Biometric Analysis of Ear Recognition using Shallow and Deep Techniques

Soumyajit Sarkar

March 24, 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-

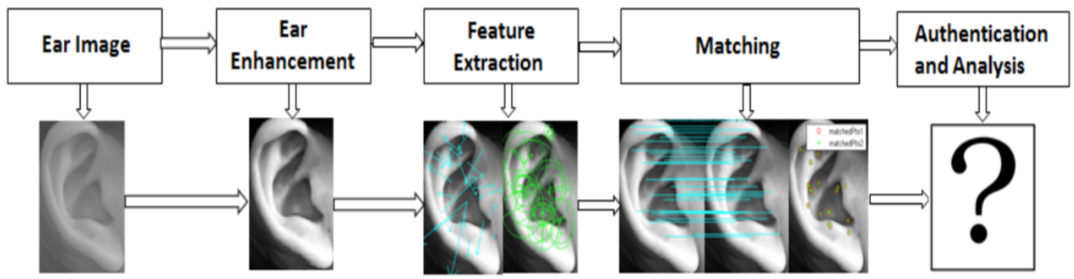

Figure 1.1. The pipeline of the proposed Ear Recognition System

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

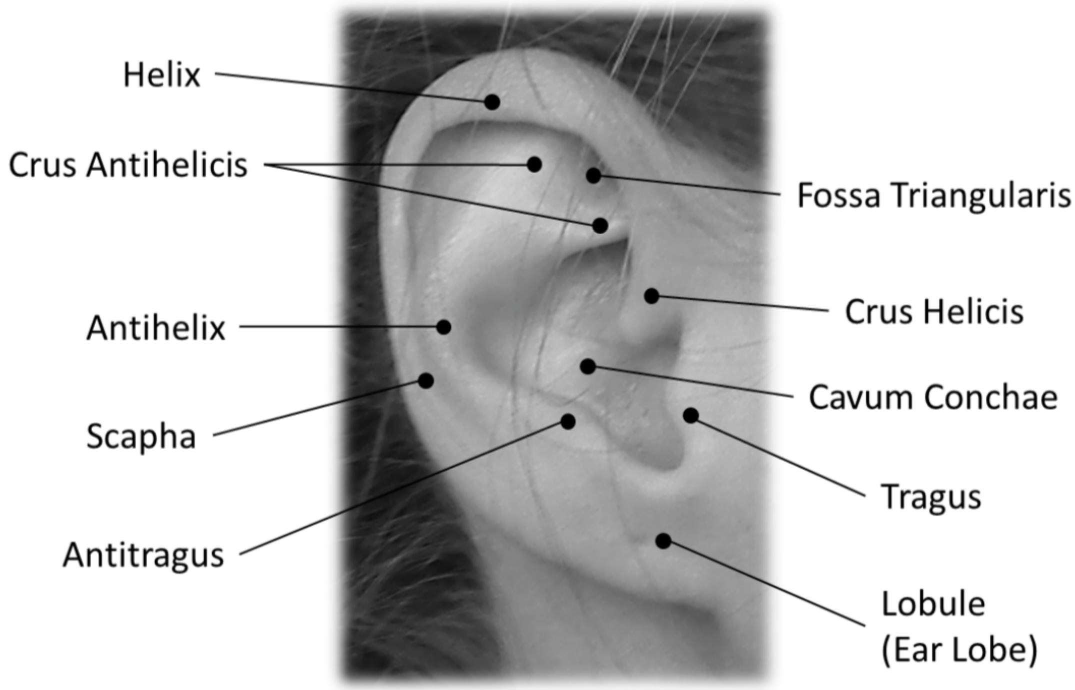

Figure 2.1. Characteristics of Human Ear [1]

thereby discovering that ears are unique even among genetically identical persons [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

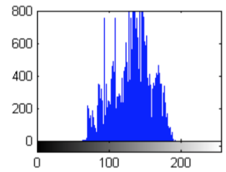

Figure 5.2. Histogram of Original Image

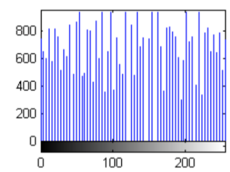

Figure 5.3. Histogram of Enhanced Image

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

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

The main goal of the second redesign of the scheduler module is to further reduce the amount of overhead and jitter involved in thread scheduling

5.3 Training Model

One of the main goals of The new hardware components used to schedule both SW and HW threads were fully integrated into the existing scheduler module. Results from simulation and synthesis tests to verify scheduler correct scheduler functionality of the new hybrid features along with O(1) ready-to-run queue structure can be found in chapter 6.

5.4 Deep Learning Approach

5.5 Convolution Neural Network

5.6 Our Deep Network and Model

Chapter 6

Implementation Results

The modules from each of the scheduler redesigns have undergone a series of tests in both simulated and synthesized forms. These tests help show the performance differences between the different ready-to-run queue structures in terms of scheduling overhead and jitter as well as the effects of the hybridization of the scheduler on the entire HybridThreads operating system.

6.1 Results of the Traditional Approach

making a scheduling decision is hidden because it occurs during a context switch. However, if a scheduling event occurs very soon after a thread is initially context switched to, the system may have to wait for the scheduler to finish calculating the next scheduling decision resulting from the context switch. This is more likely to happen with more threads in the ready-to-run queue, due to increased execution-times of scheduling operations due to the O(n) nature of the queue itself. Although the amount of jitter is in the microsecond range, it can still be further reduced by restructuring the ready-to-run queue so that its functions are able to operate in constant amounts of time.

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 presented the scheduling operations that execute in under 50 clock cycles (500 ns) or less at the base hardware level. From the system level, the hardware scheduler module provides very fast scheduling operations with an end-to-end scheduling delay of 1.9 μs with 1.4 μs of jitter with 250 active threads running on a Xilinx [?] Virtex-II Pro FPGA. The integrated system level tests have shown that the migration of scheduling services into the fabric of the FPGA have drastically reduced the amount of system overhead and jitter related

HybridThreads project [?], which furthers the future impacts of the project even more. More detailed descriptions of the hybrid threads project can be found in [?].

Acknowledgment

The work in this article is partially sponsored by National Science Foundation EHS contract CCR-0311599. The opinions expressed are those of the authors and not necessarily those of the foundation.

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.