April 2018

Identification of Individuals from Ears in Real World Conditions

Earnest Eugene Hansley
University of South Florida, earniehansley@gmail.com

Follow this and additional works at: http://scholarcommons.usf.edu/etd
Part of the Computer Sciences Commons

Scholar Commons Citation

This Dissertation is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.
Identification of Individuals from Ears in Real World Conditions

by

Earnest Eugene Hansley

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Computer Science and Engineering
College of Engineering
University of South Florida

Major Professor: Sudeep Sarkar, Ph.D.
Sanjukta Bhanja, Ph.D.
Rangachar Kasturi, Ph.D.
Miguel A. Labrador, Ph.D.
Eric Maxwell, Ph.D.

Date of Approval:
March 6, 2018

Keywords: Biometrics, Unconstrained Ear Recognition, Handcrafted Features
Learned Features, Convolutional Neural Network

Copyright © 2018, Earnest Eugene Hansley
DEDICATION

I dedicate this work to my mother (Pearlie Mae Hansley) who always sees the best in me, to my late father (David Franklin Hansley) who taught me so much, to my wife (Nilda Hansley) who has cheered me on, stood by me during this journey and pushed me gently, to my daughter (Kaila Hansley) who is my heart and is special in so many ways, to my son (Paul Hansley) who I occasionally leaned on for strength during this journey, to my namesake (Earnest Hansley, Jr) who helped me with innumerable Powerpoint presentations, to my mentor (Phillip Merrell) who made it possible for me to enter the Ph.D. program, and to my brother (David Franklin Hansley, Jr.) who gave me the gift of life.
I am thankful for Dr. Sudeep Sarkar who inherited me as a Ph.D. student and has always treated me like I were a student he chose while guiding me through a rigorous world-class Computer Science degree program and never giving up on me. I am thankful for Dr. Kevin Bowyer, who accepted me as a student, guided me while he was at USF and continued to provide academic guidance for over a decade as needed after he departed. I am thankful for Dr. Lawrence Morehouse and my Ph.D. Committee Members for honoring me by serving on my committee. I am thankful to Bernard Batson for his counsel and for connecting me to resources. I am thankful to Dr. Mauricio Pamplona Segundo for his guidance and co-authoring a paper with me. I am thankful for my initial office mates (Dr. Padmanabhan Soundararajan, Dr. Mark Powell, Dr. Isidro Robledo Vega, Dr. Jaesik Min, Dr. Kyong Chang, and Dr. Zongy Liu) for helping me get through the first couple years and sharing their insights. I am thankful to Dr. Ravi Subramaniam, Dr. Fillipe Souza, Dr. Rodrigo Minetto, Sathya, Gilbert, and Daniel for helping me race to the finish line. I am also thankful for the CSE front office support team that was there when I entered the Ph.D. program: Dee Allen, Nina Goode, and Judy. They were always courteous and helpful. And, Catherine Burton was helpful then and remains helpful ensuring I am aware of, adhere to, and meet the graduate deadlines. Plus, I am thankful to the current CSE Front Office team (Laura Smurr, Lashanda Lightbourne, and Gabriela Franco Munguia) who have always helped me without complaint even if I asked them to register me at the very last minute or requested some other academic administrative assistance. I am indebted to so many. Thank you.
# TABLE OF CONTENTS

LIST OF TABLES ........................................................................................................ iii

LIST OF FIGURES ..................................................................................................... iv

ABSTRACT ................................................................................................................. v

CHAPTER 1: INTRODUCTION ......................................................................................... 1
  1.1 Biometric Recognition ....................................................................................... 1
  1.2 Types of Biometric Recognition ....................................................................... 2
    1.2.1 Fingerprint Recognition ........................................................................... 2
    1.2.2 Iris Recognition ....................................................................................... 3
    1.2.3 Face Recognition ..................................................................................... 3
    1.2.4 Gait Recognition ..................................................................................... 4
    1.2.5 Retinal Scan Recognition ......................................................................... 4
    1.2.6 Ear Recognition ....................................................................................... 5

CHAPTER 2: RELATED WORK ....................................................................................... 6
  2.1 Ear Recognition in Computer Vision ............................................................ 6
  2.2 Principal Component Analysis for Ear Recognition .................................. 6
  2.3 Convolutional Neural Network for Ear Recognition ............................... 7

CHAPTER 3: OVERVIEW OF OUR APPROACH .......................................................... 9

CHAPTER 4: LANDMARK AND NORMALIZATION ...................................................... 12
  4.1 Landmark Detection ...................................................................................... 12
  4.2 Geometric Normalization .............................................................................. 14

CHAPTER 5: DESCRIPTION AND MATCHING .......................................................... 16
  5.1 Overview of Description and Matching ....................................................... 16
  5.2 Holistic Features ......................................................................................... 16
  5.3 Handcrafted Features .................................................................................. 17
  5.4 Learned Features ......................................................................................... 17

CHAPTER 6: FUSION ................................................................................................... 19

CHAPTER 7: DATABASE ............................................................................................. 22
  7.1 Indian Institute of Technology Delhi Ear Database .................................... 22
  7.2 West Pomeranian University of Technology Ear Database .................... 23
7.3 Annotated Web Ears Database ................................................................. 24
7.4 In-the-wild Ear Database ......................................................................... 25
7.5 Unconstrained Ear Recognition Challenge Database ............................. 25
7.6 Discussion .............................................................................................. 26

CHAPTER 8: RESULTS .................................................................................. 27
  8.1 Landmark Detection Results ................................................................. 27
  8.2 Normalization Results ......................................................................... 28
  8.3 CNN Description Results ..................................................................... 29
  8.4 Fusion Results ...................................................................................... 31
  8.5 Discussion ............................................................................................. 35

CHAPTER 9: CONCLUSION ........................................................................ 39

REFERENCES ............................................................................................ 41

APPENDIX A: COPYRIGHT CLEARANCE FORMS ................................. 45

ABOUT THE AUTHOR .................................................................................... END PAGE
LIST OF TABLES

Table 1 Network architecture for landmark detection in ear images.................................13

Table 2 Network architecture for feature extraction in ear images ..................................18

Table 3 Rank-1 and EER results for the AWE database as reported by Emersic et al. [12] and our reproduction of their experiments using images with normalization and images without (raw) normalization.........................................................29

Table 4 EER results for all testing sets using descriptors learned from each database or from all databases combined........................................................................................................31

Table 5 Individual and fusion results for all descriptors in Chapter 5 using the AWE database ..................................................................................................................................................32

Table 6 Individual and fusion results for CNN, HOG, POEM and LBP in the overall performance evaluation through the UERC protocol, as well as the top scoring participants of the UERC challenge. .................................................................................................34

Table 7 Individual and fusion results for CNN, HOG and POEM in the scalability evaluation through the UERC protocol, as well as the top scoring participants of the UERC challenge..........................................................................................................................35
LIST OF FIGURES

Figure 1 Example of a challenging task for ear recognition in an unconstrained setting ..........10

Figure 2 Diagram of our ear recognition framework ............................................................10

Figure 3 Normalization results (a)-(b) with and (c)-(d) without the same sampling rate in both axis for two ear images of the same person with pose variations...............................14

Figure 4 Example of images from (a) IIT, (b) WPUTE, (c) AWE and (d) ITWE databases ......23

Figure 5 Cumulative error distribution for landmark detection using the proposed approach and Zhou and Zaferiou’s approach [47] .................................................................28

Figure 6 CMC curves for all participants of the UERC challenge plus our best fusion results obtained by combining CNN and HOG considering the (a) overall performance evaluation and (b) scalability evaluation protocols ...........................................37
ABSTRACT

A number of researchers have shown that ear recognition is a viable alternative to more common biometrics such as fingerprint, face and iris because the ear is relatively stable over time, the ear is non-invasive to capture, the ear is expressionless, and both the ear’s geometry and shape have significant variation among individuals. Researchers have used different approaches to enhance ear recognition. Some researchers have improved upon existing algorithms, some have developed algorithms from scratch to assist with recognizing individuals by ears, and some researchers have taken algorithms tried and tested for another purpose, i.e., face recognition, and applied them to ear recognition. These approaches have resulted in a number of state-of-the-art effective methods to identify individuals by ears. However, most ear recognition research has been done using ear images that were captured in an ideal setting: ear images have near perfect lighting for image quality, ears are in the same position for each subject, and ears are without earrings, hair occlusions, or anything else that could block viewing of the entire ear.

In order for ear recognition to be practical, current approaches must be improved. Ear recognition must move beyond ideal settings and demonstrate effectiveness in an unconstrained environment reflective of real world conditions. Ear recognition approaches must be scalable to handle large groups of people. And, ear recognition should demonstrate effectiveness across a diverse population.
This dissertation advances ear recognition from ideal settings to real world settings. We devised an ear recognition framework that outperformed state-of-the-art recognition approaches using the most challenging sets of publicly available ear images and the most voluminous set of unconstrained ear images that we are aware of. We developed a Convolutional Neural Network-based solution for ear normalization and description, we designed a two-stage landmark detector, and we fused learned and handcrafted descriptors. Using our framework, we identified some individuals that are wearing earrings and that have other occlusions, such as hair. The results suggest that our framework can be a gateway for identification of individuals in real world conditions.
CHAPTER 1: INTRODUCTION

1.1 Biometric Recognition

The term "biometric" infers some sort of life measurement, as the compound word is forged from two different words that refer to life and measurement. Biometric recognition is the automatic recognition of individuals based on their physiological and/or behavioral characteristics [18]. By measuring some physiological characteristic or some part of the body, we can determine the identity of individuals and verify the identity of individuals too. At the turn of the century, it was rare to hear the term biometric recognition outside of academia. Now, with the advances in computer technology and smart devices that have embedded biometric recognition software applications on them, biometric recognition is becoming quite popular. It is not uncommon to see a popular television show like Naval Criminal Investigative Service (NCIS) or NCIS New Orleans use biometric references like facial recognition or finger print recognition. But, biometric recognition has been used for centuries.

There is evidence that fingerprints were used as early as 500 B.C. [34]. However, the first systematic capture of handprints was done in 1858 by Sir William Hershel [22]. He was working for the Civil Service of India and used the handprints on the back of contracts to

---

1 Materials from this chapter were published in “Employing Fusion of Learned and Handcrafted Features for Unconstrained Ear Recognition”, by E. Hansley, M. Pamplona Segundo, & S. Sarkar, 2018, IET Biometrics, http://dx.doi.org/10.1049/iet-bmt.2017.0210 The materials are reproduced by permission of the Institution of Engineering & Technology. Permission is included in Appendix A.
distinguish employees from others. He wanted to ensure that the people receiving pay were contract employees.

The first documented case study involving biometrics was done by a French Police Officer named Alphonse Bertillon in 1885 [3]. Bertillon used body measurements to identify criminals. The method was called anthropometrics. The combined measurements that he used included the ear [1]. In 1886, Richard Imhofer, a Czechoslovakian doctor, was able to distinguish 500 ears using only four features [1]. In 1964, Iannarelli published a book about his system of ear identification [16]. It was based on his research whereby he examined over 10000 ears during a 38-year period. These earlier biometric works and ear recognition research have established a good basis point for subsequent research.

There are multiple options for the choice of biometric to use for recognition. Some of the popular biometric methods include: fingerprint, iris, face, gait, and ear. Each method has some characteristics that algorithms exploit information from in order to recognize and identify individuals. The choice of biometric selected can be based upon the expected usage and the associated advantages. The choice can also be made based on risks associated with weaknesses of biometric recognition systems under consideration. Therefore, it is useful to know the strengths and weaknesses of biometric methods.

1.2 Types of Biometric Recognition

1.2.1 Fingerprint Recognition

One of the oldest, most used, and most reliable biometric methods is the fingerprint. Fingerprints have one of the highest reliabilities [30]. A fingerprint has ridges, patterns, and minutiae that are distinct. And, fingerprints are difficult to alter. Because of the reliability, the fingerprint is used to identify suspects of criminal activity and to prosecute criminals. It is those
distinct qualities that enable identification of individuals. Even identical twins have different fingerprints [18]. Because of the popularity of fingerprints, many smart devices have embedded fingerprint capabilities. But, a small group of the population may not be good fingerprinting subjects, and some skilled labor workers that are subject to occasional cuts on their hands may not be ideal subjects [18].

1.2.2 Iris Recognition

The iris is becoming a popular choice of biometric. The iris is a circular structure in the eye that surrounds the pupil and its patterns are used to distinguish individuals. Although it is small and sometimes difficult to image, the iris pattern variability among different individuals is enormous. The iris has unique patterns that are complex and can be recognized from a distance. Therefore, it is non-invasive to capture. Some iris recognition systems have been employed by various governments. But, acquisition of the iris for identification can be challenging because of its small size, especially if someone is moving.

1.2.3 Face Recognition

A very popular choice of biometric is face recognition. The face’s eyes, nose and mouth are some of the attributes on the face that can be used to extract features or compute distances among them for identification. Face recognition is popular because it is non-invasive and faces can be captured at a distance. The Face Recognition Technology (FERET) program was implemented to assess the state-of-the-art in face recognition, to identify future areas of research, and to test algorithms performance on a large scale [32]. The program tremendously advanced the field of face recognition research. And since the FERET program began, face recognition has been used in several settings. One such usage that received a lot of notoriety was the Super Bowl that Tampa, FL hosted in 2001. Additionally, Apple has implemented facial recognition on
its iPhone X smartphone. A lot of different face recognition algorithms have been tested over the years. But pose angle, pose variations, different expressions, and different lighting remain challenges that impact face recognition. And, it has been shown that face images taken a year apart can degrade performance [14]. Faces, do however, remain one of the most popular biometrics. And, because of that popularity, face recognition is likely to become available on more smart devices.

1.2.4 Gait Recognition

Gait recognition is recognizing people by the way they walk or run. Researchers posed the question of whether humans can be recognized by their gait in 1977 [7]. In 1999 researchers showed that humans can identify people by their gait signature [33]. And, computer vision researchers have shown that simple approaches using silhouettes can be used for gait recognition [28]. The benefit of gait recognition is that it can be used for surveillance and it can be used from a distance when another biometric may fail or not be as useful. For example, if someone is walking towards a highly secure building or towards an airport, contact methods such as fingerprints are not options. Also, face recognition requires quality images for effective recognition and the face may not always be visible if a subject is walking. Therefore, gait can be an attractive option. But, because gait recognition is done by capturing video sequence footage, it is computationally expensive [18].

1.2.5 Retinal Scan Recognition

Retina scans are another effective biometric means to identify someone. The retina is an inner coat of the eye with rich vascular structure. The unique structure of the retina enables identification [6] and, it is considered to be the most secure biometric [18]. It is difficult to change or replicate the retina’s vascular structure. Governmental agencies such as the Central
Intelligence Agency, the Federal Bureau of Investigation, and the National Aeronautical Space Agency use retina scanning. Retina scans can become more popular with commercial usage. To capture a retina scan does require some participation from the subject. That requirement has possibly slowed the commercial adaption for usage. But, a September 2017 report suggested that using the retina scan could be the future of biometrics and banking [26]. People may not mind interacting with a scanner to safeguard their assets from theft.

1.2.6 Ear Recognition

Ear recognition is another common biometric, and it is a good alternative to the biometric options discussed in the preceding paragraphs. Different parts of ear images are marked for ear recognition algorithms to detect features and compare them to different ear images. Ear images can passively be captured and the stability of the ear affords a more consistent biometric identification method. Additionally, researchers have shown that ear recognition can be used to distinguish identical twins [29]. This is significant in that recognition of identical twins is an unsolved problem for face recognition. Because ear images can be captured from a distance and because they have some advantages over other biometric methods, ear recognition could become the biometric of choice for governmental agencies, and advances in ear recognition research could accelerate commercial adaptations.
CHAPTER 2: RELATED WORK

2.1 Ear Recognition in Computer Vision

The first image-based work for ear recognition was done by Burge and Burger [4]. They made a case for ear biometrics being viable and developed a Machine Vision system as a proof of concept of the viability of ear biometrics for passive identification. Their system consists of three parts: localization and segmentation, feature extraction, and feature comparison. In that their system is a proof of concept [4], they do not provide recognition results. But, their foundational approach has resulted in their work being amongst the most cited papers in Computer Vision ear recognition research. Since Burge and Burger’s work, there has been a number of different approaches for ear recognition.

2.2 Principal Component Analysis for Ear Recognition

Victor, et al. published the first work on using principal component analysis (PCA) for ear recognition in 2002[41]. Principal component analysis is a statistical analysis method that is used to detect patterns and variations in a dataset. It is also a dimensionality reduction technique, which makes it an attractive choice for ear recognition even with the advances in computer technology.

---

2 Materials from this chapter were published in “Employing Fusion of Learned and Handcrafted Features for Unconstrained Ear Recognition”, by E. Hansley, M. Pamplona Segundo, & S. Sarkar, 2018, IET Biometrics, http://dx.doi.org/10.1049/iet-bmt.2017.0210 The materials are reproduced by permission of the Institution of Engineering & Technology. Permission is included in Appendix A.
Victor, et al. modeled the work done by Turk and Pentland who applied PCA to face recognition [40]. Turk and Pentland projects face images onto a feature space and refer to it as face space. They use PCA to find the eigenvectors corresponding to the face images. Because the eigenvectors are face-like in appearance, Turk and Pentland refer to them as eigenfaces. Victor et al. refer to the ears analyzed using PCA as eigen-ears in their work.

Victor et al.’s approach consisted of three steps: 1) Preprocessing, 2), Normalization, and 3) Identification. In the preprocessing step, images are cropped, the ear is centered, and landmarks are annotated. In the normalization step, they scale images and apply a mask to get rid of hair and anything else in the background of the image that may not be helpful. And, the image is normalized for illumination. They ensure that the training set has clean images without earrings. Their identification step consists of training and testing. Victor et al. tested PCA with their method using 76 images in the gallery set and 73 images in the probe set. This is a very small dataset compared to the size of the datasets we tested for this work. But, at the time of their research that was adequate to show that PCA can be used for ear recognition.

2.3 Convolutional Neural Network for Ear Recognition

A relatively new approach for ear recognition involves using a convolutional neural network (CNN). This approach became very popular during an image net competition [24] because the winners of the competition achieved the best error rates. Their results were significantly better than all competitors. Their Top-1 performance and Top-5 performance are 8 to 10 percent better than their closest competitors. What is equally significant is that their results were obtained using 1.3 million images with 1000 different classes.

A CNN has neurons applied across a space that enables detecting and learning features such as lines and curves. Those learned features are used to compare different ear images in
order to determine the identity of an individual. A CNN applied to ear image is different from the state-of-the-art algorithms that have traditionally been used. Some traditional ear recognition approaches search for features such as lines and curves. Other approaches collect some sort of statistics. What is different about a CNN is that it does not search for lines, curves, or statistical patterns. What CNN is searching for is not predetermined. A CNN extracts information from the data that is representative of features such as curves that could represent an animal, an ear, or something else. Then CNN learns the features and those learned features are used to help delineate whatever is in the image, such as an ear or an animal if there is an animal or ear that is present in the image.

In 2016, Tian, et al. applied CNN to ear recognition [39]. They designed a CNN with three convolutional layers, a fully-connected layer and a soft-max classifier. They used a dataset of 79 subjects. The images from the subjects have some pose angle variations, but they are a constrained set of images. There are no earrings, headsets, or similar occlusions. They did test partial occlusions that they added to the corners of the ear images. But, self-manufactured occlusions are not the same as testing with unconstrained ear images. In 2017, Emersic, et al. reported that Istanbul Technical University (ITU) used a deep learning approach based on Visual Geometry Group 16-layer CNN model (VGG-16) [36] to identify individuals from ears [11]. They used a dataset of 2304 images from 166 subjects and submitted their finding as part of the Unconstrained Ear Recognition Challenge(UERC) [11].
CHAPTER 3: OVERVIEW OF OUR APPROACH

A number of researchers have shown that ear recognition is a viable alternative to more common biometrics such as fingerprint, face and iris [5, 25, 45]. The ear is stable over time, is less invasive to capture, and does not require as much control during image acquisition as other biometrics. And, it is reasonable to assert that there are fewer privacy issues for the ear than there are for the face.

Traditionally, ear recognition research has been performed on ear images that were captured in an ideal setting. In an ideal setting, the ears are all captured in the same position, with identical lighting, and identical resolution. With the advances in computer vision and pattern recognition techniques, research of ear recognition is shifting to a more challenging scenario whereby ear images are acquired from real world (unconstrained) settings [12, 11]. It is more difficult to recognize ears in the wild. In this paper we use “ears in the wild” and “unconstrained ears” interchangeably. Figure 1 illustrates the difficulty of recognizing individuals using ears in the wild.

This example mostly illustrates the problem of pose variation, but many other factors may affect the recognition performance: different acquisition devices, low resolution, illumination variations, occlusions caused by hair and head accessories, earrings, headsets and so

---

3 Materials from this chapter were published in “Employing Fusion of Learned and Handcrafted Features for Unconstrained Ear Recognition”, by E. Hansley, M. Pamplona Segundo, & S. Sarkar, 2018, IET Biometrics, http://dx.doi.org/10.1049/iet-bmt.2017.0210 The materials are reproduced by permission of the Institution of Engineering & Technology. Permission is included in Appendix A.
on. To overcome these recognition challenges, ear recognition has to achieve good results for non-cooperative subjects. This will make ear biometric recognition very useful for practical purposes, like video surveillance and continuous authentication.

![Figure 1: Example of a challenging task for ear recognition in an unconstrained setting.](image1)

In order to carry out the task of recognizing humans through their ears, a common sequence of steps is usually followed (see Figure 2).

![Figure 2: Diagram of our ear recognition framework.](image2)

The first step is to capture a digital biometric sample using an appropriate sensor. For all of our experiments we used images from five publicly available databases. The second step is the localization step. The algorithm locates the biometric information and separates it from irrelevant parts of the acquired sample. The images we used were either already cropped or the ground truth location of the ears was provided; thus we do not perform the localization step.
However, it is possible to find successful approaches that perform ear detection in the wild in the literature [46, 10]. A very critical step is the normalization step. In this step, we are reshaping the input sample to a standard format to reduce unwanted variations. We used a landmark detector based on Convolutional Neural Networks (CNN) [23] to locate a set of 55 landmarks, which were then employed to translate, rotate and scale the input image to a standard configuration.

We used the landmarked points to aid with feature description. In the feature description step, discriminant features were extracted from a normalized sample. This usually reduces its dimensionality. We used a state-of-the-art CNN architecture that was designed for face recognition, to do the task of ear recognition in the wild, and we used different traditional ear description approaches, too.

The final step is the Recognition step. In the recognition step, we compare the results from the descriptors and determine if the same person matches or not. All images are compared to each other using the descriptor’s distance metric. All scores are normalized using min-max normalization [17]; then score level fusion [21] is used to combine results of different descriptors and inform the decision. After implementing our framework and evaluating the results, the following contributions were achieved. We designed and developed a two-stage CNN-based landmark detector that produces accurate results even in the presence of variations not seen in the training data. We used our detector to automatically normalize images and instantly observed a boost in the recognition rate. We devised a CNN-based ear descriptor based on a state-of-the-art face recognition architecture that outperformed similar state-of-the-art ear recognition works that are based on CNNs. We demonstrated that handcrafted and learned descriptors are complementary, and thus fusing them results in a considerable increase in performance.
CHAPTER 4: LANDMARK AND NORMALIZATION

4.1 Landmark Detection

Even with the recent emergence of deep learning methods for biometric recognition in uncontrolled scenarios, normalization is still necessary to achieve better results. For instance, a landmark-based orientation and scale normalization is a standard procedure in face recognition state-of-the-art works [42, 43]. With this in mind, we pursued a similar path for the ear recognition problem by investigating the use of CNNs for the landmark detection task. To this end, we used images and annotations provided in Collection A of the ITWE database for CNN training and accuracy evaluation. As only 500 images were available for training, we performed different data augmentation operations in order to avoid overfitting and increase the network generalization power. For each training image, we used principal component analysis (PCA) [2] on the 2D coordinates of the annotated landmarks to obtain the upright orientation of the ear (i.e. we assumed it corresponds to the direction of the first component). Then, we created multiple images by rotating the upright ear from $-45^\circ$ to $45^\circ$ with steps of $3^\circ$. Each ear was also transformed by a random scale change of up to 20% of the original ear size in both axes, as well as a random translation of up to 20% of the original ear size in each axis. After applying all these...

---

4 Materials from this chapter were published in “Employing Fusion of Learned and Handcrafted Features for Unconstrained Ear Recognition”, by E. Hansley, M. Pamplona Segundo, & S. Sarkar, 2018, IET Biometrics, http://dx.doi.org/10.1049/iet-bmt.2017.0210 The materials are reproduced by permission of the Institution of Engineering & Technology. Permission is included in Appendix A.
modifications, images were rescaled to 96×96 pixels, and we ended up with 15500 training images.

The architecture of our network is based on a common design nowadays, even for landmark detection [37], which consists of alternating between convolution and max pooling layers in the beginning, and then following with a sequence of fully-connected layers. We used rectified linear units in convolution and fully-connected layers to train models from scratch. We also added dropouts after all max pooling and the first fully-connected layers to avoid overfitting the training data. A complete description of our architecture is presented in Table 1. It was implemented using TensorFlow, and the optimization to minimize the mean squared error in the output was carried out by the Nesterov’s Momentum algorithm [38] for 2000 epochs.

<table>
<thead>
<tr>
<th>#</th>
<th>Type</th>
<th>Input</th>
<th>Filter</th>
<th>Stride</th>
<th>Drop</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Conv/Relu</td>
<td>96×96×1</td>
<td>3×3×1×32</td>
<td>1</td>
<td></td>
<td>96×96×32</td>
</tr>
<tr>
<td>2</td>
<td>MaxPool</td>
<td>96×96×32</td>
<td>2×2</td>
<td>2</td>
<td>10%</td>
<td>48×48×32</td>
</tr>
<tr>
<td>3</td>
<td>Conv/Relu</td>
<td>48×48×32</td>
<td>2×2×32×64</td>
<td>1</td>
<td></td>
<td>48×48×64</td>
</tr>
<tr>
<td>4</td>
<td>MaxPool</td>
<td>48×48×64</td>
<td>2×2</td>
<td>2</td>
<td>20%</td>
<td>24×24×64</td>
</tr>
<tr>
<td>5</td>
<td>Conv/Relu</td>
<td>24×24×64</td>
<td>2×2×64×128</td>
<td>1</td>
<td></td>
<td>24×24×128</td>
</tr>
<tr>
<td>6</td>
<td>MaxPool</td>
<td>24×24×128</td>
<td>2×2</td>
<td>2</td>
<td>30%</td>
<td>12×12×128</td>
</tr>
<tr>
<td></td>
<td>Flattening</td>
<td>12×12×128</td>
<td></td>
<td></td>
<td></td>
<td>18432</td>
</tr>
<tr>
<td>7</td>
<td>Fc/Relu</td>
<td>18432</td>
<td></td>
<td></td>
<td>50%</td>
<td>1000</td>
</tr>
<tr>
<td>8</td>
<td>Fc/Relu</td>
<td>1000</td>
<td></td>
<td></td>
<td></td>
<td>1000</td>
</tr>
<tr>
<td>9</td>
<td>Fc</td>
<td>1000</td>
<td></td>
<td></td>
<td></td>
<td>110</td>
</tr>
</tbody>
</table>

Table 1: Network architecture for landmark detection in ear images.

Although this network achieved an admirable accuracy considering the level of variations in unconstrained scenarios, we evaluated a two-stage solution, whereby the first network was used to create an easier landmark detection scenario by reducing scale and translation variations, and the second network is used to generate the 2D coordinates for landmarks. We used the coordinates obtained by the network described above to refine the center and the orientation of
an ear using PCA, and then fed the rectified image to a second network that was trained in a more controlled scenario. The second network has the same architecture and optimization procedure of the first one, the only difference is the training data, which uses less variation in the augmentation process. Rotations are performed from $-15^\circ$ to $15^\circ$ with steps of $1^\circ$, and random scale and translation changes are limited to up to 10% of the original ear size.

4.2 Geometric Normalization

After landmark detection, we performed a geometric normalization of the ears by applying PCA on the retrieved landmarks. We used the first component as the orientation of the ear and the center of the oriented bounding box as the center of the ear. We then interpolated a 128 $\times$ 128 image with these parameters considering that the distance between the center of the ear and the top of the image is equal to two times the square root of the first eigenvalue in the original image. However, as ears in the wild may present significant pose variations, this also occurs in width variations that may affect the recognition performance, as shown in Figures 3a and 3b. Thus, we used different sampling rates in x and y directions in a way that the distance between the center of the ear and one side of the image is equal to two times the square root of the second eigenvalue in the original image. This way, the width and the height of the normalized ear were approximately the same, as may be seen in Figures 3c and 3d, and image variations caused by pose became less intense.

Figure 3: Normalization results (a)-(b) with and (c)-(d) without the same sampling rate in both axis for two ear images of the same person with pose variations.
The ear images we experimented with contained a variety of different sizes and different orientations. In order to make a fair comparison of ear images, they must be aligned to a common reference. We used geometric normalization to align them to a common reference. The images were rotated, scaled, and translated to make them all the same size, to place them all in the same position, and to place them equidistant from a reference position.
CHAPTER 5: DESCRIPTION AND MATCHING

5.1 Overview of Description and Matching

We evaluated three different description and matching schemes based on 1) holistic image features, 2) handcrafted features and 3) learned features. We then investigated if fusing some of them would achieve a higher accuracy.

5.2 Holistic Features

PCA was one of the first methods employed to the ear recognition problem [5], as it provides a holistic description of the sample images while reducing the dimensionality of the data. However, even the pioneer works using PCA have already reported a performance drop caused by variations in pose and illumination, and such variations are much more intense in recent uncontrolled databases. We tested a PCA implementation available in the Face Identification Evaluation System [44] as a baseline approach, and its feature vectors were matched through the Mahalanobis distance. The first 20 eigenvectors were dropped to avoid illumination and background variations, and we kept 60% of the eigenvectors in our PCA descriptor.

---

5 Materials from this chapter were published in “Employing Fusion of Learned and Handcrafted Features for Unconstrained Ear Recognition”, by E. Hansley, M. Pamplona Segundo, & S. Sarkar, 2018, IET Biometrics, http://dx.doi.org/10.1049/iet-bmt.2017.0210 The materials are reproduced by permission of the Institution of Engineering & Technology. Permission is included in Appendix A.
5.3 Handcrafted Features

As holistic features are strongly affected by different variations, specialists designed different feature extraction approaches, which are known as handcrafted features, seeking to overcome some of these problems. Emersic et al. [12, 11] released a toolbox that contains the best performing state-of-the-art handcrafted features for ear recognition: local binary patterns (LBP), binarized statistical image features (BSIF), local phase quantization features (LPQ), rotation invariant LPQs (RILPQ), patterns of oriented edge magnitudes (POEM), HOG, dense scale-invariant feature transform (DSIFT) and Gabor wavelets. All descriptors were extracted using the default parameters of the toolbox. For matching, as in Emersic et al.’s work [12], we compared histogram-based descriptors using the chi-square distance and Gabor descriptors using the cosine distance.

5.4 Learned Features

Considering that the performance of handcrafted descriptors degrades when using uncontrolled ear images [12], we explored CNNs so that we could improve performance and so that we could learn more about the images, as well as how to describe them in a more discriminatory and concise way. The CNN that we implemented is a state-of-the-art CNN architecture employed for face recognition in the wild [42]. We trained it from scratch for the ear recognition in the wild problem. We present a complete description of the chosen CNN architecture as well as specific layer configurations in Table 2. This network was implemented using TensorFlow, and the Adam optimization algorithm [20] was used to minimize the weighted sum of softmax and center losses. As in Wen et al.’s work [42], we set the center loss weight to 0.003.
Table 2: Network architecture for feature extraction in ear images.

<table>
<thead>
<tr>
<th>#</th>
<th>Type</th>
<th>Input</th>
<th>Filter</th>
<th>Siphle</th>
<th>Drop</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Conv/Relu</td>
<td>128 x 128 x 1</td>
<td>3 x 3 x 1 x 128</td>
<td>1</td>
<td></td>
<td>128 x 128 x 128</td>
</tr>
<tr>
<td>2</td>
<td>Conv/Relu</td>
<td>128 x 128 x 128</td>
<td>3 x 3 x 128 x 128</td>
<td>1</td>
<td></td>
<td>128 x 128 x 128</td>
</tr>
<tr>
<td>3</td>
<td>MaxPool</td>
<td>128 x 128 x 128</td>
<td>2 x 2</td>
<td>2</td>
<td>10%</td>
<td>64 x 64 x 128</td>
</tr>
<tr>
<td>4</td>
<td>Conv/Relu</td>
<td>64 x 64 x 128</td>
<td>3 x 3 x 128 x 128</td>
<td>1</td>
<td></td>
<td>64 x 64 x 128</td>
</tr>
<tr>
<td>5</td>
<td>MaxPool</td>
<td>64 x 64 x 128</td>
<td>2 x 2</td>
<td>2</td>
<td>20%</td>
<td>32 x 32 x 128</td>
</tr>
<tr>
<td>6</td>
<td>Conv/Relu</td>
<td>32 x 32 x 128</td>
<td>3 x 3 x 128 x 256</td>
<td>1</td>
<td></td>
<td>32 x 32 x 256</td>
</tr>
<tr>
<td>7</td>
<td>MaxPool</td>
<td>32 x 32 x 256</td>
<td>2 x 2</td>
<td>2</td>
<td>30%</td>
<td>16 x 16 x 256</td>
</tr>
<tr>
<td>8</td>
<td>Conv/Relu</td>
<td>16 x 16 x 256</td>
<td>3 x 3 x 256 x 256</td>
<td>1</td>
<td></td>
<td>16 x 16 x 256</td>
</tr>
<tr>
<td>9</td>
<td>MaxPool</td>
<td>16 x 16 x 256</td>
<td>2 x 2</td>
<td>2</td>
<td></td>
<td>8 x 8 x 256</td>
</tr>
<tr>
<td>10</td>
<td>Conv/Relu</td>
<td>8 x 8 x 256</td>
<td>3 x 3 x 256 x 256</td>
<td>1</td>
<td></td>
<td>8 x 8 x 256</td>
</tr>
<tr>
<td></td>
<td>Flatten 8</td>
<td>8 x 8 x 256</td>
<td>1</td>
<td></td>
<td></td>
<td>16384</td>
</tr>
<tr>
<td></td>
<td>Flatten 9</td>
<td>8 x 8 x 256</td>
<td>1</td>
<td></td>
<td></td>
<td>16384</td>
</tr>
<tr>
<td></td>
<td>Concat 8 &amp; 9</td>
<td>16384/16384</td>
<td>1</td>
<td></td>
<td></td>
<td>32768</td>
</tr>
<tr>
<td>11</td>
<td>FC</td>
<td>32768</td>
<td>1</td>
<td></td>
<td></td>
<td>512</td>
</tr>
</tbody>
</table>

This CNN outputs 512-dimensional descriptors that can be matched through the cosine distance, making the entire processing time (i.e. description and matching) comparable to that of handcrafted descriptors. For a given training set, the network optimization was performed in batches of 128 images for 1000 epochs using softmax loss only, and then the weighted sum of softmax and center losses was used until convergence was reached (i.e. no improvement after 50 epochs). As was done during landmark detection, we performed data augmentation operations to increase the number of training images by: applying random rotation between −10° and 10°, applying random crop with 85% to 100% of the original image size and by applying a random contrast change increasing or decreasing the range of pixel intensities in up to 50%.
CHAPTER 6: FUSION

Although the goal during image acquisition is to capture perfect ear images, that is difficult to do. And, ear images captured in real world conditions are far from perfect. This has led to researchers looking for ways to improve recognition with the images that they have. Researchers have concatenated different biometrics to make multimodal images, they have used multiple samples of single biometrics, and they have used several types of fusion methods to improve biometric recognition results. As the computer vision field has progressed, recent efforts to identify individuals using unconstrained ear images reflect the realization that image acquisitions will not always be under ideal circumstances. This motivated us to consider a number of different algorithms that could help identify individuals from ears in real conditions. We reviewed numerous scientific journal publications and examined advantages and disadvantages of ear recognition methods. There are different kinds of multimodal systems that address problems associated with single modality systems [19], but a multimodal system based on multiple matchers is the most adequate one for wild scenarios. The reason is that it may not always be possible to have multiple biometric traits (e.g. face and ear), multiple units of a biometric trait (e.g. thumb and index fingerprints) or multiple samples of the same biometric trait

---

6 Materials from this chapter were published in “Employing Fusion of Learned and Handcrafted Features for Unconstrained Ear Recognition”, by E. Hansley, M. Pamplona Segundo, & S. Sarkar, 2018, IET Biometrics, http://dx.doi.org/10.1049/iet-bmt.2017.0210 The materials are reproduced by permission of the Institution of Engineering & Technology. Permission is included in Appendix A.
(e.g. face in video). But, we can always apply multiple matching techniques to a single biometric sample.

There are many choices of algorithms to use for fusing single biometric samples. And, there are multiple ways to approach which algorithms to use. One way is to study the advantages and disadvantages of multiple algorithms and try to find a complementary combination whereby a disadvantage of one algorithm may be an advantage for another approach and vice-versa. Another approach is to select algorithms that have achieved good results for the biometric than one is testing. In that we were using the UERC toolkit, we had 9 state-of-the-art algorithms available to consider. We added PCA to the experiments and that gave us a total of 10 algorithms to consider. We were interested in determining if a particular fusion combination of size two could improve performance. Since we started with 10 algorithms, we had 45 pairings to consider. We were also interested in learning whether a particular fusion scheme was best. In order to fuse matchers based on the descriptors previously presented, we evaluated different fusion schemes at score level, such as sum, min, max and product rules. We used the sum rule \[21\] as it achieved the best results in our experiments. Because we are experimenting with different algorithms, the scores generated by different algorithms are dissimilar as each algorithm captures something different during feature extraction. And, dependent upon the algorithm used, a different distance metric is used to determine the distance between images. Therefore, we normalized all the scores.

In order to normalize the scores we established probe versus gallery sets of scores based on the FERET study to make our fusion simpler to implement. Then, we used min-max normalization\[17\]. Let a gallery set for a biometric (e.g. ear) be denoted by \(Ge = \{ge_1, \ldots, ge_N\}\) and the corresponding probe set be denoted by \(Pe = \{pe_1, \ldots, pe_N\}\). Let \(de_{ij} = \)
distance between ith ear probe with the jth gallery probe. Let the normalized distance of the ear be: 
\[d_{ne_{ij}} = \frac{d_{e_{ij}} - \min_j d_{e_{ij}}}{\max_j d_{e_{ij}} - \min_j d_{e_{ij}}}.\]

Sum Fusion We apply the simple sum of scores [21] by summing the matching scores of each algorithm. Let the normalized score of Algorithm A matching scores be represented by \(A_{G_{score1}} \ldots A_{G_{scoreN}}\) for a probe versus gallery matching. Let the normalized score of Algorithm B matching points be represented by \(B_{G_{score1}} \ldots B_{G_{scoreN}}\) for a probe versus gallery matching. The simple sum of scores is the summation from 0 to max of Algorithm A score 0 and Algorithm B score 0 \ldots n. The fused scores are normalized.
CHAPTER 7: DATABASE

There are many things that can affect the performance of ear recognition and some sets of ear images are easier than others. Therefore, it is a good idea for researchers to experiment with multiple image datasets when feasible. A test may include using ideal images, then progressing to unconstrained more difficult images to recognize. In this work we used five different databases to train and evaluate our ear recognition framework. We used images from the Indian Institute of Technology Delhi Ear Database (IIT), the West Pomeranian University of Technology Ear Database (WPUTE), the Annotated Web Ears database (AWE), the In-the-wild Ear Database (ITWE) and the Unconstrained Ear Recognition Challenge database (UERC). More details about each of them are given in the subsequent sections.

7.1 Indian Institute of Technology Delhi Ear Database

The IIT database [25] was released in two different formats, a raw version and a normalized version. We used the raw version for our experiments. It contains 493 images with size $272 \times 204$ from 125 different subjects. Each image shows a small region around the left ear and was collected in an indoor environment in a well-controlled acquisition setup, which makes this database a suitable benchmark for a nearly ideal ear recognition scenario. Figure 4a shows some raw images provided by the IIT database that appear to have been taken in an ideal setting.

---

7 Materials from this chapter were published in “Employing Fusion of Learned and Handcrafted Features for Unconstrained Ear Recognition”, by E. Hansley, M. Pamplona Segundo, & S. Sarkar, 2018, IET Biometrics, http://dx.doi.org/10.1049/iet-bmt.2017.0210 The materials are reproduced by permission of the Institution of Engineering & Technology. Permission is included in Appendix A.
The WPUTE database [13] was originally created to evaluate the performance of ear recognition in the wild. It contains images that are challenging even for state-of-the-art ear recognition approaches. The images reflect the challenges associated with ear recognition, such as occlusions caused by hair, earrings, and headsets. The database also provides images with variations in gender, ethnicity, pose, illumination, and acquisition sensor. However, because the
vast majority of same subject images were acquired during a single session, intraclass variation is minimal. Thus, although the preprocessing step was heavily affected by these variations, some of the variations could in fact benefit the recognition task (e.g. a person wearing the same earring in all acquisitions). This database provides 3348 images with size $380 \times 500$ from 474 different subjects ($i.e.$ each subject has at least four images), showing a small region around the ear. However, 1388 of them are duplicates, which may have inflated the reported accuracy of some works in the literature [47], and we also found six images that were mistakenly labeled as left ears while they actually were right ears.

After removing duplicates and fixing labels, 1960 images were available for use, 982 from left ears and 978 from right ears. Some examples of WPUTE images are shown in Figure 4b.

7.3 Annotated Web Ears Database

The AWE database [12] contains 1000 images from 100 different subjects ($i.e.$ 10 images per subject) which were collected from searches for public figures on the Internet. Image size varies from $15 \times 29$ to $473 \times 1022$ pixels, with size $83 \times 160$ on average. Ears were tightly cropped, so the proportion of background pixels is the smallest among all databases used in this work. All variations presented in the WPUTEDB database are also present in the AWE database in an intenser form. Although it labels ears as left and right, with 520 and 480 images respectively, the images may have been inadvertently flipped horizontally before being released on the Internet. So, it is possible that there are some noisy labels. Figure 4c showcases samples of AWE test images that depict some of the challenges we encountered.
7.4 In-the-wild Ear Database

The ITWE database [47] is divided into two sets, Collection A and Collection B. Collection A was collected using Google image search and contains 605 images without identity reference, but with 55 manually annotated landmarks. The position of these landmarks can be observed in Figure 2. This collection was randomly split in a training set with 500 images and a test set with 105 images. It is suitable for training ear detection and normalization approaches, but not for recognition purposes. For this reason, Collection B was created for recognition evaluation and contains 2058 images from 231 different subjects taken from three public databases for face recognition in the wild: VGG-Face [31], LFW [15] and Helen Dataset [27].

Bounding boxes for each ear were obtained by a detector based on histograms of oriented gradients (HOG) [9] which were trained on images from Collection A, and these box coordinates were released together with this collection. Images in both Collection A and Collection B include cluttered backgrounds (e.g., face, body parts, scenario) and vary considerably in size and ear resolution. Variations in ear images of the ITWE database are comparable to the AWE database ones, but there is no differentiation between left and right ears (i.e., ITWE images are horizontally flipped so that all have the same orientation), which is a problem for recognizing people with asymmetric ears (i.e., about 10% of people according to Yan and Bowyer [45]). In addition, we were able to find many mislabeled samples. We did not fix any of them for comparison purposes. Some examples of ITWE images are presented in Figure 4d.

7.5 Unconstrained Ear Recognition Challenge Database

The UERC database [11] is an extension of the AWE database and was built for competition purposes. The major difference is the number of images and subjects; it has many more. The database is divided into two parts, with 2304 images from 166 subjects for training
and 9500 images from 3540 subjects for testing. The subjects designated for training have at least 10 images, while subjects in test may contain only one image. A portion of the subjects in training and testing (i.e. 150 and 180 subjects, respectively) have exactly 10 images. Ears may be left or right oriented, but ground truth annotations of the orientation are only available for training images.

7.6 Discussion

The five sets of images we used have different levels of difficulty that enabled us to conduct a fair test and evaluate the performance of our framework implementation, then compare our results to the state-of-the-art. While the IIT ear images are not unconstrained, they can be used to detect overfitting to the wild scenarios (i.e. using images from easier databases should always result in higher accuracy), a problem that was already observed in works that recognize faces in the wild [8]. Although all the remaining databases are unconstrained, based on their descriptive characteristics, we conclude that WPUTE and UERC are respectively the least and the most challenging unconstrained image sets, while AWE and ITWE are of similar difficulty.
CHAPTER 8: RESULTS

We designed our experiments to validate each module of our recognition framework. Thus, in the following sections we present separate results for landmark detection, geometric normalization, CNN-based description and descriptor fusion. We also compare our results to the state-of-the-art when possible.

8.1 Landmark Detection Results

Zhou and Zaferiou [47] evaluated different variations of Active Appearance Models (AAM) using the test set from ITWE’s Collection A. Their best result was achieved by training a holistic AAM based on SIFT features. As initialization for their landmark detector, they used a HOG-based ear detector. They computed the cumulative error distribution using the test set of the same database, where the error for an image is the normalized point-to-point error with respect to the diagonal of the bounding box for the ground truth annotations. Their best result is shown as a line with solid squares in Figure 5.

We performed the same evaluation for our proposed landmark detector in four different scenarios. In the first one, the ground truth annotations were used to obtain the center and the size of the ear. This reflects the performance of our method in a perfect scenario, in which ear’s location and size are reliably retrieved by an ear detector. Then, to simulate scenarios in which

---

8 Materials from this chapter were published in “Employing Fusion of Learned and Handcrafted Features for Unconstrained Ear Recognition”, by E. Hansley, M. Pamplona Segundo, & S. Sarkar, 2018, IET Biometrics, http://dx.doi.org/10.1049/iet-bmt.2017.0210 The materials are reproduced by permission of the Institution of Engineering & Technology. Permission is included in Appendix A.
the ear detector does not perform that well, we added random variations with up to 20%, 30% and 40% of the ear size to the ground truth values of the first scenario yielding four scenarios for us to compare. Figure 5 depicts the results of the four scenarios. As may be observed, our two-stage landmark detector performs slightly better than the single-stage one when using up to 20% of variation, and there is no significant difference in performance between ground truth initialization and an initialization with 20% of variation. This is expected, as this amount of variation was taken into account during the training stage. For larger variations that are unknown to the training, a single-stage landmark detector can have a considerable drop in performance, but our two-stage solution does not experience a considerable drop. It is able to perform at least as well as the state-of-the-art [47].

Figure 5: Cumulative error distribution for landmark detection using the proposed approach and Zhou and Zaferiou’s approach [47].

8.2 Normalization Results

Since there was no normalization ground truth, we evaluated the benefits of the normalization process by checking the difference in the recognition performance with and without normalization for different handcrafted descriptors. To this end, we normalized all images from the AWE database and followed the same protocol proposed by Emersic et al. [12]
that was released in their toolbox. We used the development set that contains 60% of the AWE images. We performed a 10-fold cross-validation, and we report the mean accuracy and standard deviation in Table 3. It is worth highlighting that even images that were not correctly normalized were still used in all recognition experiments, and that all folds are exactly the same in our and Emersic et al.’s works. Table 3 shows Emersic et al.’s results, our reproduction of their experiments and our results using normalized images in terms of Rank-1 and Equal Error Rate (EER). It is possible to observe that our results without normalization are very similar to the ones reported by Emersic et al., showing that our reproduction of their experiments was successful, and that our results with normalization yielded higher Rank-1 recognition rates for all features and a lower EER in most of the cases. These results illustrate that an effective normalization approach can help improve performance when using description techniques that are not necessarily robust to wild ear variations.

### 8.3 CNN Description Results

In order to learn features for the problem of recognizing ears in the wild, we divided each of the IIT,WPUTE, AWE and ITWE databases in two sets, one for training and one for testing.

Table 3 Rank-1 and EER results for the AWE database as reported by Emersic et al. [12] and our reproduction of their experiments using images with normalization and images without (raw) normalization.

<table>
<thead>
<tr>
<th>Method</th>
<th>Emersic et al.’s work</th>
<th>This work (raw)</th>
<th>This work (norm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rank-1</td>
<td>EER</td>
<td>Rank-1</td>
</tr>
<tr>
<td>LBP</td>
<td>43.5±7.1</td>
<td>32.0±7.4</td>
<td>43.5±7.1</td>
</tr>
<tr>
<td>BSIF</td>
<td>48.4±6.8</td>
<td>30.0±9.6</td>
<td>48.4±6.4</td>
</tr>
<tr>
<td>LPQ</td>
<td>42.8±7.1</td>
<td>30.0±7.4</td>
<td>42.6±7.0</td>
</tr>
<tr>
<td>RILPQ</td>
<td>43.3±9.4</td>
<td>34.0±6.4</td>
<td>43.5±9.2</td>
</tr>
<tr>
<td>POEM</td>
<td>49.6±6.8</td>
<td>29.1±9.1</td>
<td>49.6±6.8</td>
</tr>
<tr>
<td>HOG</td>
<td>43.9±7.9</td>
<td>31.9±7.8</td>
<td>48.1±8.8</td>
</tr>
<tr>
<td>DSIFT</td>
<td>43.4±8.6</td>
<td>30.9±8.4</td>
<td>42.2±9.0</td>
</tr>
<tr>
<td>Gabor</td>
<td>39.8±7.1</td>
<td>32.1±8.1</td>
<td>39.8±7.1</td>
</tr>
</tbody>
</table>
We segregated them in a subject-independent way (i.e. no subject has images in both training and testing sets) by taking the first half of the subjects rounded up and using their images for training, and using the remaining ones for testing. Before the automatic normalization process, we flipped both left and right ear images so that all ears would have the same orientation. Then, we normalized and transformed each training set image into a set of 20 modified images during the data augmentation stage, as described in Chapter 4.1, and we trained five different descriptors: one for each of the four databases training set and one using all these training sets combined. We evaluated the EER performance of these five descriptors by performing an all-versus-all comparison in all testing sets available. The results are presented in Table 4. As may be observed, the best performance for all unconstrained testing sets was obtained by the descriptor learned using all training sets, followed by the descriptor learned using the training set from the same database. This shows that every database has different types of variations that tend to be overrepresented in models learned from a single database. When all databases are combined, the model benefits from both a wider training set (i.e. more subjects) and less database overfitting. The models do not appear to be overfitting the unconstrained images, as the performance for the IIT test set is about 2% for all models.

We also show in Table 4 that knowing whether or not the image is of the left ear or right ear is helpful during the recognition process. If we only consider genuine matchings as the matchings between ear images from the same side of the head, the EER is reduced in about 4-6% for the WPUTE database and in approximately 2-3% for the AWE database in all tests. These results corroborate the findings of Yan and Bowyer [45] regarding ear asymmetry, but in an uncontrolled scenario. However, it is not always possible to have this information, so we did not
consider ear asymmetry in the following experiments and classified matchings between different ears of the same person as genuine.

Table 4: EER results for all testing sets using descriptors learned from each database or from all databases combined.

<table>
<thead>
<tr>
<th>TEST</th>
<th>TRAIN</th>
<th>IIT</th>
<th>WPUTE</th>
<th>WPUTE*</th>
<th>AWE</th>
<th>AWE*</th>
<th>ITWE</th>
</tr>
</thead>
<tbody>
<tr>
<td>IIT</td>
<td>1.76%</td>
<td>29.62%</td>
<td>25.85%</td>
<td>35.29%</td>
<td>33.56%</td>
<td>35.47%</td>
<td></td>
</tr>
<tr>
<td>WPUTE</td>
<td>2.12%</td>
<td>15.95%</td>
<td>9.40%</td>
<td>29.87%</td>
<td>28.33%</td>
<td>29.46%</td>
<td></td>
</tr>
<tr>
<td>AWE</td>
<td>2.12%</td>
<td>25.03%</td>
<td>20.04%</td>
<td>26.53%</td>
<td>23.52%</td>
<td>25.68%</td>
<td></td>
</tr>
<tr>
<td>ITWE</td>
<td>2.37%</td>
<td>23.50%</td>
<td>18.93%</td>
<td>27.51%</td>
<td>25.30%</td>
<td>22.09%</td>
<td></td>
</tr>
<tr>
<td>ALL</td>
<td>2.59%</td>
<td>15.17%</td>
<td>9.59%</td>
<td>25.42%</td>
<td>22.93%</td>
<td>19.68%</td>
<td></td>
</tr>
</tbody>
</table>

* As WPUTE and AWE distinguish left and right ears, we also show results considering only genuine matchings between ear images from the same side of the head.

Zhou and Zaferiou [47] used transfer learning in order to employ CNN descriptors previously trained in a different domain [35] for ear recognition. To this end, they evaluated both Support Vector Machines (SVM) and Linear Discriminant Analysis (LDA) for matching those descriptors, and achieved about 30% EER for the ITWE database. Their testing/training proportion was 80%/20%, and the division was not made in a subject-independent manner. We experimented with a much more difficult scenario, using a 50%/50% testing/training subject-independent split and still achieved a considerably lower EER in all cases where a true unconstrained database was used for training (i.e. AWE, ITWE and ALL), as may be observed in Table 4. These results show that even when a small number of images are available for the ear domain, it may be worth it to train a domain-specific CNN.

### 8.4 Fusion Results

For our first round of fusion experiments, we used the testing set of the AWE database, since this was the most challenging one in our previous experiment. We evaluated the fusion of all possible pairs of features, including all holistic, handcrafted and learned features presented in Chapter 5. The chosen CNN model was the one with the best result in Table 4 (ALL). Table 5
shows individual results for each feature, as well as the top fusion results in terms of Rank-1, Rank-5, Area Under Curve (AUC) and EER. As may be observed, although learned features and top handcrafted features perform equally well individually for Rank-1 and Rank-5, fusion results were dominated by CNN combinations. We believe this is caused by a larger correlation among handcrafted features, which usually have a similar design inspiration that was exploited in slightly different ways by different experts (e.g. quantize gradients, encode neighbors). Thus, CNN is probably learning something complementary to the experts’ knowledge, which is corroborated by the fact that nearly all combinations between CNN and one handcrafted feature perform better than all combinations between two handcrafted features.

Table 5: Individual and fusion results for all descriptors in Chapter 5 using the AWE database. Individual results were grouped by descriptor type, and handcrafted features were grouped into categories, the first one (Handcrafted I) for methods based on neighborhood encoding and the second one (Handcrafted II) for methods based on gradient orientations.

<table>
<thead>
<tr>
<th>Type</th>
<th>Descriptors</th>
<th>Rank-1</th>
<th>Rank-5</th>
<th>AUC</th>
<th>EER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Holistic</td>
<td>PCA</td>
<td>43.0%</td>
<td>64.4%</td>
<td>0.866</td>
<td>37.87%</td>
</tr>
<tr>
<td>Handcrafted I</td>
<td>POEM</td>
<td>65.2%</td>
<td>85.0%</td>
<td>0.948</td>
<td>31.68%</td>
</tr>
<tr>
<td></td>
<td>BSIF</td>
<td>63.8%</td>
<td>83.4%</td>
<td>0.939</td>
<td>30.82%</td>
</tr>
<tr>
<td></td>
<td>LBP</td>
<td>62.0%</td>
<td>82.6%</td>
<td>0.939</td>
<td>30.00%</td>
</tr>
<tr>
<td></td>
<td>LPQ</td>
<td>59.6%</td>
<td>84.2%</td>
<td>0.942</td>
<td>30.52%</td>
</tr>
<tr>
<td></td>
<td>RILPQ</td>
<td>55.0%</td>
<td>79.2%</td>
<td>0.926</td>
<td>34.07%</td>
</tr>
<tr>
<td>Handcrafted II</td>
<td>HOG</td>
<td>64.2%</td>
<td>86.2%</td>
<td>0.955</td>
<td>29.33%</td>
</tr>
<tr>
<td></td>
<td>DSIFT</td>
<td>57.8%</td>
<td>78.4%</td>
<td>0.916</td>
<td>32.99%</td>
</tr>
<tr>
<td></td>
<td>GABOR</td>
<td>50.2%</td>
<td>75.6%</td>
<td>0.911</td>
<td>32.56%</td>
</tr>
<tr>
<td>Learned</td>
<td>CNN</td>
<td>64.2%</td>
<td>86.2%</td>
<td>0.957</td>
<td>22.89%</td>
</tr>
<tr>
<td>Sum fusion</td>
<td>CNN+HOG</td>
<td>75.6%</td>
<td>90.6%</td>
<td>0.972</td>
<td>22.87%</td>
</tr>
<tr>
<td></td>
<td>CNN+POEM</td>
<td>75.4%</td>
<td>90.4%</td>
<td>0.968</td>
<td>24.29%</td>
</tr>
<tr>
<td></td>
<td>CNN+LPQ</td>
<td>72.8%</td>
<td>88.6%</td>
<td>0.966</td>
<td>23.61%</td>
</tr>
<tr>
<td></td>
<td>CNN+RILPQ</td>
<td>72.0%</td>
<td>90.6%</td>
<td>0.962</td>
<td>25.11%</td>
</tr>
<tr>
<td></td>
<td>HOG+BSIF</td>
<td>70.8%</td>
<td>88.6%</td>
<td>0.963</td>
<td>28.34%</td>
</tr>
<tr>
<td></td>
<td>CNN+BSIF</td>
<td>70.2%</td>
<td>89.8%</td>
<td>0.963</td>
<td>24.18%</td>
</tr>
<tr>
<td></td>
<td>CNN+LBP</td>
<td>70.0%</td>
<td>89.4%</td>
<td>0.964</td>
<td>23.53%</td>
</tr>
<tr>
<td></td>
<td>HOG+RILPQ</td>
<td>70.0%</td>
<td>86.4%</td>
<td>0.957</td>
<td>29.89%</td>
</tr>
<tr>
<td></td>
<td>CNN+GABOR</td>
<td>69.4%</td>
<td>88.6%</td>
<td>0.963</td>
<td>24.56%</td>
</tr>
<tr>
<td></td>
<td>HOG+LPQ</td>
<td>69.0%</td>
<td>86.4%</td>
<td>0.960</td>
<td>28.67%</td>
</tr>
</tbody>
</table>
For our second round of fusion experiments, we reproduced two experiments proposed by Emersic et al. [11] to evaluate challenge participants using the UERC database. One experiment evaluates the overall performance and the other evaluates the scalability of the recognition approaches. To this end, we normalized the UERC training images and used them to learn a sixth CNN descriptor (i.e. data augmentation was used to balance the classes in a way that each subject ended up with 200 images). Because UERC test images do not have the same orientation and ground truth annotations are not provided, we also trained a simple side classifier by changing the output size of the network presented in Table 1 to two classes (left and right) and then trained it for the UERC training images using softmax loss and the Adam optimization algorithm. Because the images from this database are already cropped, the entire testing process was fully automatic.

For the overall performance evaluation, only the first 1,800 test images from 180 subjects were used in an all-versus-all comparison. In this experiment, we only used CNN and three other handcrafted features: HOG, POEM and LBP. HOG and POEM obtained the best fusion results with CNN in Table 5, and LBP was a baseline approach for participants of the UERC challenge [11]. Table 6 shows individual results for each feature, results using sum fusion, and the best results reported in the UERC challenge. As may be observed, our normalization resulted in a considerable boost in performance for handcrafted descriptors. We achieved more than 20% improvement in Rank-1 when comparing our LBP result to its baseline version without normalization. Again, individual ranking performances of learned and handcrafted features were similar, but the CNN fusion pairings stood out. Our performance was higher than all participants of the challenge except University of Colorado Colorado Springs (UCCS), whose results we
have not verified. It seems that they may be using test images for training. CMC curves for the best performing works are presented in Figure 6a.

Table 6: Individual and fusion results for CNN, HOG, POEM and LBP in the overall performance evaluation through the UERC protocol, as well as the top scoring participants of the UERC challenge.

<table>
<thead>
<tr>
<th>Type</th>
<th>Descriptors</th>
<th>Rank-1</th>
<th>Rank-5</th>
<th>AUC</th>
<th>EER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Handcrafted I</td>
<td>POEM</td>
<td>36.83%</td>
<td>58.44%</td>
<td>0.907</td>
<td>36.17%</td>
</tr>
<tr>
<td></td>
<td>LBP</td>
<td>35.00%</td>
<td>55.11%</td>
<td>0.897</td>
<td>35.81%</td>
</tr>
<tr>
<td>Handcrafted II</td>
<td>HOG</td>
<td>39.78%</td>
<td>60.56%</td>
<td>0.916</td>
<td>35.51%</td>
</tr>
<tr>
<td>Learned</td>
<td>CNN</td>
<td>36.94%</td>
<td>60.56%</td>
<td>0.930</td>
<td><strong>26.77%</strong></td>
</tr>
<tr>
<td>Sum fusion</td>
<td>CNN+HOG</td>
<td><strong>49.06%</strong></td>
<td><strong>69.94%</strong></td>
<td><strong>0.951</strong></td>
<td>27.84%</td>
</tr>
<tr>
<td></td>
<td>CNN+POEM</td>
<td>47.28%</td>
<td>70.00%</td>
<td>0.948</td>
<td>28.21%</td>
</tr>
<tr>
<td></td>
<td>CNN+LBP</td>
<td>45.22%</td>
<td>67.44%</td>
<td>0.946</td>
<td>28.05%</td>
</tr>
<tr>
<td></td>
<td>HOG+POEM</td>
<td>43.06%</td>
<td>64.33%</td>
<td>0.926</td>
<td>35.14%</td>
</tr>
<tr>
<td></td>
<td>HOG+LBP</td>
<td>41.22%</td>
<td>60.89%</td>
<td>0.919</td>
<td>35.11%</td>
</tr>
<tr>
<td></td>
<td>POEM+LBP</td>
<td>38.56%</td>
<td>59.00%</td>
<td>0.911</td>
<td>35.39%</td>
</tr>
<tr>
<td>Literature</td>
<td>UCCS [11]</td>
<td>90.4%*</td>
<td>100.0%*</td>
<td>0.994*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IAU [11]</td>
<td>38.5%</td>
<td>63.2%</td>
<td>0.940</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ITU-II [11]</td>
<td>27.3%</td>
<td>48.3%</td>
<td>0.877</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LBP-baseline [11]</td>
<td>14.3%</td>
<td>28.6%</td>
<td>0.759</td>
<td></td>
</tr>
</tbody>
</table>

For the scalability evaluation, we matched all images from subjects with at least two images to all other test images, totaling 7,442×9,499 matching pairs. This experiment increased the number of subjects to 3,540 and also adds many images with poor quality, affecting considerably the performance of the evaluated approaches. In Table 7 we show results for CNN, HOG and POEM, for all possible fusion among two of them, and for the best performing approaches in the UERC challenge. We can see that the combination of CNN and HOG was again the best performing method for lower ranks, and that these results show that our approach is the most scalable unconstrained ear recognition approach. CMC curves for the best performing works are presented in Figure 6b and show how well our approach performs for lower ranks, outperforming all other works by at least 10% in most ranks before Rank-300.
Table 7: Individual and fusion results for CNN, HOG and POEM in the scalability evaluation through the UERC protocol, as well as the top scoring participants of the UERC challenge.

<table>
<thead>
<tr>
<th>Type</th>
<th>Descriptors</th>
<th>Rank-1</th>
<th>Rank-5</th>
<th>AUC</th>
<th>EER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Handcrafted I</td>
<td>POEM</td>
<td>17.98%</td>
<td>28.48%</td>
<td>0.851</td>
<td>40.38%</td>
</tr>
<tr>
<td>Handcrafted II</td>
<td>HOG</td>
<td>18.50%</td>
<td>28.78%</td>
<td>0.851</td>
<td>40.80%</td>
</tr>
<tr>
<td>Learned</td>
<td>CNN</td>
<td>17.13%</td>
<td>28.73%</td>
<td>0.873</td>
<td>35.92%</td>
</tr>
<tr>
<td>Sum fusion</td>
<td>CNN+HOG</td>
<td>24.17%</td>
<td>36.43%</td>
<td>0.882</td>
<td>36.24%</td>
</tr>
<tr>
<td></td>
<td>CNN+POEM</td>
<td>23.02%</td>
<td>35.70%</td>
<td>0.882</td>
<td>35.82%</td>
</tr>
<tr>
<td></td>
<td>HOG+POEM</td>
<td>20.57%</td>
<td>32.12%</td>
<td>0.856</td>
<td>40.26%</td>
</tr>
<tr>
<td>Literature</td>
<td>UCCS [11]</td>
<td>22.3%</td>
<td>26.3%</td>
<td>0.858</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IAU [11]</td>
<td>8.4%</td>
<td>14.2%</td>
<td>0.810</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ITU-II [11]</td>
<td>6.9%</td>
<td>12.8%</td>
<td>0.844</td>
<td></td>
</tr>
</tbody>
</table>

*These results still require verification

Table 7 also shows that our CNN outperformed the other two top-scoring CNN-based approaches proposed by researchers from the Islamic Azad University (IAU) and the Istanbul Technical University (ITU-II) [11]. Similar to Zhou and Zaferiou [47], IAU and ITU-II employed transfer learning approaches in a network from a different domain [35] and were not able to achieve results as high as our domain-specific CNN.

8.5 Discussion

Unconstrained ear recognition is a very challenging problem, and recent efforts to provide data for unconstrained ear images for research are helpful. Initial databases such as IIT and WPUTE were captured images instead of wild images. They do not have much intraclass and interclass variations. The initial wild databases collected images from the Internet such as AWE and ITWE still lack interclass variability due to their small number of subjects. The UERC database is a vast repository with thousands of subjects and images with intraclass variations and interclass variations. It is the most challenging ear dataset that we are aware of.

Although initial ear recognition works have consistently used ear alignment before recognition [5, 25], researches for unconstrained ear recognition were most focused on finding robust features [12]. Even among the UERC participants, only the Imperial College London
(ICL) used an alignment step, although they used an AAM-based solution [47] that may not be as successful for wild images as recent techniques such as CNNs (see Figure 5). Nevertheless, we attribute a big part of the success in our results to the normalization step. It considerably increased the performance of traditional methods, such as handcrafted features in Tables 3 and 6, and, also helped the deep learning process by letting it focus on what matters the most for the recognition task. It also helped in our cross-dataset experiments shown in Table 4, as we do not have problems with different cropping areas or noise in ear location.

Our CNN descriptors were not only comparable to the best handcrafted descriptors in terms of Rank-N results, but they performed better in terms of EER in all experiments, meaning that they were more accurate for verification purposes. In addition, our performance was favorably compared to the best performing participants of the UERC challenge, as shown in Table 7, and to the best results from the state-of-the-art approaches. There were two factors that may have helped us to achieve these results: we trained CNNs from scratch specifically to our problem domain, and we used a discriminative learning technique based on center loss that was proposed by Wen et al. [42].

Finally, as learned and handcrafted features were achieving similar ranking results for our normalized images, we decided to combine them through score fusion in order to seek a better performance rate. We discovered that the combination of our CNN descriptors and handcrafted descriptors achieved much better results in all experiments. None of the combinations between a pair of handcrafted features could get close to the top scores, which may be explained by the fact that handcrafted features are highly correlated due to their similar design. On the other hand, CNN descriptors do not follow an expert’s design and are likely learning some discriminative
information that is complementary to most handcrafted descriptors, as may be observed in Tables 5, 6 and 7.

Figure 6: CMC curves for all participants of the UERC challenge plus our best fusion results obtained by combining CNN and HOG considering the (a) overall performance evaluation and (b) scalability evaluation protocols.

Finally, as learned and handcrafted features were achieving similar ranking results for our normalized images, we decided to combine them through score fusion in order to seek a better
performance rate. We discovered that the combination of our CNN descriptors and handcrafted
descriptors achieved much better results in all experiments. None of the combinations between a
pair of handcrafted features could get close to the top scores, which may be explained by the fact
that handcrafted features are highly correlated due to their similar design. On the other hand,
CNN descriptors do not follow an expert’s design and are likely learning some discriminative
information that is complementary to most handcrafted descriptors, as may be observed in Tables
5, 6 and 7.
CHAPTER 9: CONCLUSION

Ear recognition has progressed from the initial research conducted two centuries ago whereby identification of individuals from ears was done manually by a subject matter expert. With the advances in Computer Technology, today ear recognition is an automated process. And, because of the availability of faster and cheaper computers with more storage capacity, one can assert that it is a lot easier to do ear recognition nowadays. But, ear recognition research has not reached a plateau. There is still a lot of room for improvement. The foundational ear recognition research and most prior ear recognition research has been done using ideal ear images captured in a laboratory or similar setting. Using ideal ear images has some advantages. They enable researchers to provide a proof of concept for their ideas and they provide a starting point that allows small successful steps. It is those small steps of starting with easy biometric recognition problems that enable research to progress to the point being able to experiment with very difficult images and identify individuals from them. In order to advance the field of Computer Vision, ear recognition research must move beyond using constrained images that are taken in a controlled environment to being able to handle unconstrained ear images.

Unconstrained ear recognition is a very challenging problem. To address the challenge of unconstrained ear recognition, we devised a framework that combines handcrafted features and CNN. We tested our framework using the most challenging publicly available ear databases that

---

9 Materials from this chapter were published in “Employing Fusion of Learned and Handcrafted Features for Unconstrained Ear Recognition”, by E. Hansley, M. Pamplona Segundo, & S. Sarkar, 2018, IET Biometrics, http://dx.doi.org/10.1049/iet-bmt.2017.0210 The materials are reproduced by permission of the Institution of Engineering & Technology. Permission is included in Appendix A.
we are aware of. Our results are considerably better than recently published works and less impacted by database scale. But, this isn’t the only significant thing about our research. Using our framework, we were able to identify individuals from ears using ear images that that contained earrings and other ear occlusions. This is also significant because these types of images would not have been considered just ten years ago. And, most previous ear recognition research does not use unconstrained images.

As a result of our research, we gained invaluable lessons that can further enhance unconstrained ear recognition research. Handcrafted features are not dead. Handcrafted features and CNN are complementary. Normalization is critical and enhances performance recognition of handcrafted features. CNN combined with any of the state-of-the-art descriptors that we used improves recognition.

There is still a lot to be learned in order to address the difficult challenges associated with unconstrained ear recognition. This dissertation demonstrates that CNN and handcrafted features are a good starting point. Our ear recognition framework is a major step towards identification of individuals from ears in real world conditions.
REFERENCES


APPENDIX A: COPYRIGHT CLEARANCE FORMS

Below is permission granted to use the materials in Chapter 1 through Chapter 9.

Ref: EEH/Permission/0210

FAQ: Earnest Eugene Hansley

3 April 2018

Permission to reproduce IET content.

Dear Sirs,

Employing fusion of learned and handcrafted features for unconstrained ear recognition ("the Material") to be used in the dissertation 'Identification of Individuals from Ears in Real World Conditions', by Earnest Eugene Hansley to be published under the University of South Florida ("the Work")

The Institution of Engineering and Technology ("the IET") hereby grants to Earnest Eugene Hansley (EEH) non-exclusive permission to use the Material in the Work subject to the terms set out below:

Territory: Worldwide
Languages: All languages
Term: Life of the Work
Media: All print and electronic media known or hereafter devised
Fee: None

The permission granted by this letter may not be licensed or assigned without the prior written consent of the IET.

Neither party shall be liable to the other for indirect, incidental, special or consequential damages arising out of or in relation to this agreement (unless such liability cannot be excluded or limited by law).

The parties agree that this letter constitutes the entire agreement between them relating to the use of the Material by EEH and supersedes all previous agreements, understandings and arrangements between them, whether in writing or oral in respect of its subject matter.

No variation of this agreement shall be valid or effective unless it is in writing, refers to this agreement and is duly signed or executed by, or on behalf of, each party.

This letter and any dispute or claim arising out of, or in connection with, it, its subject matter or formation (including non-contractual disputes or claims) shall be governed by, and construed in accordance with, the laws of England and Wales and the parties irrevocably
agree that the courts of England and Wales shall have exclusive jurisdiction to settle any dispute or claim arising out of, or in connection with, this letter, its subject matter or formation (including non-contractual disputes or claims).

Please sign and return the enclosed copy of this letter to confirm your agreement to it.

Yours faithfully

James Sutherland
Permissions Officer

We confirm our agreement to the terms set out above

Signed: [Signature]
Print name: [Name]
Date: 04/03/2018
ABOUT THE AUTHOR

Earnest Hansley is a career military veteran from Vidalia, Georgia. He graduated from Vidalia High School in 1977 and enlisted into the United States Army as a Medical Specialist after graduation. After serving in the army for four years as an enlisted soldier, he entered college at Fort Valley State University and joined the Reserve Officers’ Training Corps (ROTC). He was awarded a three-year Army ROTC scholarship after his Freshman year. He graduated from Fort Valley State University with honors and as a Distinguished Military Graduate in 1985. He earned a Bachelor of Science degree in Computer Science and he received a commission into the Regular Army as a Signal Corps officer. Because of his outstanding performance and outstanding academic record, early in his career, he was selected to attend the Naval Postgraduate School fulltime to pursue his master’s degree. He received a Master of Science Degree in Systems Technology from The Naval Postgraduate School in Monterey, California in 1996. Later in his career, because of his continued superior performance and scholarly acumen, he was selected to attend the University of South Florida fulltime to pursue a Doctor of Philosophy degree. After September 11, 2011, Earnest was transferred to the Pentagon. Earnest served with distinction in a variety of command and staff positions stateside and overseas during over three decades of military service. He has received numerous military awards and honors. He retired at the Pentagon in 2014 as an Army Colonel, and with this dissertation Earnest culminates his doctoral studies and earns his Ph.D. in Computer Science and Engineering at the University of South Florida.