# Biometric Analysis of Ear Recognition using Shallow and Deep Techniques

*Soumyajit Sarkar*

March 27, 2016

Submitted to the Department of Electrical Engineering & Computer Science and the Faculty of the Graduate School of the University of Kansas in partial fulfillment of the requirements for the degree of Master's of Science

**Thesis Committee:**

Dr. Guanghui Wang: Chairperson

Dr. Bo Luo

Dr. Jerzy Grzymala-Busse

Date Defended

The Thesis Committee for Soumyajit Sarkar certifies
That this is the approved version of the following thesis:

**Biometric Analysis of Ear Recognition using Shallow and Deep Techniques**

Committee:

_____

Chairperson

_____

_____

# Abstract

Biometric ear authentication has received enormous popularity in re-cent years due to its uniqueness for each and every individual, even for identical twins. In this paper, two scale and rotation invariant feature detectors, SIFT and SURF, are adopted for recognition and authentication of ear images. An extensive analysis has been made on how these two descriptors work under certain real-life conditions; and a performance measure has been given. The proposed technique is evaluated and compared with other approaches on two data sets. Extensive experimental study demonstrates the effectiveness of the proposed strategy. Deep Learning has become a new way to detect features in objects and is also used extensively for recognition purposes. Sophisticated deep learning techniques like Convolution Neural Networks(CNNs) have also been implemented and analysis has been done.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

Biometric authentication of people based on various anatomical characteristics, like eye, ear, face, iris, and fingerprint have attracted lots of attention during the past few decades, and some of these techniques have already been successfully applied for recognition and authentication. However, many systems are not very robust and may fail to work under certain conditions. Biometric ear recognition is a relatively new technique that may surpass the existing systems due to several significant reasons. For example, the acquisition of ear images is relatively easy and, unlike iris, can be captured without the co-operation of individuals [1]

Human ear contains rich and stable features which are more reliable than face fea- tures, as the structure of the ear is not subject to change with age. It has also been found out that no two ears are exactly the same even for identical twins [2]. The detailed structure of ear is not only very unique but also permanent, since the shape of a human ear never shows drastic changes over the course of life. The research on ear identification was first conducted by Bertillon, a French criminologist, in 1890. The process was refined by American police officer, Iannarelli [20], who divided the ear based on vari- ous distinctive features of seven parts: i.e. he-

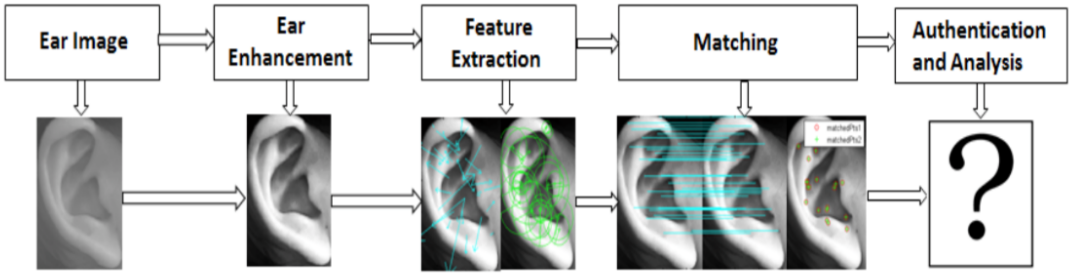



**Figure 1.1.** The pipeline of the proposed Ear Recognition System

lix, concha, antihelix, crux of helix, inter- tragic notch, tragus, and antitragus [3].

Here, we propose to use two scale and rotation invariant feature detectors, i.e. SIFT (scale invariant feature transform) and SURF (speed up robust features), for ear recognition. Both SIFT and SURF extract specific interest points from an image and generate descriptors for the feature points to a form a reliable matching results.

Extensive experiments have been carried out on two different sets of databases to evaluate their performance with respect to various rotations and scales. One of the most important feature of ear images is its easiness in acquisition, however, the acquired images may be in different scales, rotations, and illumination. The scale and rotation invariant property of the SIFT and SURF algorithms makes them perfect for ear authentication under various circumstances.

A new concept in the field of machine learning and computer vision has come up which has surpassed the traditional object recognition methods. This new approach is called deep learning. Deep Learning is a branch of Machine Learning which has multiple levels of representations and abstractions. It is basically a rebranding of the term Artificial Neural Networks. Deep Learning algorithms have already been applied in Apple's Siri, Google's Streetview etc.

The rest is organized as follows. Some background and related research are discussed in Section 2; the proposed method is presented in details in Section 3; some experimental results and analysis are given in Section 4; and the paper is concluded in Section 5.

# Chapter 2

# Statement of the Problem

## 2.1 Types of Biometrics

Biometrics has been an active field of research over the last decade. The reason behind their success is that biometric characteristics are universal, unique and permanent. Unlike other forms of authentication such as passwords or identification cards which can be stolen or faked easily.

There are many kinds of biometrics which can be used for authentication purposes. Among them the prominent being, Face, Ear, Palm, Fingerprint, Iris and others which are frequently being used these days in day to day life to authenticate an individual. Another reason biometrics have been used these days are due to terrorist activities and other fraudulent ways in which people impersonate themselves which are harder to catch. These days biometrics are used everywhere from Airports to ATMs to secured entry to corporate offices where checking the identity of an identity of an individual is mandatory before access is given. It helps to strengthen the security of an organization or country potential threat. As mentioned above, the different types of biometrics, different biometrics have different

purposes and importance. The most popular being face recognition which is being used everywhere to authenticate people, the only disadvantage being the change in facial expression and with age the face changes upto a certain extent which makes it difficult to recognize and authenticate. Fingerprint is also being used in almost any high priority zone nowadays to authenticate and is very successful but it requires complete co-operation of an individual in order to authenticate them. The same problem happens with iris authentication where it becomes very difficult to extract the iris image to match and authenticate.

Ear authentication comes to the rescue in such a situation due to many reasons. The primary being the stability in the human ear structure and ear images can easily be captured without the co-operation of an individual. Each ear is unique, so any side image of an individual is enough in order to authenticate a person.

## 2.2 Purpose of Biometric Ear Recognition

Ear authentication and recognition is being considered as one of the most innovative processes as of today. The human ear can be divided into six main parts: Outer helix, the antihelix, the lobe, the tragus, the antitragus and the concha. The shape of the outer ear evolves during the embryonic state from six growth nodules. The structure is completely random, the randomness can be observed by comparing the left and right ear of the same person - thus they are not symmetric. French criminologist Alphonse Bertillon was the first to be aware of ear to be used for human identification purposes. His work was carried on by Alfred Ianarelli whoc collected 10,000 ear images and determined 12 characteristics needed to identify a person. He also conducted studies on twins and triplets thereby discovering that ears are unique even among genetically identical persons

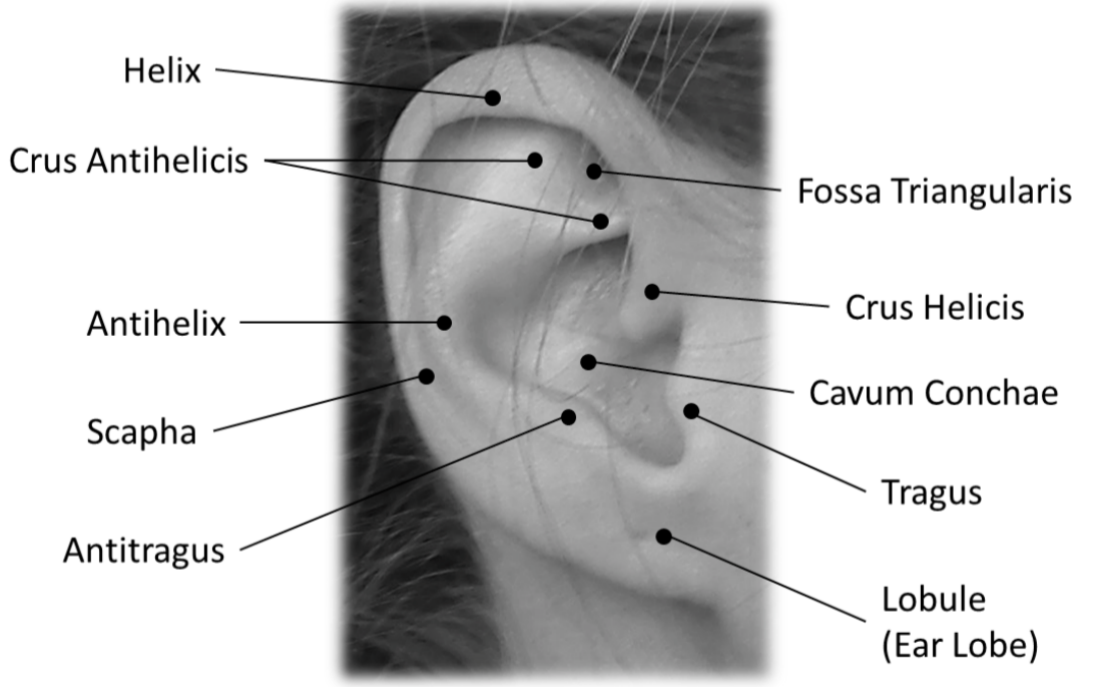



**Figure 2.1.** Characteristics of Human Ear [1]

[1]. The different parts of the ear are shown in Figure 2.1

The typical ear biometric system can be viewed as a system where an input image can be reduced to a set of features that is used to compare with the features of other images to determine its identity. The salient features of a classical ear recognition system are [2]

1. Ear Detection/ Segmentation - The first stage which is used to localize the position of the ear in the image.

2. Ear Normalization and Enhancement - The size of the ear image is normalized for standardization and enhanced using standard image processing techniques in order for more features to be extracted.

3. Feature Extraction - Feature extraction refers to a process in which the ear

image is being reduced to a mathematical model called a feature vector to get information.

4. Matching Features - The features extracted are then compared to the features that are extracted earlier and stored in the database to find a match.

5. Decision - Matching scores are generated by the model used to train the features to give a decision of whether the image is matched or not.

## 2.3 Contributions of this Project

The main goal of this work is to develop shallow and deep techniques to extract efficient features from a set of ear images in order to authenticate a human being. A thorough comparison of two traditional techniques called SIFT(Scale-Invariant Feature Transform)[] and SURF(Speed-up of Robust Features have been provided)[], then another comparison has been done with modern deep learning models constructed with the help of convolution neural networks[]. [4]. My paper is [5]

# Chapter 3

# Background

Human ears start to develop between fifth and seventh weeks of pregnancy. At this stage, the embryo face takes on more definition as mouth perforation, nostrils and ear indentations become visible. Forensic science literature reports that ear growth after the first four months of birth is highly linear [20]. The rate of stretching is five times greater than normal during the period from 4 months to the age of 8, after which, it is constant until the age of seventy when it again increases. Thus it can be said that ear remains almost unchanged during a substantial period of 62 years and, thus, it is one of the strong points of considering ear for biometric authentication.

Haar-based methods have given fairly better results for face detection as it is robust and fast. The different types of ear recognition systems include those of intensity-based, force-field based, 2D curves geometry, wavelet transformation, Gabor filters, SIFT, and 3D features. The force-field transforms gained popularity due to its uniqueness and efficiency [22]. Similar methods have also been implemented on other kinds of ear recognition systems [8][10].

Deep Methods have already come up and showing good performances on other

face recognition systems which shows that it can also be applied to ear recognition systems. Hand-crafted feature detectors have not been able to work properly and are not robust, so deep features have been extracted to improve upon the performance. But one of the few drawbacks about deep learning is that it needs a large amount of data to train the model. There are not many ear databases that are too big but an attempt has been made to apply deep learning on a small scale database and analyze the results.

# Chapter 4

# Related Works

A lot of work has been happening in the ear biometrics over the past decade. The approaches are varied with some working on Intensity-based features while others on 2-D ad 3-D curves etc. Chang et al.[2003] whole worked on the UND database and got an accuracy of 72.7p.c. using the PCA approach. A new concept called Force-Field was being brought by Hurley et al. which gave an accuracy of 99.2 p.c. on the XM2VTS dataset. Many other approaches like 3D Features, Gabor Filters, SIFT, Wavelet Transformation have been applied on different databases and results have been obtained. This project is mostly on the analysis of Biometric Humar Ear datasets on two methods - SIFT and SURF and a comparison is given on the rotation and scaling factors and how the number of features varies on such conditions keeping the real life scenarios in mind where Ear images are not obtained as compared to a dataset.

# Chapter 5

# Design of Proposed Approach

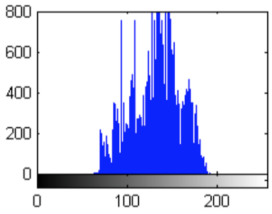
## 5.1 Traditional Approach

Real-life ear images can be acquired in various formats with different scaling and rota- tion conditions. In this paper, we propose to use scale and rotation invariant feature detectors to describe interested features and match them with other images in the data- bases. The proposed ear recognition technique is shown in Figure 1.1. Below is a brief description of each function block.

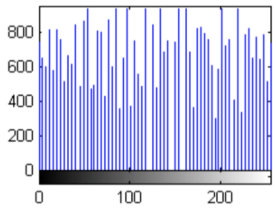
(a) Original Ear Image (b) Enhanced Ear image

**Figure 5.1.** Ear Image Enhancement

The ear enhancement process starts with contrast enhancement, where we apply histo- gram equalization to improve the contrast in an image in order to

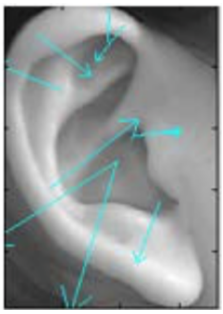



(a) Histogram of Original Image

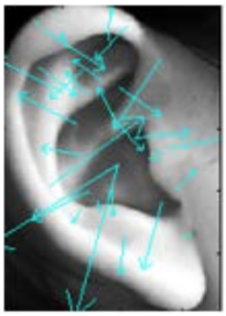



(b) Histogram of Enhanced Image

**Figure 5.2.** Histogram of Enhanced Image



(a) 10 SIFT features detected in the original image



(b) 32 SIFT features detected in the original image

**Figure 5.3.** Histogram of Enhanced Image

stretch out its intensity range, from which, we get an enhanced version of the original image by maximizing the contrast level of an image, as shown in Figure 5.1

Feature Extraction is the process of extracting salient features from the image, and each feature is described by a vector which summarizes the required information for that point [2]. Features are extracted exclusively in order for the image to be matched with the features of the input image to authenticate the ear so that a decision can be made. In this paper, two rotation and scale invariant features are studied.The details are being discussed in the next section.

table 5.1.

**Table 5.1.** Command Set of Scheduler Module, Build 1

| Type | Name | Actions |
|---|---|---|
| TMcom | Enqueue | Schedules a thread |
| TMcom | Dequeue | Removes a thread from the ready-to-run queue |
| BUScom | Get_Entry | Returns a thread's table attribute entry |
| BUScom | Toggle_Preemption | Toggle preemption interrupt on/off |
| BUScom | Get_Entry | Returns a thread's table attribute entry (for debug use) |
| BUScom | Get_Priority | Returns the priority-level of a thread |
| BUScom | Set_Priority | Sets the priority-level of a thread |
| BUScom | Set_Default_Priority | Sets the priority-level of a thread (no error-checking) |

## 5.2 SIFT and SURF Descriptor

**Speed up Robust Features(SURF) -** SURF is a high performance, fast scale and rotation invariant point detector and descriptor. It outperforms previously proposed schemes with respect to repeatability, distinctiveness and robustness [9]. The detector is based on the Hessian matrix and uses a very basic Laplacian-based detector, called difference of Gaussian (DoG). The implementation of SURF can be divided into three main steps. First, interest points are selected at distinctive locations in the image, such as corners, blobs, and T-junctions. Then, the neighborhood of every interest point is represented by a feature vector. This descriptor has to be distinctive and robust to noise, detection errors, and geometric and photometric deformations. Finally, the descriptor vectors are matched between different images. When working with local features, the issue that needs to be settled is the required level of invariance. Here the rotation and scale invariant descriptors seem to offer a good compromise between feature complexity and robustness to commonly occurring deformations, skew, anisotropic scaling, and perspective effects [9].

Given a point in an Image, the Hessian matrix is as follows:

$$H(x,\sigma) = \begin{bmatrix} L_{xx}(x,\sigma) & L_{xy}(x,\sigma) \\ L_{xy}(x,\sigma) & L_{yy}(x,\sigma) \end{bmatrix}$$

where $L_{xx}(x,\sigma)$ is the convolution of the gaussian second order derivative $\frac{d^2}{dy^2}g(\sigma)$ at the point. This method leads to a novel detection, description and subsequent matching steps. Using relative strengths and orientations of gradient reduces the effect of photometric changes. Figure 5.4 shows the detection results with respect to rotation and scale change. As shown in Section 4, it has been found that though SURF is rotation invariant, its performance in matching, i.e. matching score, decreases sharply when the im- ages are rotated or scaled. The SURF features are not stable over various rotation angles and scale changes.

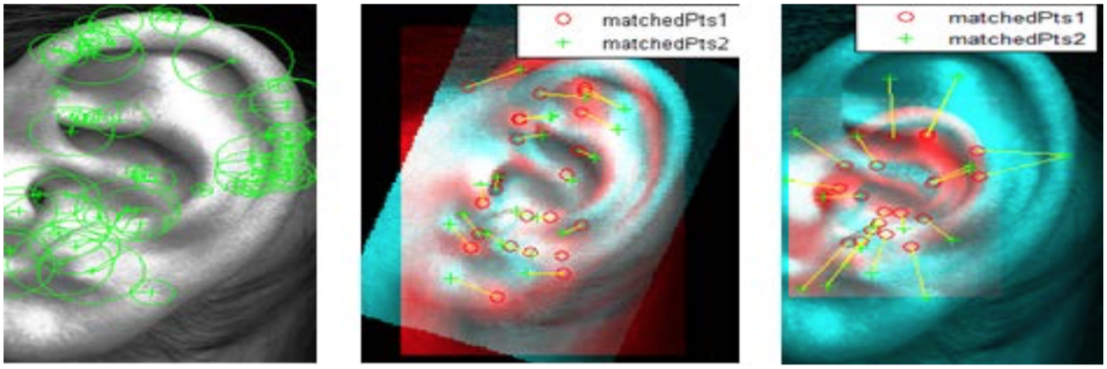



**Figure 5.4.** The detected SURF features (left) and matching result under rotation (middle) and scale change (right)

**Scale Invariant Feature Transform(SIFT) -** The SIFT features are invariant to image scaling and rotation and shown to provide robust matching across a substantial range of affine distortion, change in 3D viewpoint, addition of noise, and change in illumination. The computation stages of SIFT are as follows.

Step 1. Scale space extrema detection: The first step is to construct a Gaussian scale over all the locations. It is implemented efficiently by using a difference of Gaussian (DoG) to identify potential interest points. The 2D Gaussian operator G(x,y,) is convolved with the input image *I(x,y)*:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$$

where the DoG images are obtained by subtracting the subsequent scales in each octave.

$$G(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma)$$

Step 2. Accurate keypoint localization: Once a keypoint has been detected, a detailed model is fitted to determine its location and scale. The keypoints are selected based on measures of their stability. Further details can be found in [16].

Step 3. Orientation assignment: One or more orientations are assigned to each key- point location based on local image gradient directions. All future operations are per- formed on image data that has been transformed relative to the assigned orientation, scale, and location for each feature.

Step 4. Keypoint descriptor: The local image gradients are measured at selected scale in the region around each keypoint. They are transformed into a certain representation that allows for significant levels of local shape distortion and shape illumination.

Figure 5.5 shows an evaluation of the SIFT detector. It is evident the SIFT keypoints are very stable when the images are rotated and scaled. The scaling results are much better compared to the rotation results in our experiments.

**Figure 5.5.** The matching results of SIFT detectors under rotation (left) and scale change (right)

## 5.3 Training Model

Machine Learning models have previously worked wonders on the correct recognition of various algorithms when features extracted are fed into the model for it to figure out the false and the true cases.

For our purpose we have used Multi-class SVMs. Support Vector Machines were originally designed for binary classification. The formulation to solve Multi-class SVM must have variables which are proportional to the number of different classes. The concept of SVM was proposed by Vapnik [vapnik paper-face recog] et

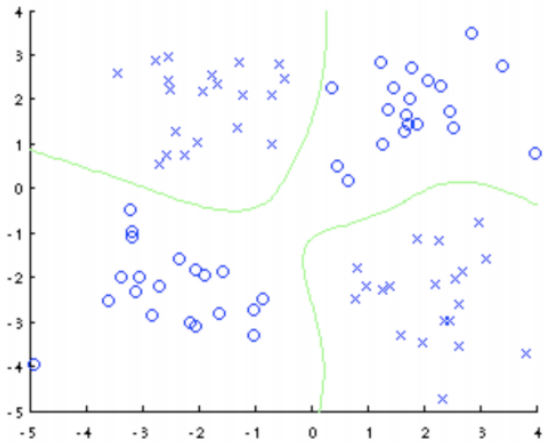



**Figure 5.6.** Multi-class SVM(from [libSVM paper])

al. They helped to classify almost everything right from linear problems to multi-dimensional problems where the kernel matrix is use to transform the conflicting points into a different dimensional space called the kernel space in order to draw a hyperplane in a conclusive manner so as to separate the different cases. Here multiclass SVM is being used in order to classify the features as extracted by the various feature extraction techniques like SURF and SIFT. After that the model is trained and the features are matched with the features obtained from the query ear image in order to find a nearest match to succeed. Since it is multi-class, thus it helps to create separate classes for different classes of images and helps to classify them when the matching process is being done. The main process of better classification of the model depends on the input features. So as a matter of fact we can say that better the feature extraction is being done, better will be the classification made by the multiclass SVM and thus better will be the results.

The Results can be found in chapter6.

## 5.4 Deep Learning Approach

## 5.5 Convolution Neural Network

## 5.6 Our Deep Network and Model

# Chapter 6

# Implementation Results

The modules from both the methods have been tested under simulated conditions. The performances of both the traditional and deep approaches have been given and a comparison has been made on the results. The conditions under which the results were tested have also been provided, detailed explanation has also been given in 6.3.

## 6.1 Results of the Traditional Approach

The proposed approach has been evaluated on two data sets. One is the AMI database [7], which consists of 175 ear images; and the other is the IIT Delhi database [5], which consists of 494 images of 125 distinct persons. The images were all converted to gray- scale images for ease of work. It has also been found out that contrast enhancement is an important factor for feature detection and matching, because it makes the feature detectors find better set of keypoints and increase the effectiveness of matching. According to the experiments performed, it has been found that upper helix, antihe- lix, and tragus are the most important regions for

feature selection compared to others. These regions contribute to about 64p.c. of the feature points. Figure 7 shows some sample images from the two databases we used for our exper- iments. The graphs in Figure 8 indicates the average number of keypoints found and matched by SIFT and SURF detectors when the images are rotated from a range of 0 to 180 degrees. The results suggest that the SIFT detector is fairly stable over a variation of angles from 20 to 160 degrees, whereas the SURF detector, though faster and rotation invariant, is not very stable. Table 1 shows the keypoints detected and matched by the SIFT and SURF detectors, where the performance ratio is the ratio of the number of matched points to that of detected features. It is obvious that the SIFT algorithm performs better when the sizes of images are decreased, while the SURF algorithm performs better when the image sizes are increased. However, the amount of detected keypoints by the SIFT detector is always higher than that by the SURF detector.

**Table 6.1.** SIFT and SURF detection and matching results at different scales

| Scaling | Methods | 0.25 | 0.5 | 0.75 | 1.0 | 2.0 | 3.0 | 4.0 |
|---|---|---|---|---|---|---|---|---|
| Number of Features | SIFT | 28 | 53 | 58 | 64 | 170 | 247 | 233 |
| | SURF | 3 | 12 | 30 | 41 | 39 | 41 | 44 |
| Number of Matches | SIFT | 24 | 45 | 54 | 64 | 53 | 47 | 51 |
| | SURF | 2 | 9 | 23 | 41 | 20 | 21 | 16 |
| Performance Ratio | SIFT | 0.85 | .85 | 0.89 | 1.0 | 0.32 | .20 | 0.22 |
| | SURF | 0.67 | 0.75 | 0.75 | 1.0 | 0.51 | 0.51 | .30 |

**Table 6.2.** Experimental results on the IIT Delhi database

| Method | Number of Images | Matched | Unmatched | Time | Recognition Rate |
|---|---|---|---|---|---|
| SIFT | 125 | 121 | 4 | 0.21 | 96.8 p.c. |
| SURF | 125 | 118 | 7 | 0.183 | 94.4 p.c. |

Table 2 shows an overview of how the two detectors work in real-life conditions where some images are not matched due to illumination changes as those images

were mostly taken at night and at different angles. Thus, the descriptors fail to find enough feature keypoints for matching. The overall recognition rates of the SIFT and SURF algorithms on the IIT Delhi database are 96.8p.c. and 94.4p.c., respectively. As a compar- ison, we also implemented other methods for ear recognition. The template matching technique yields a recognition rate of 93p.c. for [24], and 92.6p.c. for [23], whereas the recognition rate by the contour extraction technique [25] is 85p.c. It is evident that the proposed technique yields a higher recognition rate.

## 6.2 Results of the Deep Approach

Synthesis of the second redesign of the scheduler module targeting a Xilinx [**?**] Virtex-II Pro 30 yields the following FPGA resource statistics: 1,034 out of 13,696 slices, 522 out of 27,392 slice flip-flops, 1,900 out of 27,392 4-input LUTs, and 2 out of 136 BRAMs. The module has a maximum operating frequency of 143.8 MHz, which easily meets our goal of a 100 MHz clock frequency.

## 6.3 Comparison of the Approaches

Synthesis of the third redesign of the scheduler module targeting a Xilinx [**?**] Virtex-II Pro 30 yields the following FPGA resource statistics: 1,455 out of 13,696 slices, 973 out of 27,392 slice flip-flops, 2,425 out of 27,392 4-input LUTs, and 3 out of 136 BRAMs. The module has a maximum operating frequency of 119.6 MHz, which easily meets our goal of a 100 MHz clock frequency.

# Chapter 7

# Conclusion

In this paper, we have studied two scale and rotation invariant feature detecters and their application to ear recognition. Although both the SIFT and the SURF are invariant under scale and rotation changes, their performance decreases under certain conditions. The SIFT detector is more stable than the SURF detector under rotation changes. It is also found that the SIFT algorithm performs better for image decreasing, in contrast, the SURF algorithm performs better for image increasing. Experimental evaluations have demonstrated the effectiveness of the proposed techniques in ear recognition. In future study, we will further investigate how to increase the performance and reliability of the proposed approach.

## Acknowledgment

# Bibliography

[1] A. Pflug and C. Busch, "Ear biometrics: a survey of detection, feature extraction and recognition methods," *Biometrics, IET*, vol. 1, no. 2, pp. 114–129, 2012.

[2] A. Abaza, A. Ross, C. Hebert, M. A. F. Harrison, and M. S. Nixon, "A survey on ear biometrics," *ACM computing surveys (CSUR)*, vol. 45, no. 2, p. 22, 2013.

[3] A. Tariq and M. U. Akram, "Personal identification using ear recognition," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 10, no. 2, pp. 321–326, 2012.

[4] C. Harris and M. Stephens, "A combined corner and edge detector.," in *Alvey vision conference*, vol. 15, p. 50, Citeseer, 1988.

[5] S. Sarkar, J. Liu, and G. Wang, "Biometric analysis of human ear matching using scale and rotation invariant feature detectors," in *Image Analysis and Recognition - 12th International Conference, ICIAR2015, Niagara Falls, ON, Canada, July 22-24, 2015, Proceedings*, pp. 186–193, 2015.