

University of Southampton Research Repository ePrints Soton

Copyright © and Moral Rights for this thesis are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given e.g.

AUTHOR (year of submission) "Full thesis title", University of Southampton, name of the University School or Department, PhD Thesis, pagination

UNIVERSITY OF SOUTHAMPTON
FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS
School of Electronics and Computer Science

On Guided Model-Based Analysis for Ear Biometrics

by

Banafshe Arbab-Zavar

A thesis submitted for the degree of
Doctor of Philosophy

December 2009

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS

SCHOOL OF ELECTRONICS AND COMPUTER SCIENCE

Doctor of Philosophy

by Banafshe Arbab-Zavar

Ears are a new biometric with major advantage in that they appear to maintain their structure with increasing age. Current approaches have exploited 2D and 3D images of the ear in human identification. Contending that the ear is mainly a planar shape we use 2D images, which are consistent with deployment in surveillance and other planar-image scenarios. So far ear biometric approaches have mostly used general properties and overall appearance of ear images in recognition, while the structure of the ear has not been discussed. In this thesis, we propose a new model-based approach to ear biometrics. Our model is a part-wise description of the ear structure. By embryological evidence of ear development, we shall show that the ear is indeed a composite structure of individual components. Our model parts are derived by a stochastic clustering method on a set of scale invariant features on a training set. We shall review different accounts of ear formation and consider some research into congenital ear anomalies which discuss apportioning various components to the ear's complex structure. We demonstrate that our model description is in accordance with these accounts. We extend our model description, by proposing a new wavelet-based analysis with a specific aim of capturing information in the ear's outer structures. We shall show that this section of the ear is not sufficiently explored by the model, while given that it exhibits large variations in shape, intuitively, it is significant to the recognition process. In this new analysis, log-Gabor filters exploit the frequency content of the ear's outer structures.

In recognition, ears are automatically enrolled via our new enrolment algorithm, which is based on the elliptical shape of ears in head profile images. These samples are then recognized via the parts selected by the model. The incorporation of the wavelet-based analysis of the outer ear structures forms an extended or hybrid method. The performance is evaluated on test sets selected from the XM2VTS database. By results, both

in modelling and recognition, our new model-based approach does indeed appear to be a promising new approach to ear biometrics. In this, the recognition performance has improved notably by the incorporation of our new wavelet-based analysis.

The main obstacle hindering the deployment of ear biometrics is the potential occlusion by hair. A model-based approach has a further attraction, since it has an advantage in handling noise and occlusion. Also, by localization, a wavelet can offer performance advantages when handling occluded data. A robust matching technique is also added to restrict the influence of corrupted wavelet projections. Furthermore, our automatic enrolment is tolerant of occlusion in ear samples. We shall present a thorough evaluation of performance in occlusion, using PCA and a robust PCA for comparison purposes. Our hybrid method obtains promising results recognizing occluded ears. Our results have confirmed the validity of this approach both in modelling and recognition. Our new hybrid method does indeed appear to be a promising new approach to ear biometrics, by guiding a model-based analysis via anatomical knowledge.

Contents

Glossary	v
Declaration of Authorship	vi
Acknowledgements	vii
1 Context and Contributions	1
1.1 Context	1
1.2 Contributions	2
1.3 Thesis Overview	3
1.4 List of Publications	5
2 Ear Biometrics	6
2.1 The Early Work of Iannarelli and Ear Forensics	8
2.2 Burge and Burgers’ Proof of Concept	10
2.3 Force Field Approach	10
2.4 3D Ear Biometrics	12
2.4.1 Yan and Bowyers’ Approach	12
2.4.2 Chen and Bhanus’ Local Surface Patch Approach	15
2.5 Acoustic Ear Recognition	17
2.6 Ear in Forensics	18
2.7 Conclusions	20
3 Enrolment	21
3.1 Hough Transform for ellipses	22
3.2 Eliminating Erroneous Votes	24
3.2.1 Vertical Ellipses	24
3.2.2 Selective Pairing	26
3.2.3 Spectacles and Anticipated Errors	28
3.3 Ear Enrolment	29
3.4 Conclusions	32
4 Ear Features – a Biological Approach	33
4.1 Ear Anatomy	34
4.2 Ear Embryology	35
4.2.1 Motivations	36
4.2.2 Hillocks Hypothesis	37
4.3 Variations and Anomalies	39

4.3.1	Cleft Defects	40
4.3.2	Helix and Antihelix	42
4.3.3	Ascending helix	44
4.3.4	Tragus	45
4.3.5	Superior crus of antihelix	46
4.3.6	Lobe	47
4.4	Discussion	48
4.5	Conclusions	50
5	The Parts-Based Model	51
5.1	Feature Extraction	52
5.2	Learning	53
5.2.1	Updating Clusters	55
5.2.2	Revision	55
5.2.3	Termination	56
5.3	Training Set and Parameters	56
5.4	The Derived Model	58
5.4.1	Feature Subset Selection	61
5.4.2	The Ten Most Significant Parts	62
5.5	Conclusions	64
6	Robust Log-Gabor Filter for the Outer Ear Curves	66
6.1	Improving the Parts-Based Model	67
6.2	Approach	68
6.3	Preparing the Templates	70
6.4	Wavelet Approach and the Log-Gabor Filter	71
6.4.1	Log-Gabor Filter	73
6.4.2	Filter Parameters	74
6.5	Robust Matching	79
6.6	Conclusions	81
7	Hybrid Classification and Experimental Results	82
7.1	Test and Training Data	82
7.2	Ear Recognition	88
7.3	Occlusion Analysis	92
7.3.1	PCA and Robust PCA	93
7.3.2	Effects on Enrolment	95
7.3.3	Hybrid Classification and Occlusion	97
7.4	Conclusions	105
8	Conclusions and Future Work	107
8.1	Conclusions	107
8.2	Future Work	109
A	Further Work on the Model	111
	References	114

Glossary

Ear Biometrics:	methods exploiting the 2D and 3D images of the external ear in human identification.
Ear print:	prints left behind as a result of the ear touching a surface, for example while listening at a door. Ear prints have been used as forensic evidence.
Feature:	a related piece of information extracted from an ear image.
Occlusion:	any obstacle hindering the capturing of the full image of an ear, which would otherwise be visible from the side view of the face. Hair and earrings are the most probable objects to occlude an ear image.
Noise:	additive Gaussian noise which corrupts an image. Noise arises in sampling, quantization and transmission of images. It may also refer to an unwanted output of a filter, processing an ear image.
Pose:	the viewpoint from which the image of the ear is captured. Zero angle pose for the ear image is when the image is captured normal to the ear plane.
Enrolment:	the method separating the object of interest, the ear, from the background, commonly the rest of the face-profile image components, including eyes, nose, mouth, hair, neck and etc.

Declaration of Authorship

I, Banafshe Arbab-Zavar, declare that the thesis entitled ‘On Guided Model-Based Analysis for Ear Biometrics’ and the work presented in it are my own. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- parts of this work have been published as listed in Section 1.4 of this thesis.

Signed:

Date :

Acknowledgements

I would like to express my gratitude to my supervisor Professor Mark Nixon for his extensive guidance and timely encouragements, and for introducing me to an innovative and exciting approach to research. Thanks also go to Dr. David Hurley for his insightful comments and discussions.

I would also like to thank my parents and my brother, who have supported me in every possible way, and whose love and advice have accompanied me all the way through. Thanks are also due to Ann and Lionel Wardle for their guidance, support and friendship, and also to Hazel Peasley and Abelardo Clariana-Piga. I would also like to thank Dominic Buchstaller, Michela Goffredo and Gabriele Gherbaz for all the interesting discussions and for being such good friends. And finally thanks to all my friends and all the members of the ISIS group who made this a joyful place to study.

*To Mum, Dad
and Babak*

Chapter 1

Context and Contributions

1.1 Context

Biometrics is a fast, reliable, and convenient means of identifying individuals based on their physical or behavioural traits. Various biometric traits have been studied, including: finger print, iris, face, odour, voice, signature, hand geometry, gait and ear. However, there is no single biometric that can satisfy the needs of all applications. Finger prints and irises offer a high level of uniqueness, however, to capture a sample for identification the subject's cooperation is required, and their usage might not be as acceptable as faces. The face is perhaps the most natural and acceptable biometric for humans, with potentially straightforward deployment. However, there are a number of obstacles associated with an automated face recognition system, most notably the variations with aging and facial expressions. Gait offers a unique opportunity for recognition at a distance when no other biometric can be seen, whilst facing different complications such as the variations in clothing and the effects of injuries. The ear is a relatively new biometric trait. It can be captured in a similar way to the face so it is non-intrusive. However, the ear's major advantage is that it appears to be consistent with age. Potential occlusion by hair seems to be the main obstacle for ear biometric recognition.

The French criminologist Alphonse Bertillon was the first to recognize the biometric potential of human ears in 1890 [7]. He incorporated features from the ear in his 'spoken portrait' method for forensic identification. He described the ear as the most discrimi-

nant part of an individual and one which does not change with age. Empirical evidence supporting the ear's uniqueness was later provided by Iannarelli, who examined over 10,000 ear samples and developed a manual system for ear identification [41]. Ear prints have also been used as forensic evidence [55, 24]. Despite the use of ear evidence in forensics for many years, machine vision approaches to ear identification are relatively new. Burge et al. [10] were amongst the first to introduce an automatic ear biometric method. Current approaches have exploited 2D and 3D images of the ear in human identification.

1.2 Contributions

Despite various approaches to ear biometric recognition, the structure of the ear has not been explicitly understood, and discriminant features have not been identified. Current approaches, which are mainly holistic, capitalize on general properties and overall appearance of the images. Contending that the ear is mainly a planar shape we use 2D images, which are consistent with deployment in surveillance and other planar-image scenarios. By evidence from the embryological development of the ear, we propose that the ear is better described as a composite structure of separate parts. We thereby propose a new model-based approach, in which our ear model is a constellation of various ear components. Ear embryology studies attribute individual growth centres to the development of the ear, apportioning various components to the ear's complex structure. Even though there is no direct evidence to sustain the link between ear development and automated recognition, it can guide our approach and provide a basis for explicit evaluation of the proposed method.

Our model is the first model-based approach to ear biometrics. The deployment advantages of a model include robustness in noise and occlusion, which is particularly favourable since images of the ear are susceptible to occlusion, mostly by hair. Our model is learned via a stochastic clustering algorithm on a set of scale invariant features detected on a training set. Extending our model description, we also propose a new wavelet-based analysis which explores the fluctuations in the two parallel ridges of the ear boundary. We shall illustrate that the information residing in these curves has only been partially explored by the model. By embryological evidence, these embody big variations and are potentially discriminant, hence the choice of a wavelet analysis

which is targeted to capture information in these components. By localization, a wavelet can also offer performance advantage when handling occluded data. Results from both modelling and recognition indicate that our new hybrid method does indeed appear to be a promising new approach to ear biometrics.

The potential occlusion by hair and earrings is the biggest obstacle hindering the deployment of ear biometrics. However, this has not been carefully addressed in ear biometric studies, and the extent of performance impairment in occlusion has not been evaluated. As mentioned, our techniques offer advantages in handling occlusion. We shall also show that our automatic enrolment performs well in the presence of occlusion. We present a thorough evaluation of the performance of our method in occlusion and compare the results against a robust PCA (Principal Component Analysis). In this, our hybrid method achieves promising results recognizing occluded ears.

1.3 Thesis Overview

- Chapter 2 : *Ear Biometrics*

In this chapter, we shall review the main approaches to ear biometrics [A, D]. Starting with the Iannarelli's manual method for forensic recognition using ears, we then describe the most prominent machine vision approaches, exploiting 2D images and 3D data. We also look at a related study in which acoustic properties of the ear are used for identification. The use of earprint evidence in forensics has also been discussed.

- Chapter 3 : *Enrolment*

In this chapter, we shall describe our new automatic ear enrolment method [C], which enrolls ears from 2D head-profile images. In this, a reduced Hough Transform for ellipses is used for the initial location of the ear. Adding refinements and incorporating some prior knowledge regarding the specifications of the expected ear ellipse, the ear location is detected. An image patch localizing the ear is then cropped.

- Chapter 4 : *Ear Features – a Biological Approach*

The material discussed in this chapter is mainly derived from embryological and surgical accounts of the human ear, revisited from a new perspective, to be ex-

exploited in ear biometrics [F]. We shall review the different accounts of external ear embryology, and discuss in relation to the variations and components of the ear structure. In this, the abnormalities of the auricle often offer vital cues. We shall illustrate the analysis with example ears from the surgical studies and, when possible, ears from the database that will later be analyzed.

- Chapter 5 : *The Parts-Based Model*

In this chapter, we shall present our new parts-based model for ear biometrics [B, F]. Our model is learned via a stochastic clustering algorithm on a set of scale invariant features detected on the training set. By embryological evidence, although the ear appears as an intact shape, it is a confluence of individual components, which validates our modelling scheme. We shall analyze the derived parts of the model and evaluate their significance in recognition. Having compared these parts with the embryological components, we shall explicitly evaluate the extent of ear information that has been captured via the model. Our results confirm the validity of this approach not just in modelling, but also in recognition capabilities.

- Chapter 6 : *Robust Log-Gabor Filter for the Outer Ear Curves*

Extending our model description, in this chapter, we propose a new wavelet-based approach [E, F]. A specific aim of this is to capture information in the ear's outer structures, which have been under-represented in the model. In this, the target neighbourhood is captured in aligned templates. The variations between the boundary curves are then explored using log-Gabor filters. We analyze the robustness of this method to partial occlusion and add a more robust matching strategy.

- Chapter 7 : *Hybrid Classification and Experimental Results*

In this chapter, starting with the description of the database, we shall present extensive performance analysis [F]. For this, the data is divided into various test sets accommodating the need for an objective evaluation of the various aspects of our biometric approach. In recognition, the ear model and the wavelet-based analysis are employed both separately and when combined. We shall specifically focus our attention on assessing the effects of occlusion.

- Chapter 8 : *Conclusions and Future Work*

Finally in this chapter, overall conclusions are reviewed and potential future work avenues are discussed.

1.4 List of Publications

So far the following papers have arisen from this research:

- [A] David J. Hurley, Banafshe Arbab-Zavar, and Mark S. Nixon. The ear as a biometric. In *European Signal Processing Conference (EUSIPCO'07)*, Poznan, Poland, pages 25-29, September 2007.
- [B] Banafshe Arbab-Zavar, Mark S. Nixon, and David J. Hurley. On model-based analysis of ear biometrics. In *IEEE Conference on Biometrics: Theory, Applications and Systems (BTAS'07)*, pages 1-5, Washington DC, September 2007.
- [C] Banafshe Arbab-Zavar and Mark S. Nixon. On shape-mediated enrolment in ear biometrics. In *International Symposium on Visual Computing (ISVC'07)*, pages 549-558, Nevada, November 2007.
- [D] David J. Hurley, Banafshe Arbab-Zavar, and Mark S. Nixon. The ear as a biometric. In A. Jain, P. Flynn, and A. Ross, editors, *Handbook of Biometrics*, pages 131-150, Springer, 2008.
- [E] Banafshe Arbab-Zavar and Mark S. Nixon. Robust log-Gabor filter for ear biometrics. In *International Conference on Pattern Recognition (ICPR'08)*, pages 1-4, Tampa, Florida, December 2008.
- [F] Banafshe Arbab-Zavar and Mark S. Nixon. Ear biometrics by guided model-based analysis. In submission.

Chapter 2

Ear Biometrics

Ears have long been considered as a potential means of personal identification. The French criminologist Alphonse Bertillon was the first to recognize the biometric potential of human ears in 1890 [7]; ear features were included in his multimodal system for forensic recognition. Empirical evidence supporting the ear’s uniqueness was later provided in two studies by Iannarelli, examining over 10,000 ears [41]. Earprints have also been considered as a possible source of forensic evidence [55, 24].

Ears have appealing properties for personal identification: they have a rich structure that appears to be consistent throughout life from a few months after birth; clearly, ears are not affected by changes in facial expression; images of ears can be acquired without the subject’s participation, with no hygiene issues, and, ear images can be captured from a distance. However there exists a big potential obstacle — the occlusion by hair and earrings, which is almost certain to happen in uncontrolled environments.

Most ear biometric approaches have exploited the ear’s planar shape in 2D images. One of the first computer-based approaches to ear biometrics was introduced by Burge and Burger [10]. They modeled each individual ear with an adjacency graph which was calculated from a Voronoi diagram of the ear curves. However, they did not provide an analysis of biometric potential. Moreno et al. [57] used different combinations of several neural classifiers and reported a recognition rate of 93% on a dataset of 168 images. Hurley et al. [40] used force field feature extraction to map the ear to an energy field which highlights ‘potential wells’ and ‘potential channels’ as features. Achieving a recognition rate of 98.4% on a dataset of 252 images, this method proved to yield

a much better performance than PCA when the images were poorly registered. The force field transform is also robust in the presence of noise. Abdel-Mottaleb et al. [2] used the force field transform to obtain a smooth surface representation for the ear and then applied different surface curvature extractors to gather the features. Yuizono et al. [92] treated the problem as an optimization problem, and proposed a recognition system using a specially developed genetic local search with respect to the ear images. Given that their work does not include any feature extraction process, it has no invariant properties. Naseem et al. [59] have proposed the use of sparse representation, following its successful application in face recognition. The geometrical properties of ear curves have also been used for recognition [17, 18, 58]. The most prominent example of these, and arguably the first ear biometric method, proposed by Iannarelli [41], was based on measurements between a number of landmark points, determined manually. These methods are primarily reliant on accurate segmentation and positioning of the landmarks. Alvarez et al. [6] proposed a method to estimate the outer ear by an ovoid model and suggested that this can be used in segmentation and normalization stages of ear biometric systems. Bustard et al. [12] have recently proposed a robust registration technique for 2D ear images addressing problems such as pose variation and clutter. Yuan et al. [91] and Liu et al. [51] have also examined the effects of pose variation on ear biometrics.

The 3D structure of the ear has also been exploited, and good results have been obtained [88, 14, 63, 73]. Yan et al. [88] captured 3D ear images using a range scanner and having segmented the ear, they used Iterative Closest Point (ICP) registration for recognition to achieve a 97.8% recognition rate on a database of 415 individuals. Chen et al. [14] proposed a 3D ear detection and recognition system using a model ear for detection, and using a local surface descriptor and ICP for recognition, reporting a recognition rate of 90.4%. Although using 3D can improve the performance, using 2D images is consistent with deployment in surveillance or other planar image scenarios. Also, given that the human ears have a reasonably flat structure, and the inner parts – when the 3D changes really occur – are quite hard to image successfully, the choice of expensive 3D ear biometrics is debatable.

PCA (Principal Component Analysis) has been used regularly in ear biometric research [13, 84, 40]. It is a statistical holistic analysis which obtains satisfactory results in controlled situations. Other statistical methods such as ICA (Independent Component

Analysis) and LDA (Linear Discriminant Analysis) have also been utilized [93, 83]. However these methods have almost no invariance properties, thus they rely on the acquisition and pre-processing stages to window and align the data.

A number of multimodal approaches containing the ear biometric have also been considered. Iwano et al. [43] combined ear images and speech using a composite posterior probability, and showed that the performance improves using ear images in addition to speech in the presence of noise. Face and ear have also been used together in biometric recognition studies [77, 13, 66, 82]. In these, PCA was mostly used to extract the features, and an increase in performance was commonly reported using the multimodal biometric instead of the individual biometrics. Victor et al. [77] showed that the face yields a better performance than the ear using PCA. However, Chang et al. [13] conducted a similar experiment and reached a different conclusion: no significant difference was observed with face and ear biometrics. The image dataset in the Victor et al. study had less control over covariates such as earring, hair over ears, exact lighting setup over time, etc. As suggested by Chang et al., these variations in image quality confound with the covariates under study and cause this contradiction between the two experiments. Woodard et al. [82] combined 3D ear with 3D face, using the Iterative Closest Point (ICP) to compare the samples. In this study, the face outperforms the ear in the recognition task. Yan et al. [84] proposed a combination of 2D and 3D ear data. Using PCA for 2D ears and an edge-based Hausdorff distance method for 3D ear matching, they reported a 71.9% recognition rate for both 2D and 3D data. The performance was significantly improved to 90.6% when using the multimodal biometrics.

In related studies, Akkermans et al. [4] developed an ear biometric system based on the acoustic properties of the outer and middle ear. This introduces a unique opportunity for ear biometrics to combine the image-based information with acoustic data. In this chapter, we shall describe the main techniques for ear recognition, and we shall also look at the Akkermans's acoustic analysis and the use of earprints in forensics.

2.1 The Early Work of Iannarelli and Ear Forensics

In 1949, Alfred Iannarelli became interested in the ear as a means of personal identification in the context of forensic science. He subsequently developed the Iannarelli System

of Ear Identification [41], which was used by American law enforcement agencies. As shown in Figure 2.1, his system essentially consists of taking a number of measurements from a set of landmark points on the ear. In this, the ear image is projected onto a standard easel, where the images are carefully aligned by adjusting the position of some reference lines to the tragus and the crus of helix (see section 4.1 for a description of the anatomical parts of the ear). The ear images are then adjusted to a standard size. The easel's four guide lines, which are at equal 45° intervals, are reproduced on the ear image. Figure 2.1 shows an ear image with these guide lines superimposed. The intersections of the guide lines and the ear curves are then marked. The Iannarelli measurements are computed based on the positions of these markers.

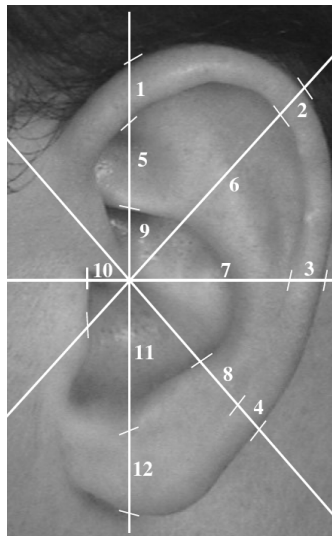


FIGURE 2.1: Iannarelli's manual ear measurement system (from [10]).

Iannarelli has appeared personally as an expert witness in many court cases involving ear evidence, or is often cited as an ear identification expert by other expert witnesses [50]. In the preface to his book Iannarelli states,

“Through 38 years of research and application in earology, the author has found that in literally thousands of ears that were examined by visual means, photographs, ear prints, and latent ear print impressions, no two ears were found to be identical - not even the ears of any one individual. This uniqueness held true in cases of identical and fraternal twins, triplets, and quadruplets”

We believe that Iannarelli was referring to the left and right ears of an individual when

he suggests that “not even the ears of any one individual are unique”. Yan et al. [88] reported that about 10% of people have significantly different left and right ears. In recognition, considering each sample as a point in the feature space, samples belonging to an individual will form a cluster. It is the extent to which these clusters are separated from each other and the extent to which the clusters are closely grouped around their own means, that determines how well a particular biometric system performs.

2.2 Burge and Burger’s Proof of Concept

Burge and Burger [10, 11] were the first to study the human ear as a biometric in the context of machine vision. Inspired by the earlier work of Iannarelli, they conducted a proof of concept study where the viability of the ear as a biometric was discussed theoretically, in terms of the uniqueness and measurability over time, and examined in practice through the implementation of a computer vision based system. In this, each subject’s ear was modeled as an adjacency graph built from the Voronoi diagram of its Canny extracted curve segments. They introduced a novel graph matching algorithm for authentication which takes into account the erroneous curve segments which can occur in the edge detected image due to changes such as lighting, shadows, and occlusion. They found that the features were robust and could be reliably extracted from a distance. Figure 2.2 shows the extracted curves, a Voronoi diagram, and a neighbourhood graph for an ear. They also identified occlusion by hair as a major obstacle for ear recognition and proposed the use of thermal imagery to resolve this problem.

2.3 Force Field Approach

An invertible linear transform has been developed by Hurley et al. [38, 39, 36, 40], who pretended that pixels have a mutual attraction proportional to their intensities and inversely proportional to the square of the distance between them. This transform is rather like the Newton’s Law of Universal Gravitation. Associated with this force field, there is an energy field, which in the case of an ear image takes the form of a smooth surface with a number of peaks joined by ridges (see figure 2.3). The peaks are the potential energy wells and the ridges correspond to potential energy channels. Considering the invertibility of this transform, and since the energy surface is otherwise

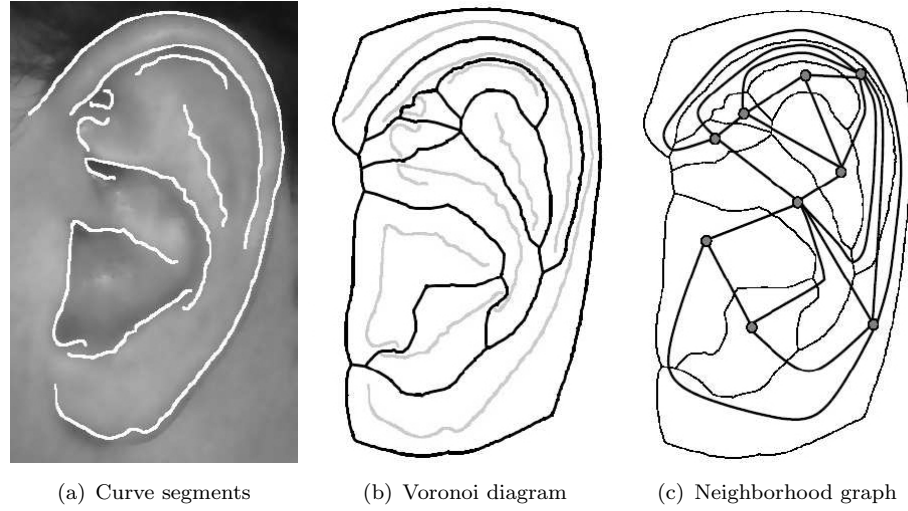


FIGURE 2.2: Stages in building the ear biometric graph model. A generalized Voronoi diagram of the curve segments is built and a neighborhood graph is extracted [10].

smooth, information theory suggests that much of the information is transferred to these features, and therefore they should make good features.

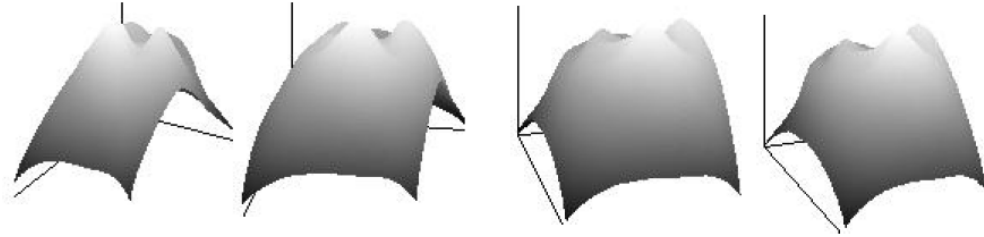


FIGURE 2.3: Energy surface for an ear viewed from below the lobe [40].

Equations 2.1 and 2.2 show how the force and energy fields, \mathbf{F} and E , are calculated at any point \mathbf{r}_j .

$$\mathbf{F}(\mathbf{r}_j) = \sum_i \left\{ P(\mathbf{r}_i) \frac{\mathbf{r}_i - \mathbf{r}_j}{|\mathbf{r}_i - \mathbf{r}_j|^3} \right\} \forall i \neq j, \quad 0 \quad \forall i = j \quad (2.1)$$

$$E(\mathbf{r}_j) = \sum_i \frac{P(\mathbf{r}_i)}{|\mathbf{r}_i - \mathbf{r}_j|} \forall i \neq j, \quad 0 \quad \forall i = j \quad (2.2)$$

Hurley et al. offered two different methods for extracting the features of the energy field. The first method, the field line map, depicted in figure 2.4a is algorithmic, where test pixels seeded around the perimeter of the force field are allowed to follow the force field direction joining together to form channels which terminate in potential wells. The second method, the convergence field, depicted in figure 2.4b is analytical, where a scalar function analyzes the divergence of the force vector field. This method was used

to obtain a recognition rate of over 99% on a dataset of 252 ear images consisting of 4 time lapsed samples from 63 subjects, extracted from the XM2VTS face-profile database [56]. We have also used this database in our experiments.

Convergence provides a more general description of channels and wells in the form of a mathematical function in which wells and channels are presented as peaks and ridges in the function value respectively. $C(\mathbf{r})$ is defined by,

$$C(\mathbf{r}) = -\text{div}\mathbf{f}(\mathbf{r}) = -\lim_{\Delta A \rightarrow 0} \frac{\oint \mathbf{f}(\mathbf{r}) \cdot d\mathbf{l}}{\Delta A} = -\nabla \cdot \mathbf{f}(\mathbf{r}) = -\left(\frac{\partial f_x}{\partial x} + \frac{\partial f_y}{\partial y}\right) \quad (2.3)$$

where $\mathbf{f}(\mathbf{r}) = \frac{\mathbf{F}(\mathbf{r})}{|\mathbf{F}(\mathbf{r})|}$ is the force direction, ΔA is an incremental area, and $d\mathbf{l}$ is its boundary outward normal. The negative values of $C(\mathbf{r})$ correspond to diverging force and present extra feature extraction opportunities.

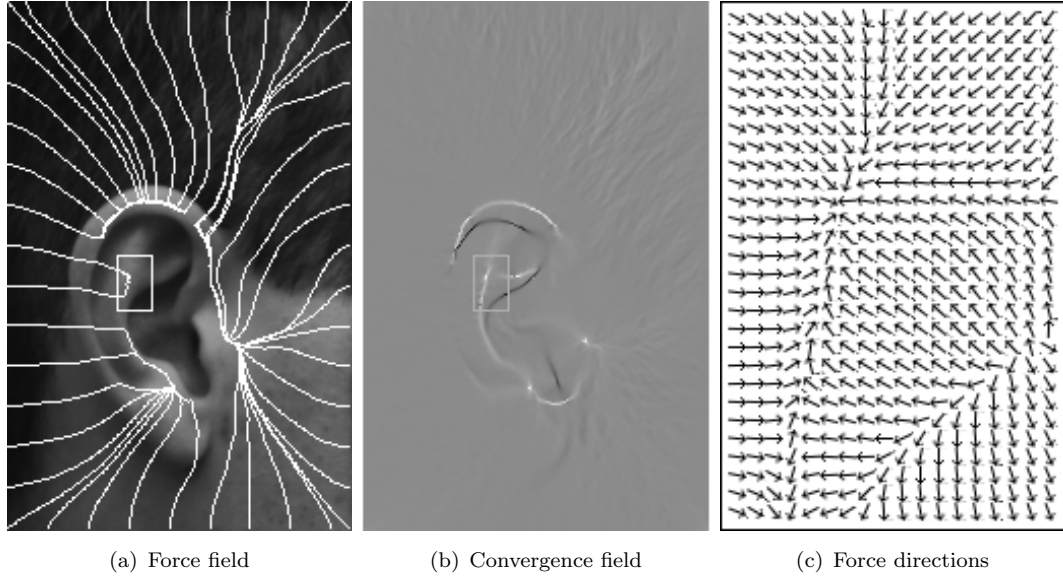


FIGURE 2.4: The force field for an ear and its corresponding convergence field. The force direction field corresponds to the small rectangular inserts surrounding a potential well on the antihelix [40].

2.4 3D Ear Biometrics

2.4.1 Yan and Bowyers' Approach

Yan et al. [88, 87, 85] use a Minolta VIVID 910 range scanner to capture both depth and colour information. This device uses a laser to scan the ear, and depth is automatically

calculated using triangulation. They have developed an automatic ear biometric system using ICP-based 3D shape matching for recognition, and using both 2D and 3D data for automatic ear extraction which not only extracts the ear image but also separates it from possible hair and earrings. They have achieved a recognition rate of 97.8% on a time-lapse dataset of 1,386 images from 415 subjects, with an equal error rate of 1.2%.

Ear Extraction uses a multistage process, in which both 2D and 3D data are used to detect the ear pit which is then used to initialize an elliptical active contour to locate the ear outline and crop the 3D ear data.

Ear pit detection itself includes various processes: *i*) geometric preprocessing to locate the nose tip which acts as the hub of a sector containing the ear with a high degree of confidence; *ii*) skin detection to isolate the face and ear region from hair and clothes; *iii*) surface curvature estimation to detect the pit regions depicted in black in the top right corner of figure 2.5; *iv*) surface segmentation and classification to select amongst possible multiple pit regions using a voting scheme to select the most likely candidate. Figure 2.5 depicts the steps for ear pit detection on an ear image. The detected ear pit is



FIGURE 2.5: 3D ear pit detection. From left to right starting from the input image; i&ii) most likely sector generation and skin detection; iii) pit detection; iv) selection and surface segmentation and classification [88].

then used to initialize an elliptical active contour algorithm to find the ear outline. Both 2D colour and 3D depth are used to drive the contour, as using either alone is inadequate

since there are cases in which there is no clear colour or depth change around the ear contour. Figure 2.6 shows the elliptical active contour and the segmented ear.



FIGURE 2.6: Ear's outline is detected via an active contour which is centred at the ear pit [88].

3D Shape Matching; ICP [8] has been widely used for 3D shape matching due to its simplicity and accuracy, however it is computationally expensive. Given a probe point set P and a gallery point set G , ICP iteratively calculates the rigid transform T that best aligns P and G . At the i^{th} iteration, the transform T_i is the transform that minimizes the mean square differences between the corresponding points of P_{i-1} and G . The corresponding points are the closest points between the two point-sets. P_{i-1} is then updated using T_i .

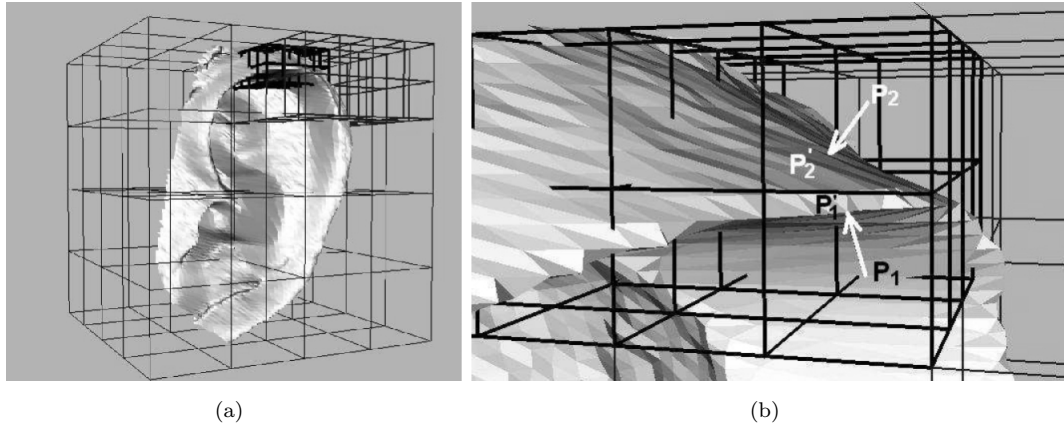


FIGURE 2.7: Voxelization; (a) 3D image space is partitioned into voxels. (b) Two voxel centres P_1 and P_2 and their closest points on the gallery surface P'_1 and P'_2 [89].

Yan et al. [89] have developed an efficient ICP registration method called “Pre-computed Voxel Closest Neighbours” which exploits the fact that subjects have to be enrolled beforehand for biometrics. Since the most time consuming task of the ICP algorithm

is finding the closest points between the probe and the gallery (of order $N_P \log N_G$) the main idea of this method is to approximate each point of the probe with a nearby point whose nearest point in the gallery point set is pre-computed. In this, they have proposed a quantized 3D volume using voxels, as shown in Figure 2.7. Placing the 3D probe image into this volume, each point of the probe falls into a voxel. Each probe point is then approximated by the voxel centre wherein it is placed. For each voxel the closest point in 3D space on the gallery surface is computed ahead of time. Figure 2.7 shows two voxel centres P_1 and P_2 and their closest point on the gallery.

2.4.2 Chen and Bhanus' Local Surface Patch Approach

Chen et al. [14, 15] have also tackled the 3D ear biometrics problem using a Minolta range scanner to acquire a 3D dataset of 52 subjects with two images per subject. In this, the ears are detected using template matching between a model template, based on the helix and antihelix curves of a single ear image, and the edge clusters of an input sample. A number of feature points are then extracted using local surface shape information. A signature called a "Local Surface Patch" based on local curvatures is computed for each feature point and is used along with ICP to achieve a recognition rate of 90.4%

Feature Point Extraction; Shape index, S_i , is a quantitative measure of surface shape [26], calculated from principal curvatures. Shape index takes a value in the $[0,1]$ interval and classifies surface shapes as one of nine basic types associated with this value. Shape index is defined as:

$$S_i(p) = \frac{1}{2} - \frac{1}{\pi} \tan^{-1} \frac{k_1(p) + k_2(p)}{k_1(p) - k_2(p)}, \quad (2.4)$$

where k_1 and k_2 are the maximum and minimum principal curvatures respectively. The points with local minimum or maximum shape index are then selected as feature points.

Local Surface Patch; Describing each feature point, P , a local surface patch (LSP) captures properties in a neighbourhood of points, N , around the feature point which are close enough to the feature point in Euclidean distance and surface normal;

$$N = \{N_i : \text{pixel } N_i, \|N_i - P\| \leq \varepsilon_1, \text{acos}(n_p \bullet n_{n_i}) < A\}. \quad (2.5)$$

For each feature point, shape index values of its neighbourhood, N , and the dot product

of surface normal vectors of the feature point and its neighbouring points are computed, and accumulated in a 2D histogram. The 2D histogram accumulates this information in bins along two axes. The shape index with a range of $[0,1]$ and the dot product of surface normal vectors with a range of $[-1,1]$ constitute the two axes of this histogram. A surface type of “concave”, “convex”, or “saddle” is also assigned to each LSP. Taken together, the 2D histogram, the surface type and the centroid of the local surface patch make up a distinctive signature for each patch.

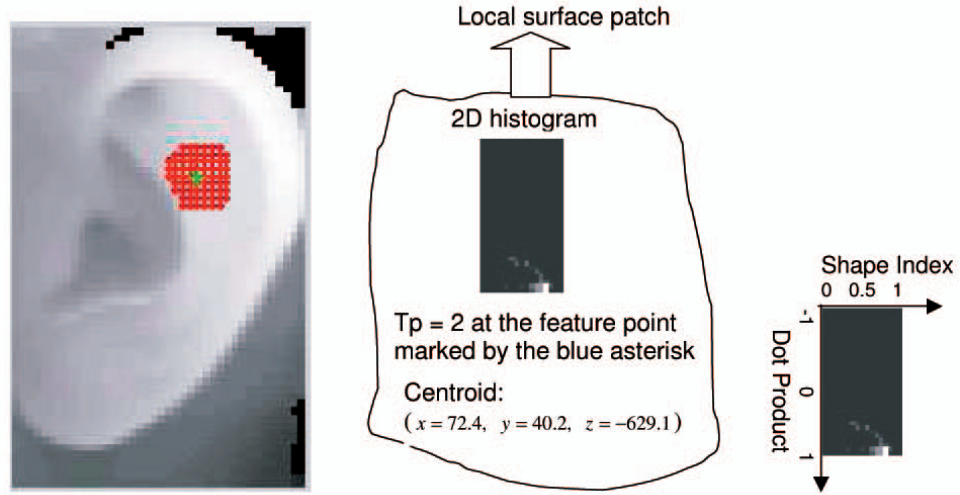


FIGURE 2.8: Local Surface Patch. The LSP constitutes a characteristic signature consisting of a 2D histogram, a surface type, and a centroid [14].

Recognition is performed in two stages. First a coarse alignment between the probe and a gallery image is obtained using the LSPs, and then fine alignment is achieved via ICP. Probe images are compared against all images in the gallery; each comparison is started by identifying the best match for each probe LSP in the gallery image. Assuming that the true set of matches, which pairs the patches that depict similar features, is a subset of all matches, a geometric constraint is applied to divide the matches into groups where each pair of matches in a group must satisfy the following condition,

$$d_{C_1, C_2} = |d_{P_1, P_2} - d_{G_1, G_2}| < \varepsilon_2 \quad (2.6)$$

where $C_1 = \{P_1, G_1\}$ and $C_2 = \{P_2, G_2\}$ are two matches between the probe and a gallery image patches, P_i and G_i , respectively. d_{P_1, P_2} and d_{G_1, G_2} are the Euclidean distances between patch centroids. The above constraint guarantees that a group of matches preserves the mutual position of the patches. In other words, d_{P_1, P_2} should be

consistent with d_{G_1, G_2} . Note that with this definition a match can be placed in more than one group. The biggest group is then declared as the true match subset.

Starting with an initial rigid transform based on the true match subset, ICP is applied to find the refined alignment between the probe and the gallery image. Having compared the probe to all the gallery images, the gallery image with least root mean square (RMS) disparity is declared as the matching sample.

2.5 Acoustic Ear Recognition

Akkermans et al. [4] have exploited the acoustic properties of the ear for recognition. By virtue of their structure, the auricle and the external ear canal behave like a filter and modify sound signals. A sound signal played into the ear is characteristically transformed and reflected dependent on the specific structure of the individual's ear. Thus this acoustic transfer function can be regarded as an acoustic ear signature. An obvious commercial use is that a small microphone might be incorporated into the earpiece of a mobile phone to receive the reflected sound signal and the existing loudspeaker could be used to generate the initial signal.

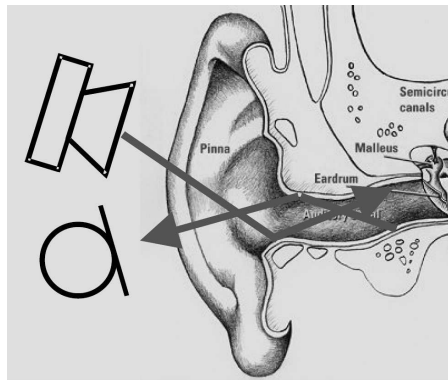


FIGURE 2.9: An ear signature is generated by probing the ear with a sound signal which is reflected and picked up by a small microphone. The shape of the pinna and the ear canal determine the acoustic transfer function which forms the basis of the signature [4].

Akkermans et al. measure the impulse response of the ear by sending a noise signal $n(t)$ with a spectrum $N(\omega)$ into the ear and measuring the response $r(t)$. Next, the response is transformed into the frequency domain by using an FFT to calculate the output frequency spectrum $R(\omega)$. Finally, an estimate is obtained of the transfer func-

tion $H(\omega) = R(\omega)/N(\omega)$, where $H(\omega)$ is the cascade of the transfer functions of the loudspeaker, pinna and ear canal, and microphone as shown in Figure 2.10.

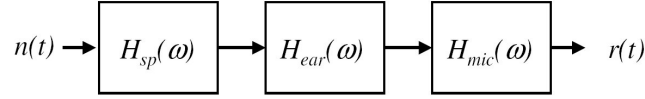


FIGURE 2.10: Calculating the impulse response of the ear [4].

The test dataset consists of 8 ear signatures collected from each of 31 subjects using headphones and a separate set of 8 signatures from 17 subjects using a modified mobile phone with a small microphone incorporated into the earpiece. The correlation metric,

$$C = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|} \quad (2.7)$$

was used for comparison where \mathbf{x} and \mathbf{y} are the feature vectors taken relative to the mean of the population. Using Fisher LDA, equal error rates of 1.5% - 7% were obtained depending on whether headphones or mobile phones were used.

2.6 Ear in Forensics

In forensics, the use of morphological properties of the external ear anatomy is mainly in the form of earprint identification and occasionally identification via images of the ear captured in surveillance cameras. Earprints, which may be found in up to 15% of crime scenes [68], are latent prints left behind as a result of the ear touching a surface, for example while listening at a door. In a legal context, the evidence regarding earprints could be utilized for various purposes including: dismissing a suspect, increasing evidence against a suspect or identifying possible suspects [55]. Dutch courts have reportedly admitted numerous cases of earprint related evidence [24], while in a number of cases the reliability of earprint evidence has been challenged [50, 19]. This is largely due to debate about the effect of pressure deformation, which does not affect biometric deployment. Dessimoz et al. [25] discussed the similarities, disparities and connections between biometrics and forensic sciences, considering various modalities including the ear.

Due to the different elevations of the external ear parts, some of the ear components

are commonly missing in earprints. Owing to the missing information, it can be debated whether earprints present less variability than ear images [25]. Figure 2.11 shows some sample earprints. The consequences of arbitrary pressure with which the ear is pressed onto the surface, to produce the print, can also cause some complications. Due to their different elevation and flexibility, ear parts react differently to the changes in pressure and cause large intra-individual variations. Considering this inherent property in missing components and potentially large intra-individual variations, earprints pose a more challenging recognition problem than ear images, and they introduce an interesting avenue for future research.

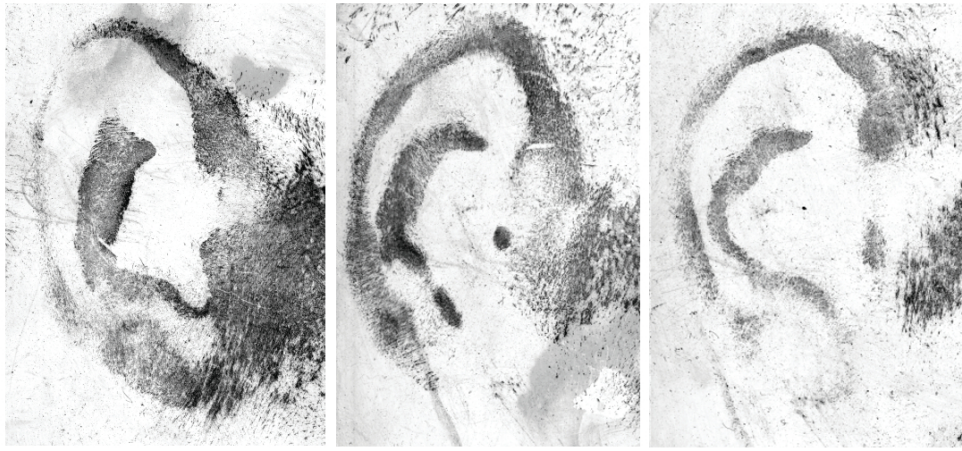


FIGURE 2.11: Sample earprints (from Meijerman's doctoral thesis [54]).

Meijerman et al. [55] have provided an overview of the earprint identification process. This study and most earprint recognition methods are based on manual analyses carried out by an expert. Alternatively, in a proof of concept paper by Rutty et al. [68], having identified the advantages of computerizing this process, they introduced a preliminary computerized earprint recognition system and discussed the problems hindering such approaches. Alberink et al. [5] have reported an equal error rate (EER) of 3.9% for laboratory quality prints and a 9.3% EER when simulated marks are compared against the prints database. In this, Alberink et al. have utilized computer vision techniques to classify the prints, although these are either manually segmented or annotated by experts.

2.7 Conclusions

In this chapter, we have reviewed the main contributions to ear biometrics. Starting with the Iannarelli's manual method for forensic recognition using ears, we then discussed the most prominent machine vision approaches, exploiting 2D images and 3D data. We also noted that new information regarding the ear structure can be obtained by acoustic analysis of the reflected sound signals from the ear, which may also be utilized for the purpose of liveness detection. The use of earprints in forensic sciences have also been mentioned as a closely connected field, which capitalizes on the structural parts of the ear that leave a print when the ear touches a surface.

Regarding the image-based methods discussed in this chapter, two main sub-tasks are distinguishable. First is the segmentation or localization of the ear, which is either manually or automatically performed. The second task is the matching or classification. In this work, we shall propose an automatic ear enrolment, and in classification, we opt for a feature extraction approach accompanied by a simple classifier as opposed to utilizing a more complex classification method on relatively raw data. We believe that this approach is intuitively more explicit in determining the information being exploited for recognition. It is worth noting that although there is no panacea in designing such systems, a fair balance in robustness and repeatability of these sub-tasks can help prevent undesirable bottleneck points. In the next chapter, we shall explain our ear localization method, where ears are automatically enrolled from a database of 2D head-profile images.

Chapter 3

Enrolment

The enrolment stage prepares an acquired sample for feature analysis and classification, by separating the object of interest from the background. Automatic enrolment is an essential part of any automated biometric system. It processes the raw data, and it deals, more so than the subsequent tasks, with the traits of the given dataset and is tuned accordingly. Non-invasive biometrics such as gait, face and ear are often acquired with standard equipment and include unnecessary background information. The input samples of an ear biometric system are often head-profile images which also include the eyes, mouth, nose, hair and a background.

Yan et al. [88] developed an automatic ear biometric system, which utilizes both 2D colour images and 3D range information of the human head-profiles. Their enrolment is a multistage process which uses the 2D and 3D data. Via curvature estimation, the ear pit is detected which is then used to initialize an elliptical active contour to locate the ear outline and segment the 3D ear data. Chen et al. [14] also use 2D and 3D face-profile images and the ears are detected by aligning the edge clusters with an ear model based on the helix and antihelix curves of a single ear image. Alvarez et al. [6] fits an ovoid model to the ear contour. However, their algorithm initializes with a manual estimate of the ear contour. Recently, Bustard et al. [12] have proposed a robust registration technique for 2D ear images addressing problems such as pose variation and clutter, and Islam et al. [42], inspired by an AdaBoost approach to face detection, used a similar method to detect the ears.

The precision of enrolment varies in these methods; some studies have opted for exact

segmentation [88, 6], while some use rectangular image patches containing the ear [14, 40, 42]. In this, the enrolment corresponds directly to the needs of the subsequent tasks. For example, exact segmentation is necessary, if matching does not distinguish between the relevant and irrelevant information. This is often a design concern which exploits the trade-off between accuracy of enrolment and complexity of matching. As mentioned before, due to the inherent tendency of the ear data to be occluded, one of our main concerns is to achieve better robustness in occlusion. Evaluation of occlusion tolerance has not been addressed by most of the current ear biometric approaches. However, recently, Islam et al. [42], showed that their AdaBoost detection approach has good occlusion tolerance properties, while Yan et al. [88] merely suggest that their method is capable of handling small amounts of occlusion by hair.

We propose the use of the Hough Transform (HT), which can extract shapes with properties equivalent to template matching and is tolerant of noise and occlusion. In this, HT for ellipses is used to find the elliptical shape of the ear in 2D head-profile images. Since other ellipses might also be present in the image and furthermore the ear does not have a perfect elliptical shape, the use of HT for ellipses produces various false positives. Therefore, we also add a number of refinements to the ellipse detection and some customizations distinguishing the ear ellipse from other erroneous detections. A rectangular image patch is then cropped localizing the ear. We shall report the enrolment results on two standard databases: *i*) a subset of XM2VTS [56] face-profile database with the total number of 764 images; *ii*) the UND ear database [86] with 942 images. The details of the enrolment performance is discussed in chapter 7 along with the description of the datasets. In this chapter, we shall describe our ear enrolment technique, which is based on a reduced version of the Hough Transform for ellipses and incorporates some prior knowledge regarding the specifications of the expected ear ellipse and also enhances the accumulator space helping to distinguish valid local peaks.

3.1 Hough Transform for ellipses

The Hough Transform (HT) can extract shapes with properties equivalent to template matching and is tolerant of noise and occlusion. As such, it appears an attractive choice for the initial location of the ears. However, known drawbacks of the HT include need for much memory and its high computational requirement. Hence, a reduced HT appears

suitable [3]. In this the centre is determined by using known ellipse properties. In these properties, given two points whose tangents intersect, then the line between the intersection point and their midpoint will pass through the centre of the ellipse. Figure 3.1 illustrates this with two points P1 and P2.

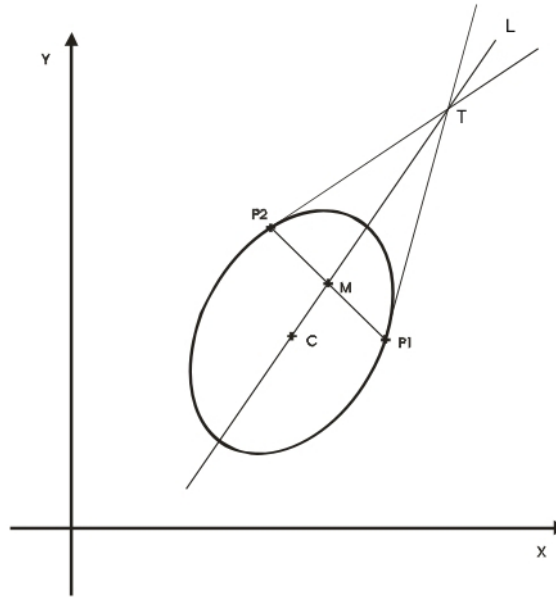


FIGURE 3.1: Tangents of the two points P1 and P2 intersect at the point T. The line L which passes through the intersection point T and the midpoint M, also passes through the ellipse centre C.

For application to ear images, first a smoothed edge detected image is derived using the Canny operator, and this is used to vote into an accumulator for the centre of the ellipse. The accumulator is constructed by voting for the lines which pass through the putative ellipse centre. When this process is completed, the locations of the peaks in the accumulator give the coordinates of the centre of the best matching ellipses.

This reduced HT for ellipses decreases the computational complexity by using known geometrical properties of the ellipse to decompose the parameter space. The space which would otherwise be a 5D has been reduced to a 2D space, gathering evidence for putative ellipse centres whilst marginalizing on scale and orientation, which causes ellipses of arbitrary scale and orientation to cast votes in single cells of the accumulator. Hence, some erroneous peaks may appear in the accumulator. Thereby while this method is effective in easing the computational and memory requirements, it may impair the accuracy.

3.2 Eliminating Erroneous Votes

Applying the reduced Hough transform to a profile image of the human head, results in an accumulator which usually presents a number of peaks. To detect the peak corresponding to the ear, we use various geometrical cues to eliminate some of the votes or even some of the voting pair points which do not comply with the properties of a prospective ear rim ellipse.

3.2.1 Vertical Ellipses

Given a non-rotated head profile image, the ear usually appears as a vertical ellipse with a vertical axis approximately twice as large as the horizontal axis. This can be described as,

$$\frac{(x - x_0)^2}{a^2} + \frac{(y - y_0)^2}{b^2} = 1, \quad b = 2a \quad (3.1)$$

where equation (3.1) denotes a non-rotated ellipse where a and b are the ellipse size along each axis and (x_0, y_0) is the centre.

By differentiation,

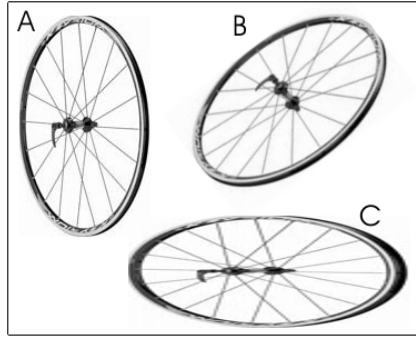
$$\frac{\partial y}{\partial x} = -\frac{(x - x_0)b^2}{(y - y_0)a^2}. \quad (3.2)$$

The edge direction can be arranged to equal $\frac{\partial y}{\partial x}$ and thereby a vote cast for a putative ellipse centre (x_0, y_0) by a point (x, y) is allowed in the accumulator if

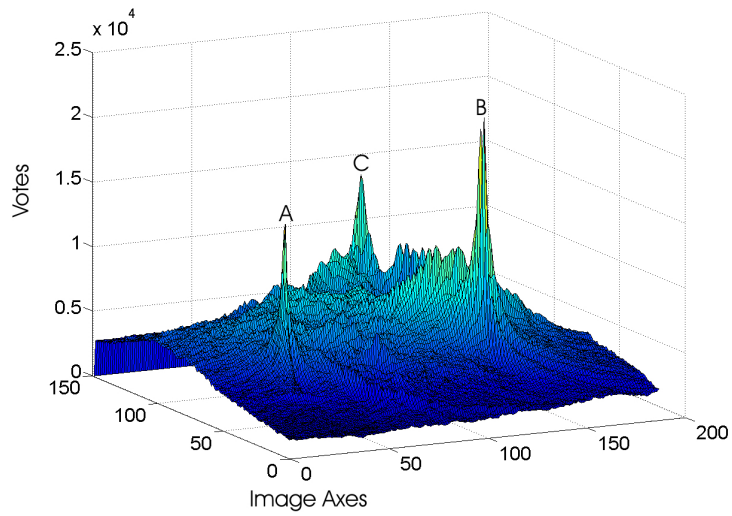
$$\frac{\partial y}{\partial x} + \frac{(x - x_0)b^2}{(y - y_0)a^2} \leq \text{threshold}. \quad (3.3)$$

This filter is set to ignore discrepancies less than a threshold value since: *i*) some ears are slightly rotated; *ii*) the exact value of $\frac{b^2}{a^2}$ for each ear is not known; *iii*) edge direction information cannot be accurately calculated. The threshold value is empirically set to 10 pixels. For this, values between 2 to 12 pixels have been examined to obtain the best signal to noise ratio.

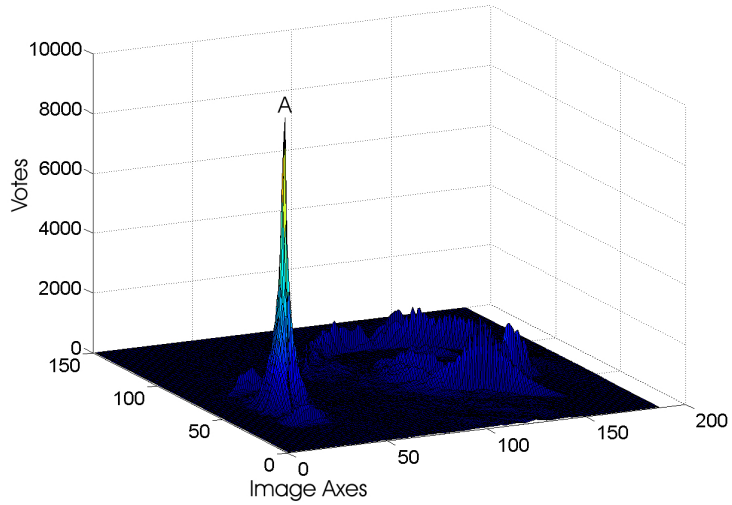
Figure 3.2(a) shows three wheels appearing as ellipses rotated and sized differently. Using the reduced HT, the resulting accumulator is depicted in figure 3.2(b), having three peaks corresponding to the three ellipses. Figure 3.2(c) shows the accumulator after filtering. In this, the peaks corresponding to non-vertical ellipses have been removed. Note that



(a) The three wheels



(b) The corresponding accumulator



(c) The accumulator after non-vertical ellipse filtering

FIGURE 3.2: Removing the non-vertical ellipses. (a) Ellipse A: a vertical ellipse with $b = 2a$. Ellipse B: the same ellipse rotated 45° . Ellipse C: a horizontal ellipse with $a = 3b$. (b) Three peaks are presented corresponding to the three ellipses. (c) After filtering only the peak corresponding to the vertical ellipse with $b = 2a$ has remained.

the noise level has also been reduced as a result of this filtering.

3.2.2 Selective Pairing

Given that a pair of edge pixels are within a scale window which controls the size of the target ellipse, they will vote for putative ellipse centres along the vote lines, as in figure 3.1. Aiming to be more selective, we use the edge direction information to determine whether or not a pair of points can possibly be part of an ellipse.

Figure 3.3 depicts our approach to rejecting invalid pairs. In this, the tangent, l , of an edge point, p , divides the image plane into two regions. l is also tangent to any ellipse containing p . Considering the edge direction at p , we speculate as to the location of the putative ellipse, and a section of the image – the gray region – is marked as being outside of the ellipse. Thereby, only the points in the white region can be considered as the pairing counterparts of the point p .

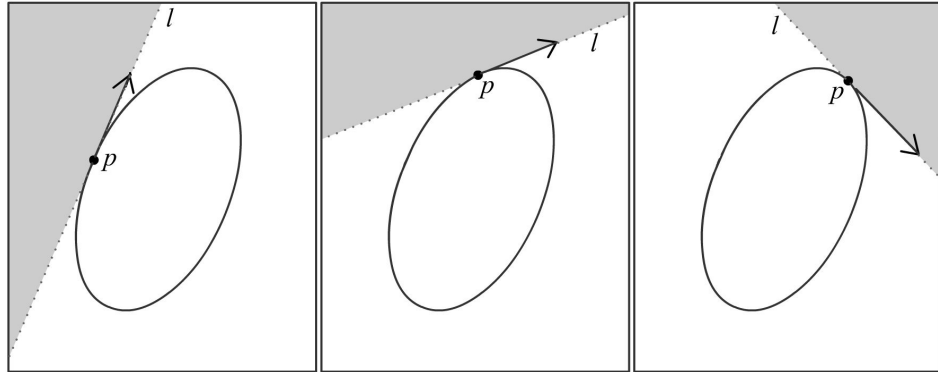
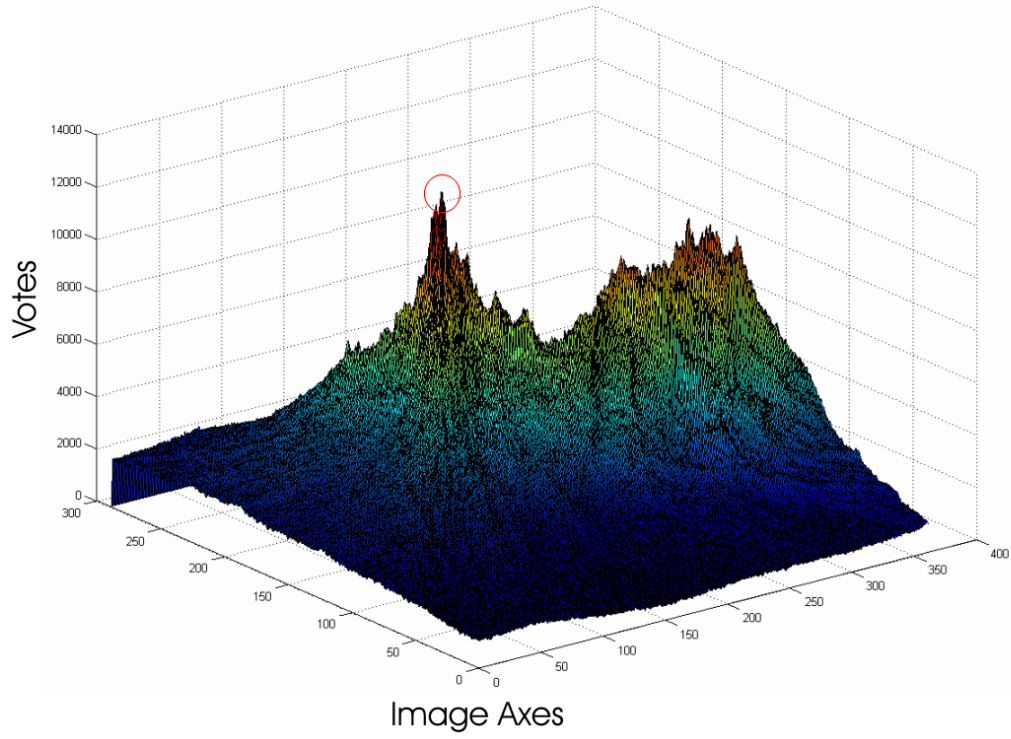
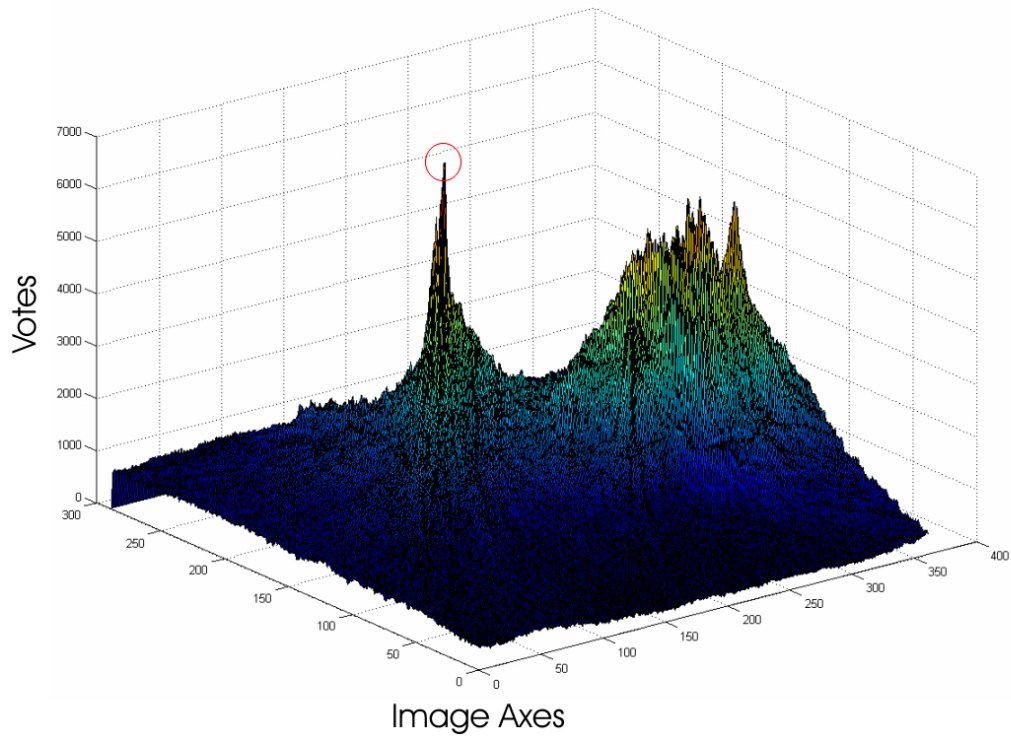


FIGURE 3.3: Distinguishing some ellipse free regions (the gray regions), to pose more selectivity on edge points pairing.

Figure 3.4 shows the improvements obtained using this refinement. A head-profile image from XM2VTS database and the edge detected map using Canny operator are shown in figure 3.5, and the HT accumulator is in figure 3.4(a). This accumulator is noisy and many points obtain relatively high votes. As seen in figure 3.4(b), filtering the pairs, via the new method, reduces the noise level of the accumulator and the peak corresponding to the ear becomes more pronounced.



(a) The accumulator corresponding to figure 3.5



(b) The same accumulator with selective pairing

FIGURE 3.4: The refined accumulator, using the selective pairing, exhibits less noise. Signal to noise ratio (calculated by dividing the value of the peak by the mean value of the accumulator) has increased to 6.36 in (b), where selective pairing is used, from 4.44 in (a). The peak corresponding to the ear is indicated with a red circle.

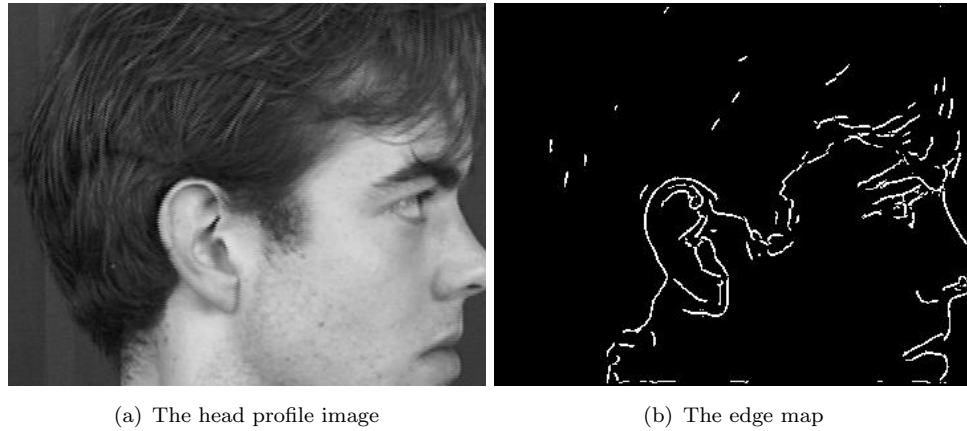


FIGURE 3.5: (a) A head profile image from the XM2VTS database [56] (the approximate head region is detected from the blue background). (b) The corresponding edge map, using Canny operator.

3.2.3 Spectacles and Anticipated Errors

Frames

The lenses of spectacles may appear as ellipses when viewed from the side, this challenges the presumption of the ear being the most prominent ellipse shape in the head profile images, see figure 3.6 for examples, and thereby may cause mis-localization. Knowing the direction of the faces in images, one third of the front of the faces is cropped to eliminate these elliptical shapes.



FIGURE 3.6: The elliptical shapes of spectacles from the side-view of the face (from XM2VTS data [56]).

Handles

Inspecting example accumulators generated by the reduced HT for ellipses, shows that the dominant orientation of the accumulator vote lines in a region mimics the dominant edge direction of the corresponding region in the image. For example, the vote lines cast by ear rim edges are mostly vertical.

Given a region with numerous edge points with similar edge directions, many pairs vote

in a bounded space along vote lines with similar orientations. This leads to a ridge of votes, and possibly a peak. The case of spectacle handles brings about a similar scenario; these handles usually pose many points in the edge detected map, which all tend to have a horizontal edge direction. Figure 3.7 shows such an edge map and the corresponding accumulator. To remove the destructive effects of this type, bearing in mind that the ear ellipses are mostly vertical thus their dominant vote lines are also vertical, we eliminate the vote lines which are close to being horizontal. Our datasets do not contain significantly rotated images, consistent with a subject looking forward.

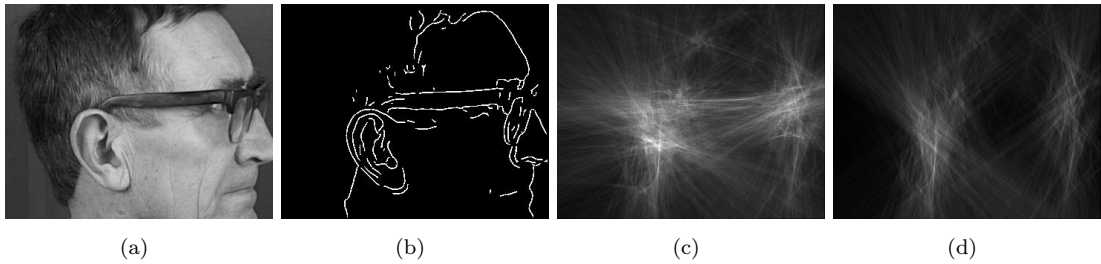


FIGURE 3.7: The destructive effect of spectacle handles. The accumulator for the edge map, shown in (b), is in the image (c) and it is filtered to eliminate the horizontal vote lines in (d). (a) is from the XM2VTS database [56].

3.3 Ear Enrolment

After these pruning stages, the ear centre is usually presented by a wide peak which has less height than expected, due to the fact that the ear is not a perfect ellipse and the edge direction information is not as accurate as desired. Occasionally, single peaks also appear in the accumulator, which may have been caused by curls in the hair. To remove the single peaks and find the centre of the widest peak which represents the ear, a 30×30 pixels averaging template is applied to the accumulator. For this, templates as small as 5×5 and as large as 50×50 are examined to produce the best aligned ear images. A 150×120 pixels image patch containing the ear is then cropped.

We use our algorithm to enrol ears from XM2VTS [56] and UND [86] data. Both these databases include indoor, 2D head-profile images. Figure 3.8 exhibits examples from XM2VTS and UND datasets along with the enrolled ear. In this, we have also included two occluded ears from XM2VTS. The XM2VTS database was not specifically targeted to the acquisition of ear images, and in many cases ears are partially or fully



FIGURE 3.8: Ear enrolment; first and second row show samples from XM2VTS [56] and UND [86] respectively, while the third row shows two examples of occluded ears from XM2VTS and the enrolled image.

occluded by hair. Hurley et al. [40] have derived a dataset of 63 individuals from the XM2VTS for their ear biometric analysis. This dataset is balanced, having four images per individual which are all taken from the left profile. Thus this set comprises $(63 \times 4 =) 252$ images. The ears in this database are fully visible and have not been obscured by hair. We also extract a further 269 images from XM2VTS and extend the number of individuals to 150. In our newly added data, each individual is represented by at least two images of the left or right ear. We also derived a separate 361-image dataset from other samples within XM2VTS which includes considerably occluded ears. This set is assembled to examine the impact of occlusion by hair. Our automatic enrolment finds the ear in $517/521 (\sim 99.2\%)$ images of the main data and in $346/361 (\sim 95.8\%)$ images of the occluded dataset. The mis-enrolled ears are shown in figure 3.9. The algorithm occasionally detects alternative elliptical shapes, like the ring-shaped earring in figure

3.9(b) or curls in the hair. However, more often the error is caused by an increased level of noise which can be caused by much hair or beard around the ear. This affects the accuracy of detection, producing misaligned samples.



(a) Main dataset



(b) Occluded dataset

FIGURE 3.9: The mis-enrolled ears of the XM2VTS-based dataset.

The UND ear biometric database [86], with 942 images from 302 subjects, also includes head-profile images, and it contains both 2D and 3D data. However, in this the emphasis is on the 3D scans, and the 2D images are of considerably lower quality compared to XM2VTS samples. In enrolment, given that the UND images include much more background information, we use skin detection in colour images and the Canny edge operator to detect the face region of the images. The method proposed by Hsu et al. [34] was used for skin detection. The head region is more easily detected in XM2VTS samples due to the blue background. Enrolment, as described, was then performed on the face region of the images. For UND data our algorithm offers a 91% enrolment success rate. In this, the edge points of the ear rim are less accurately detected which has an impairing effect on an evidence gathering algorithm such as HT which works with edge points.

3.4 Conclusions

The elliptical shape of the ear in profile images of the human head makes it possible to detect the ear by a Hough Transform. The use of the reduced HT for ellipses has produced various false positives. Therefore, we have also added a number of refinements to the ellipse detection and some constraints distinguishing the ear ellipse from other erroneous detections. Our new ear enrolment method achieves a 99% successful enrolment rate on our main data, comprising 458 images. Using a reduced HT the method is not constrained by the high computational and memory requirements of the 5D HT for ellipses and given the properties of the HT in handling noise and occlusion it obtains satisfactory results enrolling occluded ears.

In the next chapter, we shall examine the morphological properties of the human ear, using biological information and particularly the reports on the embryology of the ear. Interestingly, the formation of the ear in the human embryo is commonly discussed as the individual development of separate components. We shall later show that this provides an opportunity to describe and analyze the ear shape as a composite structure of local features.

Chapter 4

Ear Features – a Biological Approach

The human ear is an intricate structure of intertwined elevations of tissue, one which is hard to describe precisely. However, understanding this complex structure can help in designing a more efficient and effective approach to ear biometrics. In the absence of such understanding, holistic measures can be used, which provide us with general image information, albeit the task of distinguishing the relevant information is by a process of trial and error.

Identifying the components which make up the complex structure of the human external ear has been the main concern of embryological accounts studying the morphology of the external ear. We first took interest in ear embryology when we saw a report on the early development of the external ear in human embryos which showed that the external ear develops from six separate blob-like tubercles. Embryologists have long hypothesized as to the contributions from each of these six hillocks. In this, the extreme variations in some abnormal ears have offered vital cues. In biometrics, variation is the key to recognition capability. Even though there is no direct evidence to sustain the link between ear development and automated recognition, it can guide our approach and provide a basis for explicit evaluation of the proposed method.

The ear's embryology has not been previously linked with ear biometrics. It has not been studied so far to identify the components and those sites with big variations in conjunction with the ear biometric problem. In fact, the only structure-based ear recog-

nition algorithms do not go any further than to distinguish the oval shape of the ear's outer rim, and the question of 'which are the distinctive features' is rarely asked.

In this chapter, we shall review the different accounts of external ear embryology, and discuss with reference to the variations and components of the ear structure. In this, abnormalities of the auricle often offer vital cues. We shall illustrate the analysis with example ears from surgical studies and, when possible, ears from the database which we will be using. We start by reviewing the terminology of the ear's anatomy. In this chapter wherever we talk about the ear we mean the external ear.

4.1 Ear Anatomy

Figure 4.1 shows an ear with its various parts. The most prominent component of an ear is its outer rim called the helix, which merges into the lobe at the bottom. The antihelix is the rounded brim of the concha, and runs almost parallel to the helix. It forks into two branches at the top, forming the superior and the inferior cruses of antihelix. The concha is a shell-shaped cavity, surrounding the external auditory canal and merging into the incisura. The incisura has two little bumps on either side named the tragus and the antitragus. The external auditory canal is often concealed by the tragus. The concha is divided into two parts by the crus of helix which is the horizontal part of the helix, forming a transverse ridge on the floor of the concha. Immediately attached to the crus of helix is a portion of the helix which is called the ascending or anterior helix. Notice in figure 4.2, which shows a collection of ear images, that despite the inter-individual variations, these basic parts are distinguishable in most ear images.

The Darwin's tubercle ¹ and the Satyr tip are sites of possible irregularities on the helix. The Darwin's tubercle can be seen as either a tubercle or a small notch which corresponds to the true ear tip of the long-eared mammals, and is located at approximately one third of the way beyond the upper tip of the helix. The Satyr tip marks the tip of the ear toward the crown of the head. Unlike the aforementioned components, these two are not conspicuous in all ears. Two examples of these are shown in figure 4.3.

¹This feature is so called since Charles Darwin was first to speculate as to its origin.[35]

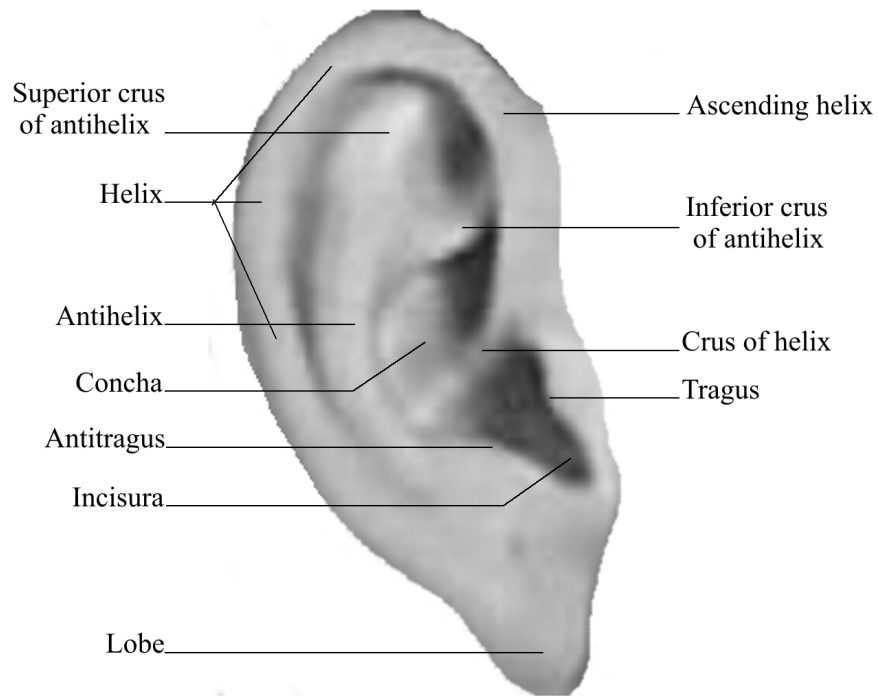


FIGURE 4.1: The anatomy of the ear

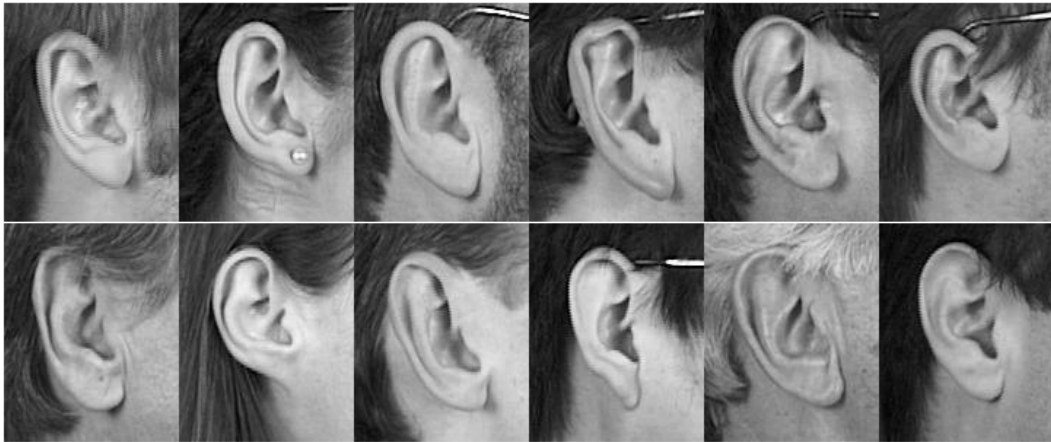


FIGURE 4.2: Notice the variety of ear shapes while all contain the basic components. Ears are from the XM2VTS database [56].

4.2 Ear Embryology

Six individual growth nodules develop to form the definitive human auricle. This is the reason for our interest in ear embryology — the premise of local and independent structures within the auricle is appealing to our classification purpose. In this section, we will first review the motivations behind the studies of ear embryology. We shall then consider the debate concerning the contributions of each of these six hillocks to the auricle structure.

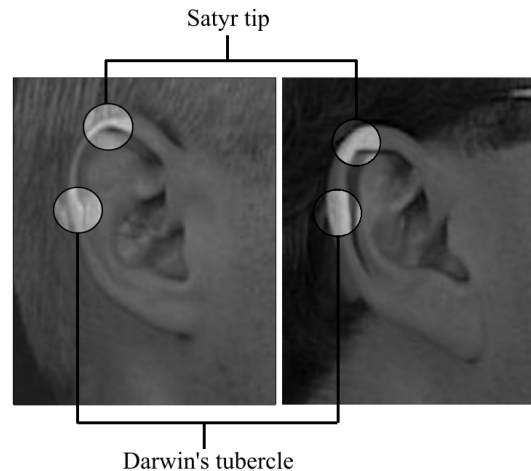


FIGURE 4.3: The Darwin's tubercle and the Satyr tip are two irregularities on the helix curve. They are not distinguishable on all ears. Notice that the Darwin's tubercle can not be accurately pinpointed in the ear on the right.

4.2.1 Motivations

Streeter [71] provided one of the most extensive accounts of external ear embryology, accompanied with images from differently aged embryos so that the reader could follow his observations. His aim was to determine the age of human embryos using the details of the external ear form as an objective criterion. He noted that length was an unreliable indication of the age, particularly in the first two months of the development, and that more accurate criteria were needed. The auricle can be used to date young embryos since it develops in the early stages of the embryogenesis; the hillocks which form the auricle, make their first appearance at the fifth week of gestation in the womb.

Understanding the origin of ear malformations has been another motive for ear embryology research. Some of the more recent works hypothesized as to the contributions of ear hillocks in an attempt to shed light on embryological derivations of ear anomalies. Better understanding of these anomalies leads to better surgical procedures to correct them, as noted by Davis [23]. He reviewed numerous congenital external ear defects and assigned their causes to anomalies in the development of different hillocks.

Microtia – meaning ‘small ear’ – is a congenital deformity of the auricle. In its severe form the external ear does not have its usual structural parts and is replaced by a small vestige or can even be totally absent (anotia). Many cases of microtia experience some degree of hearing loss. Investigations into the embryology of the ear have revealed that the external and middle ear have different embryonic origins from the inner ear. In the

majority of cases of microtia with hearing loss, the loss has conductive reasons, which are the result of malformations in the middle and outer ear causing a blockage or a narrowing of the ear canal, while the inner ear anatomy and functions remain intact and normal hearing might be restored by surgery [74].

Finally, correlations between the auricle anomalies and some specific syndromes may provide an aid to diagnosis. Most notable of these is the association between ear and kidney anomalies. Edith Potter [64] was the first to associate crumpled and flattened ears with bilateral kidney agenesis. Wang et al. [78] described various external ear malformations which suggest a greater risk of renal defects. The embryology of ear and kidneys has been studied in an attempt to describe these correlations [44].

4.2.2 Hillocks Hypothesis

The initial appearance of the external ear in the human embryo is in the shape of six individual hillocks occurring in the fifth week of embryonic life [71]. It is the unequal growth and coalescence of these six hillocks that gives the shape of the definitive auricle in a newborn baby. Figure 4.4(a) shows a six weeks old embryo with the auricular hillocks. The hillocks are numbered in the drawing 4.4(b). As illustrated in figure 4.4(c), the external ear originates from the tissue of the first and second branchial arches, called the mandibular and the hyoid arches. The six auricular hillocks consist of three mandibular hillocks and three hyoid hillocks. The mandibular and the hyoid arches are separated by the first branchial cleft, which gives rise to the external auditory canal.

Much of the literature regarding the ear formation is concerned with identifying the contributions from each of these six hillocks, and though they were first observed by His in 1882 [71], there is still disagreement as to the precise embryology of the external ear. Figure 4.5 summarizes the suggested arrangements by different authors, apportioning different hillocks and combinations to ear formation.

His [71] was the first to observe the six auricular hillocks in human embryos and to associate them with the components of the definitive auricle. Streeter [71] produced a thorough study of the development of the external ear. He recorded numerous images from the progressional development of the auricle from various specimens, and most of

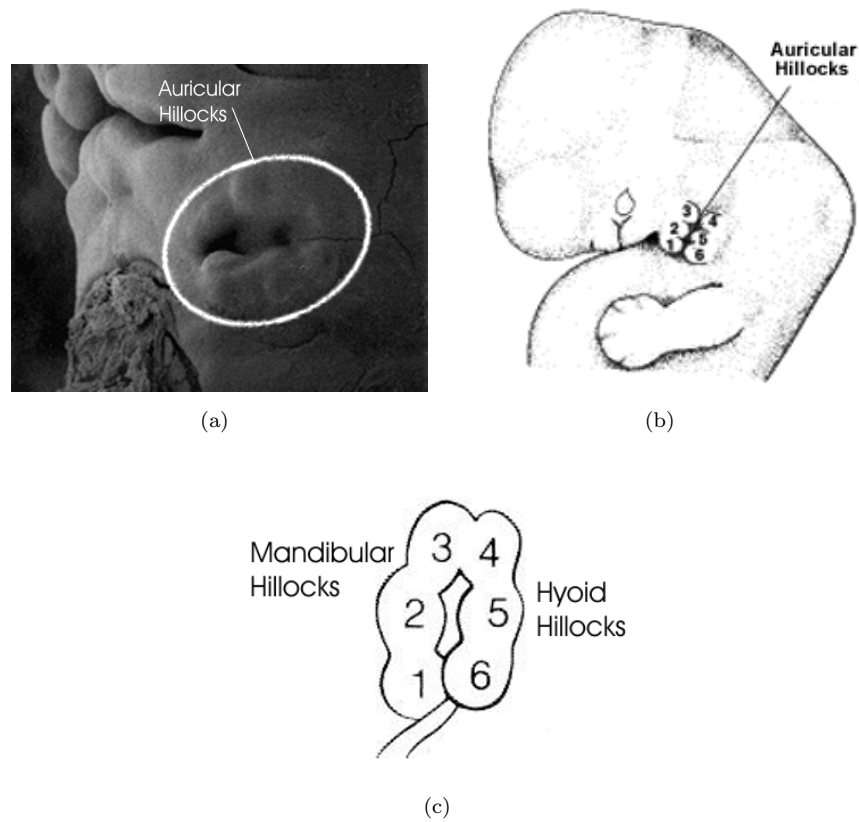


FIGURE 4.4: (a) Scanning electron micrograph of a 6 weeks old human embryo [72]. The 6 auricular hillocks can be seen in (a). They are more clearly illustrated in the drawing (b)[69]. (c) Mandibular and Hyoid hillocks [23]

his findings are still viable. His proposed arrangement is mainly based on his step-by-step observations of ear development (figure 4.5(b)). Wood-Jones et al. [81] suggested that the only mandibular component of the auricle (hillocks 1, 2 and 3) is the tragus (figure 4.5(c)). They based their argument on the evidence from two ear anomalies and also the nerve supply to the ear, which suggest a different source for the tragus from the rest of the auricle. In their study, they did not show much concern for the fate of the hyoid hillocks (4, 5 and 6). Davis [23] studied a wide range of congenital malformations of the external ear, and assigned the cause of every abnormality to defects of the auricular hillocks, in an attempt which was described by Hunter et al. [35] as ‘speculative’ (figure 4.5(d)). Sulik [72] studied the scanning electron micrographs of human embryos (figure 4.5(f)). The main point of the arrangement proposed by Park et al. [61, 62] was the assignment of a hillock exclusive to the lobe, explaining the malformation of the ear lobe (figure 4.5(g)).

Note that except the Wood-Jones et al.’s arrangement, which assigns three hillocks to

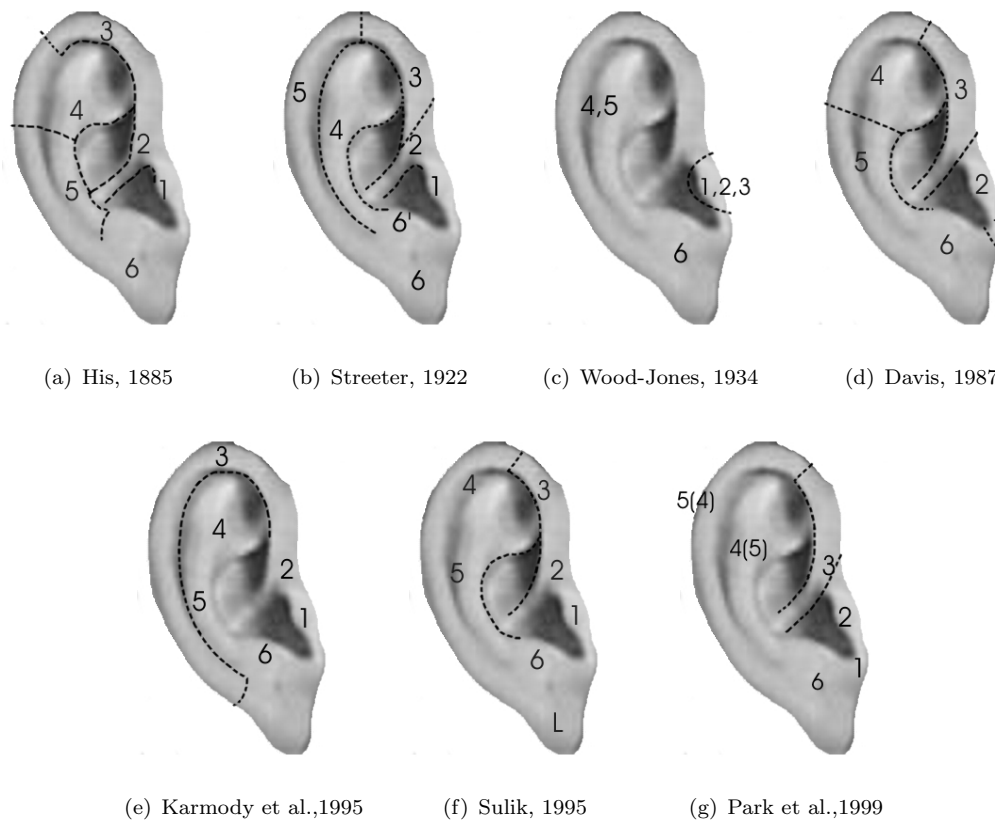


FIGURE 4.5: Differing final positions of the hillocks, suggested by different authors
(derived from: [35, 81, 23, 72, 61])

the tragus and perhaps the one by Karmody et al. [35] which assigns the origin of the helix to hillock 3, the rest of the hypotheses are similar. The disparities in these seem to be mostly in the regions of the helix and antihelix. It is worth noting that while embryologists hypothesize and argue about identifying the exact hillock which forms a specific component of the auricle, our concern is merely to identify a set of stand-alone components and also the sites at which we can expect big inter-individual variations. One of the main approaches to understanding the significance of hillocks and how they interact to form the definitive auricle is by studying the ear anomalies and variations.

4.3 Variations and Anomalies

Hunter et al. [35] provided an extensive survey of external ear syndromes and their embryological derivations. They noted that continued attention to these malformations combined with embryologic evidence may shed light on the embryology of the ear and the origin of the anomalies. Take for example one of the anomalies discussed in Hunter's

paper, a lethal condition, called otocephaly, which is accompanied by an anomaly of the auricle. In this condition, all the components of the external ear are present except for the tragus (figure 4.6), which suggests an isolated development process for the tragus, and thus supports the assignment of an individual hillock to the formation of the tragus.

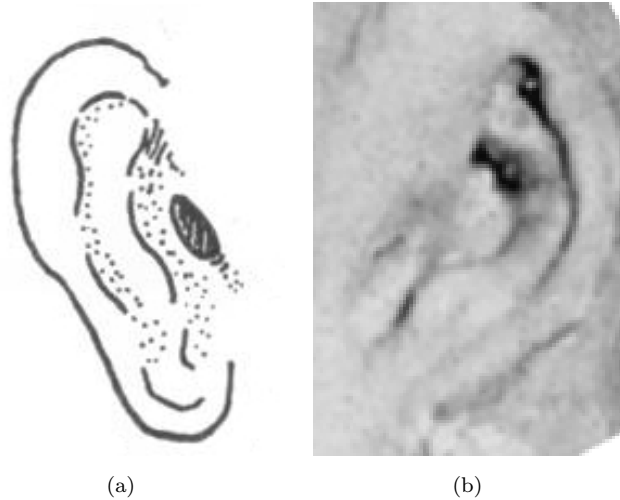


FIGURE 4.6: (a) A drawing of the ear in a case of otocephaly, described by Hannover. The ear was described complete with the exception of the tragus (Wood-Jones et al. 1934 [81]). (b) Another case of otocephaly: In this the external ear canal is missing as well as the tragus (Hunter et al. 2005 [35]). Although it has not been mentioned by the authors, it seems that the crus of helix is also missing or is underdeveloped.

Some malformations can be described as excessive growth beyond, or, underdevelopment beneath the thresholds of normality. Thereby the site of an anomaly of this kind is also where a big variance is introduced; it is unlikely that an abnormality will be observed in locations of constant structures. Take for example the crus of helix which is described by Streeter as one of the least varying parts of the auricle. In accordance with that observation, this part is not a common site for anomalies. In this section, we concentrate on anomalies which indicate specific individual components or hint as to the sites of big variations. For more information on these defects see [35] and [23].

4.3.1 Cleft Defects

Streeter argued against the significance of the auricular hillocks and suggested that the external ear comes into existence as an intact and continuous structure which elaborates into its final form. However, there is a wide range of defects which disturb the smooth

continuity of the ear structure, and clefts can be seen in various locations on the auricle. These can be best described as the failure of fusion or the lack of correct alignment of the various hillocks, which further indicates the role of separate structures in the formation of the auricle. Davis has many examples of ears with such cleft defects [23].

Figure 4.7(a) shows an example of a ‘detached upper helix’, examined by Park et al. [62]. In this, the helix is detached from the crus of helix, which can be attributed to the failure of fusion of hillocks 2 and 3 (as in the arrangement proposed by Streeter in figure 4.5(b)). There is also an abnormal confluence of the lower end of the helix and the inferior crus of antihelix. Park et al. ascribed this to a defect in the fusion of hillocks 3 and 4, which are the adjacent mandibular and hyoid components of the upper ear. Figure 4.7(b) shows an ear with a defect known as ‘upper helical irregularity’. In this, a cleft is observed near the position of the Satyr tip. Davis [23] suggested that this is the cleft between hillocks 3 and 4 (see Davis hillocks arrangement in figure 4.5(d)).

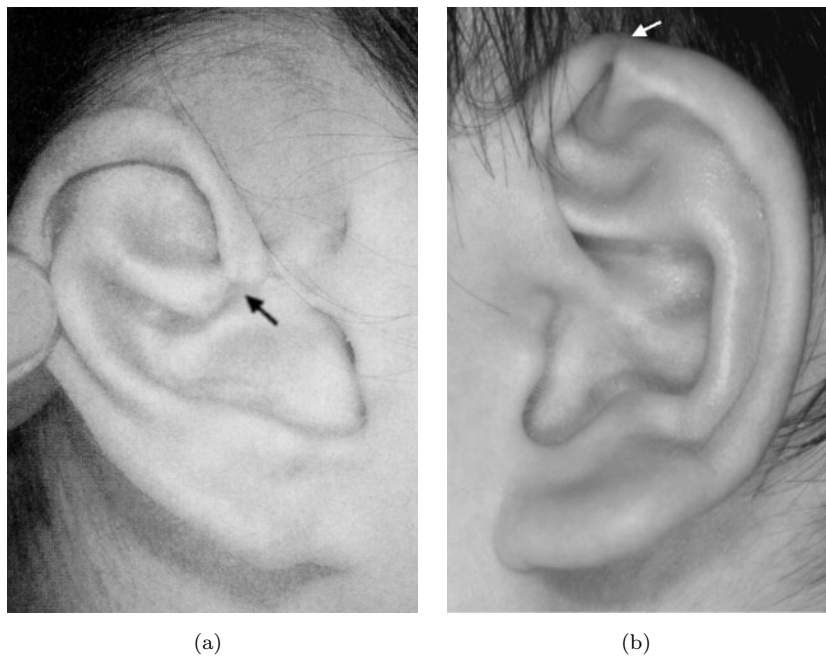


FIGURE 4.7: (a) Detachment of the upper helix (Park et al. 99 [62]). (b) A cleft is seen on the top of the ascending helix. The superior crus of antihelix also presents an abnormal wrinkle (Hunter et al. 2005 [35]).

Figure 4.8(a) shows an example of a ‘question mark ear’. The ‘question mark ear’ is a name given to a series of malformations involving the cleft between the helix and the lobe. The appearance of the malformed ears can vary. Park [60] has studied different cases of this defect. It is commonly suggested that the cleft between hillocks 5 and 6 is

responsible for this anomaly. The antitragus, which is also formed from the 6th hillock, can also be deformed or missing. Finally, as seen in figure 4.8(b), a cleft might be seen in the medial aspect of the lobe. Davis [23] and Park [61] both considered this to be the result of abnormal fusion of hillocks 1 and 6.

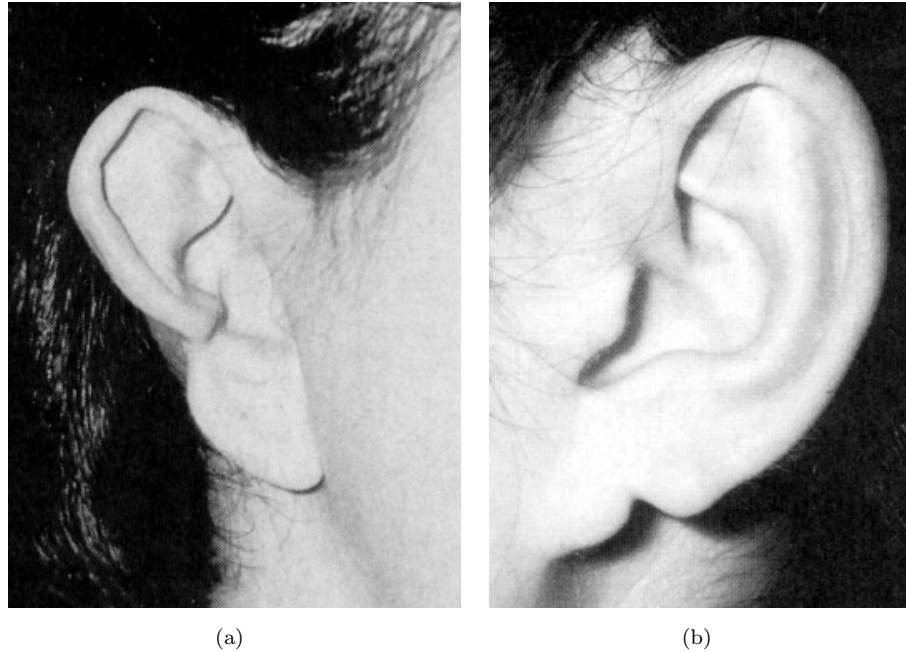


FIGURE 4.8: (a) A question mark ear (Park 98 [60]). (b) A cleft ear lobe (Park 99 [61]).

4.3.2 Helix and Antihelix

The helix and antihelix are commonly assigned to hillocks 4 and 5, and there are two major hypotheses regarding their formation: *i*) as was suggested by His, the upper and lower helical components, including the helix and antihelix, are derived from hillocks 4 and 5 respectively; *ii*) Streeter believes that a single hillock (5th) gave rise to the helix and that the antihelix is the product of hillock 4. For the purpose of this discussion, we will refer to the first hypothesis as anterior-posterior partitioning and the latter as the lateral partitioning. Figure 4.9 illustrates the two hypotheses.

Supporting His' hypothesis, the anterior-posterior partitions, Cryptotia is an abnormality of the external ear in which the superior portion of the auricle is hidden under the surrounding skin. Figures 4.10(a) and 4.10(b) show two examples of this defect. Cryptotia is associated with an underdevelopment of the upper ear which prevents the upper

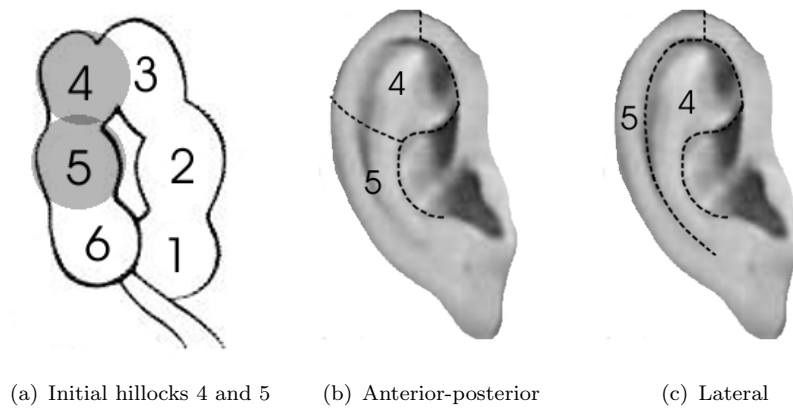


FIGURE 4.9: The two main hypotheses regarding the formation of the helix and the antihelix.

helical components from rising from against the head, thus driving them under the skin as growth happens. Another anomaly is the Satyr ear. In this, the ear has a sharp point, with a varied degree of folding forward, at the position of the Satyr tip (see figure 4.10(c)). Hunter et al. [35] noted strong intra-patient correlation, in left and right ears, and some intra-family association of Cryptotia and Satyr ear. Despite much variation in appearance of the two anomalies, Hunter suggested that both of these conditions are derivatives of underdevelopment and lack of tissue of the upper ear, ascribable to a major deficiency in hillock 4 [23]. Furthermore, we have noticed a number of individuals in our database whose upper ear appears to be excessively large compared to the lower components. Figure 4.11 shows four examples of this. Thus a considerable growth variation is observed in the upper helical components while the lower parts are relatively constant.

On the other hand, a seemingly discouraging point for Streeter’s hypothesis, of lateral partitioning, is that it is hard to see how the helix and antihelix could have been developed separately, since they are attached all along one side and each having only one free edge. An explanation could be offered via an additional component called the scapha. The scapha is the concave surface of free portion lying between the antihelix and the helix. Thus it provides a margin and allows the two to have some degree of independent development.

It is also interesting to note the Darwin’s tubercle, which is a small expansion at the position of the true ear tip, and can be found on many ears. Although it does not have any clinical importance and is not associated with any specific syndrome, it is still

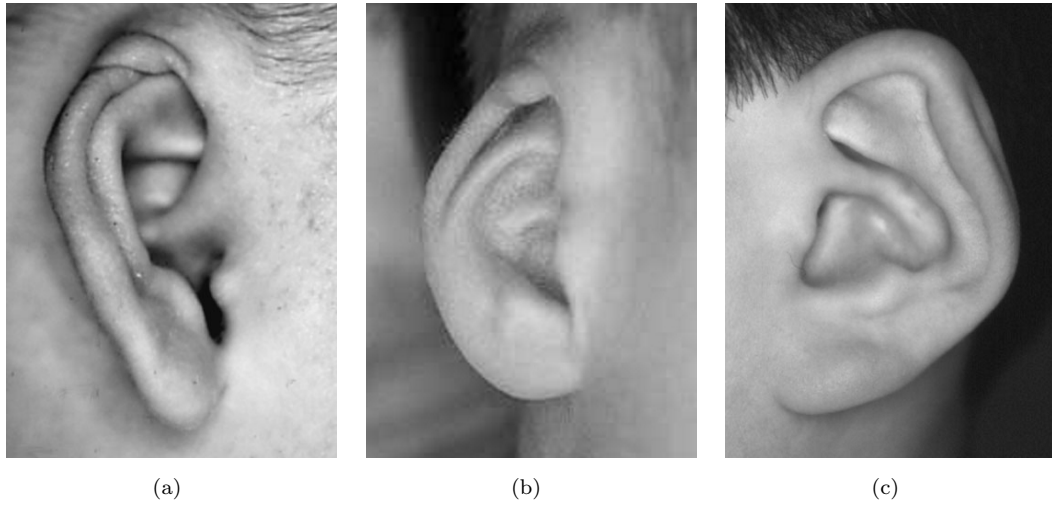


FIGURE 4.10: (a) and (b) are two examples of Cryptotia (Hunter et al. 2005 [35] and David Gault 2006 [32]). (c) a Satyr ear (Hunter et al. 2005 [35])



FIGURE 4.11: Examples from our database (XM2VTS, [56]) in which the upper ear is excessively large.

interesting for embryologists to hypothesize as to its origin, since this is roughly the point of separation between the upper and lower helix. Thus the speculations as to its origin are analogous to the hypotheses regarding the formation of the helix and antihelix. Davis [23], who has employed the anterior-posterior partitioning, regards the Darwin's tubercle as the point of fusion of hillocks 4 and 5. On the other hand, Streeter [71], who believes in lateral partitions, suggests that this expansion is a simple exuberance of hillock 5, and thus pinpoints the ultimate position of this hillock.

4.3.3 Ascending helix

One of the main points which has been captured in the arrangements suggested by both Streeter and Sulik is that they assigned an individual hillock to the ascending helix, distinguishing it from the crus of helix and the rest of the helix on the hyoid side. This

is supported by the anomaly observed in the ‘lop ear’, that describes an abnormality of the external ear in which the helix has been folded down to some degree (see figure 4.12). Davis has concluded that a deficiency or even the absence of hillock 3 has forced the remaining helix to move forward and bridge the gap [23]. Despite the deficiency of the ascending helix, the ear presents normal helix and crus of helix parts. However, excessive concavity of the concha and an arrested development of the intrinsic muscles are also suggested as possible causes.

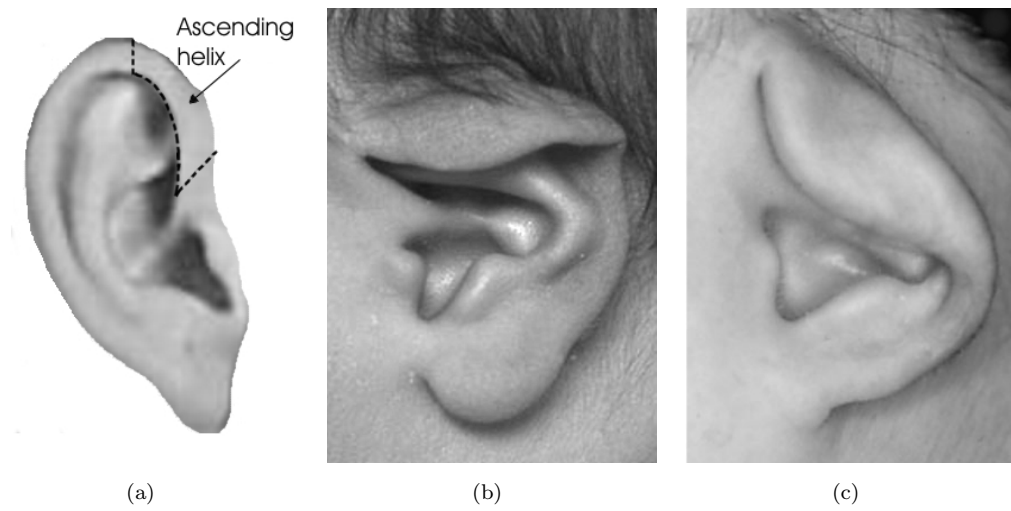


FIGURE 4.12: (a) One of the proposed causes of the lop ear is deficiency or absence of the ascending helix. (b) a lop ear where the folding is only above the Darwin’s tubercle (Aby [70]). (c) a more severe case (Hunter et al. 2005 [35]).

Two of the anomalies discussed as cleft defects, the ‘upper helical irregularity’ and the ‘detached upper helix’ (see section 4.3.1), provide further evidence as to the independence of the ascending helix. Both defects exhibit conspicuous clefts separating the ascending helix from the rest of the helical components on either side. Less conspicuous but more common, is the Satyr tip which also separates the ascending helix from the helix. Furthermore, Davis has presented examples in which the ascending helices are not conspicuous, while the rest of the auricle components have their normal shape.

4.3.4 Tragus

Otocephaly, which was mentioned at the beginning of this section, is a syndrome in which the tragus is missing. Other tragal anomalies may exhibit extensions or duplications of the tragus flesh, which might be accompanied by a separate or an underdeveloped lobe,

while the rest of the components are intact. Two ears with tragal anomaly are shown in figure 4.13. These anomalies can be indications of the independence of the tragus. Other evidence includes the nerve supply to the ear, which as was noted by Wood-Jones et al. [81], suggests a different source for the tragus from the rest of the auricle.

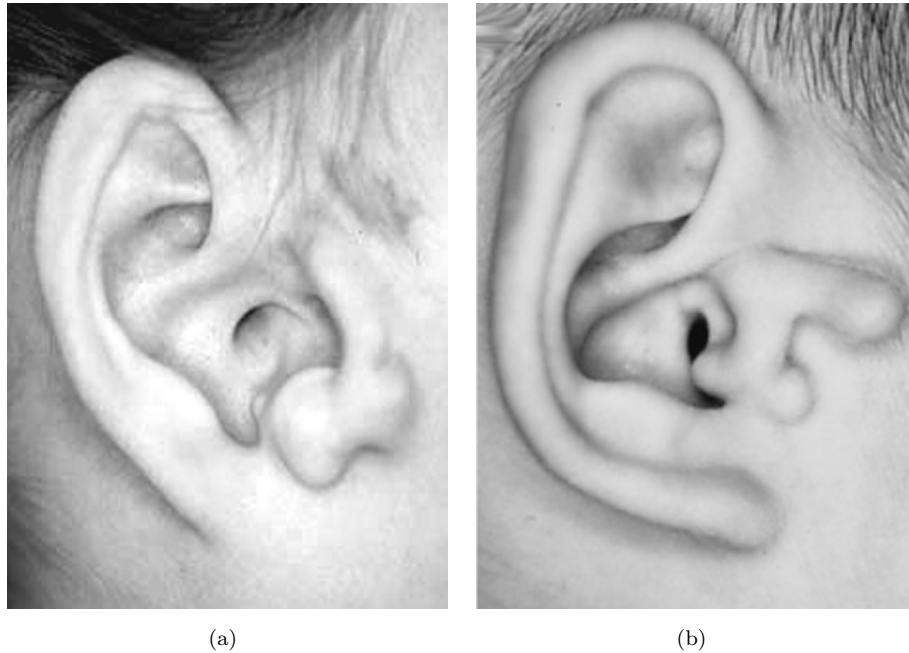


FIGURE 4.13: Tragal anomalies: The flesh of the tragus is extended or developed in abnormal directions. This has pushed the tragus out of its usual position and the opening to the external ear canal can usually be seen. ((a) is taken from Gore et al. 2006 [33] and (b) is from Hunter et al. 2005 [35])

4.3.5 Superior crus of antihelix

Stahl ear is an anomaly concerning the structure of the antihelix. In a Stahl ear, the superior crus of antihelix is not in its normal position or there is an additional crus present. This might be accompanied by an unrolled helix in some cases. Figure 4.14 shows three examples of Stahl ears, including an ear which we found in our own database (figure 4.14(c)). There is no consensus for the reasons behind the formation of Stahl ear. Various causes, including embryonic pressures and abnormalities of the intrinsic muscles have been suggested. It is also hypothesized that this might be caused by an abnormality in the coalescence of hillocks 4 and 5 [23].

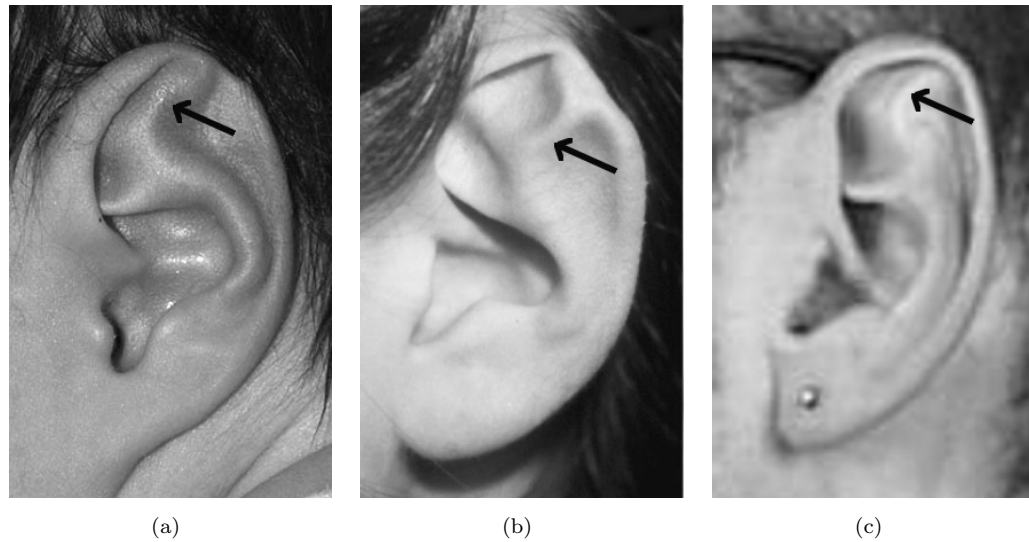


FIGURE 4.14: Stahl ears showing a third crus of antihelix. ((a) is taken from Aby [70], (b) is from Hunter et al. 2005 [35] and (c) is from our dataset [56].)

4.3.6 Lobe

The lobe is generally thought of as being derived from the lower end of hillocks 1 and 6 which form the tragus and the antitragus respectively. However, Park et al. [61, 62], who specifically studied the ear lobe malformations, suggested that hillocks 2 and 6 generate the tragus and the antitragus respectively, while the lobe is derived from hillocks 1 and 6. Davis also opted for a similar hillock contribution to the ear lobe [23].

The ear lobe is the only part of the ear which is not composed of cartilage, instead it contains fat. The shape of the lobe can vary from well-formed to attached. Whether the lobe is attached or not is an international standard for identification in Disaster Victim Identification (DVI) [76]. In some ears the lobe is almost non-existent. Figure 4.15 presents ear examples with different lobe shapes. However, the lobe seems to be the only part of the ear which continues to grow and change shape as the person grows older. Meijerman [54] looked at the lengthening of the auricle as the person ages and noted that the lobe appears to make up most of the increase. Note also that ear piercing, which is a semi-permanent body modification, was reported by Abbas et al. to occur in 46% of their population sample of 400 adults [1]. They reported that, in about 95% of the cases with ear piercing, the piercing occurs on the lobe. Although, as they noted, the presence or absence of such piercing itself is a useful attribute for forensic identification, we are more concerned with the clutter and occlusion which is caused

by wearing earrings. Therefore since the lobe’s shape varies with aging and also since we may frequently find the lobe to be occluded or cluttered by earrings, we do not pay much attention to this component of the auricle, although it could exhibit a variety of shapes, and in a database with a small time lapse between the captured samples it can be comparatively discriminant.

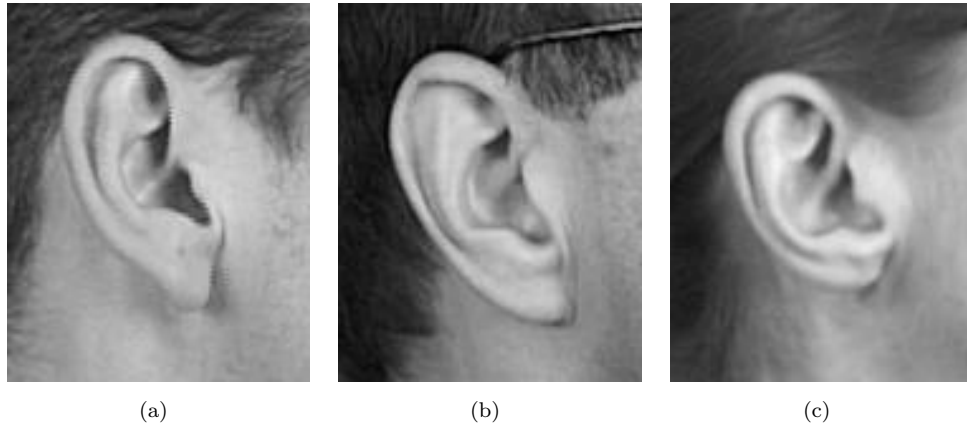


FIGURE 4.15: The different ear lobes: (a)A well-formed lobe. (b)An attached lobe. (c)The lobe is missing. Ears are from the dataset which will later be analyzed [56].

4.4 Discussion

There are many limitations as to the current knowledge of embryological details and many questions are yet to be answered. The common back-tracking method of studying the ear embryology through the observed malformations is also subject to corruption which can occur as a result of physical and environmental interactions with the samples. Although there is no consensus as to the precise embryogenesis of the external ear, certain aspects of the proposed hypotheses are similar. In these hypotheses, the helix and antihelix are derived from hillocks 4 and 5; the crus of helix is usually assigned to an individual hillock, either the second or the third; and the tragus and the antitragus are commonly attributed to hillocks 1 and 6. Streeter [71] showed that hillocks 1 and 6 are the only hillocks which remain distinguishable during the development process and their final form is similar to the initial hillock appearance.

For its more coherent approach we favour the arrangement suggested by Streeter [71]. One of the main points which has been captured in the Streeter’s hypothesis is that it assigns an individual hillock to the ascending helix, distinguishing it from the crus

of helix and the rest of the helix on the hyoid side. As discussed in section 4.3.3, the individuality of the ascending helix is well supported by various syndromes which affect this component. Streeter also has a more coherent idea as to the formation of the lobe, suggesting that early on in its development, the sixth hillock will divide in two to form the antitragus and the lobe. Although Park et al. specifically studied the ear lobe and presented some evidence for their argument [61, 62], they did not produce a comprehensive hypothesis concerning all the components.

His [71] and Sulik [72] also proposed arrangements which are similar to the one proposed by Streeter. The main discrepancy between their hypotheses is regarding the formation of the helix and the antihelix from hillocks 4 and 5. Streeter adopts the lateral partitioning, while His and Sulik employ the anterior-posterior partitions. Interestingly, there is comparative evidence supporting both arguments. Cryptotia and Satyr ear, which exhibit an underdevelopment of the upper helical region, while the lower part appears to be normal, suggest a different development rate for the two, thus reinforce the His and Sulik’s argument. On the other hand, the emergence of the scapha between the helix and antihelix supports Streeter’s point of view, which states that the helix and antihelix are separate components.

We propose a new arrangement for ear components derived from the aforementioned evidence and hypotheses. In this, the emphasis is on identifying sites of independent variations. Figure 4.16 depicts our hypothesized components of the ear. The two prominent aspects of our arrangement are: we divide the helical region into not two, but four parts – the upper helix; the lower helix; the upper antihelix and the lower antihelix – and we assign an exclusive hillock to the ascending helix. We have intentionally avoided using hillock numbers, since their imposed association is not necessary in our research. Instead, we use letters. We have also deliberately neglected the ear lobe, since it seems to be the only part of the ear which continues to grow and change as the person grows older, and thus has only marginal biometric value.

As mentioned before, we expect big variations to occur at common sites of deformities. The extent of the helix and the characteristics of its fold; the orientation and elevation of the superior crus of antihelix; the length and steepness of the ascending helix and the shape of the tragus are some of the more notable of these. On the other hand, the crus of helix and the inferior crus of antihelix (the top margin of the concha), which

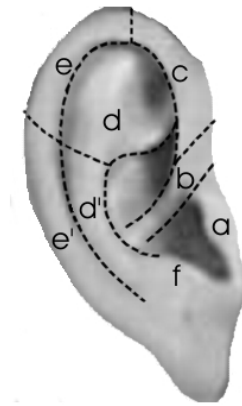


FIGURE 4.16: Our hypothesized components of the ear, which is mainly based on Streeter and Sulik arrangements. We have deliberately used letters instead of hillock numbers, since there is a disagreement between the two arrangements in the position of hillocks 4 and 5. Thus we would have needed to favour one of these arrangements if we were to use these numbers, while the hillock numbers are irrelevant to our research.

were described as the least variable of ear components by Streeter, are not common sites for anomalies. However as it happens, the shadow often caused by the concavity of the concha, creates a big contrast at the edges of the crus of helix and the inferior crus of antihelix causing them to appear as prominent image features.

4.5 Conclusions

The material discussed in this chapter is mainly derived from embryological and surgical accounts of the human ear, revisited from a new perspective, to be exploited in ear biometrics. The accounts of ear embryology attribute six individual growth centres to the formation of the ear shape, and different arrangements have been proposed for these six components in the definitive auricle. Ear abnormalities have been used as one of the main cues. We have derived a new arrangement for the ear components which is intended to highlight the sites of independent variations and is consistent with the discussed ear anomalies. We argue that these insights from ear embryology, particularly since they reveal a component-wise structure, not only assist designing and assessing methods to perform recognition, but they also reinforce the premise of ear biometric validity. Having shown that the ear is indeed a component-wise structure, in the next chapter, we employ a parts-based model to recognize ears.

Chapter 5

The Parts-Based Model

The ear's embryology inspires a localized approach which capitalizes on various ear components. Such an approach simplifies the problem by breaking down the complex structure of the ear into parts which are considered individually. To this end, we propose a parts-based model. Our model, which is the first model-based approach for ear biometrics, is learned using a stochastic clustering method and a training set of ear images. It can detect various ear components in 2D ear images. The deployment advantages of a model-based approach include robustness in noise and occlusion, which is particularly favourable since images of ears are susceptible to occlusion, mostly by hair. It also has a potential advantage in viewpoint invariance. Furthermore the model's explicit approach discards additional irrelevant elements of information which are not part of the ear structure, such as earrings.

In this chapter, we shall discuss the details of our model construction and its derived parts. The candidates for the model parts are detected using the Scale Invariant Feature Transform (SIFT). A stochastic clustering method extracts the clusters of SIFT keypoints in the training set. These clusters constitute the model parts. Although we can not detect the ear components directly from figure 4.16, this information is used to guide and inform the choice of the model. The model parts are identified and evaluated through comparison with the ear components depicted in figure 4.16, and thereby potential improvements have been identified.

5.1 Feature Extraction

The Scale Invariant Feature Transform (SIFT) [52] automatically extracts potential interest points in images in a consistent manner across scales. These features which are marked at the location of the scale-space extrema, are called the keypoints, and describe neighbourhoods of pixels. The size of each neighbourhood is proportional to the scale in which the feature is detected. SIFT's special design enables it to detect the most stable features across all possible scales. We apply SIFT following the enrolment of ears.

The keypoints determined by SIFT have assigned locations, scales and orientations. A distinctive descriptor is also assigned to each keypoint, describing the appearance of the neighbourhood. These descriptors, which are 4×4 arrays of orientation histograms, are normalized with respect to scale and orientation, therefore SIFT is scale and rotation invariant. Furthermore, they are partially invariant to changes in illumination and viewpoint. Lowe has shown that the assigned descriptors are highly distinctive, which allows a single feature to find its correct match with good probability in a large database of features [52].



FIGURE 5.1: The SIFT keypoints of an ear image.

Figure 5.1 shows an ear image with the detected keypoints superimposed. Each keypoint is depicted by a vector showing its location, scale, and orientation. Note that many of these keypoints, such as those detected on the hair, are irrelevant and many, such as the small step edges along the helix, are insignificant in recognition. However, among these are the keypoints which describe specific ear components.

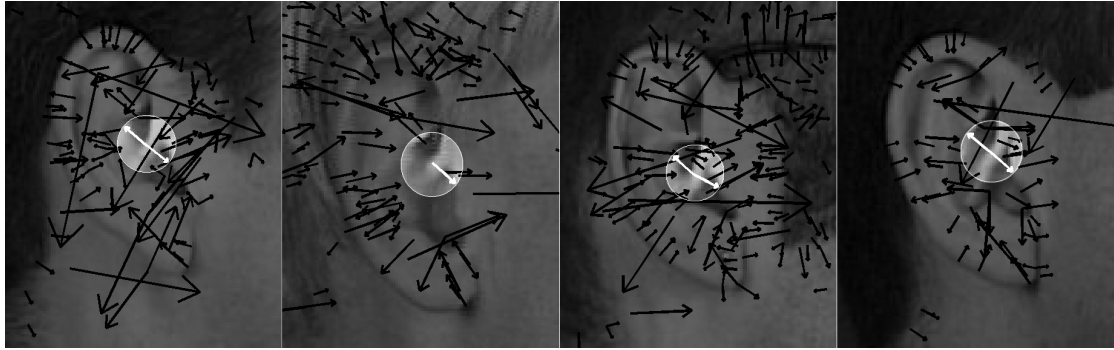


FIGURE 5.2: The crus of helix has been detected on four ear images using SIFT.

Figure 5.2 shows four ear images and their SIFT keypoints. In these, the keypoints describing the crus of helix are highlighted. The crus of helix, which is the transverse ridge on the floor of the concha, is only one of the many ear components which are detected using SIFT. Some components are more consistently detected than the others.

5.2 Learning

Our ear model is constructed using a stochastic clustering method on the SIFT keypoints of the training set. This clustering is an instance of a generic approach called leader-follower clustering [28]. In this, a cluster is initiated upon the arrival of a SIFT keypoint which does not match any of the existing clusters. A new keypoint which is added to a cluster will update the cluster's properties (appearance, location, scale and orientation) depending on the elasticity of the cluster, which is determined by its population. The SIFT keypoints of each ear image are repeatedly presented to the construction algorithm, until the model is stabilized. This is known as recycling, and each cycle is called an epoch [16].

Our learning comprises three distinct processes:

- **Updating/Adding Clusters:** Upon the introduction of a keypoint the matching cluster, if it exists, is updated or a new cluster is initiated.
- **Revision:** A hierarchical clustering algorithm scans for duplicated or similar clusters, and these are merged to avoid the division of focus.
- **Termination:** The recycling is terminated when the model's alteration is not significant in three consecutive epochs, indicating that the components of the model

have stabilized.

The main advantage of this method is that the data is not put into a presupposed number of clusters, but clusters are dynamically created and merged to accommodate the data. While the description of the database we have used will come later, it should be noted here that accurate enrolment has been used in training by cropping to the average ear size of the database to ensure the use of ear features only. Examples of the images in the training set are shown in Figure 5.3.



FIGURE 5.3: A sample of training set images.

SIFT keypoints are detected on every ear image of the training set, and thus our information includes the appearance \mathbf{a} , location \mathbf{x} , scale s , and the orientation θ of these keypoints;

$$\mathbf{P}_n = (\mathbf{a}_n, \mathbf{x}_n, s_n, \theta_n), \quad (5.1)$$

where \mathbf{P}_n is a SIFT keypoint. At each epoch a set of keypoints detected on a single training image is presented to the algorithm. We use a distance measure which combines the normalized scores (in standard deviations) of the Euclidean distances of appearance and location. Normalization is achieved by mapping the Euclidean distances to the distance distributions of matching keypoints manually marked between two ear images. In this, the training images are well registered and thus the locations of matching keypoints are correlated. Let (μ_a, σ_a) and (μ_x, σ_x) denote the mean and the standard deviation of these two distributions for matching appearances and locations respectively. Thereby the distance d between two keypoints \mathbf{P}_i and \mathbf{P}_j is:

$$d(\mathbf{P}_i, \mathbf{P}_j) = \frac{\|\mathbf{a}_i - \mathbf{a}_j\| - \mu_a}{\sigma_a} + \frac{\|\mathbf{x}_i - \mathbf{x}_j\| - \mu_x}{\sigma_x}, \quad (5.2)$$

when both normalized scores are less than three standard deviations, following the application of a filter which discards pairs that exhibit big scale or orientation disparities. Otherwise the two keypoints will be reported as mismatched keypoints. For the values

of the parameters and thresholds see table 5.1.

5.2.1 Updating Clusters

Let $\mathbf{C}_i(k)$ the cluster i in k^{th} epoch be:

$$\mathbf{C}_i(k) = \{\mathbf{ca}_i(k), \mathbf{cx}_i(k), cs_i(k), c\theta_i(k), n_i(k)\}, \quad (5.3)$$

where \mathbf{ca} denotes the cumulative average of appearances, \mathbf{cx} is the cumulative average of locations and the other components are similarly defined. Let $n_i(k)$ be the number of keypoints that have contributed to this cluster up to the k^{th} epoch. All image keypoints, considered in various epochs, contribute to their nearest cluster if it is sufficiently close,

$$i_0 = \underset{i}{\operatorname{argmin}} \{d[\mathbf{C}_i(j), \mathbf{P}_x(j)]\}$$

$$d[\mathbf{C}_{i_0}(j), \mathbf{P}_x(j)] < \varepsilon_s$$

as:

$$\mathbf{C}_{i_0}(k+1) = \frac{\sum_{j=1}^k \mathbf{P}_x(j)}{n_{i_0}(k+1)}. \quad (5.4)$$

In this, $\mathbf{C}_{i_0}(k+1)$ is derived by the contributions from different keypoints, $\mathbf{P}_x(j)$ s, in k epochs. Any keypoint which is not matched to a cluster initiates a new cluster. ε_s is the similarity threshold which determines a match.

5.2.2 Revision

As a result of these alterations some clusters might become so similar that they can no longer be considered distinct,

$$d[\mathbf{C}_i(k), \mathbf{C}_j(k)] < \varepsilon_d, \quad (5.5)$$

where ε_d is the distinction threshold. Therefore a hierarchical clustering algorithm is applied at the end of each epoch to detect these duplicated clusters. In this, the two nearest clusters are repeatedly merged until the distance between the nearest clusters is bigger than ε_d and thus all clusters are distinct. Let $\mathbf{C}_l(k)$ denote the new cluster which

replaces $\mathbf{C}_i(k)$ and $\mathbf{C}_j(k)$,

$$\mathbf{C}_l(k) = \frac{n_i(k) \times \mathbf{C}_i(k) + n_j(k) \times \mathbf{C}_j(k)}{n_i(k) + n_j(k)} . \quad (5.6)$$

5.2.3 Termination

Thus far, our ear model has been obscured by the mass of isolated clusters which were added to the structure so that their potential as a model part would be assessed, but which failed to generate well populated clusters. Applying a threshold on the clusters' cardinalities, we prune out these isolated clusters. Thus the model at the k^{th} epoch $\mathbf{Mod}(k)$ is a cumulative average composed of sufficiently populated clusters,

$$\mathbf{Mod}(k) = \{\mathbf{C}_i(k) | n_i(k) > \varepsilon_p\} , \quad (5.7)$$

where ε_p is the population threshold. The recycling is terminated when the model's alteration is not significant in three consecutive epochs. To detect this stable state an estimate of the model evolution rate $m(k)$ is obtained by measuring the distance between the model parts in adjacent epochs,

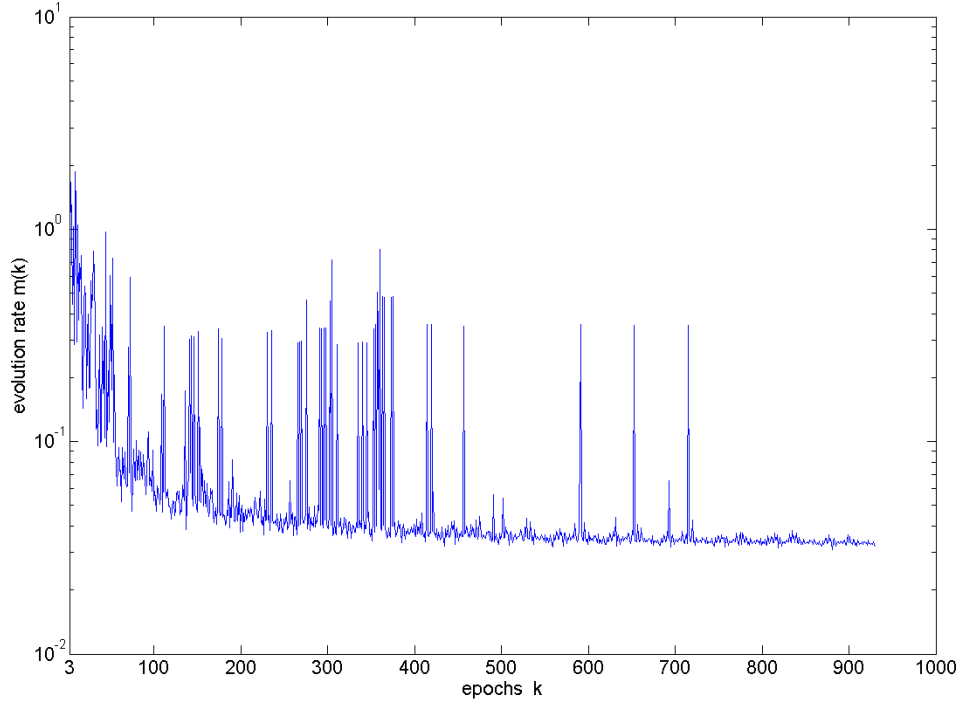
$$m(k) = \sum_i d[\mathbf{Mod}_i(k), \mathbf{Mod}_{j_0}(k-1)] \quad (5.8)$$

$$j_0 = \underset{j}{\operatorname{argmin}} \{d[\mathbf{Mod}_i(k), \mathbf{Mod}_j(k-1)]\} .$$

Figure 5.4 depicts $m(k)$ against the increasing number of epochs. A logarithmic scale is used for plotting $m(k)$ to display its fluctuations in small values more clearly. As mentioned in equation 5.7, a threshold determines the prominent clusters which constitute the model. In this the clusters are constantly changing in the background; if a cluster becomes sufficiently populated, it gets included in the model and causes an abrupt shift in the structure of the model and a spike in $m(k)$, as can be seen in figure 5.4. The frequency of these spikes gradually decreases, and the model gradually stabilizes.

5.3 Training Set and Parameters

63 ear images are manually registered and used in the training of our model. The details of our data are discussed in chapter 7. Table 5.1 summarizes the training parameters

FIGURE 5.4: An estimate of the model evolution rate $m(k)$ in adjacent epochs.TABLE 5.1: The modeling parameters and thresholds (p stands for pixels)

μ_a	σ_a	μ_x	σ_x	ε_s	ε_d	ε_p
0.33	0.12	$6.5p$	$23.3p$	5σ	4σ	$\frac{2}{5}k$

and their values. μ_a , σ_a , μ_x and σ_x are estimated by the mean and standard deviation of two distance distributions of matching keypoints between two ear images. For this we manually labelled the matching keypoints in two ear images.

ε_s and ε_d are experimentally chosen from amongst a small plausible subset, $A = \{3\sigma, 4\sigma, 5\sigma, 6\sigma\}$. Considering that these thresholds are for sums of two distinct distances (appearance and location), which are both normalized and stated in standard deviations, the subset A , in fact, indicates the largest concurrent deviations of $\{\frac{3}{2}\sigma, 2\sigma, \frac{5}{2}\sigma, 3\sigma\}$. Thereby ε_s is set to 5σ , maximizing the recognition via the model for a small set of ear images. ε_d is set to a lower value of 4σ , which delays the merging of clusters and intuitively allows the clusters more time to assume individuality. ε_p is set by evaluating the model in a recognition task in random stages during the learning.

5.4 The Derived Model

32 parts are learned for the model. Figure 5.5 depicts these parts, which are shown individually as circular regions, superimposed on the mean image of the training set, depicting the location, orientation and the scale of each part. However, this conveys no information on variance and detectability of these parts.

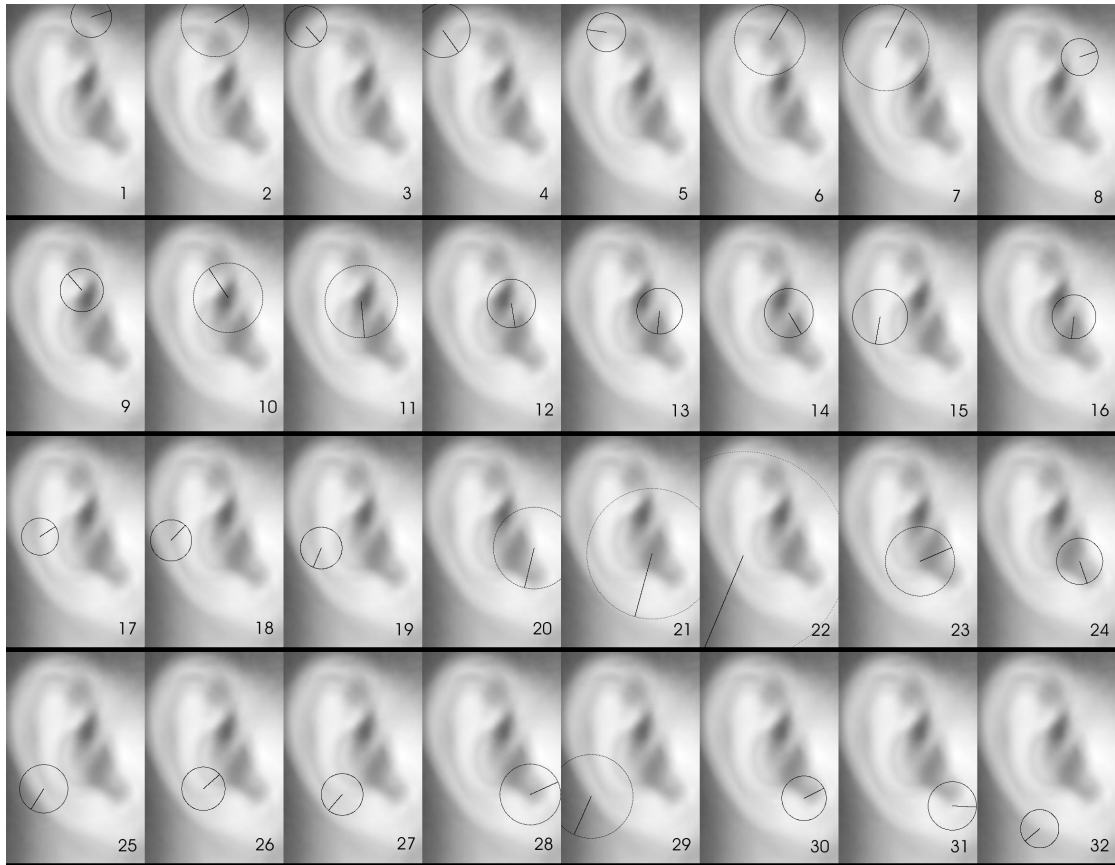


FIGURE 5.5: The 32 derived model parts superimposed on the mean image of the training set.

In recognition, these parts are detected on every ear image; only the corresponding parts are then compared. A model part can be described as the prevalent feature it detects. Figure 5.6 exhibits the detected parts on the first 37 images of our data. In this, each column presents an ear image with its parts, while rows depict the parts detected on different ear images. Clearly, not every part is detected on every image. The frequency of detection varies for different parts; part 9 is detected on most images while part 22 is rarely picked. The model parts are also different in terms of the features they capture. We divide the parts into three categories, one of which is ear components:

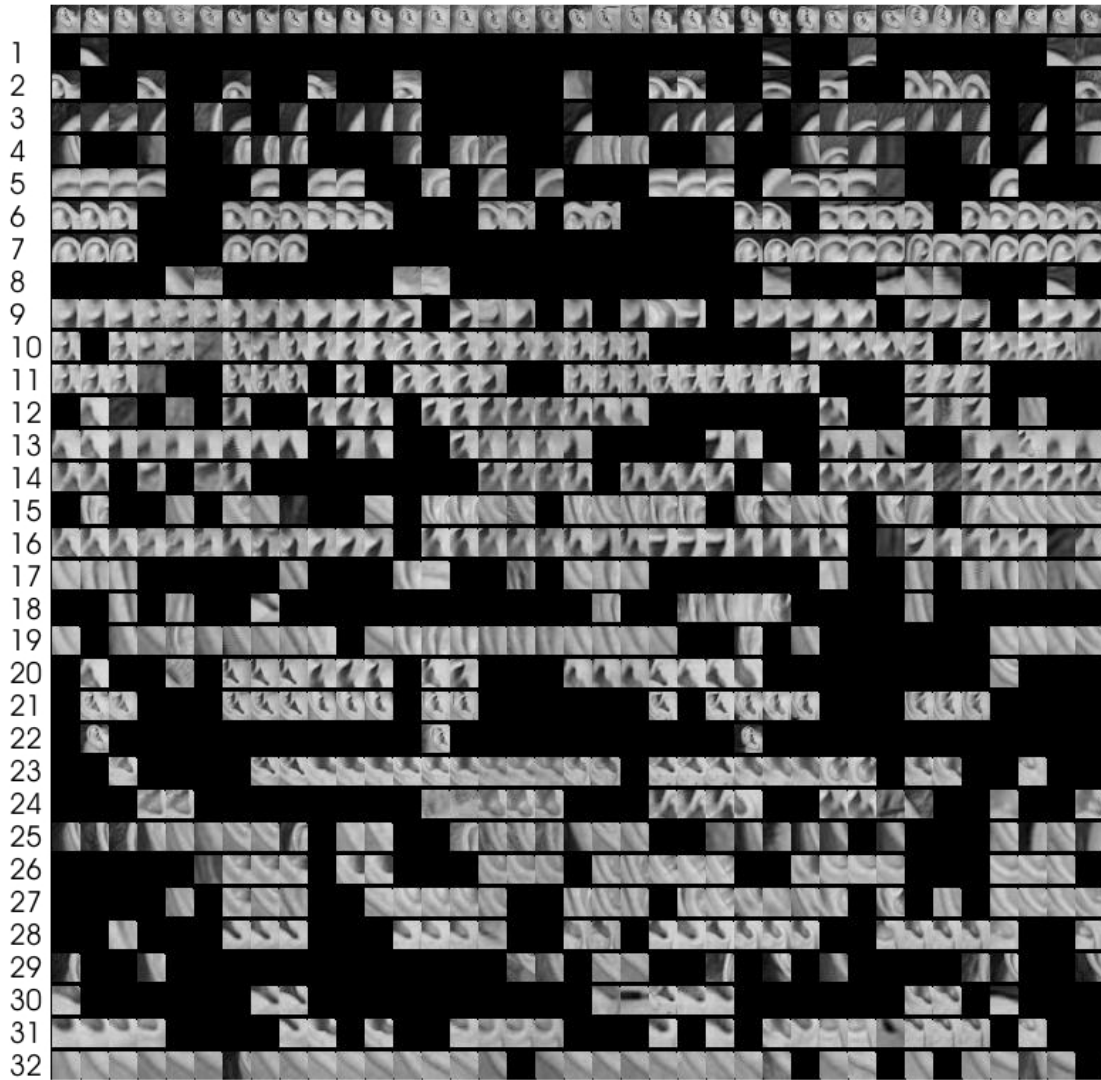


FIGURE 5.6: Detected model parts on ear images (a column presents different parts in an ear image, and a row presents a part detected on different ears).

- **Basic features:**

Common image features, such as step edges, appear in six parts – 1, 3, 4, 8, 13 and 29. Although these could be observed within complex components, they are not individually significant. Especially in the case of ear images, they often emerge as a result of the contrast introduced by dark hair. The six diagrams in figure 5.7 represent these six parts.

- **Recurring shapes:**

Similar shapes might appear repeatedly within more complex structures. Situated on the helix, are a series of small patches which include a vertical indentation adjacent to a bloated ridge, which is the rolled over edge of the helix. These are

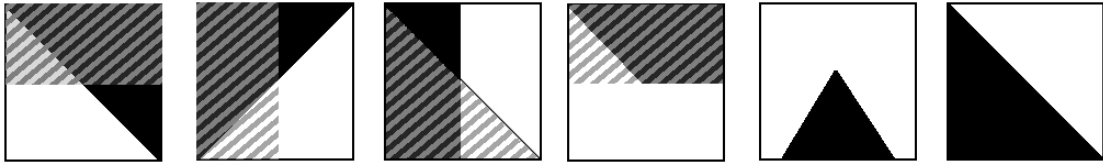


FIGURE 5.7: Diagrammatic representation of the basic features. The striped sections represent the parts' inclined direction of variation.

TABLE 5.2: The model parts which describe specific ear components.

Describing Part	Ear Component being Described	Level of Detail
2	Triangular Fossa (the depression between the superior and inferior cruses of antihelix)	High
6	Ascending helix	Medium
7	Upper helical region	Low
9	Inferior Crus of antihelix	High
10, 11	Confluence of inferior crus of antihelix and the crus of helix	Medium
14, 16	Crus of helix	High
12, 20	Concha	High, Medium
21	Lower ear region	Low
22	Entire ear	—
23	Antitragus	High
28, 30, 31	Incisura	High

aligned to form an elliptical component – the helix. Nine parts presenting such features are: 5, 15, 17, 18, 19, 25, 26, 27 and 32.

- **Ear components:**

These are sixteen parts each corresponding to either an ear component, a significant section of a component or a confluence of components. Table 5.2 summarizes these parts. The level of detail captured in each description is also included in the table. Note that the significance of each of these parts in recognition depends on the frequency and accuracy of detection as well as the featured component's ability to discriminate.

Part 24, which does not produce consistent results, is not assigned to any of these groups.

5.4.1 Feature Subset Selection

In recognition, we use our model as a mask in keypoint selection. We search for matches to our model parts amongst the detected SIFT keypoints on each ear image. The best match is selected from a set of possible matches that are ranked according to their appearance similarity and from which those with large scale or orientation disparities have already been discarded. The distance between each pair of ear images is defined as the mean distance between their corresponding parts, and a k-nearest neighbour is then used to classify the ear images. The details of the model-based recognition and the description of the training and test sets can be found in chapter 7. However, for the sake of argument, here, we refer to a model-based recognition rate on one of the test sets.

A 96.3% correct recognition rate is achieved on our dataset of 63 individuals, which comprises 189 images (testset B, section 7.1). However, as we will show here, a much smaller subset of parts can achieve similar recognition. Feature subset selection algorithms search for the best features and the most effective subsets of features for a classification. Sequential Forward Floating Search (SFFS) is a feature subset selection algorithm proposed by Pudil et al. [65], and in a feature selection evaluation study by Jain et al. [46] it was reported as the dominant method among the tested algorithms, including genetic algorithms and artificial neural networks. Given the desired number of features for the subset, SFFS starts with an empty set and searches for the best subset by sequentially adding the next best feature in conjunction with its present subset. Each feature which is added to the subset constitutes a forward step. The features are also removed in backward steps. A backward step is performed when it improves the performance of an intervening subset.

The recognition rates achieved using the subsets selected by SFFS are shown in figure 5.8. It can be seen that the model achieves 98% of its best performance with only ten parts. Nine parts out of these ten belong to our third category of parts, describing specific ear components. The remaining parts are either irrelevant to recognition, or they were rendered redundant by the chosen subset.

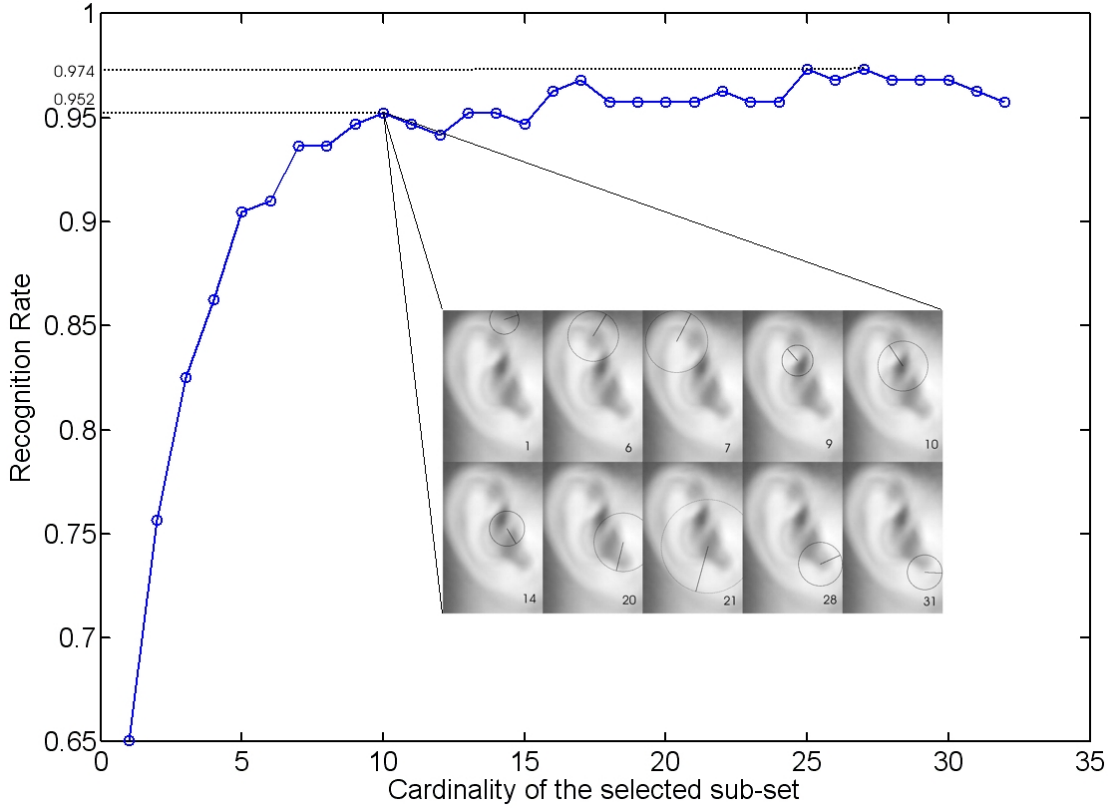


FIGURE 5.8: The recognition rates achieved by subsets with different cardinalities selected by SFFS.

5.4.2 The Ten Most Significant Parts

Table 5.3 provides a more detailed description of the ten parts which have been selected via SFFS. These parts are listed in order of their significance in recognition, which is the order by which they were picked by the algorithm. Eight matches corresponding to each part are also shown in the table. Studying these parts, we discuss the model's capability to perform recognition, and potential improvements are also identified.

According to table 5.3, the most important ear features are the inferior crus of antihelix (part 9) and its intersection with the crus of helix (part 10). This appears to contradict one of our conclusions in section 4.4, which was based on some ear malformation studies and was also observed by Streeter [71]: the inferior crus of antihelix and the crus of helix are the least variant parts of the ear. We believe that this disparity is caused by our model's proven capability in detecting the parts in question. We can see in table 5.3 that parts 9 and 10 are the most accurately detected (detection rate \times correct detection rate) parts of the ear, which means that the information regarding these parts is most


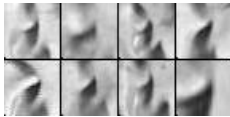
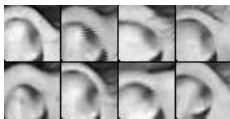

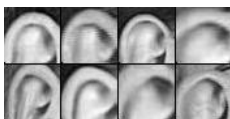


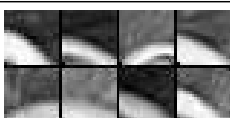

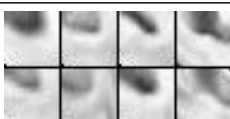
Part	Detection rate	CDR*	Description	examples
9	0.91	0.96	Inferior Crus of antihelix	
10	0.85	0.88	Confluence of inferior crus of antihelix and the crus of helix	
6	0.74	0.89	Ascending helix	
20	0.37	n/a	Concha or similar cavities	
7	0.52	n/a	Upper helical region (including the upper helix, superior crus of antihelix and the ascending helix)	
14	0.57	0.82	Crus of helix	
21	0.35	1	Lower ear region	
1	0.19	–	A basic shape	
28	0.55	0.82	Incisura	
31	0.50	0.83	Incisura	

TABLE 5.3: The details of the ten most significant parts of the model. * CDR (Correct Detection Rate) is determined by manual evaluation. CDR is not presented for parts 20 and 7, since there is a high uncertainty in labeling the erroneous matches. This is due to the fact that these parts each detect a feature but in slightly different scales.

frequently available. As to the reason for the model’s capability to detect these parts, we hypothesize that their comparative consistency, as was noted previously, helps learning these components by clustering.

The ascending helix was also discussed during the study of the ear’s embryology. It was shown that although the ascending helix is not an obvious component, there is much evidence supporting its individuality. In accordance with that, the third most significant model part (part 6) describes the ascending helix.

The model parts depicted in table 5.3 can also reveal which ear components and features are under-represented in the model. Most notable of these is the helix. Although some model parts, like part 5 and 18, describe aspects of the helical fold, the shape of the helix as a whole has not been captured, while the helix is one of ear’s most variant and hence one of its most discriminant components. As aforementioned, the scapha, which provides a margin between the helix and antihelix, increases the potential of free development for the helix. This potentially discriminant and independent component, which seems not to have been considered, marks a potential improvement for the model.

The detection of the upper helical region as an important part (part 7) is interesting from the recognition point of view, since we have shown in section 4.3 that there is evidence indicating distinct development for this section of the ear. It is also interesting to speculate why the incisura is featured in many different model parts (see table 5.2). Can it be that a big variance is imposed on this region because of the two distinct hillocks (1 and 6)? Furthermore the tragus is considered to be one of the most variant components. Thus the variance might be larger than can be expressed in a single cluster.

An alternative probabilistic approach [30], which is used to learn models for object recognition, has also been applied to learn the ear model. By preliminary evaluations of performance in modelling and recognition, this method does not present an improvement over our ear model. The details of this can be found in Appendix A.

5.5 Conclusions

In this chapter, we have proposed a new parts-based model for ear biometrics. We have learned our model from a training set of ear images. In this, a stochastic clustering

method finds the clusters of SIFT keypoints in the training set. Thus the frequently observed structures in these images are detected. As discussed in chapter 4, although the ear appears as an intact shape, embryology describes it as a confluence of individual components, which validates our modelling scheme. Even though not all the structures which appear in the model are specific ear components, a feature subset selection has revealed that the most significant model parts in recognition are in fact the ear components.

Chapter 7 contains the details of the experimental results, including a thorough evaluation of the model-based performance in recognition when applied to various test sets and with different levels of occlusion. As we shall see, the model achieves a 96.3% recognition rate on testset B, which includes 189 images from 63 individuals each represented by 3 images of the left ear. This test set is from the Hurley et al.s' selection [40] from the XM2VTS database [56]. The images in the Hurley et al.s' derived dataset are all taken from the left profile. We also extract further images from XM2VTS and extend the number of individuals to 150. In our newly added data, each individual is represented by at least two images, and since there is no preference between left and right ears both are considered, and the one with most number of unoccluded samples is chosen. Thereby, we derive testset C, with 458 images of 150 individuals, on which the model obtains an 89.1% recognition rate. One of the most significant advantages of model-based approaches is their capability in handling occlusion and noise, and we shall show that our model is robust in dealing with the disturbances caused by occlusion. Our results confirm the validity of this approach not just in modelling, but also by recognition capabilities.

Having studied the components of the ear and the location and description of the expected variations, in chapter 4, we have evaluated the extent of ear information that has been captured by the model. Thereby, we have noted that the information residing in the helix and the antihelix has not yet been sufficiently explored. In the next chapter, we shall exploit the characteristics of these curves, using a wavelet-based approach.

Chapter 6

Robust Log-Gabor Filter for the Outer Ear Curves

Our parts-based model has shown that various embryonic parts of the ear form clusters, when examined by a local feature descriptor such as SIFT, and thus they are detectable in different ear images. However, the shape and characteristics of the outer ear boundaries – the helix and the antihelix – have been left unexplored using this method. Expanding on our previous model, in this chapter we capitalize on the ear’s boundary curves. The embryology and the anatomy of these curves offer evidence of a high degree of independent variation, which leads us to expect considerable performance gain as the result of incorporating these new features.

Visualizing the helix and the antihelix as the ripples of a fluctuating surface, we explore their features by analyzing the frequency content along radial lines which are mostly normal to the helix curve. For this task we have chosen a wavelet approach, more specifically the log-Gabor filter. As in the previous stages, we are concerned with the effects of partial occlusion. By localization a wavelet can offer performance advantages when handling occluded data. A more robust matching strategy is also applied to handle erroneous coefficients if they emerge. Prior to applying the wavelet transform, a neighbourhood including the helix and the antihelix is automatically detected, rearranged and aligned into a template.

In this chapter, we shall first discuss the characteristics of the features we aim to capture with this new approach. We will then prepare aligned templates and explore them via the

log-Gabor filters. Finally, we analyze the robustness of this method to partial occlusion and add a more robust matching strategy.

6.1 Improving the Parts-Based Model

Figure 6.1 shows how the ten most significant parts of the model (see section 5.4.2) have been distributed on an ear image. Each part is shown by a circular region, coloured in a shade proportional to the scale of the part; bigger parts have more lighter shade indicating that a lesser amount of detail has been captured. It can easily be seen that the model contains mainly the mandibular components (mainly the components seen on the right-hand side of the image) of the ear. The hyoid parts, especially the helix, are under-represented. We hypothesize that this is caused by utilizing SIFT points as candidates for model parts. Being a circular-based region descriptor, SIFT is incapable of describing the boundaries and stretched curves in comparative detail. Therefore, Lowe suggests the use of image contour features in addition to the SIFT features [52]. Fergus also adds curve descriptors to his region-based features in his approach to object recognition [29].

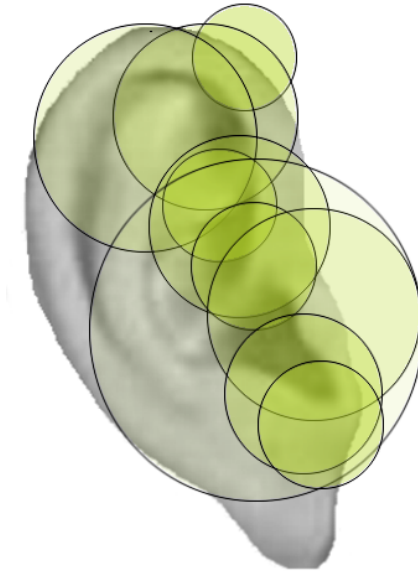


FIGURE 6.1: The model parts coverage map. Note that the helix and parts of the antihelix are outside the concentration zone.

The helix and antihelix are developed from two out of six embryonic hillocks. By neglecting these parts we would potentially lose one third of the information stored in an

ear image. It is also interesting to note, as was observed by Streeter [71], the helix and the antihelix take longer to develop into their definitive shape and are more detailed and occupy a much bigger area of the ear compared to the parts on the mandibular side.

The helix is attached to the antihelix along all of one side. Thus, it appears that the two should have highly correlated structures. However, on this note, we should also consider the scapha. The scapha is a concave surface of *free portion* lying between the antihelix and the helix. Note figure 6.2 in which three differently shaped and sized scaphas are highlighted. The scapha is almost inconspicuous in figure 6.2(a), while it occupies a larger area of the ear in figure 6.2(b). Figure 6.2(c) shows that scapha is not necessarily parallel to the antihelix. The scapha provides a margin and allows the helix to be rather shape-independent from the curves and ridges within it. Thus the possibility of finding new and discriminant information arises from the potential of independence in the helix.

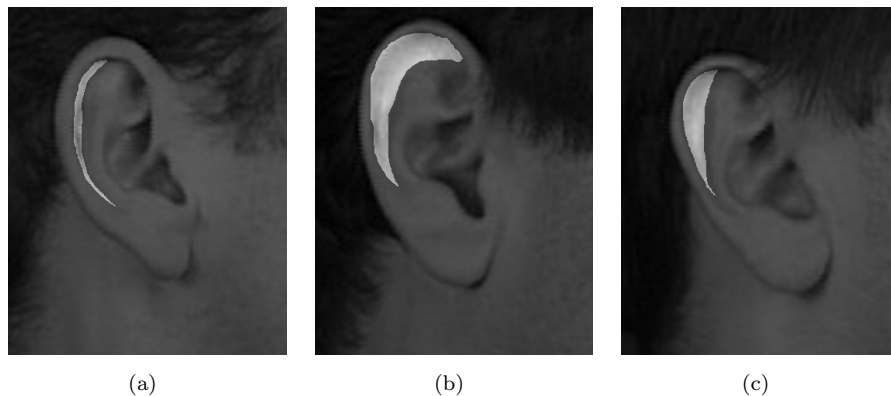


FIGURE 6.2: The scapha is highlighted in these images. The scapha is a surface of free portion, and thus it assumes various shapes.

6.2 Approach

The ear boundary curve has been previously explored by other researchers mainly for the purpose of segmentation. Alvarez et al. [6] fitted an ovoid model to the ear contour, thereby described the shape of the helix. Yan et al. [88] used an elliptical active contour to locate the ear outline. These detected curves can then be represented by Fourier descriptors. However, Alvarez et al. initialize their algorithm with a manual estimate of the ear contour, and Yan et al.'s method relies on the presence of both coloured 2D and

3D information and is susceptible to occlusion. Furthermore, apart from the information regarding the curvature of the helix, other features of the boundary curves include: *i*) the characteristics of the helix rolled-in margin; *ii*) the curvature and the width of the antihelix; *iii*) the depth and the shape of the scapha. Figure 6.3 displays four images from our database. In this, the first ear on the top left has rather normal shaped parts, but the other three each exhibit a discernible deviation from the normal shape in one of the traits mentioned above.

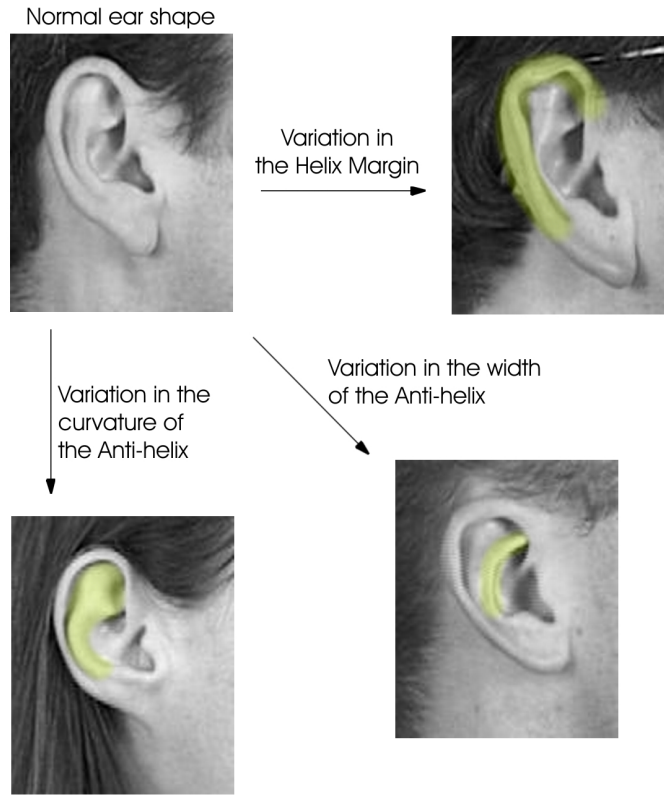


FIGURE 6.3: Some of the possible variations in the shapes of the helix and antihelix. Compare the characteristics of the traits mentioned in the deviated samples to the corresponding region in the normal ear shape. These images are taken from our data.

We propose to describe the boundary information in terms of the frequency content of the radial signals of the image intensities outwards from the concha. These radii, which are mostly normal to the helix curve, capture the characteristics of the helix and the antihelix as well as the variations between the two at each angle. A wavelet technique is chosen to extract the frequency information. Note that wavelets offer decimation in space as well as in frequency, which is their main advantage over the Fourier transform. Thus by localization, they are intuitively beneficial when handling partially occluded data.

6.3 Preparing the Templates

Prior to applying the wavelets, we prepare aligned templates of image data. Our template is the sampled image intensities in a semi-circular region which includes the helix. The chosen centre of this semi-circle is where the crus of helix curves inwards, which is almost the midpoint of the ear height and is situated on the outermost part of the ear, opposite to the helix (see figure 6.4).

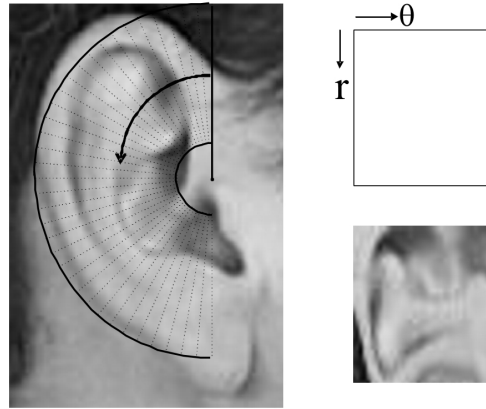


FIGURE 6.4: The input template

In this template, the ridge of the antihelix and the scapha reside in the rows, and thus the columns exhibit the variation between the antihelix and the helix in each specific angle. Figure 6.5 highlights two columns of the template in figure 6.4. In this the helix and the antihelix are featured as hills, while the scapha is presented as a furrow. The helix and the antihelix curvature information is also implicitly incorporated as the accumulated location information of the corresponding ridges across all angles. Small discrepancies in rotation and scale of the templates can be accounted for in matching, with little loss of information, via shifting the template horizontally or/and vertically. However, in the controlled environments in which databases are usually acquired, scale does not change and ears are not rotated, notably.

The first step in building these templates is to detect the centre of the semi-circle or the crus of helix. This is done using some of the previously detected model parts which vote for this location. Note that although parts 14 and 16 of the model directly feature the crus of helix, they are not accurately detected on every ear image. Thus a further set of parts contribute to the task via voting for the correct position of the crus of helix. This set contains parts: 6, 7, 9, 10 or 11, 12 and 14 or 16 (see the model parts in section

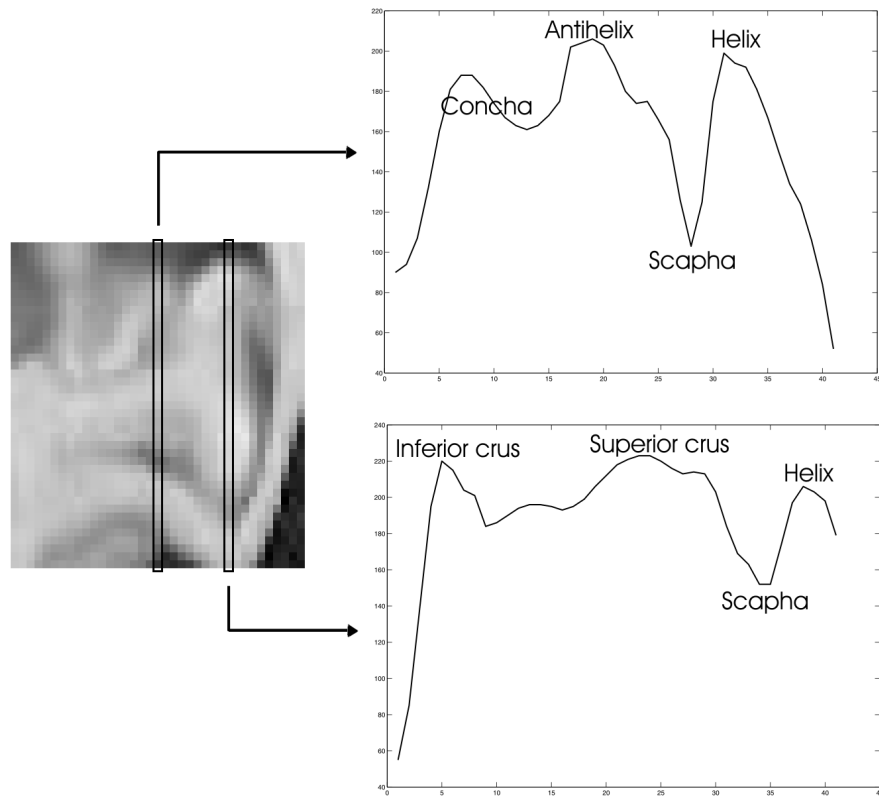


FIGURE 6.5: Radial signal fluctuations in two different angles. In these, the helix and the antihelix are clearly presented as hills, while the scapha appears as a furrow. In the second signal the antihelix consists of two parts, the inferior crus and the superior crus of antihelix, which are visible as minor elevations.

5.4), which indicate nearby locations and are relatively more reliably detected. Figure 6.6 shows two ear images plus images showing how the votes accumulate to locate the centre. In this, each part votes in a neighbourhood most likely to contain the crus of helix. The votes in this neighbourhood are weighted, firstly proportional to the part's accuracy rate, which has been determined via manually assessing the detected parts in a small set of ear images, and secondly inversely proportional to the distance from the putative centre. The location with most number of votes has been highlighted on the initial image with a cross. Once the position of the hub has been determined, the image is sampled along radial lines via interpolation.

6.4 Wavelet Approach and the Log-Gabor Filter

The Fourier transform is the main tool for frequency analysis. However, it is not good for localizing frequency information spatially. In this, the sinusoid basis function spans

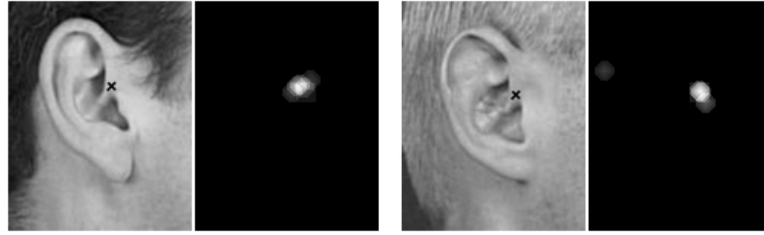


FIGURE 6.6: Six model parts vote for the position of the hub. Larger weights are assigned to the votes of the parts which are expected to be more accurately detected.

the entire space with no hint as to the spatial significance of each frequency component. Alternatively, wavelets have been proposed to obtain localized frequency content. They achieve this by decomposing the signal over waveforms which are well-localized in the spatial-frequency plane. However, in doing so they lose some of the precision of the frequency analysis. Indeed, any frequency-based analysis is governed by an inevitable trade-off between the precision of spatial and frequency content [53].

For applications in computer vision, Gabor wavelets have been the most popular choice of wavelet. This is somewhat justified as Daugman showed that the impulse responses of a class of cells in the visual cortex can be approximated by Gabor wavelets [20]. The Gabor basis function is essentially a sine/cosine wave modulated by a Gaussian windowing function. The wavelet transform provides a means for multi-scale analysis of data, in which image projections into wavelets with various scales, highlight the specific frequency content in the spatial domain. Daugman has successfully used the Gabor wavelet for iris recognition [21].

An alternative to the Gabor wavelet is the log-Gabor wavelet proposed by Field [31]. Field has looked at the statistical properties of the natural images and showed that an alternative coding – the log-Gabor filter – is a much more efficient method to describe this class of images, which are characterized by their long tails in the frequency domain. Furthermore, log-Gabor filters retain zero DC value regardless of the bandwidth. The DC component is the response regarding the illumination or intensity of the image, which is preferably set to zero giving invariance to illumination. In contrast, for the Gabor filters we need to keep within one octave of bandwidth to obtain filters with zero DC value. Log-Gabor filters have been comprehensively used by Kovess [49, 48] to calculate phase congruency. They have also been used as a feature extraction method in face and iris recognition [67, 90]. Here we use the log-Gabor filter to extract the features of the helix and antihelix.

6.4.1 Log-Gabor Filter

The log-Gabor filter is defined as having a frequency response which is Gaussian in a logarithmic frequency scale as opposed to the standard Gabor which has Gaussian frequency response in a linear scale. Figure 6.7(a) and 6.7(b) show the log-Gabor's frequency response in linear and logarithmic scales respectively. Note the filter's zero DC component and its long tail in 6.7(a). Figure 6.7(c) and 6.7(d) display the real and the imaginary parts of the filter in the spatial domain. Kovesi noted that the appearance of the Gabor and the log-Gabor filters in the spatial domain are almost identical for bandwidths of less than one octave, however the log-Gabor filter becomes much sharper as the bandwidth increases [48].

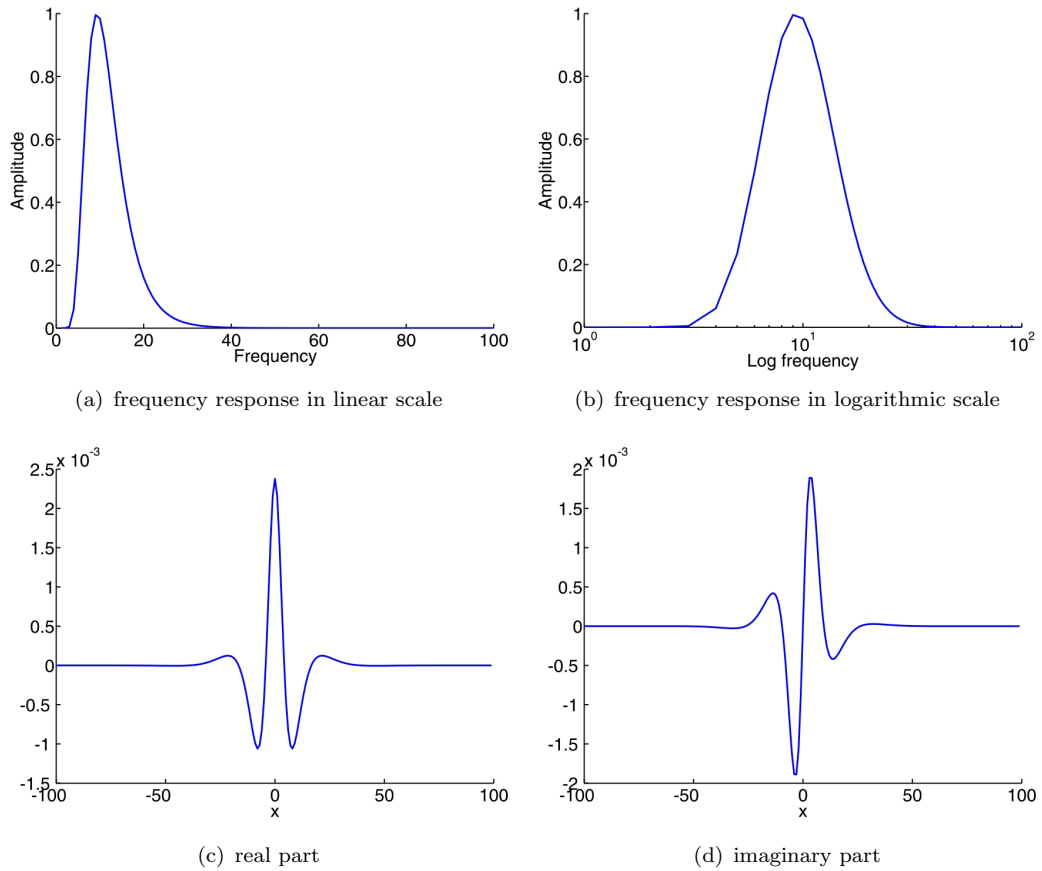


FIGURE 6.7: A log-Gabor filter profile in frequency and spatial domain. The frequency response in linear and logarithmic scales are shown in (a) and (b). The real and imaginary parts of the log-Gabor in spatial domain are shown in (c) and (d).

The log-Gabor frequency response is:

$$G(\omega) = e^{\frac{-(\log(\omega/\omega_0))^2}{2(\log(k/\omega_0))^2}} \quad (6.1)$$

where ω_0 is the centre frequency of the filter and the bandwidth is determined by the k/ω_0 term. We will discuss the centre frequency and the bandwidth in greater detail in the next section, which looks at assigning values to the filter parameters. We use a one dimensional log-Gabor filter on the columns of the templates, $T(r, \theta)$. Thus the projected image $PI(r, \theta)$ is obtained by:

$$PI_n(r, \theta) = F^{-1}[F(T(r, \theta)) \times G_n(\omega)], \quad (6.2)$$

where F and F^{-1} denote the Fourier and the inverse Fourier transforms, and $G_n(\omega)$ is a log-Gabor filter at scale n . This corresponds to convolving a template with the wavelet in the spatial domain. Both the real and the imaginary responses are used in matching.

6.4.2 Filter Parameters

In this section, we describe how we obtain and tune our log-Gabor filter bank. A filter bank is a collection of individual filters whose overlapping frequency responses span a frequency spectrum. In this, we have followed the guidelines provided by Kovesi [48, 47]. In choosing the best values for the parameters of the log-Gabor filters, the main question we have to answer is what range of frequencies we need to analyze. Figure 6.5 shows two typical examples of our input signals, in which we expect the antihelix and the helix to appear individually as hills in image intensity values. Roughly, the wavelengths of the components representing these two are between one third to two thirds of the duration of the signal (=41 pixels). On the other hand, Kovesi suggests that the smallest wavelength value, and thus the biggest frequency, that can be captured is the Nyquist limit of 2 pixels. Thus to find the most distinctive frequency span we examine the following intervals: min wavelength of 3 to 18 pixels and maximum wavelength of about $27(\sim 41 \times \frac{2}{3})$ pixels or larger.

Once we have an estimate of the frequencies we want to cover, we have to produce a bank of filters which provides an even coverage of this spectrum. Designing an effective log-Gabor filter bank requires the solution of the dilemma between providing an even coverage of the spectrum and retaining independence in the filter responses. To this end, Kovesi provides a table of empirically determined values for the scaling factor between the centre frequencies of adjacent filters, achieving minimal overlap for a fairly even coverage.

TABLE 6.1: Plausible intervals for the best value for each of the log-Gabor input parameters.

	Min wavelength	Max wavelength	Bandwidth
Search intervals	[3, 4, ..., 18] pixels	$\lambda_{max} \geq 27$ pixels	$\sim 0.5, 1, 1.5, 2$ octaves

The other parameter to be set is the bandwidth. As mentioned before, the upper limit of one octave for Gabor filters has been overcome by the special design of log-Gabor, thus filters with arbitrary bandwidths can be defined. Normally bandwidth represents the width of the filter in the frequency domain. However, for log-Gabor filters changing the bandwidth does not result in a simple linear stretch in the frequency domain. Kovesi shows that increasing the bandwidth does not always create more localized log-Gabor wavelets in spatial domain. He found that a bandwidth of two octaves minimizes the spatial width of the filter and that, interestingly, for bandwidths larger than three octaves the spatial width starts to increase considerably.

Table 6.1 lists the plausible intervals which were examined to determine the best values for these parameters. The ‘decidability’ d' , which was introduced by Daugman [22], is used as a measure of fitness of these parameters. d' evaluates the potential decisiveness of a biometric task by examining the separability of the clusters which represent individuals in the feature space.

The bandwidth is controlled by the k/ω_0 parameter in equation 6.1. As calculated by Kovesi, $k/\omega_0 = 0.75$ results in a filter with approximately one octave bandwidth, and $k/\omega_0 = 0.55$ results in a two octave bandwidth. To design a filter bank, Kovesi couples these with a multiplier, m , which determines the scaling between adjacent filters. m is optimized to achieve minimal overlap and even coverage simultaneously. Details of this can be found in [47]. The maximum wavelength is then determined;

$$\lambda_{max} = \lambda_{min} \times m^{(n-1)}, \quad (6.3)$$

where m is the multiplier and n is the number of filters in the filter bank. Here, n is set to the smallest value for which $\lambda_{max} \geq 27$ is satisfied.

Figure 6.8 displays the decidability values in four trajectories corresponding to four bandwidth values. In this, we plot the decidabilities against increasing λ_{min} . The parameter values corresponding to the tests with $k/\omega_0 = 0.85$ and $k/\omega_0 = 0.55$ are

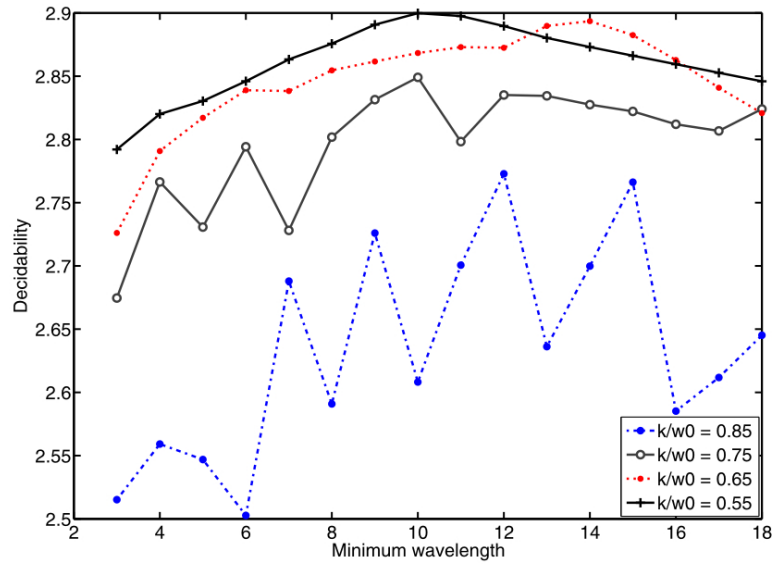


FIGURE 6.8: Decidability values in four different bandwidths. Note that bandwidth increases as the k/ω_0 decreases.

TABLE 6.2: List of parameters in two different bandwidths

(a) $k/\omega_0=0.85$						(b) $k/\omega_0=0.55$					
λ_{min}	k/ω_0	m	n	λ_{max}	d'	λ_{min}	k/ω_0	m	n	λ_{max}	d'
3	0.85	1.3	10	32	2.52	3	0.55	3	3	27	2.79
4	0.85	1.3	9	33	2.56	4	0.55	3	3	36	2.82
5	0.85	1.3	8	31	2.55	5	0.55	3	3	45	2.83
6	0.85	1.3	7	29	2.50	6	0.55	3	3	54	2.85
7	0.85	1.3	7	34	2.69	7	0.55	3	3	63	2.86
8	0.85	1.3	6	30	2.59	8	0.55	3	3	72	2.88
9	0.85	1.3	6	33	2.73	9	0.55	3	2	27	2.89
10	0.85	1.3	5	29	2.61	10	0.55	3	2	30	2.90
11	0.85	1.3	5	31	2.70	11	0.55	3	2	33	2.90
12	0.85	1.3	5	34	2.77	12	0.55	3	2	36	2.89
13	0.85	1.3	4	29	2.64	13	0.55	3	2	39	2.88
14	0.85	1.3	4	31	2.70	14	0.55	3	2	42	2.87
15	0.85	1.3	4	33	2.77	15	0.55	3	2	45	2.87
16	0.85	1.3	3	27	2.59	16	0.55	3	2	48	2.86
17	0.85	1.3	3	29	2.61	17	0.55	3	2	51	2.85
18	0.85	1.3	3	30	2.65	18	0.55	3	2	54	2.85

listed in table 6.2. Note that, m is determined following the choice of k/ω_0 and n is obtained from equation 6.3 via conditioning λ_{max} . It is evident from figure 6.8, that the decidability values grow and also become more stable as the bandwidth becomes larger, peaking around $\lambda_{min} = 10$ and then reducing slightly. It seems that the fluctuations in d' at $k/\omega_0 = 0.85$ reflect the fluctuations in λ_{max} values. This sensitivity to λ_{max} value is decreased as the bandwidth becomes larger. From amongst the best decidability values we chose the one which obtains a better classification rate using a simple nearest



TABLE 6.3: Eight ear images along with their corresponding templates, featuring four individuals each represented by two samples.

neighbour classifier. The chosen combination of parameters are highlighted in table 6.2. These parameters are: $\lambda_{min} = 9 \text{ pixels}$; $\text{bandwidth} \sim 2 \text{ octaves}$; $m = 3$; $n = 2$; $\lambda_{max} = 27 \text{ pixels}$. Note that this method is sensitive to the accurate detection of the crus of helix which acts as the hub to generate the templates. Thus, for each probe image four additional templates are derived where the position of the hub has been slightly shifted horizontally or vertically. In each comparison, the best matching template is used.

Table 6.3 presents eight ear images along with their corresponding templates. In this there are four individuals each represented by two samples. The real and imaginary projections of these templates using the log-Gabor filters are shown in table 6.4. Note that the helix and the edge of the antihelix adjacent to the concha are the most discernible on the projections of filters with 9 pixels wavelength, while the body of the antihelix is featured in the responses of the larger scaled filter.












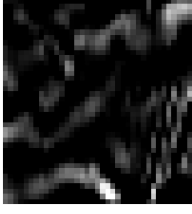



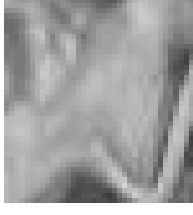
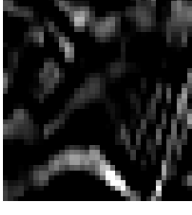
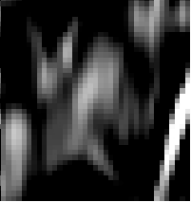

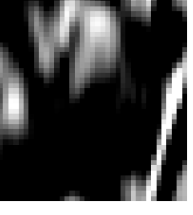

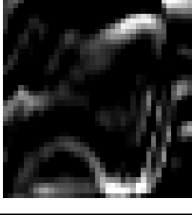
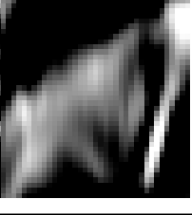



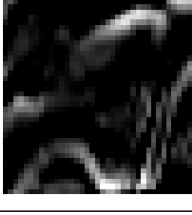







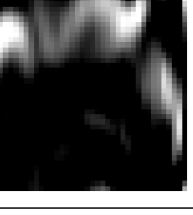





		Real Projections		Imaginary Projections	
		$\lambda = 9$	$\lambda = 27$	$\lambda = 9$	$\lambda = 27$
(1)					
(2)					
(3)					
(4)					
(5)					
(6)					
(7)					
(8)					

TABLE 6.4: The real and imaginary projections of templates in table 6.3.

6.5 Robust Matching

The spatially localized frequency information provided by log-Gabor filters is intuitively beneficial when analyzing ear images with occlusion. In this, the local frequency is not contaminated by the clutter of surrounding regions. However, the projected image still contains invalid information where data is corrupted. For classification purposes, we use a simple nearest neighbour approach, for which we could use the Euclidean distance, but this corresponds to least squares estimation, which is known to be intolerant of outliers. Thus, here we shall consider a more robust alternative.

The least squares method minimizes the error function E :

$$E(\mathbf{r}) = \sum_i \rho(r_i) \quad , \quad \rho(r_i) = r_i^2, \quad (6.4)$$

where r_i is a residual error. To demonstrate the influence of the residuals on the error function, an influence function, ψ , is defined as:

$$\psi(r_i) = \frac{d\rho(r_i)}{dr_i} . \quad (6.5)$$

ψ -function for a given ρ -function expresses the influence of different residual values on the estimation. In the least squares method with $\psi = 2r_i$, the data points retain an influence proportional to their residual value, no matter how far from the mean they are. This unconditional weighting of the distant residuals, which are the potential outliers, is the cause of the lack of robustness. Estimation can be made more robust by restraining the influence of large residuals. It should also be noted that the optimum choice for the influence function is also dependent on the characteristics of the noise in the data.

Black and Jepson [9] have identified a similar problem in PCA, and used a robust ρ -function with good effect. They showed that the standard mapping of images into the eigenspace corresponds to the least squares estimate, and thus the mapping of noisy images produces poor results. The ρ -function of their choice is:

$$\rho(r_i, s) = \frac{r_i^2}{s^2 + r_i^2} \quad (6.6)$$

where ρ is a function of s as well as the residuals. The corresponding ψ -function is:

$$\psi(r_i, s) = \frac{2r_i s^2}{(s^2 + r_i^2)^2} . \quad (6.7)$$

Figure 6.9 shows this influence function. In this, the influence of potential outliers begins to decrease at:

$$\psi' = 0 \rightarrow |r_i| > \frac{s}{\sqrt{3}} . \quad (6.8)$$

Thus the parameter s determines a soft threshold.

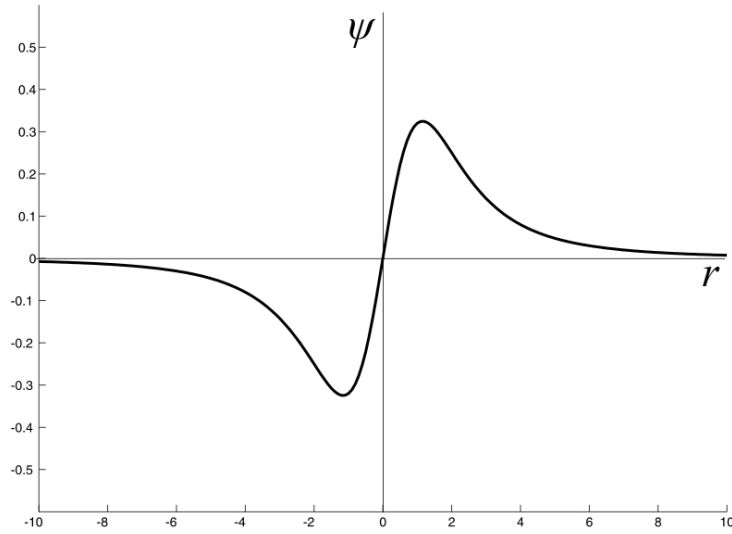


FIGURE 6.9: The robust influence function previously used by Black and Jepson to handle the noise in the data when acquiring PCA projections.

Black and Jepson showed that this ρ -function is appropriate to handle the disturbances caused by structured noise and occlusion. This is very similar to the type of clutter and occlusion we are dealing with in ear biometrics. Inspired by this, we use the same ρ -function in our distance measure. Here the residual vector \mathbf{r} consists of the variations between the components of two image projections i and j :

$$\mathbf{r} = [PI_n(r, \theta)]_i - [PI_n(r, \theta)]_j \quad (6.9)$$

Thus minimizing E corresponds to finding the nearest neighbour for image i . The value of s is set to maximize the correct classification rates on a dataset of ears – testset B

(see section 7.1) – both in its original state and with 30% of synthetic occlusion.

6.6 Conclusions

In this chapter, we have introduced a new complementary analysis to span the features which have been neglected using our parts-based model. The specific aim of the new approach is to capture information in the ear’s outer structure. In this, we have captured the target neighbourhood in aligned templates and explored the variations between the boundary curves using log-Gabor filters. The undesirable effects of occlusion were moderated first via the localized approach of the log-Gabor filter and then by applying a more robust matching strategy which restricts the influence of erroneous projections.

We have conducted our experiments on a database of 521 images of 150 individuals. A dataset of occluded samples have also been examined. The data is divided into various testsets evaluating our techniques in different settings. The details of our experiments, including the recognition performance of the parts-based model and the wavelet approach, both separately and fused, are discussed in the next chapter. As we shall find, on our main testset – testset C (see section 7.1) – our new wavelet method achieves a 91.9% correct classification rate and when combined with the model, it improves the model’s performance by 8% achieving a hybrid recognition rate of 97.4%. On this testset, the robust matching has improved the performance of the wavelet-based method by about 5% for original samples. However, when the samples were occluded by about 20% this improvement increased to about 17%, as will be described in the next chapter.

Chapter 7

Hybrid Classification and Experimental Results

In the preceding chapters, we have discussed the details of our ear biometric approach, including: automatic enrolment, model learning and a complementary method of feature extraction. In this chapter, starting with the description of the database, we shall present extensive performance analysis. For this purpose, the data is divided into various testsets to satisfying the need for an objective evaluation of the various aspects of our biometric approach. In recognition, the ear model and the wavelet-based analysis are employed both separately and when combined. We shall specifically focus our attention on assessing the effect of occlusion. As was mentioned earlier, we consider occlusion to be the main obstacle hindering the deployment of ear biometrics. Thus we aim to minimize the impairments caused by occlusion.

7.1 Test and Training Data

Our ear images are selected from face-profile images of the XM2VTS database [56]. XM2VTS is a multi-modal database, comprised of digital video recordings of the face and speech of 295 subjects. In this, we use the frames of the left and right profile views of the head. Figure 7.1 shows some face-profile images from XM2VTS. This database was not specifically targeted to the acquisition of ear images, and in many cases ears are partially or fully occluded by hair. Hurley et al. [40] have derived a dataset of 63



FIGURE 7.1: Examples of XM2VTS face-profile images.

individuals from the XM2VTS for their ear biometric analysis. This dataset is balanced, having four images per individual which are all taken from the left profile. Thus this set comprises $(63 \times 4 =) 252$ images. The ears in this database are fully visible and have not been obscured by hair. Hurley et al. have also rotated, scaled and registered these ears to 111×73 sized images. This registration, which is manually performed, is specially developed to reinforce recognition capability in PCA for comparison purposes. We make use of this data along with their manual registration. We also extract a further 269 images from XM2VTS and extend the number of individuals to 150. In our newly added data, each individual is represented by at least two images of the left or right ear; it also includes ears that are slightly occluded. We also derived a separate 361-image dataset from other samples within XM2VTS which includes considerably occluded ears. This set is assembled to examine the impact of occlusion by hair.

We divide this data into various sets to evaluate our ear biometric technique in different

settings. Table 7.1 summarizes the details of these testsets. The probability of random match in a leave-one-out cross-validation scheme ¹, which takes into account the size of the database and the number of samples per individual, is also shown for each testset. Note that this probability is presented as a measure of difficulty of the testsets. The other difficulty factors include errors in enrolment, image quality and occlusion.

TABLE 7.1: Summary of the testsets properties.

Testset	Size	Subjects	Automatic Enrolment	Enrolment Error	Probability of random match
Testset A	189	63	×	–	0.0106
Testset B	189	63	✓	0/189 = 0%	0.0106
Testset C	458	150	✓	4/458 ~ 1%	0.0048
Testset D	269	87	✓	4/269 ~ 1.5%	0.0086
Testset E - probe	148	104	✓	7/148 ~ 5%	0.0088
Testset E - gallery	213	104	✓	8/213 ~ 4%	

We have developed an automatic ear enrolment method, in chapter 3, which finds the position of the ear. An image including the ear is then cropped. Figure 7.2(a) shows some enrolled ears using this method, Hurley’s manual registration of these ears is shown in figure 7.2(b). Our automatic enrolment finds the ear in 517/521 (~ 99.2%) images of



(a) Automatic enrolment



(b) Hurley’s manual enrolment

FIGURE 7.2: Our automatic enrolment and Hurley’s manual enrolment

the main data (testset C plus the training set) and in 346/361 (~ 95.8%) images of the occluded dataset (testset E).

¹In a leave-one-out cross-validation scheme, each sample is compared against the rest of the dataset. Thus the probability of random match is: $\sum_{n=2}^4 \frac{m_n}{N} \times \frac{n-1}{N-1}$, where N is the testset size, n is the number of samples per subject, and m_n is the number of subjects having n number of samples. Each subject is represented by at least two and at most four samples.

The detailed description of our datasets is as follows:

- **Training set:**

- 63 images, 63 individuals, from Hurley's selection [40]
- Manual registration, from Hurley's registration

These images are used to train the ear model.

- **Testset A:**

- 189 images, 63 individuals, from Hurley's selection
- Manual registration, from Hurley's registration

This testset is used to display the PCA's best performance.

- **Testset B:**

- 189 images, 63 individuals, same ears as in testset A
- Automatic enrolment

This dataset is used in conjunction with testset A to evaluate the significance of pinpoint registration in the overall performance of different recognition methods. The manual registration provides accurate enrolment of the ears and furthermore the ears are normalized in terms of rotation and scale, while our automatic enrolment crops a wider image patch where the ear is roughly placed in the centre. Some of the surrounding hair and skin are also captured in these patches. Note that the redundancy in the enrolled patches provides a margin for error, increasing the chance of producing workable samples when the ears are occluded or otherwise cluttered.

- **Testset C:**

- 458 images, 150 individuals, the entire unoccluded dataset
- Automatic enrolment

This is the entire unoccluded dataset excluding the training samples. This testset is unbalanced; it includes four, three or two samples per individual.

- **Testset D:**

- 269 images, 87 individuals,
- Automatic enrolment

This dataset includes the ear samples that we have added to Hurley's data from XM2VTS. Compared to testset B, testset D is generally more affected by occlusion, out-of-focus and interlacing problems.

- **Testset E:**

- 361 images in 213 gallery and 148 probe images, 104 individuals
- Automatic enrolment

This testset is assembled to assess the impacts of occlusion by hair. It consists of a probe set of ears which are occluded by hair, while the gallery includes good ear samples of the individuals. This testset is comparable to our main testset (testset C) in terms of number of images. Unfortunately, it is not straight forward to assess the degree of occlusion in probe images. The extent of occlusion in these images varies from about 10% to as much as 60%; the cases of severe or total occlusion in XM2VTS are not included. Figure 7.3 shows the entire set of occluded probes of testset E which have been automatically enrolled. Given the properties of the HT in handling noise and occlusion it obtains satisfactory results enrolling occluded ears.

The UND ear biometric database [86] also includes head-profile images, and it contains both 2D and 3D data. This database, with 942 images from 302 subjects, is one of the biggest available datasets of ears. However, in this the emphasis is on the 3D scans, and the 2D images are of considerably lower quality compared to XM2VTS samples. Yan et al. [88] have reported a high recognition rate on this data. However, note that their analysis is mainly based on the 3D data and the use of 2D information is very limited. They use 2D colour data in segmentation as one of the two conditions to stop the expansion of an active contour. The 3D scans and 2D colour images are obtained nearly simultaneously. Example images from UND are shown in figure 7.4. In this, a Minolta VIVID 910 range scanner is used to capture both 3D and 2D data. A rotating filter captures the 2D colour images, which suffers when the subject moves slightly. We find that the blurring effect caused by the long exposure time and poor indoor lighting conditions produce 2D images which are of considerably lower quality than the XM2VTS images. Because of these limitations in quality of UND data we mainly use XM2VTS data in performance evaluation, although we have tried our automatic enrolment on UND data (see section 3.3).



FIGURE 7.3: The entire probe set of testset E, automatically enrolled. The ears of this set are all occluded to various degrees, and in some the enrolment has been unable to detect the ear. The errors in enrolment have been highlighted.



FIGURE 7.4: Examples of 2D images in the UND dataset.

7.2 Ear Recognition

In recognition, we use the parts-based model and the log-Gabor coefficients both separately and jointly. We combine the two using a simple decision fusion technique of the weighted sum of the normalized distances. Thus the hybrid distance, D , between two ear images is defined as:

$$D = w_m \times D_{model} + w_o \times D_{outer\ ear}. \quad (7.1)$$

where $D_{outer\ ear}$ is the distance of the log-Gabor coefficients, and is so named since it presents the characteristics of the ear's outer structures.

We use our model as a mask in keypoint selection. The model parts are detected on every ear image. Thereafter only the corresponding parts are compared, and thus only the featured components in the model are used in recognition. The model searches for the model parts amongst the detected SIFT keypoints. The best match is selected from a set of possible matches that are ranked according to their appearance similarity and from which those with large scale or orientation disparities have already been discarded.

Let \mathbf{MP}_x be the set of model parts detected on image x . \mathbf{MP}_{x_k} is set to null if part k is not detected. Thereby the part-wise distance between each pair of ear images is

estimated as the mean distance between their corresponding parts:

$$D_{part-wise}(\mathbf{MP}_i, \mathbf{MP}_j) = \sum_{k|\mathbf{MP}_{i_k}, \mathbf{MP}_{j_k} \neq \emptyset} \frac{\|\mathbf{MP}_{i_k}, \mathbf{MP}_{j_k}\|}{m_{ij}} \quad (7.2)$$

where m_{ij} is the number of parts which were detected in both images. In comparing a probe image to the gallery, $D_{part-wise}$ is biased in that potentially different subsets of parts contribute to assessing the distances of the probe to different gallery images. Thus, we also include a measure of structural similarity.

The Tanimoto (Jaccard) coefficient, which was first introduced by Paul Jaccard [45], is used to measure the similarity between sets. It is defined as the size of the intersection divided by the size of the union of the sets. This metric is commonly used in comparing the structural features of molecules in chemical databases, where vectors of binary data are compared [80]. Here, we use it to assess the similarity between each pair of ear images in terms of presence or absence of various parts. The absence of a part in an ear image could mean that the part is occluded. However for unoccluded ear images it can mean that the part is significantly different from the normal shape, thus it has not been detected. We define a status set for each ear image which indicates whether different parts have been detected. The similarity between these status sets quantifies the similarity in fitting the model to the images. In a status set, \mathbf{S} , a part k can be either detected, (k, on) , or not detected, (k, off) ;

$$\mathbf{S}_i = \{(k, \text{on}) \text{ or } (k, \text{off})\}, \quad k = 1 \dots N, \quad (7.3)$$

where N is the number of the model parts. We expect a bigger correlation between the status sets of images from the same ear, and we use the Tanimoto distance, which is derived from the Tanimoto coefficient (Tanimoto distance = $1 - \text{Tanimoto coefficient}$), to determine the distance between two status sets:

$$D_{tanimoto}(\mathbf{S}_i, \mathbf{S}_j) = \frac{n_i + n_j - 2n_{ij}}{n_i + n_j - n_{ij}}, \quad (7.4)$$

where n_i and n_j are the number of elements in the sets \mathbf{S}_i and \mathbf{S}_j respectively, and n_{ij} is the number of elements in both sets. Here, $n_i = n_j = N$, where N is the number of the model parts. We define the model-based distance, D_{model} , as the weighted sum of the

two aforementioned distances:

$$D_{model} = w_p \times D_{part-wise} + w_t \times D_{tanimoto}, \quad (7.5)$$

where w_p and w_t are the weights of the part-wise and Tanimoto distances respectively. These weights are determined empirically. Table 7.2 lists the number of correct classifications as well as the Daugman's decidability values [22], d' , as the weight of the part-wise metric increases. The evident improvement in performance as w_p increases indicates the comparative significance of the part-wise distance in recognition. A small improvement is obtained as the result of incorporating the Tanimoto distance at $w_p = 0.85$ and $w_t = 0.15$.

TABLE 7.2: Tuning the weights of the components in the model-based distance by classification rates and decidability on testset B

w_p	w_t	Correct/189	d'
0	1	39	1.19
0.1	0.9	50	1.40
0.2	0.8	61	1.63
0.3	0.7	87	1.89
0.4	0.6	108	2.17
0.5	0.5	143	2.45
0.6	0.4	162	2.70
0.7	0.3	174	2.89
0.75	0.25	177	2.95
0.8	0.2	179	2.99
0.85	0.15	182	3.00
0.9	0.1	182	2.99
0.95	0.05	181	2.97
1	0	178	2.93

On the other hand, the $D_{outer\ ear}$, as was described in chapter 6, is the distance between two projected images PI_i and PI_j (from equation 6.2), which is robustly estimated by:

$$D_{outer\ ear}(PI_i, PI_j) = \sum_{r, \theta, n} \rho\{[PI_n(r, \theta)]_i - [PI_n(r, \theta)]_j\}, \quad (7.6)$$

where ρ is the robust ρ -function in equation 6.6.

Table 7.3 lists the number of correctly classified ears as well as the respective decidability values examining different combinations of weights for D_{model} and $D_{outer\ ear}$ on two testsets B and C. The best values are highlighted in this table; $w_m = 0.4$ and $w_o = 0.6$.

Note that this does not correspond to the maximum d' value, but it is close to it. Note also that the d' is a measure of the decisiveness of the feature space and is independent of the classifier, hence the disparity when the classifier does not exploit the full potential of the identity clusters.

TABLE 7.3: Tuning the model-based and outer ear weights in the hybrid distance metric

w_m	w_o	Testset B		Testset C	
		Correct/189	d'	Correct/458	d'
0	1	181	2.80	423	2.23
0.1	0.9	184	3.05	433	2.46
0.2	0.8	187	3.28	439	2.69
0.3	0.7	187	3.46	444	2.88
0.4	0.6	188	3.58	447	3.02
0.5	0.5	188	3.61	444	3.09
0.6	0.4	187	3.57	441	3.10
0.7	0.3	187	3.47	436	3.05
0.8	0.2	183	3.33	424	2.96
0.9	0.1	183	3.17	417	2.84
1	0	182	3.00	408	2.72

Table 7.4 displays the recognition results obtained via the model and the outer ear process as well as the fusion of the two. In this, a k-nearest neighbour with $k = 1$ is used to classify the ear images. All images in the testsets are considered in a leave-one-out cross-validation framework, which makes maximum use of the data. In all four datasets, the hybrid classification exhibits a significant improvement over the model and the wavelet-based method, thereby supporting our hypothesis that the outer ear analysis captures some new and independent information from that already in the model. On testset C, which is our main testset, the model obtains an 89.1% recognition rate, the outer ear method achieves a 91.9% recognition rate and the hybrid classifier exhibits a 97.4% recognition rate.

TABLE 7.4: Correct classification rates (CCR) on various testsets

	Testset A	Testset B	Testset C	Testset D
Model	181/189 ~ 95.8%	182/189 ~ 96.3%	408/458 ~ 89.1%	237/269 ~ 88.1%
Outer ear	181/189 ~ 95.8%	182/189 ~ 96.3%	421/458 ~ 91.9%	250/269 ~ 92.9%
Hybrid	187/189 ~ 98.9%	188/189 ~ 99.5%	446/458 ~ 97.4%	261/269 ~ 97.0%

Similar correct classification rates (CCR) were obtained on testsets A and B with the model and the outer ear analysis. This clearly demonstrates that our approach does not require exact registration, and hence it can handle enrolment misalignments more efficiently. Also notice in this table that in the more challenging testsets, testset C and D, the model's performance degrades by about 7% and the wavelet-based performance by about 3%, but the hybrid recognition still maintains a recognition rate greater than 97%.

7.3 Occlusion Analysis

Clearly, classifying occluded samples is more challenging, because of the missing information. However, another significant impact of occlusion is the general disturbance it causes. In this, the expected overall appearance which is observed in typical unoccluded samples, which make up the training set and the gallery, does not emerge. Holistic methods measure general properties of the samples, which is intuitively more susceptible to the disturbances caused by noise and occlusion. For example, utilizing PCA, an occluded ear might appear more like an ear which is similarly occluded than an unoccluded sample of itself. In a model-based approach, a model is provided which imitates the structure or the dynamics of the object. This is potentially more effective in predicting beyond the training set and distinguishing an object's features despite occlusion and clutter. Occlusion can also affect the pre-processing and registration stages, which would subsequently impair feature extraction and classification.

We have previously mentioned that perhaps the biggest challenge in ear biometric recognition is handling occluded samples. In fact, handling the loss of data and more importantly the disturbance in the data's general structure was the main incentive in opting for a model-based approach. Thus we also require occluded ear samples and datasets which present these challenges. To this end, we have constructed testset E, in which the probes are between 10% and 60% occluded. The occluded probes of testset E have been shown in figure 7.3. However, it is not straightforward to arrive at an objective conclusion as to the extent of the occlusion related impairment, since the samples are occluded in different manners and to various extents. For this reason, we also synthetically occlude the samples of the other testsets. In this, we can control the degree of occlusion, and since the same ear samples are increasingly occluded, we obtain an unbi-

ased assessment of the performance as the occlusion intensifies. However, it should be noted that the occlusion which is caused by hair often introduces more complications, in that it might interfere with the features of the ear. In our scenario of occlusion, the dataset includes two separate parts: the gallery; and the probes. The gallery includes the good samples — the unoccluded ears which have been taken in advance and under supervision when the individual was enrolled into the system. The probes are acquired with less supervision, and thus they might be occluded. Therefore we require both unoccluded samples, which constitute the gallery, and occluded samples, which would be the probes.

In this section, we look into the impacts of occlusion in ear biometrics. We use a PCA-based method as a representative of holistic methods, and generate some benchmark results. PCA is a well-defined and widely used method for recognition, and obtains a 98.4% recognition rate on testset A, which is unoccluded and accurately registered. However, since PCA is well-known to be sensitive to outliers we use a robust version of PCA. We shall also look into the impact of occlusion on enrolment and its subsequent effects on recognition.

7.3.1 PCA and Robust PCA

The CCRs obtained by PCA on increasingly occluded ears are shown in figure 7.5. The samples are synthetically occluded from the top. Also see an occluded ear in figure 7.6(b). Clearly, the high recognition performance of PCA (98.4%) does not hold for occluded samples. The recognition rate drops to 12% for only 20% occlusion. In this, the eigenvectors were determined using the method described in [75, 37], and the distance metric is the Manhattan distance, which in this case performs better than the Euclidean distance. Classification is performed using the nearest neighbour method. Although the sudden drop in performance displays an impairment which can be caused by occlusion, it hardly provides a benchmark to compete against. In fact sensitivity to outliers is a well-known limitation of PCA.

Many approaches have been suggested to tackle PCA's problems with sensitivity to outliers, and although they all reside under the common name of Robust PCA (RPCA), they are different in terms of what they are attempting to achieve. These methods mainly consider the case of occlusion and varying background in association with two

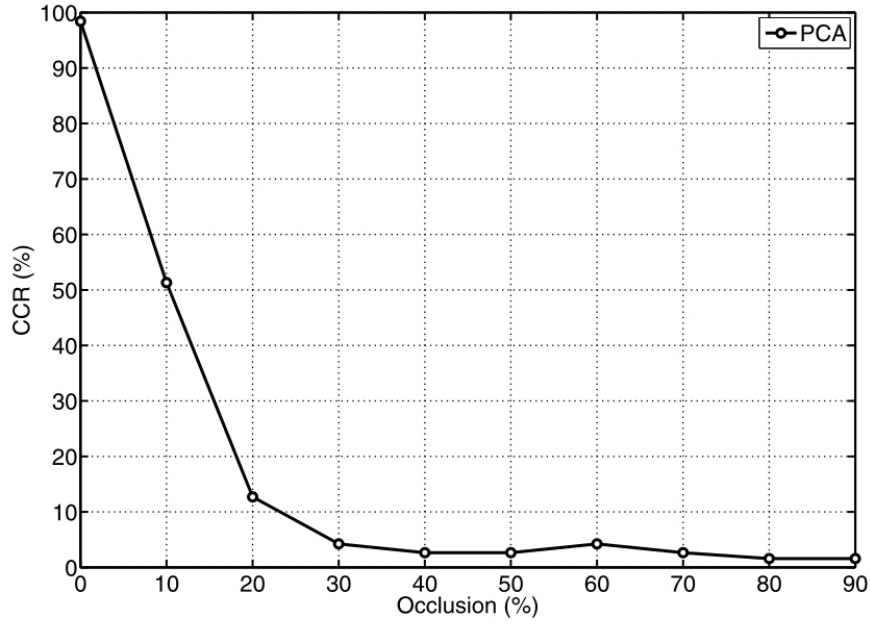


FIGURE 7.5: PCA performance drops as the samples are occluded.

distinct robustness issues: *i*) outliers in the training set, which undermine the learning of the eigenspace; *ii*) outliers in the test set, which are often poorly reconstructed by the eigenvectors. Considering our occlusion scenario, which is based on relatively good training samples and poor test samples, we require methods from the second category. We have used the approach suggested by Black and Jepson [9], which reformulates the standard eigenspace reconstruction by a more robust estimator.

This robust version of PCA, provided by Black and Jepson, has also inspired the use of a robust matcher in chapter 6 (see section 6.5). Here, minimizing the error function corresponds to finding the image projection in the eigenspace. This is done using a gradient descent algorithm. Starting with a large value for s (in equation 6.6), initially no data is rejected, then as we approach the best answer, s is gradually decreased and so is the influence of outliers. See figure 7.6 in which RPCA provides a much better reconstruction of the occluded ear than the standard PCA. Figure 7.7 displays the CCRs obtained via RPCA on testset A. Note the significant improvement which has been gained.

Other PCA problems include variant background and image transformations such as translation, scaling and rotation. In other words, an accurate segmentation and registration is also required. We have accounted for this by using testset A, for which PCA

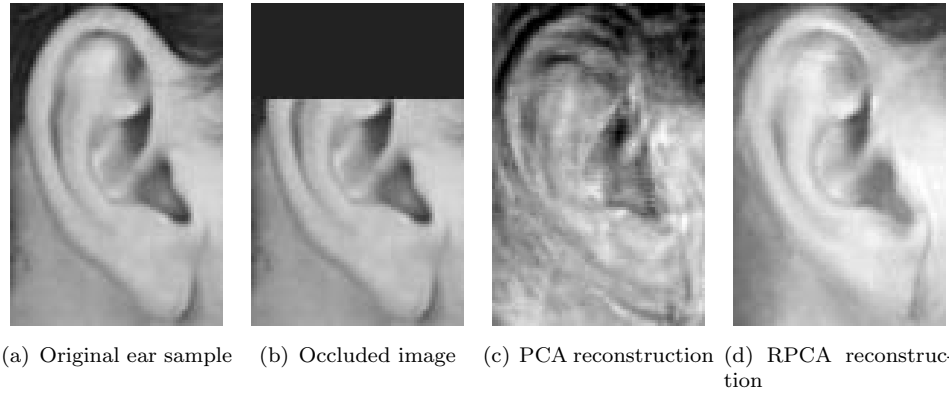


FIGURE 7.6: RPCA provides a much better reconstruction of the occluded ear.

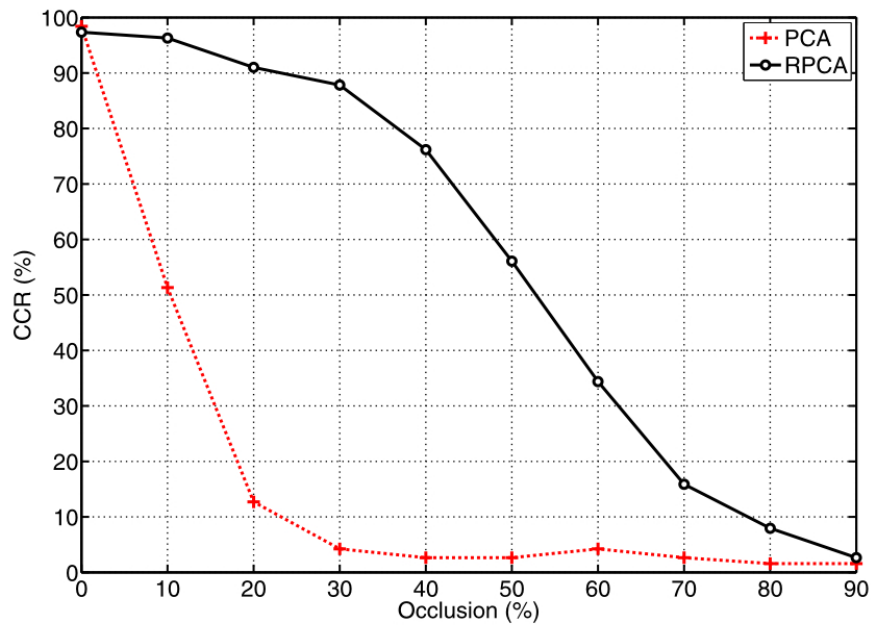


FIGURE 7.7: Using RPCA the performance in occlusion improves significantly.

obtains a 98.4% recognition rate for unoccluded samples. The PCA performance degrades to 76% when we classify the images of testset B. RPCA performs slightly better with a 79.9% recognition rate. Note that testset A and B contain the same ear images; they only differ in registration. Testset A is accurately registered to average ear size, while less refined enrolment is automatically obtained for testset B.

7.3.2 Effects on Enrolment

Although the synthetic occlusion on testset A impairs the recognition performance, it should be noted that it does not spoil the enrolment since in that dataset the enrolment is

performed manually. The same is not true for testset B and other testsets with automatic enrolment. In these, the ear region of the face-profiles are synthetically occluded from the top and then the occluded ears are enrolled, thus it is more likely that the enrolment would produce erroneous results. Figure 7.8 shows an ear which has been occluded, increasingly from the top by roughly 10% to 50%, and subsequently enrolled.

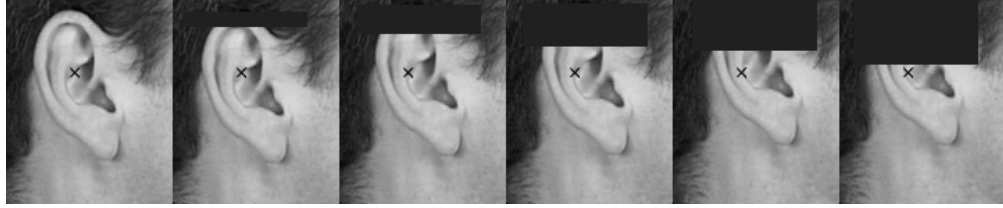


FIGURE 7.8: An increasingly occluded ear (10% to 50%), enrolled successfully. The detected ellipse centres – the presumed ear centres – are shown by a cross.

Figure 7.9 depicts the enrolment success rate of testset B against increasing synthetic occlusion similar to the occlusion in figure 7.8. For this, the erroneous outputs are

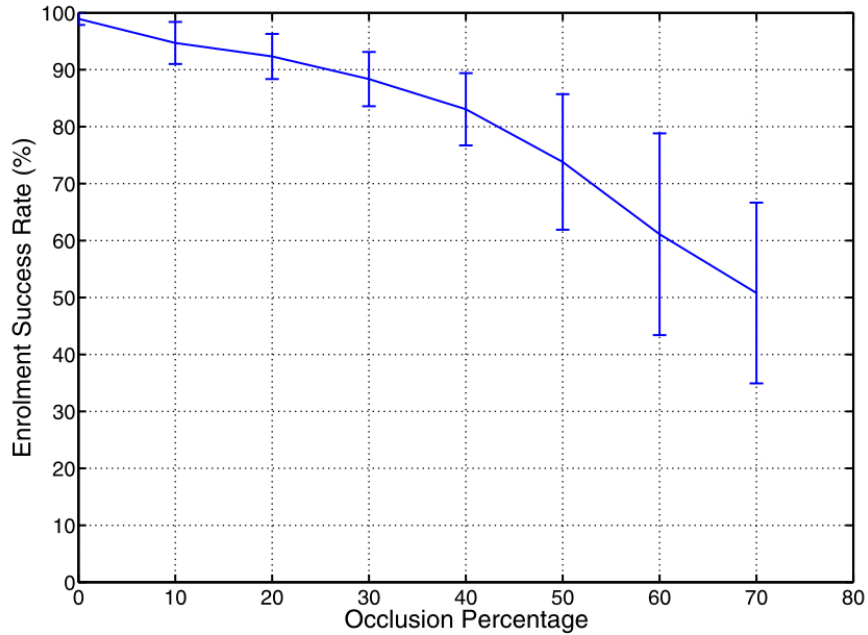


FIGURE 7.9: Enrolment of occluded ears of testset B. The upper and lower bound of the error correspond to the different degrees of sensitivity to misalignment.

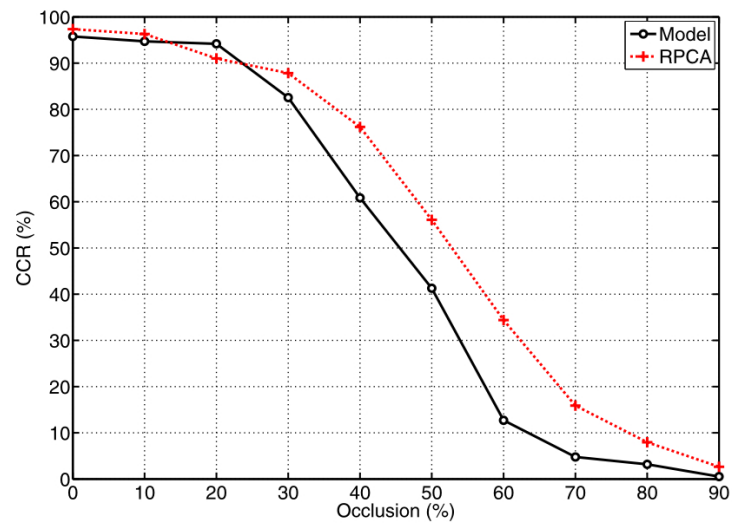
manually labeled. The error bars in this figure represent the uncertainty in labeling the erroneous outputs. Error can be defined either as when parts of the ear are missing in the enrolled image, or when the centre of the enrolled patch does not correspond to the centre of the ear. The latter, which can be considered as the alignment error,

poses a more difficult problem, and more outputs are marked as erroneous. This error constitutes the lower bound of the error in figure 7.9 while the upper bound is the error rate according to the first definition. Overall, occlusion reduces performance and increases uncertainty, but by little up to 40% occlusion.

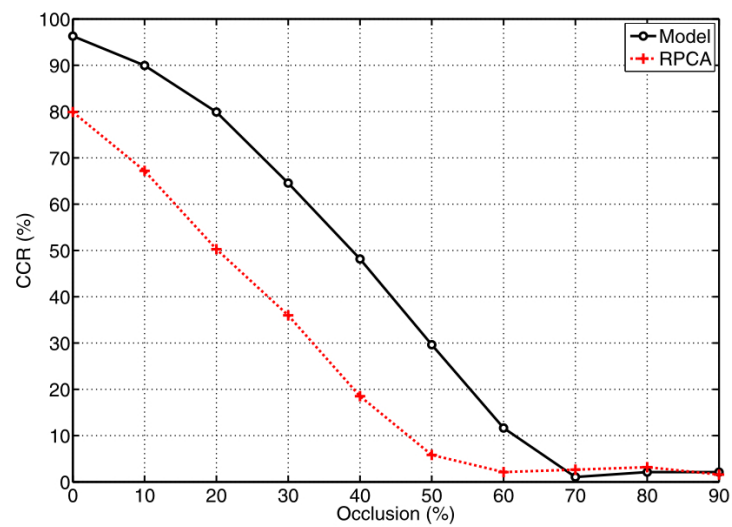
7.3.3 Hybrid Classification and Occlusion

Both our parts-based model and wavelet analysis capture local information, making them potentially reliable when images are occluded. In this section, we shall analyze and discuss the performance of these methods both separately and when combined in comparison to RPCA in occlusion. We shall also discuss various relevant issues, including enrolment errors and scalability

Figure 7.10 presents the model's recognition performance in occlusion, comparing it to RPCA on testsets A and B. On testset A, our model performance is similar to RPCA (see figure 7.10(a)), indicating the occlusion handling capabilities of our model. However, the model shows a more rapid rate of decline from 20% occlusion. Despite the high recognition rates achieved by RPCA, it cannot be readily used, since it is dependent on a pre-processing stage which aligns and registers the samples, such as in testset A, which is manually registered and optimized to accommodate PCA's requirements of good samples. When RPCA is used to classify the ears in testset B, which has the same images as in testset A, but with automatic enrolment, the performance for unoccluded ears drops to 79.9%. In contrast, the model-based recognition shows almost no change, even a slight improvement, at 96.3%. Figure 7.10(b) shows the recognition results on testset B. Although RPCA's gentle decline still depicts robustness to occlusion, there is now a relatively large gap between the two. Note that, as shown in figure 7.9, the automatic enrolment is also impaired by the occlusion, thus the errors are accumulated and a more rapid decline is observed in the model's performance in figure 7.10(b).



(a) Testset A



(b) Testset B

FIGURE 7.10: Model versus RPCA in occlusion on testsets A and B.

On the other hand, our analysis of the outer ear components comprises log-Gabor projections which are compared using a robust matcher. The robust matching is specifically added to handle occluded samples. Figure 7.11 displays the obtained CCRs on testset B both with and without the robust matcher. For unoccluded samples, an improvement of about 3% is obtained as the result of incorporating the robust matcher. However, at 20% occlusion the improvement is 13%, thereby exhibiting the benefits of using the robust matcher in occluded samples. The model-based analysis also has an extra component,

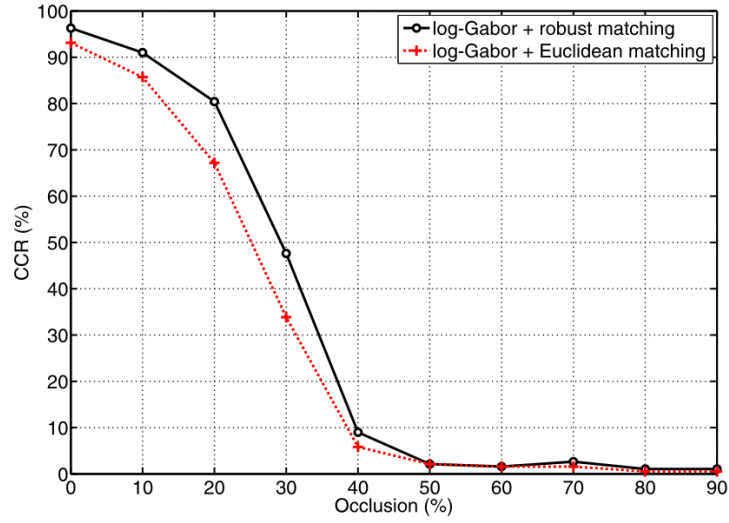
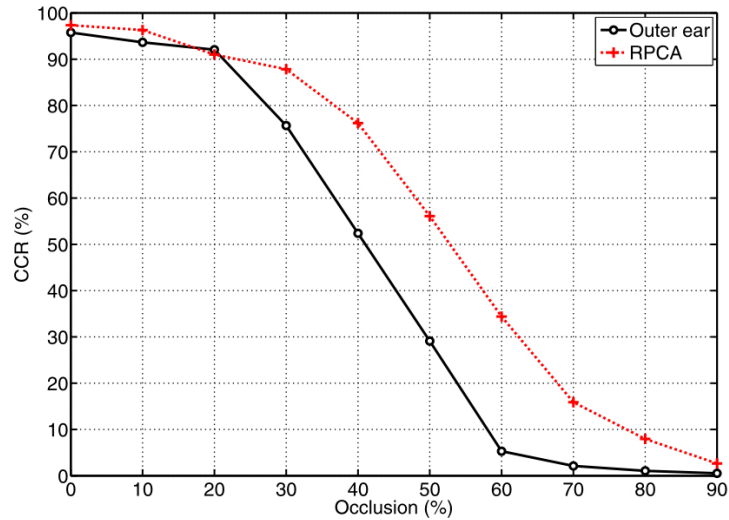


FIGURE 7.11: Significance of the robust matcher in the wavelet analysis on testset B.

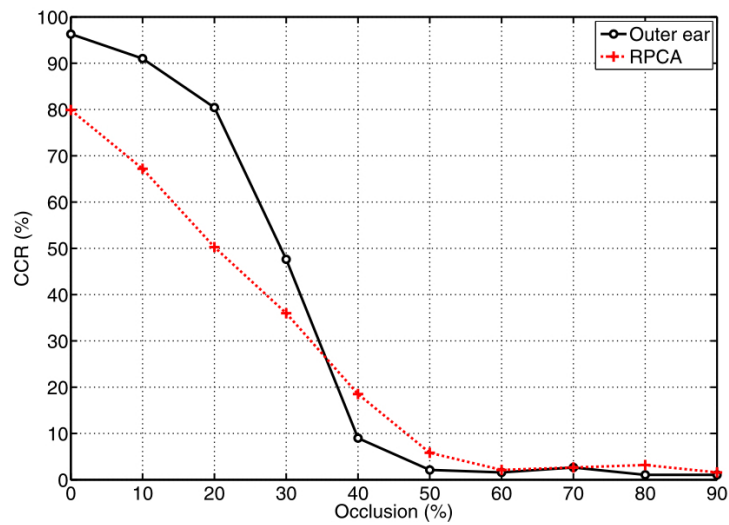
the Tanimoto distance, which evaluates the similarity in the state of detection of the parts. However, in this case, we observe no correlation between the significance of the Tanimoto metric and the extent of the occlusion. We obtain a steady improvement of about 5% in CCRs as the result of incorporating the Tanimoto metric.

Figure 7.12 shows the CCRs of the outer ear analysis and RPCA on testsets A and B. On testset A, the wavelet-based analysis also exhibits a gentle decline in performance which is comparable to RPCA. However, its performance degrades at a much faster rate on testset B. As mentioned before, the enrolment error on testset B forces more erroneous classifications on this set. Additionally, we attribute some of the error to the hierarchical arrangement of the recognition components. In this arrangement, the hub of the target area analyzed by the wavelets is determined using a set of model parts. Thus the process is disturbed not only by the errors in enrolment, but also by the erroneous detection of the model parts, albeit, the CCRs obtained by this method on testset B

still outperform RPCA for occlusions less than 40%.



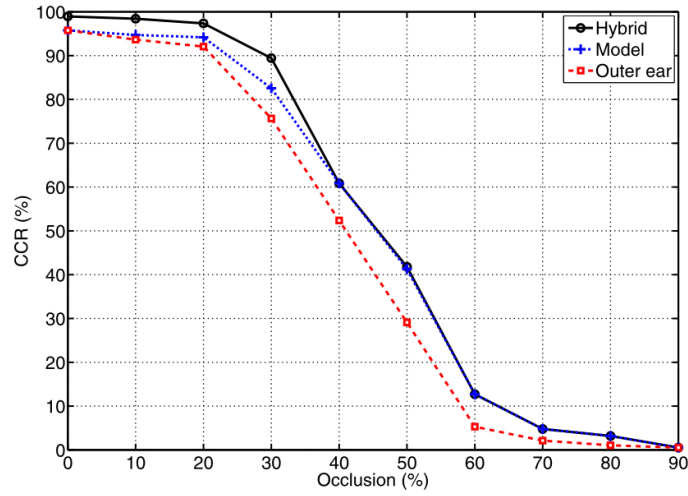
(a) Testset A



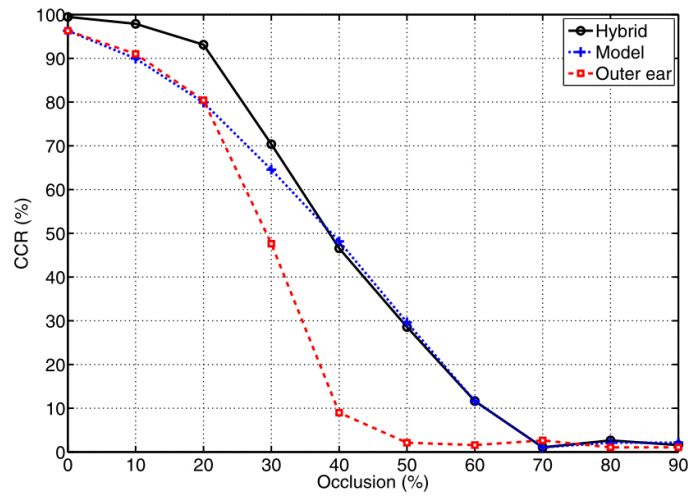
(b) Testset B

FIGURE 7.12: The outer ear analysis versus RPCA in occlusion on testsets A and B.

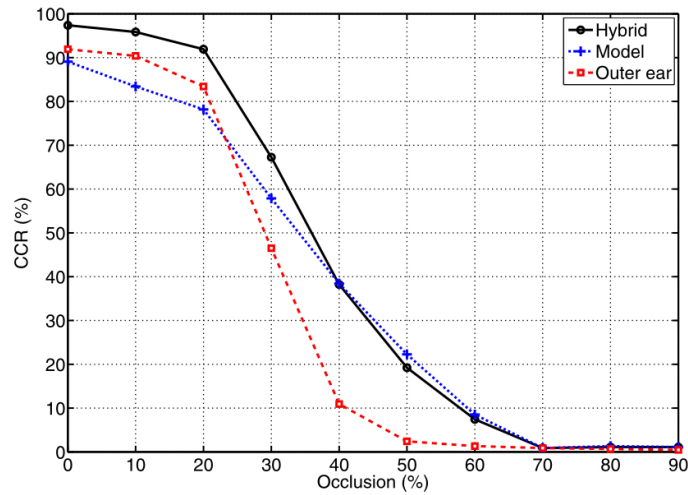
In section 7.2, we described the hybrid classification, which is formed by the addition of the outer ear analysis to the model. We showed that this hybrid classifier is superior to both, the model and the wavelet-based technique, separately. Here, we shall look at the hybrid performance in occlusion. Figure 7.13 shows the hybrid CCRs in occlusion on testsets A, B and C. It can be seen that the hybrid classification improvement is maintained as the ears are increasingly occluded. However, due to the more rapid rate of decline in the outer ear analysis, especially on testsets B and C, the hybrid performance gradually approaches that of the model. At about 40% of occlusion, the



(a) Testset A



(b) Testset B



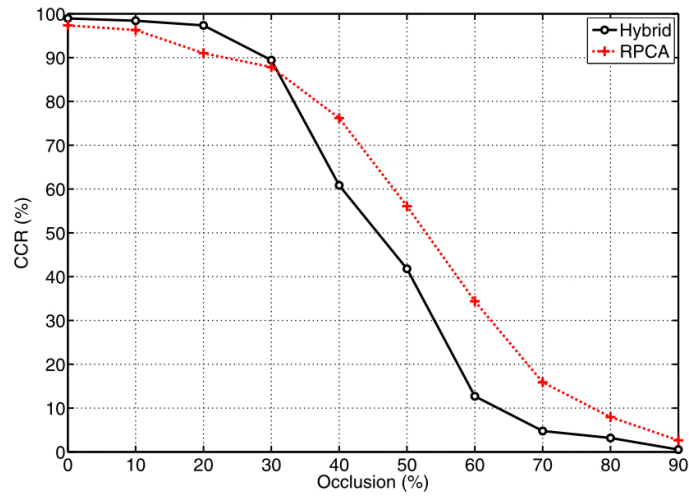
(c) Testset C

FIGURE 7.13: Improvement obtained by fusing the model-based and outer ear metrics; hybrid classification in occlusion on testsets A, B and C.

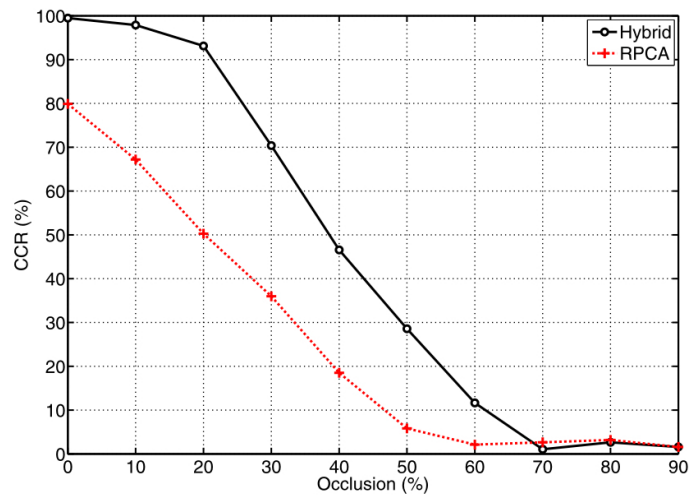
hybrid performance is identical or slightly worse than the model.

Figure 7.14 displays the CCRs of our hybrid classification and RPCA on testsets A, B and C. On testset A, our hybrid method performs better than RPCA for as much as 30% of occlusion; although for more occluded samples RPCA offers better recognition. Clearly, as the occlusion from the top intensifies the significance of the lower structures increases. The ear lobe is one of the most prominent and distinctive parts of the lower ear. However, in our analysis we intentionally neglected the lobe, due to the likelihood of change in its shape with age, whereas in RPCA all image information is treated identically, thus making optimal use of all information available. The results on testsets B and C mainly exhibit the degrading effect of less accurate registration on RPCA. In contrast, the hybrid classification maintains good performance, and clearly outperforms RPCA.

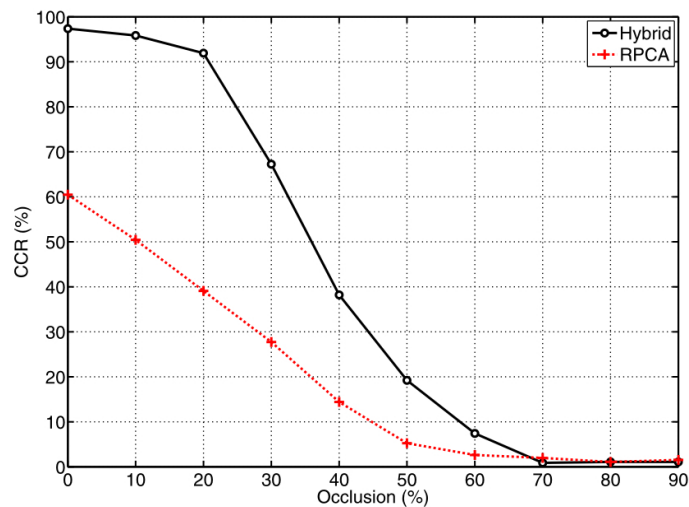
Figure 7.15 compares the hybrid classification results on testsets A,B and C and similarly the RPCA performance on these three testsets. Recall that testset C comprises 458 images to testset B's 189 images. It also introduces more complications in terms of the number of individuals, number of samples per individual, overall image quality and occlusion. In spite of the increased complexity, the point to point comparison (0% to 50% occlusion) of hybrid CCRs show a mean decrease in performance of only 3.4% between testsets B and C, while RPCA performance drops by a mean of 11.9%. For unoccluded samples, the hybrid recognition rate drops by 2.1%, while RPCA recognition rate drops by 19.4%. Therefore our hybrid classification exhibits better scalability traits compared to RPCA.



(a) Testset A

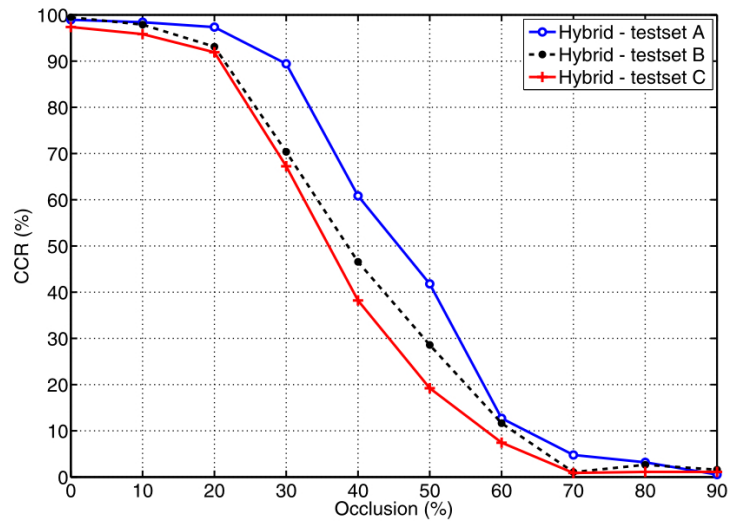


(b) Testset B

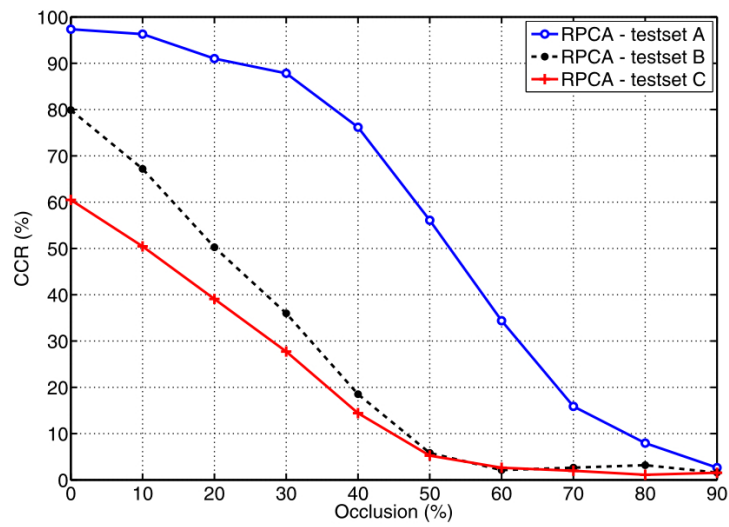


(c) Testset C

FIGURE 7.14: The hybrid classifier versus RPCA in occlusion on testsets A, B and C.



(a) Hybrid recognition



(b) RPCA recognition

FIGURE 7.15: The hybrid classification and RPCA scalability across different testsets.

Table 7.5 summarizes the CCRs at 10% and 30% occlusion obtained by our methods as compared to RPCA. The results achieved on testset E, which is the testset with real occlusion, are shown in table 7.6. In this, hybrid classification obtains a 68.9% recognition rate to RPCA’s 26.4%. It can be seen that the hybrid classifier performs better than RPCA in all cases particularly when automatic enrolment is used.

TABLE 7.5: Comparing the correct classification rates in occlusion

Testsets	A		B		C		D	
Occlusion	10%	30%	10%	30%	10%	30%	10%	30%
Model	94.7%	82.5%	89.9%	64.6%	83.4%	57.9%	84.0%	64.3%
Outer ear	93.7%	75.7%	91.0%	47.6%	90.4%	46.5%	90.7%	44.2%
Hybrid	98.4%	89.4%	97.9%	70.4%	95.9%	67.2%	95.5%	68.0%
RPCA	96.3%	87.8%	67.2%	36.0%	50.4%	27.7%	49.8%	30.1%

TABLE 7.6: Comparing the correct classification rates on testset E

Testset E	
Model	70/148 \sim 47.3%
Outer ear	96/148 \sim 64.9%
Hybrid	102/148 \sim 68.9%
RPCA	39/148 \sim 26.4%

7.4 Conclusions

In this chapter, we have evaluated the performance of the various stages of our ear biometric approach. The ears of the main data are automatically enrolled with a 99.2% success rate. Also on the occluded dataset we obtain a 95.8% correct enrolment. Synthetically occluded ears have also been enrolled. In recognition, our hybrid classifier, which incorporates the outer ear analysis into the model, improves the performance significantly. This method obtains 99.5% and 97.4% recognition rates on testsets B and C respectively. On testset C, which is our largest dataset, the model and the outer ear analysis have obtained 89.1% and 91.9% recognition rates respectively.

In occlusion, a Robust PCA is used as an example of holistic methods, and it performs well in occlusion on testset A. The model and the outer ear analysis also exhibit occlusion tolerant traits. However, in contrast to RPCA, they are also tolerant of potential misalignments in enrolled samples. The hybrid recognition maintains its advantage over the model and outer ear analysis in occlusion. However, it approaches the model’s

performance if large parts of the ear are occluded. Our hybrid classifier mostly outperforms RPCA, even to some extent on testset A, which is tuned for PCA-based methods. Notably, on testset E, wherein the probe samples are occluded by hair, the hybrid classification obtains a 68.9% recognition rate compared to RPCA's 26.4%. The hybrid performance also exhibits better scalability; it degrades by only 2% between testsets B and C, where the testset size grows by two and a half times, while RPCA degrades by 19%.

Chapter 8

Conclusions and Future Work

8.1 Conclusions

This thesis capitalizes on a new guided model for ear biometrics. This is the first model-based approach to ear biometric recognition. By evidence from embryonic development of the human ear, it has been shown that the complex structure of the ear is in fact composed of individual components. We have thus proposed a new parts-based model for ears. Although there is no consensus as to the precise embryology of the external ear, certain aspects of the proposed hypotheses, which apportion different hillocks and combinations to ear formation, are similar. External ear anomalies also offer cues as to the sites and types of variations in the ear structure. For the purpose of ear biometric recognition, we have derived a new arrangement of parts for the ear structure, highlighting the sites of independent variations.

Our new model is learned using a stochastic clustering method which detects the clusters of SIFT keypoints in the training set. We have illustrated that even though not all the structures which appear in the model correspond with specific ear components, a feature subset selection has revealed that the most significant model parts in recognition are in fact exhibiting ear components. This also shows that various embryonic parts of the ear form clusters, when examined by a local feature descriptor such as SIFT, and are thereby detectable in different ear images. Our results have confirmed the validity of this approach not just in modelling, but also by recognition capabilities. By analysis on an extended selection from the XM2VTS database the new approach achieves a 96.3%

correct classification rate (CCR) on a dataset (Testset B) which includes 189 images. On a more challenging dataset, with 458 images including samples with lower image quality (Testset C), the model obtains an 89.1% CCR.

Although it has been shown that the recognition capability of the model is derived from a set of specific ear components, not all components of the ear are sufficiently explored by this model. It appears that the helix and the antihelix are under-represented in the model. Extending our model description, we have used a new wavelet-based approach with a specific aim to capture information in these boundary structures. In this, the frequency content of the radial signals of the image intensities outwards from the concha are explored via log-Gabor filters. These radii, which are mostly normal to the helix curve, capture the variation in the helix and the antihelix at each angle. We have illustrated that this new approach is suitable for capturing the information in these components. By embryological evidence, much shape variation can occur within the body of the helix and the antihelix, emphasizing their fruitfulness in recognition. Using only our new wavelet-based method, we have obtained promising results, achieving a 96.3% CCR on testset B and a 91.9% CCR on testset C. We have then added the wavelet-based analysis to our model using a simple decision fusion technique. The resulting hybrid method exhibits significant improvement with 99.5% and 97.4% CCRs on testsets B and C, respectively, thereby supporting our hypothesis that the outer ear analysis captures some new and discriminant information. Note that our new automatic enrolment offers an error-free enrolment for testset B, and on testset C it achieves a 99% success rate.

Considering that the potential occlusion by hair is one of the main obstacles in the deployment of ear biometrics, throughout this thesis our techniques were chosen from amongst the methods which offer occlusion tolerant properties. The main deployment advantage of a model is its robustness in noise and occlusion. Also, our Hough transform based enrolment accrues tolerance to occlusion. Wavelets offer decimation in space which is beneficial when handling partially occluded data. Finally, a more robust matching has been applied to handle erroneous wavelet projections. We have evaluated the performance in occlusion on synthetically occluded data as well as on a dataset of ear images occluded by hair. For comparison purposes, a robust PCA (RPCA) is used as a representative of holistic methods, and is shown to have good results on testset A, for both original samples and when occluded. Considered separately, our initial model

and our wavelet-based technique exhibit occlusion tolerant traits on this testset. Our hybrid method obtains a better performance than RPCA on testset A for as much as 30% occlusion. Despite being tolerant of occlusion, RPCA – which extracts no image-based features – has no other invariant properties. Thus it can only maintain this high performance, if the images are well aligned and normalized for scale and rotation, as in testset A. In testset B, which includes the same images of testset A, but only automatically enrolled, although no error has been noted in enrolment, the exact registration and normalization, which is obtained manually for testset A, is not offered. In this, the RPCA performance drops by 17%, while our model and wavelet-based method show no change and even a slight improvement. It has been shown that occlusion increases the uncertainty in the automatic enrolment results, meaning that a larger proportion of enrolled samples are misaligned, and thus the RPCA performance degrades at a faster rate. Our hybrid method clearly outperforms RPCA on testsets B and C, which are enrolled automatically. Notably, on testset E, wherein the probe samples are occluded by hair, the hybrid classification obtains a 68.9% CCR to RPCA's 26.4%. The hybrid performance also exhibits a good scalability; it degrades by only 2% between testsets B and C, where the testset size grows by two and a half times.

8.2 Future Work

One of the main issues to be addressed is the need for larger datasets of ear images, wherein a more accurate estimate of the recognition performance can be obtained, and potential variations in the performance can be analyzed. Ear biometrics also benefits from working with datasets which address the covariates (exploratory variables) of ear data directly. One of the main covariates to be analyzed is the effects of aging on the ear structure. Although it appears that the ears are consistent with age, for a concrete analysis, we need a dataset with big time lapses between the captured samples. Other factors to be considered for the future datasets include: partial occlusion; pose variation; lighting and background changes. The correlation between the shape of the left and right ears can also be examined in a dataset which captures images from both ears for each individual.

In this thesis, we have examined the effects of partial occlusion on ear recognition. We have also noted that our method does not require exact segmentation and normaliza-

tion of the ear samples. It is also interesting to examine the effects of pose variation and lighting changes, which potentially alter the visual characteristics of the structure. Model-based approaches have potential advantage in handling pose variations. Furthermore, SIFT keypoints offer partial invariance to view point changes. Our method can also be tuned to handle the expected scale and rotation variations in the dataset. However, this property has not been tested since our data does not exhibit large scale or rotation changes.

Given that the components of the ear exhibit different types of visual characteristics, we have found that the model parts are less suited to capture information in some of these components. Noting that the helix and the antihelix are only partially explored by the model, we have applied a new wavelet-based technique to capture the information in the helix and the antihelix. Further employing the same idea, the recognition performance might benefit from using other specialized methods to capture information in other ear components. For example, the tragus, which is shown to be one of the most variant parts of the ear, by surgical studies, is not highlighted by the model as a particularly useful part in recognition. Thus incorporating the features of the tragus directly may improve the performance. Also note that in the model the significance of a part in recognition depends on the rate and accuracy of detection as well as the discriminating capabilities of the captured component. Alternatively, the model parts can be used in robust detection of ear components via voting. Subsequently, by capitalizing on individual ear components, we may obtain a better estimate of the significance of each component in recognition.

Finally, the major avenue for our future research is to build an automatic ear recognition system for real-time applications. This research suggests that it is indeed feasible to achieve recognition by planar ear structure, and the research herein could guide this development.

Appendix A

Further Work on the Model

In a related field of research, parts-based models have been used to recognize object categories [30, 79, 27]. The main question is whether an instance of a certain object is present in an image. In a work by Fergus et al. [30], the model consists of not only the appearance, location and the scale of each part, but it also contains the variance and the covariances; the probabilities of arbitrary occlusion patterns and an estimate of the background. A probabilistic approach based on an expectation maximization (EM) algorithm is used to learn this model. This method has potential benefits over our parts-based model in that it extracts additional information from the training set. Also, the use of probabilistic methods reduces the need for tuning various thresholds. Here, we further investigate the benefits of incorporating this method into our ear model.

Given the model parameters, θ_{fg} , Fergus et al. detect the model parts in an image by computing the posterior density of valid hypotheses. A hypothesis, \mathbf{h} , is an allocation of image features to model parts. The posterior density of a hypothesis, \mathbf{h} , is calculated by:

$$p(\mathbf{h}|\mathbf{X}, \mathbf{S}, \mathbf{A}, \theta_{fg}) = \frac{p(\mathbf{X}, \mathbf{S}, \mathbf{A}, \mathbf{h}|\theta_{fg})}{\sum_{h \in H} p(\mathbf{X}, \mathbf{S}, \mathbf{A}, \mathbf{h}|\theta_{fg})} \quad (\text{A.1})$$

where, \mathbf{X} , \mathbf{S} and \mathbf{A} are the locations, scales and appearances of the feature points in the image respectively. H is a set of valid hypotheses. Despite the benefits, this method also introduces some limitations. Its high computational complexity imposes a practical limit on the number of parts in the model [30], and it also requires a good initialization. For initialization, considering that much of the recognition potential of our model is derived from a small subset of parts, as discussed in section 5.4.2, we choose the seven

most important parts from table 5.3.

Commencing learning, we have found that some of the model parts deviate from these initial values, and assume new identities. Figure A.1 compares the performance of our model with the new modified version in part detection. The matches to the six most important parts of our model are shown in figure A.1(a). The first row in this figure shows the input images. A CCR of 89.9% is obtained using these parts. Figure A.1(b) displays the new matches obtained by the modified model using the parts in figure A.1(a) as initial values. It can be seen that part 3 has deviated significantly from its original value, and parts 4 and 5 produce inconsistent results. We hypothesize that the deviation from these clusters is caused by the fact that, although, these are the most discriminant parts, as chosen by the SFFS feature selection algorithm, they are not necessarily the best clusters. In recognition, the performance drops to a 51.9% CCR. We can partially prevent this deviation by tuning the initial variance and covariances to smaller values. Thereby, we obtain an estimate for the extra parameters in θ_{fg} , while retaining the initial clusters. However, early performance evaluations again exhibit no significant improvement in part allocation or in recognition. Figure A.1(c) displays the matches to the model parts obtained using this method. In this, the parts retain their original identity and the performance is similar to the current model. However, part 4 obtains inconsistent matches. In recognition, the new restrained method obtains a 77.2% CCR. By removing the noisy part, part 4, the performance increases to 84.1% with five parts.

We hypothesize that the specific properties of the parts in our ear model have caused the new information to appear insignificant. Note that the joint probability density $p(\mathbf{X}, \mathbf{S}, \mathbf{A}, \mathbf{h}|\theta_{fg})$, in equation A.1, is determined as follows:

$$p(\mathbf{X}, \mathbf{S}, \mathbf{A}, \mathbf{h}|\theta_{fg}) = \overbrace{p(\mathbf{A}|\mathbf{X}, \mathbf{S}, \mathbf{h}, \theta_{fg})}^{\text{Appearance}} \times \overbrace{p(\mathbf{X}|\mathbf{S}, \mathbf{h}, \theta_{fg})}^{\text{Shape}} \times \overbrace{p(\mathbf{S}|\mathbf{h}, \theta_{fg})}^{\text{Scale}} \times \overbrace{p(\mathbf{h}|\theta_{fg})}^{\text{Occlusion}} \quad (\text{A.2})$$

where the likelihood components include appearance, shape, scale and occlusion likelihoods, respectively. We have found that in our model, $p(\mathbf{X}, \mathbf{S}, \mathbf{A}, \mathbf{h}|\theta_{fg})$ is dominated by the appearance likelihood. Recall that the parts constituting the ear model are highly distinctive parts, each describing a specific ear component. The difference between the first and the second appearance matches for these parts is often large, thus it is viable to detect these parts mainly on the basis of appearance similarity. In contrast,

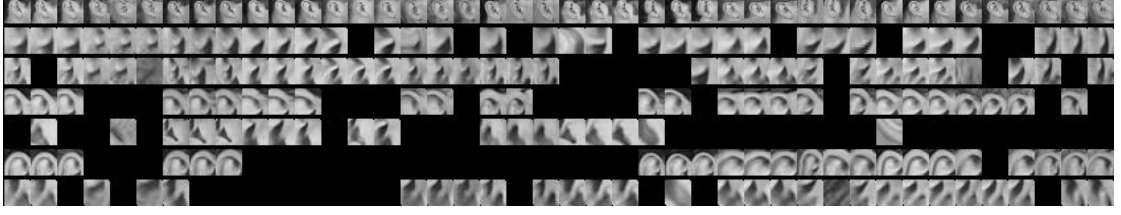
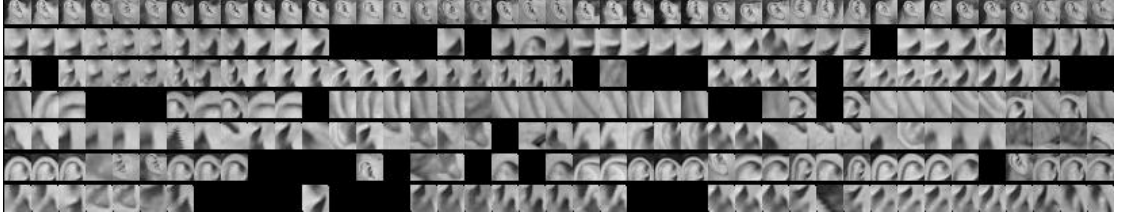
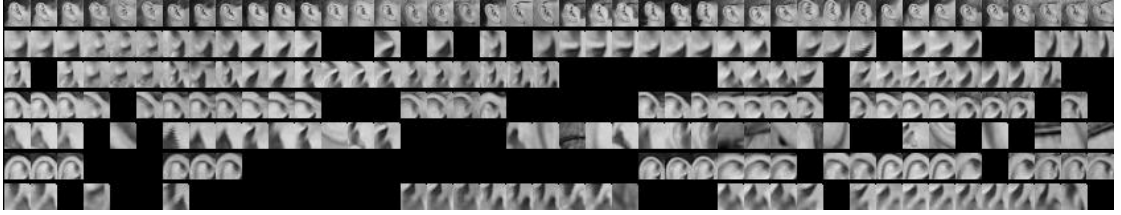
(a) Our current model — CCR: (170/189 \sim) 89.9%(b) New modified version using the probabilistic approach — CCR: (98/189 \sim) 51.9%(c) New restrained version — CCR: (146/189 \sim) 77.2% and (159/189 \sim) 84.1% excluding part 4.

FIGURE A.1: The six most important parts of our model are displayed in (a). The new modified version, using the probabilistic approach, detects the parts depicted in (b). It can be seen that the third part has deviated from its original position, and parts 4 and 5 are noisy. By partially restraining the appearance deviation, the parts in (c) have been derived, which are similar to the parts in (a). However, in (c) part 4 produces inconsistent results. The first row of these figures show the input images.

a part which is categorized as a ‘basic feature’ (see section 5.4) would exhibit a good appearance match with many more image features, and therefore to obtain the correct correspondence the other likelihood components are essential. Given the specific properties of our model and by preliminary analysis, we propose that the contributions of this probabilistic model and the new information it adds are not particularly beneficial to our research at this stage.

References

- [1] A. Abbas and G. N. Ritty. Ear piercing affects earprints: the role of ear piercing in human identification. *Journal of Forensic Sciences*, 50(2):386–392, 2005.
- [2] M. Abdel-Mottaleb and J. Zhou. Human ear recognition from face profile images. In *Proc. 1st International Conference on Biometrics (ICB'06)*, pages 786–792, 2006.
- [3] A. S. Aguado, E. Montiel, and M. S. Nixon. On using directional information for parameter space decomposition in ellipse detection. *Pattern Recognition*, 29(3):369–381, 1996.
- [4] A. H. M. Akkermans, T. A. M. Kevenaer, and D. W. E. Schobben. Acoustic ear recognition for person identification. In *Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'05)*, pages 219–223, 2005.
- [5] I. Alberink and A. Ruifrok. Performance of the fearid earprint identification system. *Forensic Science International*, 166:145–154, 2007.
- [6] L. Alvarez, E. Gonzalez, and L. Mazorra. Fitting ear contour using an ovoid model. In *Proc. of 39 IEEE International Carnahan Conference on Security Technology*, pages 145–148, 2005.
- [7] A. Bertillon. *La photographie judiciaire, avec un appendice sur la classification et l'identification anthropométriques*. Gauthier-Villars, Paris, 1890.
- [8] P. J. Besl and N. D. McKay. A method for registration of 3-d shapes. *IEEE Trans. Pattern Analysis and Machine Intelligence*, pages 239–256, 1992.
- [9] M. J. Black and A. D. Jepson. Eigentracking: Robust matching and tracking of articulated objects using a view-based representation. *International Journal of Computer Vision*, 26(1):63–84, 1998.

- [10] M. Burge and W. Burger. Ear biometrics. In A. Jain, R. Bolle, and S. Pankanti, editors, *BIOMETRICS: Personal Identification in a Networked Society*, pages 273–286. Kluwer Academic, 1998.
- [11] M. Burge and W. Burger. Ear biometrics in computer vision. In *International Conference on Pattern Recognition(ICPR'00)*, pages 822–826, 2000.
- [12] J. D. Bustard and M. S. Nixon. Robust 2D ear registration and recognition based on sift point matching. In *IEEE Conference on Biometrics: Theory, Applications and Systems(BTAS'08)*, Washington DC, September 2008.
- [13] K. Chang, K. W. Bowyer, S. Sarkar, and B. Victor. Comparison and combination of ear and face images in appearance-based biometrics. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 25(9):1160–1165, 2003.
- [14] H. Chen and B. Bhanu. Human ear recognition in 3D. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 29(4):718–737, 2007.
- [15] H. Chen, B. Bhanu, and R. Wang. Performance evaluation and prediction for 3-d ear recognition. In *Proc. of International Conference on Audio and Video based Biometric Person Authentication*, pages 748–757, NY, 2005.
- [16] V. Cherkassky and F. Mulier. *Learning from Data: Concepts, Theory, and Methods*. John Wiley & Sons, Inc., NY, USA., 1998.
- [17] M. Choras. Ear biometrics based on geometrical feature extraction. *Electronic Letters on Computer Vision and Image Analysis (Journal ELCVIA)*, 5(3):84–95, 2005.
- [18] M. Choras. Further developments in geometrical algorithms for ear biometrics. In F.J Perales and B. Fisher, editors, *Articulated Motion and Deformable Objects*, pages 58–67. Springer Berlin, 2006.
- [19] R v. Dallagher [2002] EWCA Crim 1903, 2002.
- [20] J. Daugman. Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Research*, 20(10):847–856, 1980.
- [21] J. Daugman. High confidence visual recognition of persons by a test of statistical independence. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 15(11):1148–1161, 1993.

- [22] J. Daugman. Biometric decision landscapes. Technical Report TR482, University of Cambridge, Computer Laboratory, 2000.
- [23] J. Davis. Surgical embryology. In *Aesthetic and reconstructive otoplasty*, pages 93–125. Springer Verlag, New York, 1987.
- [24] C. Van der Lugt. *Earprint Identification*. Elsevier Bedrijfsinformatiem, Gravenhage, 2001.
- [25] D. Dessimoz and C. Champod. Linkages between biometrics and forensic science. In A. Jain, P. Flynn, and A. Ross, editors, *Handbook of Biometrics*, pages 425–459. Springer, 2008.
- [26] C. Dorai and A. Jain. Cosmos-a representation scheme for free-form surfaces. In *Proc. IEEE Conf. Computer Vision*, pages 1024–1029, 1995.
- [27] G. Dorko and C. Schmid. Selection of scale-invariant parts for object class recognition. In *Proc. of the 9th International Conference on Computer Vision*, pages 634–640, Nice, France, 2003.
- [28] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern classification*. Wiley-Interscience, New York, 2001.
- [29] R. Fergus. *Visual Object Category Recognition*. PhD thesis, University of Oxford, 2005.
- [30] R. Fergus, P. Perona, and A. Zisserman. Weakly supervised scale-invariant learning of models for visual recognition. *International Journal of Computer Vision*, 71(3):273–303, March 2007.
- [31] D. J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America*, 4(12):2379–2394, 1987.
- [32] London center for ear reconstruction, 2006. <http://www.earreconstruction.co.uk/index.html>.
- [33] S. Gore, S. Myers, and D. Gault. Mirror ear: A reconstructive technique for substantial tragal anomalies or polyotia. *Journal of Plastic, Reconstructive & Aesthetic Surgery*, 59(5):499–504, 2006.

- [34] R. Hsu, M. Abdel-Mottaleb, and A. K. Jain. Face detection in color images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(5):696–706, 2002.
- [35] A. Hunter and T. Yotsuyanagi. The external ear: More attention to detail may aid syndrome diagnosis and contribute answers to embryological questions. *American Journal of Medical Genetics*, 135A:237–250, 2005.
- [36] D. J. Hurley. *Force Field Feature Extraction for Ear Biometrics*. PhD thesis, Electronics and Computer Science, University of Southampton, 2001.
- [37] D. J. Hurley, B. Arbab-Zavar, and M. S. Nixon. The ear as a biometric. In A. Jain, P. Flynn, and A. Ross, editors, *Handbook of Biometrics*, pages 131–150. Springer, 2008.
- [38] D. J. Hurley, M. S. Nixon, and J. N. Carter. Force field energy functionals for image feature extraction. In *Proc. 10th British Machine Vision Conference*, pages 604–613, 1999.
- [39] D. J. Hurley, M. S. Nixon, and J. N. Carter. Force field energy functionals for image feature extraction. *Image and Vision Computing, Special Issue on BMVC 99*, 20:311–317, 2002.
- [40] D. J. Hurley, M. S. Nixon, and J. N. Carter. Force field feature extraction for ear biometrics. *Computer Vision and Image Understanding*, 98:491–512, 2005.
- [41] A. Iannarelli. *Ear Identification*. Paramount Publishing Company, Freemont, California, 1989.
- [42] S. M. S. Islam, M. Bennamoun, and R. Davies. Fast and fully automatic ear detection using cascaded adaboost. In *Proc. of IEEE Workshop on Applications of Computer Vision(WACV’08)*, pages 1–6, 2008.
- [43] K. Iwano, T. Hirose, E. Kamibayashi, and S. Furui. Audio-visual person authentication using speech and ear images. In *Proc. of Workshop on Multimodal User Authentication*, pages 85–90, 2003.
- [44] H. Izzedine, F. Tankere, V. Launay-Vacher, and G. Deray. Ear and kidney syndromes: Molecular versus clinical approach. *Kidney International*, 65(2):369–385, 2004.

- [45] P. Jaccard. Étude comparative de la distribution florale dans une portion des alpes et des jura. *Bulletin del la Société Vaudoise des Sciences Naturelles*, 37:547–579, 1901.
- [46] A. Jain and D. Zongker. Feature selection: Evaluation, application, and small sample performance. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(2):153–158, 1997.
- [47] P. Kovesi. What are log-gabor filters and why are they good? School of Computer Science & Software Engineering, The University of Western Australia. <http://www.csse.uwa.edu.au/~pk/research/matlabfns/PhaseCongruency/Docs/convexpl.html>.
- [48] P. Kovesi. *Invariant Measures of Image Features From Phase Information*. PhD thesis, University of Western Australia, 1996.
- [49] P. Kovesi. Image features from phase congruency. *Videre: Journal of Computer Vision Research, MIT Press*, 1(3), 1999.
- [50] State v. David Wayne Kunze, 1999. Court of Appeals of Washington, Division 2, 97 Wash. App. 832, 988 P.2d 977.
- [51] H. Liu and J. Yan. Multi-view ear shape feature extraction and reconstruction. In *IEEE Conference on Signal-Image Technologies and Internet-Based System(SITIS'07)*, pages 652–658, December 2007.
- [52] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [53] S. Mallat. *A wavelet tour of signal processing*. Academic Press, San Diego, second edition, 2001.
- [54] L. Meijerman. *Inter- and intra individual variation in earprints*. PhD thesis, Dept. of Anatomy and Embryology, Leiden University, 2006.
- [55] L. Meijerman, S. Sholl, F. De Conti, M. Giacon, C. van der Lugt, A. Drusini, P. Vanezis, and G. Maat. Exploratory study on classification and individualisation of earprints. *Forensic Science International*, 140:91–99, 2004.

- [56] K. Messer, J. Matas, J. Kittler, J. Luetttin, and G. Maitre. XM2VTSDB: The extended M2VTS database. In *Second International Conference on Audio and Video-based Biometric Person Authentication(AVBPA)*, Washington D.C., 1999.
- [57] B. Moreno and A. Sanchez. On the use of outer ear images for personal identification in security applications. In *Proc. IEEE 33rd Annual Intl. Conf. on Security Technology*, pages 469–476, 1999.
- [58] Z. Mu, L. Yuan, Z. Xu, D. Xi, and S. Qi. Shape and structural feature based ear recognition. In *Advances in Biometric Person Authentication*, pages 663–670. Springer, Berlin / Heidelberg, 2005.
- [59] I. Naseem, R. Togneri, and M. Bennamoun. Sparse representation for ear biometrics. In *International Symposium on Visual Computing (ISVC’08)*, pages 336–345, Las Vegas, Nevada, 2008.
- [60] C. Park. Correction of the unilateral question mark ear. *Plastic and Reconstructive Surgery*, 101(6):1620–1623, 1998.
- [61] C. Park. Lower auricular malformations: Their representation, correction, and embryologic correlation. *Plastic and Reconstructive Surgery*, 104(1):29–40, 1999.
- [62] C. Park and T. Suk Roh. Congenital upper auricular detachment. *Plastic and Reconstructive Surgery*, 104(2):488–490, 1999.
- [63] G. Passalis, I. A. Kakadiaris, T. Theoharis, G. Toderici, and T. Papaioannou. Towards fast 3D ear recognition for real-life biometric applications. In *IEEE Conference on Advanced Video and Signal Based Surveillance(AVSS’07)*, pages 39–44, September 2007.
- [64] E. L. Potter. Bilateral renal agenesis. *The Journal of Pediatrics*, 29(1):68–76, 1946.
- [65] P. Pudil, J. Novovicova, and J. Kittler. Floating search methods in feature selection. *Pattern Recognition Letters*, 15:1119–1125, 1994.
- [66] M. M. Rahman and S. Ishikawa. Proposing a passive biometric system for robotic vision. In *Proc. of the Tenth International Symposium on Artificial Life and Robotics*, Oita, Japan, 2005.

- [67] N. Rose. Facial expression classification using gabor and log-gabor filters. In *Automatic Face and Gesture Recognition (FGR)*, pages 346–350, Southampton, UK, 2006.
- [68] G. N. Rutty, A. Abbas, and D. Crossling. Could earprint identification be computerised? an illustrated proof of concept paper. *International Journal of Legal Medicine*, 119(6):335–343, 2005.
- [69] T. W. Sadler. *Langman’s medical embryology*. Lippincott Williams & Wilkins, tenth edition, 2006.
- [70] Johnson center for pregnancy and newborn services, Department of Pediatrics, Stanford School of Medicine. <http://newborns.stanford.edu/PhotoGallery/Ears.html>.
- [71] G. L. Streeter. Development of the auricle in the human embryo. *Contribution to Embryology*, (69):111–139, 1922.
- [72] K. K. Sulik. Embryology of the ear. In R. J. Gorlin, H. V. Toriello, and M. M. Cohen, editors, *Hereditary hearing loss and its syndromes*, pages 22–42. Oxford University Press, 1995.
- [73] T. Theoharis, G. Passalis, G. Toderici, and I. A. Kakadiaris. Unified 3D face and ear recognition using wavelets on geometry images. *Pattern Recognition*, 41(3):796–804, 2007.
- [74] C. Thorne. Otoplasty and ear reconstruction. In C. Thorne, W. Grabb, J. Smith, R. Beasley, S. Bartlett, S. Aston, and G. Gurtner, editors, *Grabb and Smith’s Plastic Surgery*, pages 297–312. Lippincott Williams & Wilkins, 2006.
- [75] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.
- [76] International Criminal Police Organization INTERPOL. Disaster victim identification, 2008. <http://www.interpol.int/Public/DisasterVictim/Default.asp>.
- [77] B. Victor, K. W. Bowyer, and S. Sarkar. An evaluation of face and ear biometrics. In *International Conference on Pattern Recognition(ICPR’02)*, pages 429–432, 2002.
- [78] R. Y. Wang, D. L. Earl, R. O. Ruder, and J. M. Graham. Syndromic ear anomalies and renal ultrasounds. *Pediatrics*, 108(2):e32, 2001.

- [79] M. Weber, M. Welling, and P. Perona. Towards automatic discovery of object categories. In *Proc. of the Computer Vision and Pattern Recognition (CVPR'00)*, pages 101–109, June 2000.
- [80] P. Willett, J. M. Barnard, and G. M. Downs. Chemical similarity searching. *Journal of Chemical Information and Computer Sciences*, 38(6):983–996, 1998.
- [81] F. Wood-Jones and W. I-Chuan. The development of the external ear. *Journal of Anatomy*, 68:525–533, 1934.
- [82] D. L. Woodard, T. C. Faltemier, P. Yan, P. J. Flynn, and K. W. Bowyer. A comparison of 3D biometric modalities. In *Conference on Computer Vision and Pattern Recognition Workshop(CVPRW'06)*, 2006.
- [83] Z. Xiaoxun and J. Yunde. Symmetrical null space LDA for face and ear recognition. *Neurocomputing*, 70:842–848, 2007.
- [84] P. Yan and K. W. Bowyer. 2D and 3D ear recognition. In *Biometric Consortium Conference*, 2004.
- [85] P. Yan and K. W. Bowyer. Empirical evaluation of advanced ear biometrics. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*, page 41, 2005.
- [86] P. Yan and K. W. Bowyer. Empirical evaluation of advanced ear biometrics. In *IEEE Computer Society Workshop on Empirical Evaluation Methods in Computer Vision*, San Diego, 2005.
- [87] P. Yan and K. W. Bowyer. A fast algorithm for ICP-based 3D shape biometrics. In *Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID)*, pages 213–218, NY, 2005.
- [88] P. Yan and K. W. Bowyer. Biometric recognition using 3D ear shape. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 29(8):1297–1308, 2007.
- [89] P. Yan and K. W. Bowyer. A fast algorithm for ICP-based 3D shape biometrics. *Computer Vision and Image Understanding*, 107(3):195–202, 2007.
- [90] P. Yao, J. Li, X. Ye, Z. Zhuang, and B. Li. Iris recognition algorithm using modified log-gabor filters. In *International Conference on Pattern Recognition(ICPR'06)*, pages 461–464, August 2006.

-
- [91] L. Yuan and Z. Mu. Ear recognition based on 2D images. In *IEEE Conference on Biometrics: Theory, Applications and Systems(BTAS'07)*, Washington DC, September 2007.
- [92] T. Yuizono, Y. Wang, K. Satoh, and S. Nakayama. Study on individual recognition for ear images by using genetic local search. In *Proc. of the 2002 Congress on Evolutionary Computation*, pages 237–242, 2002.
- [93] 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. of the Fourth International Conference on Machine Learning and Cybernetics*, pages 4511–4515, 2005.