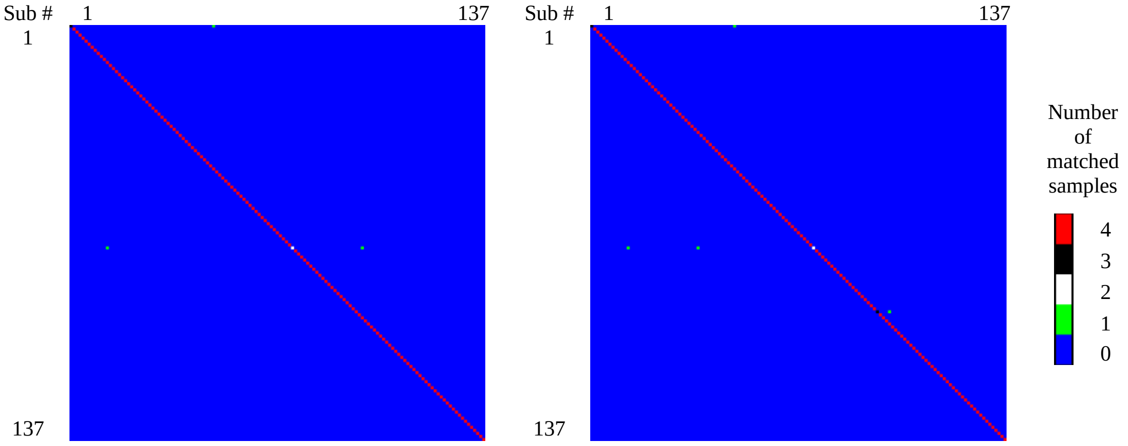
(a) XM2VTS database with manual enrolment (intersection)

(b) XM2VTS database with manual enrolment (difference)



(c) XM2VTS database with automatic enrolment (intersection)

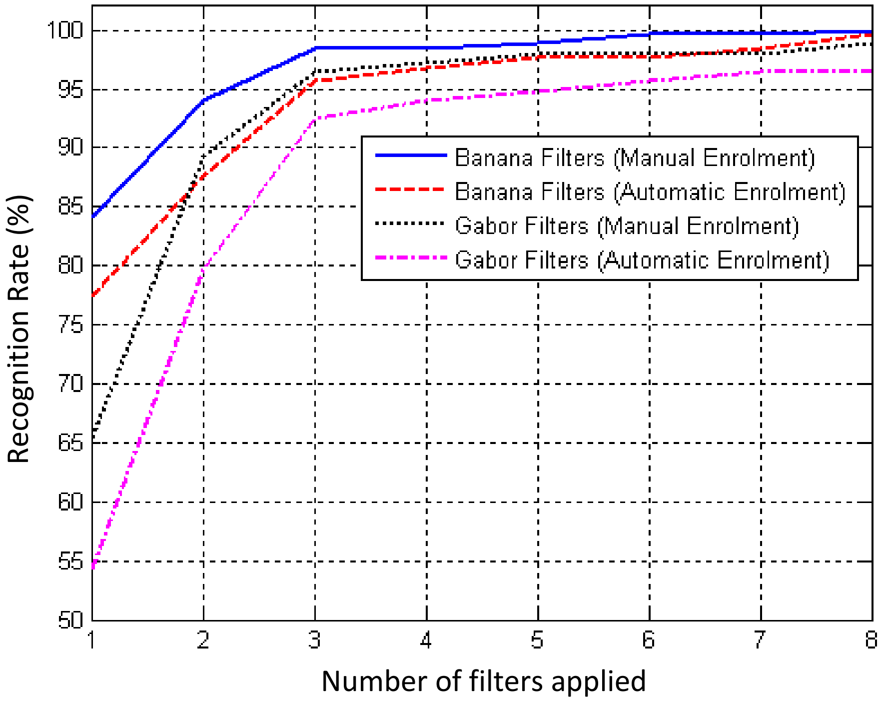
(d) XM2VTS database with automatic enrolment (difference)



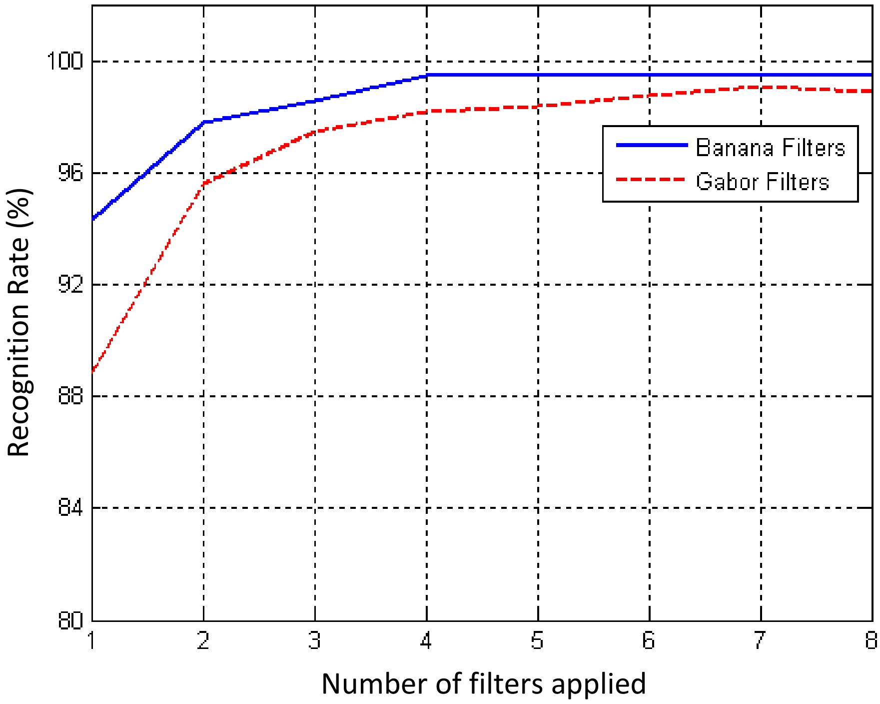
(e) SOTON dataset 2 (intersection)

(f) SOTON dataset 2 (difference)

Figure 5.13:Confusion matrix of ear recognition for XM2VTS database and SOTON dataset 2 using intersection and difference

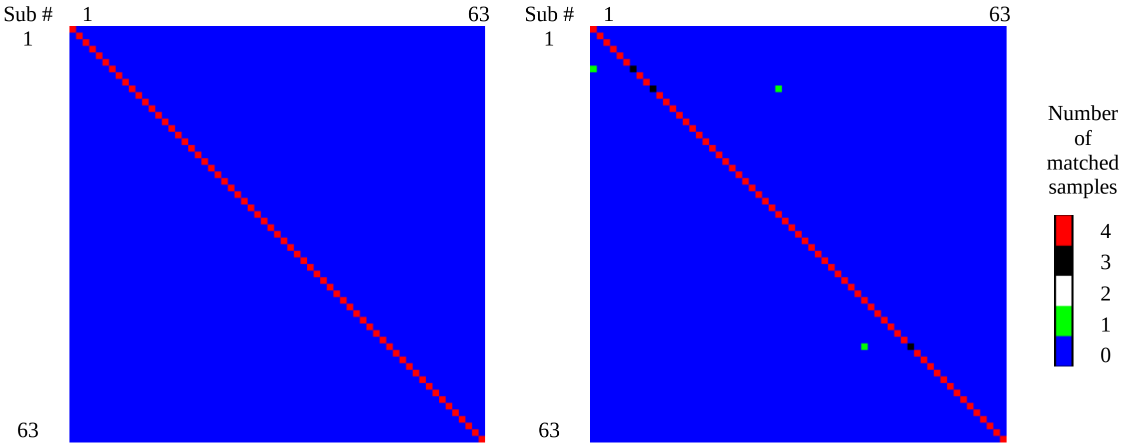


(a) XM2VTS database



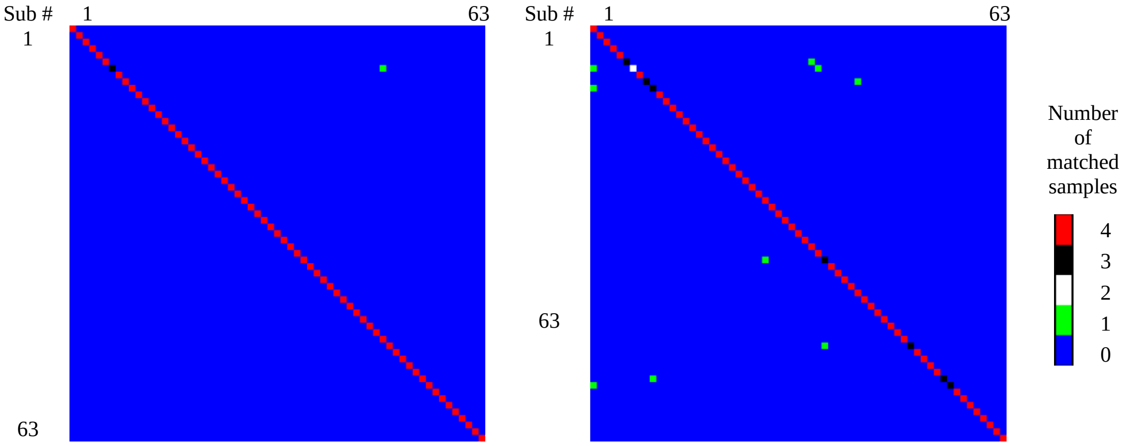
(b) SOTON dataset 2

Figure 5.14: Recognition rate versus number of filters applied to XM2VTS database and SOTON dataset 2 using banana and Gabor filters



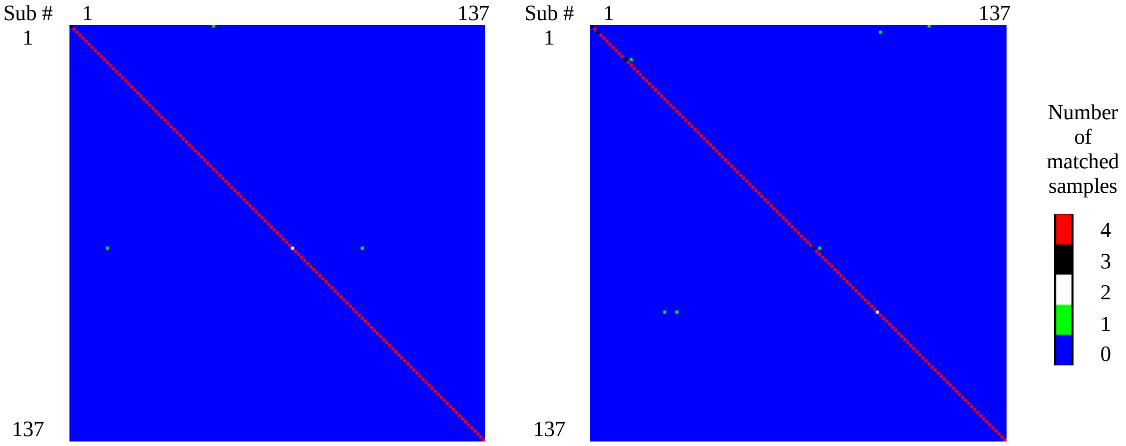
(a) XM2VTS database with manual enrolment (banana)

(b) XM2VTS database with manual enrolment (Gabor)



(c) XM2VTS database with automatic enrolment (banana)

(d) XM2VTS database with automatic enrolment (Gabor)



(e) SOTON dataset 2 (banana)

(f) SOTON dataset 2 (Gabor)

Figure 5.15:Confusion matrix of ear recognition for XM2VTS database and SOTON dataset 2 using banana and Gabor filters

5.4.4 Occlusion

To show the effect of occlusion on the performance of our recognition technique, occlusion from top and left are applied to XM2VTS database as discussed before in chapter 3. The results of the occlusion analysis are provided in Figure 5.16, which illustrates the recognition rate of the XM2VTS database against increasing synthetic occlusion. As shown in the figure, occlusion from left obtains better recognition rate than occlusion from top until 50% but after that the performance of occlusion from top becomes more successful than occlusion from the left.

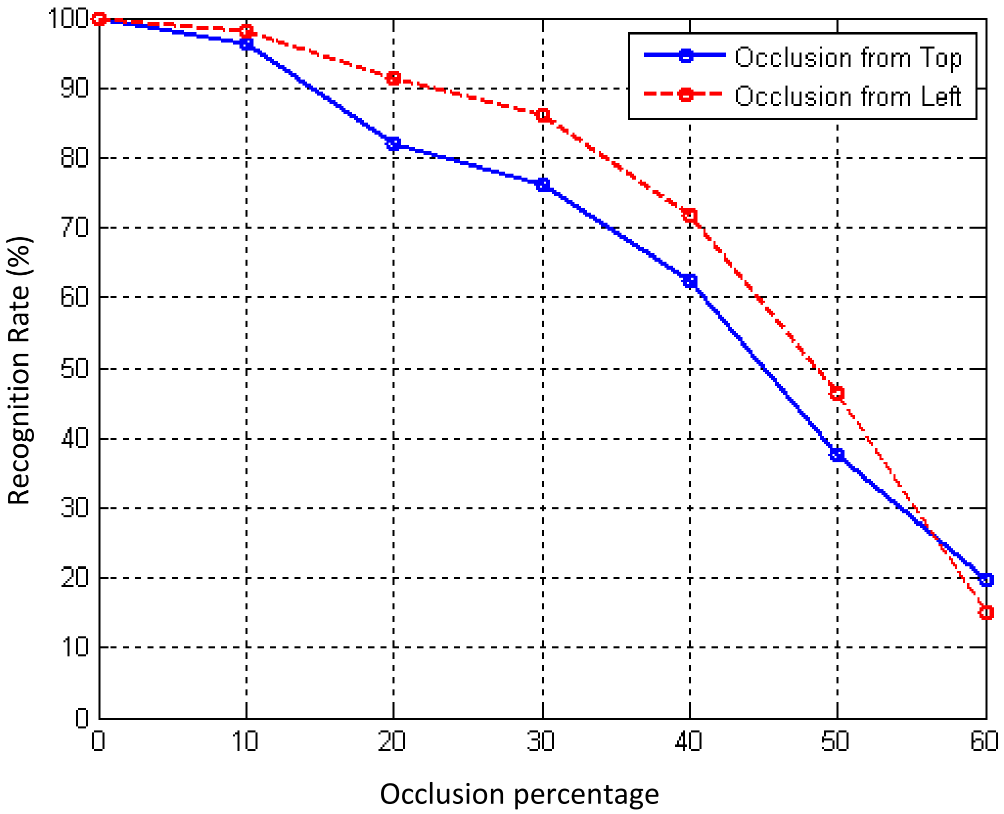


Figure 5.16: Recognition rate for ears with occlusion from top and left for XM2VTS database

5.4.5 Noise

An additive zero mean Gaussian noise, with various standard deviations, is added to the XM2VTS database to evaluate the performance of our recognition technique in the presence of noise. The results of the noise analysis are provided in Figure 5.17. We note that for noise-free images the recognition rate is 100%. As expected, the recognition rate drops with increasing noise.Figure 5.17 also shows a comparison between the performance of the banana filters and Gabor filters in the presence of noise. As the noise increases, the advantages associated with using curved banana filters become less significant and Gabor filters are more successful than banana filters. However, that such noise is highly unlikely in a realistic scenario.

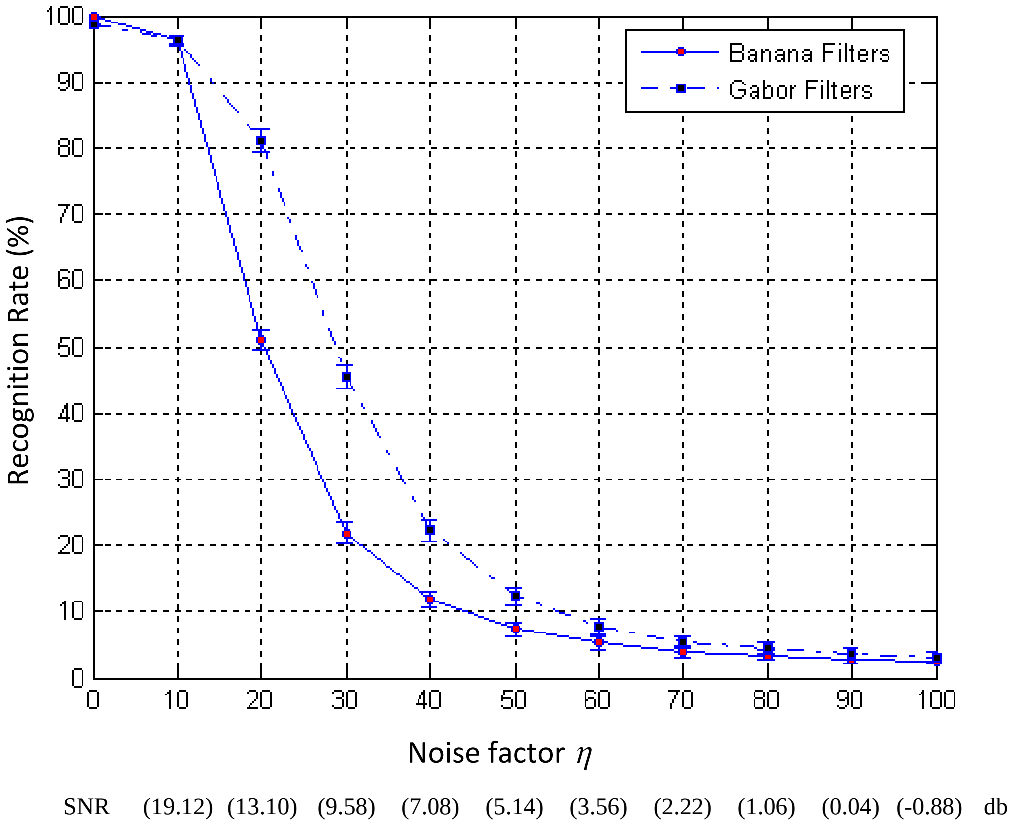


Figure 5.17: Recognition rate for banana filters and Gabor filters in the presence of noise for XM2VTS database

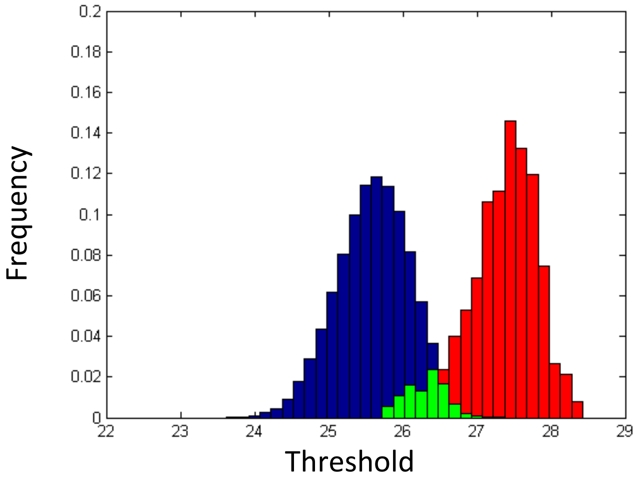
5.5 Verification Performance

The verification performance of the proposed technique is measured in terms of the relationship between the intra-class and inter-class distributions of the histograms intersections (see Figure 5.18). In this figure, the distribution on the left in the graph shows the intra-class distribution which different ear images from different subjects are compared. But when different ear images from the same subject are compared, the inter-class distribution on the right in the graph is the result. Any overlap between the intra-class and inter-class distributions affects on the verification performance, so these two distributions should be well separated.

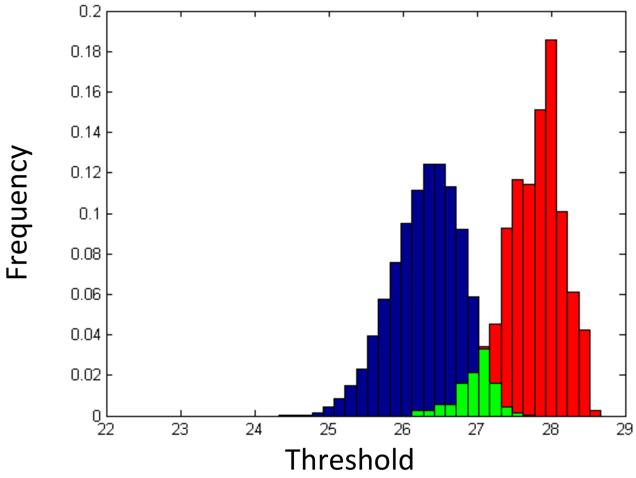
A decidability index d' is a good measure of how well the intra-class and inter-class distributions are separated [69]:

$$d' = \frac{|\mu_1 - \mu_2|}{\sqrt{\left(\sigma_1^2 + \sigma_2^2\right)/2}} \tag{5.21}$$

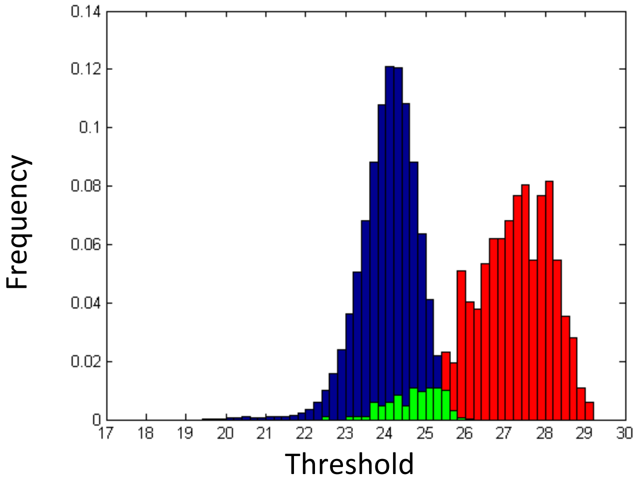
where μ_1 and σ_1 are the mean and standard deviation of the inter-class distribution, respectively, and μ_2 and σ_2 are the mean and standard deviation of the intra-class distribution, respectively. This is similar to Eq. (5.8) but the $\mathbf{F\text{-}ratio}$ reflects more the difference between means, whereas Eq. (5.21) also includes standard deviation (in a manner similar to the Mahalanobis distance).Table 5.4 shows the measured decidability values when applying banana filters and Gabor filters in our proposed technique for the XM2VTS database (for manual and automatic enrolment) and SOTON dataset 2. Figure 5.19 illustrates the decidability versus the number of banana filters and Gabor filters applied for different databases. The results show that the intra-class and inter-class distributions for our technique by applying banana filters are better separated than those by applying Gabor filters. For a database, the higher the decidability measure, the better the algorithm performance is.



(a) XM2VTS database (Manual Enrolment)

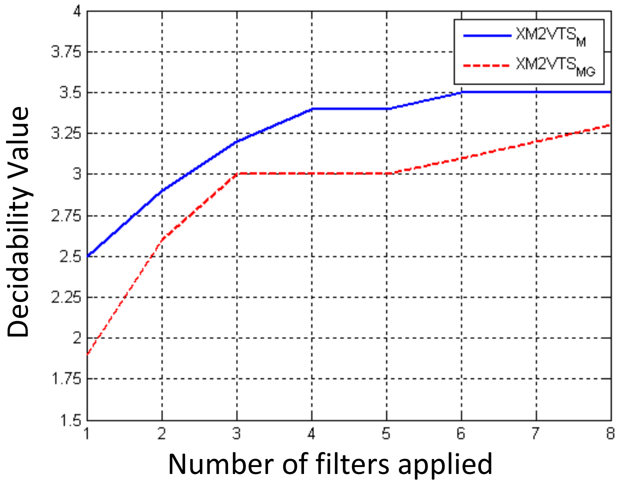


(b) XM2VTS database (Automatic Enrolment)

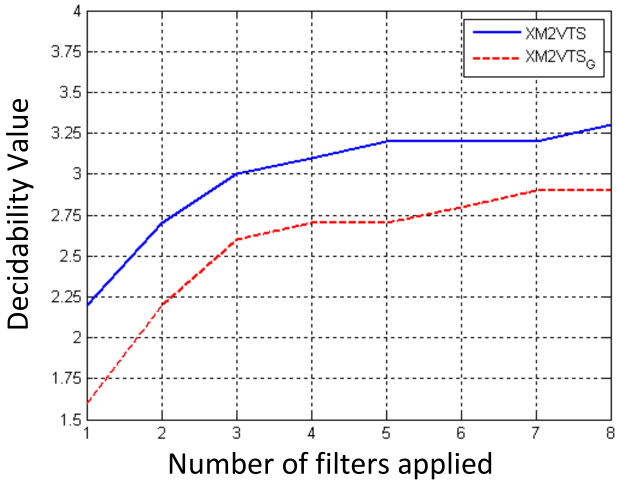


(c) SOTON dataset 2

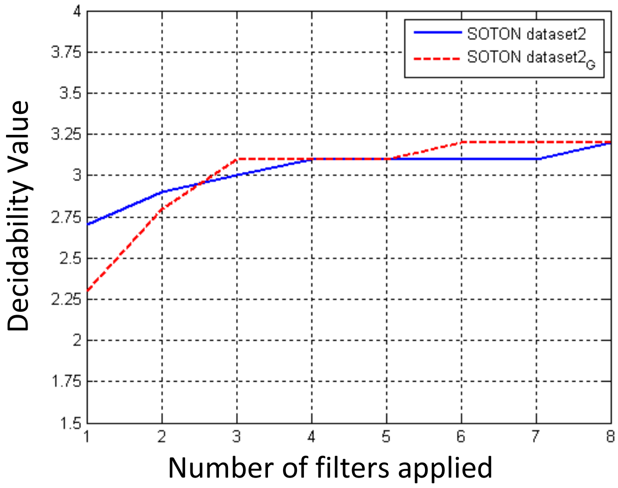
Figure 5.18: Intra-class and inter-class distributions for different databases.



(a) Decidability for XM2VTS database (Manual Enrolment)



(b) Decidability for XM2VTS database (Automatic Enrolment)



(c) Decidability for SOTON dataset 2

Figure 5.19: Decidability measure versus number of filters applied for different databases.

	μ_1	μ_2	σ_1	σ_2	d'
XM2VTS (automatic enrolment - Banana)	27.83	26.40	0.41	0.47	3.3
XM2VTS (automatic enrolment - Gabor)	27.21	25.65	0.53	0.53	2.9
XM2VTS (manual enrolment - Banana)	27.41	25.68	0.50	0.49	3.5
XM2VTS (manual enrolment - Gabor)	26.81	24.81	0.62	0.60	3.3
Dataset 2 (Banana)	27.15	24.17	1.12	0.72	3.2
Dataset 2 (Gabor)	26.74	23.06	1.39	0.82	3.2

Table 5.4: The measured decidability for different databases

The verification performance is also measured in terms of the receiver operating characteristic (ROC) curve and the equal error rate (EER).The ROC curve describes a relation between the false accept rate (FAR) and the false reject rate (FRR) at different thresholds. False accept occurs when the system accepts an impostor ear and false reject occurs when the system refuses a true ear. The FAR and the FRR are calculated using the following equations [70]:

$$FAR = \frac{N_{FA}}{N_{\mathrm{Im}}} \tag{5.22}$$

$$FRR = \frac{N_{FR}}{N_{Tr}} \tag{5.23}$$

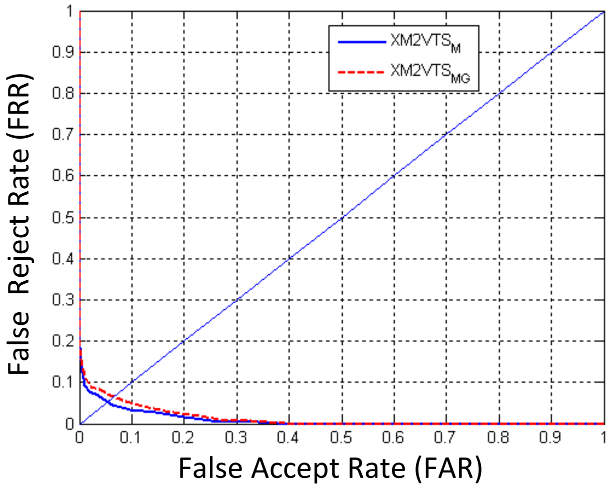
where N_{FA} is number of false accept, N_{FR} is number of false reject, N_{Im} is Number of imposter access, and N_{Tr} is number of true claimant accesses. The threshold can determine the level of required security to the system. If the threshold is set to high value, the system will be high security which the FAR will be low and the FRR will be high. However, when the threshold is set to low value, the system will be low security which the FRR will be low and the FAR will be high.

Figure 5.20 shows the ROC curve of the proposed technique on XM2VTS face profile database and SOTON ear dataset 2. As illustrated in the figure, the verification performance when applying banana filters are better than the verification performance when applying Gabor filters.

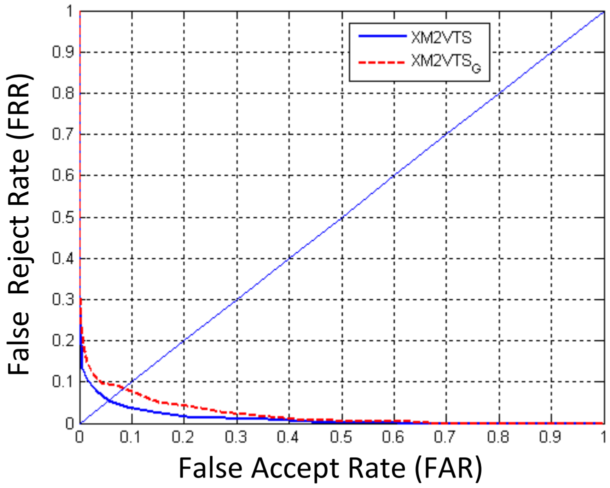
The equal error rate (EER) of the experiments is 5.4% for XM2VTS database (for manual and automatic enrolment) and 5.8% for SOTON ear dataset 2. The EER is the location on the ROC curve (Figure 5.20) where the FAR and the FRR are equal.

5.6 Parameters Sensitivity

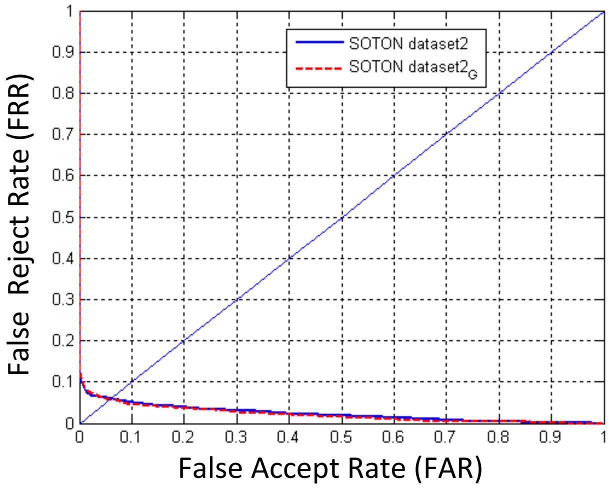
ANOVA technique is used to identify the best banana filter parameters for the recognition process. A sensitivity analysis is carried out to these parameters to understand how the variations in banana filter parameters affect the recognition rate. To measure the sensitivity, each parameter is changed by a specific amount while all the other parameters are kept constant and the recognition result is recorded for each change in the parameter. In experiments, we find that the recognition rate is sensitive to all parameters, especially to f (frequency) parameter (as shown in Table 5.5). Parameters sensitivity is also applied to test the effect of the parameters on verification performance. As expected, the experiments show that f parameter is the most sensitive parameter on the verification performance which the decidability drops to 1.3 and EER increases to 23.7% for the changing in f parameter.



(a) XM2VTS database (manual enrolment)



(b) XM2VTS database (automatic enrolment)



(c) SOTON dataset 2

Figure 5.20: The ROC curve for XM2VTS and SOTON databases

θ		f		c		s		R, C	
change	CRR	change	CRR	change	CRR	change	CRR	change	CRR
$-\frac{2\pi}{16}$	98.4%	-0.36	83.3%	-0.04	99.6%	$\frac{s}{2}$	98.8%	-20	94.8%
$-\frac{\pi}{16}$	98.8%	-0.24	96.0%	-0.02	100%	$\frac{3s}{4}$	99.2%	-10	96.8%
0	100%	-0.12	97.6%	0	100%	0	100%	0	100%
$+\frac{\pi}{16}$	99.2%	0	100%	+0.02	99.2%	$\frac{3s}{2}$	99.6%	+10	98.4%
$+\frac{2\pi}{16}$	97.6%	+0.12	98.4%	+0.04	97.6%	$2s$	99.2%	+20	98.0%
		+0.24	98%						
		+0.36	97.2%						

Table 5.5: How changes in parameter values affect the correct recognition rate (CCR)

5.7 Conclusions

In this chapter we describe our proposed ear recognition technique. In our technique, ANOVA is used as a feature selection algorithm to select the best banana filters for the recognition process. The selected banana filters are convolved with the detected ear. Then the LBP is applied to the convolved images. After that, the produced image is described as histogram sequences in a LBP framework. The experiments show that when our proposed recognition technique is applied to XM2VTS face profile database (manual and automatic ear detection) and SOTON ear dataset 2, a recognition rate of 100%, 99.6% and 99.5% are achieved; respectively. Here, for the XM2VTS database, the ear is extracted manually to avoid the effects of ear detection process on the recognition, and automatically to get the result when the whole process is automatic.

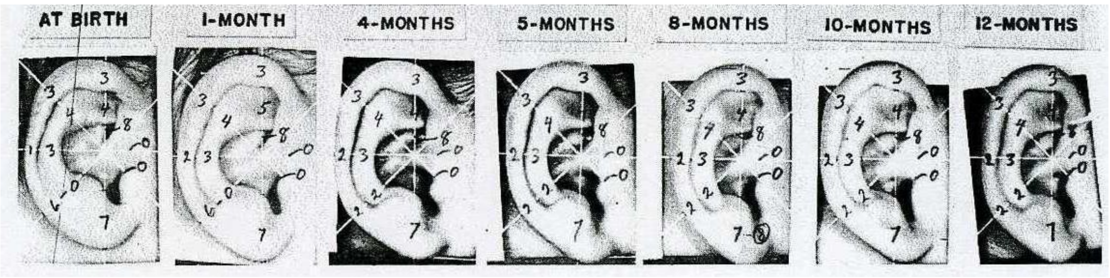
Chapter 6

The Effect of Time on Ear Biometrics

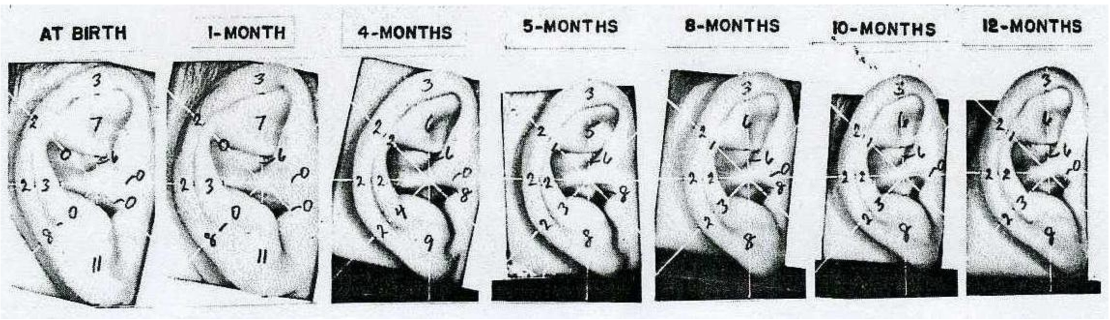
The notion that ear is relatively time invariant was originally noted in [71]. However there is no experimental study to show the time invariance property associated with ear biometric. In this chapter, the first experimental framework is proposed to investigate the effect of time on ear recognition.

6.1 Introduction

After birth, the matured ear remains unchanged from the ninth month until death, unless there is damage, disease, or surgery on ear [71]. Figure 6.1 illustrates the morphological change in configuration of male and female ears during the first twelve months. As shown in the figure, the ear changes after birth until month 8, after which no change is noted in the ear configuration. Although ears grow throughout life in certain respects, the growth does not affect anthropometric measurements of the ear for purposes of identification as growth simply changes scale and not structure [71]. This was the first notion that ear structure does not change significantly over time. Ear recognition techniques have mainly focused on recognition rate. But there is no experimental study to show the effect of time on ear recognition. We present an experimental study to demonstrate the effect of the time difference in image acquisition for gallery and probe on the performance of ear recognition. This experimental research is the first study on the time effect on ear biometrics. For the purpose of recognition, we use our ear recognition algorithm explained in the previous chapter.



(a) Male



(b) Female

Figure 6.1: Morphological change in configuration of male and female ears during first twelve months.

6.2 Database Description

The database (SOTON dataset 3) is selected from SOTON temporal database [72] from which a selection of images is shown in Figure 6.2. The advantage of this database is that, it is created to enable analysis to be performed over time which is captured in five different sessions over a period of eleven months. There are 25 subjects used in these sessions. The number of subjects in each session is not equal, and the number of the images for each subject is also not equal, because we select only the images without ear occlusion. The number of subjects and images available for each session is shown in Table 6.1.

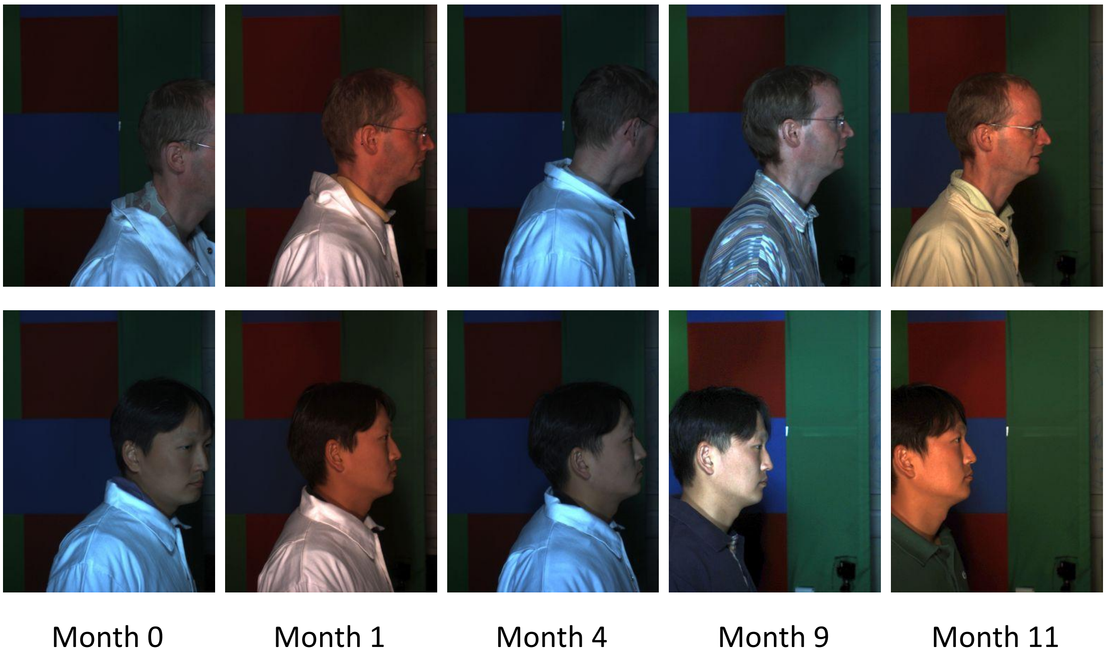


Figure 6.2: Samples from SOTON dataset 3

Month	Number of subjects	Number of images
0	18	337
1	18	350
4	14	234
9	12	231
11	15	339

Table 6.1: Number of Subjects and Images in each Session

6.3 The Effect of Time on Recognition

The effect of time on the performance of ear recognition is investigated by applying our proposed recognition technique to SOTON ear dataset 3. The average recognition rate achieved over eleven months is 98.8%. Table 6.2 shows the combinations of probe and gallery used in this experiment. The recognition rate for each combination is shown in Figure 6.3.

Probe (month)	Gallery (month)	Time Difference (months)
0	0	0
0	1	1
9	11	2
1	4	3
0	4	4
4	9	5
4	11	7
1	9	8
0	9	9
1	11	10
0	11	11

Table 6.2: The Combinations of Probe and Gallery

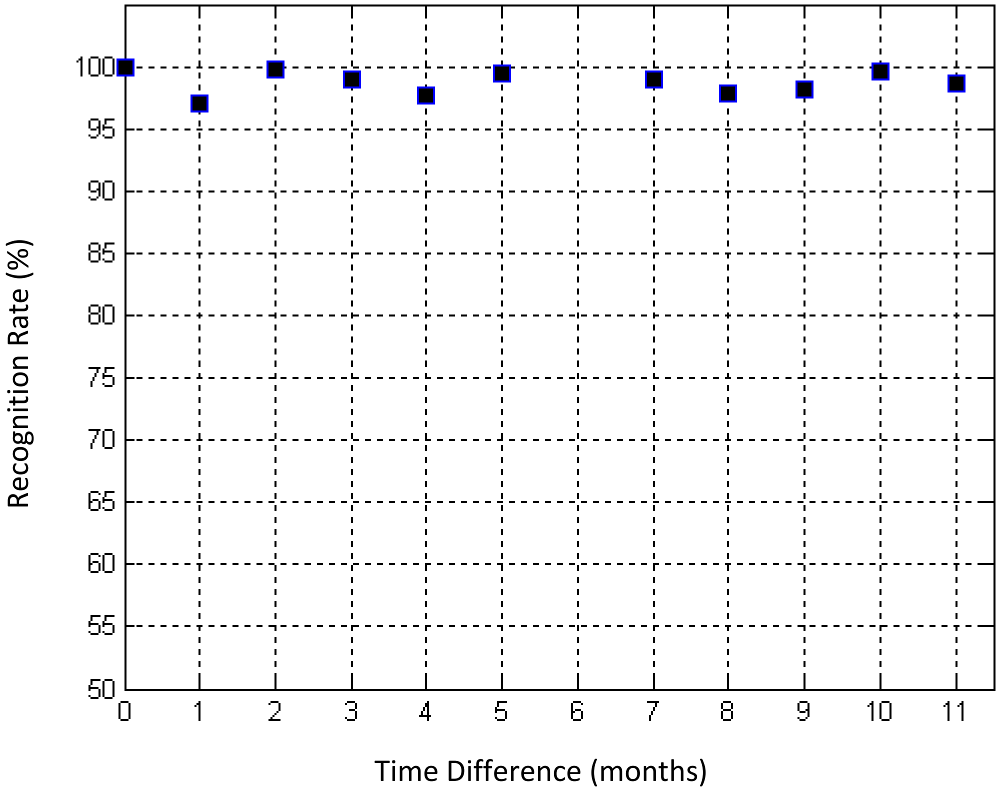


Figure 6.3: Recognition rate over time

The recognition rate for the time difference of 0, 2, 3, 5, 7, 10 and 11 months is nearly 100%. But the recognition rate drops by a value between 2% and 3% for the time difference corresponding to 1, 4, 8 and 9 months due to the higher values of rotation (Figure 6.4-a). Our recognition algorithm is relatively immune to slight rotation (Figure 6.4-b).

(a) Highly rotated ears for the same person

(b) Slightly rotated ears for the same person

Figure 6.4: Examples of highly and slightly rotated ears

6.4 The Effect of Time on Verification

The verification performance over time of our ear recognition technique is measured in terms of the decidability d' measure, EER value and ROC curve (these measures are described in Chapter 5). The average decidability and EER values are 2.5 and 11.0%, respectively, for applying the recognition technique to SOTON dataset 3. The decidability and EER values over eleven months are shown in Table 6.3. As expected, time difference 1, 4, 8 and 9 have the lowest d' values and the highest EER values (they also have the lowest recognition rate). Figure 6.5 illustrates the average ROC curve over eleven months. The figure also shows the variations in FAR and FRR values. Figure 6.6 shows two examples of ROC curves between months 0 and 11, and months 4 and 9.

Time Difference (months)	d'	EER
0	2.9	7.5%
1	2.3	13.4%
2	3.1	7.3%
3	2.3	12.2%
4	2.0	14.1%
5	2.3	12.0%
7	2.4	11.6%
8	2.1	17.8%
9	2.0	16.5%
10	3.0	6.8%
11	3.0	7.1%

Table 6.3: The Decidability and EER values over time

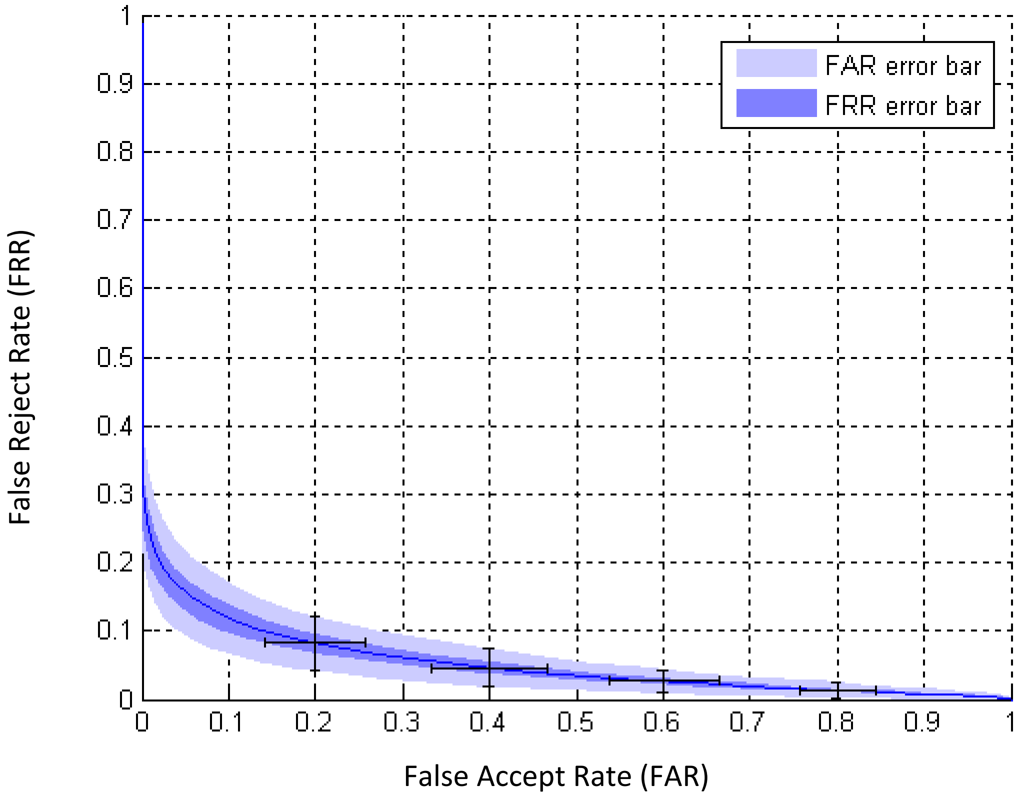
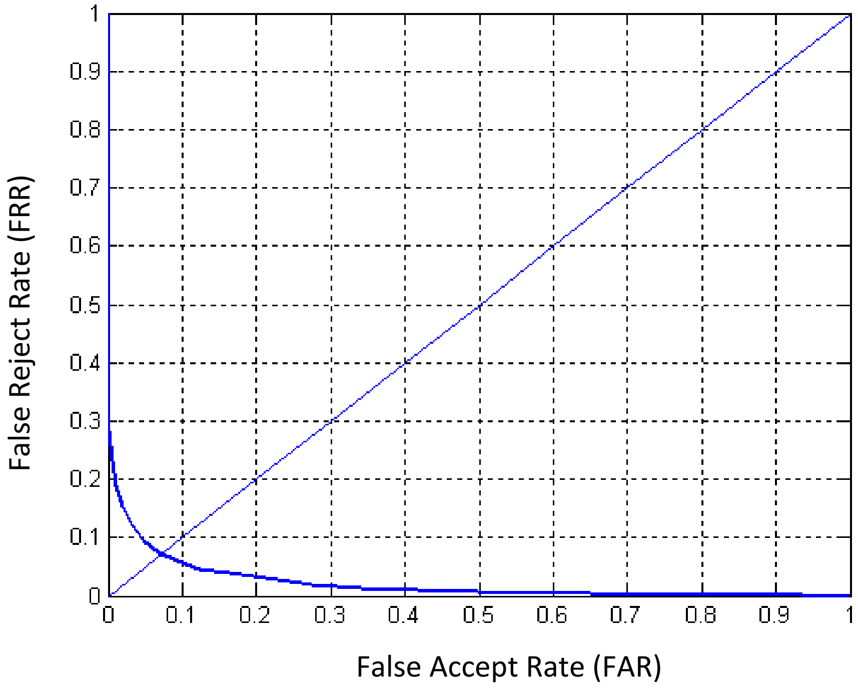
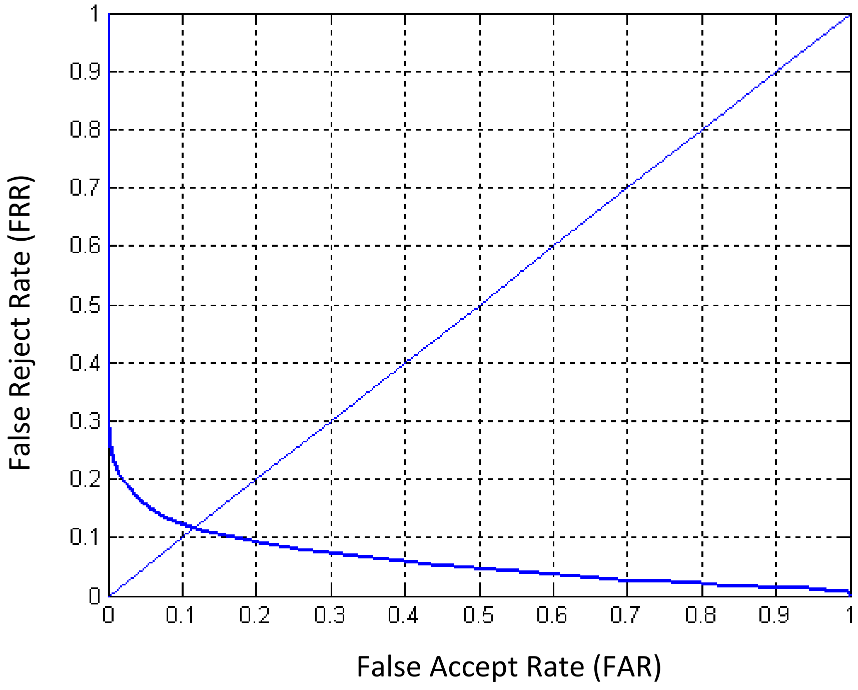


Figure 6.5: The ROC curve SOTON dataset 3 with error bar



(a) The ROC curve between month 0 and month 11



(b) The ROC curve between month 4 and month 9

Figure 6.6: The ROC curve for SOTON dataset 3

6.5 Conclusions

In this chapter, we measure the effect of time on the performance of the ear recognition. Here we present the first biometric experimental study that examines the effect of elapsed time on ear biometrics. The recognition approach is based on our recognition technique which is described in chapter 5. The experiments show that the recognition rate is not affected considerably over eleven months. The average recognition rate over eleven months is 98.8% for SOTON ear dataset 3. We can therefore conclude that ear can be used in various applications as a time invariant biometric. The verification performance over time is also examined in this chapter by computing the decidability measure, EER value and ROC curve.

Chapter 7

Conclusions and Future Work

7.1 Overall Conclusions

This thesis explores the use of frequency domain decomposition for ear biometrics by using banana wavelets to exploit the curvilinear structure of the ear. The complexity of the task has been reflected in the fact that ear images can vary in appearance under different viewing and illumination conditions. Our experiments show that the technique proposed here is effective for ear detection. Our technique is robust in the presence of noise and it does not depend on a controlled lighting conditions or skin colour; it therefore appears suitable for general applications in ear biometrics. The technique proposed here is applied to a more complex database known as SOTON database in which the image acquisition is largely uncontrolled. The result of the ear detection for the SOTON database is good enough according to the uncontrolled conditions, and shows the expected performance in occlusion. The performance of our banana wavelet based technique is compared with that of Gabor wavelet based technique showing that banana wavelets can capture the curved structures better than Gabor wavelets. The detection technique proposed here is fully automated and does not require any manual intervention to detect the ear. The success of our technique relies on the fact that the selected curvilinear structures are generic for all ears. To improve the detection results, a preprocessing stage is employed before applying our ear detection technique. The preprocessing stage segments skin and non-skin regions in an image using texture and colour. Then the adaptive filter is applied to the skin regions. There is an improvement in the detection rate when applying the preprocessing stage.

For recognition purposes, the recognition approach is based on the description of the ear image as histogram sequences. Our method to describe the ear depends on the local binary pattern algorithm. It is therefore robust to local image transformation due to variations of lighting and shifting. One major strength of our proposed method is that it appears to be general for various databases. The general structure of the ear has many curvilinear structures which match the structures of the banana filters. We apply a feature selection algorithm by using ANOVA based feature selection method to select the best filters from a bank of banana filters.

There is no experimental study to show the time invariance property associated with ear biometrics. The researchers have mainly focused on recognition rate. In this thesis, we have shown for the first time an experimental study to report the recognition rate over time and measure the effect of time on the performance of the ear recognition and verification. Our results show that ear can be used as reliable biometric over time. Clearly a database gathered over a longer timeframe could extend the results achieved here; however an 11 month timeframe is all that is available in current databases worldwide.

7.2 Future Work

The probe images considered in this study usually contain only one subject (i.e. one ear per image). An extension to our work can be to develop an algorithm to cope with multiple subjects per probe image. Such an extended algorithm can therefore automatically detect any number of ears with various scales (sizes) in a given image and then perform the recognition task on each detected ear.

Another future work could be to investigate and develop algorithms which can deal with more challenging outdoor cluttered and complex images. The complexities of such outdoor images include variations in lighting and shades, the presence of various objects characterised with similar structure as ears and so on.

Future research should therefore concentrate on the investigation of building an automatic ear biometric system for real time applications. Because the application will be used in real time system, it requires not only accurate detection and recognition results but also fast processing.

It is also interesting to quantitatively investigate the time invariance property of Ear biometrics with longer time difference than eleven months between the gallery and probe considered in this work. Such an investigation can be challenging due to dataset construction by finding the same subjects used in the gallery for the probe images after a long period of time.

Another future work would be to develop a method to automatically find a more accurate size of ears in the ear detection process. Such an algorithm can be implemented by employing a knowledge-based technique exploiting the relative size of heads and ears obtained from statistical measures in a population. The development of such an algorithm also demands a method to extract the head of a subject in an image.

In this thesis we have examined the performance of our ear detection and recognition algorithms on two databases. It would be interesting to test the performance of our automatic system on other databases and compare our results with those obtained from other researchers. It is also interesting to apply other techniques to our database (SOTON database) to compare the results from these techniques with our results.

The recognition technique used in this thesis is relatively immune to slight rotation but fails in certain cases due to the higher values of rotation. Therefore, a further study could be to investigate other classifiers. Different classification algorithms can be investigated and tested to obtain a more precise and robust classifier to higher values of rotation, occlusion and noise.

For more robust and accurate results for ear detection/recognition, a 3D ear representation can be employed to be able to capture high non planar rotation angles. This task could be challenging due to the requirement of the 3D image reconstruction.

Another interesting line of research for further studies could be to fuse ear biometric with other biometrics such as gait and face to improve the recognition rate especially for real data of outdoor images where one biometric is obscured but when another is available.

Appendix A

Adaptive Filter

The adaptive filter [59] structure consists of two adaptive filters with different cutoff frequencies and the DC (zero frequency) gains of the two filters differ by one. The two adaptive filters are connected in parallel-like structure to perform image edge enhancement. The filter structure is illustrated in Figure A.1.The first adaptive filter is of higher order, smaller gain and lower cutoff frequency. The second filter is of lower order and of higher cutoff frequency. The output of the first adaptive filter is subtracted from the output of the second adaptive filter, resulting in an output image with enhanced edges. The first filter size (order) can be any size 3×3, 5×5, or any required filter size. The second filter size is fixed to a cross shaped window of five weights centred at the same centre point of the first filter.

The adaptive filter structure has two inputs, the primary input d (the original image) and the reference input x (delayed version of the input image). The reference input, x, is used as an input to two adaptive filters. The first filter is assumed to have cutoff frequency ω_{c1}, the second filter is of lower order and has cutoff frequency ω_{c2}, where $\omega_{c2} > \omega_{c1}$. The second filter DC gain is larger than the first filter gain by 1.

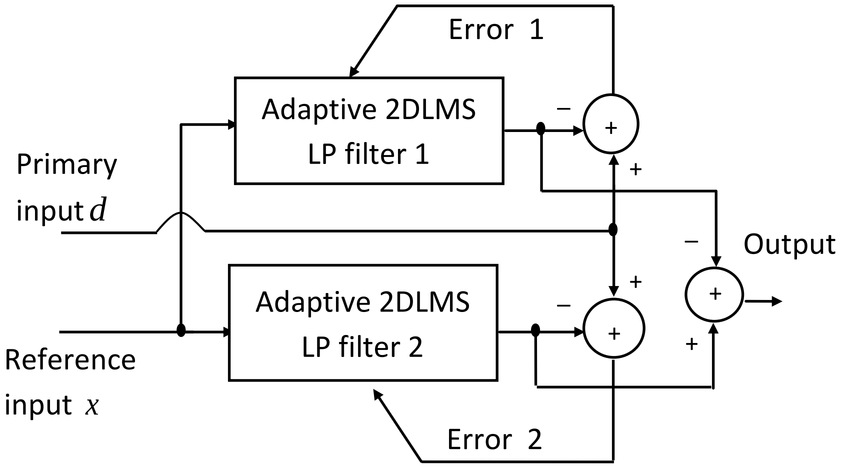


Figure A.1: The adaptive filter structure

For the parallel filter structure, if the output of the first filter is $y_1 = h_1 * x$ and the output of the second filter is $y_2 = h_2 * x$, where h_1 and h_2 are the impulse responses of the first and second filters respectively. The filter structure output is obtained as:

$$y = y_2 - y_1 = h_2 * x - h_1 * x \tag{A.1}$$

$$y = (h_2 - h_1) * x \tag{A.2}$$

Consider the following 3x3 adaptive coefficient matrix

$$\begin{bmatrix} H_{-1,-1} & W_{-1,0} & H_{-1,1} \\ W_{0,-1} & W_{0,0} & W_{0,1} \\ H_{1,-1} & W_{1,0} & H_{1,1} \end{bmatrix} \tag{A.3}$$

Imposing the constraint

$$\{\sum_{l}\sum_{\substack{k \\ (l,k)\neq(p,q)}}(-H_{l,k})\} + \{\sum_{p}\sum_{\substack{q \\ (l,k)\neq(p,q)}}(W_{p,q})\} = 1 \tag{A.4}$$

where the values of l,k,p,q for the 3×3 case are as shown in the matrix Eq. (A.3) and the coefficients $H_{l,k}$ and $W_{p,q}$ are the adaptive coefficients of the modified adaptive filter. The elements $W_{p,q}$ of the matrix in Eq. (A.3) represent an adaptive noise smoothing filter implemented inside the adaptive filter mask. Both coefficients $H_{l,k}$ and $W_{p,q}$ are updated independently according to the Two-Dimensional Least Mean Square (TDLMS) algorithm in the following equations [73]:

$$W_{j+1}(p,q) = W_j(p,q) + 2 \cdot \mu_{Wj} \cdot e_{Wj} \cdot X(m-p,n-q) \tag{A.5}$$

where μ_{Wj} is the convergence factor.

$$e_{Wj} = d_j - \{\sum_{p}\sum_{\substack{q \\ (p,q)\neq(l,k)}} W_{p,q} \cdot X(m-p,n-q)\} \tag{A.6}$$

where the values of (p,q) are obtained from the set (-1,0), (0,-1), (0,0), (0,1) and (1,0), and

$$H_{j+1}(l,k) = H_j(l,k) + 2 \cdot \mu_{Hj} \cdot e_{Hj} \cdot X(m-l,n-k) \tag{A.7}$$

where μ_{Hj} is the convergence factor.

$$e_{Hj} = d_j - \{\sum_{l}\sum_{\substack{k \\ (l,k)\neq(p,q)}} -H_{l,k} \cdot X(m-l,n-k)\} \tag{A.8}$$

for the remaining set of coefficients (l,k) in the filter mask of any size.

For the sum of the centre elements $\sum_{p,q} W_{p,q}$, we can assume any suitable value. This value depends on the amount of sharpening required by the filter. If the value of $\sum_{p,q} W_{p,q}$ assumes a positive value > 1, this means that the sum of the other elements in the matrix is negative and is less than the magnitude of $\sum_{p,q} W_{p,q}$ by 1. This form gives a filter matrix with adaptive coefficients.

The weights $H_{l,k}$ and $W_{p,q}$ are adaptively adjusted independently to satisfy Eq. (A.4) according to the choice of the sum $\sum_{p,q} W_{p,q}$ and according to the image statistics using the TDLMS algorithm, Eq. (A.5) and Eq. (A.7).

References

[1] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, *Handbook of Fingerprint Recognition*, Springer, New York, 2003.

[2] D. J. Hurley, M. S. Nixon, and J. N. Carter, "Force field feature extraction for ear biometrics," *CVIU*, vol. 98, no. 3, pp. 491-512, 2005.

[3] S. Cadavid, S. Fathy, J. Zhou, and M. Abdel-Mottaleb, "An Adaptive Resolution Voxelization Framework for 3D Ear Recognition," in *Proc. IJCB,*Washington DC, USA, 2011.

[4] F. Juefei-Xu, K. Luu, M. Savvides, T. Bui, and C. Suen, "Investigating age invariant face recognition based on periocular biometrics," in *Proc. IJCB,*Washington DC, USA, 2011.

[5] D. Smeets, J. Keustermans, J. Hermans, P. Claes, D. Vandermeulen, and P. Suetens, "Symmetric surface-feature based 3d face recognition for partial data," in *Proc. IJCB,*Washington DC, USA, 2011.

[6] J. Daugman, "New methods in iris recognition," *IEEE Trans. Systems, Man, Cybernetics B*, vol. 37, no. 5, pp. 1167-1175, 2007.

[7] J.Ko, Y. Gil, J. Yoo, and K. Chung, "A Novel and Efficient Feature Extraction Method for Iris Recognition," *ETRI Journal,* vol.29, no.3, pp.399-401, June 2007.

[8] L. Shen, and A. Kot, "A New Wavelet Domain Feature for Fingerprint

Recognition," *Biomedical Soft Computing and Human Sciences,*vol.14, no.1, pp.55-59, 2009.

[9] Z. M. Win and M.M. Sein, "Texture Feature Based Finger-Print Recognition for Low Quality Images,"*MHS 2011,* Nagoya, Japan, Nov, 2011.

[10] D. J. Hurley, B. Arbab-Zavar, and M. S. Nixon, "The ear as a biometric," in *Handbook of Biometrics*, A. Jain, P. Flynn, and A. Ross, Eds. 2008.

[11] D. J. Hurley, "Force Field Feature Extraction for Ear Biometrics," PhD Thesis, *Electronics and Computer Science*, University of Southampton, 2001.

[12] A. Iannarelli, *Ear Identification*. Paramount Publishing Company, Freemont, California, 1989.

[13] A. Bertillon, *La PhotographieJudiciaire*, avec unAppendicesur la Classification et l'Identification Anthropometriques, Gauthier-Villars, Paris, 1890.

[14] A. Abaza, A. Ross, C. Hebert, M. A. F. Harrison, and M. Nixon, "A Survey on Ear Biometrics,"*ACM Computing Surveys, Accepted,* (2011) (In Press)

[15] S. Islam, M. Bennamoun, R. Owens, and R. Davies, "Biometric Approaches of 2D-3D Ear and Face: A Survey," *In Advances in Computer and Information Sciences and Engineering*, T. Sobh, Eds. Springer Netherlands, pp. 509-514, 2008.

[16] K. Pun and Y. Moon, "Recent Advances in Ear Biometrics," *In Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition,* pp. 164-169, May 2004.

[17] M. Choras, "Image feature extraction methods for ear biometrics—A survey," *In Proc. 6th International Conference on Computer Information Systems and Industrial Management Applications*, pp. 261-265, Minneapolis-MN, USA, 2007.

[18] M. Burge, and W. Burger, "Ear biometrics," in *Biometrics: Personal ID in Networked Society*, A. Jain, R. Bolle, and S. Pankanti, Eds. Kluwer, pp. 273-286,

1998.

[19] M. Abdel-Mottaleb and J. Zhou, "Human Ear Recognition from Face Profile Images," in *ICB*, pp. 786-792, *2006*.

[20] L. Yuan and Z. C. Mu, "Ear detection based on skin-color and contour information," in *International Conference on Machine Learning and Cybernetics,* vol. 4, pp. 2213-2217, 2007.

[21] H. Chen and B. Bhanu, "Human ear detection from side face range images," in *Proc. International Conference on Pattern Recognition*, vol. 3, pp. 574-577, 2004.

[22] S. Ansari and P. Gupta, "Localization of ear using outer helix curve of the ear," in *Proc. the International Conference on Computing: Theory and Applications*, pp. 688-692, 2007.

[23] S. Prakash, U. Jayaraman, and P. Gupta, "Ear Localization from Side Face Images using Distance Transform and Template Matching," in *IPTA,* pp. 1-8, Sousse, Tunisia, Nov. 2008.

[24] S. Prakash, U. Jayaraman, and P. Gupta, "*A Skin-Color and Template Based Technique for Automatic Ear Detection*," in *Proc. ICAPR*, pp. 213-216, Kolkata, India, Feb. 2009.

[25] A. Cummings, M. Nixon, and J. Carter, "A Novel Ray Analogy for Enrolment of Ear Biometrics," in *IEEE Fourth Conference on Biometrics: Theory, Applications and Systems*, Washington D.C., Sep. 2010.

[26] S. M. S. Islam, M. Bennamoun, and R. Davies, "Fast and Fully Automatic Ear Detection Using Cascaded AdaBoost," in *Proc. IEEE Workshop on Application of Computer Vision*, USA, pp. 1-6, Jan. 2008.

[27] B. Arbab-Zavar, and M. S. Nixon, "On shape mediated enrolment in ear biometrics," in *Proc. ISVC'07*, pp. 549-558, Nevada, California, Nov. 2007.

[28] B. Moreno and A. Sanchez, "On the Use of Outer Ear Images for Personal

Identification in Security Applications," in *Proc. IEEE 33rd Intl. Conf. on Security Technology,* pp. 469-476, 1999.

[29] B. Victor, K. W. Bowyer, and S. Sarkar, "An evaluation of face and ear biometrics," in *Proc. ICPR*, pp. 429-432, 2002.

[30] K. Chang, K.W. Bowyer, S. Sarkar, and B. Victor, "Comparison and combination of ear and face images in appearance-based biometrics," *IEEE Trans. PAMI*, vol. 25, no. 9, pp. 1160-1165, 2003.

[31] H. Zhang, Z. Mu, W. Qu, L. Liu, and C. Zhang, "A novel approach for ear recognition based on ICA and RBF network," in *Proc. the Fourth International Conference on Machine Learning and Cybernetics*, pp. 4511-4515, 2005.

[32] Z. Mu, L. Yuan, Z. Xu, D. Xi, and S. Qi, "Shape and Structural Feature Based Ear Recognition," *Advances in Biometric Person Authentication,* pp. 663-670, China, 2004.

[33] M. Choras, "Ear Biometrics Based on Geometrical Feature Extraction," *Electronic Letters on Computer Vision and Image Analysis*, vol. 5, no. 3, pp. 84-95, 2005.

[34] B. Arbab-Zavar, M. S. Nixon, and D. J. Hurley, "On model-based analysis of ear biometrics," in *Proc. IEEE BTAS'07*, pp. 1-5, Washington DC, September 2007.

[35] B. Arbab-Zavar and M. S. Nixon, "Robust Log-Gabor Filter for Ear Biometrics," in *ICPR,* pp. 1-4, Tampa, Florida, 2008.

[36] B. Arbab-Zavar and M. S. Nixon, "On Guided Model-Based Analysis for Ear Biometrics," in *CVIU,* vol. 115, no. 4, pp. 487-502, 2011.

[37] J. D. Bustard and M. S. Nixon, "Robust 2D Ear Registration and Recognition Based on SIFT Point Matching," in *Proc. IEEE BTAS'08, 2008*.

[38] J. D. Bustard and M. S. Nixon, "Toward Unconstrained Ear Recognition From Two-Dimensional Images,"*IEEE Transactions on Systems, Man and Cybernetics* (A), vol. 40, no. 3, pp. 486-494, 2010.

[39] H. Chen and B. Bhanu, “Contour matching for 3-D ear recognition,” in *Proc. IEEE Workshop on Applications of Computer Vision*, pp. 123-128, Colorado, 2005.

[40] P. Yan and K. W. Bowyer, “ICP-based approaches for 3D ear recognition,” in *Biometric Technology for Human Identification II*, vol. 5779, pp. 282–291, 2005.

[41] S. Cadavid and M. Abdel-Mottaleb, “3D Ear Modeling and Recognition from Video Sequences using Shape from Shading,” *IEEE Transactions on Information Forensics and Security*, pp.1-4, Dec. 2008.

[42] M. M. Rahman and S. Ishikawa, “Proposing a Passive Biometric System for Robotic Vision,” in *Proc. the Tenth International Symposium on Artificial Life and Robotics,* Oita, Japan, 2005.

[43] K. Iwano, T. Hirose, E. Kamibayashi, and S. Furui, “Audio-Visual Person Authentication Using Speech and Ear Images,” in *Proc. Workshop on Multimodal User Authentication*, pp.85-90, 2003.

[44] P. Yan, K. W. Bowyer, “Ear biometrics using 2d and 3d images,” *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 121-128, 2005

[45] R. Katiyar and V. K. Pathak, “Recognition Based on Fusion of Gait, Ear and Face Features Using KPCA Method,”*In proc. International Conference on Intelligent Computing,* pp. 412-419, China, 2011

[46] Nixon, M.S., Bouchrika, I., Arbab-Zavar, B., Carter, J.N., “On use of biometrics in forensics: gait and ear,” *In Proc. European Signal Processing Conference*, pp. 1655-1659, Aalborg, Denmark, Aug. 2010.

[47] D. R. Kisku, P. Gupta, J. K. Sing, and M. Nasipuri, “Fusion of Gaussian Mixture Densities for Face and Ear Biometrics Using Support Vector Machines,” *In Proc Future Generation Information Technology*, pp. 344-351, Korea, Dec, 2010.

[48] N. Krüger, M. Pötzsch, and G. Peters, “Principles of Cortical Processing Applied

to and Motivated by Artificial Object Recognition,”in *Information Theory and the Brain,* Cambridge University Press, 2000.

[49] N. Krüger, and N. Lüdtke, “ORASSYLL: Object recognition with autonomously learned and sparse symbolic representations based on local line detectors,” in *British Machine Vision Conference,* 1998.

[50] N. Krüger, and G. Peters, “ORASSYLL: Object recognition with autonomously learned and sparse symbolic representations based on metrically organized local line detectors,” in *CVIU*, vol. 77, pp. 48-77, 2000.

[51] J. Rodrigues and J. du Buf, “Cortical object segregation and categorization by multi-scale line and edge coding,” in *Proc. the International Conference on Computer Vision, Theory and Applications*, vol. 2, pp. 5-12, Portugal, 2006.

[52] P Berkes, and L Wiskott, “Applying slow feature analysis to image sequences yields a rich repertoire of complex cell properties,” in *Proc. International Conference on Artificial Neural Networks*, pp. 81-86, 2002.

[53] F. Mendels, P. Vandergheynst , and J. P. Thiran, “Matching pursuit-based shape representation and recognition using scale-space,” *International Journal of Imaging Systems and Technology*, vol. 6, no. 15, pp. 162–180, 2006.

[54] M. S. Nixon and A. Aguado,*Feature extraction and image processing.* Academic Press, 2nd Ed., 2008.

[55] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. H. Romeny, J. B. Zimmerman, and K. Zuiderveld, “Adaptive histogram equalization and its variations,” in *Computer Vision Graphics Image Processing*, vol. 39, no. 3, pp. 355-368 , 1987.

[56] Z. Xu, X. Liu, and X. Chen, “Fog Removal from Video Sequences Using Contrast Limited Adaptive Histogram Equalization,” in *International Conference on Computational Intelligence and Software Engineering*, pp. 1-4, 2009.

[57] M.C. Shin, K.I. Chang and L.V. Tsap, “Does Colorspace Transformation Make

Any Difference on Skin Detection?” in *Proc. IEEE Workshop Applications of Computer Vision*, pp. 275-279, 2002.

[58] Z. Zhengzhen, and S. Yuexiang, “Skin color detecting unite YCgCb color space with YCgCr color space,”in *Proc. IASP*, pp. 221-225, China, 2009.

[59] M. M. Hadhoud, W.S. ElKilani, and M. I. Samaan, “An adaptive algorithm for fingerprints image enhancement using gabor filters,” in *Proc. International Conference on Computer Engineering and Systems*, pp. 227-236, 2007.

[60] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, “XM2VTSDB: The Extended M2VTS Database,” in *Proc. AVBPA'99*, Washington D.C., 1999.

[61] S. Samangooei, J. Bustard, M. S. Nixon, and J. N. Carter, “On Acquisition and Analysis of a Dataset Comprising of Gait, Ear and Semantic Data,” in *Multibiometrics for Human Identification*. B. Bhanu, and V. Govindaraju, CUP, 2011.

[62] G. M. Clarke and D. Cooke, *A Basic Course in Statistics*. Arnold, 1998.

[63] J. D. Shutler, “velocity moments for holistic shape description of temporal features,” PhD Thesis, *Electronics and Computer Science,* University of Southampton, 2002.

[64] T. Ojala, M. Pietikainen, and D. Harwood, “A comparative study of texture measures with classification based on feature distributions,”*Pattern Recognition*, vol. 29, no. 1, pp. 51-59, 1996.

[65] T. Ojala, M. Pietikainen, and T. Maenpaa, “Multiresolution grayscale and rotation invariant texture classification with local binary patterns”, in *IEEE PAMI*, vol. 24, no. 7, July 2002.

[66] T. Maenpaa, “The local binary pattern approach to texture analysis extensions and applications,” PhD Thesis, *University of Oulu*, 2003.

[67] W. Zhang, S. Shan, W. Gao, X. Chen and H. Zhang, “Local Gabor binary pattern histogram sequence (lgbphs): a novel non-statistical model for face

representation and recognition," in *Proc. ICCV*, pp. 786-791, 2005.

[68] M. Swain and D. Ballard. "Color indexing," in *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11-32, 1991.

[69] J. Daugman, "Biometric decision landscapes," University of Cambridge, Computer Laboratory, *Technical Report TR482*, 2000.

[70] M. Ismail, R. A. El-Khoribi,"HMT of the Ranklet Transform for Face Recognition and Verification,"in *ICGST's Graphics Vision and Image Processing Journal*, vol.6, no. 3,pp. 7-13, Dec. 2006.

[71] A. Iannarelli, *TheIannarelli system of ear identification*, Foundation Press, Brooklyn, New York, 1964.

[72] D. Matovski, M. Nixon, S. Mahmoodi, and J. Carter, "The Effect of Time on the Performance of Gait Biometrics," in *Proc. IEEE BTAS'10,* pp. 1-6, Washington DC, USA, 2010.

[73] M. M. Hadhoud and D. W. Thomas, "The two-dimensional adaptive LMS (TDLMS) algorithm," *IEEE Trans. Circuits Syst.*, vol. 35, pp. 485-494 , 1988.