SOURCES OF ERROR IN IRIS BIOMETRICS

A Thesis

Submitted to the Graduate School
of the University of Notre Dame
in Partial Fulfillment of the Requirements
for the Degree of

Master of Science
in Computer Science and Engineering

by

Karen P. Hollingsworth, B.S.

____________________________

Kevin W. Bowyer, Director

____________________________

Patrick J. Flynn, Director

Graduate Program in Computer Science and Engineering
Notre Dame, Indiana
April 2008
SOURCES OF ERROR IN IRIS BIOMETRICS

Abstract

by

Karen P. Hollingsworth

Two decades ago, the United States issued the first patent claiming the idea of an automated iris biometrics system [1]. Today, multiple companies offer commercial biometrics products. However, there are still many unanswered questions about the workings of the iris biometrics algorithms. Before iris recognition systems are more widely employed, we must ask, “when do these algorithms fail?”

Previous research has assumed that all parts of an iris code are equally valuable. Alternatively, some researchers claim that parts of the iris are more valuable, but they still use the same portions of the iris for all subjects. No previous researcher has attempted to experimentally determine how different parts of a particular subject’s iris code may be more or less valuable. I obtained multiple images of 24 subjects’ eyes to study the reliability of individual bits in the iris codes. I develop a theoretical explanation of the sources of inconsistencies, based on the coarse quantization of complex coefficients in creating the iris code.

Another source of inconsistency in the iris code is dilation of the iris. The majority of iris research ignores the degree of dilation in processing iris images for biometric purposes. I experimentally quantify how much the Hamming distance is affected by iris dilation.
CONTENTS

FIGURES ................................................................. iv

ACKNOWLEDGMENTS .................................................... viii

CHAPTER 1: INTRODUCTION ............................................. 1

CHAPTER 2: BACKGROUND ............................................. 3
  2.1 Performance of Biometric Systems .............................. 3
  2.2 Eye Anatomy ..................................................... 5
  2.3 Early Research in Iris Biometrics ............................... 7
  2.4 Recent Research in Iris Biometrics ............................ 13
  2.5 Iris Biometrics Evaluations .................................... 18

CHAPTER 3: FRAGILE BITS ............................................. 23
  3.1 Data and Software .............................................. 24
  3.2 Existence of Fragile Bits ...................................... 30
  3.3 Outliers .......................................................... 38
  3.4 Inner vs. Outer Radial Bands of the Iris ...................... 45
  3.5 Effect of Granularity of Iris Alignment on Consistency .... 49
  3.6 Effect of Filter on Consistency ................................ 51
  3.7 Theoretical Impact of Fragile Bits on False Reject Rate ... 54
  3.8 Empirical Evidence of Improved Accuracy .................... 58
  3.9 Discussion ....................................................... 62

CHAPTER 4: DILATED PUPILS .......................................... 63
  4.1 Research in Iris Dilation ...................................... 63
  4.2 Data and Software .............................................. 65
  4.3 Measuring Dilation ............................................. 68
  4.4 Degree of Dilation Affects Performance ...................... 72
  4.5 Comparisons between iris images with varying degrees of dilation 78
  4.6 Recommendations .............................................. 80
FIGURES

2.1 Image 02463d1276 from the Iris Challenge Evaluation Dataset. Elements seen in a typical iris image are labeled here. 6
2.2 Major steps in iris biometrics processing. 12
2.3 This “Iris on the Move” portal acquires an iris image as subjects walk through a portal at normal walking speed. The portal itself contains infrared lights to illuminate the subject. Three high-zoom video cameras in the far cabinet take video of the subject. 15
3.1 Histogram of an image acquired by LG 2200, using 171 of 256 intensity levels. 25
3.2 Example image 04233d1145 from our dataset. This image has no occlusion by eyelids or eyelashes. However, part of the sclera is included in the segmentation of this iris region. Also, some specular highlight is evident in the iris region near the lower eyelid. 26
3.3 Example image 02463d1268 from our dataset. This iris contains specular highlights near the bottom of the iris, and near the pupil. 27
3.4 In this image, parts of the top and bottom of the iris were masked before the iris code was computed. The corresponding parts of the iris code of all images of this iris were dropped from the experimental analysis. This is image 04239d1060 from our dataset. 28
3.5 Despite including some poorly focused images in the dataset, on average 15.96% of bits in the iris codes were perfectly consistent. This is image 04336d692 from our dataset. 29
3.6 Black areas in each rectangle are inconsistent parts of the iris code, and white areas are consistent. Each rectangle represents data from the iris codes of at least 15 different images of the same eye. Two rectangles are shown for each subject; one rectangle shows the real bits in the iris code, and the next rectangle shows the imaginary bits. Light gray regions are masked regions. The numbers on the side are the subject numbers associated with images in the ICE dataset. 32
3.7 The number of fragile bits depends on the threshold used. There are fewer bits that flip in at least 40% of the iris codes than there are that flip in at least 20% of the iris codes. 33

3.8 There is no difference in the average fragility of iris code bits between the two different genders. 35

3.9 The four Asian subjects in our dataset tended to have more fragile bits in their iris codes, but with only four subjects, we cannot conclude that this is a general trend. 36

3.10 These 54 complex numbers, each from the same region on 54 different images of the same subject’s eye, all correspond to the same location on the iris code. Each complex number is mapped to two bits. This particular part of the iris code had a highly consistent real bit, and a highly inconsistent imaginary bit. 37

3.11 Close-up view of image 02463d1323 showing specular highlights. Triangles mark the positions on the iris that resulted in outliers after the 1D log-Gabor filter was applied. A zoomed-in portion of the corresponding normalized iris is also shown. 40

3.12 Faint reflections often appear close to brighter reflections. This image contains (A) bright reflections on the eyelid, (B) a nearby, faint reflection on the iris, (C) a bright reflection in the pupil, and (D) a nearby, faint reflection on the iris. Both of the faint reflections marked in this image (B and D) caused outliers in the complex data. (Image 02463d1329). 42

3.13 The positions of the sampling points around a specular highlight affect how extreme an outlier is produced. Green circles represent the sampling points. Red, upward-pointing triangles are outliers greater than 4 standard deviations from average. Blue, downward-pointing triangles are outliers greater than 3 standard deviations from average. Yellow, right-pointing triangles are outliers greater than 2 standard deviations from average. A sample point in the center of a highlight corresponds with an extreme outlier (A). Sample points falling on the edge of a specular highlight cause less-extreme outliers (B). (Image 02463d1249) 43

3.14 Some outliers showed up on dark regions of the iris image. Small dots show the sample points on the iris used to create the normalized image. (02463d1276) 44

3.15 This figure shows the percent of fragile bits in each row of the iris code. Rows in the middle of the iris code (rows 5 through 12) are the most consistent. 46
3.16 Mask used for evaluating the consistency of the inner versus outer bands of the iris. When we consider only the right and left sides of the iris, and disregard the top and bottom sections, which had the most specular highlights and occlusion, there is not as much spread in the fraction of fragile bits across the rows. There still seems to be a high percentage of fragile bits in rows 1 and 2 of the iris code; however, the general trend shows that all rows of the iris code have a high amount of consistent information.

3.17 When the regions most affected by specular highlights and occlusion are masked, the consistency of the outer rows of the iris code improves. (The black horizontal line is drawn to aid in comparison between outer and inner rows.)

3.18 In matching a pair of iris codes, multiple different orientations of the probe iris code are considered. Allowing for smaller possible rotations decreases the percent of fragile bits.

3.19 Each rectangle shows the fragile bits from part of the OSIRIS iris code. Black bits represent fragile bits. The rectangles shown here are the part of the iris code created with the imaginary part of the largest Gabor filter used.

3.20 For the large filter, inconsistent regions of the iris code (shown in yellow) often fall in between regions consistently equal to one (shown in red) and regions consistently equal to zero (shown in blue). In the lower pane, the corresponding black and white figure is shown, with inconsistent regions drawn in black.

3.21 The largest filters had the highest number of bits in the first bin, and therefore, in general, the largest filters seem to produce the most consistent bits.

3.22 This figure shows the match and non-match distributions for our software, without any masking of fragile bits.

3.23 This figure shows the match and non-match distributions for our software, when bits close to the complex axis are masked. The match distribution has moved a significant amount to the left, closer to 0, as desired. In addition, the non-match distribution has also widened slightly.

4.1 Distribution of normalized Hamming distances for all iris comparisons in our data set. Two histograms are shown. The one on the left represents comparisons between two images of the same iris. The one on the right represents comparisons between two different irises.
4.2 Histogram showing different degrees of dilation for irises in our data set. A dilation ratio of 0.2 corresponds to a very small pupil (contracted eye), and a dilation ratio of 0.7 corresponds to a very large pupil (dilated eye). 

4.3 The dilation ratio is the pupil radius divided by the iris radius. Some subjects’ eyes naturally tended to have higher dilation ratios than other subjects’ eyes. Also, some subjects showed more variation in their pupil size than others. Subject 9 had the largest variation (0.307), with a maximum dilation ratio of 0.655 and a minimum dilation ratio of 0.348. The subjects have been ordered in this plot by increasing minimum dilation ratio.

4.4 It is easy to obtain a difference in pupil dilation. The figure on the left is taken with an extra lamp in the room, and the figure on the right is taken wearing sunglasses.

4.5 This subject (subject number 05288) showed the biggest difference in pupil size in the data set. The smallest dilation ratio (pupil radius/iris radius) for this subject was 0.3478 and the largest dilation ratio was 0.6545.

4.6 This subject (subject number 02463) had the smallest pupils in the dataset. The smallest dilation ratio for this subject was 0.2459, and the largest dilation ratio was 0.4400.

4.7 This subject (subject number 05379) had the largest pupils in the dataset. The smallest dilation ratio for this subject was 0.5391, and the largest was 0.7009.

4.8 The match distribution for comparisons between eyes with large pupils has a larger mean than the distribution of comparisons between eyes with small pupils.

4.9 The non-match distribution of comparisons between eyes with large pupils has a smaller mean than the distribution of comparisons between eyes with small pupils.

4.10 \( \Delta \) is a quantity that refers to the difference in dilation between two eyes in a comparison.

4.11 The match distribution of comparisons between eyes with different degrees of dilation has a larger mean than the distribution of comparisons between eyes with the same degrees of dilation. The non-match distribution is not noticeably affected by the difference in dilation in the comparisons.
ACKNOWLEDGMENTS

This thesis would have been an impossible task without the assistance of my dedicated and supportive advisors. Dr. Bowyer and Dr. Flynn have spent countless hours in teaching, guiding, proofreading my papers, and encouraging me in my research.

I also thank my husband, Nathaniel, for listening to my daily reports of my progress, for staying up late with me when I have had deadlines to meet, and for being my number one fan.

This effort is sponsored all or in part by the Central Intelligence Agency and the National Science Foundation under Grant CNS 01-30839. The opinions, findings and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the sponsors.
CHAPTER 1

INTRODUCTION

The proper functioning of many of our social, financial, and political structures today relies on correct identification of people. However, reliable and unique identification of people is a difficult problem. People typically use identification cards, usernames, or passwords to prove their identities. However, passwords can be forgotten, and identification cards can be lost or stolen. Biometric methods, which identify people based on physical or behavioral characteristics, are of interest because people cannot forget or lose their physical characteristics in the way that they can lose passwords or identity cards.

Each individual human iris has a unique texture of sufficient complexity that it can be used for identification \[2\]. In the past twenty years, researchers have developed several automated methods of identification relying on the human iris. A survey of some of these efforts is included in Chapter 2. Several commercial iris biometric systems exist \[3–6\]. The companies providing these systems claim high rates of recognition. The largest current deployment of iris recognition algorithms is in the United Arab Emirates \[7\]. The approximately twelve thousand people entering the country each day must submit to having their iris scanned and compared to a national watchlist.

Despite the amazing progress of iris biometric systems to date, the field of iris recognition still requires more research. The United States Customs and Borders
Protection processes approximately 1.13 million individuals daily [8], or three orders of magnitude more people than the United Arab Emirates handles. If the United States wished to implement an iris biometrics system at its ports and international airports, it would require higher confidence in the performance of the systems, and more flexible and faster acquisition environments.

My research investigates some of the sources of error in iris biometrics algorithms. Chapter 3 discusses how errors arise in an iris code at a bit-level. Chapter 4 details how dilation of the iris affects error rates. Finally, Chapter 5 gives conclusions and recommendations for future research.
CHAPTER 2

BACKGROUND

This chapter introduces basic terminology used in evaluating the performance of biometric systems. Since this work focuses on iris biometrics, an explanation of some parts of the eye is given. Next, a typical iris recognition algorithm is explained. We highlight some recent research in iris biometrics and conclude with an overview of iris biometrics evaluations. This chapter includes content that has previously been published in [9].

2.1 Performance of Biometric Systems

Biometrics is the process of identifying a person based on physiological or behavioral characteristics. Biometrics can be used in at least two different types of applications. In a verification scenario, a person claims a particular identity and the biometric system is used to verify or reject the claim. Verification is done by matching a biometric sample acquired at the time of the claim against the sample previously enrolled for the claimed identity. If the two samples match well enough, the identity claim is verified, and if the two samples do not match well enough, the claim is rejected. Thus there are four possible outcomes. A true accept occurs when the system accepts, or verifies, an identity claim, and the claim is true. A false accept occurs when the system accepts an identity claim, but the
claim is not true. A true reject occurs when the system rejects an identity claim and the claim is false. A false reject occurs when the system rejects an identity claim, but the claim is true. The two types of errors that can be made are a false accept and a false reject. Biometric performance in a verification scenario is often summarized in a receiver operating characteristic (ROC) curve. The ROC curve plots the verification rate on the Y axis and the false accept rate on the X axis, or, alternatively, the false reject rate on the Y axis and the false accept rate on the X axis. The equal-error rate (EER) is a single number often quoted from the ROC curve. The EER is where the false accept rate equals the false reject rate. The terms verification and authentication are often used interchangeably in this context.

In an identification scenario, a biometric sample is acquired without any associated identity claim. The task is to identify the unknown sample as matching one of a set of previously enrolled known samples. The set of enrolled samples is often called a gallery, and the unknown sample is often called a probe. The probe is matched against all of the entries in the gallery, and the closest match, assuming it is close enough, is used to identify the unknown sample. Similar to the verification scenario, there are four possible outcomes. A true positive occurs when the system says that an unknown sample matches a particular person in the gallery and the match is correct. A false positive occurs when the system says that an unknown sample matches a particular person in the gallery and the match is not correct. A true negative occurs when the system says that the sample does not match any of the entries in the gallery, and the sample in fact does not. A false negative occurs when the system says that the sample does not match any of the entries in the gallery, but the sample in fact does belong to someone in
the gallery. Performance in an identification scenario is often summarized in a cumulative match characteristic (CMC) curve. The CMC curve plots the percent correctly recognized on the Y axis and the cumulative rank considered as a correct match on the X axis. For a cumulative rank of 2, if the correct match occurs for either the first-ranked or the second-ranked entry in the gallery, then it is considered as correct recognition, and so on. The rank-one-recognition rate is a single number often quoted from the CMC curve. The terms identification and recognition are often used interchangeably in this context.

2.2 Eye Anatomy

Many different physical characteristics can be used in a biometrics system. This work focuses on iris biometrics. The iris is the “colored ring of tissue around the pupil through which light...enters the interior of the eye.” The iris’s function is to control the amount of light entering the eye. Two muscles in the iris, the dilator and the sphincter muscles, control the size of the pupil, and therefore, the amount of light passing through the pupil. Figure 2.1 shows an example image acquired by an LG 2200 commercial iris biometrics system at the University of Notre Dame. The sclera, a white region of connective tissue and blood vessels, surrounds the iris. A clear covering called the cornea covers the iris and the pupil. The pupil region generally appears darker than the iris. However, the pupil may have specular highlights, and cataracts can lighten the pupil. The iris typically has a rich pattern of furrows, ridges, and pigment spots. The surface of the iris is composed of two regions, the central pupillary zone and the outer ciliary zone. The collarette is the border between these two regions.

The minute details of the iris texture are believed to be determined randomly
Figure 2.1: Image 02463d1276 from the Iris Challenge Evaluation Dataset. Elements seen in a typical iris image are labeled here.
during the fetal development of the eye. They are also believed to be different between persons and between the left and right eye of the same person [11]. The color of the iris can change as the amount of pigment in the iris increases during childhood. Nevertheless, for most of a human’s lifespan, the appearance of the iris is relatively constant.

2.3 Early Research in Iris Biometrics

The idea of using the iris as a biometric is over 100 years old [12]. However, the idea of automating iris recognition is more recent. In 1987, Flom and Safir obtained a patent for an unimplemented conceptual design of an automated iris biometrics system [1]. Johnston [13] published a report in 1992 on an investigation of the feasibility of iris biometrics conducted at Los Alamos National Laboratory after Flom and Safir’s patent. Iris images were acquired for 650 persons, followed up over a 15-month period. The pattern of an individual iris was observed to be unchanged over the 15 months. The complexity of an iris image, including specular highlights and reflections, was noted. It was concluded that iris biometrics held potential for both verification and identification scenarios, but no experimental results were presented.

The most important work in the early history of iris biometrics is that of Daugman. Daugman’s 1994 patent [14] and early publications (e.g., [15]) described an operational iris recognition system in some detail. Iris biometrics as a field has developed with the concepts in Daugman’s approach becoming a standard reference model. Also, due to the Flom and Safir patent and the Daugman patent being held for some time by the same company, nearly all existing commercial iris biometric technology is based on Daugman’s work.
Daugman’s patent states that “the system acquires through a video camera a digitized image of an eye of the human to be identified.” A 2004 paper [2] said that image acquisition should use near-infrared illumination so that the illumination could be controlled, yet remain unintrusive to humans. Near-infrared illumination also helps reveal the detailed structure of heavily pigmented (dark) irises. Melanin pigment absorbs much of visible light, but reflects more of the longer wavelengths of light.

Systems built on Daugman’s concepts require subjects to position their eye within the camera’s field of view. The system assesses the focus of the image in real time by looking at the power in the middle and upper frequency bands of the 2-D Fourier spectrum. The algorithm seeks to maximize this spectral power by adjusting the focus of the system, or giving the subject audio feedback to adjust their position in front of the camera. More detail on the focusing procedure is given in the appendix of [2].

Given an image of the eye, the next step is to find the part of the image that corresponds to the iris. Daugman’s early work approximated the pupillary and limbic boundaries of the eye as circles. Thus, a boundary could be described with three parameters: the radius $r$, and the coordinates of the center of the circle, $x_0$ and $y_0$. He proposed an integro-differential operator for detecting the iris boundary by searching the parameter space. His operator is

$$\max(r, x_0, y_0) \left| G_\sigma(r) \ast \frac{\partial}{\partial r} \oint_{r, x_0, y_0} \frac{I(x, y)}{2\pi r} \, ds \right|$$

(2.1)

where $G_\sigma(r)$ is a smoothing function and $I(x, y)$ is the image of the eye.

All early research in iris segmentation assumed that the iris had a circular boundary. However, often the pupillary and limbic boundaries are not perfectly
circular. Recently, Daugman has studied alternative segmentation techniques to better model the iris boundaries [16]. Even when the inner and outer boundaries of the iris are found, some of the iris still may be occluded by eyelids or eyelashes.

Upon isolating the iris region, the next step is to describe the features of the iris in a way that facilitates comparison of irises. The first difficulty lies in the fact that not all images of an iris are the same size. The distance from the camera affects the size of the iris in the image. Also, changes in illumination can cause the iris to dilate or contract. This problem was addressed by mapping the extracted iris region into a normalized coordinate system. To accomplish this normalization, every location on the iris image was defined by two coordinates, (i) an angle between 0 and 360 degrees, and (ii) a radial coordinate that ranges between 0 and 1 regardless of the overall size of the image. This normalization assumes that the iris compresses or stretches linearly in the radial direction when the pupil dilates or contracts, respectively. A paper by Wyatt [17] explains that this assumption is a good approximation, but it does not perfectly match the actual deformation of an iris.

The normalized iris image can be displayed as a rectangular image, with the radial coordinate on the vertical axis, and the angular coordinate on the horizontal axis. The left side of the normalized image marks 0 degrees on the iris image, and the right side marks 360 degrees. The division between 0 and 360 degrees is somewhat arbitrary, because a simple tilt of the head can affect the angular coordinate. Daugman accounts for this rotation later, in the matching technique.

Directly comparing the pixel intensity of two different iris images could be prone to error because of differences in lighting between two different images. Daugman uses convolution with 2-dimensional Gabor filters to extract the texture
from the normalized iris image. In his system, the filters are “multiplied by the raw image pixel data and integrated over their domain of support to generate coefficients which describe, extract, and encode image texture information.” [14]

After the texture in the image is analyzed and represented, it is matched against the stored representation of other irises. If iris recognition were to be implemented on a large scale, the comparison between two images would have to be very fast. Thus, Daugman chose to quantize each filter’s phase response into a pair of bits in the texture representation. Each complex coefficient was transformed into a two-bit code: the first bit was equal to 1 if the real part of the coefficient was positive, and the second bit was equal to 1 if the imaginary part of the coefficient was positive. Thus after analyzing the texture of the image using the Gabor filters, the information from the iris image was summarized in a 256 byte (2048 bit) binary code. The resulting binary “iris codes” can be compared efficiently using bitwise operations.\(^1\)

Daugman uses a metric called the normalized Hamming distance, which measures the fraction of bits for which two iris codes disagree.\(^2\) A low normalized Hamming distance implies strong similarity of the iris codes. If parts of the irises are occluded, the normalized Hamming distance is the fraction of bits that disagree in the areas that are not occluded on either image. To account for rotation, comparison between a pair of images involves computing the normalized Hamming distance for several different orientations that correspond to circular permutations of the code in the angular coordinate. The minimum computed normalized

\[^1\text{The term, “iris code” was used by Daugman in his 1993 paper. We use this term to refer to any binary representation of iris texture that is similar to Daugman’s representation.}\]

\[^2\text{The Hamming distance is the number of bits that disagree. The normalized Hamming distance is the fraction of bits that disagree. Since normalized Hamming distance is used so frequently, many papers simply mention “Hamming distance” when referring to the normalized Hamming distance. We also follow this trend in subsequent sections of this work.}\]
Hamming distance is assumed to correspond to the correct alignment of the two images.

An iris biometrics system following Daugman’s general approach could be described in four basic steps (1) acquisition, (2) segmentation, (3) texture analysis, and (4) matching. These basic modules are depicted in Figure 2.2. The goal of image acquisition is to acquire an image that has sufficient quality to support reliable biometrics processing. The goal of segmentation is to isolate the region that represents the iris. The goal of texture analysis is to derive a representation of the iris texture that can be used to match two irises. The goal of matching is to evaluate the similarity of two iris representations. The distinctive essence of Daugman’s approach lies in conceiving the representation of the iris texture to be a binary code obtained by quantizing the phase response of a texture filter. This representation has several inherent advantages. Among these are the speed of matching through the normalized Hamming distance, easy handling of rotation of the iris, and an interpretation of the matching as the result of a statistical test of independence [15].

Wildes [18] describes an iris biometrics system developed at Sarnoff Labs that uses a very different technical approach from that of Daugman. Whereas Daugman’s system acquires the image using “an LED-based point light source in conjunction with a standard video camera,” the Wildes system uses “a diffuse source and polarization in conjunction with a low light level camera.” When localizing the iris boundary, Daugman’s approach looks for a maximum in an integro-differential operator that responds to circular boundary. By contrast, Wildes’ approach involves computing a binary edge map followed by a Hough transform to detect circles. The Hough transform considers a set of edge points and finds the circle
that best fits the most edge points. In matching two irises, Daugman’s approach involves computation of the normalized Hamming distance between iris codes, whereas Wildes applies a Laplacian of Gaussian filter at multiple scales to produce a template and computes the normalized correlation as a similarity measure. Wildes briefly describes [18] the results of two experimental evaluations of the approach, involving images from several hundreds of irises. This paper demonstrates that multiple distinct technical approaches exist for each of the main modules of an iris biometrics system.

There are advantages and disadvantages to both Daugman’s and Wildes’ designs. Daugman’s acquisition system is simpler than Wildes’ system, but Wildes’ system has a less-intrusive light source designed to eliminate specular reflections. For segmentation, Wildes’ approach is expected to be more stable to noise pertur-
bations; however, it makes less use of available data, due to binary edge abstraction, and therefore might be less sensitive to some details. Also, Wildes’ approach encompassed eyelid detection and localization. For matching, the Wildes approach made use of more of the available data, by not binarizing the bandpass filtered result, and hence might be capable of finer distinctions; however, it yields a less compact representation. Furthermore, the Wildes method used a data-driven approach to image registration to align two instances to be compared, which might better respond to the real geometric deformations between the instances, but comes at increased computation.

In 1996 and 1998, Wildes et al. filed two patents [19] which described their automated segmentation method, the normalized spatial correlation for matching, and an acquisition system allowing a user to self-position his or her eye. A more recent book chapter by Wildes [20] largely follows the treatment in his earlier paper [18]. However, some of the technical details of the system are updated and there is discussion of some experimental evaluations of iris biometrics done since the earlier paper. Earlier and less detailed descriptions of the system appear in [21, 22].

2.4 Recent Research in Iris Biometrics

Initially, image acquisition for iris biometrics required extremely controlled conditions. In order to extend the possible applications of iris biometrics, recent work has attempted to reduce the constraints required for image acquisition. The least-constrained system to date is described by Matey et al. [23]. They aim to acquire iris images as a person walks at normal speed through an access control point such as those common at airports. The image acquisition is “based on
high-resolution cameras, video synchronized strobed illumination, and specularity based image segmentation.” The system aims to be able to capture useful images in a volume of space 20 cm wide and 10 cm deep, at a distance of approximately 3 meters. The height of the capture volume is nominally 20 cm, but can be increased by using additional cameras. The envisioned scenario is that “subjects are moderately cooperative; they look forward and do not engage in behavior intended to prevent iris image acquisition, such as squinting or looking away from the acquisition camera. Subjects may be required to remove sunglasses, depending on the optical density of those sunglasses. Most subjects should be able to wear normal eyeglasses or contact lenses.” Experiments were performed with images from 119 Sarnoff employees. Results were that “the overall recognition rate (total number of successful recognitions divided by the total number of attempts) for all subjects was 78%.” The paper concludes, “the Iris on the Move system is the first, and at this time the only, system that can capture iris images of recognition quality from subjects walking at a normal pace through a minimally confining portal.” An example of such a portal is shown in Figure 2.3.

Algorithms enabling the processing of off-angle images may further reduce acquisition constraints in iris biometric systems. A recent trend in segmentation aims at dealing with off-angle images. Dorairaj et al. [24] assume that an initial estimate of the angle of rotation is available, and then use Daugman’s integro-differential operator as an objective function to refine the estimate. Once the angle is estimated, they apply a projective transformation to rotate the off-angle image into a frontal view image. Li [25] fits an ellipse to the pupil boundary and then uses rotation and scaling to transform the off-angle image so that the boundary is circular. It is shown that the proposed calibration step can improve the separation
Figure 2.3: This “Iris on the Move” portal acquires an iris image as subjects walk through a portal at normal walking speed. The portal itself contains infrared lights to illuminate the subject. Three high-zoom video cameras in the far cabinet take video of the subject.
between intra-class and inter-class differences that is achieved by a Daugman-like algorithm. Abhyankar et al. [26] propose using active shape models for finding the elliptical iris boundaries of off-angle images. A recent paper by Daugman [16] presents alternative methods of segmentation based on active contours, a way to transform an off-angle iris image into a more frontal view, and a description of new score normalization scheme to use when computing Hamming distance that would account for the total amount of unmasked data available in the comparison.

In another effort to improve performance, there has been an army of literature presenting different modifications on how to extract texture information from the segmented iris region. A number of these studies involved implementing a different filter and comparing the filter to the two-dimensional Gabor filter proposed by Daugman. A detailed comparison of seven different filter types is given by Thornton et al. [27]. They consider the Haar wavelet, Daubechies wavelet, order three, Coiflet wavelet, order one, Symlet wavelet, order two, Biorthogonal wavelet, orders two and two, circular symmetric filters, and Gabor wavelets. They applied a single bandpass filter of each type and determined that the Gabor wavelet gave the best equal error rate. They then tune the parameters of the Gabor filter to optimize performance. They report that “Although we conclude that Gabor wavelets are the most discriminative bandpass filters for iris patterns among the candidates we considered, we note that the performance of the Gabor wavelet seems to be highly dependent upon the parameters that determine its specific form.”

Once an iris image is acquired and segmented, and relevant texture information extracted, the resulting feature vector or iris code is compared with enrolled iris codes. A key concept of Daugman’s approach to iris biometrics is the linking of
the Hamming distance to a confidence limit for a match decision. The texture computations going into the iris code are not all statistically independent of each other. However, given the Hamming distance distributions for a large number of true non-matches, the distribution can be fit with a binomial curve to find the effective number of degrees of freedom. The effective number of degrees of freedom then allows the calculation of a confidence limit associated with any recognition decision.

Daugman and Downing [11] describe an experiment to determine the statistical variability of iris patterns. Their experiment evaluates 2.3 million comparisons between different iris pairs. The mean Hamming distance between two different irises is 0.499, with a standard deviation of 0.032. This distribution closely follows a binomial distribution with 244 degrees of freedom. The distribution of Hamming distances for the comparisons between the left and right irises of the same person is found to be not statistically significantly different from the distribution of comparisons between different persons. Daugman’s 2003 paper [28] presents similar results as [11], but with a larger dataset of 9.1 million iris code matches. This number of matches could derive from matching each of a set of just over 3,000 iris images against all others.

Bolle et al. [29] also study the theory behind the performance rates of an iris biometric system. Following on concepts developed by Daugman, they consider the probability of bit values in an iris code and the Hamming distance between iris codes to develop an analytical model of the false reject rate and false accept rate as a function of the probability \( p \) if a bit in the iris code being ‘flipped’ due to noise. The model predicts that “the iris FAR performance is relatively stable and is not affected by \( p \).” They indicate that the “FAR performance predicted by the
foregoing analytical model is in excellent agreement with the empirical numbers reported by Daugman.” In contrast, their theoretical FRR accuracy did not meet the empirical numbers presented by Daugman. This observation motivated the detailed study, presented in chapter 3, of the consistency of individual bits in an iris code.

2.5 Iris Biometrics Evaluations

There have been few publicly-accessible, large-scale evaluations of iris biometrics technology. There are a number of papers that compare a proposed algorithm to “Daugman’s algorithm.” However, this generally means a comparison to a particular re-implementation of Daugman’s algorithm as described in his earliest publications. Thus the “Daugman’s algorithm” used for comparison purposes in two different papers may not be exactly the same algorithm and may not give the same performance on the same dataset. Also, it is very likely that neither would match the performance of Daugman’s own most recent algorithms. There are also, as mentioned earlier, many research papers that compare different texture filters in a relatively controlled manner. However, the datasets used in such experiments have generally been small relative to what is needed to draw conclusions about statistical significance of observed differences, and often the experimental structure confounds issues of image segmentation and texture analysis.

As one example of a research-level comparison of algorithms, Vatsa et al. [30] implemented and compared four algorithms. They looked at Daugman’s method [31]; Ma’s algorithm which uses circular symmetry filters to capture local texture information and create a feature vector [32]; Sanchez-Avila’s algorithm based on zero-crossings [33]; and Tisse’s algorithm which uses emergent frequency and
A comparison of the four algorithms, using the CASIA 1 database, showed that Daugman’s algorithm performed the best with 99.90% accuracy, then Ma’s algorithm with 98.00%, Avila’s with 97.89%, and Tisse’s algorithm with 89.37%.

A widely-publicized evaluation of biometric technology done by the International Biometric Group in 2004 and 2005 [35] had a specific and limited focus: “The scenario test evaluated enrollment and matching software from Iridian and acquisition devices from LG, OKI, and Panasonic” [35]. Iris samples were acquired from 1,224 individuals, 458 of whom participated in data acquisition again at a second session several weeks after the first. The report gives failure to enroll (FTE) rates for the three systems evaluated, where “FTE was defined as the proportion of enrollment transactions in which zero [irises] were enrolled. Enrollment of one or both [irises] was considered to be a successful enrollment.” The report also gives false match rates (FMR) and false non-match rates (FNMR) for enrollment with one system and recognition with the same or another system. One conclusion is that “cross-device equal error rates, while higher than intra-device error rates, were robust.” With respect to the errors encountered in the evaluation, it is reported that “errors were not distributed evenly across test subjects. Certain test subjects were more prone than others to FTA, FTE, and genuine matching errors such as FNMR.” It is also reported that “one test subject was unable to enroll any [irises] whatsoever.” Some of these high-level patterns in the overall results may be representative of what would happen in general application of iris biometrics.

Authenticorp released a report in 2007 [36] that evaluates three commercial iris recognition systems in the context of three main questions: (1) What are the
realistic error rates and transaction times for various commercial iris recognition products? (2) Are ISO-standard iris images interchangeable (interoperable) between products? (3) What is the influence of off-axis user presentation on the ability of iris recognition products to acquire and recognize iris images? The experimental dataset for this report included about 29,000 images from over 250 persons. The report includes a small, controlled off-axis experiment in addition to the main, large scenario evaluation, and notes that “the current generation of iris recognition products is designed for operational scenarios where the eyes are placed in an optimal position relative to the product’s camera to obtain ideal on-axis eye alignment.” The data collection for the experiment includes a time lapse of up to six weeks, and the report finds that this level of time lapse does not have a measurable influence on performance. The report also notes that, across the products tested, there is a tradeoff between speed and accuracy, with higher accuracy requiring longer transaction times.

A different sort of iris technology program, the Iris Challenge Evaluation (ICE), was conducted under the auspices of the National Institute of Standards and Technology (NIST) [37]: “The ICE 2005 is a technology development project for iris recognition. The ICE 2006 is the first large-scale, open, independent technology evaluation for iris recognition. The primary goals of the ICE projects are to promote the development and advancement of iris recognition technology and assess its state-of-the-art capability. The ICE projects are open to academia, industry, and research institutes.” The initial report from the ICE 2006 evaluation is now available [38], as well as results from ICE 2005 [37].

One way in which the ICE differs from other programs is that it makes source code and data sets for iris biometrics available to the research community. As
part of ICE, source code for a baseline Daugman-like iris biometrics system and a dataset of approximately 3,000 iris images had been distributed to over 40 research groups by early 2006. The ICE 2005 results that were presented in early 2006 compared self-reported results from nine different research groups [37]. Participants included groups from industry and from academia, and from several different countries. The groups that participated in ICE 2005 did not all submit descriptions of their algorithms, but presentations by three of the groups, Cambridge University, Tohoku University, and Iritech, Inc., are online at http://iris.nist.gov/ice/presentations.htm.

Iris images for the ICE program were acquired using an LG 2200 system, with the ability to save raw images that would not ordinarily pass the built-in quality checks. Thus this evaluation seeks to investigate performance using images of less-than-ideal quality. The ICE 2006 evaluation was based on 59,558 images from 240 subjects, with a time lapse of one year for some data. A large difference in execution time was observed for the iris biometrics systems participating in ICE 2006, with a factor of 50 difference in speed between the three systems whose performance is included in the report. The ICE 2006 report is combined with the Face Recognition Vendor Test (FRVT) 2006 report, and includes face and iris results for the same set of people [38].

In [39], Newton and Phillips compare the findings of the evaluations by NIST, Authenticorp, and the International Biometrics Group [35–37]. They note that all three tests "produced consistent results and demonstrate repeatability." The evaluations may have produced similar results because most of the algorithms were based on Daugman’s work, and Daugman-based algorithms dominate the market. The best performers in all three evaluations achieved a false non-match
rate of about 0.01 at a false match rate of 0.001.
CHAPTER 3

FRAGILE BITS

Many researchers have developed iris biometric algorithms following the typical steps outlined by Daugman. They apply a texture filter to an image of the iris and extract a representation of the texture, the iris code. Each bit of the iris code indicates whether a given texture filter applied at a given point on the iris image has a negative or non-negative result. Even though these typical steps in Daugman’s algorithm have been implemented and used in numerous systems, the theoretical explanation for the false reject rates of these systems had yet to be explained satisfactorily prior to this research.

As mentioned in the previous chapter, Bolle et al. [29] attempted to model the theoretical lower bounds for the false accept rate and false reject rate of an iris biometric system. They found that “the reported empirical FRR performance degradation is significantly more stable with respect to the system threshold variation than predicted by the theory. This implies that the invariant bits in the iris code representation are dramatically robust to the imaging noise.” As a possible explanation for the difference between the empirical and theoretical results, they suggested “...that perhaps not all bits are equally likely to flip, that there are some particularly ‘fragile’ bits.” For a given iris, a bit in its iris code is “fragile” (i.e. inconsistent) if there is any substantial probability of it ending up a 0 for some images of the iris and a 1 for other images of the iris. To investigate the existence
of fragile bits, we compared iris codes from multiple images of the same eye. Some research presented in this chapter has previously been published in [40].

3.1 Data and Software

The data used for this experiment was collected at the University of Notre Dame and provided to the National Institute of Standards and Technology (NIST) for use in the Iris Challenge Evaluation (ICE) [37]. All images were acquired using an LG2200 EOU iris imaging system [4]. One unusual aspect of images taken with the LG 2200 is that the intensity values are automatically contrast-stretched by the LG 2200 to use 171 gray levels between 0 and 255. A histogram of the gray values in the image used in Figure 2.1 is given in Figure 3.1.

The subset of the ICE dataset used for the experiments in this paper contains over a hundred different images for each of several irises. No more than six images were acquired for a particular iris of a particular subject in any given week. To test the fragility of individual bits in an iris code, we selected images that were mostly unoccluded by eyelids or lashes. Example images from this dataset are shown in Figures 3.2 through 3.5. We selected a subset of the ICE data that contains 24 subjects, with between 15 and 118 images of the left eye of each subject, for a total dataset of 1226 left iris images. This dataset contains 4 Asian and 20 Caucasian subjects. Ten of the subjects are female and 14 are male.

We used software similar to the open source software, IrisBEE [37], to create the iris codes. This software uses one dimensional log-Gabor wavelets to create a 240x20x2-bit iris code. The software we used contains improvements to the segmentation as described in [41]. Our software automatically found correct inner and outer boundaries of the iris in all 1226 images selected for the study (verified
by visual inspection). The images were selected to minimize the effects of segmentation errors; however, some minor imperfections in the segmentations could not be avoided. One difficulty in iris segmentation lies in the fact that iris boundaries are not exactly circular. Therefore, even the best circle caught part of the sclera or eyelid above and below some of the irises (Figure 3.2). Also, there are specular highlights on irises of some of the images (Figure 3.3). To deal with these segmentation difficulties, we later consider how masking parts of the iris affects our results.

In cases where the segmentation software detected occlusion by eyelids, parts of the iris region were masked (Figure 3.4). If part of the iris code was masked in even one image of a subject, the bits from that part of the iris code were left out of all computations in our experiments on that subject. Some of the images have sharp focus, but a few images are less well focused (Figure 3.5). Our experiments show that some bits are consistent across all images, even when a few poorly-focused images are included in the experiment. This result is consistent
Figure 3.2: Example image 04233d1145 from our dataset. This image has no occlusion by eyelids or eyelashes. However, part of the sclera is included in the segmentation of this iris region. Also, some specular highlight is evident in the iris region near the lower eyelid.
Figure 3.3: Example image 02463d1268 from our dataset. This iris contains specular highlights near the bottom of the iris, and near the pupil.
Figure 3.4: In this image, parts of the top and bottom of the iris were masked before the iris code was computed. The corresponding parts of the iris code of all images of this iris were dropped from the experimental analysis. This is image 04239d1060 from our dataset.
Figure 3.5: Despite including some poorly focused images in the dataset, on average 15.96% of bits in the iris codes were perfectly consistent. This is image 04336d692 from our dataset.
with Bolle’s assertion that “the invariant bits in the iris code representation are dramatically robust to the imaging noise.” [29]

Two different iris codes from the same eye are not necessarily in the same orientation when acquired from the raw images, and therefore may have to be rotated in order about the optical axis to be aligned correctly. We wrote a program that would rotate the iris codes so that all codes from the same eye could be aligned and compared for consistency. We used the first iris code as a reference, and aligned each subsequent code to the first code.

From this set of aligned binary iris code images we created a 240x20x2 matrix, where each entry in the matrix contains the average value for the corresponding bit in the iris code. Half of these entries corresponded to the real part of the output from the texture filtering process, and half of the entries corresponded to the imaginary part. In order to see patterns in the output, we divided each matrix into two 240x20x1 matrices where one matrix represented the real bits of the iris code, and the other represented the imaginary bits.

3.2 Existence of Fragile Bits

All subjects had three different regions apparent in their iris codes: areas consistently equal to 0, areas consistently equal to 1, and inconsistent areas. The inconsistent areas tended to occur at the boundaries between regions of zeros and regions of ones. Examples of the inconsistent regions in the iris codes are shown in Figure 3.6. If a bit was equal to one the majority of the time, but was equal to zero 30% of time, then we say that the bit “flipped” in 30% of the iris codes. If a bit was zero the majority of the time, but one for 30% of the time, we also say that the bit “flipped” in 30% of the iris codes. In this figure, the black regions
correspond to bits that were flipped in at least 30% of the iris codes. If we are less strict in our definition of what constitutes an inconsistent bit, then there will be a greater number of inconsistent bits. Figure 3.7 shows that if we consider any bit that flips more than 20% of the time to be fragile, then there will be many fragile bits. If we consider any bit that flips more than 40% of the time, then there will be fewer fragile bits. In our study we found that, on average across our set of images, 15.96% of the bits in an iris code were perfectly consistent; that is, 15.96% of the unmasked bits were always equal to 1 or always equal to 0, for all iris codes for an iris. The subject with the smallest fraction had 4.74% of the bits perfectly consistent, and the subject with the largest fraction had 33.2% of the bits perfectly consistent. The sample standard deviation across the 24 irises was 6.39%.

It is natural to wonder why one subject would have more perfectly consistent bits than another. This phenomenon can be explained by considering the number of images available of each subject. We computed Pearson’s correlation between the number of images we had for a subject and the percent of perfectly consistent bits. Pearson’s correlation reflects the amount of linear relationship between two variables, and ranges between -1 (perfect negative correlation) and 1 (perfect positive correlation). The correlation coefficient between number of images and percent of perfectly consistent bits is -0.5369. We applied a two-tailed Student’s t test to test the null hypothesis of no correlation against the alternative that there is a nonzero correlation, and found evidence that there is a statistically significant correlation (p-value 0.0068). This result makes sense because as we acquire more images of an iris, there are more opportunities to acquire outliers in any given bit in the iris code.
Figure 3.6: Black areas in each rectangle are inconsistent parts of the iris code, and white areas are consistent. Each rectangle represents data from the iris codes of at least 15 different images of the same eye. Two rectangles are shown for each subject; one rectangle shows the real bits in the iris code, and the next rectangle shows the imaginary bits. Light gray regions are masked regions. The numbers on the side are the subject numbers associated with images in the ICE dataset.
In addition to considering perfectly consistent bits, we also looked at the average fragility of the unmasked bits of a subject. For each bit, the percentage of the images in which the bit flips must lie between 0% and 50%, or between 0 and 0.5. We found the frequency that each bit flipped and took the average across all bits for a subject. This average fragility for each subject is not correlated with the number of images we had for the subject. The correlation coefficient between these factors is only -0.1730, which is not statistically significant (p-value 0.4188).

Fragile bits show up in about equal amounts in both males and females. We use average fragility in making comparisons between genders because average fragility did not appear to be dependent on number of images. Figure 3.8 shows the average fragility of the 24 subjects when divided into groups of males and females. There is no apparent difference between the two genders. Figure 3.9 shows average fragility versus race. Our dataset suggests that Asian subjects had more fragile bits than
Caucasian subjects. However, we only have four Asian subjects in our dataset, so it is quite possible that this trend may not be representative of the general population.

The spatial pattern apparent in these consistent and inconsistent regions likely comes as a result of how the iris image is processed to generate the code. A Fourier transform is applied to the iris image, and then the values are multiplied by a log-Gabor filter. Next an inverse Fourier transform is applied, yielding 4800 complex filter responses in the spatial domain. Rather than storing the complex numbers as the iris codes, the phase of each complex number is quantized to one of the four quadrants. A complex number in the first quadrant of the complex plane is mapped to the value, 11; the second quadrant, 01; the third quadrant, 00; and the fourth quadrant, 10. If a region of the iris image were associated with a complex number near the negative imaginary axis, small changes in the complex number could make that region of the iris map to a 00 some of the time, and a 10 at other times. In this case, we would expect the real bit to be fragile because half of the time that bit would be a 0, and half of the time that bit would be a 1. Furthermore, we would expect the imaginary bit to be consistent, because the imaginary bit is equal to zero no matter whether the complex number is in the third or fourth quadrants.

Figure 3.10 shows an example of the distribution of 54 complex numbers from 54 different images of the same iris. Each of these 54 complex numbers is associated with the same location in the iris code. In particular, this location on the iris code had a highly inconsistent imaginary bit and a highly consistent real bit. As we expected, the complex numbers associated with these two bits of the iris lay close to the positive x-axis.
Figure 3.8: There is no difference in the average fragility of iris code bits between the two different genders.
Figure 3.9: The four Asian subjects in our dataset tended to have more fragile bits in their iris codes, but with only four subjects, we cannot conclude that this is a general trend.
Figure 3.10: These 54 complex numbers, each from the same region on 54 different images of the same subject’s eye, all correspond to the same location on the iris code. Each complex number is mapped to two bits. This particular part of the iris code had a highly consistent real bit, and a highly inconsistent imaginary bit.
3.3 Outliers

Upon first inspection, it appeared to us that many of the distributions of complex numbers corresponding to particular points on the iris images followed a two-dimensional Gaussian distribution. In fact, upon randomly selecting one such set of complex numbers and testing for normality, there was no statistically significant evidence that the distribution was not normal. However, when we tested another such set of complex numbers, it did not have a Gaussian distribution. Either the distributions were truly non-Gaussian, or there were sufficient outliers to provide statistical evidence for non-normality.

In order to better understand our results, we studied what might be causing the outliers in the sets of complex numbers used to create the iris codes for one of our subjects. Using multiple images of a subject, we can get multiple different measurements that each correspond to the same position. We have 60 images from subject 02463, and thus we have sets of 60 complex numbers each, and we have 4800 such data sets from the different positions of the iris.

We began our analysis by finding the average spread of the data for all the data sets. The data sets are not all centered at the same position. In order to estimate the standard deviation for all the complex numbers, the data sets first need to be shifted to all be centered at the same location. For each data set, the points were translated so that the average of each set was centered at 0+0j. Next, the standard deviation for all 4800x60 numbers was estimated. For subject 02463, the sample standard deviation was 0.00815.

We manually examined all complex numbers that fell more than 4 sample standard deviations away from the mean in their data set. We marked the 1478 positions on the original iris images that yielded such outliers. The majority of
the outliers were due to spots on the iris that were lighter than the surrounding region. These light spots seem to be specular highlights. Some of the brightest specular highlights have a dark border surrounding them. The dark borders also cause outliers in the data. While most outliers occurred on or near light regions of the iris image, some outliers appeared on shadows in the image that were not near any highlights. 85.9% (1270 out of 1478) of the outliers for subject 02463 were due to light regions on the eye. 11.8% (174 of the 1478) of the outliers were due to some dark imaging artifacts around specular highlights, and 2.3% of outliers (34 of the 1478) were due to shadows on the image that seemed unrelated to any highlights.

Figure 3.11 displays a portion of an example iris image showing outliers caused by specular highlights and dark borders. The positions of the outliers are marked with triangles. Complex numbers that fell more than 4 standard deviations away from the average are marked with red, upward-pointing triangles. Complex numbers that were between 3 and 4 standard deviations from the average are marked with blue, downward-pointing triangles. This particular example appears near the lower eyelid. However, dark borders also commonly appear near the upper eyelid, near the tear duct, and around the main specular highlight near the pupil.

Because of the way the images are processed, an abnormality in one part of a row of the normalized iris image causes a string of outliers along that row. In applying the log-Gabor filters, a Fourier transform is applied to one row of the normalized iris image. Next, the values are multiplied by the filter, and then the inverse Fourier transform is applied. Since the Fourier transform is done on the entire row, a highlight in one part of the row can affect numbers in other parts of the row. Thus, there are outliers to the right and left of the specular highlight as
Figure 3.11: Close-up view of image 02463d1323 showing specular highlights. Triangles mark the positions on the iris that resulted in outliers after the 1D log-Gabor filter was applied. A zoomed-in portion of the corresponding normalized iris is also shown.
well as immediately on the highlight.

There are multiple smooth surfaces in the eye that can reflect light and thus create the highlights which are causing the outliers. Light can be reflected off of the 1) outer and 2) inner surfaces of the cornea, and off the 3) outer and 4) inner surfaces of the lens of the eye. These four types of reflections are called Purkinje images [42]. Since the lens is located behind the iris, the light regions seen on the iris are probably not reflections from the lens, but they could be reflections from the cornea. In some instances, there appears to be faint specular reflections close to and in line with brighter specular reflections. Figure 3.12 shows examples of these types of highlights.

Since the majority of outliers for this subject came from highlights, it is natural to inquire whether all the light regions of the images manifest outliers. We found a highlight in one image (02463d1249) that was not marked as an outlier when we looked at outliers greater than three standard deviations from average. However, when we looked at where the highlight fell with respect to the 4800 sampling points on the iris, we found that the sample points fell on the edges of the highlight, rather than in its center. Furthermore, the complex numbers around this highlight still were greater than two standard deviations from average. Figure 3.13 illustrates this situation.

The remaining outliers were less easy to explain. These outliers showed up on dark regions in the images. We surmise that these outliers are due to shadows on the iris, or perhaps features that are small enough that they only show up in the best-focused images. Figure 3.14 shows an example of this type of outlier.
Figure 3.12: Faint reflections often appear close to brighter reflections. This image contains (A) bright reflections on the eyelid, (B) a nearby, faint reflection on the iris, (C) a bright reflection in the pupil, and (D) a nearby, faint reflection on the iris. Both of the faint reflections marked in this image (B and D) caused outliers in the complex data. (Image 02463d1329)
Figure 3.13: The positions of the sampling points around a specular highlight affect how extreme an outlier is produced. Green circles represent the sampling points. Red, upward-pointing triangles are outliers greater than 4 standard deviations from average. Blue, down-ward pointing triangles are outliers greater than 3 standard deviations from average. Yellow, right-pointing triangles are outliers greater than 2 standard deviations from average. A sample point in the center of a highlight corresponds with an extreme outlier (A). Sample points falling on the edge of a specular highlight cause less-extreme outliers (B). (Image 02463d1249)
Figure 3.14: Some outliers showed up on dark regions of the iris image. Small dots show the sample points on the iris used to create the normalized image. (02463d1276)
3.4 Inner vs. Outer Radial Bands of the Iris

Several researchers [34, 43, 44] have suggested using only the inner part of the iris for recognition. Du et al. [45] suggested that “a more distinguishable and individually unique signal is found in the inner rings of the iris,” and that “as one traverses to the limbic boundary of the iris, the pattern becomes less defined and ultimately less useful in determining identity.” Since the inner part of the iris is less likely to be occluded, we suspect that it is the occlusion, rather than the lack of texture, that might make outer bands less valuable. To test this idea, we graphed the percent of fragile bits that occurred at each radial band in the iris.

Using a consistency threshold of 40%, we calculated the percent of bits that were fragile for each row of the “unwrapped” iris image (corresponding to a radial band), for all 24 iris images. The results are shown in Figure 3.15.

Despite our efforts to obtain unoccluded images for the dataset, we suspected that specular highlights close to the eyelids could be affecting our measurements of the consistency of the different bands of the iris code. Therefore, we applied a mask to the data to disregard the upper and lower quarters of the iris, and then we looked at the percent of fragile bits for each row of the iris code in the unmasked quarters of the iris (left and right sides of the iris). The mask used is shown in Figure 3.16, and the result is shown in Figure 3.17.

We also performed the same test using fragility thresholds of 30% and 20%. In all cases, masking the top and bottom quarters improved the consistency of the outer rows. Clearly, the quality of the segmentation affects the value of the outer rows of the iris.
Figure 3.15: This figure shows the percent of fragile bits in each row of the iris code. Rows in the middle of the iris code (rows 5 through 12) are the most consistent.
Figure 3.16: Mask used for evaluating the consistency of the inner versus outer bands of the iris. When we consider only the right and left sides of the iris, and disregard the top and bottom sections, which had the most specular highlights and occlusion, there is not as much spread in the fraction of fragile bits across the rows. There still seems to be a high percentage of fragile bits in rows 1 and 2 of the iris code; however, the general trend shows that all rows of the iris code have a high amount of consistent information.
Figure 3.17: When the regions most affected by specular highlights and occlusion are masked, the consistency of the outer rows of the iris code improves. (The black horizontal line is drawn to aid in comparison between outer and inner rows.)
3.5 Effect of Granularity of Iris Alignment on Consistency

Since we have experimentally shown the existence of fragile bits, it is reasonable to ask how small modifications in the iris recognition algorithm might affect the consistency of the bits in the iris code. One important requirement of an iris recognition algorithm is that the algorithm must be rotation-invariant. That is, a small tilt of the head should not cause the recognition algorithm to fail. Daugman [2] suggested that to achieve rotation invariance, an enrolled iris code could be compared to several different shifts of the probe iris code, and the shift that yielded the smallest Hamming distance could be taken as the correct orientation of the probe image.

Our software shifts the iris code by two angular steps at a time by default, which is equivalent to rotating the original iris image by 3 degrees. In order to be as thorough as possible, we used shifts of the iris code as small as one step (1.5 degrees) when testing for the existence of fragile bits. We subsequently needed to check whether changing the shift size would affect the consistency of the bits.

We tested for the existence of fragile bits using shift sizes as small as 1.5 degrees and as large as 7.5 degrees. The general trend showed that using a finer resolution tended to yield a larger number of consistent bits. However, if we extrapolate on the data, it is clear that the y-intercept of the graph is greater than zero, implying that fragile bits exist regardless of shift resolution. Therefore, even if we could have an infinitely small shift size, there would still be fragile bits. A graph showing the effect of shift size is shown in Figure 3.18. Each line in the graph represents one of our 24 subjects. As in the previous section, we used a threshold of 40% to determine which bits were fragile.
Figure 3.18: In matching a pair of iris codes, multiple different orientations of the probe iris code are considered. Allowing for smaller possible rotations decreases the percent of fragile bits.
3.6 Effect of Filter on Consistency

Our software uses one dimensional log-Gabor wavelets in creating the iris code, but there are many different types of filters that can be used in an iris recognition algorithm [27]. We explored the effect of a different filter on the fragile bit patterns in the iris code.

In order to try a different filter implementation, we obtained an open-source iris recognition system, OSIRIS [46]. For segmentation, we used the centers and radii of the irises and pupils generated by our own software. OSIRIS does not currently mask eyelids and eyelashes. However, the images in this dataset were selected because they did not contain much occlusion due to eyelids. In addition, in creating the iris code, we chose to ignore the 20% of the iris closest to the limbic boundary, instead using sample points closer to the pupil so we would avoid the regions affected by the eyelid.

To create the iris code, OSIRIS first processes the segmented iris image to yield a normalized image that is 64 by 512. Next, OSIRIS applies two-dimensional Gabor filters to selected sample points in this image. The filter bank that comes with OSIRIS contains 6 filters, three for real parts and three for imaginary parts. The filters come in three different sizes: 9 by 15, 9 by 27, and 9 by 51. We selected 1280 sample points in a 16 by 80 grid pattern on the 80% of the normalized image closest to the pupil. Therefore, the resultant iris code is 80x16x6. Once again, we graphed which parts of the iris code were fragile. As in Figure 3.6, the black regions correspond to bits that were flipped in at least 30% of the iris codes. Since OSIRIS uses 3 filters, and each filter has a real and imaginary part, we have 6 rectangles to display the fragile bits for each iris. The fragile bits patterns for the imaginary parts of the largest filter are shown in Figure 3.19 for a few subjects.
Clearly the phenomenon of fragile bits is apparent even when a different filter is used.

In section 3.2, we remarked that inconsistent areas in our iris codes tended to occur at the boundaries between regions of zeros and regions of ones. The same pattern is apparent in many of the iris codes for the large filter of OSIRIS. One example of this phenomenon is illustrated in Figure 3.20. In this figure, areas consistently equal to 1 are marked in red and areas consistently equal to zero are marked in blue. The yellow and green areas represent inconsistent areas, that are 1’s in some of the iris codes for the subject, and 0’s in other of the iris codes for the subject. The yellow regions clearly are sandwiched between an area of red on one side and blue on the other. Interestingly, this trend is not apparent when a smaller filter is used. We suspect that this trend would appear if we selected sample points that were closer together.

Our initial impressions of the fragile bits maps were that the smaller filters had fewer consistent bits than the larger filters. To test this idea, we made histograms showing the consistency of the bits for each of the six filters, across all images of all subjects. These histograms are displayed in Figure 3.21. The x-axis shows the percent of times that bits flipped. Bits that fell in the first bin are the most consistent and flipped between 0 and 5% of the time. Bits that fell in the last bin are the most fragile and flipped between 45 and 50% of the time. The largest filters had the highest number of bits in the first bin, and therefore, in general, the largest filters seem to produce the most consistent bits.
Figure 3.19: Each rectangle shows the fragile bits from part of the OSIRIS iris code. Black bits represent fragile bits. The rectangles shown here are the part of the iris code created with the imaginary part of the largest Gabor filter used.
Figure 3.20: For the large filter, inconsistent regions of the iris code (shown in yellow) often fall in between regions consistently equal to one (shown in red) and regions consistently equal to zero (shown in blue). In the lower pane, the corresponding black and white figure is shown, with inconsistent regions drawn in black.

3.7 Theoretical Impact of Fragile Bits on False Reject Rate

Knowing which bits in a particular subject’s iris code are fragile could improve recognition performance. At enrollment, a sequence of iris images could be taken and analyzed to create the enrollment iris code and a mask which would mask out inconsistent bits. Such a mask could be used in addition to a mask for masking out eyelids and eyelashes. This section presents the theoretical false reject rate for an iris recognition scenario, and then shows how this rate can be improved by masking out the fragile bits in the iris code. We try to follow as closely as possible the notation presented by Bolle et al. [29] in their calculations of FRR. We make two different calculations for FRR. The first calculation presents the error rate for a traditional method that uses all bits in the iris code, and the second calculation presents the error rate for a method that masks fragile bits and only uses the more
Figure 3.21: The largest filters had the highest number of bits in the first bin, and therefore, in general, the largest filters seem to produce the most consistent bits.
consistent bits. Let $Q$ and $R$ be “ground truth” iris codes of length $N$, both from the same iris. When iris code $Q$ is calculated by the iris acquisition system, some of the bits flip so that the result, $\hat{Q}$, is only an approximation to the true iris code, and does not match iris code $R$ exactly. If enough of the bits flip so that the Hamming distance $h(\hat{Q}, R)$ exceeds the decision threshold, $d_T$, then a false reject occurs. Bolle et al. [29] showed that the false reject rate is

$$FRR(d_T) = Pr\left(h(\hat{Q}, R) > d_T | h(Q, R) = 0\right)$$

$$= \sum_{i=d_T+1}^{N} \binom{N}{i} p^i (1-p)^{N-i}$$

where $p$ is the probability that a bit will flip, and $(1-p)$ is the probability of not flipping a bit. Since we showed that not all bits have the same probability of flipping, we modify Bolle et al.’s equation to include two different probabilities. Suppose that $k$ of the bits have probability $p_1$ of flipping, and $N-k$ have probability $p_2$ of flipping. The first $k$ bits will be termed “Set 1” and the second $k$ bits are “Set 2”. Let $p_1 > p_2$ so that Set 1 contains the more fragile bits. Let $i$ be the number of bits that flip when iris code $\hat{Q}$ is acquired. The normalized Hamming distance is the fraction of bits in the two iris codes that disagree, so the probability of a false reject is the probability that $i/N > d_T$, or equivalently, that $i > N \cdot d_T$. Let $j$ be the number of bits from Set 1 that flip, and let $i-j$ be the number of bits from set 2 that flip. The probability of $j$ bits flipping, out of $k$ bits from Set 1 is

$$\binom{k}{j} (p_1)^j (1-p_1)^{k-j}.$$  \hspace{1cm} (3.1)
The probability of $i - j$ bits flipping, out of the $N - k$ bits in set 2 is

$$\binom{N - k}{i - j} (p_2)^{i-j} (1 - p_2)^{N-k-i+j}. \quad (3.2)$$

Thus, the total FRR when using all $N$ bits from the iris code is

$$\binom{k}{j} \binom{N - k}{i - j} (p_1)^j (1 - p_1)^{k-j} (p_2)^{i-j} (1 - p_2)^{N-k-i+j}. \quad (3.3)$$

summed over all possible values for $i$ and $j$: $i$ ranges from $N \cdot d_T$ to $N$, and $j$ ranges from $\max\{i - N + k, 0\}$ to $\min\{i, k\}$.

Now consider the situation where the $k$ fragile bits from Set 1 are masked out. Let $i^*$ be the number of remaining bits that flip. The probability of a false reject is the probability that $i^*/(N - k) > d_T$. The total FRR when using $N - k$ bits is

$$\sum_{i^*= (N-k) d_T}^{N-k} \binom{N - k}{i^*} (p_2)^{i^*} (1 - p_2)^{N-k-i^*}. \quad (3.4)$$

To compare the situation of using all $N$ bits to the second situation of using only $N - k$ bits, we assume some numerical values. Bolle et al. [29] used $N = 173$ because Daugman’s 1993 paper [15] found 173 independent degrees of freedom in the iris code. In order for our results to be easily comparable to Bolle et al., we also take $N = 173$. We analyzed the experimental consistency data so we could choose reasonable values for $k$, $p_1$, and $p_2$. For each subject, we divided the bits into two groups: fragile bits with probability greater than 0.4 of flipping, and consistent bits with probability less than 0.4 of flipping. We counted how many bits fell into each group. On average, 15.0% of bits had probability greater than 0.4 of flipping,
and 85.0% of bits had probability less than 0.4 of flipping. The probability of a bit from the first group flipping averaged 0.4484, and the probability of a bit from the second group flipping averaged 0.1467. 15.0% of 173 is approximately 26, so we let $k = 26$. We also let $p_1=0.4483$ and let $p_2=0.1467$. Assuming these numerical values, the FRR for the traditional scenario is 0.0000132, and the FRR for the scenario which masks the fragile bits is 0.0000000333. Thus, detecting fragile bits at enrollment and masking them out at matching can improve accuracy.

3.8 Empirical Evidence of Improved Accuracy

To truly test the situation described in the previous section, we would have to select some subset of images from each subject and assign them to be “enrollment images” while using the remaining images as “probe images.” Alternatively, we could try to predict which bits in the iris code would be fragile, by using our knowledge that complex numbers near the axes of the complex plane yielded inconsistent bits in the iris code. Such an approach would only be an approximation to truly detecting and masking fragile bits, but it would be simpler because it would only require one iris image to decide which bits to mask.

In section 3.2, we demonstrated that complex numbers near the imaginary axis of the complex plane resulted in fragile real bits, and complex numbers near the real axis resulted in fragile imaginary bits. This idea suggests that one simple way to mask out fragile bits is to mask real bits from complex numbers too close to the imaginary axis, and mask imaginary bits too close to the real axis. We ran two recognition experiments; the first experiment used our normal iris recognition software, and in the second experiment, we masked bits close to the axes.

For the gallery in our experiment, we took the first image from each subject.
The remaining images for each subject were used as the probes. Since we have 24 subjects, we had a total of 24 images in our gallery, and 1202 images in our probe. For both experiments, we had 100% rank-one recognition. Therefore, in order to distinguish between the two experiments, we graphed the histograms of the match and non-match distributions for each experiment. The experiment resulting in a better separation between the match and the non-match distributions is the better method. The result of the first experiment, with no masking of fragile bits, gives the histogram shown in Figure 3.22. The mean Hamming distance for all match comparisons was 0.2772, and the mean Hamming distance for all non-match comparisons was 0.4649.

For the second experiment, we modified the code for creating the mask of our iris code templates. We took the real parts of all 4800 complex numbers for the image, took their absolute value, and then sorted them. Next, we identified all the numbers in the lowest quartile of this set. For each complex number with its real part in the lower quartile of the data, we masked the corresponding real bit in the iris code. Finally, we applied the same procedure to sort and mask the lower quartile of the imaginary numbers. This procedure had the effect of masking all real bits close to the imaginary axis of the complex plane, and masking all imaginary bits close to the real axis of the complex plane\(^1\). The result of this second experiment yields the histogram shown in Figure 3.23. The mean Hamming distance for all match comparisons was 0.1689, and the mean Hamming distance for all non-match comparisons was 0.4459.

The non-match distribution grew slightly wider in the second experiment. The

\(^1\) We would like to thank John Daugman for the idea of masking the lower quartile of numbers (as opposed to trying to tune our program using a specific cut-off threshold). From personal communication with Daugman, we understand that he has previously developed and used techniques described in this section.
Figure 3.22: This figure shows the match and non-match distributions for our software, without any masking of fragile bits.
Figure 3.23: This figure shows the match and non-match distributions for our software, when bits close to the complex axis are masked. The match distribution has moved a significant amount to the left, closer to 0, as desired. In addition, the non-match distribution has also widened slightly.
mean of the non-match distribution has shifted slightly to the left, by 0.0190. However, the match distribution has also shifted a significant distance to the left. The mean of the match distribution has decreased by 0.1083. In the first experiment, the distance between the two means was 0.1877, and the distance between the two means in the second experiment was 0.2770, a large improvement over the previous performance.

3.9 Discussion

The consistency of the different bits in an iris code has not been studied by any other previously published work. Our experiments prove the existence of fragile bits. Our investigations into some of the causes of fragile bits showed that reflections on the iris cause many of the outliers in the data. This study naturally led us to consider in more detail the effects of lighting when an iris image is acquired. In addition to causing reflections on the iris, lighting also affects the dilation of the pupil. The next chapter delves into the effect of lighting on pupil dilation, and in turn, how dilation affects performance of iris recognition.
CHAPTER 4

DILATED PUPILS

Multiple factors affect the size of the iris in an image. For instance, the zoom of the camera and the distance between the camera and the subject alters the imaged iris size. In order to effectively compare two irises, iris biometric systems normalize two iris images to the same size. In Daugman’s early work [15], he assumed a “rubber sheet” model. “The homogenous rubber sheet model assigns to each point in the iris, regardless of size and pupillary dilation, a pair of dimensionless real coordinates \((r, \theta)\) where \(r\) lies on the unit interval \([0,1]\) and \(\theta\) is the usual angular quantity that is cyclic over \([0, 2\pi]\)” [15]. This transformation assumes that when the pupil dilates, the stretch of the iris tissue in the radial direction is linear. This normalization of the iris makes it possible to compare two images of different size. However, information about the degree of pupil dilation is discarded. No prior work has quantified how differing degrees of dilation affect the performance of an iris biometrics system. For this research, we collected a dataset of iris images of varying degrees of dilation and measured the resultant affect in performance. The content of this chapter has been submitted for publication.

4.1 Research in Iris Dilation

Wyatt [17] developed a mathematical model to explain how the collagen fibers in the iris deformed as the pupil dilated. Initially he restricted his model to require
linear deformation along the radial direction of the iris, and then he later relaxed that constraint. He compared his models to measurements of surface features from several human irises. He reports that “some of the data ... appear to be a better match to the lines indicating linear behavior; other data ... appear to be a better match to the lines indicating nonlinear behavior.” He further compared his model to some measurements of angles of collagen fibers measured on an iris, and found that the nonlinear model matched better.

Some iris biometrics researchers have noted that pupil dilation affects the quality of a match between iris images [47–49]. Ma et al. [47] characterized how many of their false non-matches were due to pupil dilation. They explained that “under the extreme conditions (namely the iris texture is excessively compressed by the pupil), the iris after normalization still has many differences with its normal state (i.e., the iris has a pupil of normal size). Therefore, the matching distance between such a pair of iris images is very large. In our experiments, 10.7% false non-matches result from the pupil changes. This is a common problem in all iris recognition methods.” Thornton et al. [48] add an extra processing step to account for the nonlinear deformations of the iris that occur when the pupil dilates. They find the maximum a posteriori probability estimate of the parameters of the relative deformation between a pair of images. Their results show that estimating the relative deformation between the two images improves performance. Wei et al. [49] account for dilation by modeling nonlinear iris stretch as a sum of linear stretch, and a Gaussian deviation term. Their training set includes multiple images of a subject taken under gradually varying illumination. They compare their algorithm with two previous algorithms which use the simple rubber-sheet model and show that their model achieves a lower equal error rate. Both Thornton and
Wei focus on comparing algorithms. Neither work reports experimental results employing subsets of images with different degrees of dilation.

Other than the few papers mentioned above [47–49], the large majority of iris biometrics literature assumes that the “rubber sheet” approach of Daugman is sufficient to deal with the differences in dilation. Furthermore, even though a few researchers have looked at pupil dilation, we have not found any work that quantifies how much dilation really affects recognition. This work is the first that we are aware of to examine the impact of dilation on the performance of an iris recognition algorithm.

4.2 Data and Software

In order to measure how dilation affects recognition, a special iris image data set was collected at the University of Notre Dame between July 2007 and September 2007, using an LG 2200 camera. This data set includes 632 left eye images from 18 different subjects. This set includes ten males and eight females. Twelve subjects were Caucasians and six were Asians. In order to have images of varying pupil dilation, we turned off the ambient lighting for acquiring 28% of the images. The LG 2200 system, like all commercial cameras, uses infrared LEDs to actively illuminate the iris and therefore can still take high-quality iris images with the overhead lights in the room off. Iris dilation is driven by the visible light level, not by infrared light. The LEDs in the LG 2200, while visible, are not bright enough to significantly affect pupil size.

We used iris recognition software developed by Liu [41]. This software, based on ideas from Daugman [15] and Wildes [18] and on the implementation of Ma- sek [50], uses a Canny edge detector and a Hough transform for segmentation. The
boundaries of the iris are approximated using two circles that are not necessarily concentric. We visually inspected the segmentation for all the 632 images. The iris and pupil appeared correctly located in all but one of the images, and for that image, we replaced the machine segmentation with a hand-specified segmentation. Once an image is segmented, log-Gabor filters are used to analyze the texture of the iris and extract a binary iris code. If parts of the iris are occluded by eyelids or eyelashes, the corresponding bits in the iris code are masked, or excluded from future computations.

To show how this software performs for iris recognition, a graph of the match and non-match distributions for this image set is shown in Figure 4.1. The match distribution shows the histogram of Hamming distances from comparisons in which both images are of the same iris. The non-match distribution shows distances from comparisons in which the two images come from different irises. The amount of overlap between the two distributions is related to the error rates in the system. As the figure shows, the match distribution clearly has a lower mean than the non-match distribution, but there is some overlap between the distributions. In commercial biometric systems, this overlap is typically handled by setting a threshold on the Hamming distance so that the probability of a false match is arbitrarily low; e.g., less than one in one million.

Daugman [2] documents the expected non-match distribution for an iris recognition system. He explains that “because any given bit in the phase code for an iris is equally likely to be a 1 or 0, and different irises are uncorrelated, the expected proportion of agreeing bits between the codes for two different irises is HD = 0.500.” However, in order to account for tilt of a person’s head when an image is taken, many iris recognition algorithms try multiple possible rotations of
Figure 4.1: Distribution of normalized Hamming distances for all iris comparisons in our data set. Two histograms are shown. The one on the left represents comparisons between two images of the same iris. The one on the right represents comparisons between two different irises.
an iris during a comparison, and assume the best possible match to be the correct alignment of the iris. Daugman notes that “this ‘best of n’ test skews the distribution to the left and reduces its mean from about 0.5 to 0.458.” [2] Likewise, our non-match distribution is also skewed slightly to the left, with a mean of 0.462, and with the majority of Hamming distance values falling between 0.4 and 0.5.

Our match distribution shows that our software does not perform as well as Daugman’s. In Figure 9 of Daugman’s paper [2], his match distribution has a mean Hamming distance of 0.110. The mean Hamming distance for our match distribution is 0.317. However, our dataset intentionally includes wide variation in pupil dilation. Thus we argue that our software performs well enough that it can be used to show interesting observations about how dilation affects the performance. As demonstrated in [41], this system achieves a 97.1% rank-one recognition rate on an experiment of 4249 images.

4.3 Measuring Dilation

Our first step was to measure the degree of dilation for each image. The segmentation step of the analysis provided the radius of the pupil and the radius of the iris. To measure dilation, we divided the pupil radius by the iris radius. Since the pupil radius is always less than the iris radius, this dilation ratio must fall between 0 and 1. In our 632 images, all dilation ratios were between 0.2459 and 0.7009. The distribution of dilation ratios is shown in Figure 4.2.

Winn et al. [51] report that “several factors are known to affect pupil size, including the level of retinal illuminance, the accommodative state of the eye, and various sensory and emotional conditions. In addition, the size of the pupil tends to change as a function of the individual’s age, with smaller pupils being
Figure 4.2: Histogram showing different degrees of dilation for irises in our data set. A dilation ratio of 0.2 corresponds to a very small pupil (contracted eye), and a dilation ratio of 0.7 corresponds to a very large pupil (dilated eye).
predominant in the elderly population.” In our data, the two oldest subjects had the smallest pupil size. The minimum dilation ratios for the two oldest subjects were 0.25 and 0.28. Wyatt [17] reports that typical pupil diameters during waking hours fall within 12 to 60 percent of iris diameters. We do not have any images with a pupil diameter as small as 12% of the iris diameter. However, since pupil size is somewhat correlated with age [51], the lack of images with very small pupils may be due to the fact that the majority of the subjects in our dataset are between 19 and 32 years old, and we have no subjects older than 52 in our dataset. Two of the subjects attended only one acquisition session. Figure 4.3 plots the minimum and maximum dilation ratios for the remaining 16 subjects. For one subject, the pupil radius varied as much as 31% of the iris radius. The subject showing the least change in pupil dilation still had the pupil radius varying by 15%.

As mentioned in the previous section, some of the images were taken with the overhead lights in the room off. Therefore it is natural to wonder whether the variation in pupil size might occur under real-world conditions. One realistic scenario that causes pupils to contract is a higher level of illumination in the room. One simple way to get pupils to dilate is to use optometrist’s eye drops. Such drops can cause a person’s pupils to dilate and remain dilated for hours. An attacker might use such drops to try to fool the biometric system, or a legitimate user may have dilated pupils after a routine visit to the optometrist. An even more common way of causing pupils to dilate is to wear sunglasses. We avoided using sunglasses for our main dataset because we did not want the complicating factor of extra specular reflections from the glasses. However, we took two pictures, one of an author wearing sunglasses and one of the same person with an extra lamp next to the camera (Figure 4.4). With sunglasses, the pupil dilation ratio was
Figure 4.3: The dilation ratio is the pupil radius divided by the iris radius. Some subjects’ eyes naturally tended to have higher dilation ratios than other subjects’ eyes. Also, some subjects showed more variation in their pupil size than others. Subject 9 had the largest variation (0.307), with a maximum dilation ratio of 0.655 and a minimum dilation ratio of 0.348. The subjects have been ordered in this plot by increasing minimum dilation ratio.
0.54 which was nearly as large as the largest dilation ratio from that subject with the lights off (0.56). With the extra lamp in the room, the dilation ratio was 0.31. The pupil radius had varied by 23% between the two pictures. Therefore, we conclude that such a range of pupil variation is not uncommon and should be expected and evaluated. Figures 4.5, 4.6, and 4.7 show the pupil variation for three of the subjects in our dataset.

4.4 Degree of Dilation Affects Performance

In evaluating the impact of dilation on iris biometrics, the first question to consider is whether the iris recognition software performs as well on a set of data made up of entirely dilated pupils. One might expect that if all images in a dataset showed a large degree of pupil dilation, but were all consistently dilated, the system would still perform well.

Since all dilation ratios in our dataset fall between 0.2 and 0.8, we divided the images into three sets. Set 1 contains images with a dilation ratio less than 0.4. Images in this set have small pupils. Set 2 contains images with a dilation ratio between 0.4 and 0.6. Finally, set 3 contains the most dilated pupils, with dilation ratios greater than 0.6. We obtained the match and non-match distributions for each of the three sets. The match distributions for all three sets are shown in Figure 4.8.

Surprisingly, even though all irises within each set have somewhat consistent degrees of dilation, not all sets showed the same performance. In set 1, the mean Hamming distance for the match distribution is 0.2933. For set 2, the mean Hamming distance is 0.3218. Set 3 has a mean Hamming distance of 0.3415. Apparently, as pupils get larger, the mean of the match distribution increases,
Figure 4.4: It is easy to obtain a difference in pupil dilation. The figure on the left is taken with an extra lamp in the room, and the figure on the right is taken wearing sunglasses.

Figure 4.5: This subject (subject number 05288) showed the biggest difference in pupil size in the data set. The smallest dilation ratio (pupil radius/iris radius) for this subject was 0.3478 and the largest dilation ratio was 0.6545.
Figure 4.6: This subject (subject number 02463) had the smallest pupils in the dataset. The smallest dilation ratio for this subject was 0.2459, and the largest dilation ratio was 0.4400.

Figure 4.7: This subject (subject number 05379) had the largest pupils in the dataset. The smallest dilation ratio for this subject was 0.5391, and the largest ratio was 0.7009.
Figure 4.8: The match distribution for comparisons between eyes with large pupils has a larger mean than the distribution of comparisons between eyes with small pupils.
getting closer to the non-match distribution and increasing the rate of false rejects, or false non-matches.

One probable reason for the degraded performance in the set of dilated eyes involves the simple fact that there is less iris area visible. Typical iris recognition algorithms take the annular region of the eye and create a rectangular normalized region. In a dilated image, less iris area is available for creating each pixel in the normalized image. With less iris data available the eyes will be characterized more poorly.

The creation of the rectangular iris image involves sampling the original image. For an eye with a small pupil, the sample points on the original image will be spaced out along a line going from the pupil, radially outward. For an eye with a large pupil, the sampling density along such a line will be higher. As pupil dilation increases, the radial width of the iris decreases. Thus, there are fewer distinct pixels along the annular width of the iris, but the number of sample points for the normalized iris image remains the same.

In creating the normalized image, our system uses a square of four pixels and interpolates a single value from those four pixels to create one pixel of the normalized image. The normalized image that is created is 20 pixels by 240 pixels. Therefore, the system is expecting an iris with a width at least 40 pixels across. The most dilated pupil in the images analyzed for this paper has an iris that is only 35 pixels in annular width. For this image, the values of some pixels are used more than once in creating the normalized image.

The International Standards Organization has specified that an iris image used for recognition should have 200 pixels across the diameter of the iris [52]. However, we suggest that the diameter of the iris is an inadequate way to measure the total
amount of iris data available. Instead, we recommend that the annular width of the iris is a more correct measure of iris size. Annular width is easily computed by taking the iris radius and subtracting the pupil radius. For systems that allow non-circular pupillary and/or limbic boundaries, the distance between the two can be averaged over a number of radial samples. Even if two circles describing the pupil and iris are not concentric, this measure of annular width is still easily computed, and still represents an average annular width of the iris. However, recording the minimum annular width may also be useful in determining the quality of an image.

Figure 4.9 shows the non-match distributions for the three sets. In set 1, the mean Hamming distance for the non-match distribution is 0.4663. The mean Hamming distance for set 2 is 0.4600 and the mean Hamming distance for set 1 is 0.4472. The difference in means for the non-match distribution was not as large as the difference in means for the match distribution. Therefore, we wished to test whether this difference was statistically significant. To simplify the test, we split the data into two groups, where the first group contains all eyes with dilation ratios below the median, and the second group contains all eyes with dilation ratios above the median. In this test, the null hypothesis is that the means of the non-match distributions for the two groups are equal, and the alternative hypothesis is that they are not equal. We used a balanced, one-factor ANOVA test for comparing the means of the two groups of data. The factor in this test is the dilation ratio. We got an F-statistic of 16092, and a corresponding p-value of 0.0000. Thus, we concluded that the difference between the two groups was indeed significant. That is, the non-match distribution for dilated eyes has a different mean than the non-match distribution for non-dilated eyes. In our experimental results, as pupils get larger, the mean Hamming distance of the non-match distribution decreases,
moving closer to the match distribution and therefore increasing the rate of false accepts or false matches.

4.5 Comparisons between iris images with varying degrees of dilation

Our next experiment deals with comparisons in which the two irises being compared have varying degrees of relative dilation. In order to discuss varying degrees of dilation, we defined a quantity, \( \Delta \), which measured the difference in dilation between two eyes (Figure 4.10).

For a comparison in which one eye has a very small pupil, and one eye has a very large pupil, \( \Delta \) is very large. Within the comparisons done in our experiments, we measured a \( \Delta \) value as large as 0.455. A comparison between two eyes taken under identical lighting conditions could have \( \Delta = 0 \).

Again, we separated all comparisons into sets. For this experiment, the first set contained comparisons with small values of \( \Delta \). The last set contained comparisons with large \( \Delta \), or comparisons between two eyes of very different degrees of dilation. We ran this experiment for both the match distribution (same subject) and the non-match distribution (different subjects). The distributions are shown in Figure 4.11. For the match distribution, there were almost no comparisons with \( \Delta \) in the 0.3 to 0.4 range, so Figure 4.11(a) only shows three sets. The non-match distribution had more comparisons in that range, and therefore, four sets are shown.

The match distribution shifts to the right as the difference in pupil dilation increases. The mean of the distributions for comparisons with \( \Delta \) between 0 and 0.1 is 0.3096. The mean of the distribution for comparisons with \( \Delta \) between 0.1 and 0.2 is 0.3434. When \( \Delta \) is between 0.2 and 0.3, the mean is 0.3834. In contrast, the
Figure 4.9: The non-match distribution of comparisons between eyes with large pupils has a smaller mean than the distribution of comparisons between eyes with small pupils.
Figure 4.10: Δ is a quantity that refers to the difference in dilation between two eyes in a comparison.

The non-match distribution does not shift noticeably. The means of the non-match distributions do not show either an increasing or decreasing trend. The means of the distributions were, in order from smallest Δ to largest Δ, 0.4615, 0.4621, 0.4621, 0.4613.

4.6 Recommendations

We have quantitatively characterized the effect of pupil dilation on iris recognition performance. We found that comparisons between two dilated eyes followed a distribution with a mean normalized Hamming distance of nearly 0.06 higher than the mean of the distribution for non-dilated eyes. The means of both the match and the non-match distributions are expected to fall between 0 and 0.5. Therefore, a shift of 0.06 is nontrivial, amounting to twelve percent of this range. The difference in performance may be partially due to the fact that there is less iris area available in an image of a dilated eye. Points on an eye image are also sampled more closely together for a dilated eye image as compared with a non-dilated image.
Figure 4.11: The match distribution of comparisons between eyes with different degrees of dilation has a larger mean than the distribution of comparisons between eyes with the same degrees of dilation. The non-match distribution is not noticeably affected by the difference in dilation in the comparisons.
We further found that the difference in dilation between an enrollment image and an image to be recognized has a marked affect on the comparison. Comparisons between images with widely different degrees of dilation follow a distribution with a mean about 0.07 higher than the mean of the distribution of images with similar degrees of dilation.

Based on our results, we recommend that a measure of pupil dilation should be created as meta-data to be associated with each generated iris code. This would allow systems to characterize the reliability of an iris code match as a function of the pupil dilations in the underlying images.

One possible line of future work suggested by our results concerns pre-processing the iris image to create an iris code. When the degree of pupil dilation is large, so that the width of the pupil is small in pixels, it may be worthwhile to include a super-resolution step in the pre-processing.
CHAPTER 5

CONCLUSIONS

This research is the first to experimentally demonstrate that different parts of a particular subject’s iris code may be more or less valuable. Our experiments prove the existence of fragile bits and show that the phenomenon of fragile bits appears across races, genders, and different filter types. The fragile portions of the iris code appear as a result of the interaction between the iris and the filter and are not due to iris texture alone. There appears to be no gender difference in the consistency of iris code bits. The evidence is not conclusive as to whether there might be an Asian/Caucasian race difference. There does seem to be a difference in consistency of iris code bits based on the size of filter used.

Contrary to some conventional wisdom in the iris biometrics field, we find no significant difference in the value of the inner rings of the iris versus the outer rings. Perhaps, surprisingly, our results indicate that the middle bands may be slightly better than either the inner or the outer bands.

We show that there are “outliers” in the distribution of values underlying a given iris code bit, and that these are largely due to specular reflections. Some of these specular reflections are faint and have generally been overlooked; however, they have a noticeable, negative impact on the consistency of those regions of the iris code.
We present a theoretical argument showing that masking out fragile bits can reduce the false reject rate in an iris code system by three orders of magnitude. Subsequently, we present an experiment that masks many of the fragile bits in the iris code. This modification of the algorithm significantly shifts the match distribution away from the nonmatch distribution. In the future, iris biometrics could potentially be used with extremely large populations. Any application of iris biometrics on a nationwide scale would necessitate extremely low error rates. Based on our numerical evaluations, we expect that we could reduce the false reject rate by three orders of magnitude by using our knowledge of consistent and inconsistent bits.

This research is also the first to quantify exactly how much impact dilation has on iris biometric performance. One simplistic way to handle strongly dilated eye images would be to simply classify such images as “poor-quality,” and require some minimum quality metric before accepting an image for processing. However, such a decision would restrict the acquisition environment, and there is a current interest in achieving more flexible acquisition conditions. Another way to partially handle the challenge of dilated eyes would be to come up with some transformation, perhaps similar to the one proposed by Thornton [48] or Wei [49], to account for the deformations in the iris tissue.

We propose that a true solution to this challenge would require some investigation into the sampling rates of the iris image. Perhaps using iris video data, we can use super-resolution techniques to acquire more pixels in the iris image. Additionally, a different size filter might be better when there is a smaller annular ring of tissue to analyze. Accounting for pupil dilation is certainly an open question for future iris biometrics research.


