2007

Quality of iris segmentation as a predictor of verification performance

Syvale Lee
West Virginia University

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Recommended Citation

Lee, Syvale, "Quality of iris segmentation as a predictor of verification performance" (2007). *Graduate Theses, Dissertations, and Problem Reports*. 1857.
<https://researchrepository.wvu.edu/etd/1857>

This Thesis is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Thesis in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Thesis has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

QUALITY OF IRIS SEGMENTATION AS A PREDICTOR

OF VERIFICATION PERFORMANCE

by

Syvale Lee

Thesis submitted to the College of Engineering and Mineral Resources

at West Virginia University

in partial fulfillment of the requirements

for the degree of

Master of Science

in

Electrical Engineering

Approved by

Bojan Cukic, Ph.D.

Arun Ross, Ph.D.

Natalia Schmid, D.S.

Lane Department of Computer Science and Electrical Engineering

Morgantown, West Virginia

2007

Keywords: biometrics, iris recognition, segmentation,
evaluating segmentation quality, curve fitting, f-test

Copyright 2007 Syvale Lee

ABSTRACT

QUALITY OF IRIS SEGMENTATION AS A PREDICTOR OF VERIFICATION PERFORMANCE

by Syvale Lee

In this thesis, the accuracy of iris segmentation algorithms will be compared using quality measures based on both homogeneity and heterogeneity of the resulting image regions. After calculation of the quality measures, Principal Component Analysis (PCA) is performed. The resulting pairwise distances between the genuine pairs are compared to the hamming distance scores of the iris templates. The relationship between the two samples will be examined, and curve fitting by least squares regression is shown to be an adequate method of prediction for genuine Hamming distance score. The work proposes a segmentation quality metric that is highly correlated with the distance score.

By determining the relationship between the segmentation quality values and hamming distances, it is possible to apply the same quality measures for future segmented iris images to determine the approximate genuine Hamming distance score without having to first encode the iris template and then calculate the actual hamming distance. This would be of the most practical use in a verification biometric system, where the hamming distance score is used to confirm the identity of a user. The raw data obtained from the user can

be examined and determined whether it should be used for verification after segmentation, without needing to have the hamming score calculated.

ACKNOWLEDGMENTS

I would like to acknowledge my family and friends for their support during this time. In addition, I would like to thank my advisor Dr. Bojan Cukic and my advisory committee for their continual advice and encouragement.

Contents

Table of Contents	v
List of Tables	vii
List of Figures	viii
1 Introduction	1
1.1 Biometrics System	1
1.2 Image Segmentation	2
1.3 Quality Measures	3
1.4 Objective	4
1.5 Thesis Organization	5
2 Related Work	6
2.1 Ideal Iris Segmentation	6
2.2 Non-Ideal Iris Segmentation	8
2.3 Quantitative Measures of Segmentation	9
2.4 Evaluating Image Segmentation	10
2.5 Evaluating Image Quality	11
3 Iris Segmentation Quality Measures	14
3.1 Energy; A Piecewise Model	15
3.2 Energy; A Probabilistic Model	15
3.3 Heterogeneity Magnitude	16
3.4 Homogeneity Magnitude	16
3.5 Regional Heterogeneity Magnitude	17
3.6 Regional Homogeneity Magnitude	17
3.7 Concentric Gradient Magnitude	17
3.8 Canny-based Heterogeneity Magnitude	18
3.9 Summary on Quality of Iris Segmentation	18
4 Analysis	20
4.1 Purpose of Work	20
4.2 West Virginia University Iris Database	21

<i>CONTENTS</i>	vi
4.3 Procedural Outline	22
4.4 Distance Results	23
4.5 Curve Fitting	26
4.6 Hypothesis Testing	29
5 Conclusion	35
5.1 Conclusion	35
5.2 Future Work	36
Bibliography	39
A Correlation Matrix	41

List of Tables

4.1	Polynomial coefficients from curve fitting	27
4.2	Segmented regression coefficients	28
4.3	Results from Statistical Tests after Curve Fitting	31

List of Figures

1.1	Typical human eye	3
2.1	Results of segmentation using Masek's algorithm	7
2.2	Results of segmentation using Zuo's algorithm	8
2.3	Directions of 5x5 block homogeneity analyzer	13
4.1	Iris images that failed to segment properly	22
4.2	Genuine score distributions	24
4.3	Genuine score distributions divided according to segmentation	24
4.4	Pairwise distance distributions for quality measures of Masek's algorithm	25
4.5	Pairwise distance distributions for quality measures of Zuo's algorithm	25
4.6	Masek genuine score plots with curve fitting	33
4.7	Zuo genuine score plots with curve fitting	34
A.1	Graphical representation of absolute value of correlation matrix	42

Chapter 1

Introduction

Personal identification has progressed from authenticated documents to the current state of biometrics, where a person is uniquely and automatically identified based upon physical and behavioral characteristics that are perceived to be inimitable. This advancement is due to the belief that physiological personal identification is more difficult to falsify than a written document. The human iris is currently being researched for its potential as a unique physical identifier. The iris is considered unique as it is one of the few internal organs that is externally visible, and does not significantly change over the course of a typical adult's lifetime.

Using the iris database collected at West Virginia University (WVU), the performance of various iris segmentation algorithms and their resulting relationship with genuine pair Hamming distances will be shown.

1.1 Biometrics System

A basic biometrics system consists of the following blocks: *acquisition*, *feature extraction*, and *comparison* [1]. The first step during feature extraction of an iris based

biometric system must be image segmentation. This is necessary to ensure that only the region of interest will be used in the subsequent processing steps.

After image segmentation, the features of the iris are extracted in the form of a binary iris code which is stored as a template [2], and this iris code is then used as a means of comparison between two system users. In this thesis, the Hamming distance between two iris codes is calculated to produce a quantifiable measure of the difference between the two.

An imposter distance score is defined to be the distance between two templates from two different classes (users). A genuine distance score is defined to be the distance between two templates from the same class [3]. In this thesis, only the genuine distance scores were used to analyze the quality of the segmentation algorithms.

1.2 Image Segmentation

The image of a human eye presents an interesting challenge to computer image segmentation algorithms. Digital images typically segment to produce both a foreground and a background. Sometimes various portions of the foreground can be further segmented based on gradients, edge detection and color components. The WVU iris database consists of 3100 black and white images of various eyes, each one often including any combination of the iris, pupil, sclera, eyelashes, eyelids, and eyebrow (see Fig. 1.1).

Typical image segmentation algorithms are not constrained by the shape (circular or elliptical) or location (between the eyelids and surrounding the pupil) of the foreground object, such as the case with the iris. This means that iris segmentation needs to locate and identify important features in order to be successful. Another consideration is the presence of objects, such as eyelids and eyelashes, that occlude

Figure 1.1 Typical human eye

the desired iris area. Therefore it is necessary to develop an algorithm specifically designed to segment eye images.

Two different iris segmentation algorithms were implemented and tested with the WVU database, one developed by Libor Masek at the University of Western Australia [4] and the other developed by Zuo et. al. at West Virginia University [5].

1.3 Quality Measures

To develop a segmentation quality measure, it is necessary to examine the homogeneity and heterogeneity of the resulting regions after an iris template is segmented. Homogeneity in image processing is defined as the degree to which that the intensity values in a given scalar region remains constant. Heterogeneity is an opposing measure, and is defined as the rate of change of the intensity values in the scalar field. Ideally, in a perfectly segmented image with an unlimited number of regions, each region would have infinite homogeneity and zero heterogeneity, as each region would have no intensity changes. But in a constrained system where there is a limit to the

number of regions produced by segmentation, both homogeneity and heterogeneity are definite and measurable. Both homogeneity and heterogeneity can be measured by calculating the gradient of the scalar field. The magnitude of the gradient at a single point determines how quickly the intensity values change at the given location.

After segmentation using Masek's algorithm, the homogeneity and heterogeneity of different regions of each iris template were measured by calculating the gradient. Both quantities were used as quality of segmentation scores with additional analysis of gradient mapping and edge detection results. PCA was then performed using a training set consisting of one set of quality measures from each unique user class, and subsequent pairwise distances between the genuine pairs were calculated. This process was then repeated using the segmentation algorithm developed by Zuo.

After fitting different curves to the pairwise distances and hamming scores and comparing the goodness of fit, it is seen that the pairwise distances can be used to estimate the Hamming distance between genuine pairs with a reasonable amount of accuracy. There is a direct correlation between the segmentation quality scores and the Hamming distances.

1.4 Objective

To determine the most accurate representation of the relationship between the pairwise distances and the hamming scores, the p-values can be calculated and used to compare the goodness of fit of the polynomial equations. These p-values are calculated for all three pairwise distance measures for both segmentation algorithms. If it can be shown that there is a high degree of correlation between the segmentation quality measure distances and the Hamming distance scores for genuine pairs, then the polynomial equation can be used to predict future genuine pair scores when the

segmentation quality is known.

1.5 Thesis Organization

The rest of the thesis is organized as follows. Chapter 2 presents a literature review of previous works relating to determining the quality of image segmentation. Chapter 3 presents the various segmentation quality measures that we developed. Chapter 4 shows the analysis of the results of the segmentation quality measures when applied to two different segmentation algorithms. Finally, Chapter 5 outlines the conclusions drawn from this study and necessary future work. Appendix A examines the correlation between the developed segmentation quality measures.

Chapter 2

Related Work

2.1 Ideal Iris Segmentation

Ideally, the photograph of an iris should be captured from directly in front of the eye, as this results in an image with circular iris and pupil shapes instead of elliptical. In addition, there is also less distortion of the iris image when it is viewed from the front instead of from the side. If the iris is circular in shape, then a large circle can be used to indicate the boundary between the iris and sclera, and a smaller circle located interior of the other for the iris and pupil boundary. Eyelid and eyelash occlusions are usually located across the top and lower portions of the image, and can be represented with a horizontal line across the upper and lower regions of the iris. Masek's algorithm implements this design and sample results can be seen in Fig. 2.1.

The formula for a circular region is given by:

$$\frac{(x - x_{p0})^2 + (y - y_{p0})^2}{r^2} = 1 \quad (2.1)$$

where

$$x_{p0}, y_{p0} = \text{coordinates of the circle center}$$

Figure 2.1 Results of segmentation using Masek's algorithm

The circular Hough transform is used in conjunction with the edge mapping to determine the edge map. The Hough transform is a feature extraction technique which identifies lines (or in this particular case, circles) in an image. The locations of circles in an image of an eye logically correspond to the locations of the iris and pupil boundaries. The Hough transform also calculates the edge map and gradients of the image in order to reduce the number of possible circle candidates as edges will be orthogonal to the local gradient. Further restrictions are enforced by placing a larger emphasis on vertical and horizontal rates of change. These gradients are biased in the vertical direction for the iris and sclera boundary, whereas both horizontal and vertical gradients are equally biased for the iris and pupil boundary. For the iris boundary, the left and right edges should be most prominent, whereas sides, top and bottom edges should be equally prominent for the pupil boundary. For each eyelid, a linear Hough transform is used to approximate the eyelid boundary, but in our final output, this line is substituted by a horizontal line.

Although Masek's algorithm allows for the isolation of eyelashes by simply elimi-

Figure 2.2 Results of segmentation using Zuo’s algorithm

nating regions of the iris that fall below a grey level threshold, it was determined by experimentation that this greatly reduces the size of the iris template. This is because the WVU iris database often does not have dark colored, distinct eyelashes that are easily distinguishable from the iris background. Therefore in this implementation, eyelash isolation is not performed.

2.2 Non-Ideal Iris Segmentation

Typical iris images are not ideal after capture, as they are prone to various factors such as specular reflection, off-angle capture, defocus blur, motion blur and occlusions by eyelids and lashes. The algorithm developed by Zuo et. al. [5] addresses the limitations of a circular fit by using ellipses to fit the iris and pupil. By using a combination of shape, intensity and location information, Zuo demonstrated an increase in segmentation performance of 26.61% over Masek’s algorithm [5].

For segmentation, Zuo et. al. assume that the pupil area has the darkest intensities

and a circular or elliptical shape of the following form:

$$\frac{((x - x_{p0}) * \cos \phi + (y - y_{p0}) * \sin \phi)^2}{a^2} + \frac{(-(x - x_{p0}) * \sin \phi + (y - y_{p0}) * \cos \phi)^2}{b^2} = 1 \quad (2.2)$$

where

x_{p0}, y_{p0} = coordinates of the ellipse center

ϕ = angle of rotation

a, b = major and minor axes

In addition, the pupil boundary takes into consideration the possibility of distortion of specular reflections and eyelash occlusion. After locating the pupil boundary, the search for the iris boundary is limited to a small search space due to the anatomical constraints of the human eye. The iris is assumed to have a similar shape to the pupil. Using both normalized gradient and connected edge detection, the boundary of the iris is determined.

After segmentation, the iris is unwrapped from an elliptical shape to a rectangular one, using cubic interpolation to calculate pixel intensity.

2.3 Quantitative Measures of Segmentation

Qian Huang and Byron Dom [6] propose that there is one set of measures that can be used to quantitatively measure segmentation results when the ground truth is not available. When a segmentation algorithm produces M non-overlapping regions of an image plane F , it maximizes both the homogeneity within and heterogeneity among the M regions. Between-region heterogeneity can be measured using gradient magnitude. Low internal homogeneity can also be referred to as high within-region heterogeneity.

When the segmentation algorithm performs poorly, the within-region heterogeneity H_w grows larger while the between-region heterogeneity H_b approaches zero. The authors show that both measures of internal homogeneity and between-region heterogeneity provide a reasonably accurate assessment of the quality of segmentation.

2.4 Evaluating Image Segmentation

Sylvie Philipp-Foliguet and Laurent Guigues propose a method of evaluating segmentation of color images when there is no ground truth [7]. Segmentation of an image can be viewed as an optimization problem, where the partitioning of an image minimizes a total energy.

$$E(k, P) = \sum_{R \in P} E_D(R) + k \cdot E_C(R) \quad (2.3)$$

where

k = contribution parameter

R = region R of partition P of an image

E_D = modeling quality (distance between partition model and image)

E_C = complexity energy (prevents oversegmentation)

Since the output of both iris segmentation algorithms produces only one region of interest, the degree of segmentation E_C does not need to be measured. That is, the algorithms are designed to produce a single iris region and are incapable of oversegmenting an input image to produce multiple regions of interest. Therefore the only energy that needs to be measured is the one that pertains to modeling quality, where the distance between the partition model and the image is calculated.

E_D can also be referred to as the internal energy, and can be calculated in two different methods. The first piecewise model is measured using L_2 norm. For a

segmented region R containing A_i pixels $(X_1, X_2, \dots, X_{A_i})$, let μ be the average value of the pixels. Therefore the variance matrix V of matrix X_p is defined as:

$$V = \frac{1}{A_i} \sum_{p=1}^{A_i} (X_p - \mu)(X_p - \mu) \quad (2.4)$$

The vector which minimizes the distance to the initial data is the mean, which is found by:

$$Q_{R_i} = A_i \cdot \text{Trace}(V) = A_i \sum_j \lambda_j \quad (2.5)$$

where λ_j is the j -th eigenvalue of the matrix V .

The second model is probabilistic, where the pixel values are assumed to be i.i.d. samples of a Gaussian law and the energy of a region is:

$$G = A_i \log(\det V) \quad (2.6)$$

Although these models were originally derived for use with pictures with three color components, they have been simplified for use with black and white photos with 256 intensity levels. A normalization coefficient for each model is also required. Since each image has $2n$ possible intensity values, then n^2 is an upper bound for the variance and covariance values. Therefore Q has a normalization coefficient of $\frac{1}{n^2 \times A \times 100}$, and G has a coefficient of $\frac{1}{A}$.

2.5 Evaluating Image Quality

Various methods used for determining image quality include mean squared error and peak signal to noise ratio. Kebin An et al. proposed another method for determining image quality by using a measure which utilizes second derivative masks to calculate local image homogeneity [8]. It assumes that there are neighborhoods with smooth intensities in a given image.

Let the pixels $I(i, j)$ in an intensity homogeneous region be:

$$B_{kl} = I(i, j) \mid (i, j) \in W_{kl}$$

$$W_{kl} = \left\{ (i, j) \mid k - \frac{W-1}{2} < i < k + \frac{W-1}{2}, l - \frac{W-1}{2} < j < l + \frac{W-1}{2} \right\} \quad (2.7)$$

where

$$I(i, j) = \text{image signal}$$

$$W_{kl} = \text{rectangular window of size } W \times W \text{ centered at } (k, l)$$

In a homogeneous region, the signal variation is close to zero. This method divides the image into overlapping blocks of the same size, where each block uses a local uniformity analyzer to compute a measure of homogeneity. The local uniformity analyzer measures homogeneity in eight different directions, and if the image is homogeneous in a given direction, then the output is close to 0 (see Fig. 2.3). For each pixel, the resulting homogeneity measure ξ_{kl} is found by finding the total sum of the absolute value of each output, where L_r and $mask_r$ are the corresponding luminance value and mask at each direction as in Fig. 2.3.

$$\xi_{kl} = \sum_{r=1}^8 |L_r - mask_r|$$

The homogeneity analyzer can be expressed as an operator on the image function I .

$$I_o(i, j) = -I(i-2, j) + 2I(i, j) - I(i+2, j) = -(I'(i, j) - I'(i-2, j)) \quad (2.8)$$

The analyzer is seen to be the output of a second order finite difference operator, which acts as a high pass operator on the image I .

Figure 2.3 Directions of 5x5 block homogeneity analyzer

Chapter 3

Iris Segmentation Quality Measures

We developed a total of twenty-four quality measures to determine the accuracy of the segmentation of an iris image. This has been done by examining both the homogeneity and heterogeneity of the segmented output. The output of an iris segmentation algorithm defines two borders, one that separates the iris of an eye from the sclera, the other the pupil from the iris. These borders are represented in two different manners, dependent on the segmentation algorithm.

For Masek's algorithm, the iris and pupil areas are represented by two circles, each with a row and column coordinate for the center and a positive value for the radius. The occlusions are represented by horizontal rectangular areas, which when unwrapped to a normal coordinate system, produces semicircular binary masks.

For Zuo's algorithm, the iris and pupil areas are represented by two ellipses, each with a row and column coordinate for the center, two positive values for the major and minor axes, and a rotation angle ϕ . The occlusions are represented with binary masks.

Using both wrapped and unwrapped results from both segmentation algorithms, the following quality measures were calculated. These measures are based on the previous works as mentioned in Chapter 2.

3.1 Energy; A Piecewise Model

As outlined in Section 2.4, the internal energy of a given region can be found using a piecewise model.

$$Q_i = \frac{A_i \times \text{trace}(V)}{n^2 \times A_i \times 100} \quad (3.1)$$

where

i = pupil or iris region combined with occlusion mask

A_i = number of non-zero pixels

V = eigenvalue of vectorized intensity values

n = 127 (number of gray levels divided by 2)

3.2 Energy; A Probabilistic Model

As outlined in Section 2.4, the internal energy of a given region can be found using a probabilistic model.

$$G_i = \log \{\det \{V\}\} \quad (3.2)$$

where

i = pupil or iris region combined with occlusion mask

V = eigenvalue of vectorized intensity values

3.3 Heterogeneity Magnitude

The gradient heterogeneity magnitude at any given point in the image is determined by finding the x and y directional gradients (px, py) of the image. The magnitude $p = \sqrt{px^2 + py^2}$ is calculated for every point in the image.

$$\text{Heterogeneity Magnitude}_i = \frac{\sum H_t}{C_{Ht}}; \quad (3.3)$$

where

i = pupil or iris region edge (one unit thick) combined with occlusion mask

H_t = vector of gradient magnitude values p

C_{Ht} = area of iris or pupil region

3.4 Homogeneity Magnitude

Using the same gradient magnitudes p as in Section 3.3, gradient homogeneity magnitude for a given region i is found with the following expression.

$$\text{Homogeneity Magnitude}_i = \frac{\sum H_m}{C_{Hm}}; \quad (3.4)$$

where

i = pupil or iris region combined with occlusion mask without edge (one unit thick)

H_m = vector of gradient magnitude values p

C_{Hm} = area of iris or pupil region minus area of edge

3.5 Regional Heterogeneity Magnitude

We found that the original 3x3 mask, as outlined in [8], was too sensitive to iris texture to calculate local image homogeneity. Instead, we implemented a 5x5 mask (Fig. 2.3), which was more robust when the iris was highly textured yet still responded well to the presence of eyelids and eyelashes.

The homogeneity measure is found by convolving the eight masks with the original iris image. The absolute values of the outputs are summed at each point along the iris edge for a diameter of 1 pixel to find the heterogeneity measure.

$$\xi_{kl} = \sum_{r=1}^8 |L_r \text{ mask}_r| \quad (3.5)$$

3.6 Regional Homogeneity Magnitude

The same steps as in Section 3.5 were carried out to find the homogeneity measure. However, the region of interest was limited to the iris area excluding the edge unlike the approach in Section 3.5.

3.7 Concentric Gradient Magnitude

Using the unwrapped iris templates that are 20x240 pixels in size, we divided the templates into ten regions of incrementally larger size, beginning with the outer iris edge and progressing towards the pupil (2, 4, 6, ..., 18, 20 pixel rows). Eq. 3.1 is then used again for each concentric region.

$$R_i = \frac{A_i \times \text{trace}(V)}{n^2 \times A_i} \quad (3.6)$$

where

- i = iris region combined with occlusion mask
- A_i = number of non-zero pixels
- V = eigenvalue of vectorized non-zero intensity values
- n = 127 (number of gray levels divided by 2)

3.8 Canny-based Heterogeneity Magnitude

Canny edge detection is used to detect any continuous edges in the original image. Using the region immediately around the iris boundary (3 pixels in width), the following metric is found.

$$C_i = \frac{\sum E}{A} \times \frac{S_E}{A} = \frac{S_E \sum E}{A^2} \quad (3.7)$$

where

- i = pupil edge or iris edge region combined with occlusion mask
- E = canny edge values in iris boundary region
- S_E = number of non-zero pixels
- A = area of pupil edge or iris edge region

3.9 Summary on Quality of Iris Segmentation

In total, we developed eight different iris segmentation quality measures based on previous related work, where one measure was applied to concentric regions of the iris. The other measures were applied to each region of an iris image, iris and pupil. Two measures reflect the internal energy of a given region, and the remaining six measures

determine either the heterogeneity or homogeneity of a region. Each of the various segmentation measures produced a quantitative number that is an indication of either the homogeneity or heterogeneity of either the iris or pupil region. An examination of the linear dependency and correlation between the segmentation quality measures is provided in Appendix A.

To compare individual images, the measures can be normalized from 0 to 1 using linear transformation, where 0 and 1 correspond to the lowest and highest quality segmentation score respectively. In this manner, it is possible to compare the quality of segmentation of a single region between various images. However, in order to compare the total segmentation quality of any pair of segmentation results, it is necessary to calculate an overall segmentation quality score that combines the information from each individual measure. This method is outlined in the next chapter.

Chapter 4

Analysis

4.1 Purpose of Work

Currently the established method for determining the level of performance of a given iris segmentation algorithm is by analyzing an iris database of significant size and calculating both the failure and success rates. Therefore the final result is not an indication of solely how well the segmentation algorithm performs, but the entire iris recognition system which includes algorithms contained in each block (*acquisition*, *feature extraction*, and *comparison*) in the biometric system, such as encoding the iris features. In addition, the Receiver Operating Curve (which summarizes the failure and success rates) does not indicate how well the segmentation algorithm performed for a particular given image.

Since the biometric system consists of a series of blocks, the resulting performance curve depends on the total combined performance of each individual block. If the other implemented algorithms perform poorly, then it is difficult to determine which segmentation algorithm performs better than others. Since a combination of the blocks which contain the least amount of error propagation would produce the best

performing biometric system, a method to determine which segmentation algorithm performs better would be beneficial.

In addition, if the performance of an iris segmentation algorithm could be determined for an individual image, the results could be used to help determine necessary future improvements for the specified algorithm. By examining problems common to images that are identified as segmenting poorly, the implemented segmentation algorithm can be changed to handle difficult images more robustly.

4.2 West Virginia University Iris Database

The West Virginia University Iris Database consists of 3100 black and white frontal images of both left and right eyes. Upon receipt of the database, we had to rename 152 of the templates as they had been mislabeled with an incorrect class designation. The correct class names were determined by Hamming distances and verified by visual inspection. After renaming, there were 472 different classes.

After elimination of all iris templates that fail to segment by either Masek's or Zuo's segmentation method, we were left with a total of 3082 images distributed among 466 classes, producing a total of 10,474 genuine pairs. Masek's algorithm failed to segment fifteen images, whereas Zuo's algorithm failed to segment one image and also produced inaccurate iris masks for two other images.

The following figure (Fig. 4.1) shows the eighteen images that did not produce meaningful segmentation results.

Figure 4.1 Iris images (a) through (o) failed to segment with Masek's algorithm, (p) through (r) failed to segment with Zuo's algorithm

4.3 Procedural Outline

Both Masek's and Zuo's algorithms for ideal and non-ideal irises were implemented. We applied the same segmentation quality measures (as described in Ch. 3) to the results of each algorithm. The resulting values of segmentation quality were then used with PCA [9], where the training set consisted of one set of measures from each class.

PCA was chosen as a means of fusing the segmentation quality scores together. Separately, each segmentation quality measure does not present an overall idea of the quality of iris segmentation. In addition, the measures by themselves do not indicate how the segmentation performance varies within and between user classes. Using

PCA, each user class is considered distinct from the others, and intra-class variation of segmentation measures are minimized.

After performing PCA, we determined the pairwise distances between each genuine pair using Mahalanobis, Cityblock and Euclidean methods. Using linear regression methods and curve fitting, a variety of different curves were calculated in order to determine the relationship between the pairwise distances and the hamming scores. By comparing the results of a test statistic which examined the predicted values and actual values of genuine Hamming distance scores, the best equation and pairwise distance method used to describe this relationship was found.

4.4 Distance Results

The histograms of the genuine pair Hamming distance scores from using both Masek's and Zuo's methods are presented in Fig 4.2. These show that the algorithm developed by Zuo et al. has better performance for the WVU iris database. Therefore it can be expected that a histogram of segmentation quality measures for Zuo's algorithm will show a sharper and narrower peak than one for Masek's.

Although it appears that the genuine histogram distribution (in Fig 4.2) using Masek's algorithm contains imposter scores, it can be shown that the high genuine scores are derived from a combination of poor segmentation results and poor quality images. When a subset of 1000 templates was chosen and manually classified as either a successful or failed segmentation, the distribution was split in two as shown in Fig 4.3. Using the first 1000 templates of the WVU database, it was found that 354 images segmented very well, while the remaining 646 images displayed some degree of error with their segmentation results.

Given the quality measures as an $m \times n$ data matrix $X = (x_1, x_2, \dots, x_m)$, each row

Figure 4.2 Genuine score distributions **Figure 4.3** Genuine score distributions divided according to segmentation

vector x_r and x_s is used to calculate pairwise distances:

$$\text{Cityblock: } d_{rs} = \sum_{j=1}^n |x_{rj} - x_{sj}| \quad (4.1)$$

$$\text{Euclidean: } d_{rs}^2 = (x_r - x_s)(x_r - x_s)' \quad (4.2)$$

$$\text{Mahalanobis: } d_{rs}^2 = (x_r - x_s)V^{-1}(x_r - x_s)' \quad (4.3)$$

where V is the sample covariance matrix.

As predicted, the histograms from the different pairwise distance methods used

for Masek's algorithm show a large range in quality scores. These are presented in Fig 4.4.

Figure 4.4 Pairwise distance distributions for quality measures of Masek's algorithm

Also as predicted, the histograms from the different pairwise distance methods used for Zuo's algorithm show a much narrower range (Fig 4.5).

Figure 4.5 Pairwise distance distributions for quality measures of Zuo's algorithm

Curve fitting will be used to determine whether or not there is an observable relationship between the pairwise distances and the hamming distance scores. This process is outlined in the next Section.

4.5 Curve Fitting

Correlation between the pairwise distances (X) and hamming scores (Y) can be observed by plotting the two data sets with MATLAB, however the most optimal equation which can be used to describe the relationship between the variables is unknown. We found 4 different equations, and calculated their goodness of fit. The following curves were chosen to be implemented: linear regression, segmented regression, quadratic, and cubic. In each case, the coefficients of the polynomial (and the breakpoint for segmented regression) were determined using a least squares method where the sum of the residuals is minimized [10] [11].

The four equations used to describe the fitted curves can be expressed as follows, where c is the calculated breakpoint used in segmented regression. The breakpoint c minimizes the mean squared error between the sample values and the segmented regression model. The slopes and intercepts of the linear functions change below and above the breakpoint c .

$$\begin{aligned}
 \hat{y} &= p_1x + p_2 && \text{linear} \\
 \hat{y} &= p_{1a}x + p_{2a} && \text{segmented, for } x \leq c \\
 \hat{y} &= p_{1b}x + p_{2b} && \text{segmented, for } x > c \\
 \hat{y} &= p_1x^2 + p_2x + p_3 && \text{quadratic} \\
 \hat{y} &= p_1x^3 + p_2x^2 + p_3x + p_4 && \text{cubic}
 \end{aligned} \tag{4.4}$$

Model plots of the four equations for Masek's algorithm are presented in Fig 4.6, and for Zuo's algorithm in Fig 4.7. The calculated polynomial coefficients are summarized in Table 4.1, and the segmented regression coefficients and breakpoint values in Table 4.2.

By visual inspection of the plots of Masek's algorithm (Fig 4.6), it appears that the Mahalanobis pairwise distance results gives the highest degree of correlation of

Alg	Dist	Curve	Polynomial Coefficients			
Masek	Cit	Lin	0.31157674	0.29457139		
		Quad	-0.46229627	0.56765183	0.26620307	
		Cubic	0.51485609	-0.94333874	0.69572465	0.25690630
	Euc	Lin	0.30017820	0.30931277		
		Quad	-0.65603973	0.62904757	0.27872354	
		Cubic	1.35379522	-1.84059666	0.91407657	0.26103594
	Mah	Lin	0.50695063	0.24443238		
		Quad	-1.53462473	1.37319981	0.13755377	
		Cubic	2.02109754	-3.45185028	1.90812060	0.09398305
Zuo	Cit	Lin	0.10006261	0.24846237		
		Quad	-0.33094676	0.24204584	0.23804295	
		Cubic	0.48083671	-0.71937036	0.31752155	0.23454611
	Euc	Lin	0.07238424	0.24991185		
		Quad	-0.19894604	0.17707982	0.24069218	
		Cubic	0.12901935	-0.32454965	0.20665560	0.23905880
	Mah	Lin	0.26232267	0.23991506		
		Quad	-0.80401739	0.57647941	0.22188576	
		Cubic	1.47383376	-2.05345429	0.78542307	0.21381065

Table 4.1 Polynomial coefficients from curve fitting

the three distance methods, as the Mahalanobis distances appear to follow the fitted curves more closely than the other two. This hypothesis is confirmed later in Section 4.6.

For the model plots for Zuo's algorithm (Fig 4.7), it appears that there is not a significant difference in correlation between the pairwise distance methods. Therefore it will also be necessary to calculate the degree of correlation to determine which curve models the data more closely. The actual values for correlation are presented in Section 4.6.

To verify that the Mahalanobis pairwise distance method provides the highest degree of correlation between Hamming and Mahalanobis distances, the numerical

Alg.	Dist.	Breakpoint c	p_1 for $\leq c$	p_2 for $\leq c$	p_1 for $> c$	p_2 for $> c$
Masek	Cit	0.265	0.43687831	0.27429861	0.20132069	0.33672138
	Euc	0.165	0.62446444	0.27177832	0.18791252	0.34380939
	Mah	0.267	0.87322166	0.17432671	0.13426251	0.37162880
Zuo	Cit	0.256	0.17408088	0.24080192	0.00978516	0.28286163
	Euc	0.365	0.11587476	0.24436352	0.00253688	0.28573185
	Mah	0.102	0.60149187	0.21668325	0.01318530	0.27669052

Table 4.2 Segmented regression coefficients

values of correlation can be compared for each distance method. The values of correlation for each algorithm and pairwise distance method are not directly compared, but they are used in a statistical test which allows us to either accept or reject the hypothesis that the fitted curve can be used to predict the hamming scores. If the two samples are correlated, then the curve which demonstrates the highest degree of correlation can be used to predict genuine pair Hamming distance using solely the calculated segmentation quality scores as input.

It was observed that for Zuo's algorithm, all three of the distance methods have similar plots where they did not produce a large range of segmentation quality values. We interpret these results as demonstrating the inter-class variance in segmentation performance due to differences in image quality. These differences can be attributed to changes in lighting, occlusion, defocus, and types of blur common in iris imaging. This corresponds to the earlier observation that there is a smaller range of hamming distance scores when using Zuo's algorithm, which confirms that there is a high degree of correlation between the segmentation quality measures and the hamming scores.

4.6 Hypothesis Testing

To determine if the segmentation quality measures and the calculated curves can be used to predict genuine pair hamming distance, it is also necessary to determine if there is a strong degree of correlation between pairwise distances X and hamming distance Y . To determine the amount of correlation, it is necessary to determine the goodness of fit of the curve equations. This is done by comparing the projected hamming scores $\hat{Y}$ (using the segmentation quality measures as an input to the curve equation) with the actual hamming scores. Each set of hamming scores is of size $n = 10474$. Using hypothesis testing, the null and alternative hypotheses are defined as follows:

- H_o : The sample sets (hamming distance and pairwise distance) are not correlated
- : The polynomial equation cannot be used to predict hamming distance
- : $|F| < f^*$
- : p-value $> \alpha = 0.05$
- H_1 : The sample sets (hamming distance and pairwise distance) are correlated
- : The polynomial equation can be used to predict hamming distance
- : $|F| \geq f^*$
- : p-value $\leq \alpha = 0.05$

F is the chosen test statistic and summarizes the information in the sample set, and f^* is the critical value that the test statistic F must exceed in order for the null hypothesis to be rejected. The chosen test statistic is the F-test, which is used to determine if the standard deviations of two populations are unequal. In our case, we have chosen to see whether the standard deviation of the first set is greater than the second set standard deviation [12]. The first set is chosen to be $\hat{Y}$, which is the

projected hamming distance values using pairwise distance as input X to a polynomial equation. The second set is chosen to be Y , which is the actual corresponding hamming distance values.

To test H_o , we first assume that H_o is true and apply an appropriate test statistic. The value of the significance level α is defined as follows and is traditionally equal to 0.05 for hypothesis testing.

$$\begin{aligned}\alpha &= P(\text{rejecting } H_o \text{ (accepting } H_1) \text{ when } H_o \text{ is true}) \\ &= P(H_1 | H_o)\end{aligned}$$

The value of the test statistic is used to find a corresponding probability (p-value) of obtaining such a statistic given that H_o is true [13]. If the p-value $\leq \alpha$, then the null hypothesis H_o is rejected, and H_1 is accepted. Also, if the test statistic F exceeds the critical value f^* , then the null hypothesis H_o is rejected. The test statistic F is defined as:

$$F = \frac{R^2}{1 - R^2} \times \frac{df_2}{df_1} \quad (4.5)$$

where

$$R^2 = \text{coefficient of determination R-square} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

$$df_1 = \text{numerator degrees of freedom in an F distribution} = \text{degree of polynomial}$$

$$df_2 = \text{denominator degrees of freedom in an F distribution}$$

$$= n - df_1 - 1 = 10473 - df_1$$

The critical value f^* of the F distribution with df_1 and df_2 degrees of freedom and significance level α is defined as:

$$f^* = F(\alpha, N_1 - 1, N_2 - 1) = F(\alpha, df_1, df_2) \quad (4.6)$$

and the resulting p-value can be calculated as:

$$p - \text{value} = P(F(\alpha, df_1, df_2) > F) = 1 - P(F(\alpha, df_1, df_2) \leq F) \quad (4.7)$$

As can be seen in Table 4.3, the calculated values of F are much larger than the corresponding values of f^* (and every p-value is approximately equal to 0 and less than α), therefore the null hypothesis is rejected for each curve. It is determined that every fitted curve can be used to model the relationship between the segmentation quality measures and the resulting hamming score.

		Masek			Zuo		
Curve		City	Euc	Mah	City	Euc	Mah
Linear	R	0.4285	0.4001	0.5534	0.1047	0.0971	0.1702
	R^2	0.1836	0.1601	0.3063	0.0110	0.0094	0.0290
	F	2354.9966	1996.1659	4623.0788	115.8122	99.5219	311.9737
	f^*	3.8423	3.8423	3.8423	3.8423	3.8423	3.8423
	p	0	0	0	0	0	0
Segmented	c	0.2650	0.1650	0.2670	0.2560	0.3650	0.1020
	R	0.4384	0.4232	0.6076	0.1211	0.1099	0.2176
	R^2	0.1922	0.1791	0.3691	0.0147	0.0121	0.0474
	F	2491.5529	2284.7691	6127.4469	155.5415	127.8290	519.5881
	f^*	3.8423	3.8423	3.8423	3.8423	3.8423	3.8423
	p	0	0	0	0	0	0
Quadratic	R	0.4393	0.4219	0.6062	0.1241	0.1148	0.2124
	R^2	0.1930	0.1780	0.3675	0.0154	0.0132	0.0451
	F	1252.2976	1133.7891	3041.6226	81.7168	69.7470	246.8501
	f^*	2.9966	2.9966	2.9966	2.9966	2.9966	2.9966
	p	0	0	0	0	0	0
Cubic	R	0.4398	0.4252	0.6086	0.1249	0.1150	0.2152
	R^2	0.1934	0.1808	0.3704	0.0156	0.0132	0.0463
	F	836.9745	770.1152	2053.1009	55.2242	46.6543	169.1425
	f^*	2.6058	2.6058	2.6058	2.6058	2.6058	2.6058
	p	0	0	0	0	0	0

Table 4.3 Results from Statistical Tests after Curve Fitting

By definition, R^2 is the fraction of the total squared error that is explained by the curve model and therefore it can be used to compare the goodness of fit for different models. The closer the value of R^2 to 1, the better the model approximates the observed data. As seen in Table 4.3, Mahalanobis pairwise distance produces

the highest value of R^2 for both segmentation algorithms, therefore it is concluded that Mahalanobis pairwise distance provides a greater degree of correlation between segmentation quality measures and hamming distance than the other two pairwise distance methods. When looking at R^2 for only Mahalanobis distances, it appears that linear regression is the poorest model, and there is no significant difference between the segmented regression, quadratic and cubic models.

When comparing the values of R^2 between the two different algorithms, it is seen that the values for Zuo's algorithm are much lower and do not display a significant improvement between the different curve models. This supports our conclusion that the variation in the segmentation quality scores is caused by variations in the experiment environment and are not measurable with the iris segmentation quality metrics. Since there is not a significant improvement in modeling the data between the different curves, all of the curves appear to be much flatter for Zuo's algorithm than for Masek's.

(a) Hamming distance vs cityblock distance

(b) Hamming distance vs euclidean distance

(c) Hamming distance vs mahalanobis distance

Figure 4.6 Masek genuine score plots with curve fitting

(a) Hamming distance vs cityblock distance

(b) Hamming distance vs euclidean distance

(c) Hamming distance vs mahalanobis distance

Figure 4.7 Zuo genuine score plots with curve fitting

Chapter 5

Conclusion

5.1 Conclusion

By examining the heterogeneity and homogeneity of the results of segmentation performed on an iris image, it is possible to determine the quality of segmentation. Using two different segmentation algorithms allows the quality measure results to be compared and verified. The model which best describes the relationship between the transformed segmentation quality measures and the genuine pair hamming distances determines which pairwise distance method provides the best predictive measure of the resulting distance scores. Using the F-test statistic and their resulting p-values, it is shown that the segmentation quality measures have a high correlation to the genuine hamming distances. One can use the segmentation quality measures with a chosen polynomial equation to effectively predict the resulting hamming distance between genuine pairs.

Determining the level of performance of an iris segmentation algorithm has many advantages other than the elimination of having to use visual inspection to determine if an algorithm has performed correctly. It can be utilized as a means to isolate

problem areas of existing algorithms as a tool for design improvement, as problem images can be readily identified. It can also be used as a predictor to determine whether a query image should be discarded, as the genuine pair hamming distance can be estimated without having to compare the iris to all the classes contained in the database. By being able to compare the efficiency of various algorithms, their individual contributions in a complete biometric system can be compared. Also by extension, it is shown that the performance of an iris recognition system is significantly related to that of the iris segmentation algorithm. If a poor iris segmentation algorithm is implemented, the resulting genuine hamming distance scores will show a greater degree of intra-class variance.

When comparing the quality measures for both Masek's and Zuo's segmentation algorithms, it can be seen that Zuo's segmentation algorithm performs more robustly for the WVU iris database. Masek's algorithm results in a larger range of segmentation quality values than Zuo's algorithm. The result is that the genuine pair distribution histogram of an iris recognition system using Zuo's segmentation algorithm has a more normal distribution than that of the same system using Masek's algorithm, meaning that the failures rates would be smaller and the system would perform better.

5.2 Future Work

Although two segmentation algorithms were implemented and tested, other algorithms (such as the method designed by John Daugman) need to be implemented in future work to provide further verification of the claim of direct correlation between the segmentation quality measures and Hamming distance.

In addition, the quality measures could be improved in many ways. One example

problem that was encountered was the presence of specular reflections within or along a region, which artificially decreased the quality score. Removal of the specular reflections could be carried out by performing additional segmentation and using an image processing technique such as inpainting. Once the reflection has been inpainted, then the iris template would have to exclude the processed regions.

Other quality measures could also be developed, utilizing segmented region attributes other than homogeneity and heterogeneity. For example, a well defined boundary in an image can be represented by a solid continuous line, whereas a poorly defined boundary is more likely to be broken and discontinuous. This property could potentially be exploited and an measure could be designed to measure the degree of continuity.

Even though the segmentation results did not always produce eyelid and eyelash occlusion masks, the information contained in them when they are actually generated should also be used. This could be very difficult to implement, as the occlusion mask is typically a combination of featureless eyelid and feature-rich eyelashes. Therefore it is a region which contains up to two unsegmented subregions. The occlusion mask and other features of the background region could be used to calculate future quality measures.

Testing the quality measures on more than one iris database would also determine if the technique can be adapted to different sources with varying image quality. If the same quality measures could be applied to different databases, then they could be used to identify the most efficient segmentation algorithm for a particular capture method, instead of using the genuine and imposter score distributions.

Finally, the results of the iris segmentation quality measures should be compared to those of other techniques, should they be developed. Although visual inspection can provide the ground truth, it is not feasible to carry this out on large databases,

and visual inspection does not provide a quantitative means to compare the two methods. Comparison to other measures would provide an indication of the relative performance.

Bibliography

- [1] D. Maltoni, D. Maio, A. Jain, and S. Prabhakar, *Handbook of Fingerprint Recognition*, 1st ed. (Springer-Verlag, New York, NY, 2003).
- [2] J. Daugman, “How Iris Recognition Works,” *IEEE Transactions on Circuits and Systems for Video Technology* **14**, 21–30 (2004).
- [3] A. K. Jain, A. Ross, and S. Prabhakar, “An Introduction to Biometric Recognition,” *IEEE Transactions on Circuits and Systems for Video Technology*, Special Issue on Image- and Video-Based Biometrics **14**, 4–20 (2004).
- [4] L. Masek, “Recognition of Human Iris Patterns for Biometric Identification,” B.Sc. dissertation (University of Western Australia, 2003).
- [5] J. Zuo, N. Kalka, and N. Schmid, “A Robust Iris Segmentation Procedure for Unconstrained Subject Presentation,” 2006 Biometric Consortium Conference (BCC 2006) (Sept. 19-21, 2006).
- [6] Q. Huang and B. Dom, “Quantitative methods of evaluating image segmentation,” *ICIP '95: Proceedings of the 1995 International Conference on Image Processing (Vol. 3)* p. 3053 (1995).

- [7] S. Philipp-Foliguet and L. Guigues, “New Criteria For Evaluating Image Segmentation Results,” 2006 IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP 2006 Proceedings. 2 (2006).
- [8] K. An, J. Sun, and W. Du, “Homogeneity Based Image Objective Quality Metric,” IEICE - Trans. Inf. Syst. **E89-D**, 2682–2685 (2006).
- [9] I. Jolliffe, *Principal Component Analysis* (Springer, 1986).
- [10] R. J. Oosterbaan, in *Chapter 6 in Drainage Principles and Applications*, 2nd ed., H. P. Ritzema, ed., (ILRI Publication 16, Wageningen, The Netherlands, 1994).
- [11] S. E. Ryan and L. S. Porth, “A tutorial on the Piecewise Regression Approach Applied to Bedload Transport Data,” General Technical Report RMRS-GTR-189 (Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station 2007).
- [12] “NIST/SEMATECH e-Handbook of Statistical Methods,” <http://www.itl.nist.gov/div898/handbook/> (Accessed December 9, 2007).
- [13] W. J. Conover, *Practical Nonparametric Statistics*, 3rd ed. (John Wiley and Sons, 1998).

Appendix A

Correlation Matrix

To examine the linear relationship between each of the segmentation quality measures, we calculated the correlation matrix for all 24 measures using the standard definition of correlation ρ ,

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\left[\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2 \right]^{\frac{1}{2}}} \quad (\text{A.1})$$

producing a 24x24 correlation matrix. A 8x8 subsection of the correlation matrix from the upper left is shown below:

1.000	-0.00199	0.82911	-0.09532	0.09486	-0.00403	0.12986	-0.39102
	1.000	-0.03649	0.67161	-0.07875	-0.18900	0.10471	0.26087
		1.000	-0.10515	0.12243	0.10964	0.13032	-0.44901
			1.000	-0.06878	-0.09540	0.15162	0.47156
				1.000	0.14040	-0.04289	0.04760
					1.000	0.28744	-0.02405
						1.000	-0.05259
							1.000

A graphical output of the absolute value of both correlation matrices is provided

in Fig A.1, where the regions of low and high correlation correspond to the darker and lighter areas respectively.

(a) Masek correlation matrix

(b) Zuo correlation matrix

Figure A.1 Graphical representation of absolute value of correlation matrix

The square region which contains the largest correlation values corresponds to segmentation quality measures 13 through 22, which is the concentric gradient magnitude values R_i as outlined in Sec 3.7. Since each successive measure of R_i examines the same iris region as the previous measure but only incrementally larger in size, it is logical that they would be highly correlated.

In comparison, the remainder of the correlation matrix contains values that are relatively small, indicating that these segmentation quality measures are not significantly correlated. Since the values of correlation are not equal to 0, the variables are not fully independent but still have some linear dependence.