5-2015

Periocular Region-Based Biometric Identification

Philip Miller
Clemson University

PERIOCULAR REGION-BASED BIOMETRIC IDENTIFICATION

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Computer Science

---

by
Philip Miller
May 2015

Accepted by:
Dr. Sophie Jörg, Committee Chair
Dr. Timothy Davis
Dr. D. E. Stevenson
Dr. Andrew Duchowski

Abstract

As biometrics become more prevalent in society, the research area is expected to address an ever widening field of problems and conditions. Traditional biometric modalities and approaches are reaching a state of maturity, and their limits are clearly defined. Since the needs of a biometric system administrator might extend beyond those limits, new modalities and techniques must address such concerns. The goal of the work presented here is to explore the periocular region, the region surrounding the eye, and evaluate its usability and limitations in addressing these concerns. First, a study of the periocular region was performed to examine its feasibility in addressing problems that affect traditional face- and iris-based biometric systems. Second, the physical structure of the periocular region was analyzed to determine the kinds of features found there and how they influence the performance of a biometric recognition system. Third, the use of local appearance based approaches in periocular recognition was explored. Lastly, the knowledge gained from the previous experiments was used to develop a novel feature representation technique that is specific to the periocular region. This work is significant because it provides a novel analysis of the features found in the periocular region and produces a feature extraction method that resulted in higher recognition performance over traditional techniques.

Dedication

This dissertation is dedicated to my wife Kristen and my family. Without their support I would not have had the perseverance to complete this task given the obstacles I have faced.

Acknowledgments

I would like to thank all of the members of my committee for agreeing to serve on such short notice. Many of you have only recently joined my committee and I appreciate your contribution. I would like to especially thank Dr. Sophie Joerg for being the chair of my committee and for her invaluable help in making my dissertation more accessible to those outside of the biometrics research community. I would also like to thank Dr. Damon Woodard for directing my initial research and helping to formulate the thesis and structure of my dissertation. Special thanks also to the Office of the Director of National Intelligence (ODNI), Center for Academic Excellence (CAE) for the multi-university Center for Advanced Studies in Identity Sciences (CASIS), who funded the majority of this research.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

There are three general ways to verify the identity of an individual. These ways are based on something you have, something you know, or something you are. Something you have refers to a token and can come in the form of a key or an ID badge, while something you know refers to a password. The field of biometrics studies the third means of identification: something you are.

In a sense, each person has their own biometric system. Though sometimes unconsciously, people use faces to identify each other. When a person's back is turned people use anthropometrics, body measurements such as height and arm length, to identify a person. When an unseen person is calling out, people use voice to identify the caller. Humans do all of these things subconsciously, yet researchers have a limited understanding of how the brain processes these tasks. The natural human biometric system is a vast set of characteristics used to identify individuals; however, this is not easily replicated by computers. Biometric research attempts to study each characteristic separately to identify individuals.

Through the years, researchers and professionals have proposed different modalities of biometric identifiers, where each modality serves an original purpose or specific situation. For example, Sir William Herschel used handprints to visually distinguish between his employees before paying them [24]. Alphonse Bertillon recorded anthropometrics to discover repeat criminal offenders that would give new aliases at each arrest [42]. Fingerprints quickly replaced anthropometrics because the increased accuracy helped avoid cases of mistaken identity [15]. Similarly, modern technology allows facial recognition systems, that use data in the form of images, to attain similar accuracy to fingerprint identification while using data from a different domain [15]. Iris patterns, originally conceived by ophthalmologist Frank Burch, is a biometric modality that is considered highly accurate [14]. Unlike the characteristics described here, not all human characteristics

make useful biometrics.

There are a number of parameters that define whether a human characteristic is useful at serving as a biometric identifier. Jain et al. [18] suggested that any human physiological characteristic can be used as a biometric identifier as long as it satisfies the requirements of:

- Universality: where the characteristic is present in each individual;
- Distinctiveness: where the characteristic, in any two persons, is sufficiently different;
- Permanence: where the characteristic is sufficiently invariant to change over a period of time;
- Collectability: where the characteristic can be quantitatively measured.

There are also practical considerations that need to be taken into account when building a biometric system from a human characteristic. Jain et al. [18] offered the following additional considerations of:

- Performance: the attainable accuracy of the system, given the operational requirements;
- Acceptability: the extent to which people are comfortable using the biometric identifier;
- Circumvention: the ease at which the system can be spoofed using fraudulent methods.

Once a biometric system is built using a particular biometric identifier, it is evaluated through the two types of biometric problems: identification and authentication. Identification is the problem of choosing the best identity match for a biometric sample from a database of enrolled identities. A real world example includes examining face images of unidentified people walking through the terminals of an airport. The images are compared with biometric data from individuals on a no-fly list to ensure that there are no matches. Authentication is the problem of verifying an identity claim of an individual within a certain degree of certainty. A real world example is when a government official attempts to enter a classified area by entering a pin number into a biometric lock (identity claim). The biometric lock takes a picture of the person's iris and compares it to a previously stored template, before authenticating their identity.

Figure 1.1 shows the flow of the identification and authentication problems and the typical components of a biometric system. The process of setting up a biometric experiment involves enrolling known identities into a biometric database. The first step in accomplishing this is to acquire biometric data. In the familiar case of a face recognition system, data comes in the form of an image of the face. In the second step, feature extraction, a number of different types of approaches can be used that all create a vector representation

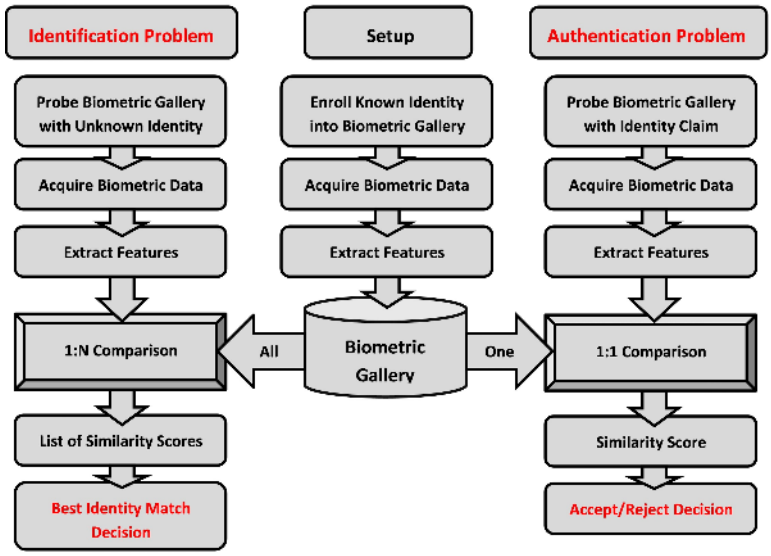


Figure 1.1: The flow of the two biometric problems: Identification and Authentication.

of the data. These first two steps are also a part of each biometric problem. The feature representations that are previously enrolled in a biometric database are known as the gallery or the target set. The feature representations that are compared to the biometric database are known as probes or queries. The third step is to compare feature representations. Typically, the comparisons generate match scores, also known as similarity scores, by the use of a mathematical distance function such as Euclidean distance,

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + \ldots + (p_n - q_n)^2}, \tag{1.1}$$

where p and q are templates representing an image in the probe set and an image in the gallery set that have feature dimensionality of n. The resulting scores are used to make a decision in each of the biometric problems. In an authentication problem, the decision is between accepting and rejecting an identity claim. One probe is compared to one identity in the gallery — a 1:1 comparison. A threshold is typically set so that a score above the threshold gives an accept decision and a score below the threshold gives a reject decision. In an identification problem, the decision is between asserting a true or false match for the system. One probe is compared to the entire gallery — a 1:N comparison. The decision is made by evaluating a sorted ordering of match scores.

There are several ways to visually and numerically evaluate the performance of the identification and authentication problems described above. A Cumulative Match Characteristic (CMC), the cumulative sum of correct identifications in the top x closest matches for each image, gives a visual representation of

the performance of an identification experiment. As seen in Figure 1.2, a given recognition rate (plotted on the y-axis) at Rank x (plotted on the x-axis) corresponds to the percentage of times the true match for an individual's identity is made in the top x ranked images. Rank 1 recognition rates show the successfulness of the system at choosing the best match for an individual. Another example, a Detection Error Tradeoff (DET) seen in Figure 1.3, shows the performance of an authentication experiment. A DET plots the false reject rate (FRR), along the y-axis, against the false accept rate (FAR), along the x-axis. The FRR is the rate at which a biometric system commits a type I error by incorrectly rejecting an identity claim. The FAR is the rate at which a biometric system commits a type II error and incorrectly accepts an identity claim. The single measure that best evaluates the performance shown in a DET is the Equal Error Rate (EER), the rate at which the FRR and FAR are the same. A lower EER corresponds to a more accurate authentication system. A Match Score Distribution (MSD) plot, seen in Figure 1.4, is another way of visually representing the performance of a biometric experiment. An MSD shows the similarity score or match score (plotted on the x-axis) versus the frequency at which the score is found (plotted on the y-axis) of both the genuine score distribution and the imposter score distribution. Recognition errors are found in a system where the distributions overlap. Type I errors are represented in this figure as genuine scores (green) to the right of a threshold and type II errors are imposter scores (red) to the left of the threshold. A single measure that evaluates the separation of the two distributions is called the decidability index or d,

$$d' = \frac{\bar{g} - \bar{i}}{\sqrt{\sigma_g^2 + \sigma_i^2}}, \tag{1.2}$$

where g is the population of genuine scores and i is the population of imposter scores. The decidability index is a ratio of the difference of the two distributions means over a function of their standard deviations. A large d value represents a system where there is little overlap between the two distributions.

The identification and authentication problems are essential to evaluating the performance of a system using a biometric characteristic such as the face and any other unexplored characteristic. These standard evaluation tools can be used to explore a specific area of the field of biometrics as we develop new biometric technology.

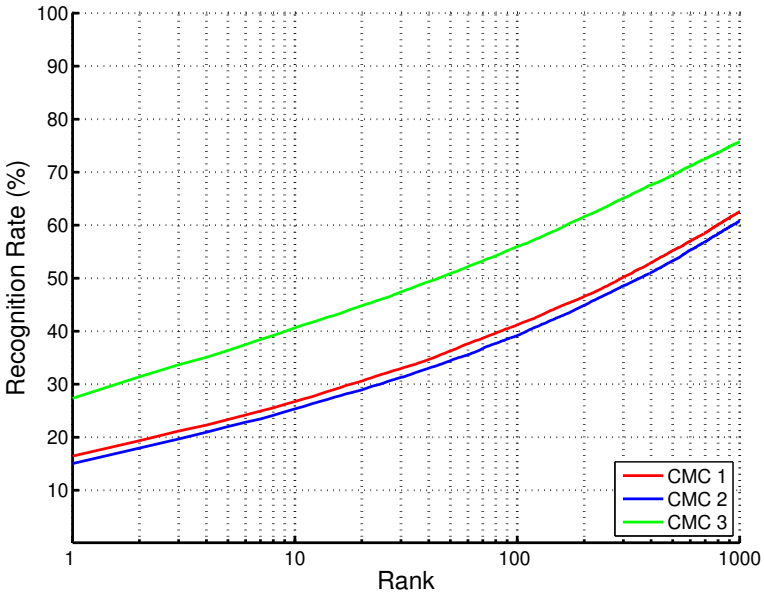


Figure 1.2: An example of a Cumulative Match Characteristic.

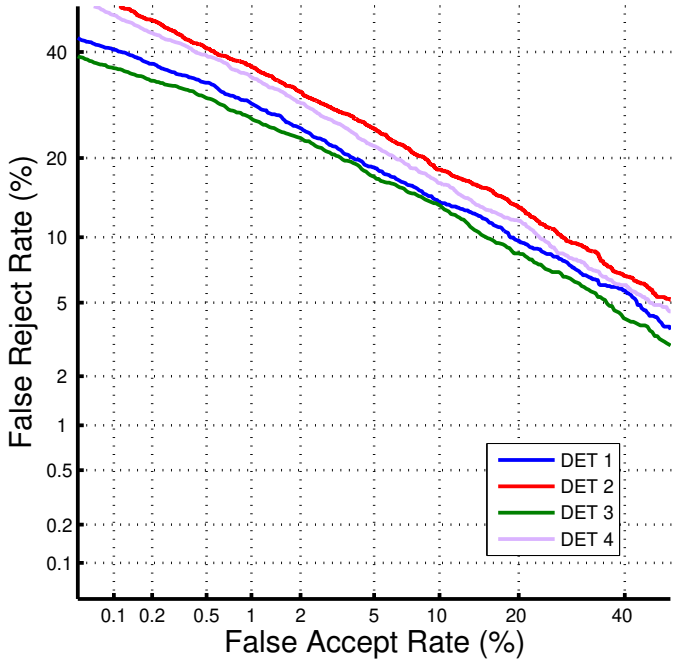


Figure 1.3: An example of a Detection Error Tradeoff.

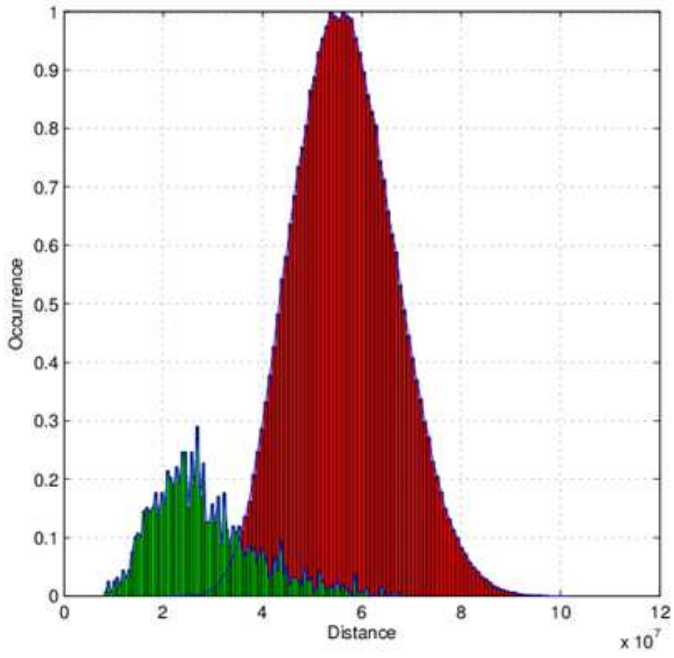


Figure 1.4: An example of a Match Score Distribution.

1.1 The Periocular Region

Biometric research has existed for almost a century and as a result the research regarding early biometric modalities in constrained experimental conditions has matured. The success of biometric research has led to an increased interest in biometric systems for use in atypical situations where the limits of constraints are pushed. Researchers are discovering that traditional face- and iris-based biometric systems may not be suitable for all of the demands of an ever growing field, and new approaches are being considered. The periocular region, a new biometric modality, is the focus of this dissertation.

The periocular region is defined as the area situated around the orbit of the eye, the bony cavity in which the eyeball sits together with its associated muscles, blood vessels, and nerves. In practice this includes two separate, bordering, rectangular regions centered on the eyes that may or may not include the eyebrows. Figure 1.5 shows an example of two periocular region images beside the face image from which they were extracted.

As seen in Table 1.1, the periocular region is very similar, in regards to certain traits, to the face and iris biometric modalities. Table 1.1 lists Jain's seven traits of a biometric identifier and specifically evaluates

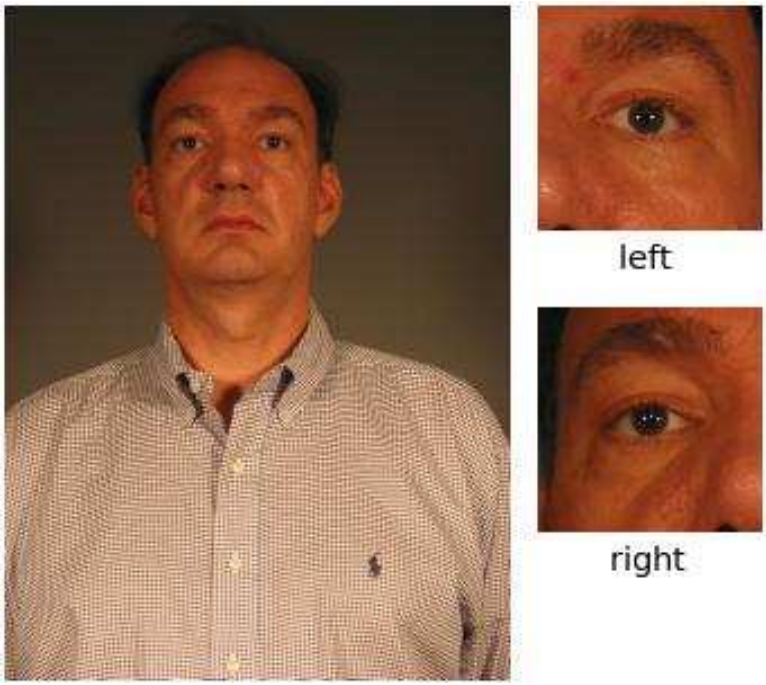


Figure 1.5: Examples of periocular region images obtained from a face image.

	Face	Iris	Periocular
Universality	High	High	High
Distinctiveness	High	High	Undetermined
Permanence	Medium	Very High	Undetermined
Collectability	High	High	High
Performance	High	Very High	Undetermined
Acceptability	Medium	Medium	Medium
Circumvention	Low	Low	Low

Table 1.1: Jain's seven traits of a biometric identifier for the face, iris, and periocular region.

the face, iris, and periocular modalities in their ideal conditions. The periocular region is a sub-region of the face that is present in all faces; therefore, it has a measure of universality similar to the face biometric. I hypothesize that the periocular region also has similar collectability, acceptability, and circumvention to the face biometric because it is found within the same image.

The face and iris biometrics offer high performance. Many commercial biometric systems that employ these modalities boast of almost 100% accuracy. This high performance is attributed to the modalities working extremely well in constrained experimental conditions. As a result of the success of face and iris, researchers are attempting to use these modalities in increasingly non-ideal situations. Through such research, many are discovering that traditional face and iris biometric systems are not meeting the new tasks as well as

Figure 1.6: Examples of a mask (left) and beard (right) that occlude a face.

needed.

There are a few examples where face and iris biometric systems might fail. First among them is occlusion. Common causes for occlusion in biometric data include a beard on a man, makeup on a woman, sunglasses, or a mask covering the face, as seen in Figure 1.6. These obstructive elements prevent the full face from being useful. Similarly, failure can occur in an iris-based system that attempts to read a dilated pupil. The enlarged pupil will subsequently reduce the area of the iris, causing distortion of the area of interest, and result in a failure to confidently make a positive identification. Another problem in iris-based biometric systems is the need for up-close, high-resolution data. This is not usually a problem when the subject is cooperative, but it severely limits the effectiveness of the iris when data is captured from a distance, on the move, or from a non-cooperative source.

Recent research hypothesizes that the periocular region may be able to address some of the failings of face and iris biometric systems. For example, the difficulty of occlusion of a person's face by a beard may be addressed by a new biometic system that utilizes the periocular region when the original face biometric system fails. The periocular region might also be able to assist in the failed iris recognition scenario. If a new biometric system were created that fused the periocular features with the iris features from the non-ideal scenario, the person with the dilated pupils might not be falsely rejected. The periocular region has the potential to handle some situations that would otherwise result in failure in traditional biometric systems. The periocular region would not be of any help if the person is wearing sunglasses, of course, which illustrates the point that the periocular region is not intended to be a complete solution to all of the issues found in other biometric systems.

1.2 Literature Review

The periocular region is a relatively new area of study, so few references were made to it before 2009. This fact does not mean that its proposed usefulness in biometrics research is without origin. The notion that the periocular region could be useful in the adverse situations described earlier grew from four existing works with the face biometric.

The periocular region first appeared in the literature within the context of partial face recognition. Sato *et al.* [46] were interested in discovering if a certain region of the face performed better than others within their biometric system. They considered the partial face regions of the nose, eyes, and ear and collected the data of these partial face regions themselves. In total, their work utilized 720 images per partial face region from 120 subjects and the size of the eye images was 12x8 pixels. The number of images and their size are low relative to experimental expectations of today. The authors used raw pixels as features to build a radial basis function (RBF) network. The RBF network was used as a classifier to determine if a query was enrolled in the system. Using eye images, the system achieved a "recognition rate" of 92%. "Recognition rate" was defined as the rate at which the system correctly determines if a query is enrolled in the system and is not required to identify the individual. Though the experimentation was not rigorous, the periocular region was studied and yielded higher performance that the other regions. The performance of the periocular region was also comparable to the full face. This early work suggested that the periocular region could be a valuable region of the face, containing significant discriminative power.

Another work, by Kamgar-Parsi *et al.* [22], looked at synthetic images to aid facial recognition systems. This work was motivated by the poor performance of systems when limited training data per subject was available. The authors proposed a system where synthetic images were created to capture variations in appearances. Their hypothesis was that the variations in these synthetic images would help improve performance. Each of the synthetic images was created by manipulating the eye region of face images to approximate varying appearances. To do so, they created a number of operators that would produce a desired appearance when applied to an image. These operators were named "Lower eyebrow", "Raise eyebrow", "Tilt eyebrow", "Arch eyebrow", "Lower Upper eyelid", "Raise Upper eyelid", "Raise Lower eyelid", "Lower Lower eyelid", and "Cast Shadows". The authors used face images from the Facial Recognition Technology (FERET) dataset [41] as well as images they personally collected. Images were compared by taking the sum of squared differences of the raw pixels. When the synthetic images were used as enrolled members of the biometric database, recognition performance reached 100% accuracy in all tests. The authors report that their work "suggests that

the eye is rich in discriminative information".

The work of Savvides *et al.* [47] also looked at partial face recognition and its fusion with holistic face recognition as part of a larger work. Three equal sized horizontal segments of the face (eyes, nose, and mouth) were considered and results showed much higher performance when using the eye region images than the nose or mouth images. The authors explained that the purpose of this work was to "observe which parts of the face provide the best discrimination information". Within the context of this approach the eye region has the most discriminating information. One interesting note of this work is that it utilized the Facial Recognition Grand Challenge (FRGC) Experiment 4 dataset which consists of 12,776 images from 222 people for training and 24,042 images from 466 people for testing [40]. This dataset is both public and much larger than others before it. Many different holistic subspace projection methods were applied to the partial face images such as Principle Component Analysis (PCA) or eigenfaces, Gram-Schmidt Linear Discriminant Analysis (GSLDA), class-dependence feature analysis (CFA), and Kernel approaches. This work was much more experimentally rigorous than previous works that used the periocular region. Therefore, it offered the best support that the periocular region has high discriminative power within the face compared to any previous work.

Teo *et al.* [51] also looked at partial face recognition, highlighting the eye region, and gave comparisons to full face recognition. This work had a very similar concept to Savvides *et al.* [47], but with two interesting additions. First, Teo *et al.* [51] used non-negative matrix factorization (NMF)-based techniques to extract features from the data. Using this technique suggested that the periocular region's discriminative power is not found only in the context of a single type of approach. Second, it motivated the use of the eye region by discussing how non-cooperative subjects can cause face occlusion and posited that the occlusion is least likely to occur in the eye region. This work used the smaller Faces-94 Essex University Face Database which contains only 3060 images from 153 subjects [49].

Although there was little reference to the periocular region in the literature, there is much to be learned from previous work. These works suggest a number of different techniques that may be used to extract features from the periocular region. Among these methods were RBF networks, eigenfaces, GSLDA, CFA, kernel approaches, and NMF. These are all examples of global appearance-based approaches which were frequently used at the time of publication. The use of each of these techniques is significant, as not every technique works well with every modality. The results of these experiments suggest that many of the techniques that are used with great effect in conjunction with face recognition can also be used with the periocular region. The previous works also used a wide variety of data for their experiments. For the most part, the corpus used in each work was a small collection of controlled images created by the authors of the work.

These small collections, especially in subject count, cannot be said to represent a real world environment but do provide enough variability to suggest that the periocular region is worth further investigation.

Though periocular recognition is a new field of study, recent works have addressed many interesting aspects of it. Park *et al.* studied the effects, on a periocular based recognition system, that different segmentation techniques have, as well as comparisons of performance with the face [37, 38]. They provided a feasibility study and the first attempt at a solely periocular-based recognition system. The method in this research implemented Local Binary Patterns and Gradient Orientation Historgrams as feature extraction techniques for the periocular region data. These local appearance based approaches became the most widely used feature representation in periocular research. Miller *et al.* provided similar analysis with a different dataset and different feature extraction methodology [31].

Lyle *et al.* and Merkow *et al.* both studied the use of the periocular region in soft biometric recognition [27, 28, 29]. Both works showed that the periocular region was capable of classifying a periocular image based on gender and/or ethnicity in their experiments. Neither work offered anything new in terms of feature extraction as Lyle *et al.* followed Miller *et al.*'s approach and Merkow *et al.* do not explain how they use Local Binary Patterns (though it can be inferred from the figures that they are using Ojala's original technique [34]).

Woodard *et al.*, Santos and Hoyle, and Tan *et al.* explored the fusion of iris and periocular features [45, 50, 53, 54]. The fusion of the periocular region with the iris was explored soon after the periocular region was introduced as it offers a natural way of boosting the performance of an iris recognition system that would fail due to poor quality data. Each of these works showed a significant performance increase in their experiments as a result of fusing periocular region features to iris features. Woodard *et al.*'s approach to extracting periocular features was based on the approach by Miller *et al.* while both Santos and Hoyle and Tan *et al.* borrowed from Park *et al.*'s approach.

Miller *et al.* considered the effects of data quality in a periocular biometric system [30]. This research used Miller *et al.*'s approach to feature extraction and analyzed the performance of a biometric system when the data varied in terms of camera focus, image size, or illumination pattern. For the experiments in this research, the periocular region greatly outperformed the face when the difference in data quality between gallery and probe images was high.

1.3 Problem Statement and Impact

There are many conditions in which traditional biometric modalities might fail or provide poor discriminative performance. Some of these conditions that affect the biometric data were listed in Section 1.1. Sometimes these conditions can not be overcome or avoided and the biometric system must make use of the data at hand.

The goal of this dissertation is to explore the usefulness of the periocular region in addressing these shortcomings. To accomplish this goal a detailed study of the periocular region will be performed that examines the performance of a periocular-based biometric system in the presence of facial occlusion, various poor quality iris data scenarios, and other data quality concerns related to image acquisition.

Through the lessons learned while experimenting with the periocular region, a novel approach for periocular feature representation will be proposed. None of the techniques for periocular feature extraction consider how each subregion of the periocular region affects performance or whether there is a specific physical feature, like the pattern of the eyelashes or wrinkles in the corners of the eye, that contributes the most discriminative information. Periocular-based recognition is still relatively new and many of the approaches used for it are very simple adaptations from other biometric modalities. A more detailed study of the periocular region is needed in order to develop algorithms specific to the periocular region.

This work is significant in a number of ways. No previous work has tried to use physical aspects of the periocular region to optimize performance. In each of the previous works, features are extracted from the periocular region using traditional approaches that give no special consideration to whether the features come from the skin around the eye, the eyebrow, or any other sub-region of the periocular region. I hypothesize that if the physical aspects of the periocular region are allowed to influence the feature selection, the overall performance will increase. Additionally, no previous work explores the performance of multiple approaches in sub-parts of the periocular region. Addressing these points will give the biometric research community a better understanding of the periocular region by contributing to a model for the most discriminative subregions of the periocular region. This knowledge may help guide future research in part-based periocular biometric algorithms as well as aid in optimization efforts.

1.4 Dissertation Outline

This dissertation is organized as follows. Chapter 2 will introduce a basic approach to periocular recognition to use it to examine some basic aspects of the periocular region such as the performance of sub-regions of the periocular region, the effect of the iris on performance, and the affect the quality of the biometric image has on a periocular-based biometric system. Chapter 3 will analyze the physical structure of the periocular region and take a closer look at its sub-regions in order to determine what kinds of features are present in each. Chapter 4 will take a look at local appearance based approaches and their use in periocular recognition with the intent of modifying them for optimal use in periocular recognition. The lessons learned from the two previous chapters will be combined to develop a novel periocular feature extraction algorithm in Chapter 5.

Chapter 2

An examination of common feature extraction techniques using the periocular region

2.1 Introduction

As mentioned in Section 1.1, traditional face and iris biometric systems can fail in the presence of non-ideal data. The periocular region may be able to address the failings of these biometric systems by allowing otherwise non-ideal data to be used within an alternate protocol. Occlusion is a common cause of failure in face- and iris-based biometric systems. Some examples of occlusion in facial images include facial hair, cosmetics, and closed eyelids. These obstructions prevent typical face- and iris-based biometric systems from being able to collect discriminative features from the entire biometric image. The presence of a beard or closed eyelids does not occlude the periocular region. Though it is likely that using only the periocular region will result in lower recognition performance than using the whole face, because the periocular region uses inherently less information, the whole face is not available in the presence of facial occlusion. In this scenario, the usefulness of the periocular region is determined in comparison to the complete failure of the facial recognition system. Similarly, the periocular region may also be used when an iris recognition system is hampered by images with closed eyelids. The periocular region is collected with the iris and can be used as a substitute when the iris is obviously obscured.

The periocular region can be seen as a sub-region of the face. Therefore, one could expect common facial feature extraction algorithms to produce discriminative feature sets from the periocular region as they do from the face. However, the degree of discrimination of periocular features remains to be seen. The early works described in Section 1.2 indicate that the periocular region is the most discriminative region of the face. They also suggest that the periocular region might be nearly as discriminative as the entire face. Unfortunately, the experiments performed in those works were rudimentary. Even if a periocular-based biometric experiment can be shown to have similar performance to a face-based biometric experiment, there is still reason to explore why this would be the case. An extensive look into the features themselves and the patterns they produce in the face and periocular region need to be performed.

The quality of data plays an influential role in the performance of biometric recognition systems. Early research of popular modalities, such as the face and iris, use high quality data in constrained environments. As expected, techniques tested under these constraints yield higher performance than when using low quality data. Scenarios exist where high quality data can be acquired reliably in real-world applications; however, many others exist where reliable data collection is not possible. Reacquiring biometric data when the first acquisition yields poor quality data is not always possible, therefore research is needed to make non-ideal data useful. Non-ideal data can be characterized by either subject-influence or environment-influence. Occlusion, pose variation, facial expression and pupil dilation are causes of non-ideal data influenced by the subject. Environmental influences include uncontrolled lighting, image focus, and image resolution (pixel count) variation.

The periocular region has been introduced as a biometric that is capable of addressing some subject-influenced concerns. For example, in the case of partially occluded face images, the periocular region may be used to identify individuals. Also, when iris recognition fails due to pupil dilation, the periocular region may be used to increase performance. The periocular region may be used in these situations because it is captured along with the face or iris without the need for additional capture equipment or procedures. However, subject-influenced concerns are not independent of environment-influenced concerns; environment-influenced concerns will likely appear in certain subject-influenced non-ideal data. If a subject is actively trying to avoid biometric recognition by hiding their face or running away from the camera then the environment-influenced concerns of blur and inconsistent image resolution will likely be present.

The goals of this chapter are as follows: (1) to examine common facial recognition data and common facial feature extraction algorithms as they are implemented in a periocular-based biometric system, (2) to determine where in the face and periocular region the most discriminative features are extracted, and

Figure 2.1: An example recording session from FRGC [40].

(3) to assess the robustness of the periocular region with respect to the environment-influenced concerns a periocular-based biometric system will likely encounter.

2.2 Data

Two different datasets were selected for use in this chapter: The Facial Recognition Grand Challenge (FRGC) Database and The Facial Recognition Technology (FERET) Database.

The FRGC database [40] consists of high resolution color images of a large number of subjects mostly between ages 18 and 22, collected over a two year period from multiple recording sessions involving controlled and uncontrolled lighting conditions, and with an expression and without. A recording session is the set of all images of a subject taken each time the subject's biometric data is collected. A typical FRGC recording session consists of four frontal face, controlled lighting still images, two frontal face, uncontrolled lighting still images, and one three-dimensional image. Figure 2.1 shows a set of controlled lighting images for one recording session. The controlled lighting images were taken in a studio setting (two or three studio lights) and with two facial expressions (smiling and neutral). In controlled conditions, the distance between the subject and the camera is approximately the same. The still images were taken with a 4 Megapixel Canon PowerShot G2 and have a pixel resolution of either 1704×2272 or 1200×1600 pixels. The images are stored in JPEG format with storage sizes ranging from 1.2 Mbytes to 3.1 Mbytes.

FRGC Experiment 1 is an experimental protocol and data subset that is widely used to compare different biometric recognition methods. FRGC Experiment 1 is a set of 16,029 still, high resolution, frontal face images taken under controlled lighting conditions. It was chosen for this work because the large face images will lead to relatively large periocular region images. FRGC Experiment 1 measures performance on the

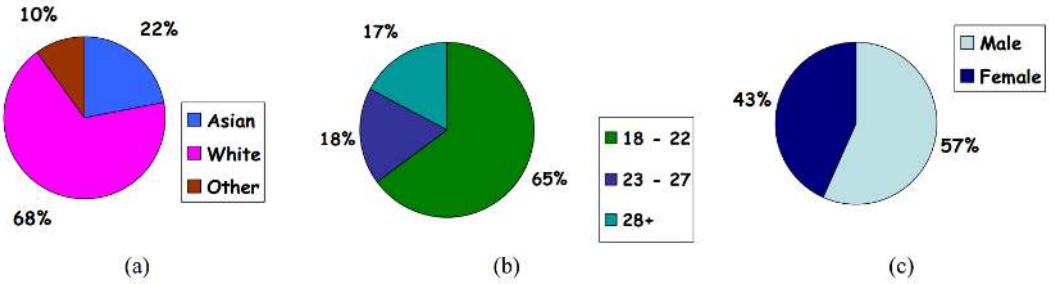


Figure 2.2: Demographics of FRGC validation partition by (a) race, (b) age, and (c) sex [40].

classic face recognition problem: recognition from frontal facial images taken under controlled illumination. FRGC Experiment 1 data is divided into training and validation partitions. Images in the validation partition were collected during the 2003-2004 academic year and cover 466 subjects from 4,007 subject sessions. The demographics of the validation partition are given in Figure 2.2. The training set consists of an additional 12,776 images taken from 226 of the same subjects in a single recording session in the Spring 2003 semester and is only used with feature extraction algorithms that require a trained model.

This work also uses the FRGC Experiment 4 data subset, a collection of images taken under uncontrolled lighting situations. Experiment 4 is a set of 8,014 still, high resolution, frontal face images taken from the same 466 subjects as Experiment 1. Images were taken either in an indoor hallway with only the overhead ceiling lights, or outside with only the sun illuminating the face. The Experiment 4 protocol calls for using the Experiment 1 dataset as the gallery set and the Experiment 4 set as the probe set.

The FERET database [41] consists of gray-scale and color images of faces captured between 1993 and 1997. The mission of FERET was to assist researchers in the development of early facial recognition systems by providing the best set of test data available at the time. Many of the subjects present in the FERET dataset were photographed in many different poses and with different facial expressions. The experiments of this chapter only make use of frontal face images. This subset consists of 1,980 frontal face images taken from 990 subjects. An example of some of the FERET images can be seen in Figure 2.3.

2.3 Method

The basic biometric experiment template followed by all experiments in this chapter is composed of the following steps: data preprocessing, testing/training partitioning, gallery/probe partitioning, feature extraction, feature comparison, and computation of performance statistics.

Figure 2.3: Example images from the FERET database [41].

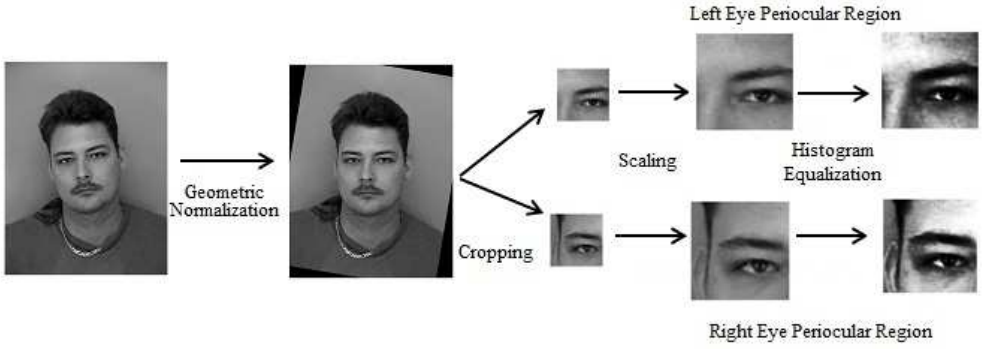


Figure 2.4: The flow of preprocessing steps applied to a face image to extract its periocular regions.

2.3.1 Data Preprocessing

Image preprocessing is an important step in the operation of a biometric system. All images from each subset discussed in Section 2.2 undergo the same process described below. These standard face-based biometric data preprocessing steps convert a raw color facial image into the preprocessed periocular images used in the experiments of this section. Figure 2.4 shows the flow of the preprocessing steps.

2.3.1.1 Geometric Normalization

The first step in preprocessing a facial biometric image is geometric normalization, which aligns all of the images in the dataset to the same spacial coordinates. An image of a face can be rotated along three axes: an axis from the face to the observer (in-plane rotation), a vertical axis dissecting the face (out-of-plane rotation), and a horizontal axis dissecting the face (out-of-plane rotation). The most commonly used points to correct in-plane rotation in facial images are the eye centers. Therefore the location of the eye centers must

be known to perform correction. Eye center locations for FRGC and FERET were provided by the curators of the databases.

Using the eye centers, a rotation matrix R was generated as follows and applied to an image to correct in-place rotation

$$R = \begin{bmatrix} cos(\theta) & sin(\theta) \\ -sin(\theta) & cos(\theta) \end{bmatrix}$$

where

$$\theta = \tan^{-1}\left(\frac{|r_y - l_y|}{|l_x - r_x|}\right)$$

l is the pixel location of the left eye, and r is the pixel location of the right eye.

2.3.1.2 Periocular Region Extraction

The next preprocessing step is to extract periocular region images from normalized and equalized facial images. This is accomplished by placing a square bounding box around each eye, centered on the post geometric normalization eye center locations. The length of the sides of this square bounding box is equal to the distance between the two eye center coordinates, also known as the interocular distance. In an average face, a single bounding box will cover the region from the center of the face to the ear and from the bottom of the nose to the middle of the forehead. The resulting periocular images are then re-sized to 200×200 pixels for use in the experiments of this chapter. 200×200 pixels was chosen because it is the size of the smallest periocular region extracted from the face images used in this chapter.

2.3.1.3 Histogram Equalization

The last step is histogram equalization, where the contrast of the image is enhanced. This step normalizes the relative illumination levels between images by changing pixel intensity values. The bins of a histogram of an image that has been equalized will contain approximately the same number of pixels. This process ensures that images that are darker or lighter than a neutral image are altered so that they appear to have been taken under the same lighting conditions.

Histogram equalization is comprised of the following three steps: calculating the histogram of an image, calculating the cumulative distribution function of the histogram, and modifying the necessary pixels. The histogram of an image is defined as the sum of the occurrences of each intensity value and is calculated by

$$H(I)[k] = \sum_{i=0}^{w} \sum_{j=0}^{h} \left\{ \begin{array}{ll} 1 & \text{if I(i,j) = k} \\ 0 & \text{otherwise.} \end{array} \right\}$$

where I is an image of $w \times h$ pixels and k is an intensity value. The cumulative distribution function (CDF) of a histogram is defined as the cumulative sum of the values of the bins of a histogram. A transform is created that will produce a new image such that the CDF of that new image is approximately linear. The application of this transform to the original image results in an image with an equalized histogram.

2.3.2 Dataset Partitioning

As mentioned in Section 2.2, the FRGC dataset comes with images designated for training biometric systems that require a training set. Therefore, no special considerations were needed to partition the dataset into training and testing subsets. The FERET dataset was not very large, so a decision was made not to perform any experiments that required a training set with the FERET data.

There is also no need for a probe/gallery partition in FRGC Experiment 1 due to the suggested experimental protocol given by the organizers of the FRGC dataset. FRGC Experiment 1 is intended to compare every image to every other image. For experiments using the FRGC Experiment 4 data, the images in Experiment 4 are used as the probe set while images from the Experiment 1 set are used as the gallery set. For FERET experiments, the earliest collected image per subject was chosen as the gallery and the other image chosen as the probe.

2.3.3 Feature Extraction

The feature extraction step in the basic biometric experiment is the main point of distinction between any two different experiments in this dissertation. Features are extracted from each image in the dataset and from each partition in the set. Each feature extraction technique transforms a two dimensional image into a 1 dimensional feature vector through its own unique process.

The periocular region is essentially a sub-part of the face; therefore, it should be expected that common facial feature representations would provide a useful means of classifying individuals using the periocular region. Classes of facial feature representations include local appearance, keypoint-based, and holistic. The experiments in this chapter will use basic approaches from each class of feature representations in an attempt to provide comparisons between the ability of the classes of feature representations to

produce discriminating features from the periocular region alone. Among the local appearance-based feature representations are Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG), Local Phase Quantization (LPQ), and Weber Local Descriptor (WLD). Scale Invariant Feature Transform (SIFT) and Speeded Up Robust Features (SURF) were chosen as the keypoint-based representations. For the holistic approach, Eigenfaces was chosen. Each technique is described below.

Each of these algorithms represents the basis for their class of feature representation techniques. While other, more sophisticated, methods might exist that would provide better performance, the purpose of these experiments is not to find the best performance for a given set of data but instead to discover how each class might respond to the problem of occlusion in facial images. This knowledge will guide later work with the periocular region in how to best develop an algorithm specific to the periocular regions' needs.

2.3.3.1 Local Appearance-Based Approaches

Local Appearance-Based Feature Extraction Methods are a class of feature extraction techniques that accumulate statistics within local neighborhoods around each pixel of an image. These statistics include the occurrences of certain textures, patterns, and information and are typically stored in a one-dimensional feature vector. These experiments make use of four local appearance-based feature extraction methods: Local Binary Patterns, Histogram of Oriented Gradients, Local Phase Quantization, and Weber Local Descriptors.

There are some differences in the way these types of features are used in biometric applications, as opposed to texture classification. Typical data used in texture classification experiments contain a single texture that is uniform and repeated across the image, however, images of the face used in biometric experiments do not fit this description. Since the face and other biometrics modalities have more physical variability than texture patterns, these features are extracted from subsections of the image, called patches, so that there is a distinction between the textures present in different parts of a biometric modality. The feature vectors from each patch are concatenated together to form the final feature vector.

Let I be a preprocessed image that is divided into N non-overlapping rectangular patches of M pixels each. The overall feature representation of the image is given by a vector $\mathscr{T}(I) = \left\{T^{(1)}, \ldots, T^{(N)}\right\}$, where $T^{(1)}, \ldots, T^{(N)}$ are the feature vectors computed from the N patches. In this work, all patches are 20×20 pixels and are placed in a non-overlapping pattern starting in the top left corner of the image. 20×20 pixel patches were chosen because this is the smallest size patch possible. Some of the methods detailed below require that the local neighborhoods around a pixel have at least a 9 pixel radius.

Local Binary Patterns Local binary patterns (LBP) is a texture classification method that was developed by Ojala *et al.* [33]. LBP accumulates texture information from an image into a feature vector. This is accomplished by labeling pixels with a binary number that is a function of placing a threshold on the neighborhood around each pixel. A histogram of these values becomes the output feature vector. Due to the success of LBP as a texture classification method, it has been used extensively for both facial recognition [2, 57, 43, 7, 48] and periocular recognition [38, 31, 30, 53, 5, 55, 54, 37, 20, 17, 19, 21, 27, 29, 28, 45].

A local binary pattern is a numerical representation of a pixel that encodes intensity changes of a local neighborhood of P pixels found on a circle of radius R around a pixel of interest. A pixel of interest is any single pixel in an image I for which a texture representation is desired. The LBP score at that pixel is a function of the changes of intensity patterns in the local neighborhood. Let a pixel of interest be represented by x_k. Then the LBP score at this location is given by

$$LBP(x_k) = \sum_{j=0}^{P-1} s\left(I(x_j) - I(x_k)\right) 2^j, \tag{2.1}$$

where

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}, \tag{2.2}$$

x_j is a pixel on the circle around x_k, and $I(x_j)$ is the intensity value of the pixel at x_j.

The occurrences of each LBP score are accumulated into a 59 bin histogram ($size(T^{(i)}) = P(P-1)+3$, where $P = 8$) which collects uniform patterns, as they are called by Ahonen *et al.* [1]. Ahonen *et al.* defines a uniform local binary pattern as one that contains at most two bitwise transitions in the binary string. In this work, all LBP calculations are made from 8 pixels along a circle with a radius of 1 pixel. These parameters were used because the original works using LBP used these parameters. The purpose of the experiments presented in this chapter is not to find the optimal LBP parameters, so there was no reason to change the parameters.

Histogram of Oriented Gradients Histogram of Oriented Gradients (HOG) is an edge and gradient based feature descriptor originally developed by Dalal and Triggs [9] to detect humans in images. HOG is a local appearance-based approach that counts the occurrences of different gradient orientations in localized portions of an image. Even though HOG was originally intended for object detection, it has been used for both facial and periocular recognition [38, 37, 44, 28, 48, 11]. HOG is a simple technique that, with the exception of

object orientation, is invariant to geometric and photometric transformation.

A modified HOG algorithm is used for extracting features from the periocular region. The first step is computing the gradient of the image using a Prewitt convolution kernel. The gradient magnitude, G_{mag}, and gradient angle, G_{ang}, are computed from the image gradient, G_x in the horizontal direction and G_y in the vertical direction, as defined by

$$G_{mag} = \sqrt{G_x^2 + G_y^2}, \tag{2.3}$$

$$G_{ang} = atan2(G_y, G_x). \tag{2.4}$$

In the next step, the values of G_{mag} and G_{ang} at each pixel location are accumulated into a histogram. Each bin of the histogram represents an evenly spaced segment of possible gradient orientations. A 12 bin histogram is used in the experiments in this dissertation, so a pixel with a G_{mag} between 0 and $\pi/6$ would correspond to the first histogram bin. For each pixel, P, the orientation bin corresponding to $G_{ang}(P)$ is incremented by $G_{mag}(P)$.

Local Phase Quantization Local Phase Quantization (LPQ) is a texture descriptor recently presented by Ojansivu *et al.* [35]. This method quantizes the phase information of a discrete Fourier transform (DFT) in patch-sized neighborhoods of an image. The main strength of this local appearance-based feature extraction method is that it is proposed to be robust to image blurring. Though originally used for texture classification in the presence of blur, LPQ has also been used for facial recognition [3]. Like LBP and HOG, the resulting LPQ codes are compiled into a histogram. The final LPQ code is a 256 bin histogram where the quantized DFT phase coefficients are accumulated after binary coding.

In LPQ the local spectra at a pixel x is computed from a short-term Fourier transform defined as

$$F(u,x) = \sum_{y \in P_x} f(x-y) e^{-j2\pi u^T y}, \tag{2.5}$$

where P_x is a pixel in a $M \times M$ neighborhood around x. The local Fourier coefficients are computed at frequency points $u_1 = [a,0]^T$, $u_2 = [0,a]^T$, $u_3 = [a,a]^T$, $u_4 = [a,-a]^T$, where a is $1/M$. The phase portion of

the Fourier coefficients is defined as the sign of the real and imaginary components of $F(u,x)$ given by

$$q_j(x) = \begin{cases} 1, & if q_j(x) \geq 0 \\ 0, & otherwise \end{cases} \tag{2.6}$$

The LPQ score is the binary coding of the eight binary coefficients $q_j(x)$. In this dissertation, all LPQ calculations were made on a 9×9 pixel window.

Weber Local Descriptor Weber Local Descriptor (WLD) is a texture descriptor developed by Chen *et al.* [8]. It was inspired by Weber's Law, a psychological law that refers to the perception of change in a signal. The law states that the change in a signal that will be just noticeable is proportional to the magnitude of the original signal. WLD is concerned with the ratio between the intensity value of a pixel and the relative intensity differences of the pixel to its neighbors, also called the differential excitation, and the gradient orientation of a pixel. Chen *et al.* initially used it for texture classification and human face detection. Like each of the feature extraction techniques discussed in this section, the WLD feature vector is a histogram of the occurrences of each excitation and orientation.

Let the differential excitation at a pixel, x_a, be defined as

$$\sigma = arctan\left[\sum_{i=0}^{p-1}\left(\frac{x_i - x_a}{x_a}\right)\right], \tag{2.7}$$

where x_i corresponds to the ith neighbor of x_a and p is the number of neighboring pixels. Let the orientation of a pixel, $x(i,j)$, be defined as

$$\theta = arctan\left(\frac{x(i+1,j) - x(i-1,j)}{x(i,j-1) - x(i,j+1)}\right). \tag{2.8}$$

The WLD descriptor for an image is defined as

$$WLD(c,t) = \sum_{i=0}^{n}\sum_{j=0}^{m}\begin{cases} 1, & \lceil(\sigma \cdot \frac{180}{\pi})/360 \cdot \phi\rceil = c \;\;\&\&\;\; \lceil(\theta \cdot \frac{180}{\pi})/360 \cdot \xi\rceil = t \\ 0, & otherwise \end{cases}, \tag{2.9}$$

where c is the excitation variable, ϕ is the maximum number of excitations, t is the orientation variable, ξ is the maximum number of orientation, n is the rows of pixels in an image, and m is the columns of pixels in an image. In this dissertation, all WLD calculations were made using 14 excitation and 14 orientation levels, so

the WLD feature vector would be comprised of 196 elements.

2.3.3.2 Keypoint-Based Approaches

Keypoint-Based Approaches are the class of feature extraction techniques that find interest points within an image. These interest points are typically maxima or minima of filter responses applied to the image. Two different keypoint-based approaches are used in this dissertation. These approaches are Scale Invariant Feature Transform (SIFT) and Speeded Up Robust Features (SURF).

Scale Invariant Feature Transform Lowe [25], the developer of Scale Invariant Feature Transform (SIFT), states that SIFT is an approach for detecting and extracting local feature descriptors that are reasonably invariant to changes in illumination, image noise, rotation, scaling and small changes in viewpoint. Interest points generated by the SIFT technique correspond to local extrema of Difference of Gaussians (DoG) filters at varying scales. It was originally developed to be used in image matching problems and is reported to be invariant to uniform scaling, orientation, and partially invariant to affine distortion and illumination changes. For each keypoint location, the keypoint descriptor is defined as a histogram of orientations weighted by the gradient magnitude computed from a local window of 8×8 pixels around the keypoint location. SIFT has been used numerous times for facial recognition [6, 26, 23] and periocular recognition [38, 37, 39, 44, 45, 50].

Speeded Up Robust Features Originally inspired by SIFT, Speeded Up Robust Feature (SURF) descriptors are also a popular keypoint-based method. It was developed by Bay *et al.* [4] and claims to be faster and more robust to different transforms of an image than SIFT. SURF keypoints are found from maxima in the determinant of the Hessian matrix of images, a matrix made up of the convolution of the Gaussian second order derivatives with the image. Along with the keypoint's locations, an orientation is assigned to the point and the descriptor is based on the sum of Haar wavelet responses. SURF was originally used for object recognition and has also been used in facial recognition contexts over the years [12, 13, 56].

2.3.3.3 Holistic Approaches

Eigenfaces is perhaps the most commonly known facial recognition algorithm [52]. It uses the mathematical procedure of principle component analysis to produce a low-dimensional representation of a face from a set of higher-dimensional training images. Unlike the other algorithms, Eigenfaces is a holistic approach to feature representation because it considers the whole face at once.

2.3.4 Feature Comparison and Computation of Performance Statistics

The FRGC experimental protocol suggests the use of distance measurement to determine the similarity between two feature vectors. Two different feature vectors with the closest measure of difference within a set of vectors are said to be the closest match. In an identification experiment the feature vector of a probe image is compared to every feature vector in the gallery set and the Top-N results are returned. In a verification experiment the feature vector of a probe image is compared to every feature vector in the gallery set and those that are above a threshold are declared to be a match. The results reported for each experiment in this dissertation are generated using the cityblock distance metric,

$$d(p,q) = \sum_{i=1}^{n} |p_i - q_i|, \tag{2.10}$$

where p and q are feature vectors extracted from two images that have a feature dimensionality of n.

There are other ways of comparing features that would likely provide better performance, such as using a trained classifier like a Support Vector Machine (SVM). Trained classifiers have restrictions that make them unusable with the FRGC and FERET data that is available for these experiments. SVMs work best when there are multiple images per subject in the gallery set to train on. The FERET data only has 1 image per subject available in the gallery set. The FRGC experimental protocol calls for each image to be compared to every other image. With SVMs there must be a clear distinction between gallery and probe because training the classifier on test data results in unrealistically high performance.

Performance statistics for the experiments presented in this dissertation can be given in the form of a Detection Error Trade-off (DET) as well as Rank-1 recognition rate, Equal Error Rate (EER), Verification Rate (VR) at 0.1% False Accept Rate (FAR), and D', all of which are discussed in Chapter 1.

2.4 Baseline experiments and examination of features

The first step that must be performed is to establish a baseline of performance for a periocular-based biometric experiment. The FRGC Experiment 1 protocol provides an excellent means of establishing that baseline. Using the Experiment 1 protocol, a comparison can be made between face and periocular performance and from there the feature sets can be evaluated.

The authors of the FRGC protocol suggest that the verification rate at 0.1% false accept rate be used to compare the performance between two experiments using the Experiment 1 protocol. This is mostly due

to the structure of the experiment where there are typically 30 images per subject. Rank-1 recognition results are usually greater than 99% for experiments using Experiment 1 regardless of how well the biometric system does in other tasks because only 1 of 30 possible true matches needs to have the best match score. Therefore, the Rank-1 recognition rate of this experiment is a poor differentiator of performance. In analyzing the performance of FRGC experiments presented in this chapter, the attainable rate of true matches, given a low rate of false matches, will be presented as opposed to the likelihood that a true match will have the highest match score for a given query. Equal Error Rate can also be used as it is a more commonly used extension of the same measure of performance. Additionally, D' shows the separability between the similarity score distributions of the set of true matches and false matches.

A total of 21 baseline experiments, seven feature extraction methods applied to three image segments, were conducted using the FRGC Experiment 1 dataset. Four performance metrics were given for each of the experiments. Table 2.1 shows the performance statistics for the experiments using the FRGC Experiment 1 dataset. Rank-1, EER, VR at 0.1% FAR, and D' are given for each feature extraction method used on left eye, right eye, and face images. Figures 2.5 - 2.11 show the DET of the same experiments.

The results indicate that the experiments using features from the face perform better than experiments using features from the periocular region in most cases. This is somewhat intuitive, regardless of the feature extraction algorithm or the dataset used, because there is more information available in the face than there is in the periocular region, and the face contains all of the periocular region. In the experiment that use HOG features, the right eye periocular images produce a better VR @ 0.1% FAR than the face images. Although the periocular region produces a better performance result in this specific experiment, it is not the best performer overall. The best performance comes from using LPQ features taken from the face images. The point of emphasis of these experiments is not whether the periocular region performed better than the face but to observe if any particular feature extraction technique produces greatly different performance results. If so, this could indicate that the periocular region and the face do not produce similarly discriminative features using a particular feature representation. For these FRGC experiments, the greatest decrease in performance when comparing face and periocular images is seen with SURF features, where the decrease is 4.8237%. There were three feature extraction techniques that resulted in greater performance for the periocular region: HOG, WLD, and Eigenfaces. Even though WLD and Eigenfaces produce better performance results when using periocular region images as opposed to face images, their potential usefulness is diminished by the relatively low performance compared to other feature extraction methods. The algorithm with the best overall performance for both modalities was LPQ and it showed a decrease in performance of 3.2634% when

	Rank-1	EER	VR @ 0.1% FAR	D'
LBP				
Left Eye	99.7068	8.8323	64.8841	2.7331
Right Eye	99.7005	8.2016	69.7151	2.8250
Face	99.9189	7.0848	69.9886	2.9429
HOG				
Left Eye	99.6069	8.0829	69.6951	2.8350
Right Eye	99.6444	7.5245	72.2473	2.9378
Face	99.8815	6.7912	70.3877	2.9692
LPQ				
Left Eye	99.7692	7.1183	75.9181	2.8654
Right Eye	99.7816	6.7227	76.6574	2.9439
Face	99.9438	5.4146	79.9208	2.9912
WLD				
Left Eye	99.3699	10.5697	56.5907	2.5122
Right Eye	99.4010	9.3778	61.5161	2.6390
Face	99.6007	14.5565	41.9101	2.1189
SIFT				
Left Eye	99.7567	8.9744	63.9085	2.3810
Right Eye	99.7255	8.4396	64.5694	2.3822
Face	99.9314	7.9453	68.2645	2.4639
SURF				
Left Eye	99.6506	9.6860	60.4454	2.2050
Right Eye	99.5383	9.4569	59.8147	2.2189
Face	99.9314	7.6792	64.6384	2.5892
Eigenfaces				
Left Eye	98.9893	15.1419	48.4156	2.0179
Right Eye	99.0766	13.8010	51.0956	2.1237
Face	99.6194	13.0819	47.7257	2.1642

Table 2.1: Performance statistics of the experiments using the FRGC Experiment 1 dataset.

comparing face and periocular images.

In most cases, the different performance metrics agree with each other as to which feature extraction method performs the best for each image segment. For instance, LPQ features produce the best performance result when extracted from face images for each of the different measures. Though there is some difference, it mostly occurs at the bottom of the list of best performing features.

Baseline experiments were also performed with images from the FERET dataset. The setup of the FERET database is different than the FRGC database and as a result, verification rate at 0.1% false accept rate is not the best performance measure to compare two experiments. In the FERET experiments there are two images per subject. One image is in the gallery and the otherone is in the probe set. Rank-1 recognition rate is the best measure of performance for experiments of this nature. However, other measures of performance

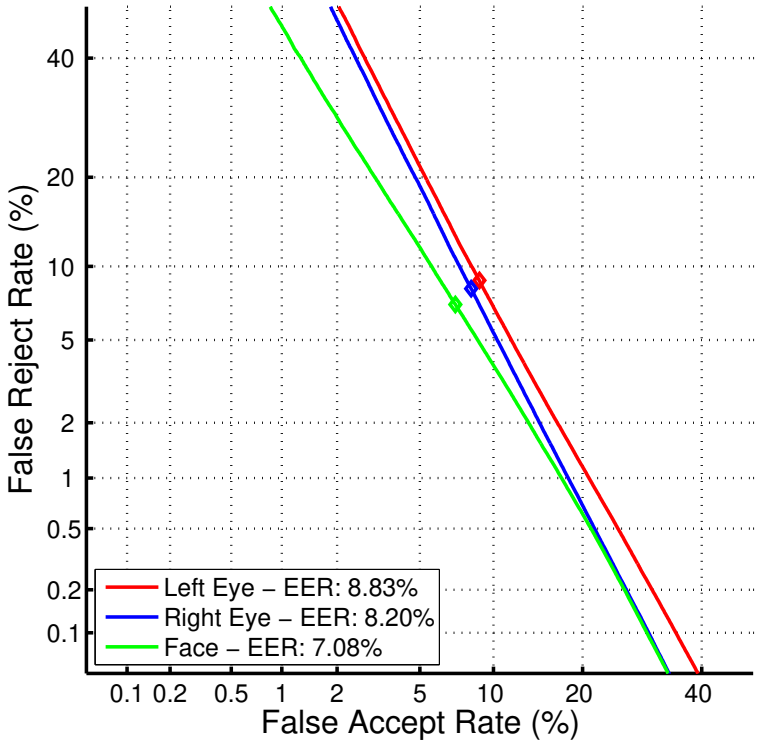


Figure 2.5: DET of Experiments using LBP features on the FRGC Experiment 1 dataset.

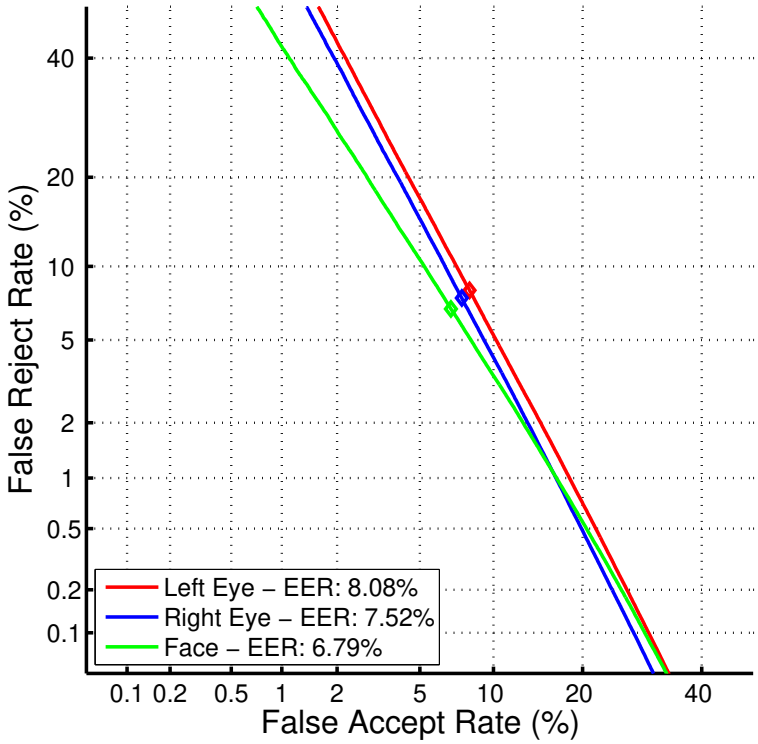


Figure 2.6: DET of Experiments using HOG features on the FRGC Experiment 1 dataset.

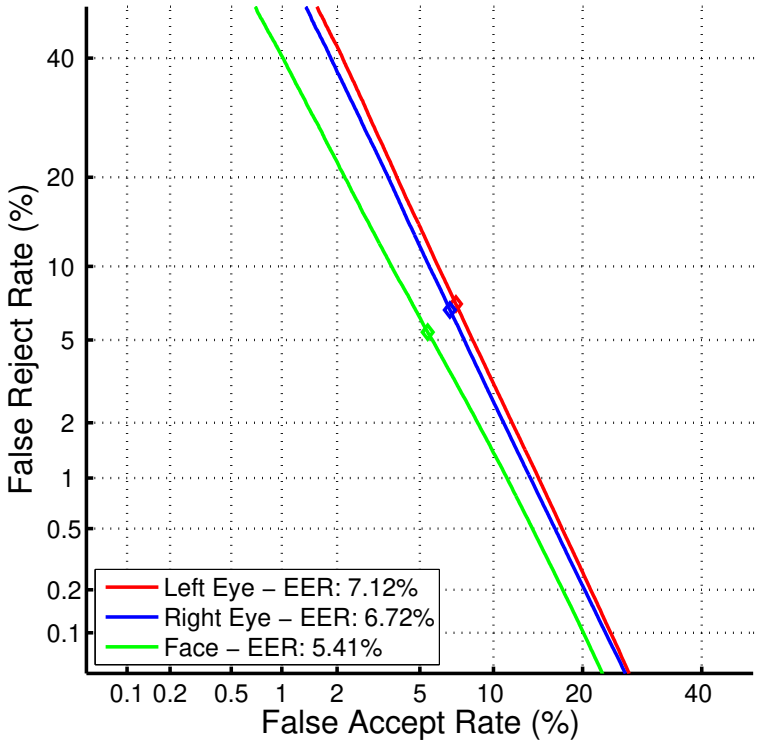


Figure 2.7: DET of Experiments using LPQ features on the FRGC Experiment 1 dataset.

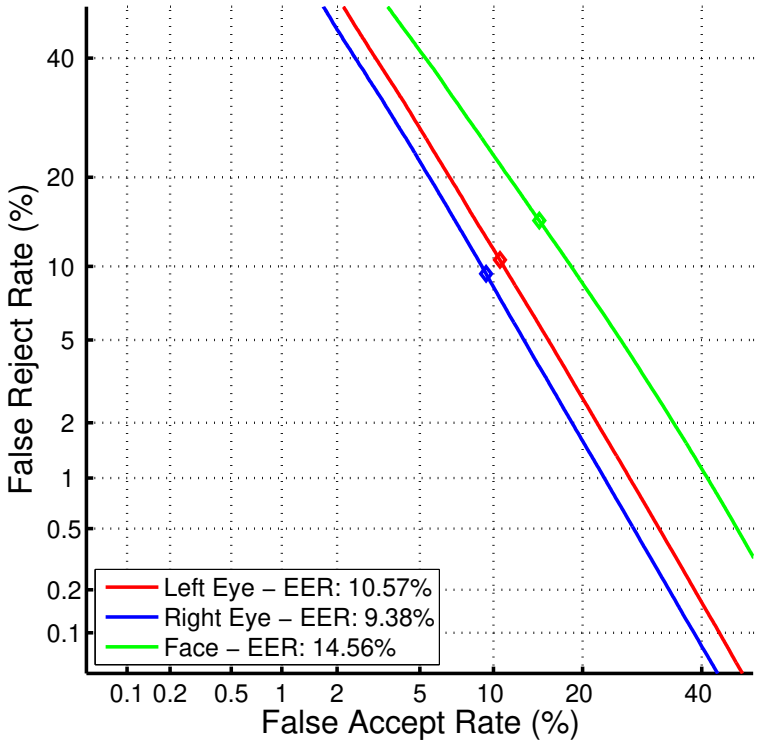


Figure 2.8: DET of Experiments using WLD features on the FRGC Experiment 1 dataset.

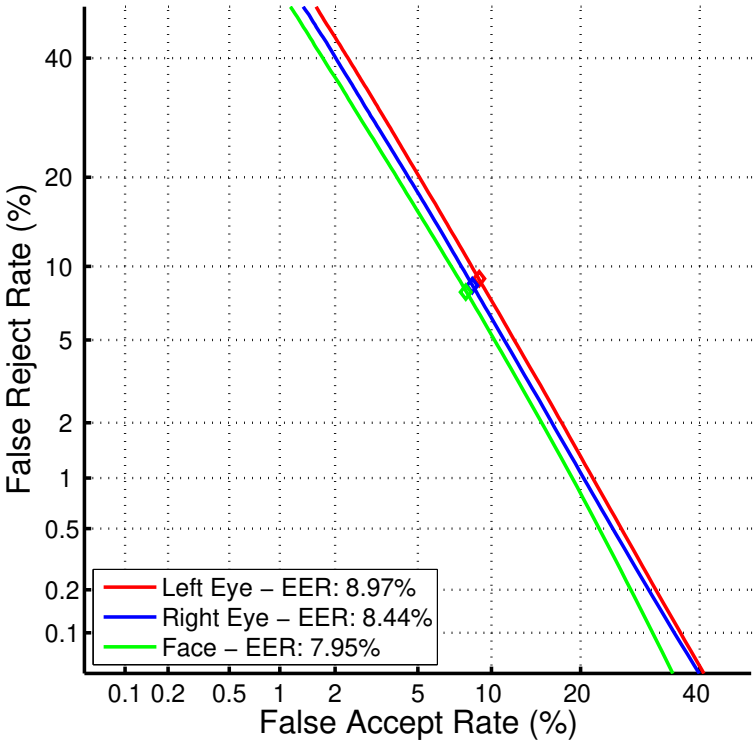


Figure 2.9: DET of Experiments using SIFT features on the FRGC Experiment 1 dataset.

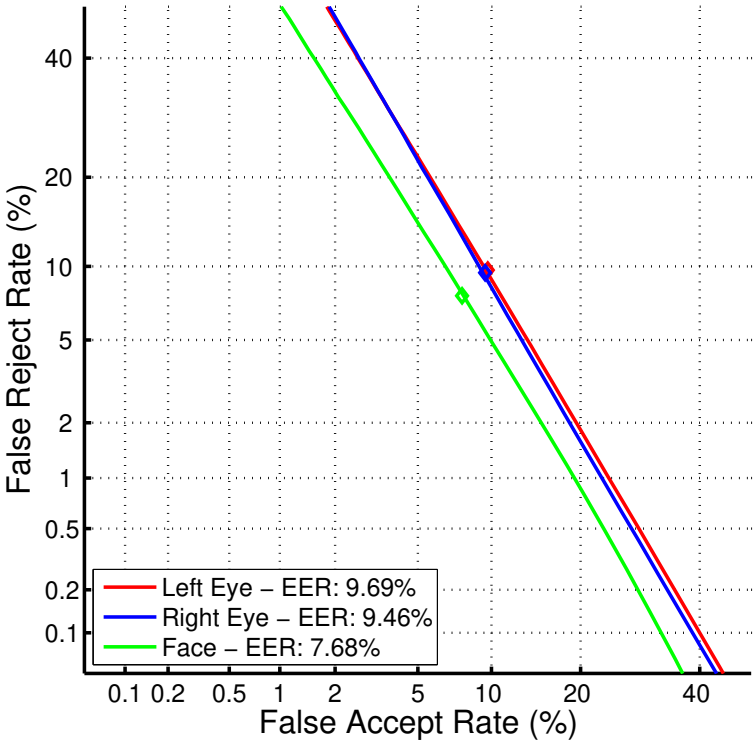


Figure 2.10: DET of Experiments using SURF features on the FRGC Experiment 1 dataset.

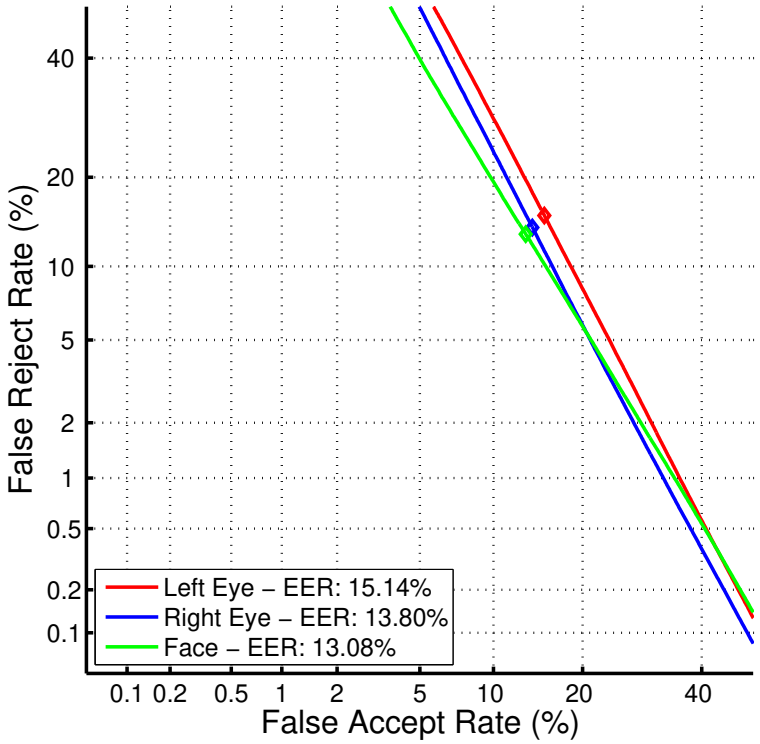


Figure 2.11: DET of Experiments using Eigenfaces on the FRGC Experiment 1 dataset.

	Rank-1	EER	VR @ 0.1% FAR	D'
LBP				
Left Eye	90.2020	4.9208	80.7071	3.2319
Right Eye	87.2727	5.2477	77.3737	3.2113
Face	96.4646	3.3375	87.7778	3.8307
HOG				
Left Eye	87.1717	5.0597	80.6061	3.3462
Right Eye	86.6667	5.3547	78.6869	3.3305
Face	96.8687	1.6138	92.4242	4.3129
LPQ				
Left Eye	92.0202	4.6486	83.9394	3.2062
Right Eye	92.2222	4.8796	83.4343	3.1487
Face	98.1818	1.1050	96.1616	4.1220
WLD				
Left Eye	85.0505	5.0035	77.4747	3.4546
Right Eye	84.0404	4.8389	75.4545	3.3978
Face	96.2626	1.6915	92.4242	4.4010
SIFT				
Left Eye	95.0505	3.1806	91.1111	4.2568
Right Eye	94.6465	2.8910	91.4141	4.2725
Face	97.2727	1.2118	96.9697	4.6119
SURF				
Left Eye	92.0202	3.9936	89.1919	4.1111
Right Eye	91.7172	3.6010	88.2828	4.1298
Face	96.1616	1.9630	95.6566	4.4745

Table 2.2: Performance statistics of the experiments using the FERET dataset.

are included in the results.

A total of 18 baseline experiments were conducted using the FERET dataset. Four performance metrics were computed for each experiment. Table 2.2 shows the performance statistics for these experiments. Rank-1, EER, VR at 0.1% FAR, and D' are given for each feature extraction method used on left eye, right eye, and face images. Figures 2.12 - 2.17 show the CMC of those experiments.

The results of the experiments using FERET images display mostly the same trends as the FRGC experiments. The face experiments perform better than the periocular region ones, though the separation in performance between the two is slightly larger for FERET experiments than FRGC experiments. LPQ is still the top performing feature extraction method. One noticeable difference between FERET and FRGC experiments is the relatively better performance of the keypoint-based methods in FERET experiments. This phenomenon will be explored later in this section.

Given that a baseline performance difference between the face and periocular region is quantified in these experiment sets, possible reasons for the scale of the difference and explanations for why the periocular

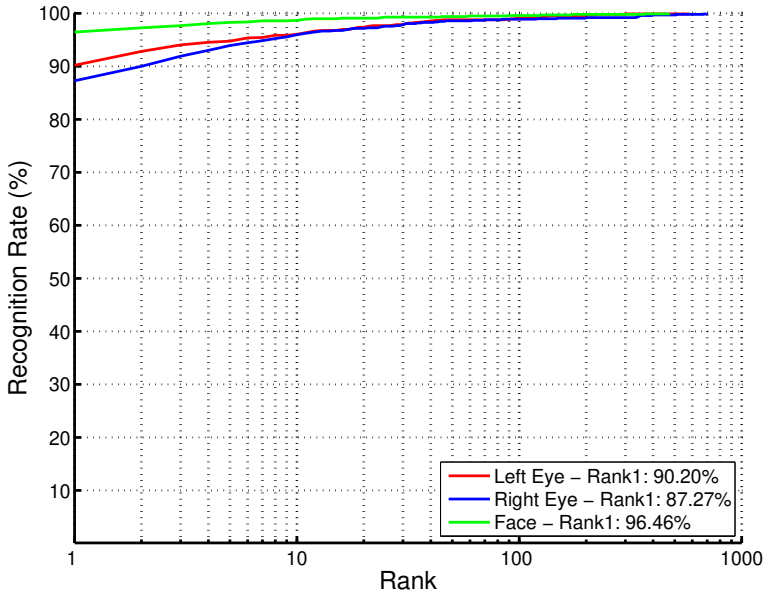


Figure 2.12: CMC of Experiments using LBP features on the FERET dataset.

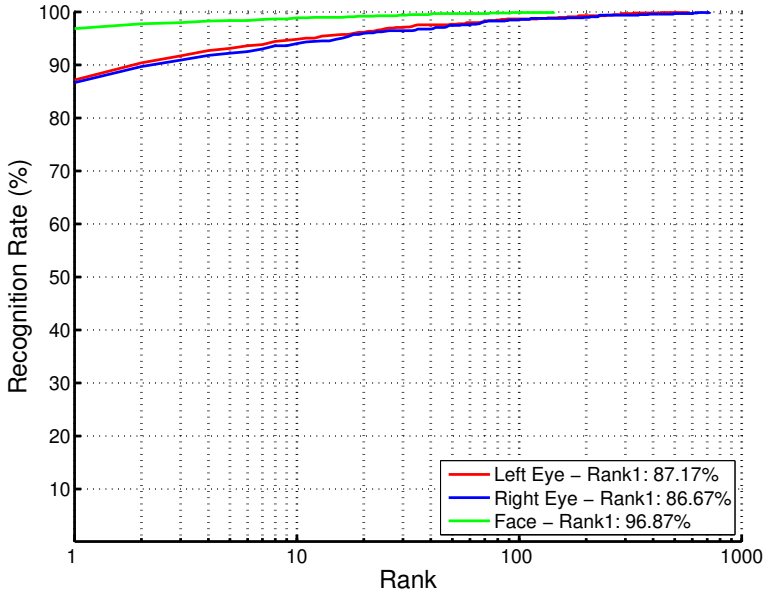


Figure 2.13: CMC of Experiments using HOG features on the FERET dataset.

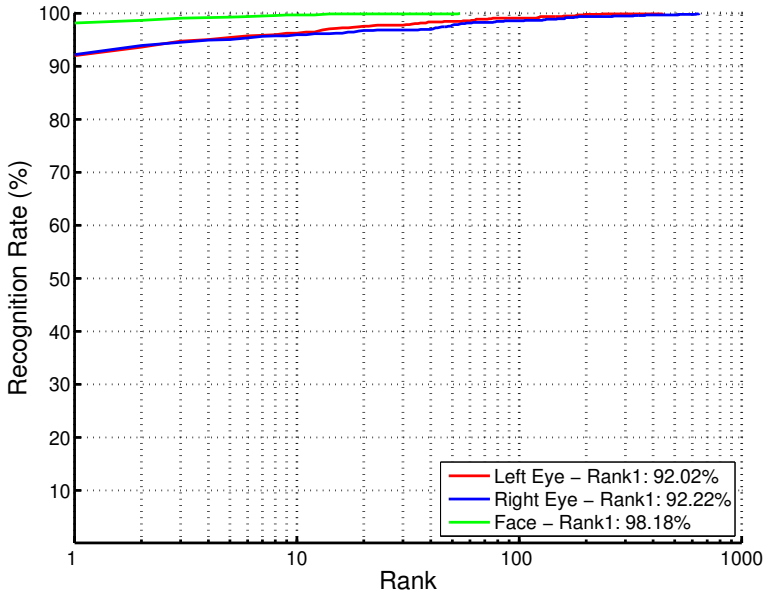


Figure 2.14: CMC of Experiments using LPQ features on the FERET dataset.

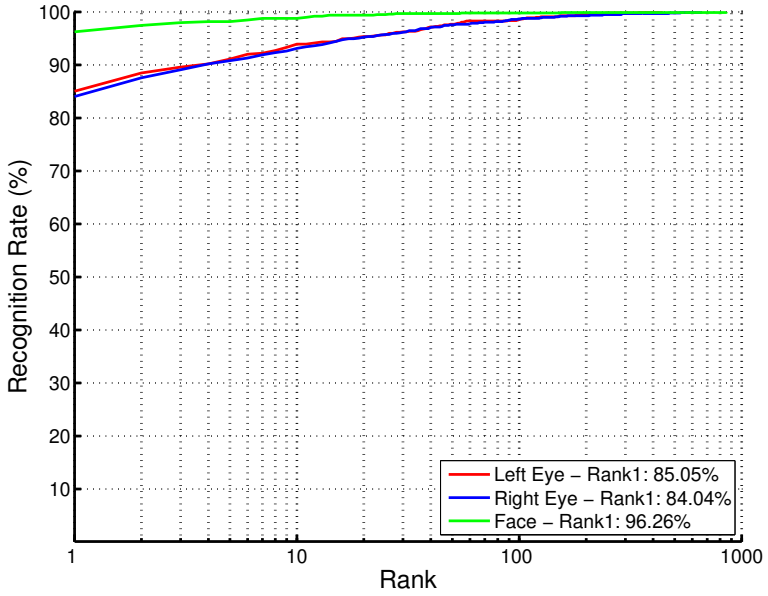


Figure 2.15: CMC of Experiments using WLD features on the FERET dataset.

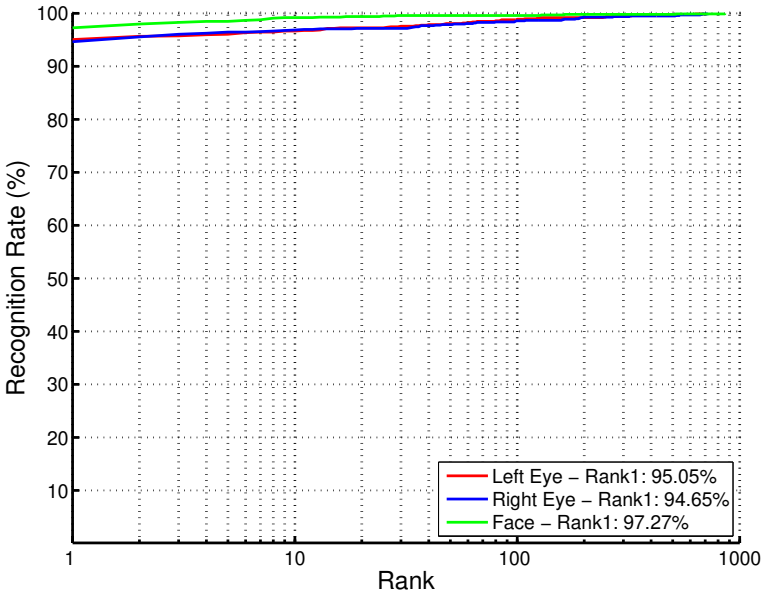


Figure 2.16: CMC of Experiments using SIFT features on the FERET dataset.

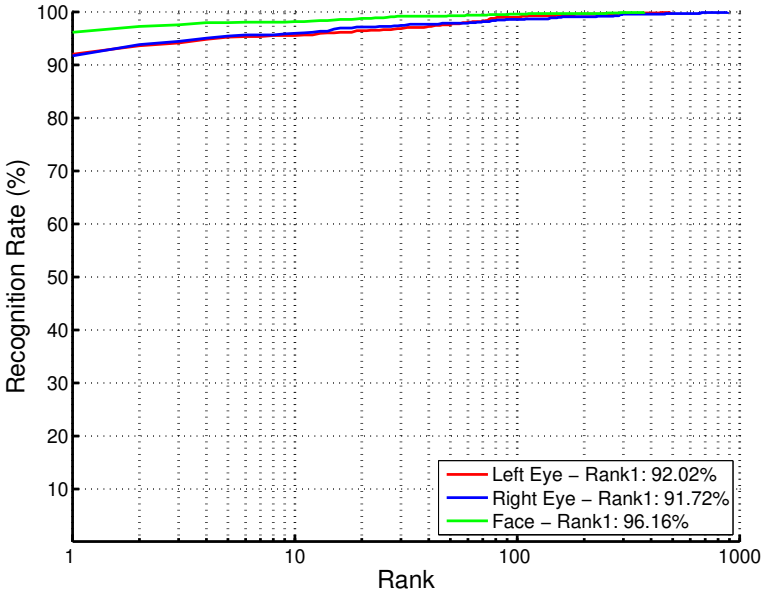


Figure 2.17: CMC of Experiments using SURF features on the FERET dataset.

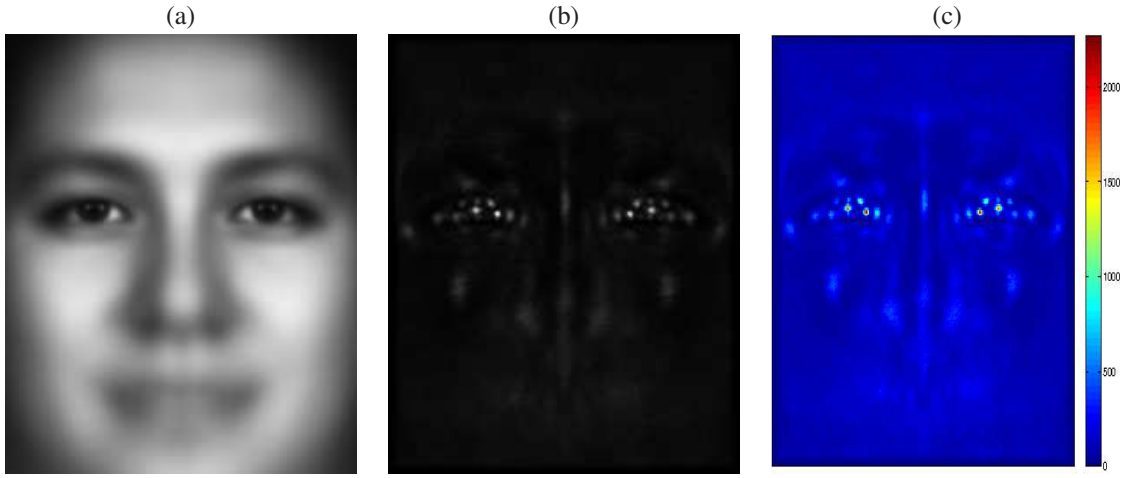


Figure 2.18: Distribution of SIFT keypoint locations from FRGC images.

region is a highly discriminative part of the face are explored.

The small difference in recognition performance between experiments using face features and experiments using periocular features suggests that the periocular region contains much of the face's discriminative information. Some examples of discriminative physical features present in the periocular region are the shape of the eyelid, the shape of the fold above the eyelid, the presence of wrinkles around the eye, the shape of the eyebrow, the thickness of the eyebrow, the texture of the eyebrow, and the texture of the skin. Both the keypoint-based approaches and the local appearance-based approaches have ways of quantifying the extent to which the physical features of the periocular region are the most discriminative in the face.

Figure 2.18 shows the distribution of SIFT keypoints found in FRGC images. Figure 2.19 shows the same for SURF keypoints. Figures 2.20 and 2.21 show SIFT and SURF keypoints for FERET images. In each figure, (a) is the mean of all images used in the experiments described above, (b) shows the percentage of keypoints found at every pixel location, with white corresponding to the maximum number of keypoints for a single pixel location and black representing zero keypoints, and (c) shows the absolute number of keypoints at each pixel location represented as a heat map. It can be easily seen in both figures where the majority of keypoints are found. Most keypoints are concentrated around the eye, in either the iris, eye corners, or eye lids. The area around the eye often contains most of the lines in the face. The effect of this concentration of keypoints around the eye is seen in the performance of the experiments using FRGC periocular images and their similarity to the experiments using the face images.

Figures 2.22 - 2.25 and 2.26 - 2.29 show the performance of experiments using local appearance-based features from only one patch in FRGC images and in FERET images, respectively. These experiments

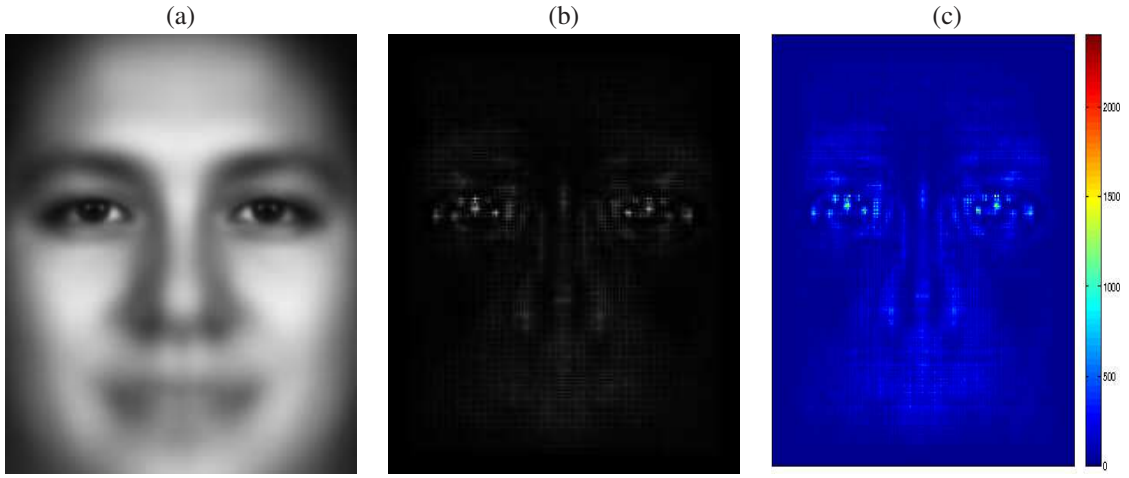


Figure 2.19: Distribution of SURF keypoint locations from FRGC images.

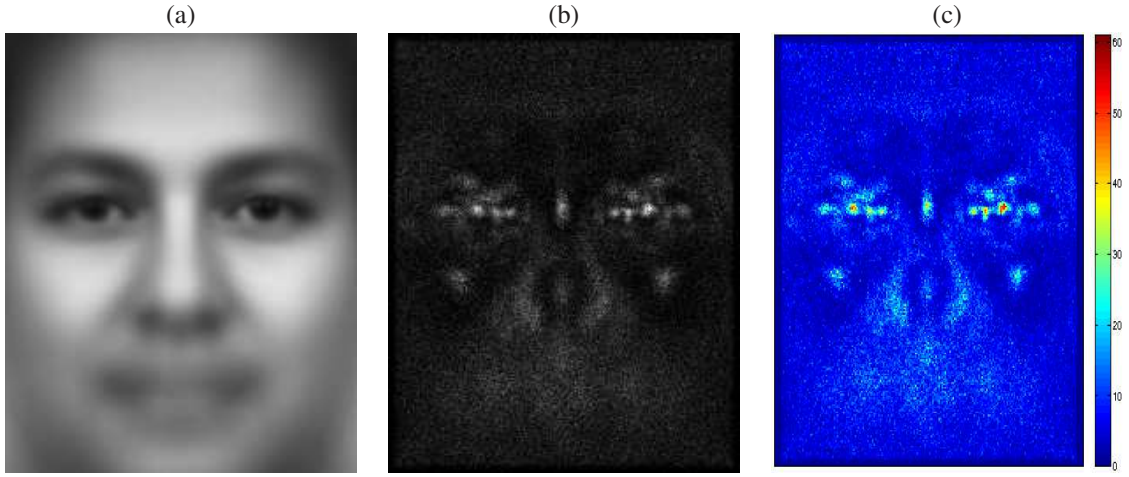


Figure 2.20: Distribution of SIFT keypoint locations from FERET images.

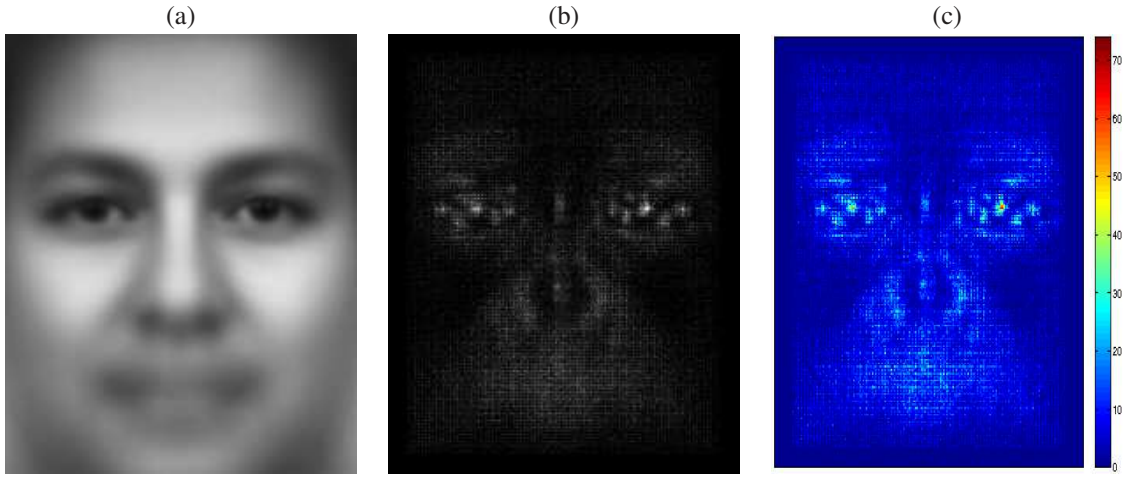


Figure 2.21: Distribution of SURF keypoint locations from FERET images.

follow the same method as all of the others in this chapter with the exception that feature vectors from multiple patches are not concatenated together to form a single feature vector for an image. Instead, the features from only one patch are compared between all images. Each face image was sub-divided into 120 patches through the method described in Section 2.3.3.1. The patches were arranged as seen in Figure 2.22. As in the previous figures, (a) is the mean of all images used in the experiments, (b) shows the relative performance of each patch, with white corresponding to the highest VR at 0.1% FAR in FRGC images and the highest Rank 1 in FERET images and black representing the lowest performance, and (c) shows the absolute value of the performance of the experiment, performed with only features from each patch, represented as a heat map. The figures shown here display the same trend as those using keypoint-based feature extraction methods. The best performing patches are found around the eye, with a few exceptions. The pattern of the results seen in these figures supports the hypothesis that the periocular region has intrinsic discriminatory power observed through the fact that two different types of methods generate similar observations.

At first glance it might appear that most of the keypoints are found within the iris and, given the known high performance of most iris recognition systems, that those keypoints are causing the high performance of the periocular region. The same can be said for the local appearance-based approaches and the fact that regions closest to the eye perform the best. Park *et al.* investigated this claim in their work [37]. In it, they performed multiple experiments using a constant set of data and feature extraction techniques. The difference between each experiment was the presence of a mask on top of the (a) iris and (b) entire eye region. Figure 2.30 is the figure seen in their paper that shows the portion of the periocular region that was masked.

Park *et al.* used two different local appearance-based feature methods and one keypoint-based

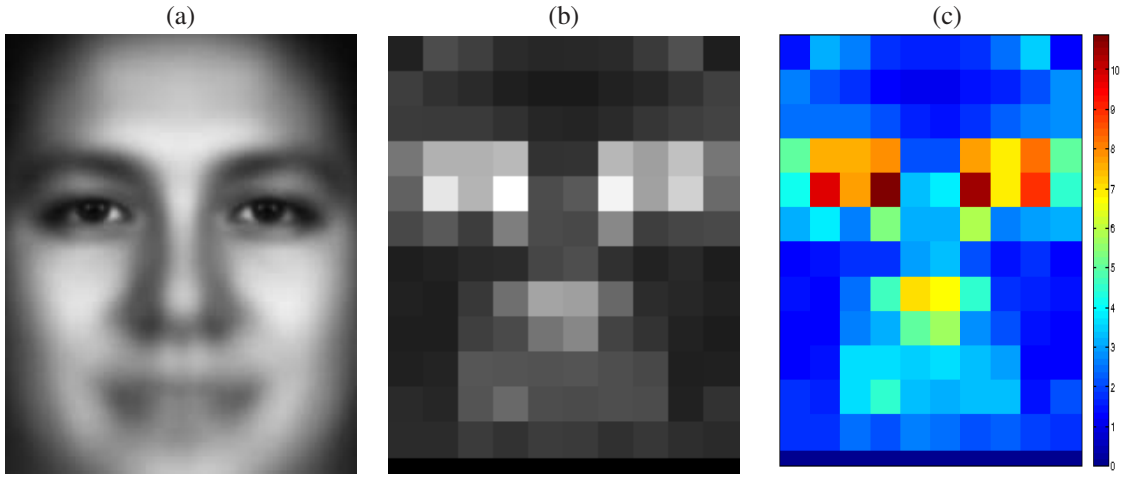


Figure 2.22: Performance of experiments using LBP features from only one patch in FRGC images.

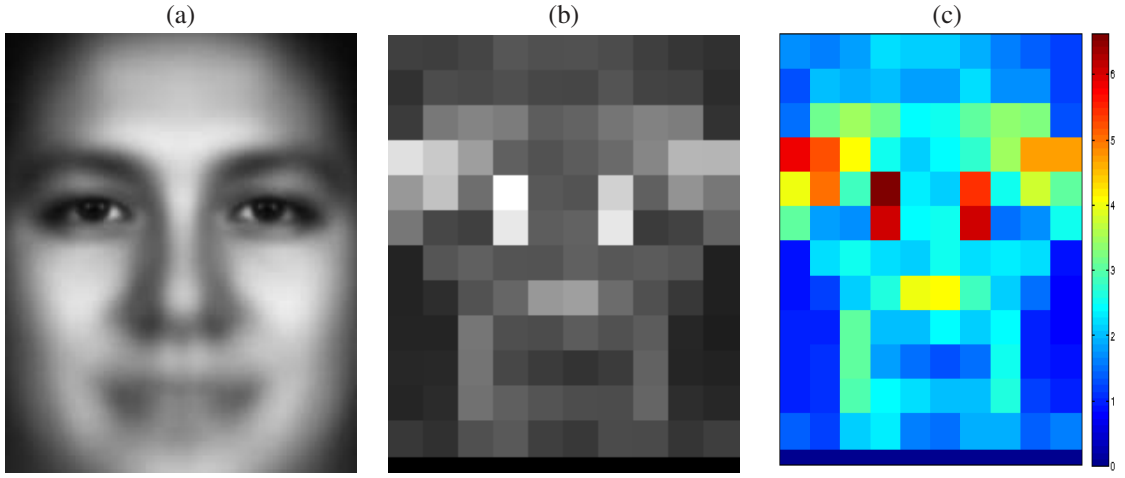


Figure 2.23: Performance of experiments using HOG features from only one patch in FRGC images.

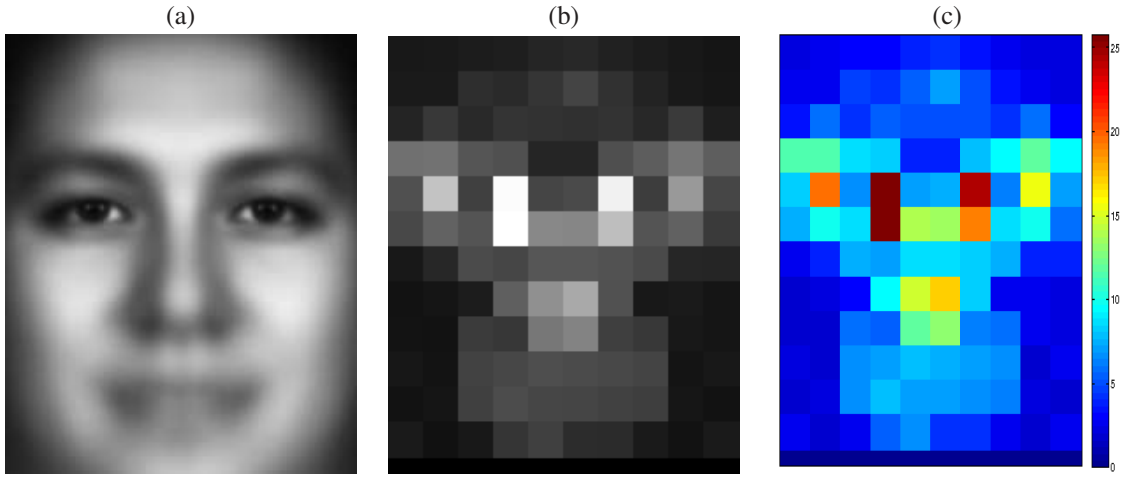


Figure 2.24: Performance of experiments using LPQ features from only one patch in FRGC images.

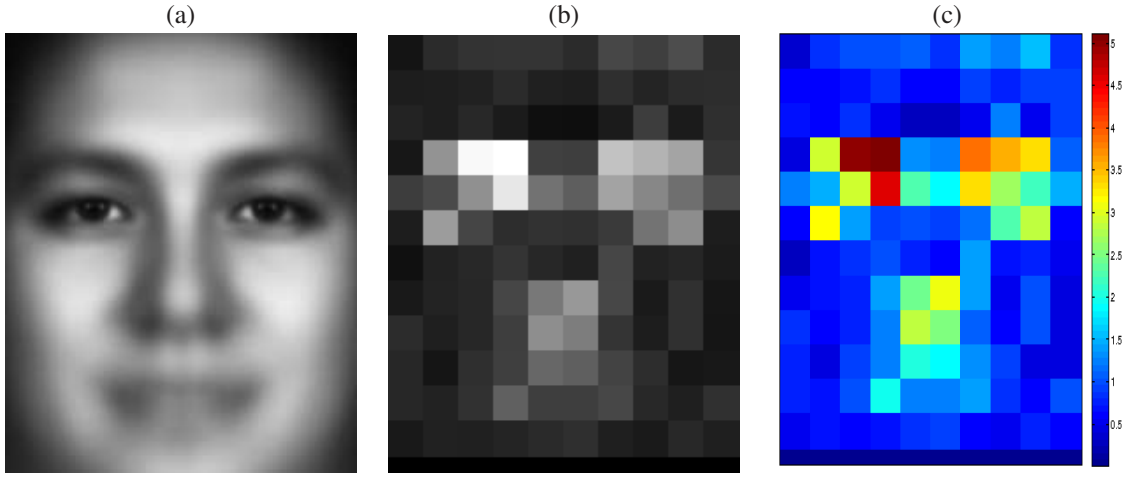


Figure 2.25: Performance of experiments using WLD features from only one patch in FRGC images.

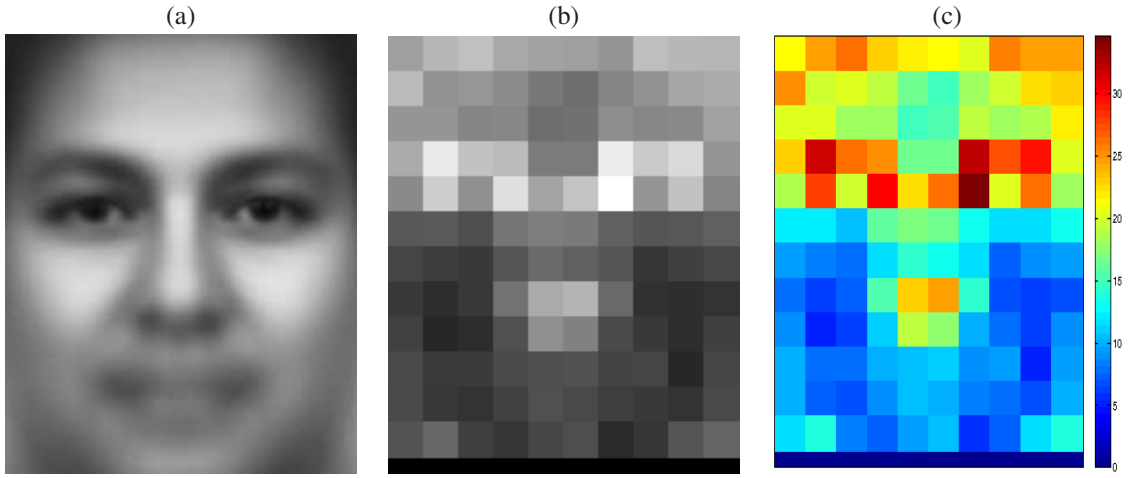


Figure 2.26: Performance of experiments using LBP features from only one patch in FERET images.

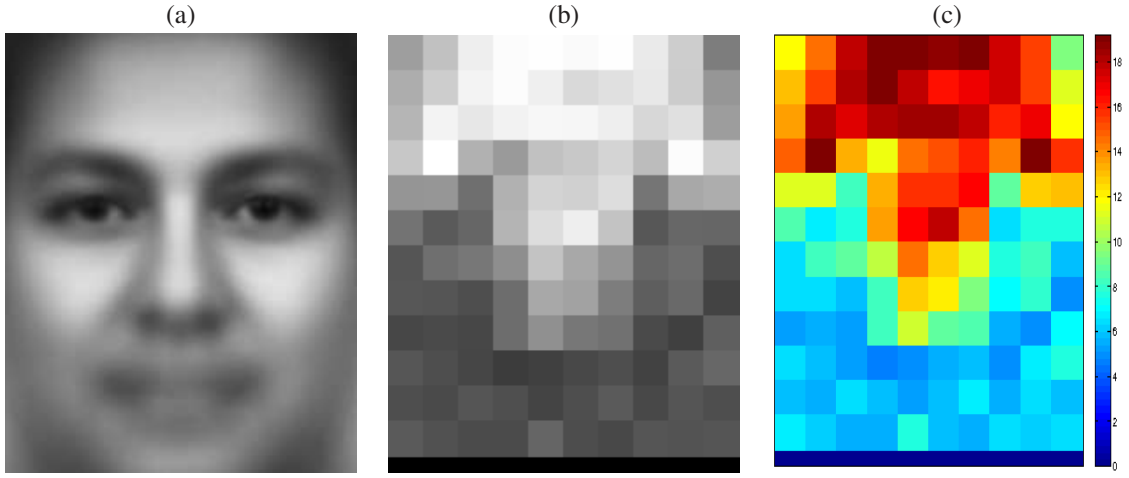


Figure 2.27: Performance of experiments using HOG features from only one patch in FERET images.

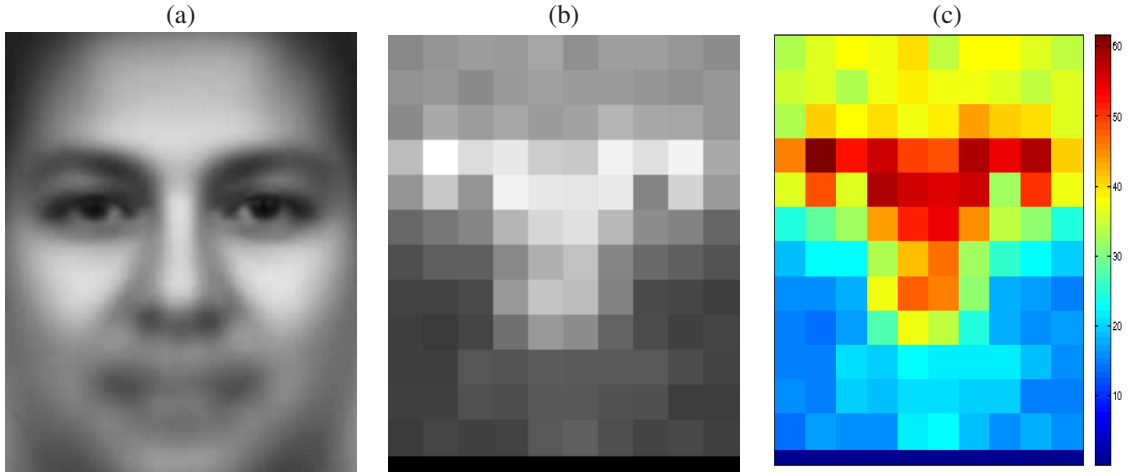


Figure 2.28: Performance of experiments using LPQ features from only one patch in FERET images.

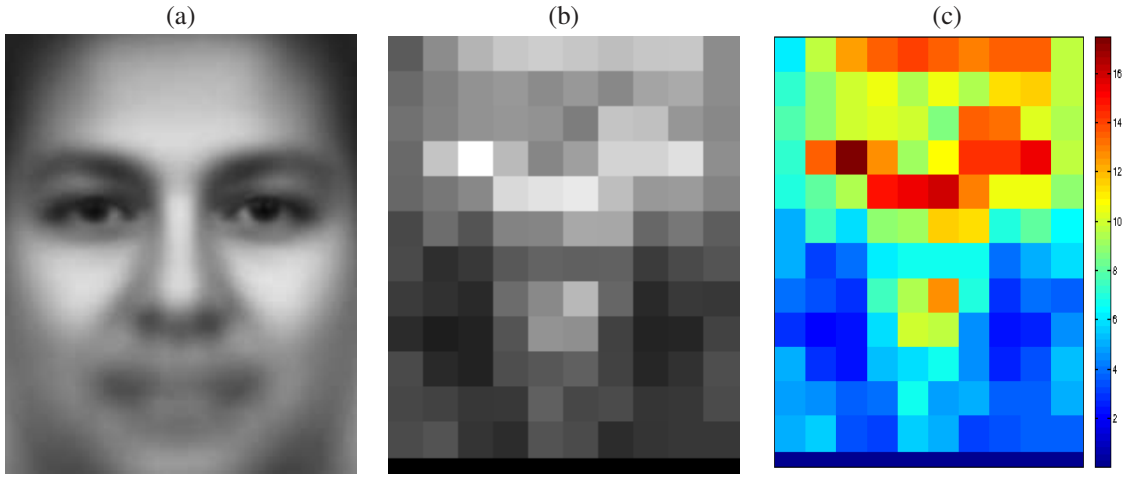


Figure 2.29: Performance of experiments using WLD features from only one patch in FERET images.

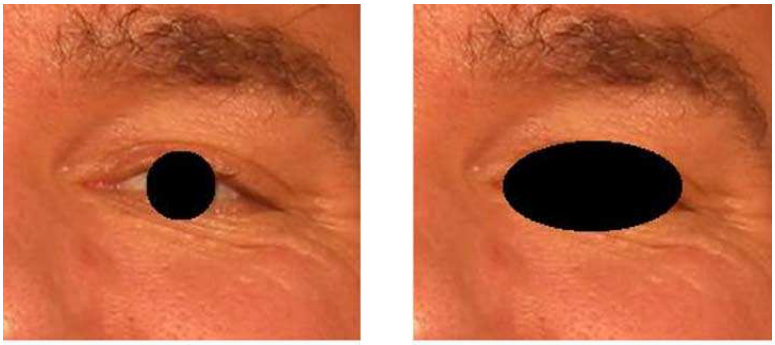


Figure 2.30: Masks used by Park *et al.*

	Without mask	Iris mask	Eye Region mask
LBP			
Left Eye	64.8841	65.2177	65.4859
Right Eye	69.7151	69.9823	70.0687
HOG			
Left Eye	69.6951	69.4327	68.9206
Right Eye	72.2473	72.0297	71.6228
LPQ			
Left Eye	75.9181	75.7238	75.6111
Right Eye	76.6574	76.3947	76.3221
WLD			
Left Eye	56.5907	56.5090	56.3560
Right Eye	61.5161	61.3175	61.1875
SIFT			
Left Eye	63.9085	63.4849	59.7951
Right Eye	64.5694	64.7299	59.6715
SURF			
Left Eye	60.4454	61.2750	58.8559
Right Eye	59.8147	61.6970	56.9514
Eigenfaces			
Left Eye	48.4156	48.2977	48.8258
Right Eye	51.0956	51.2759	52.2109

Table 2.3: Performance statistics (VR @ 0.1% FAR) of the masking experiments using the FRGC Experiment 1 dataset.

method. For the two local appearance-based methods, performance accuracy decreased only slightly by masking only the iris. Masking the entire eye resulted in a somewhat lower performance score. The keypoint-based method showed greater change between masking the iris only and the entire eye. These experiments used only a portion of the FRGC Experiment 1 images. It is worth repeating these experiments on a larger scale with more data and more feature extraction techniques.

Table 2.3 shows the VR @ 0.1% FAR results of experiments performed by masking the iris, masking the eye region, and without masking in FRGC images. Table 2.4 shows the Rank 1 recognition rates of experiments on FERET images. The masking procedure followed the procedure performed by Park *et al.* in their experiments.

The results of the experiments presented here differ only slightly from the experiments of Park *et al.*. For FRGC experiments, there is a negligible difference in performance between data with the iris masked and data without a mask and there is a slight decrease in performance when masking the entire eye region. There is no difference in performance in the FERET experiments. The distribution of SIFT keypoints within the different regions might shed some light on this observation. 22.47% of all SIFT keypoints in FRGC

	Without mask	Iris mask	Eye Region mask
LBP			
Left Eye	90.2020	90.2020	90.2020
Right Eye	87.2727	87.2727	87.2727
HOG			
Left Eye	87.1717	87.1717	87.1717
Right Eye	86.6667	86.6667	86.6667
LPQ			
Left Eye	92.0202	92.0202	92.0202
Right Eye	92.2222	92.2222	92.2222
WLD			
Left Eye	85.0505	85.0505	85.0505
Right Eye	84.0404	84.0404	84.0404
SIFT			
Left Eye	95.0505	95.0505	95.0505
Right Eye	94.6465	94.6465	94.6465
SURF			
Left Eye	92.0202	92.0202	92.0202
Right Eye	91.7172	91.7172	91.7172

Table 2.4: Performance statistics (Rank-1) of the masking experiments using the FERET dataset.

images fell within the periocular region of the experiments whose performance scores are given in Table 2.1. Using 22% of the keypoints space, SIFT only saw an approximately 4% decrease in performance. This would suggest that the keypoints in the periocular region are particularly discriminative. 2.35% of all SIFT keypoints from the face in FRGC images fell within the iris mask. This is not a lot of information to lose, so it seems reasonable that the performance would not decrease much if the iris was masked. Additionally, FRGC and FERET images are not ideal data for iris recognition systems, so you can expect the few keypoints that are found in the iris will not be particularly discriminative. 6.29% of SIFT keypoints from the face in FRGC images fell within the eye mask. This represents $\approx$30% of the keypoints found in the periocular region. This loss of keypoints contributes to the decrease in performance when masking the eye region. It is not a large decrease though, being similar to the decrease seen between using the entire face and just the periocular region. This result suggests that the remaining physical regions found in the periocular region, the shape of the eyelid, the shape of the fold above the eyelid, the presence of wrinkles around the eye, the shape of the eyebrow, the thickness of the eyebrow, the texture of the eyebrow, and the texture of the skin, are very discriminative. This assumption is supported by the figures showing the performance of individual patches using the local appearance-based methods. Many of the patches around the eye, but not those in the center, provide high performance relative to the rest of the face.

While the performance results using FRGC images shown in Table 2.3 change very little based on the masking, the results produced from using FERET images show in Table 2.4 do not change at all. This is not an unexpected result for two reasons. First, the Rank-1 performance metric used to quantify the performance of the experiments using FERET images is only reporting the relationship between the query image and its top match. The ranking of any number of matches below the top match can change without affecting the Rank-1 as long as the top match is still the top match. This is not so with VR @ 0.1% FAR, which is influenced by every match. Second, there are much fewer images in the FERET dataset compared to the FRGC dataset. Notice that in Table 2.3, some of the performance metrics changed by fractions of a percent. The performance measurements of experiments using FERET images will not change in such small amounts because of the smaller number of images.

It remains to be explored why any particular feature extraction technique performs better than the others when using periocular region images. The analysis for this question will be presented in Section 3.6 where it will contribute to a novel method for periocular recognition.

The analysis of the experiments presented in this section confirms the works that test the performance of the periocular region in the context of partial face recognition that were discussed in Section 1.2. The work presented here adds additional experimental evidence to the assertion that the periocular region is the most discriminative region of the face. It also serves to address the first two goals of this chapter as outlined in Section 2.1: (1) to examine common facial recognition data and common facial feature extraction algorithms as they are implemented in a periocular based biometric system and (2) to determine where in the face and periocular region the most discriminative features are extracted.

The previous work performed by Park *et al.* explored the effect that different segmentation techniques have on periocular recognition. It also used two different local appearance based approaches as feature extraction techniques. The work presented in this chapter provides additional insight in the form of much more test data and many more feature extraction techniques. It provides evidence that features generated using multiple classes of feature representation techniques from the periocular region are discriminative within the same set of test data. Additional experiments were performed and analysis given as to what subparts of the periocular region give the whole its discriminative power using both local appearance-based and keypoint-based approaches.

2.5 The Effect of Camera Quality

The experiments conducted in this section test the effect of deteriorated image focus, image resolution change, and uncontrolled illumination between the gallery and probe images. The experiments involving the first two conditions are further segmented into experiments where the image quality deterioration comes from natural means (environment influenced) and experiments where it comes from artificial means (post-processing to simulate deterioration). Experiments that are conducted using the unmodified periocular images serve as a baseline and follow the basic biometric experiment protocol explained above.

The biometric experiments performed in this section mostly follow the same steps as laid out in Section 2.3: data preprocessing, testing/training partitioning, gallery/probe partitioning, feature extraction, feature comparison, and computation of performance statistics. Data preprocessing remains the same. Testing/training partitioning and gallery/probe partitioning of the dataset for experiments involving image focus and image resolution change follow the FRGC Experiment 1 protocol defined in Section 2.3. The partitioning of the experiment testing uncontrolled illumination follows the Experiment 4 protocol described above. The feature extraction techniques used in Section 2.4 are used here as well. The process of feature comparison is different and explained in the corresponding sections below. Performance statistics are given in the form of verification rate at 0.1% false accept rate and equal error rate.

The experiments that test artificially deteriorated images use a different set of data. For the experiment testing image focus, the two images with the highest measure of focus were selected from each subject in the Experiment 1 dataset. For the experiment testing image resolution change, the two highest resolution periocular images were selected from each subject. The methods to deteriorate an image and the effects they have on the periocular feature extraction are described in Section 2.5.1.

2.5.1 Performance Evaluation Considerations

2.5.1.1 Image Focus

A number of factors can lead to an unfocused image. The quality of the image will suffer if the camera is unsteady when the image is captured, the subject moves out of the focus area of the camera, or the subject is moving faster than the camera's shutter speed allows. An important aspect in any biometric recognition system is robustness to variable focus. The experiments in this chapter evaluate focus in the following way.

The Fourier energy spectrum can be used to quantify the level of focus of an image [10]. Images

Table 2.5: Focus metric differences between images.

Focus difference range	true matches	false matches
0 - 4.96	592,762	135,590,852
4.97 - 9.92	156,376	79,650,906
9.93 - 14.88	36,890	29,765,650
14.89 - 19.84	9,068	8,333,278
19.85 - 24.80	2,486	2,087,510
24.81 - 29.76	734	526,754
29.77 - 34.72	296	110,328
34.73 - 39.69	64	15,212
39.70 - 44.65	0	1,412
44.66 - 49.61	0	178

Figure 2.31: A comparison of unfocused images and their focus metrics. From left to right: Un-deteriorated Image (171.46), one convolution (167.25), five convolutions (161.36), ten convolutions (156.83).

from the FRGC dataset have focus metrics ranging from 182.96 (in focus) to 141.25 (out of focus). The distribution of focus metrics is approximately normal with a mean of 163.87 and a standard deviation of 5.26.

Experiments were designed with the goal of evaluating the effect on recognition performance of the natural deterioration of image focus. The absolute difference in focus metric between each image in the FRGC Experiment 1 dataset was calculated. Image comparisons were sorted into 10 evenly spaced focus metric difference ranges. Table 2.5 shows the difference in focus metric between every image comparison in the dataset. It also shows the number of true matches and false matches in that range. Results will be reported for experiments using the image comparisons in each of the 10 different focus metric ranges.

Utilizing artificial defocus allows the periocular region to be evaluated at lower levels of focus and also provides a uniform distribution of images across all levels of focus. An input image is artificially unfocused by convolving the image with a Gaussian filter defined by a kernel size of 5 and a sigma of 1.0. An image is repeatedly convolved with the Gaussian filter until the desired focus metric is reached. Gaussian blur does not change the number of pixels in the image, therefore no changes are needed in the feature extraction methods that have been used so far. Figure 2.31 gives an example of successively defocused images compared to the original.

Table 2.6: Focus metric differences between images.

Number of convolutions	true matches	false matches
0	932	866,760
1	932	866,760
3	932	866,760
5	932	866,760
10	932	866,760
15	932	866,760
20	932	866,760
30	932	866,760
40	932	866,760

Nine experiments were conducted using artificially defocused images. The gallery set of images remains unaltered while a duplicate set of probe images are convolved with the Gaussian filter. Table 2.6 shows the number of convolutions and the number of true and false matches associated with each experiment.

2.5.1.2 Image Resolution Change

Two images taken of the same subject for the purpose of biometric recognition are unlikely to have the same number of pixels between eye centers (or any other measure of image size). In uncontrolled settings, achieving the desired resolution (image size) becomes even less likely. Provided that an algorithm has the means to compensate for changes in resolution, the biometric modality must have comparable information across resolution sizes. Resolution change in the periocular region is evaluated in the following way.

Since all images used in this section are square, a simple metric of pixel count is used to quantify the resolution of the image. Images from the FRGC dataset have pixel counts ranging from 116,281 (large image) to 36,481 (small image). The distribution of images across resolutions is approximately normal with a mean of 69,122 and a standard deviation of 9,698.

The pixel resolution difference of the image comparisons of the FRGC Experiment 1 dataset were calculated, in the same way as the image focus experiments. Table 2.7 shows the difference in resolution between every image comparison in the dataset. It also shows the number of true matches and false matches in each range. Results for these experiments will be reported using the image comparisons in each of the 10 different pixel count difference ranges.

Artificial resolution change is performed by down-sampling an image from its original size. This method can be used for whatever percentage of original size is desired. Modifications to the experimental

Table 2.7: Resolution differences between images.

Resolution difference range	true matches	false matches
0 - 7,980	588,860	114,263,582
7,981 - 15,959	183,232	80,202,500
15,960 - 23,939	25,106	40,934,850
23,940 - 31,919	1,434	14,982,054
31,920 - 39,900	44	4,401,082
39,901 - 47,880	0	1,060,138
47,881 - 55,860	0	202,302
55,861 - 63,840	0	32,580
63,841 - 71,820	0	2,882
71,821 - 79,800	0	110

Figure 2.32: A comparison of down-sampled images and their pixel counts. From left to right: Un-deteriorated Image (80,089), down-sampled to 80% (64,071), down-sampled to 40% (32,035).

setup described in Section 2.5 are made for experiments using local appearance-based feature extraction techniques because down-sampling reduces the number of pixels in the image. To accommodate for the reduction in data in each block, the size of the local block is reduced relative to the size of the image. This accommodation results in a uniform number and placement of blocks across image sizes. The feature vector calculated from each block is normalized by dividing its elements by the sum of the vector. This normalized feature vector maintains the same size and relative range of values across images of different sizes. This procedure allows for a comparison to be made between a gallery image and a smaller probe image. Figure 2.32 shows an example image next to its down-sampled versions.

Seven experiments were conducted using down-sampled images. The gallery set of images remains unaltered while a duplicate set of probe images are down-sampled as prescribed. Table 2.8 shows the percentage total pixel change and the number of true and false matches associated with each experiment.

2.5.1.3 Uncontrolled Illumination

Uncontrolled illumination is a variable that has been hard to address through traditional means, when compared to controlled illumination. Feature extraction methods that intend to be useful in this uncontrolled

Table 2.8: Image resolution differences between images.

Percent of size	true matches	false matches
100%	932	866,760
90%	932	866,760
80%	932	866,760
70%	932	866,760
60%	932	866,760
50%	932	866,760
40%	932	866,760

Figure 2.33: A comparison of a controlled illumination image (left) and an uncontrolled illumination image (right).

setting must be constant across varying lighting conditions. The FRGC Experiment 4 dataset contains frontal face images of subjects captured in uncontrolled lighting. In these images, the uneven lighting casts shadows that are not present in controlled illumination images across the face of subjects. Figure 2.33 shows an image of a subject from the FRGC dataset under controlled illumination beside an image of the same subject under uncontrolled illumination.

2.5.2 Results and Discussion

2.5.2.1 Image Focus

A total of 28 experiments were conducted using images from the FRGC Experiment 1 dataset for the purpose of evaluating the effect of changing image focus on a periocular-based biometric recognition system. Seven experiments were performed using images from each eye and for both natural and artificial image deterioration.

Section 2.5.1.1 explains how focus is quantified in an image and how it occurs naturally in the images of the FRGC Experiment 1 dataset. It is reasonable to assume that comparisons made between images over a small range of focus difference will have a different recognition performance than comparisons made between images over a large range of focus. To observe the extent of this difference, experiments were

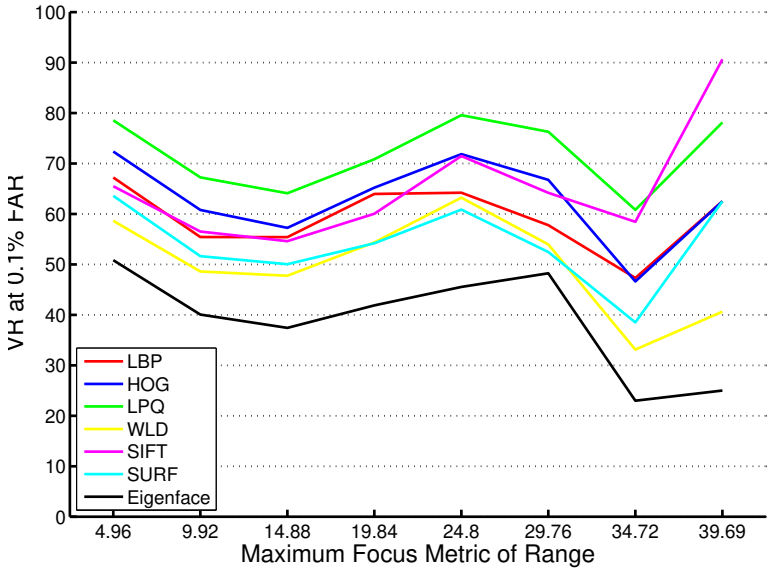


Figure 2.34: VR at 0.1% FAR from experiments using right periocular images from the FRGC dataset across difference ranges of natural focus.

conducted where each image comparison was placed into one of ten different intervals based on the focus value difference between the two images as explained in Section 2.5.1.1 and seen in Table 2.5. Figure 2.34 plots the VR at 0.1% FAR results of these experiments using the right periocular images from the FRGC Experiment 1 dataset. Left periocular image results are not shown because the results are similar and offer no additional insight on the observations of this experiment.

One thing that should be noted initially is that the number of true and false comparisons decreases at each interval of focus range. It was previously mentioned that the focus metric distribution is skewed to the largest value which explains this observation. This causes certain ranges to not have valuable information because of the lack of true comparisons. Even though ten ranges were evaluated, only eight ranges are shown in the figures presented here. The two remaining ranges have 0% VR for all features.

A decline in VR at 0.1% FAR can be seen for the first three focus metric ranges. This is in line with the expectation that the difference in focus would cause a decrease in performance. At this point the results become unexpected. The verification rate begins to increase before decreasing again. The fourth focus metric range has only 1.5% of the true matches of the first range. Therefore, these results are much less representative of the true performance of this focus difference range. Regardless, some interesting observations can be made. This experiment is comparable to the FRGC Experiment 1 performed in Section 2.4, which includes the same image set. The LPQ feature experiments performed best in the experiments from Section 2.4 with a right eye

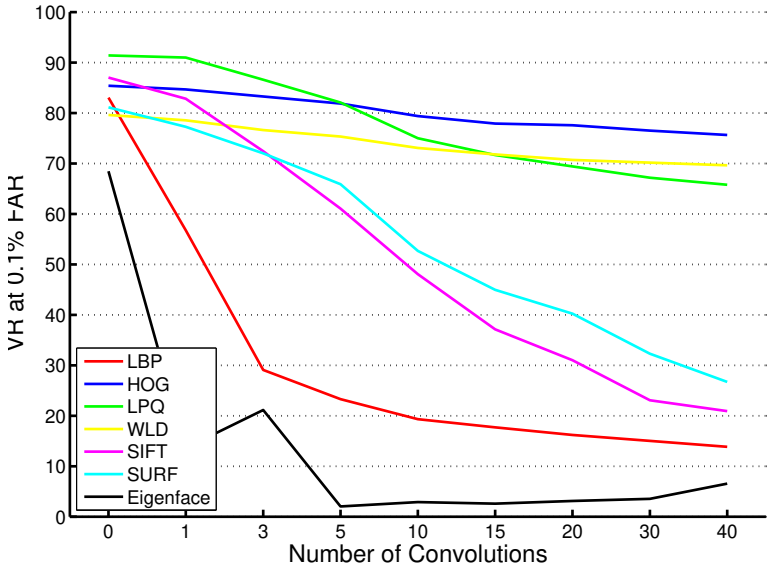


Figure 2.35: VR at 0.1% FAR from experiments using right periocular images from the FRGC Experiment 1 dataset across difference ranges of natural focus.

periocular region system performance of 76.6574% VR at 0.1% FAR. An experiment performed with those same images and presented in this section that only counted comparisons between the most similar images in focus metric showed a VR at 0.1% FAR of 78.7271%, which represents an increase of approximately 2%. The performance of experiments using HOG, SIFT, and SURF also increases while LBP, WLD, and Eigenfaces decreases. HOG, SIFT, and SURF all either claim to be partially invariant to blur or convolve the data with a Gaussian filter as part of the algorithm.

One drawback of the experiments described above is that the number of comparisons decreases as the range of focus difference increases. This changing variable can be accounted for by starting with a set of images that have a small difference in focus and introducing artificial defocus to each image, so the number of comparisons stay the same as the focus range increases. Here, the target set of images remains the same across all the experiments. Artificial defocus is introduced in increments to the same set and the new images serve as the query set. A set of images that have a small difference in focus was chosen by selecting the two images from each subject that have the highest focus metric. This method has two advantages. First, the two images with the highest focus metric always have a small difference. Second, a higher focus metric will allow a greater amount of artificial defocus to be introduced.

Section 2.5.1.1 explains how artificial defocus is applied to an image. Figure 2.35 shows the VR at 0.1% FAR results using only right periocular images from the FRGC Experiment 1 dataset.

The results seen here show a noticeable separation in performance between the different feature representation techniques as the number of convolutions increase. HOG, LPQ, and WLD features show the least amount of performance decrease. The performances of LBP, SIFT, and SURF appear to decrease at a similar rate. Only Eigenfaces immediately decreases considerably in performance.

HOG features show the best VR at 0.1% FAR at the highest number of convolutions. One step in HOG is convolving the image with a Prewitt filter. This filter is traditionally used in edge detection. It is not an uncommon step in edge detection problems to perform Gaussian blur on an image before detecting edges. This step is used to reduce noise. Therefore, the performance using HOG features in this experiment does not suffer greatly from defocus.

The fact that WLD features perform well at a high degree of blur compared to the unblurred image can be expected based on the nature of the algorithm. The two parts of the WLD algorithm are the relative intensity of a pixel compared to its neighbors and the gradient orientation of each pixel. Convolving an image with a Gaussian filter has little effect on either of these components.

The developers of LPQ claim that "The phase can be shown to be a blur invariant property under certain commonly fulfilled conditions.". In their work on the recognition of blurred faces, the authors assert that "The experimental results on CMU PIE and FRGC 1.0.4 datasets show that the LPQ descriptor is highly tolerant to blur but still very descriptive outperforming LBP both with blurred and sharp images." [3]. The experiments performed in this section provide a similar analysis of both LPQ and LBP.

2.5.2.2 Image Resolution Change

A total of 28 experiments were also conducted to show the effect of resolution change. The number of images and comparisons per experiment were given in Section 2.5.1.2. As with focus, we hypothesize that comparisons made between images in the same range of image resolution will have a different performance measure than comparisons made between images with a larger difference in image resolution. Given the diverse nature of the feature representation techniques discussed thus far, the expectation is that certain features will perform better than others when being used to compare images with a large difference in resolution. To observe this difference, EER and VR at 0.1% FAR are measured using comparisons within ten different equally spaced ranges (from minimum to maximum) of image pixel count difference. Figure 2.36 shows the VR at 0.1% FAR results using right periocular images from the FRGC Experiment 1 dataset. Here again, left periocular image results are not shown because the results are similar and offer no additional insight on the observations of this experiment.

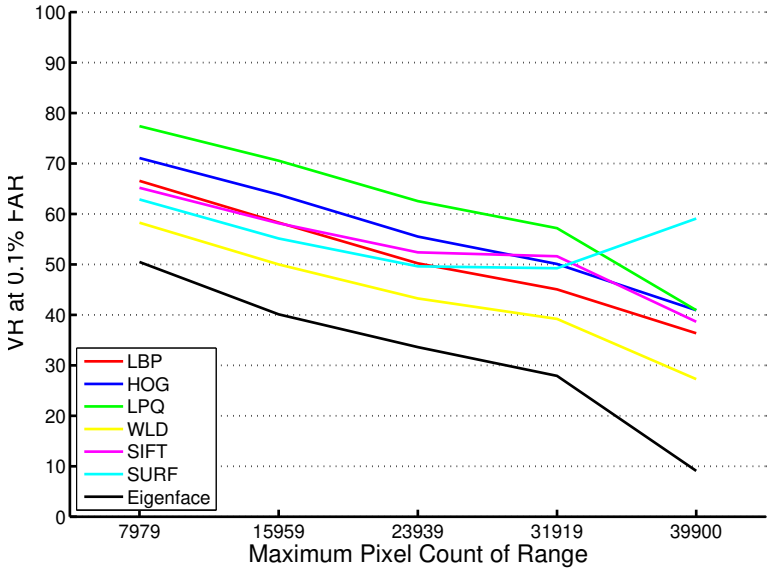


Figure 2.36: VR at 0.1% FAR from experiments using right periocular images from the FRGC dataset across ranges of natural resolution difference.

As with the experiments involving changes in focus, the number of true and false comparisons decreases at each interval of pixel count range. Even though ten ranges were evaluated, ranges with zero true matches are not shown in the figures of this work.

A decline in VR at 0.1% FAR can be seen for the first three pixel count ranges for all feature extraction techniques. This is in line with the expectation that the difference in pixel count would cause a decrease in performance. The features show the same rate of decreasing performance with the exception of SIFT and SURF. These two methods decrease at a slower rate. At the high pixel count ranges SURF show an increase in performance while all other methods continue to decline. However, the results from these ranges are less reliable.

SIFT and SURF features both claim to be scale-invariant. This invariance is achieved through the use of a difference of Gaussians approach where keypoints are found at the minima/maxima of the difference of a pyramid of images (the image in question is sub-sampled many times and analyzed at many scales).

One drawback of the experiments described above is that the number of comparisons decreases as the difference in pixel count increases. This changing variable can be accounted for by starting with a set of images that have a small difference in pixel count and down-sampling each image so that as the pixel count difference range increases the number of comparisons stay the same. In all of these experiments the target set of images remains the same while the query set is down-sampled in intervals of 10%. A set of images that

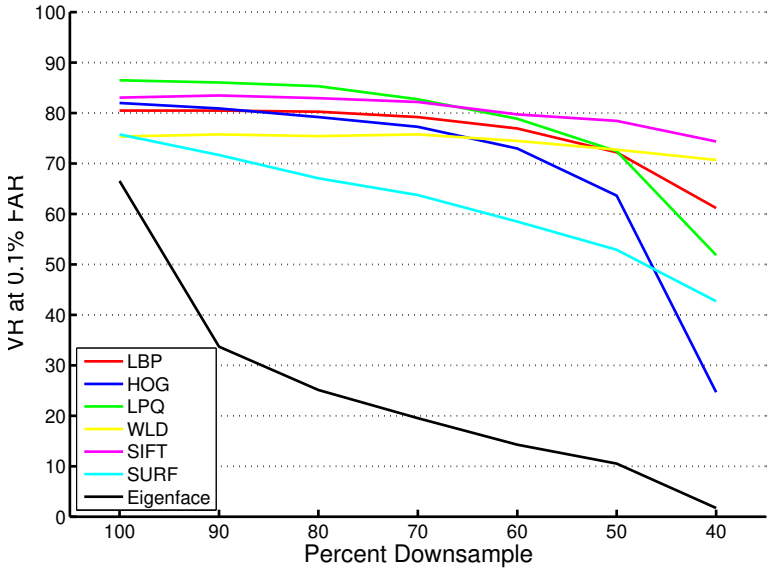


Figure 2.37: VR at 0.1% FAR from experiments using right periocular images from the FRGC Experiment 1 dataset across ranges of artificial resolution difference.

have a small difference in pixel count was chosen to serve as the target and query sets by choosing the two images from each subject that have the highest pixel count.

Section 2.5.1.2 explains how down-sampling is applied to an image. Figure 2.37 shows the VR at 0.1% FAR results using right periocular images from the FRGC Experiment 1 dataset.

The results seen here suggest that, for a small amount of down-sampling, the performance of the biometric recognition experiment with no difference in image resolution is the best predictor of the performance of the experiment with greater image resolution difference. It isn't until there is a great difference in image resolution that a method like SIFT shows it usefulness.

2.5.2.3 Uncontrolled Illumination

An additional 14 experiments were conducted to show the effects of uncontrolled illumination. An explanation of the data and the experiment was given in Section 2.5.1.3. The data used for this quality concern only allows for testing the natural occurrence of the uncontrolled illumination. The presentation of results for these experiments differs from the previous quality concerns and instead resembles the FRGC Experiment 1 results from Section 2.4. Figure 2.38 shows the DET of the experiments using right periocular images from the FRGC Experiment 4 dataset.

It should be noted that the uncontrolled illumination variable is not independent from the other

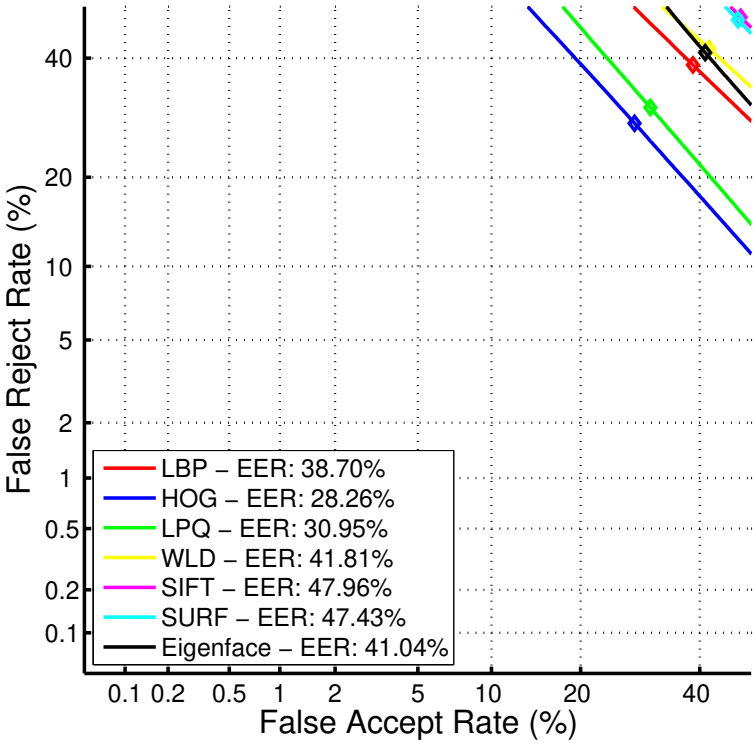


Figure 2.38: DET of Experiments testing uncontrolled illumination.

variables. The images used in the probe set have focus metrics ranging from 80.98 to 149.66 with a mean of 113.54, where the gallery set focus metrics range from 150.67 to 182.00 with a mean of 166.57. The probe set also has pixel counts ranging from 11,025 to 40,401 with a mean of 21,275, as opposed to the gallery set with pixel counts ranging from 37,249 to 114,921 with a mean of 69,323. The affect of focus and resolution change on performance has already been discussed and the uncontrolled illumination images come from a very large range of those metrics. All three of these factors affect the experimentation of this section, not just the factor of uncontrolled illumination. Therefore, poor recognition results can be expected.

All of the feature representation techniques performed so poorly at this task that nothing can be determined based on these results. However, this is not unexpected; of all the submissions to the FRGC the minimum performing algorithm performed slightly worse than the HOG implementation (5.10% VR at 0.1% FAR) presented here. The median performer was around 30% VR at 0.1% FAR. The best performing algorithm had a verification rate just below 80%. This algorithm performed this well on Experiment 4 only, which suggests it was particularly tuned to this problem. Unfortunately, it is a proprietary algorithm and it is not open for inspection [32].

2.6 Conclusion

The work presented in this chapter aimed to address three goals. The first goal was to examine common facial recognition data and common facial feature extraction algorithms as they are implemented in a periocular based biometric system. Experiments were conducted with seven different feature representations from three different classes of feature extraction methods. Each of the classes of feature extraction methods have a long history of use in facial recognition systems. Using these methods, the performance of a facial recognition system and a periocular recognition system were compared.

Establishing this baseline performance result allowed for addressing the second goal, the further exploration of where in the face the most discriminative features were found. The experiments presented in this chapter supported previous research on this question that the periocular region is the most discriminative region of the face. Some examples of discriminative physical features present in the periocular region are the shape of the eyelid, the shape of the fold above the eyelid, the presence of wrinkles around the eye, the shape of the eyebrow, the thickness of the eyebrow, the texture of the eyebrow, and the texture of the skin. The Keypoint-based feature methods found keypoints heavily concentrated around the eye corners and eye lid. The local appearance-based approaches found discriminative patches around the eyebrow, eye corners,

and the skin under the eye.

The third goal of this chapter was to assess the robustness of the periocular region with respect to the environment-influenced concerns a periocular-based biometric system will likely encounter. It was suggested that a periocular-based recognition system would likely encounter environment-influenced data quality concerns such as blur, image resolution changes, and uncontrolled illumination. Experiments were conducted to examine the extent of the performance degradation incurred by these factors when using a number of different types of features extracted from the periocular region. These experiments largely upheld the claims of the originators of each feature, such as LPQ being invariant to blur or SIFT being invariant to scale.

In later parts of this dissertation these observations will be used to develop a novel method of extracting features from the periocular region for the purpose of increasing the performance of a periocular-based biometric system over the baseline performance found by using methods originally designed for facial recognition systems.

Chapter 3

The physical structure of the periocular region

3.1 Introduction

As discussed in Section 1.2, the periocular region was first explored within the context of partial face recognition. Each of these studies discovered that the periocular region held the most discriminative information in their experimentation, which led researchers to focus on the periocular region.

Though periocular recognition is a new field of study, recent research has addressed many interesting aspects of it. Absent from the list of explored aspects of the periocular region is the motivation behind the original concept of partial face recognition. None of the current periocular studies have looked at how each sub-region of the periocular region affects performance or whether there is a specific physical feature, like the pattern of the eyelashes or wrinkles in the corners of the eye, that contribute the most discriminative information. Periocular-based recognition is still relatively new and many of the approaches used with it are very simple adaptations from other biometric modalities. A more detailed study of the periocular region is needed in order to develop algorithms that leverage specific structural aspects of the periocular region.

In addition to these early studies, a more recent investigation by Hollingsworth *et al.* explored useful physical locations of the periocular region, according to human participants [16, 17]. Human participants were presented with periocular region images and asked to decide if two images came from the same subject. During this process, the study participants were asked to identify how helpful a number of different aspects

of the periocular region were in making a decision. These aspects included “Eye shape”, “Tear Duct”, “Outer Corner”, “Eyelashes”, “Skin”, “Eyebrow”, “Eyelid”, “Color”, “Blood vessels”, and “Other”. This type of research is a step in the right direction, but there are many questions left unresolved in regard to automatic (machine) biometric recognition.

The experiments in this chapter are designed to address the following questions:

- Are there sub-regions within the periocular region that provide more discriminative information than others?
- Are these sub-regions similarly discriminative using different feature representation methods?
- Could the knowledge gained here be used to adapt the local approaches to use the physical structure of the periocular region?

The answers to these questions will give the biometric research community a better understanding of the periocular region by contributing to a model for the most discriminative sub-regions of the periocular region. This knowledge may help guide future research in part-based periocular biometric algorithms, as well as aid in optimization efforts.

3.2 Data

The experimentation in this section used a subset of the FRGC Experiment 1 data set. Two images (one probe, one gallery) were used from 466 subjects in the FRGC data set, for a total of 932 images. The two selected images were chosen for each subject from the first available recording session in order to minimize the time lapse between images, as time lapse is not the area of focus in this section. Only images with a neutral facial expression were selected, as the impact of facial expression was not the focus in this section. The set of FERET images detailed in Section 2.2 was also used.

3.3 Method

The nature of local appearance-based approaches can guide the determination of discriminative areas of the periocular region. This class of local features describes an input image by generating a histogram of predefined statistics, where the count of the occurrences of these statistics has great discriminative power. However, there are some drawbacks to using local features. Originally, most of these features were conceived

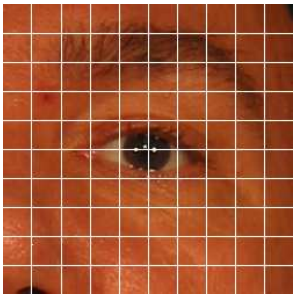


Figure 3.1: Example of a periocular region image divided into blocks.

to be used in pattern detection where a window containing a constant or repetitive pattern is compared to another window with a pattern. The local nature of these features is not well suited to describe a surface of variable structure, like the human face. Computing some features on an entire biometric image also removes spatial information. For example, the statistics of the eyelids might be different from the statistics of the skin on the cheek, yet the values are accumulated into one, unpartitioned data set. This method prevents differentiation between a particular statistic that comes from one region or the other.

A simple approach was introduced to combat these drawbacks. A biometric image can be subdivided into regions called blocks, seen in Figure 3.1. The local features can be computed in each block and the results can be concatenated together to form the feature vector. An early study that used this approach in face recognition, formatted the blocks in a non-overlapping pattern that covered the image [2].

Performance measures can be given for each block individually as well as for the entire collection. The individual blocks that perform the best may likely be the blocks that have the most discriminative information. To this end, a biometric experiment was conducted to quantify the performance of a single 20×20 pixel block from each image for blocks centered at every pixel in the image. Blocks centered on the border of the image consider only pixels within the image. 40,000 experiments were conducted, one for each pixel in the 200×200 pixel periocular image.

The biometric experiment performed in this chapter follows mostly the same steps as the one laid out in Section 2.3: data preprocessing, testing/training partitioning, gallery/probe partitioning, feature extraction, feature comparison, and computation of performance statistics. All data is used in the test set as there are no local appearance-based feature extraction techniques that require a training set. There are two images per subject in the dataset and 1 image from each is used in the gallery partition and the other in the probe partition. LBP, HOG, and LPQ features are used in this work. WLD features were left out as they have consistently underperformed in all previous work and would not add anything to the experiments conducted here. Feature

comparison is performed as in Section 2.3. Performance statistics are given in the form of Rank-1 recognition rate.

3.4 Results

Following the method described in Section 3.3, figures were created to show the relative performances of blocks centered on each pixel in the image. Figure 3.2 shows the Rank-1 recognition rate for a block centered at each pixel using local appearance-based features (LABF) extracted from FRGC images. Figure 3.3 shows the same for FERET images. As a means of anchoring the results to physical structures of the periocular region, mean images are displayed above the other figures. There were 12 different combinations of feature representations, periocular images, and datasets. Counting each of the experiments separately, these figures represent the results of 480,000 biometric experiments.

Iris Some initial periocular studies were concerned with the impact of the iris in periocular recognition [31, 37]. It was reasonable to assume that the performance of the periocular region could be greatly influenced by the iris because the iris has already been shown to be a useful and accurate biometric modality. Some research has shown that there was minimal change in performance between experiments where the iris had been masked and experiments where the iris had not been masked. Figure 3.2 supports these results. In these experiments, there was very low performance on blocks centered near the iris when using LABF. One possible reason for this is that the visible light spectrum and small iris diameter of images from FRGC were not well suited to extract texture information from the iris, hence the inclusion or exclusion of the iris had little impact on the performance of the periocular region when using LABF in these experiments. It should be noted that there are many preprocessing steps in a typical iris recognition system that were not followed in this method. It is not surprising that without these steps the iris will not produce high performance using LABF.

Skin The skin underneath the eye is a much larger area than any other part of the periocular region, yet it seems to be the least discriminative, regardless of the type of feature used. The results show no points of high performance when using LABF. Additionally, there does not appear to be any pattern to the performances in this area. To expand on this thought, the skin has the most surface area with which to find unique textures, moles, scars, or other facial features, and we expect the pattern of high performance and high key-point concentration to vary with the data to a greater degree than any other physical area of the periocular region.

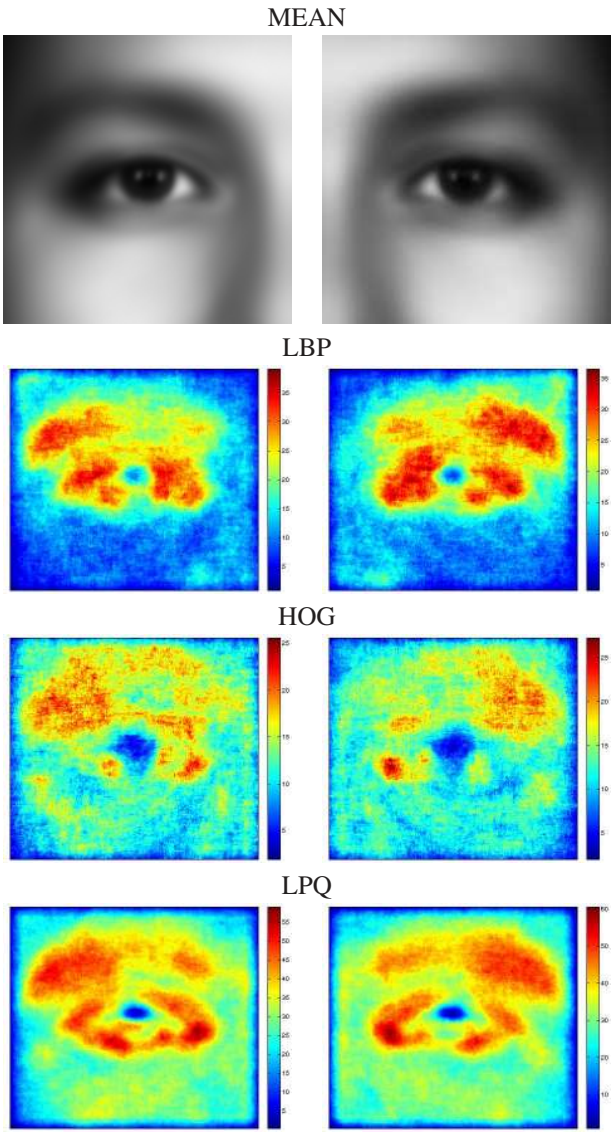


Figure 3.2: Individual Rank-1 recognition rate for a block at each pixel. Results are displayed for each feature representation method in an experiment using FRGC periocular images. Results for right periocular images are shown on the left and vice-versa.

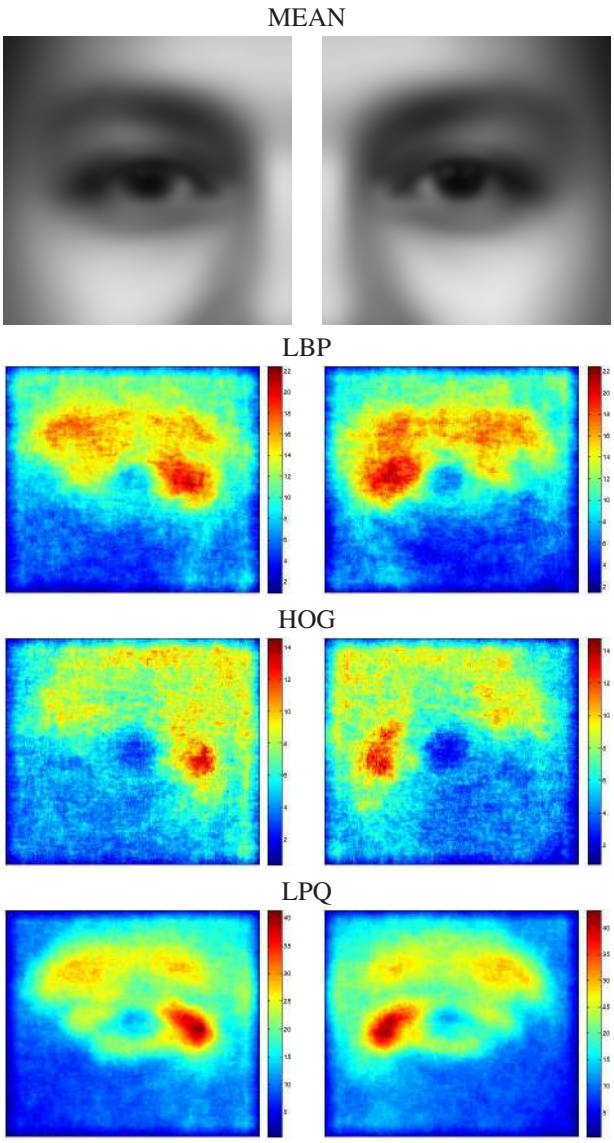


Figure 3.3: Individual Rank-1 recognition rate for a block at each pixel. Results are displayed for each feature representation method in an experiment using FERET periocular images. Results for right periocular images are shown on the left and vice-versa.

Although any one pixel location does not appear to be very discriminative relative to other locations, the skin underneath the eye is a very large region and, when taken as a whole, could provide much better discriminative power than any one smaller portion of it.

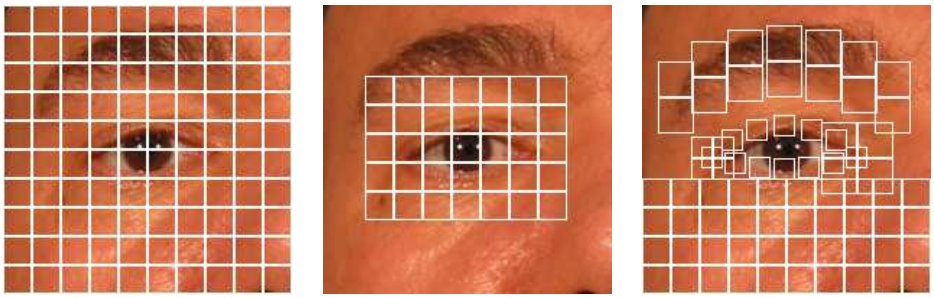
Eye Corners The corners of the eye, particularly the inner corner, appear to be the most discriminative area of the periocular region. The eye corners have properties that could explain its discriminative power. First, the location of the eye corner relative to the center of the eye has a high degree of variability, which provides useful inter-class variability. Second, the eye corners cover the smallest amount of area within the periocular region, which leads to high performance for blocks centered at the few pixels in the eye corners. Third, there is a high density of both shape and texture information present in the eye corners which all the feature types utilized in this work use.

Eyelids The eyelids have perhaps the most variance in discriminative power across different feature types using FRGC and FERET images. Both LBP and LPQ features produce a distinct eyelid shape while HOG features do not, as seen in Figures 3.2 and 3.3. The observation that HOG features do not find eyelids very distinctive may seem counter-intuitive at first because HOG features, in part, utilize something similar to an edge detection algorithm, and the eyelids are perhaps the most visible edges in the periocular region. This may indicate that the textures found by LBP and LPQ are more distinctive than the edges or contours of the eyelid.

Eyebrows Eyebrows appear to have a high degree of discriminative power compared to other physical areas of the periocular region when using LABF, perhaps second only to the eye corners. With that said, it also appears that the outer half of the eyebrow contains the majority of the eyebrow's discriminative power. This observation is slightly stronger in the FRGC experiments than the FERET ones.

3.5 Placement of blocks based on the structure of the periocular region

Previous research involving the periocular region has used the same basic block placement method as the original LBP implementation of face recognition [2]. The blocks have all been rectangular regions of the same size that border each other in a grid pattern and do not overlap. While this is a suitable approach for a



(a) (b) (c)

Figure 3.4: Different models for block placement when using LABF: (a) Miller (b) Park (c) Proposed

basic biometric system, it does not appropriately model the structure of the periocular region as evidenced by the results in Section 3.4. The periocular region is not uniformly discriminative; therefore, a block placement strategy that treats it as so is unlikely to achieve optimal performance.

Section 3.4 described five sub-regions of the periocular region that best represent the different structural areas within the periocular region. After discovering the relatively poor performance of the iris and the difference between the inner and outer portions of the eyebrow, a new model is presented. The block placement in the new model intends to cover seven sub-regions of the periocular region: upper eyelid, lower eyelid, tear duct, outer corner, inner eyebrow, outer eyebrow, and the skin under the eye.

Figure 3.4 shows the proposed block arrangement. There are two important factors in the novelty of this block arrangement. First, the blocks are not placed in a grid pattern like previous approaches. Blocks are placed so that they correspond to the sub-regions listed in Section 3.4. The purpose of this arrangement is to focus the feature extraction on structural elements of the periocular region. It is expected that these sub-regions will be more discriminative than previous standard approaches. Second, the blocks are of variable size, because not all structural elements of the periocular region are the same size. For instance, the eyebrow is much wider than the eyelid. The size of the each block is determined so that a particular sub-region of the periocular region and only that sub-region is contained in the block. The location of the blocks that cover a sub-region of the periocular region as placed so that they cover the physical feature they are intended to cover based on the mean images shown in Figures 3.2 and 3.3.

	Left			Right		
Sub-region	LBP	HOG	LPQ	LBP	HOG	LPQ
upper eyelid	71.2446	53.6481	69.0987	70.8155	52.5751	68.8841
lower eyelid	59.2275	36.0515	57.2961	60.9442	42.7039	59.4421
tear duct	70.3863	47.6395	59.8712	68.6695	45.4936	58.7983
outer corner	81.9742	59.4421	74.4635	82.8326	58.7983	75.1073
inner eyebrow	77.8970	71.4592	87.3391	78.3262	69.9571	84.9785
outer eyebrow	81.3305	72.7468	87.3391	80.4721	74.4635	88.6266
skin	90.9871	90.1288	93.7768	91.4163	91.8455	95.7082

Table 3.1: Rank-1 recognition rates of experiments using FRGC images and only features from certain sub-regions of the periocular region.

3.6 Algorithm performance analysis in different sub-regions

As seen in Section 3.4, not all LABF perform equally in each sub-region within the experiments. Each feature extraction method has types of patterns that it is intended to quantify, and these patterns express themselves differently in the different sub-regions of the periocular region. Experiments should be conducted to test the relative performance of each of the LABF within the different sub-regions. If the performance of each LABF is significantly different within the sub-regions, then an algorithm could be developed, using the proposed periocular structure based block arrangement, to make the best use of the specific patterns found in each sub-region, instead of treating the whole periocular region the same. This approach has the potential to improve the performance of periocular-based biometric systems by using a method fitted to the unique aspects of the periocular region.

To perform an analysis of the algorithmic performance of different LABF within different sub-regions of the periocular region, 42 basic biometric experiments were conducted as previously described in Section 3.3. In these experiments, features extracted from only one sub-region of the periocular region are considered so that the performance of each sub-region can be analyzed individually. These experiments use the same parameters as those listed in Section 3.3. The results presented here are different than those in Section 3.4. In Section 3.4, results are presented for a block placed around an arbitrary pixel. The results in this section are from biometric experiments using concatenated feature vectors of multiple blocks from the same sub-region. Table 3.1 shows the Rank-1 recognition rates of these experiments performed using images from the FRGC dataset. Table 3.2 shows the same using FERET images.

For the upper eyelid, lower eyelid, tear duct, and outer corner, LBP features give the best performance in these experiments using both left and right periocular region images and using both FRGC and FERET images. These results point to the observation that LBP features are more discriminate in this region

	Left			Right		
Sub-region	LBP	HOG	LPQ	LBP	HOG	LPQ
upper eyelid	45.4545	27.9798	44.1414	47.4747	28.6869	46.2626
lower eyelid	34.9495	20.3030	30.8081	34.2424	18.4848	32.0202
tear duct	53.5354	32.7273	48.4848	52.1212	32.5253	49.0909
outer corner	54.1414	34.0404	45.7576	53.6364	35.5556	47.7778
inner eyebrow	69.6970	62.8283	78.4848	69.3939	62.3232	80.0000
outer eyebrow	63.9394	56.8687	73.2323	64.5455	54.8485	73.0303
skin	73.4343	71.3131	75.2525	71.0101	68.5859	73.4343

Table 3.2: Rank-1 recognition rates of experiments using FERET images and only features from certain sub-regions of the periocular region.

than other feature types.

For the inner eyebrow, outer eyebrow, and skin, LPQ features give the best performance in these experiments using both left and right periocular region images and using both FRGC and FERET images. This would suggest that LPQ features are more discriminate in these regions than other feature types.

The performance numbers seen within a single feature extraction method do not seem to correspond to the figures from Section 3.4 at first glance. In most cases the skin under the eye area is the best performing sub-region, but Figures 3.2 and 3.3 do not show patches in the skin area to be very high performing. The figures are showing the performance of a single patch, while the results shown here come from a concatenation of the features from 40 patches. In fact, the skin area contains the most patches while other areas, like the eyebrow, contain less patches and do not see a drastic decrease in performance compared to the skin. One explanation for this behavior would be that the features found in the eyebrow, for instance, are more discriminative than the features found in the skin, and that the performance of the skin sub-region is due in large part to the large number of patches.

3.7 Discriminative ability of LABFs in different sub-regions

To explore the discriminative nature of each feature within each physical sub-region of the periocular region, further analysis is performed on the feature vectors extracted in Section 3.6. It is likely that an individual element within a feature vector, e.g., feature 5 from the 59 feature length LBP feature vector, will present a different range of values when derived from different sub-regions of a periocular image. As an example, assume that a particular element in a feature vector represents a line. Extracting features from the eyelid might cause this element to have a high value because of the presence of lines in the eyelid. Extracting

features from smooth skin might cause a low value for this element. The mean value of this element in the feature vector corresponds to the average presence of a certain feature. D' can also be calculated for an individual element to show the separability of this element when comparing images from the same subject and images from different subjects.

The method for attaining these metrics follows. For the experiments in Section 3.6, features were extracted from 7 different sub-regions of the periocular region. The sub-regions are comprised of a variable number of patches. For instance, the tear duct sub-region contains 4 patches, as seen in Figure 3.4. The complete feature vector from the tear duct sub-region would then be the concatenation of the features extracted from the 4 patches. As discussed previously in Section 2.3.3.1, an LBP feature vector is 59 features long, a HOG feature vector contains 12 elements, and a LPQ feature vector consists of 256 elements. All of the feature vectors for a single image, extracted from patches within each sub-region were summed to get a count of each feature extracted from the different sub-regions within a single image. So, after this step there would be 7 feature vectors for each image per feature extraction method. The feature vector computed from the tear duct sub-region of an image using the LBP feature extraction method would contain 59 elements. Each element within that 59 element feature vector would be the sum of that element taken from the 4 patches found in the tear duct sub-region. This would apply to all sub-regions and all feature extraction methods. Now that each image has a mean feature vector for each sub-region, the feature vectors are summed across all images. So for LBP features, the result is a feature vector with 59 elements where each element in the feature vector represents the sum of the average number of single element features that were extracted from all images. The standard deviation of this population was also calculated.

D' was also calculated for each individual element in the feature vector. This is done by finding the difference in number of individual feature elements across every image comparison in the dataset. These comparisons were then split into two groups: comparisons between images of the same subject and comparisons between images of different subjects. D' is then calculated to show the separability of this single feature element with the particular periocular sub-region.

3.7.1 Local Binary Patterns

The 59 different elements in the LBP feature vector correspond to 59 different patterns of pixel intensity difference relative to the center pixel. Figure 3.5 shows each pattern. A black circle represents a pixel that is darker than the center pixel while a white circle represents a pixel that is lighter than or equal to the center pixel. The last pattern represents all the occurrences of patterns not specified by the first 58. As

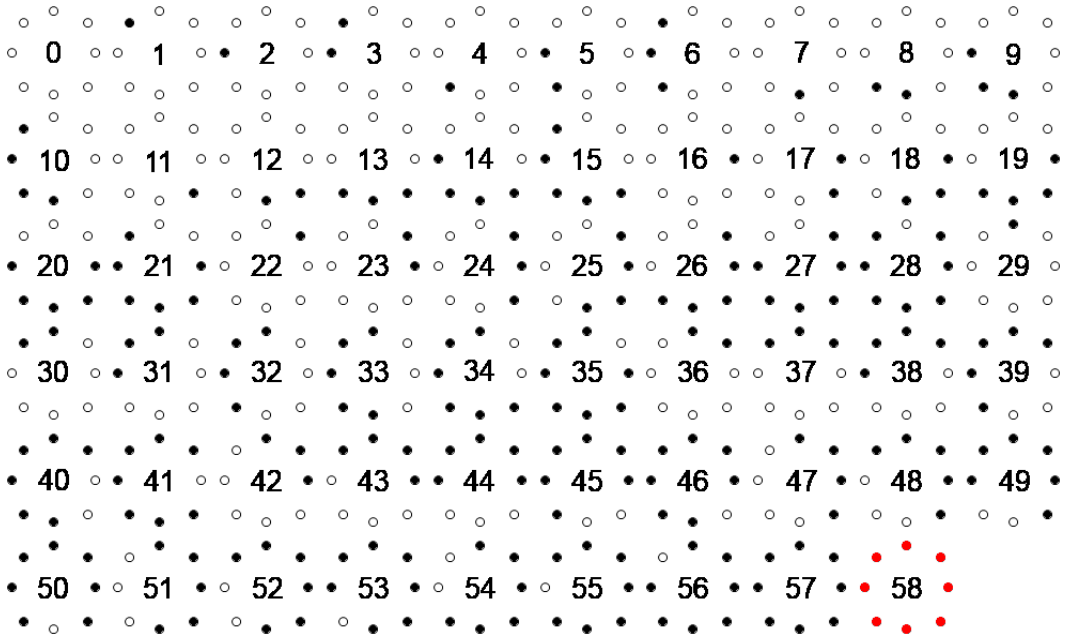
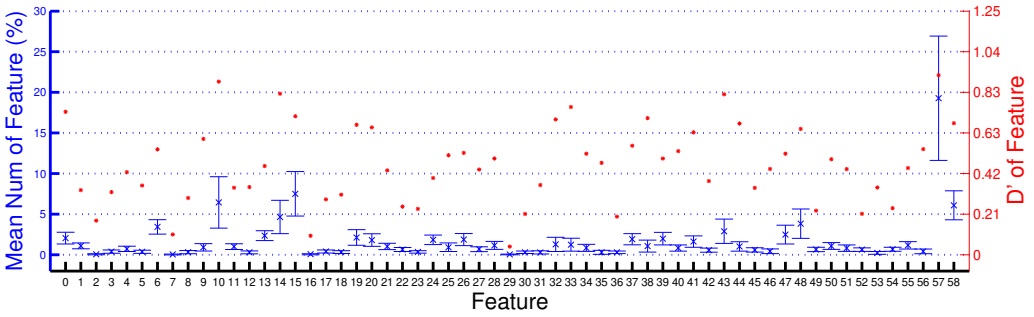


Figure 3.5: Visualization of the 59 different LBP patterns

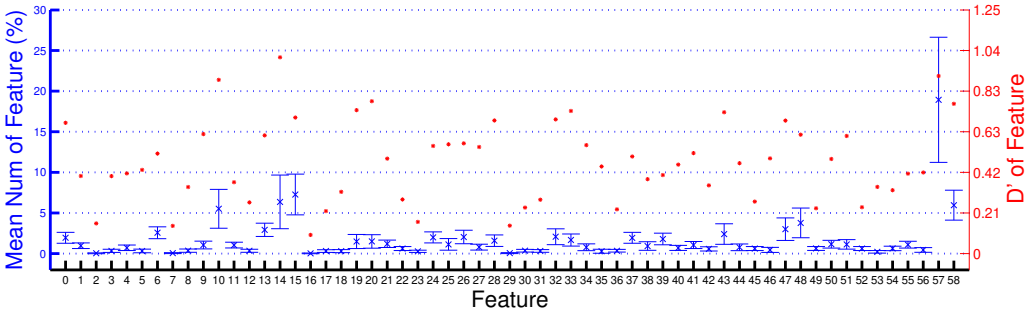
an example, pattern 44 would represent a part of a line found at the bottom of a dark area, pattern 57 would represent a bright spot within a dark area, etc..

Figure 3.6 shows the mean occurrence of each of the 59 LBP features that were extracted from the Upper Eyelid as a percent of the total number of features. The error bars around the mean point represent one standard deviation above and below the mean. Additionally, the D' value for each individual feature is displayed. A pattern is immediately apparent across all 4 types of images. The mean values are very similar across each. Using a single-tailed t-test, it is possible to determine which features occur at a statistically significant higher rate than the mean of all features at p = 0.05. From the four different data sets (left eye FRGC images, right eye FRGC images, left eye FERET images, and right eye FERET images), the following features were extracted at a statistically significant higher rate in all four data sets: 6, 10, 13, 14, 15, 24, 26, 37, 39, 43, 47, 48, 57, and 58. Only two other features were found at a significantly higher rate than average in less than three of the data sets. This level of similarity suggests that the occurrences of certain LBP features within the Upper Eyelid are universal and likely to be found in other datasets as well. Even though Tables 3.1 and 3.2 show a great difference in Rank-1 recognition rates between FRGC and FERET images, the pattern of features found in the Upper Eyelid are mostly constant across the four datasets. This pattern of highly occurring features could be viewed as an LBP feature footprint for the Upper Eyelid.

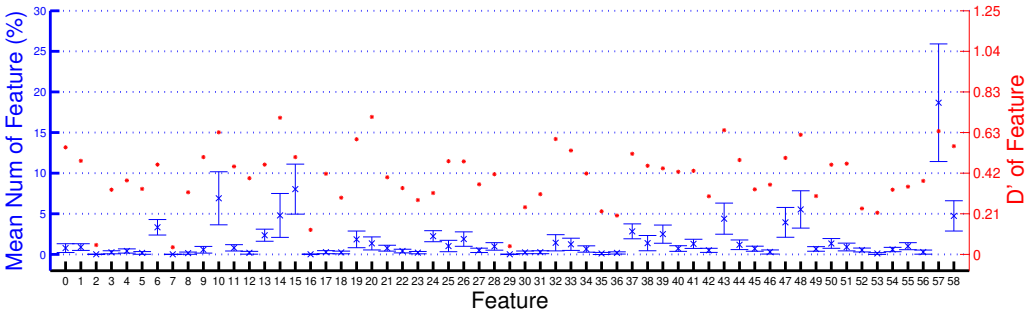
(a) Upper Eyelid, FRGC, left eye



(b) Upper Eyelid, FRGC, right eye



(c) Upper Eyelid, FERET, left eye



(d) Upper Eyelid, FERET, right eye

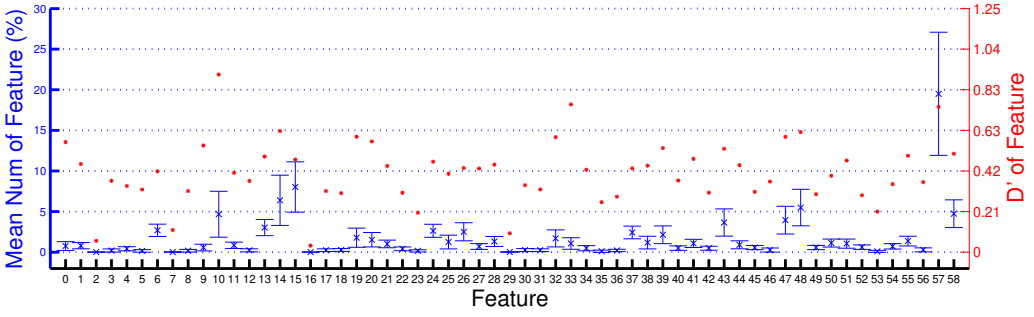
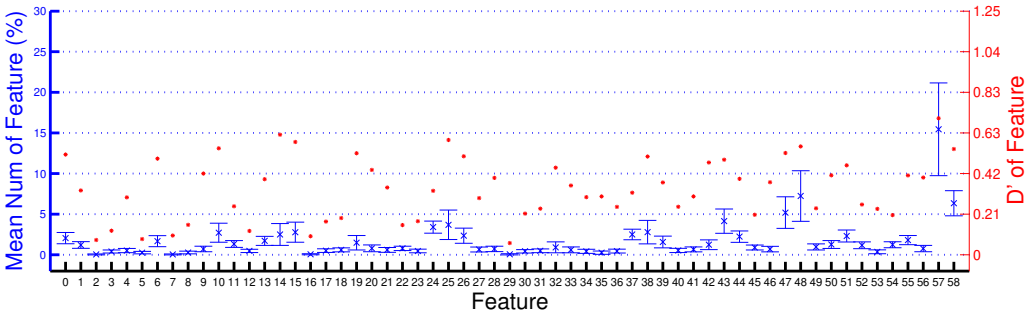
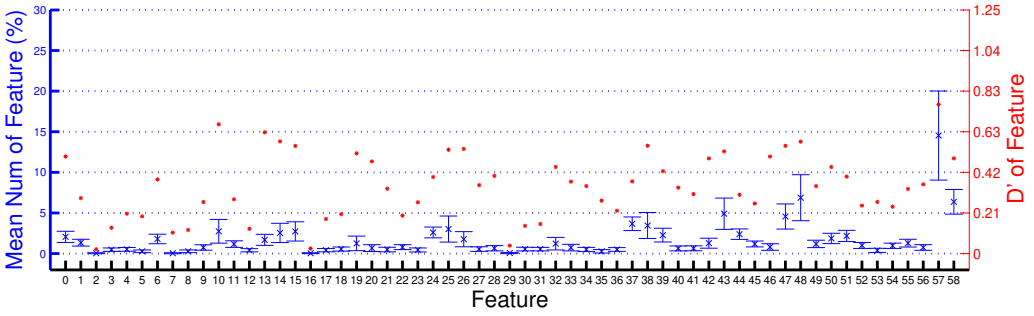


Figure 3.6: Mean occurrence and D' of the LBP patterns found in the **Upper Eyelid. Features extracted from (a) left eye images of the FRGC dataset (b) right eye images of the FRGC dataset (c) left eye images of the FERET dataset (d) right eye images of the FERET dataset**

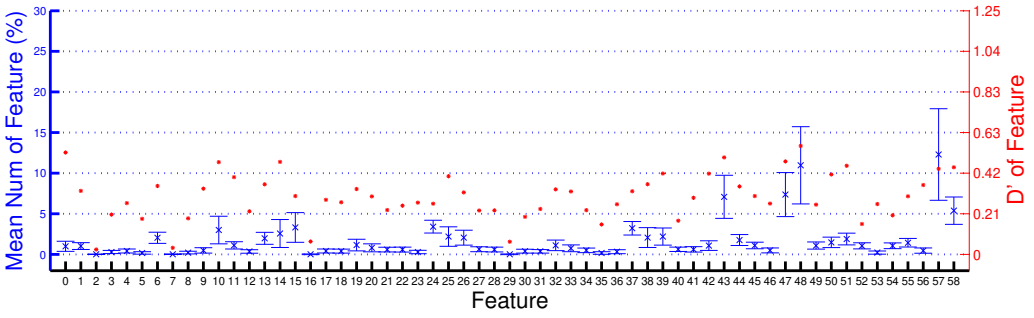
(a) Lower Eyelid, FRGC, left eye



(b) Lower Eyelid, FRGC, right eye



(c) Lower Eyelid, FERET, left eye



(d) Lower Eyelid, FERET, right eye

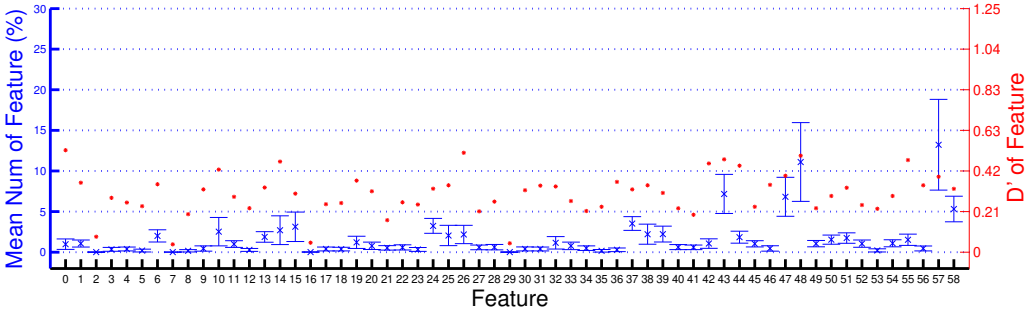
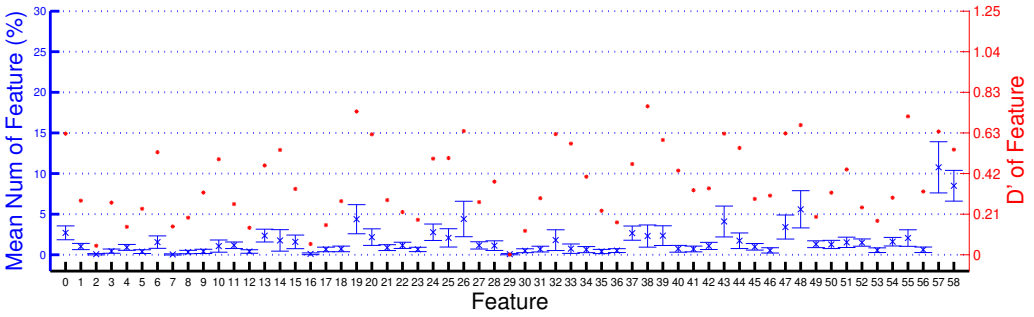
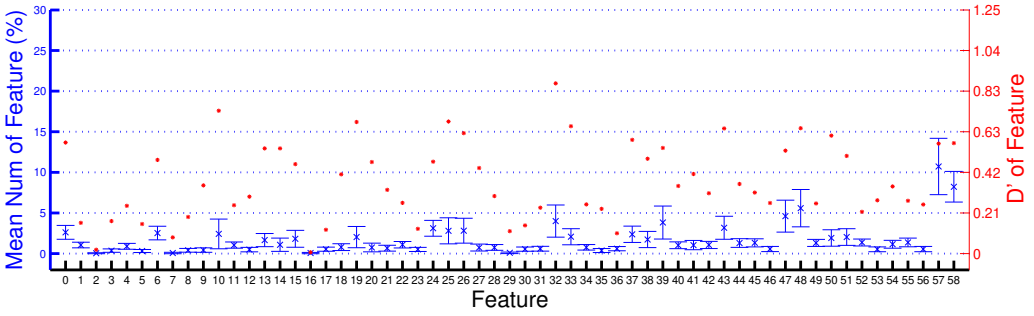


Figure 3.7: Mean occurrence and D' of the LBP patterns found in the **Lower Eyelid. Features extracted from (a) left eye images of the FRGC dataset (b) right eye images of the FRGC dataset (c) left eye images of the FERET dataset (d) right eye images of the FERET dataset**

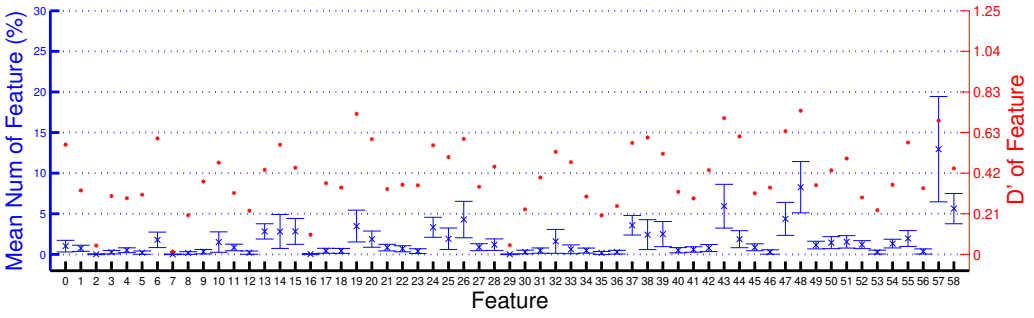
(a) Tear Duct, FRGC, left eye



(b) Tear Duct, FRGC, right eye



(c) Tear Duct, FERET, left eye



(d) Tear Duct, FERET, right eye

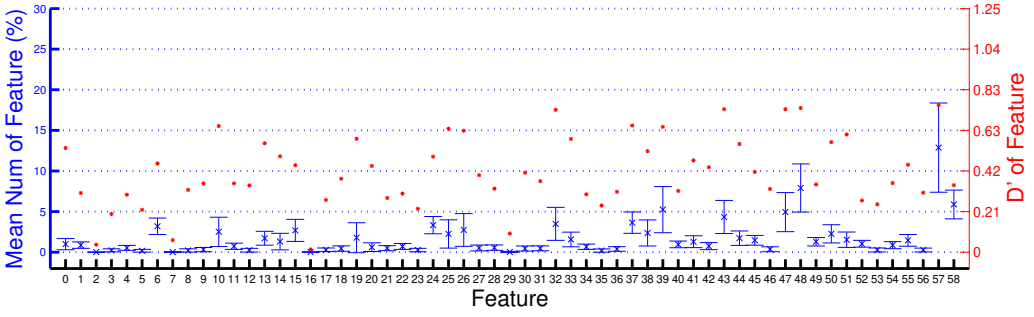
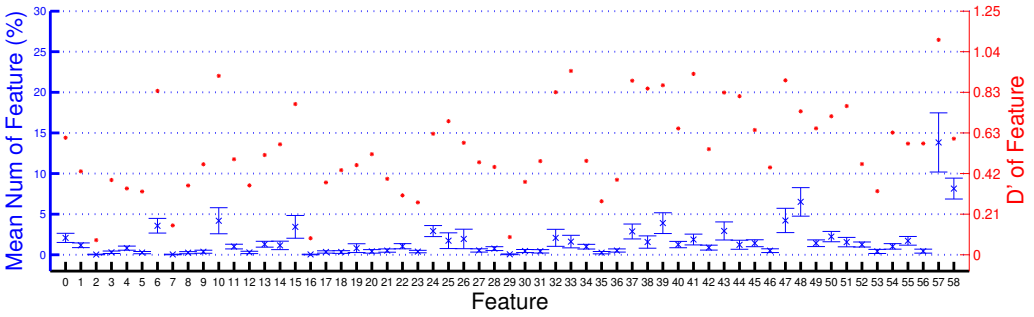
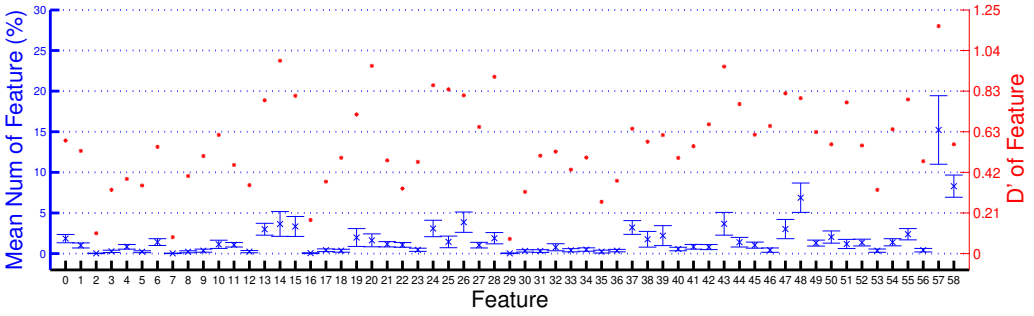


Figure 3.8: Mean occurrence and D' of the LBP patterns found in the **Tear Duct. Features extracted from (a) left eye images of the FRGC dataset (b) right eye images of the FRGC dataset (c) left eye images of the FERET dataset (d) right eye images of the FERET dataset**

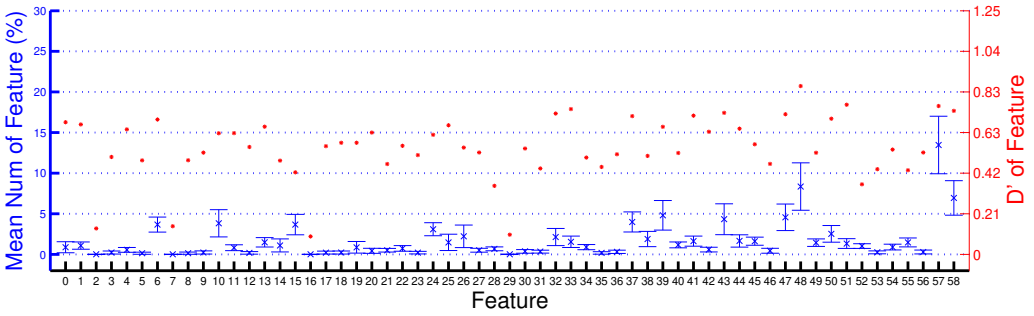
(a) Outer Corner, FRGC, left eye



(b) Outer Corner, FRGC, right eye



(c) Outer Corner, FERET, left eye



(d) Outer Corner, FERET, right eye

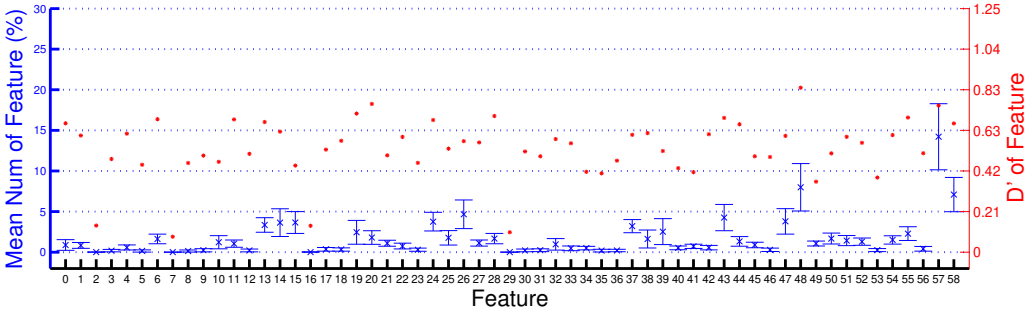
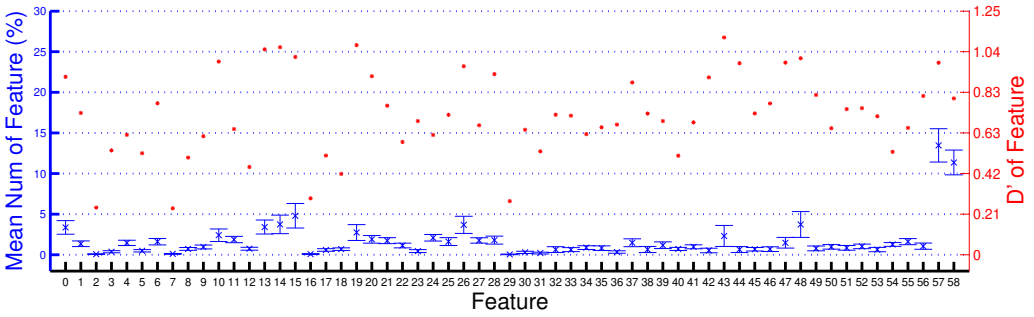
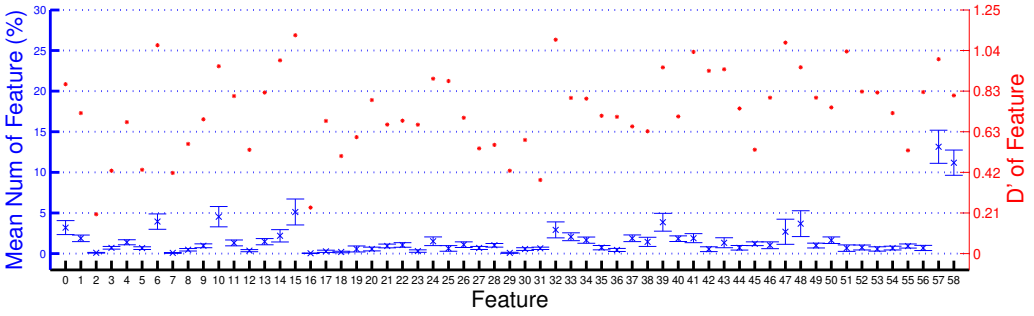


Figure 3.9: Mean occurrence and D' of the LBP patterns found in the **Outer Corner. Features extracted from (a) left eye images of the FRGC dataset (b) right eye images of the FRGC dataset (c) left eye images of the FERET dataset (d) right eye images of the FERET dataset**

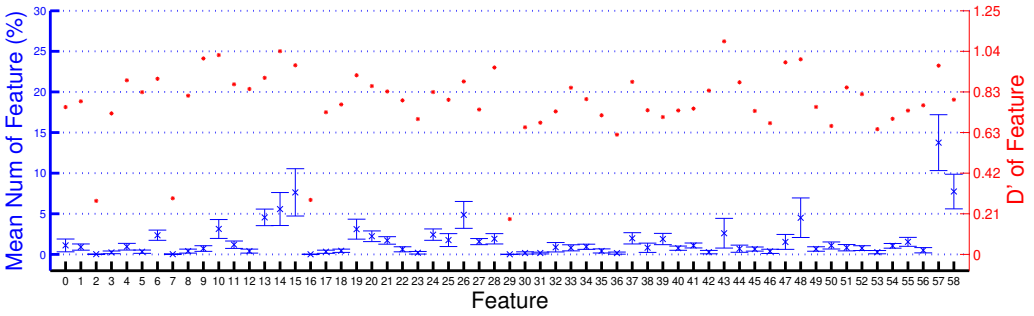
(a) Inner Eyebrow, FRGC, left eye



(b) Inner Eyebrow, FRGC, right eye



(c) Inner Eyebrow, FERET, left eye



(d) Inner Eyebrow, FERET, right eye

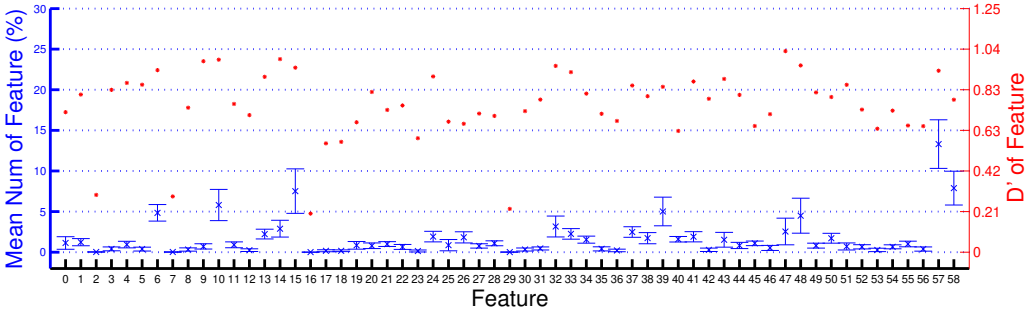
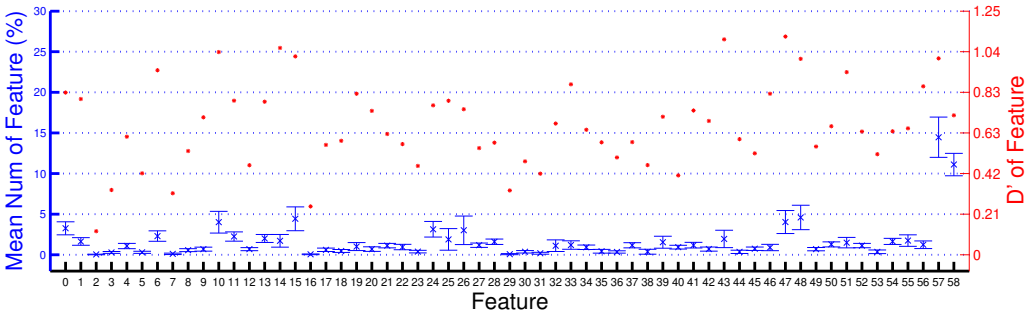
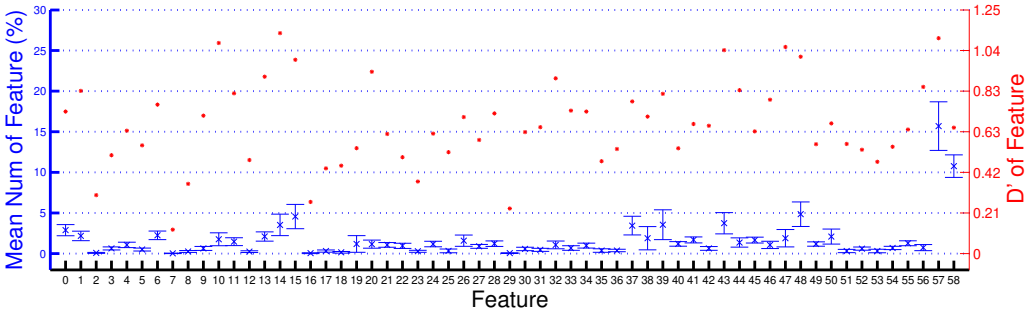


Figure 3.10: Mean occurrence and D' of the LBP patterns found in the **Inner Eyebrow. Features extracted from (a) left eye images of the FRGC dataset (b) right eye images of the FRGC dataset (c) left eye images of the FERET dataset (d) right eye images of the FERET dataset**

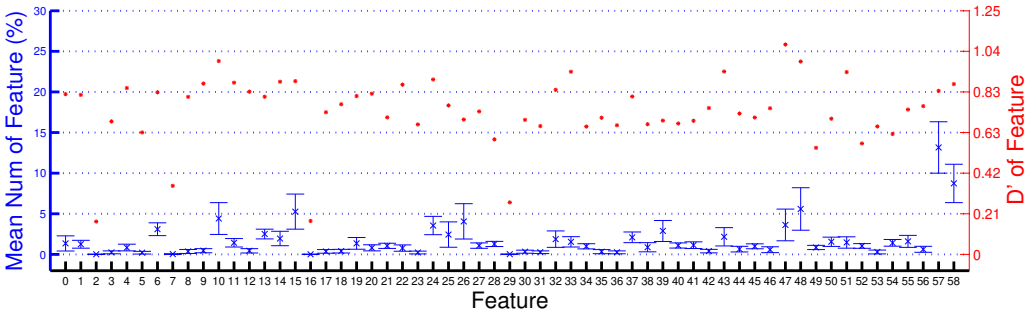
(a) Outer Eyebrow, FRGC, left eye



(b) Outer Eyebrow, FRGC, right eye



(c) Outer Eyebrow, FERET, left eye



(d) Outer Eyebrow, FERET, right eye

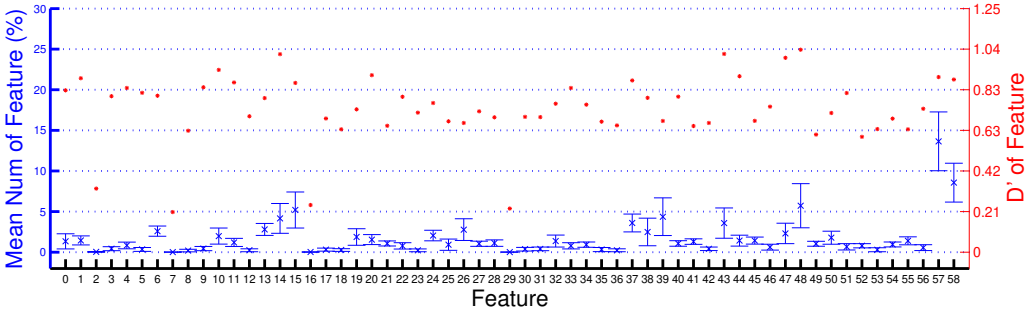
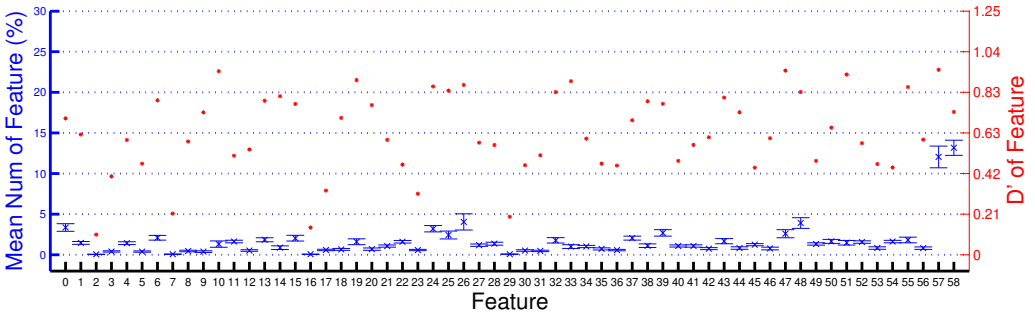
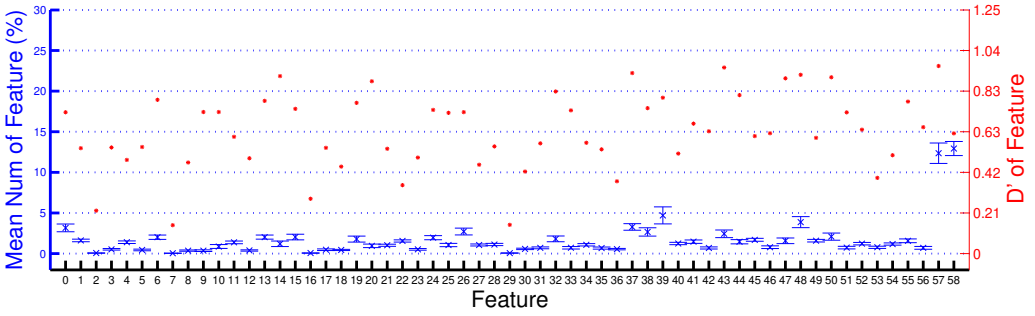


Figure 3.11: Mean occurrence and D' of the LBP patterns found in the **Outer Eyebrow. Features extracted from (a) left eye images of the FRGC dataset (b) right eye images of the FRGC dataset (c) left eye images of the FERET dataset (d) right eye images of the FERET dataset**

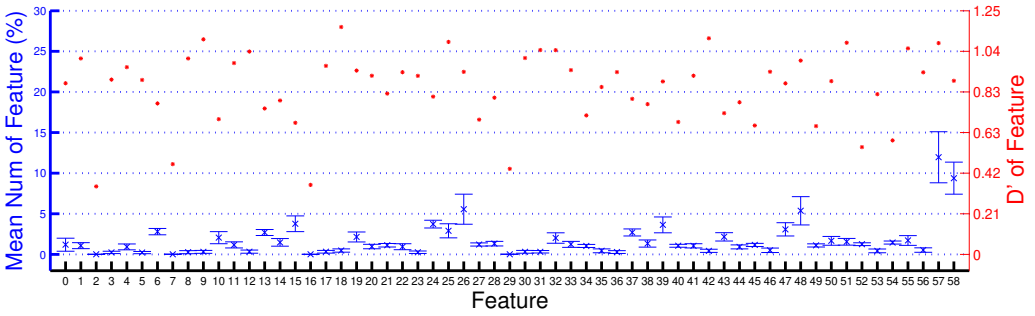
(a) Skin, FRGC, left eye



(b) Skin, FRGC, right eye



(c) Skin, FERET, left eye



(d) Skin, FERET, right eye

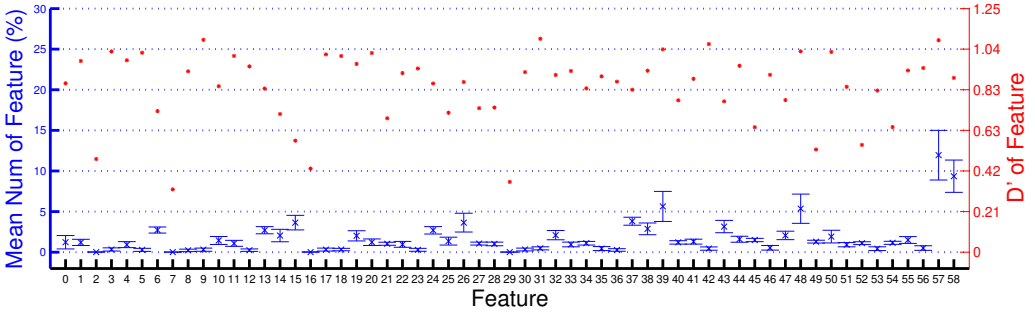


Figure 3.12: Mean occurrence and D' of the LBP patterns found in the **Skin. Features extracted from (a) left eye images of the FRGC dataset (b) right eye images of the FRGC dataset (c) left eye images of the FERET dataset (d) right eye images of the FERET dataset**

Sub-region	FRGC Left Eye	FRGC Right Eye	FERET Left Eye	FERET Right Eye
upper eyelid	0.3708	0.4054	0.2833	0.3152
lower eyelid	0.4264	0.4369	0.3859	0.2366
tear duct	0.4034	0.3872	0.4800	0.4727
outer corner	0.4070	0.3894	0.2818	0.2692
inner eyebrow	0.2162	0.2164	0.2080	0.2206
outer eyebrow	0.2605	0.2637	0.1702	0.2029
skin	0.1767	0.1598	0.0293	0.0366

Table 3.3: Coefficients of determination from a fitted regression line of mean and D' values for LBP experiments

There is not a strong correlation between the mean values and the D' values. Table 3.3 displays the coefficients of determination (R^2) for a fitted regression line to a scatter plot of mean and D' values for each of the experiments analyzed in this section. The average R^2 is 0.2897. To explain this low correlation it would be helpful to review the formula for computing D'. The relative difference of the mean and the standard deviation influences the separability of the feature. While the pattern of means can tell us what an Upper Eyelid looks like in terms of LBP features, the D' explains how useful each feature is in identifying a subject.

Figures 3.7 - 3.12 show the results for the remaining sub-regions of the periocular region. It might not be immediately apparent in the figures, but some of the sub-regions produce feature footprints that are slightly different when the features are extracted from the left eye versus the right eye. We can determine the extent of this difference by computing the sum of the absolute difference between features extracted from the left eye and features extracted from the right eye. Table 3.4 shows the distance calculated for each sub-region. The numbers in this table are a unitless representation of the difference between the occurrence of features from the right eye images and the left eye images, where a larger number represents a larger difference. Notice that the four largest distance sub-regions are the four sub-regions that have inner/outer counter-parts. These results show two important aspects of the periocular region. First, the upper and lower eyelids have a high degree of symmetry. This high symmetry is seen even when averaging the left features from both FRGC and FERET before comparing them to the right features from the same datasets. Second, features extracted from regions such as the tear duct are different when comparing features from the left eye and the right eye. The shape of the tear duct from the left eye mirrors the tear duct on the right eye. This results in a different feature footprint and suggests that it would not be helpful to talk about the feature footprint of these sub-regions without distinguishing which side of the face they came from.

Table 3.5 shows the LBP feature footprint for all of the sub-regions. There are only three features

Sub-region	left/right distance
upper eyelid	16.1167
lower eyelid	8.2918
tear duct	34.7406
outer corner	40.0833
inner eyebrow	49.4215
outer eyebrow	42.9436
skin	22.8981

Table 3.4: Distance between the features extracted from the left eye and features extracted from the right eye per sub-region

Sub-region	Feature Footprint
upper eyelid	6, 10, 13, 14, 15, 24, 26, 37, 39, 43, 47, 48, 57, 58
lower eyelid	10, 14, 15, 24, 25, 26, 37, 38, 43, 44, 47, 48, 51, 57, 58
left tear duct	13, 14, 19, 20, 24, 25, 26, 37, 38, 39, 43, 47, 48, 55, 57, 58
right tear duct	6, 10, 15, 19, 24, 25, 26, 32, 37, 39, 43, 47, 48, 50, 57, 58
left outer corner	6, 10, 15, 24, 26, 32, 37, 39, 43, 47, 48, 50, 57, 58
right outer corner	13, 14, 15, 19, 24, 26, 37, 39, 43, 47, 48, 55, 57, 58
left inner eyebrow	10, 13, 14, 15, 19, 20, 21, 24, 26, 28, 43, 48, 57, 58
right inner eyebrow	6, 10, 14, 15, 32, 33, 37, 39, 41, 47, 48, 57, 58
left outer eyebrow	6, 10, 13, 15, 24, 25, 26, 43, 47, 48, 57, 58
right outer eyebrow	6, 10, 13, 14, 15, 37, 38, 39, 43, 47, 48, 50, 57, 58
skin	6, 13, 15, 24, 26, 32, 37, 39, 48, 57, 58

Table 3.5: LBP feature footprint for each sub-region

that are seen in every sub-region, 48, 57, and 58. These three features are only a few of the total number of features found in the different LBP feature footprints. There is a lot of variable information with which to distinguish between the different sub-regions.

The Upper Eyelid and the Lower Eyelid share eleven of the same features in their feature footprint. This large overlap is to be expected considering both sub-regions are looking at eyelids. The Upper Eyelid's feature footprint has three features that are not found in the Lower Eyelid and there are four features vice versa. One possible reason for the differences, small as they are, is the presence of the upper eye fold that is often caught with the Upper Eyelid. This physical aspect of the periocular region adds more and different information to discriminate when extracting LBP features.

The right tear duct and the left tear duct also share eleven of the same features in their feature footprint. However, the right tear duct's feature footprint has five features that are not found in the left tear duct and five features vice versa. There is more to differentiate the left and right feature footprints of the same physical aspect, in this case, than there is to differentiate between two different physical aspects in the upper and lower eyelid. This supports the need to treat the left and right sides of these particular sub-regions separately when discussing the feature footprint.

Table 3.6 shows the mean D' of the LBP features extracted from each sub-region. Table 3.7 displays the Rank-1 recognition results from Section 3.6 in a format that mirrors Table 3.6 for easy comparison. It might appear that Tables 3.6 and 3.7 show inconsistent results. For instance, the mean D' values of the skin in FERET experiments is much higher than the mean D' of the skin in FRGC experiments while the Rank-1 recognition rates do not show the same relationship. These results are not inconsistent, they just cannot be directly compared to each other. Table 3.6 shows the mean D' of 59 features which are themselves the sum of all features extracted from a particular sub-region of the periocular region. The Rank-1 recognition rates shown in Table 3.7 are computed from the comparison of the unmodified, raw feature vector. The more patches in the sub-region, the less related the mean D' and Rank-1 results become; the skin has the largest number of patches of all sub-regions of the periocular region.

It is possible that there are some features in the LBP feature vector that, by being included in computing the similarity of biometric feature vectors, reduce the recognition performance of a biometric system. This hypothesis comes from the observations about certain individual features having a higher mean occurrence or D' value when compared to all of the features. Unfortunately, testing all 2^{59} possible feature vector configurations to find the best performing set of features is computationally infeasible. Using a genetic algorithm to address this problem would produce an optimal or near-optimal set of features that performs best

Sub-region	Left Eye FRGC	Right Eye FRGC	Left Eye FERET	Right Eye FERET
upper eyelid	0.4619	0.4707	0.4019	0.4121
lower eyelid	0.3393	0.3471	0.2976	0.2996
tear duct	0.3732	0.3729	0.4027	0.4130
outer corner	0.5519	0.5672	0.5499	0.5326
inner eyebrow	0.7168	0.7334	0.7793	0.7516
outer eyebrow	0.6645	0.6638	0.7367	0.7350
skin	0.6243	0.6264	0.8526	0.8482

Table 3.6: Mean D' of LBP features extracted from each sub-region

Sub-region	Left Eye FRGC	Right Eye FRGC	Left Eye FERET	Right Eye FERET
upper eyelid	71.2446	70.8155	45.4545	47.4747
lower eyelid	59.2275	60.9442	34.9495	34.2424
tear duct	70.3863	68.6695	53.5354	52.1212
outer corner	81.9742	82.8326	54.1414	53.6364
inner eyebrow	77.8970	78.3262	69.6970	69.3939
outer eyebrow	81.3305	80.4721	63.9394	64.5455
skin	90.9871	91.4163	73.4343	71.0101

Table 3.7: Rank 1 recognition rate of LBP features extracted from each sub-region

with many less computations.

Genetic algorithms (GA), inspired by the natural process of evolution, seek to approach an optimal solution to a problem through iterative changes to a group of candidate solutions. The GA used in this work is comprised of six steps. The first step is to initialize a set of candidate solutions. The problem of selecting a subset of the 59 LBP features can be interpreted as producing a 59 bit mask where 1 represents a feature being used and 0 represents a feature not being used. So, an initial candidate solution would be a 59 bit vector where the values at each bit were randomly selected. An initial population of 32 candidate solutions was produced by randomly generating bit masks. The size of the population can be any power of 2 where a larger population allows for better possible candidate solutions at a greater computational expense. The second step is to compute the performance of each candidate solution. This is done for each of the sub-regions of the periocular region in each of the datasets. For each patch within a sub-region, the elements of the LBP feature vector are either included or excluded based on the values in the candidate solution mask. The masked feature vectors are then used in the standard biometric experiment as performed everywhere else in this dissertation. The third step is to place the entire population of candidate solutions into a binary tournament where the candidate solution with the best performance score moves on to each successive round of the tournament until there are two remaining candidate solutions. These two winning candidate solutions

	All Features	Candidate Solution
Left Eye FRGC	71.2446	73.8197
Right Eye FRGC	70.8155	73.1760
Left Eye FERET	45.4545	48.0808
Right Eye FERET	47.4747	49.6970

Table 3.8: Rank 1 recognition rate of LBP features extracted from the upper eyelid

are used in step four, uniform crossover. In this step, a child candidate solution is created by randomly selecting the value of each bit from one of the two parent candidate solutions. The fifth step is random mutation. Each bit of the child candidate solution is either flipped to the opposite value or remains the same based on a rate of mutation variable. A mutation rate of 10% was used here. The sixth step is to replace the worst scored candidate solution with the newly created child candidate solution. Steps three through six are then repeated 100,000 times. At the end of this process the genetic algorithm will produce a population of candidate solutions that have been tuned to solve the problem at hand.

This approach was used to produce an optimal selection of features from the LBP feature vectors of each sub-region of the periocular region. Table 3.8 shows the Rank-1 recognition rate of features taken from the upper eyelid of each dataset as well as the Rank-1 recognition rate of the best performing candidate solution. These results show that the approach of using optimally selected features is capable of producing better results than using all features. In addition it shows that there are some features that, when used, lower the recognition performance of the biometric system. Even though the results of LBP features from the upper eyelid only are shown in Table 3.8, a candidate solution from every other sub-region produced better results than all features as well.

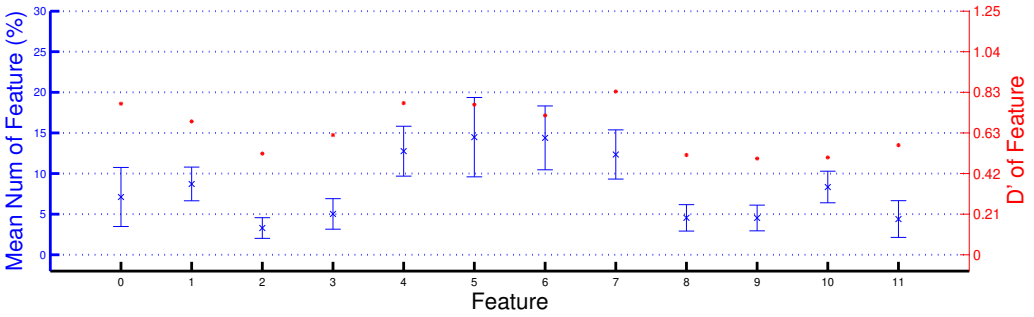
3.7.2 Histogram of Oriented Gradients

The 12 different elements in the HOG feature vector correspond to 30 degree segments of a circle in which the magnitude of the gradient of the biometric image at each pixel is applied to the segment representing the gradient angle of the pixel.

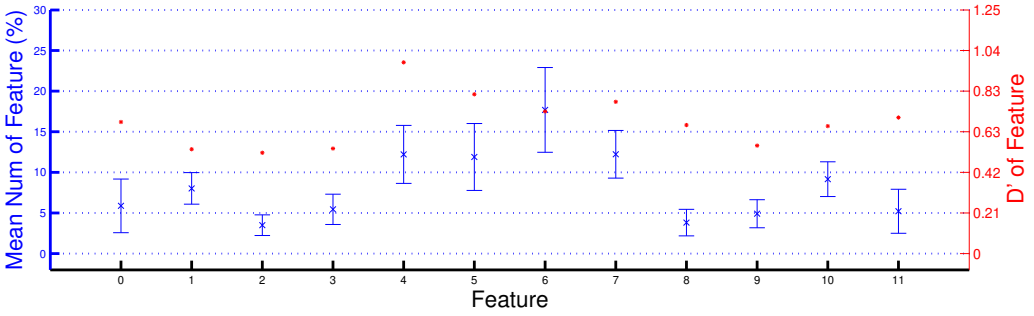
Figures 3.13 - 3.19 show the mean occurrence of each of the 12 HOG features and their D' that were extracted from each sub-region of the periocular region. Strong patterns are visible in these figures as in the LBP figures, though the difference between left and right features seems more pronounced. Table 3.9 shows the HOG feature footprint for all of the sub-regions.

Like the feature footprints from LBP features, the HOG feature footprints show a lot of variety.

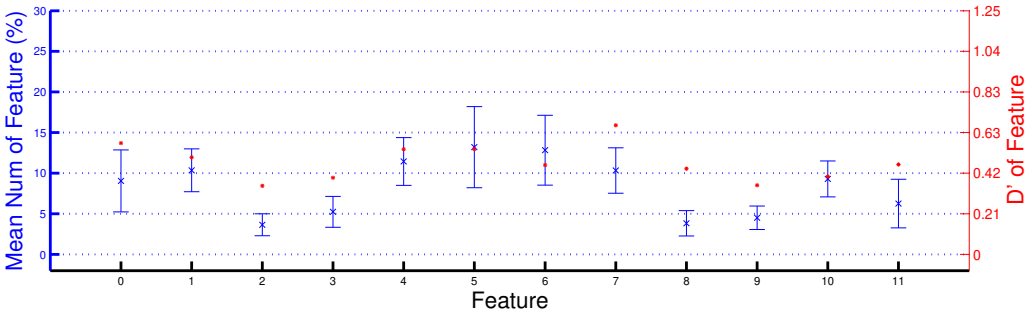
(a) Upper Eyelid, FRGC, left eye



(b) Upper Eyelid, FRGC, right eye



(c) Upper Eyelid, FERET, left eye



(d) Upper Eyelid, FERET, right eye

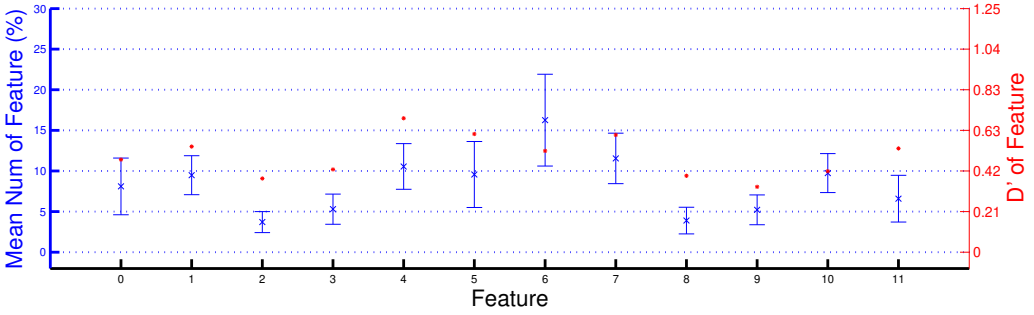
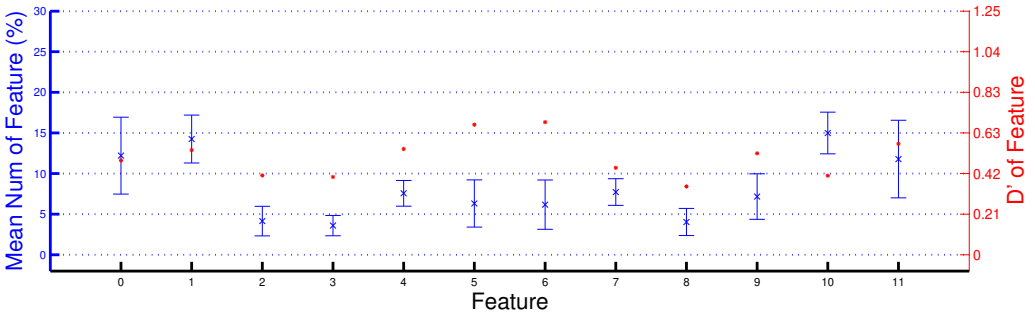
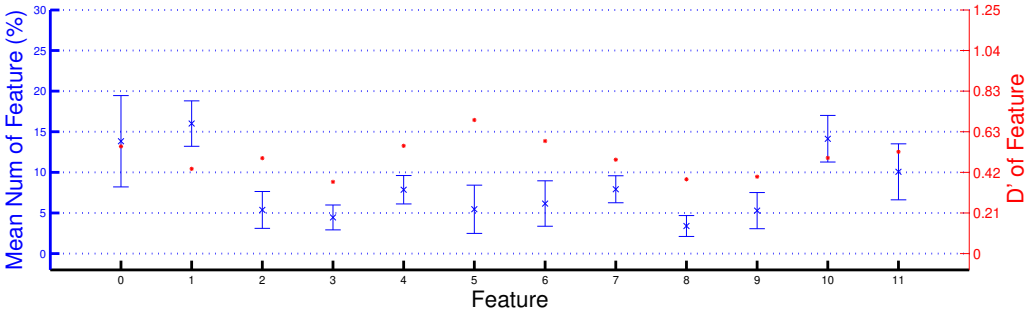


Figure 3.13: Mean occurrence and D' of the HOG patterns found in the **Upper Eyelid. Features extracted from (a) left eye images of the FRGC dataset (b) right eye images of the FRGC dataset (c) left eye images of the FERET dataset (d) right eye images of the FERET dataset**

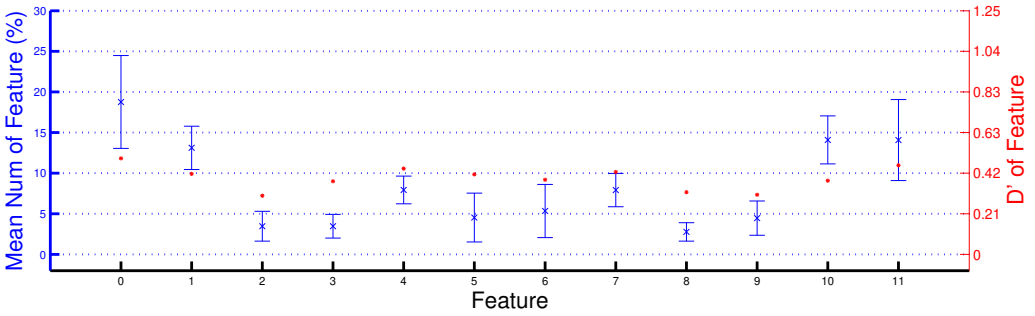
(a) Lower Eyelid, FRGC, left eye



(b) Lower Eyelid, FRGC, right eye



(c) Lower Eyelid, FERET, left eye



(d) Lower Eyelid, FERET, right eye

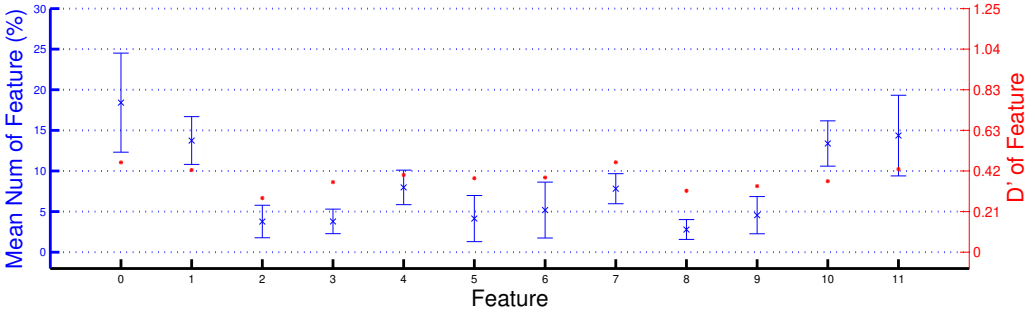
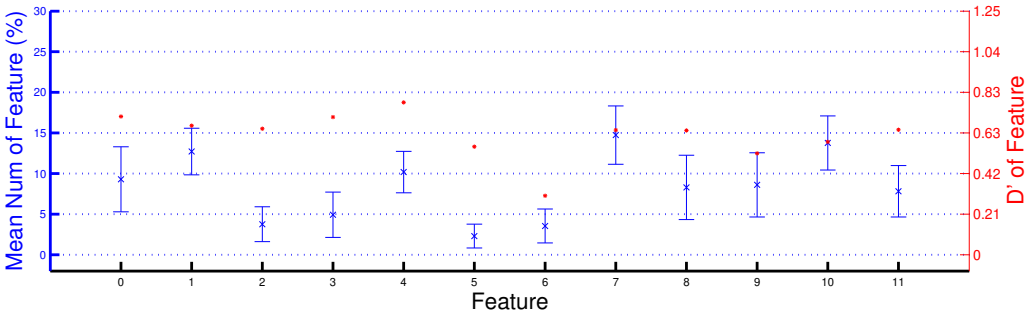
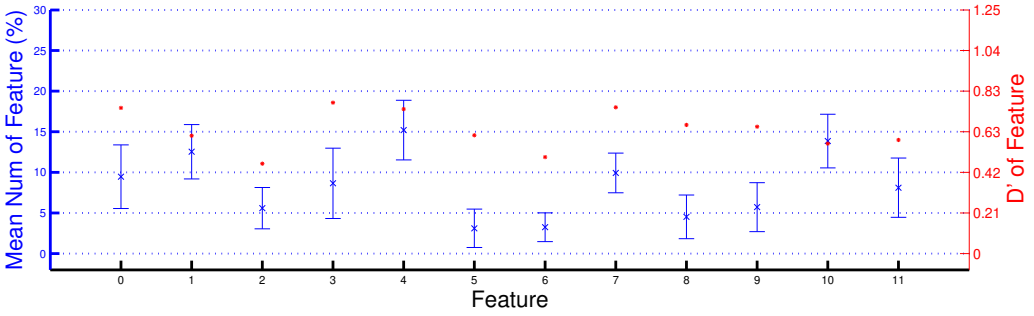


Figure 3.14: Mean occurrence and D' of the HOG patterns found in the **Lower Eyelid. Features extracted from (a) left eye images of the FRGC dataset (b) right eye images of the FRGC dataset (c) left eye images of the FERET dataset (d) right eye images of the FERET dataset**

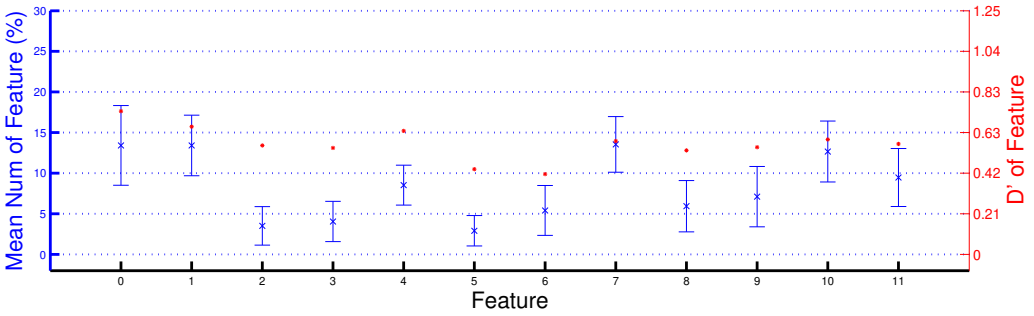
(a) Tear Duct, FRGC, left eye



(b) Tear Duct, FRGC, right eye



(c) Tear Duct, FERET, left eye



(d) Tear Duct, FERET, right eye

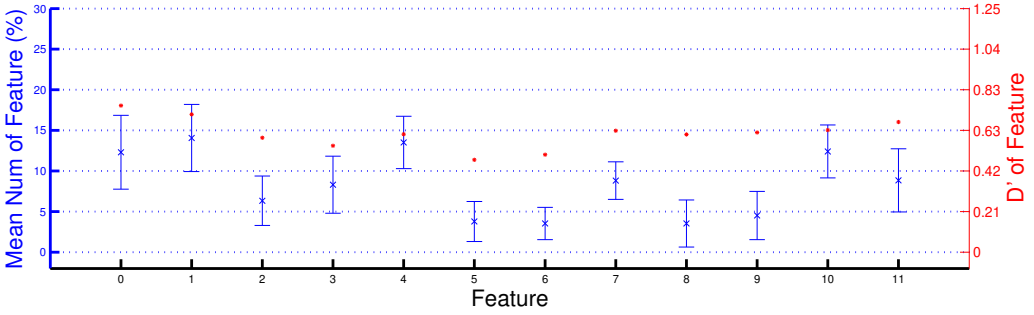
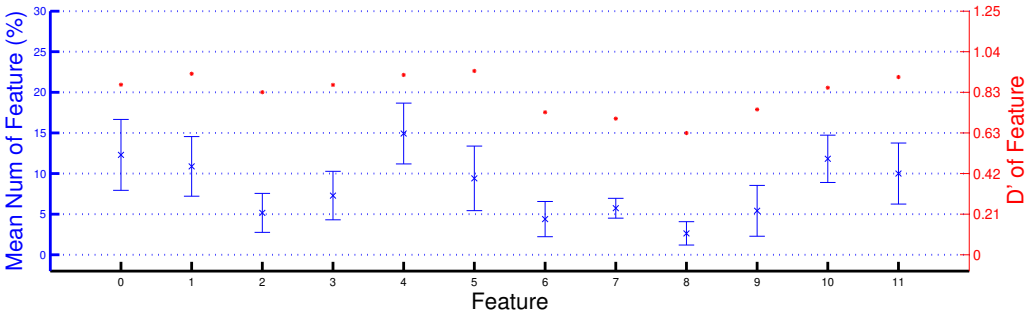
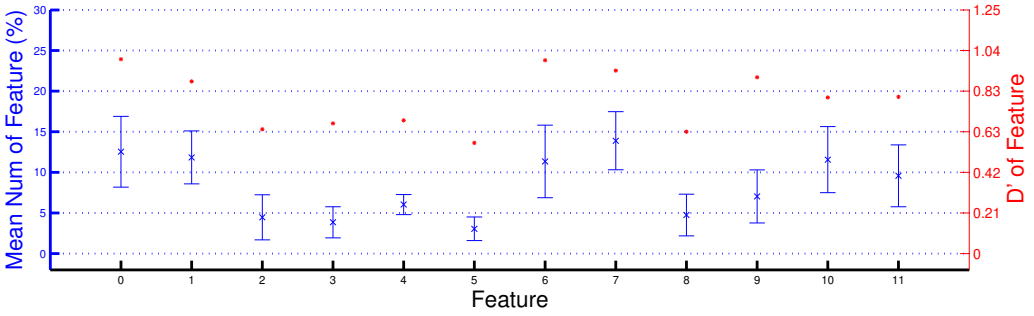


Figure 3.15: Mean occurrence and D' of the HOG patterns found in the **Tear Duct. Features extracted from (a) left eye images of the FRGC dataset (b) right eye images of the FRGC dataset (c) left eye images of the FERET dataset (d) Fright eye images of the FERET dataset**

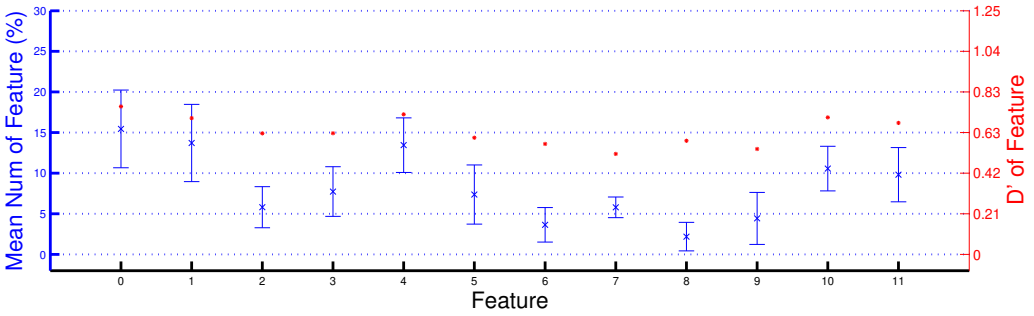
(a) Outer Corner, FRGC, left eye



(b) Outer Corner, FRGC, right eye



(c) Outer Corner, FERET, left eye



(d) Outer Corner, FERET, right eye

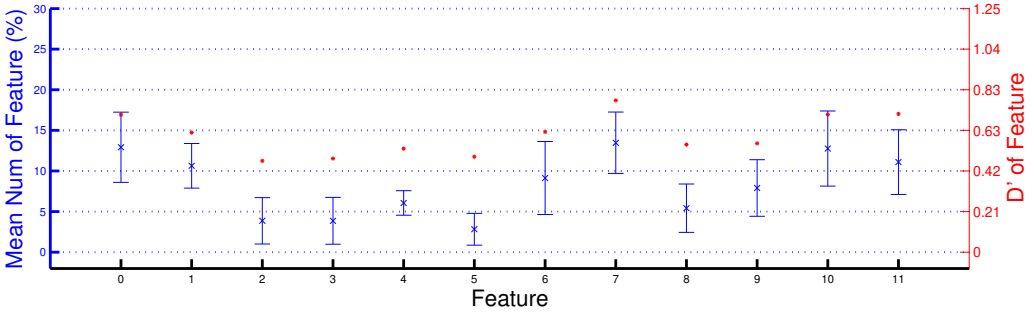
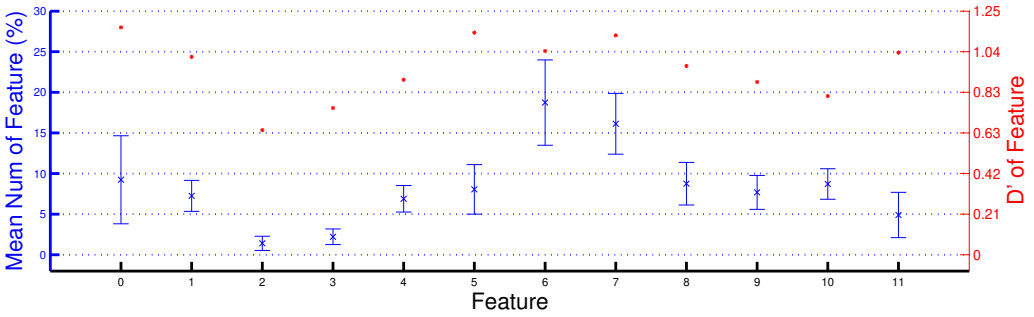
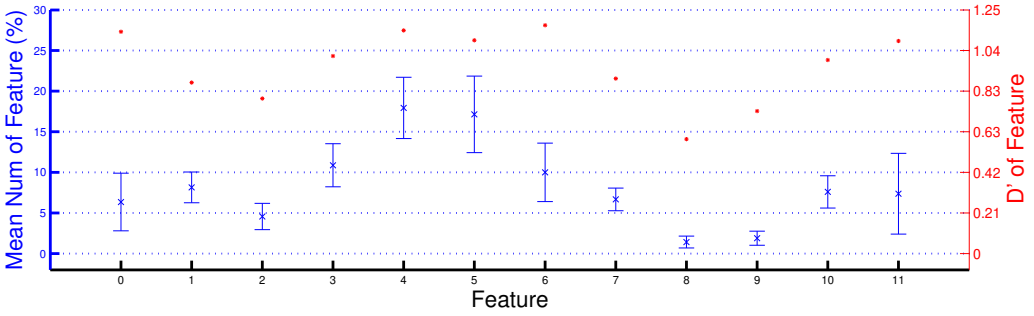


Figure 3.16: Mean occurrence and D' of the HOG patterns found in the **Outer Corner. Features extracted from (a) left eye images of the FRGC dataset (b) right eye images of the FRGC dataset (c) left eye images of the FERET dataset (d) right eye images of the FERET dataset**

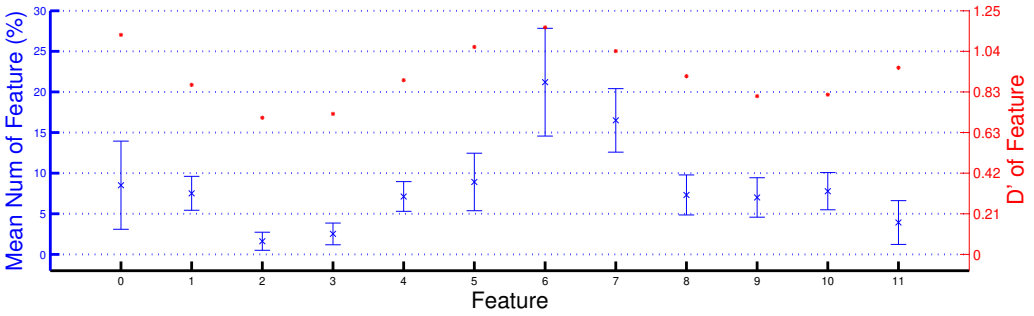
(a) Inner Eyebrow, FRGC, left eye



(b) Inner Eyebrow, FRGC, right eye



(c) Inner Eyebrow, FERET, left eye



(d) Inner Eyebrow, FERET, right eye

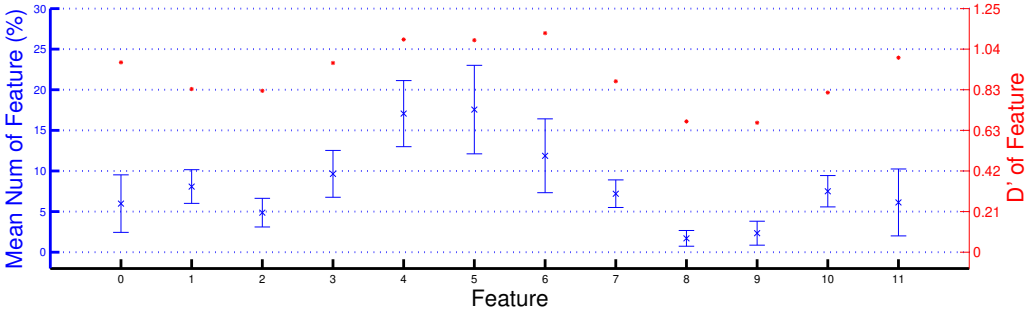
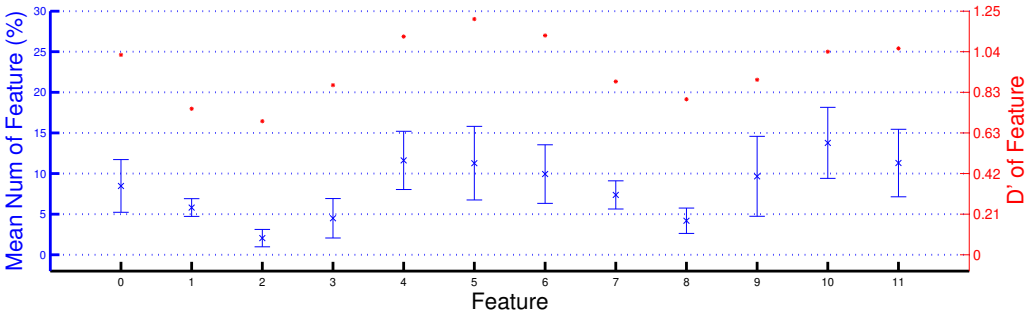
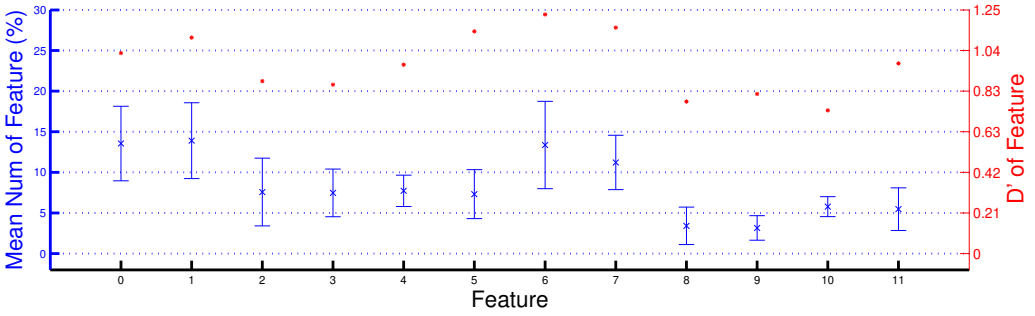


Figure 3.17: Mean occurrence and D' of the HOG patterns found in the **Inner Eyebrow. Features extracted from (a) left eye images of the FRGC dataset (b) right eye images of the FRGC dataset (c) left eye images of the FERET dataset (d) right eye images of the FERET dataset**

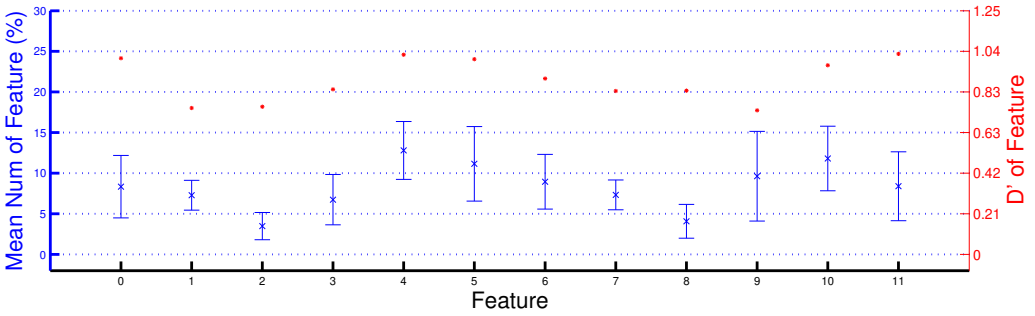
(a) Outer Eyebrow, FRGC, left eye



(b) Outer Eyebrow, FRGC, right eye



(c) Outer Eyebrow, FERET, left eye



(d) Outer Eyebrow, FERET, right eye

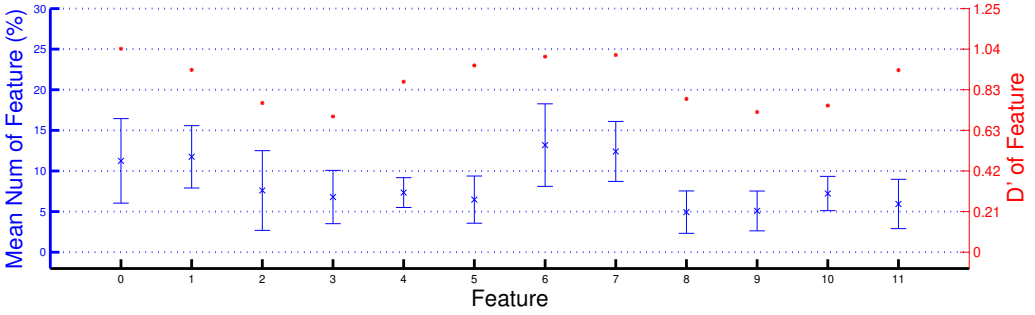
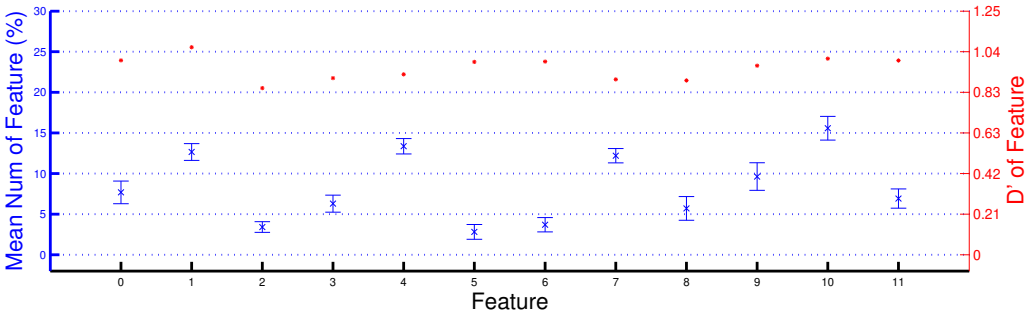
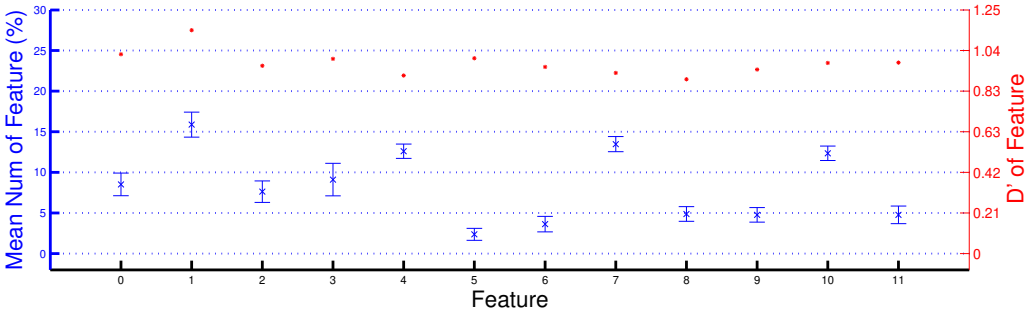


Figure 3.18: Mean occurrence and D' of the HOG patterns found in the **Outer Eyebrow. Features extracted from (a) left eye images of the FRGC dataset (b) right eye images of the FRGC dataset (c) left eye images of the FERET dataset (d) right eye images of the FERET dataset**

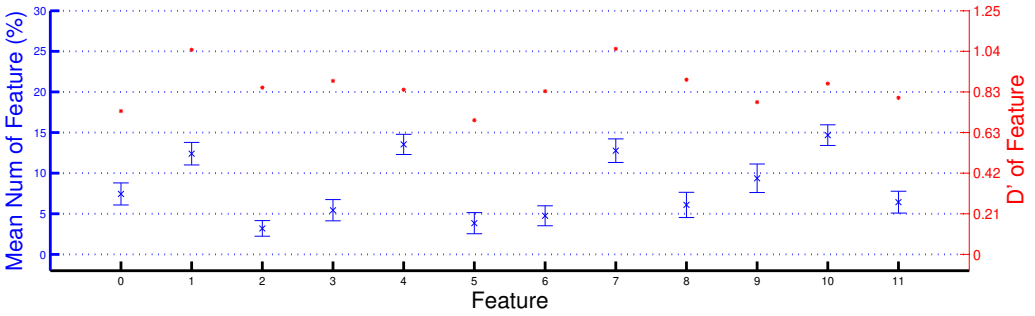
(a) Skin, FRGC, left eye



(b) Skin, FRGC, right eye



(c) Skin, FERET, left eye



(d) Skin, FERET, right eye

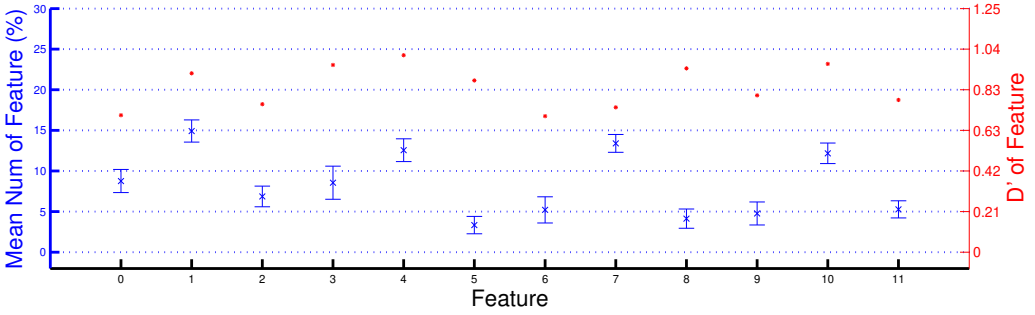


Figure 3.19: Mean occurrence and D' of the HOG patterns found in the **Skin. Features extracted from (a) left eye images of the FRGC dataset (b) right eye images of the FRGC dataset (c) left eye images of the FERET dataset (d) right eye images of the FERET dataset**

Sub-region	Feature Footprint
upper eyelid	4, 5, 6, 7
lower eyelid	0, 1, 10, 11
left tear duct	0, 1, 4, 7, 10
right tear duct	0, 1, 4, 7, 10
left outer corner	0, 1, 4, 10, 11
right outer corner	0, 1, 6, 7, 10, 11
left inner eyebrow	6, 7
right inner eyebrow	3, 4, 5, 6
left outer eyebrow	4, 5, 6, 9, 10
right outer eyebrow	0, 1, 6, 7
skin	1, 4, 7, 10

Table 3.9: HOG feature footprint for each sub-region

None of the features appear in all sub-regions at a significantly higher than average rate.

One noticeable difference between LBP and HOG features is the relationship between the upper and lower eyelids. In LBP, the feature footprints shared 11 features. In HOG, they share none. Note that the upper eyelid has features 4, 5, 6, and 7, while the lower eyelid has features 0, 1, 10, and 11. The upper eyelid feature footprint consists of the segments of the circle closest to π, while the lower eyelid consists of the segments closest to 0 and 2π. This seems to indicate that there is a direction to the gradient of the upper and lower eyelid and that the directions are orthogonal to each other.

The left and right tear duct and outer corner feature footprint, are similar in both LBP and HOG features. Many of the same features are shared between the left and right side with a smaller number of features that are different. The feature footprints of the left and right parts of both eyebrow sub-regions have very little in common.

Table 3.10 shows the mean D' of the HOG features extracted from each sub-region. Table 3.11 displays the Rank-1 recognition results from Section 3.6 in a format that mirrors Table 3.10 for easy comparison. Note that the average D' values of HOG features are much higher than their LBP counter-parts yet the Rank-1 recognition rates are much lower. This is caused by the difference in feature vector length. The shorter feature vector length of HOG leads to each individual feature having a higher measure of separability.

HOG features are shorter than LBP features, having a feature length of 12 instead of 59. This aspect makes generating an optimal usage of certain feature in the HOG feature vector through genetic algorithms mostly fruitless. The candidate solution of HOG features that produces the best results is the candidate solution that includes all of the features for every sub-region of the periocular region and for every dataset. The small number of features in a HOG feature vector lead to every feature contributing positively to the total

Sub-region	Left Eye FRGC	Right Eye FRGC	Left Eye FERET	Right Eye FERET
upper eyelid	0.6469	0.6784	0.4726	0.4927
lower eyelid	0.5009	0.4942	0.3916	0.3821
tear duct	0.6154	0.6367	0.5667	0.6090
outer corner	0.8286	0.7920	0.6336	0.6010
inner eyebrow	0.9575	0.9613	0.9239	0.9118
outer eyebrow	0.9559	0.9745	0.8922	0.8734
skin	0.9583	0.9773	0.8594	0.8475

Table 3.10: Mean D' of HOG features extracted from each sub-region

Sub-region	Left Eye FRGC	Right Eye FRGC	Left Eye FERET	Right Eye FERET
upper eyelid	53.6481	52.5751	27.9798	28.6869
lower eyelid	36.0515	42.7039	20.3030	18.4848
tear duct	47.6395	45.4936	32.7273	32.5253
outer corner	59.4421	58.7983	34.0404	35.5556
inner eyebrow	71.4592	69.9571	62.8283	62.3232
outer eyebrow	72.7468	74.4635	56.8687	54.8485
skin	90.1288	91.8455	71.3131	68.5859

Table 3.11: Rank 1 recognition rate of HOG features extracted from each sub-region

biometric system performance.

3.7.3 Local Phase Quantization

The 256 different elements in the LPQ feature vector correspond to the phase information of a discrete Fourier transform (DFT) in patch-sized neighborhoods of the biometric image.

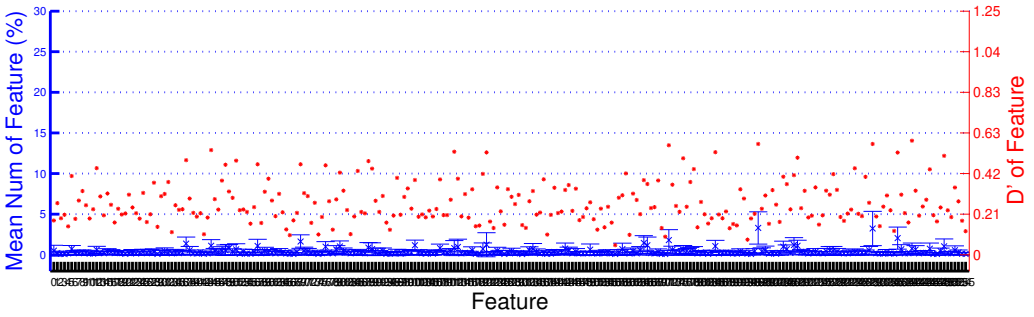
Figures 3.20 - 3.26 show the mean occurrence of each of the 256 LPQ features and their D' that were extracted from each sub-region of the periocular region. The large feature vector makes visualizing this feature extraction method difficult, yet the figures are still shown for completeness. The LPQ feature footprint shown in Table 3.12 will better illuminate any patterns.

There is much variability in the feature footprints of the LPQ features. One feature, 143, is found in all sub-regions at a significantly higher than average rate.

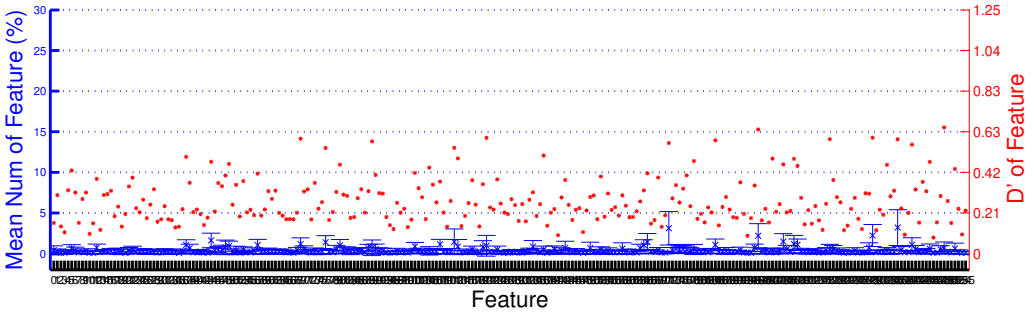
The upper and lower eyelids share 22 features from the LPQ feature footprint. This is out of 48 upper eyelid features and 62 lower eyelid features. LPQ features extracted from the upper and lower eyelids have much fewer features in common from their feature footprint than LBP features; however, the number of features in common is not as low as HOG's zero common features.

The left and right tear duct share 29 features from the LPQ feature footprint out of 74 features in

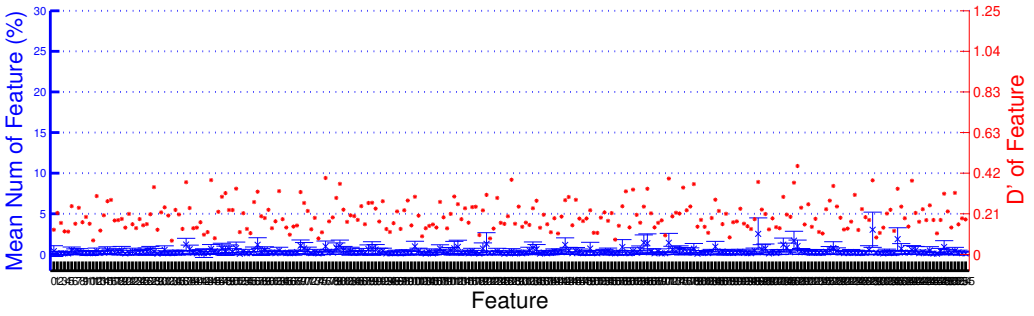
(a) Upper Eyelid, FRGC, left eye



(b) Upper Eyelid, FRGC, right eye



(c) Upper Eyelid, FERET, left eye



(d) Upper Eyelid, FERET, right eye

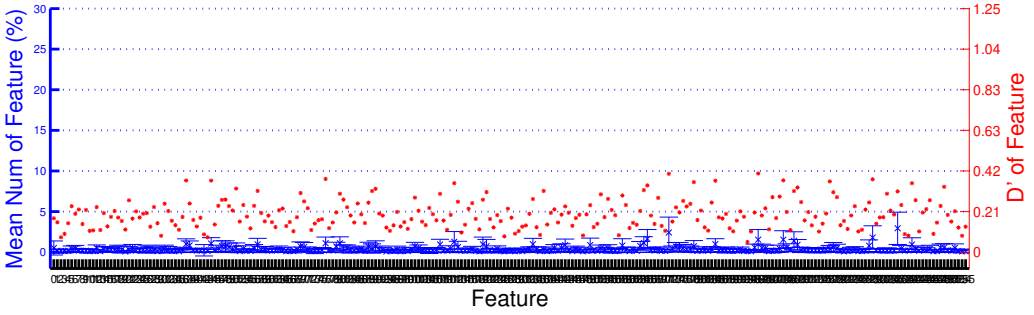
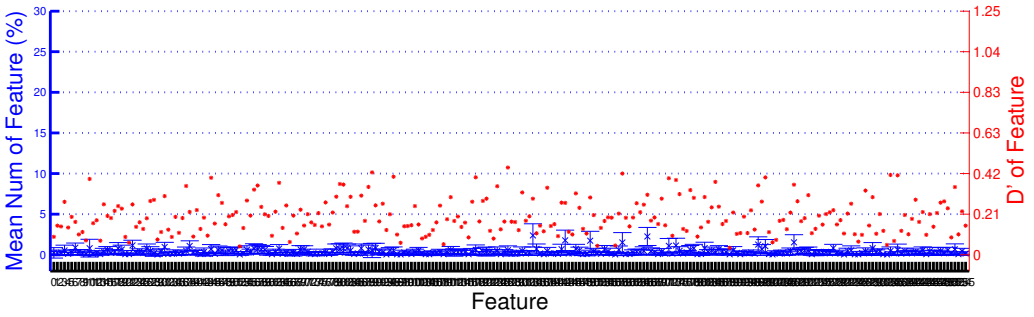
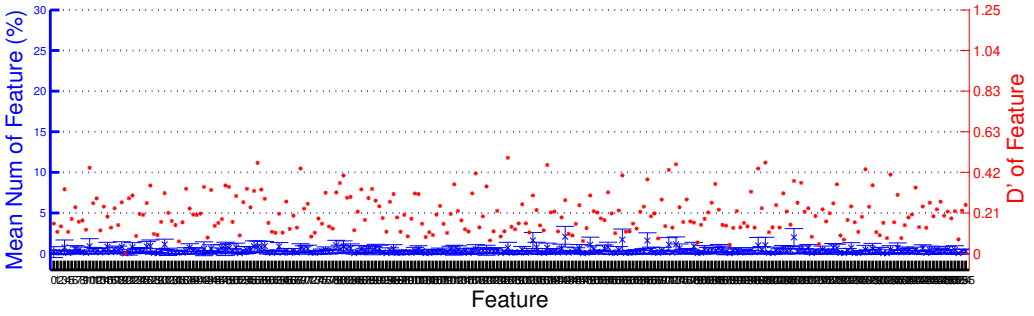


Figure 3.20: Mean occurrence and D' of the LPQ patterns found in the **Upper Eyelid. Features extracted from (a) left eye images of the FRGC dataset (b) right eye images of the FRGC dataset (c) left eye images of the FERET dataset (d) right eye images of the FERET dataset**

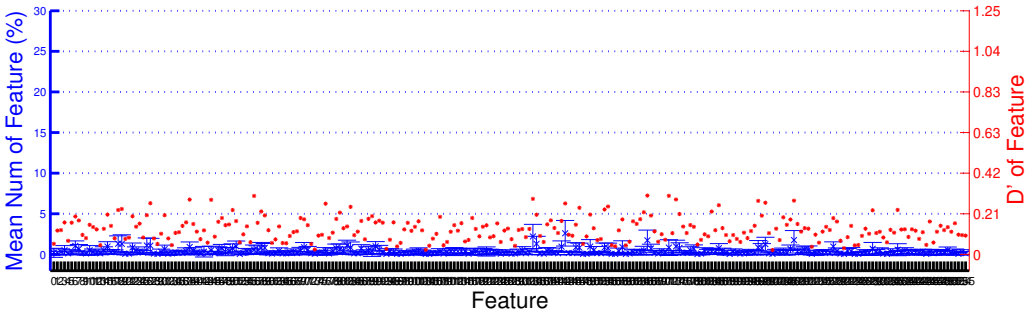
(a) Lower Eyelid, FRGC, left eye



(b) Lower Eyelid, FRGC, right eye



(c) Lower Eyelid, FERET, left eye



(d) Lower Eyelid, FERET, right eye

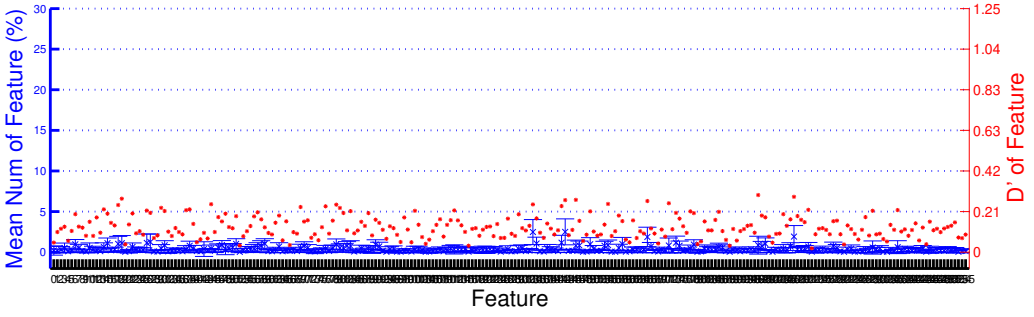
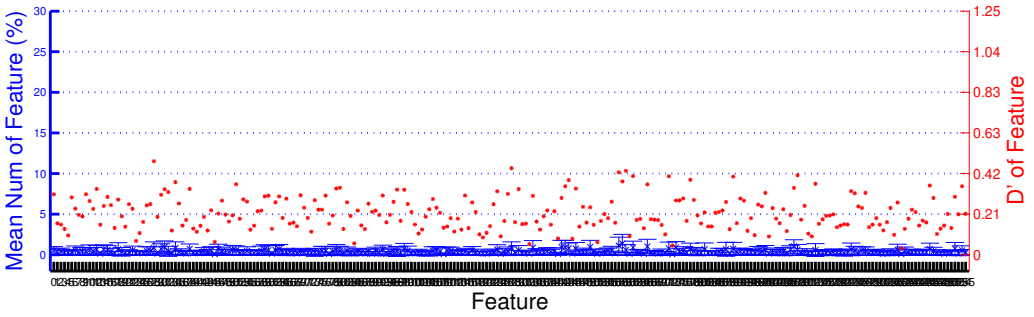
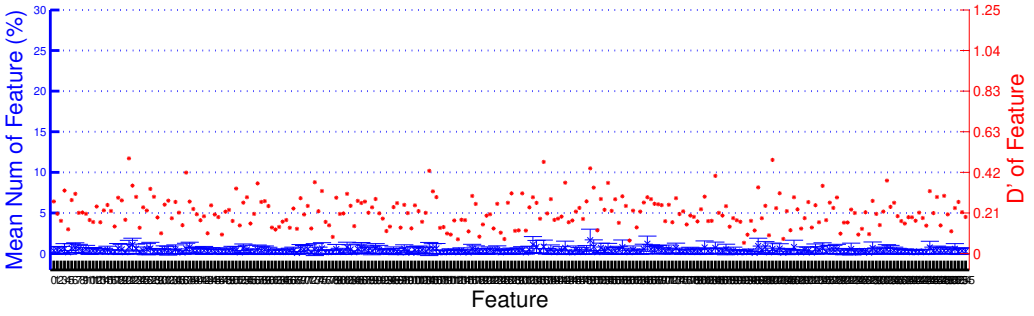


Figure 3.21: Mean occurrence and D' of the LPQ patterns found in the **Lower Eyelid. Features extracted from (a) left eye images of the FRGC dataset (b) right eye images of the FRGC dataset (c) left eye images of the FERET dataset (d) right eye images of the FERET dataset**

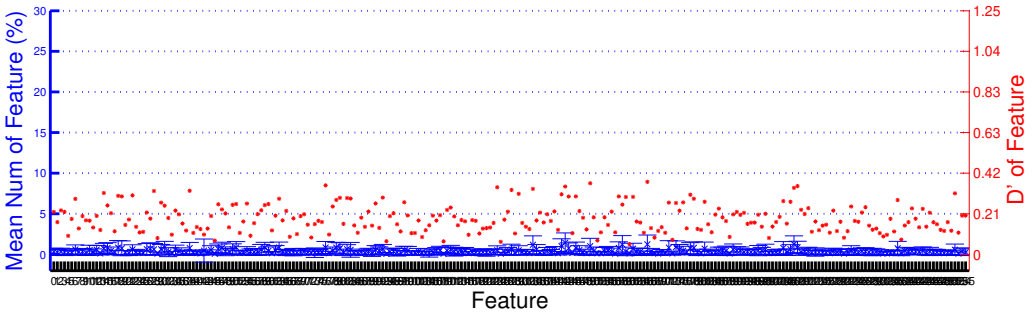
(a) Tear Duct, FRGC, left eye



(b) Tear Duct, FRGC, right eye



(c) Tear Duct, FERET, left eye



(d) Tear Duct, FERET, right eye

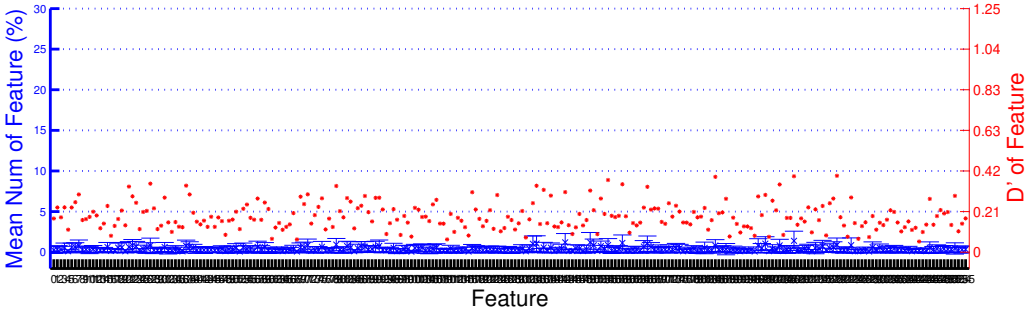
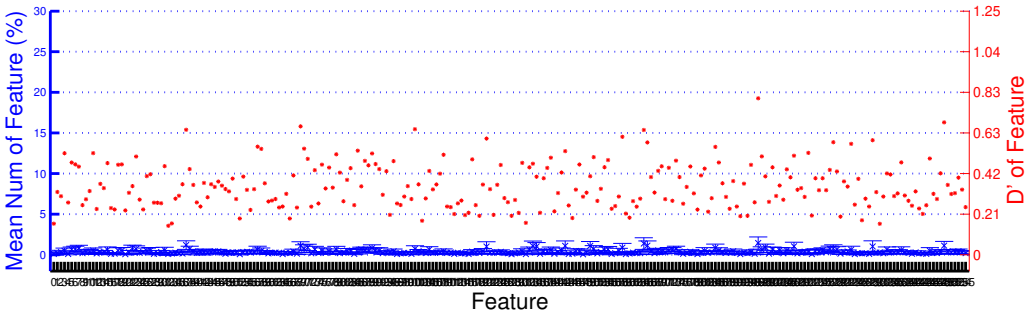
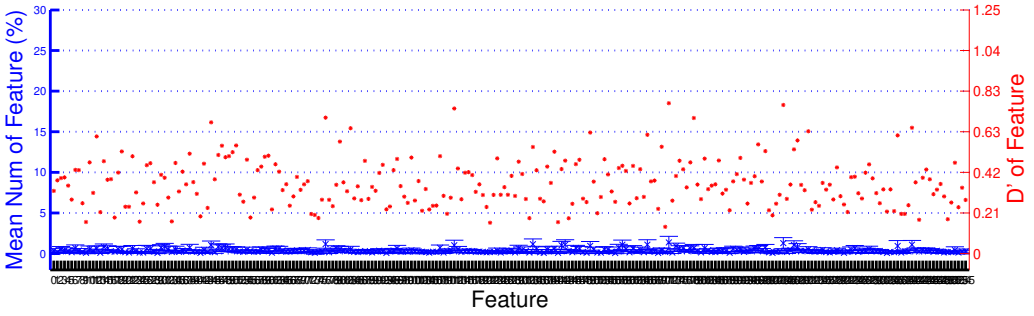


Figure 3.22: Mean occurrence and D' of the LPQ patterns found in the **Tear Duct. Features extracted from (a) left eye images of the FRGC dataset (b) right eye images of the FRGC dataset (c) left eye images of the FERET dataset (d) right eye images of the FERET dataset**

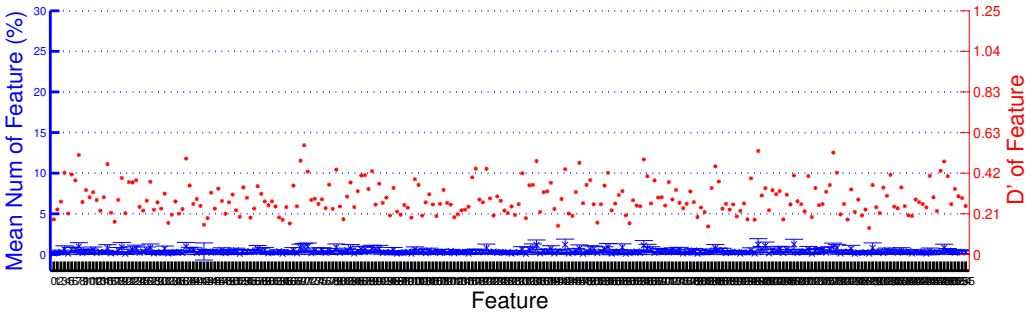
(a) Outer Corner, FRGC, left eye



(b) Outer Corner, FRGC, right eye



(c) Outer Corner, FERET, left eye



(d) Outer Corner, FERET, right eye

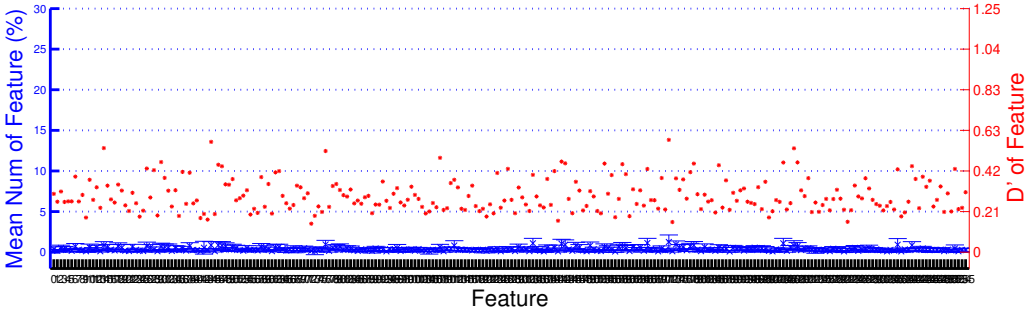
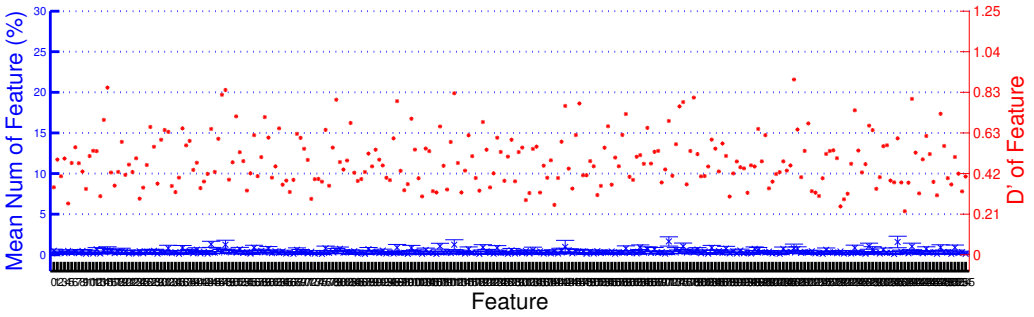
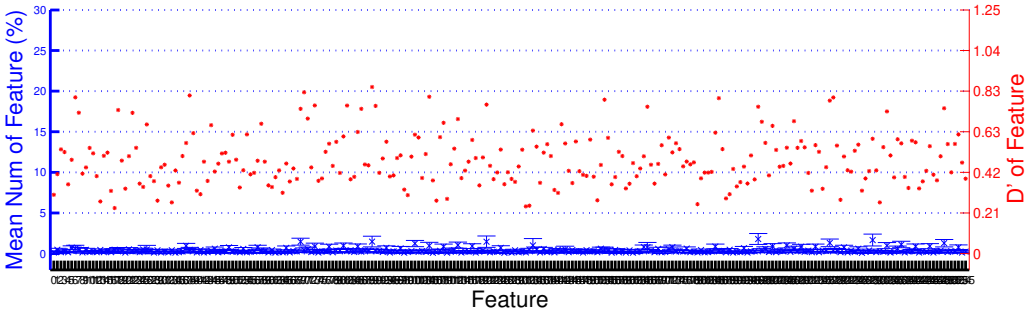


Figure 3.23: Mean occurrence and D' of the LPQ patterns found in the **Outer Corner. Features extracted from (a) left eye images of the FRGC dataset (b) right eye images of the FRGC dataset (c) left eye images of the FERET dataset (d) right eye images of the FERET dataset**

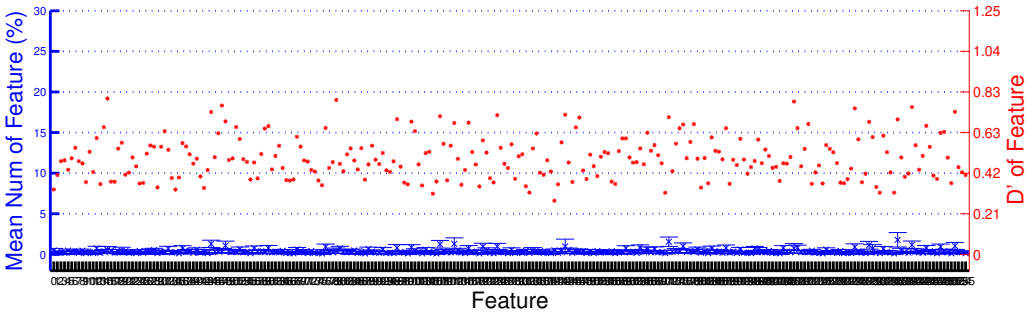
(a) Inner Eyebrow, FRGC, left eye



(b) Inner Eyebrow, FRGC, right eye



(c) Inner Eyebrow, FERET, left eye



(d) Inner Eyebrow, FERET, right eye

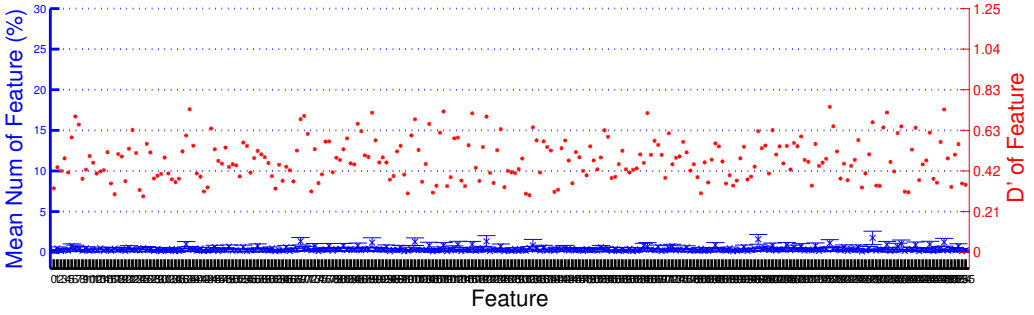
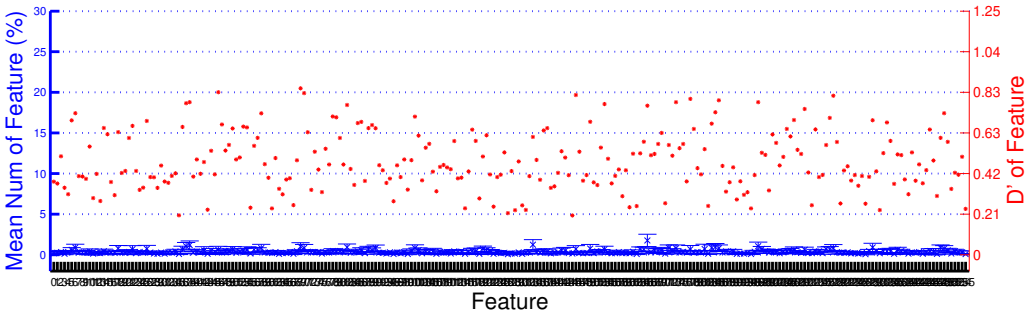
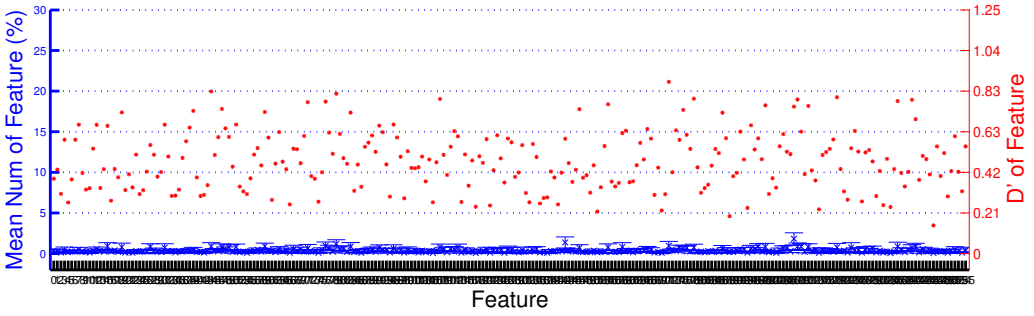


Figure 3.24: Mean occurrence and D' of the LPQ patterns found in the **Inner Eyebrow. Features extracted from (a) left eye images of the FRGC dataset (b) right eye images of the FRGC dataset (c) left eye images of the FERET dataset (d) right eye images of the FERET dataset**

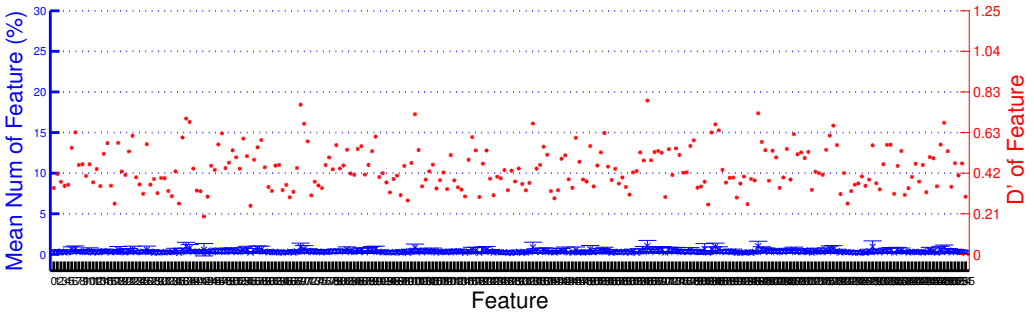
(a) Outer Eyebrow, FRGC, left eye



(b) Outer Eyebrow, FRGC, right eye



(c) Outer Eyebrow, FERET, left eye



(d) Outer Eyebrow, FERET, right eye

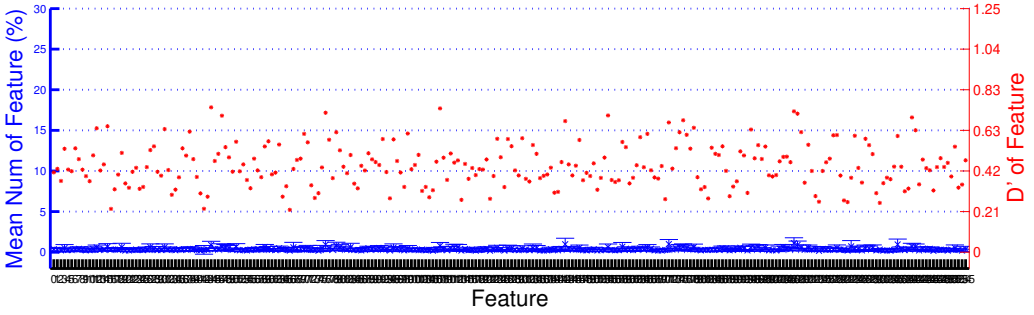
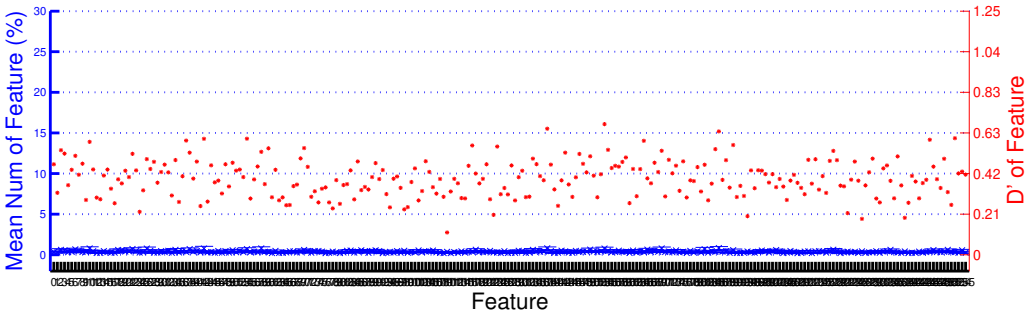
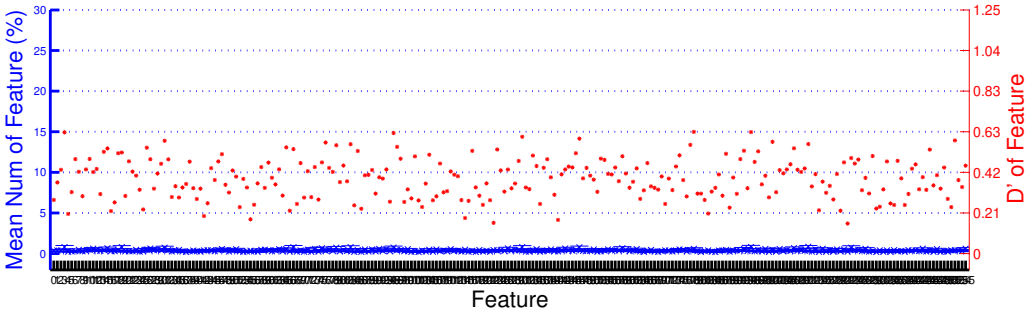


Figure 3.25: Mean occurrence and D' of the LPQ patterns found in the **Outer Eyebrow. Features extracted from (a) left eye images of the FRGC dataset (b) right eye images of the FRGC dataset (c) left eye images of the FERET dataset (d) right eye images of the FERET dataset**

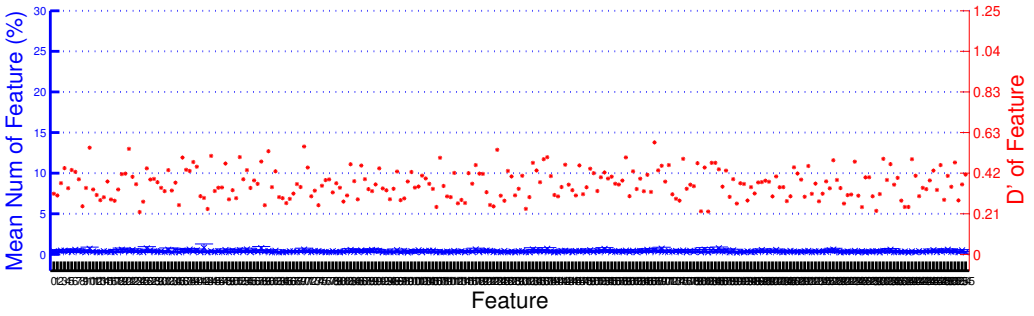
(a) Skin, FRGC, left eye



(b) Skin, FRGC, right eye



(c) Skin, FERET, left eye



(d) Skin, FERET, right eye

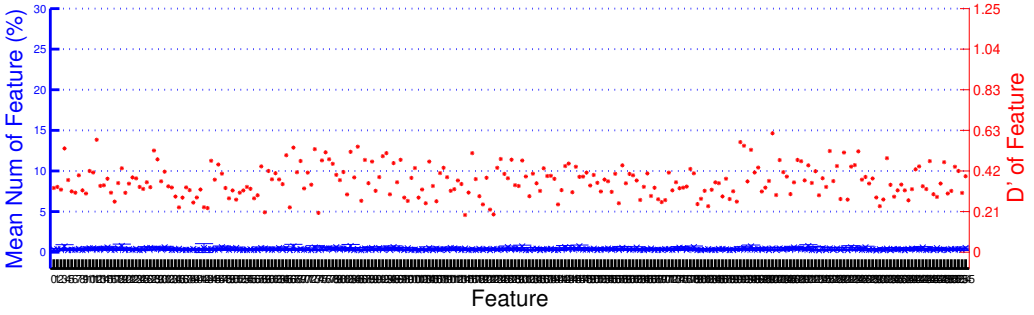


Figure 3.26: Mean occurrence and D' of the LPQ patterns found in the **Skin. Features extracted from (a) left eye images of the FRGC dataset (b) right eye images of the FRGC dataset (c) left eye images of the FERET dataset (d) right eye images of the FERET dataset**

Sub-region	Feature Footprint
upper eyelid	5, 12, 37, 38, 44, 47, 48, 49, 57, 69, 70, 76, 79, 80, 88, 89, 101, 108, 112, 113, 120, 121, 134, 143, 150, 159, 164, 165, 166, 172, 173, 179, 185, 196, 197, 204, 205, 207, 208, 217, 218, 228, 229, 236, 237, 240, 245, 249
lower eyelid	3, 6, 10, 15, 18, 19, 22, 26, 27, 31, 38, 44, 47, 49, 51, 56, 57, 58, 59, 63, 69, 70, 79, 80, 81, 82, 83, 86, 88, 90, 131, 134, 135, 138, 142, 143, 147, 150, 151, 154, 155, 158, 159, 166, 167, 172, 174, 175, 179, 182, 186, 197, 198, 199, 206, 207, 211, 218, 229, 231, 236, 238
left tear duct	6, 10, 12, 14, 15, 18, 22, 26, 27, 28, 30, 31, 34, 38, 44, 46, 47, 50, 51, 54, 58, 59, 60, 62, 63, 76, 78, 79, 80, 82, 90, 92, 98, 99, 107, 108, 110, 111, 124, 128, 130, 134, 140, 142, 143, 144, 146, 150, 155, 158, 159, 162, 166, 172, 174, 176, 178, 179, 182, 187, 188, 190, 192, 197, 198, 199, 206, 207, 208, 223, 226, 236, 238, 252
right tear duct	3, 5, 6, 7, 15, 18, 19, 21, 22, 23, 26, 27, 31, 37, 38, 53, 57, 59, 69, 70, 71, 75, 79, 82, 83, 85, 86, 87, 90, 91, 98, 101, 102, 103, 106, 107, 117, 133, 134, 135, 137, 139, 143, 146, 150, 151, 153, 155, 159, 165, 166, 167, 169, 174, 182, 185, 197, 198, 199, 203, 207, 210, 213, 214, 215, 217, 218, 219, 223, 229, 231, 245, 247, 252
left outer corner	3, 5, 6, 7, 11, 15, 18, 19, 22, 23, 26, 27, 31, 37, 38, 39, 57, 58, 67, 69, 70, 71, 73, 75, 77, 79, 82, 83, 86, 87, 89, 90, 91, 95, 101, 102, 105, 121, 131, 133, 134, 135, 139, 143, 146, 147, 150, 151, 154, 155, 159, 165, 166, 167, 169, 173, 185, 186, 197, 198, 199, 201, 203, 205, 207, 215, 217, 218, 219, 223, 229, 233, 237, 245, 249, 250
right outer corner	2, 6, 10, 12, 14, 15, 18, 19, 22, 26, 27, 30, 31, 32, 34, 38, 42, 44, 46, 47, 48, 50, 51, 54, 58, 59, 62, 63, 76, 79, 80, 83, 96, 108, 112, 130, 131, 134, 135, 138, 140, 142, 143, 146, 147, 150, 154, 155, 158, 159, 160, 162, 164, 166, 172, 174, 175, 176, 178, 179, 182, 186, 190, 192, 196, 199, 204, 206, 207, 208, 211, 224, 228, 236, 240, 243, 252
left inner eyebrow	0, 4, 8, 12, 15, 16, 19, 20, 24, 28, 32, 36, 38, 40, 44, 46, 48, 52, 56, 60, 64, 68, 76, 79, 80, 84, 88, 92, 96, 100, 104, 108, 112, 116, 120, 121, 124, 128, 132, 134, 136, 140, 142, 143, 144, 148, 152, 156, 160, 164, 166, 168, 172, 174, 176, 178, 180, 184, 188, 192, 196, 197, 200, 204, 207, 208, 212, 216, 220, 224, 228, 229, 232, 236, 240, 244, 248, 252
right inner eyebrow	1, 5, 6, 7, 9, 13, 17, 21, 25, 29, 33, 37, 38, 41, 45, 49, 53, 57, 61, 65, 69, 70, 71, 73, 77, 81, 85, 89, 93, 97, 101, 105, 109, 112, 113, 117, 121, 125, 129, 133, 134, 135, 137, 141, 143, 145, 149, 153, 157, 165, 166, 169, 172, 173, 177, 181, 185, 189, 193, 197, 199, 201, 205, 207, 209, 213, 217, 219, 221, 225, 229, 233, 236, 237, 241, 245, 249, 253
left outer eyebrow	5, 6, 18, 22, 26, 34, 36, 37, 38, 42, 44, 46, 49, 50, 53, 54, 56, 57, 58, 69, 70, 78, 82, 85, 86, 88, 89, 90, 100, 101, 102, 112, 117, 118, 120, 121, 122, 134, 137, 142, 143, 146, 150, 153, 154, 164, 165, 166, 168, 169, 170, 172, 173, 174, 178, 182, 184, 185, 186, 197, 198, 199, 205, 206, 207, 210, 216, 217, 218, 228, 229, 232, 233, 234, 236, 237, 245, 248, 249, 250, 252
right outer eyebrow	3, 7, 12, 15, 19, 27, 31, 39, 44, 47, 48, 49, 51, 59, 60, 63, 67, 69, 71, 75, 76, 77, 79, 80, 81, 83, 88, 91, 92, 95, 99, 108, 109, 111, 112, 113, 115, 121, 124, 127, 128, 131, 134, 135, 143, 144, 147, 155, 159, 164, 166, 167, 172, 174, 175, 176, 177, 179, 187, 192, 193, 195, 196, 197, 199, 203, 204, 205, 207, 208, 209, 211, 215, 219, 223, 224, 225, 227, 228, 229, 236, 237, 239, 240, 241, 243, 245, 252
skin	3, 7, 10, 14, 18, 19, 26, 27, 32, 46, 48, 50, 58, 71, 73, 83, 89, 91, 96, 105, 131, 135, 142, 143, 146, 147, 155, 178, 201, 211, 217, 219

Table 3.12: LPQ feature footprint for each sub-region

Sub-region	Left Eye FRGC	Right Eye FRGC	Left Eye FERET	Right Eye FERET
upper eyelid	0.2662	0.2707	0.2003	0.1954
lower eyelid	0.1975	0.2079	0.1309	0.1301
tear duct	0.2148	0.2140	0.1850	0.1853
outer corner	0.3543	0.3624	0.2881	0.2909
inner eyebrow	0.4831	0.4813	0.4954	0.4760
outer eyebrow	0.4796	0.4765	0.4403	0.4463
skin	0.3919	0.3852	0.3627	0.3648

Table 3.13: Mean D' of LPQ features extracted from each sub-region

Sub-region	Left Eye FRGC	Right Eye FRGC	Left Eye FERET	Right Eye FERET
upper eyelid	69.0987	68.8841	44.1414	46.2626
lower eyelid	57.2961	59.4421	30.8081	32.0202
tear duct	59.8712	58.7983	48.4848	49.0909
outer corner	74.4635	75.1073	45.7576	47.7778
inner eyebrow	87.3391	84.9785	78.4848	80.0000
outer eyebrow	87.3391	88.6266	73.2323	73.0303
skin	93.7768	95.7082	75.2525	73.4343

Table 3.14: Rank 1 recognition rate of LPQ features extracted from each sub-region

each feature footprint. This is a much lower rate of feature similarity between left and right sides of a sub-region than either LBP or HOG produced. The left and right outer corner showed the same pattern by sharing 26 features from the LPQ feature footprint out of the 76 features in the left outer corner and 77 features in the right outer corner.

There is a lower rate of similarity between left and right inner and outer eyebrow feature footprints. The left and right inner eyebrow feature footprints share 11 of the same features. The left and right outer eyebrow feature footprints share 22 of their features. This low level of similarity corresponds with the trend seen when comparing LBP and HOG feature footprints.

Table 3.13 shows the mean D' of the LPQ features extracted from each sub-region. Table 3.14 displays the Rank-1 recognition results from Section 3.6 in a format that mirrors Table 3.13 for easy comparison. Note that the average D' values of LPQ features are much lower than either LBP or HOG, yet the Rank-1 recognition rates are much higher. As discussed with the other two feature extraction methods, the longer feature vector allows for higher recognition rates with lower single feature separability.

A GA approach was used to produce an optimal selection of features from the LPQ feature vectors of each sub-region of the periocular region as well. Table 3.15 shows the Rank-1 recognition rate of features taken from the upper eyelid of each dataset as well as the Rank-1 recognition rate of the best performing

	All Features	Candidate Solution
Left Eye FRGC	69.0987	73.3906
Right Eye FRGC	68.8841	72.5322
Left Eye FERET	44.1414	45.5556
Right Eye FERET	46.2626	47.8788

Table 3.15: Rank 1 recognition rate of LPQ features extracted from the upper eyelid

candidate solution. As with the LBP experiments, this GA approach produced better results than using all features for each sub-region of the periocular region and each dataset.

3.8 Conclusion

The goal of this chapter was to address the following questions: Are there sub-regions within the periocular region that provide more discriminative information than others? Are these sub-regions similarly discriminative using different feature representation methods? Could the knowledge gained here be used to adapt the local approaches to use the physical structure of the periocular region? The experiments presented in Section 3.4 suggest that the most discriminative sub-regions of the periocular region are near the eyebrows and eye corners, while the irises provided comparably little discriminative information. Using this knowledge, a novel block placement method for LABFs is proposed in Section 3.5. This block placement method is based on the location of the structural sub-regions of the periocular region. When using only features extracted from each individual sub-region, the results of biometric experiments presented in this chapter suggest that LBP features are most discriminative in the upper eyelid, lower eyelid, tear duct, and outer corner, while LPQ features are most discriminative in the inner eyebrow, outer eyebrow, and skin.

The knowledge gained from these experiments will contribute to a new method for periocular feature extraction that leverages the physical structure of the periocular region. We hypothesize that taking these structural elements into consideration will result in an increase in total system accuracy in a baseline biometric experiment. This method could also be applied to more difficult problems, such as recognition over a large time-lapse.

Chapter 4

Features from Multiple Scales

4.1 Introduction

When extracting LABFs from periocular biometric data, the current body of research utilizes these features from a single scale. The scale, in this context, is typically determined by the size of the window over which the LABF performs its primary computation. In LBP, for instance, the binary pattern for a pixel is computed over a local neighborhood of pixels. Previous research that uses LABFs with periocular region data extracts features from a single scale [5, 17, 20, 27, 29, 30, 31, 54, 53, 55] or simply does not mention this aspect of their methodology [21, 28, 37, 38, 44, 45].

LABFs are intended to be local and were designed, in part, to give a counter perspective to global, holistic approaches. To this end, LABFs typically use the smallest possible scale from which to execute their algorithms. The intent of this use of small scales is to quantify the relationship between the closest of the neighboring pixels with the expectation that these close pixel relationships will provide more discriminative information than an overall view in specific situations.

While LABFs have been successful in specific situations, there is potential for further useful and discriminative information to be extracted from using multiple scales. When using a single (typically small) scale, the algorithm can only describe the relationship between pixels that are close together. Quantifying the relationship between pixels in multiple scales could provide additional discriminative information. This information could lead to an increase in performance that could outweigh the offsetting increase in necessary computations.

The goals of this Chapter are as follows:

- Modify the LABF algorithms to extract features from multiple scales.
- Determine if there is a scale or set of scales that provides the best performance when used with each of the LABFs.
- Determine if these scales show different performance trends across each of the sub-regions of the periocular region.

This work is a much needed exploration of an unexplored part of LABF extraction algorithms. We hypothesize that these modified LABF methods will provide an increase in performance over methods used so far in a typical biometric experiment.

4.2 Data

The experimentation in this section uses a subset of the FRGC Experiment 1 data set. Two images (one probe, one gallery) were used from 466 subjects in the FRGC data set, for a total of 932 images. The two selected images were chosen for each subject from the first available recording session in order to minimize the time lapse between images, as time lapse was not the area of focus in this section. Only images with a neutral facial expression were selected, as the impact of facial expression was not the focus in this section. This subset is the same subset that was described in Section 3.2. The set of FERET images detailed in Section 2.2 is also used.

4.3 Method

The experimental method of this chapter follows the design presented in Section 3.6. Features are extracted from the seven sub-regions of the periocular region detailed in Section 3.5.

To test the performance of LABFs across multiple scales, the algorithms must be modified to accept a new scale parameter. The means to accomplish this modification for each of the feature extraction methods is detailed below.

4.3.1 LBP

The LBP algorithm is typically visualized as a variable number of points along the circumference of a circle placed around a center point. The pattern of the difference in grayscale intensity between each

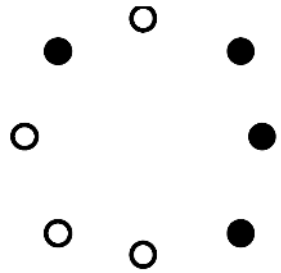


Figure 4.1: Representation of the local neighborhood on which LBP is calculated

point on the circle and the center point is used to calculate the LBP value at a pixel. In Figure 4.1, the white circles have pixel intensities less than the center pixel while the black pixels have pixel intensities greater than the center pixel. This circular pattern is unwrapped and translated into a string of ones and zeros, such as 01110001, if the top pixel is the first in the string and we move clockwise.

The ability to vary the radius of the circle on which the pixels reside is the predicate for allowing LBP to be computed from multiple scales. However, simply computing the LBP operator using pixels that are far apart does not correctly account for a change in scale. LBP is meant to be a measurement of local features and using pixels that are farther apart would not result in measuring a local pattern. Therefore, the value of each pixel on the circle is represented by the average intensity of all the pixels contained in a window of variable size around the focus pixel. The pseudo-code for the proposed approach to compute the LBP for a given pixel is listed in Figure 4.2 with red text to signify differences from the original implementations. The R and W variables are set equal to each other, so as the scale of the algorithm increases, both R and W increase equally together.

4.3.2 HOG

The HOG algorithm is a function of the gradient magnitude and gradient angle of the input data. The HOG descriptor is a histogram of the values of the gradient angle of an image convolved with a 3×3 Prewitt filter and scaled by the value of the gradient magnitude. This is a simple algorithm that has the ability to quantify small lines within an image. These lines are very fine details that only represent one small scale of the images.

Like LBP, the HOG algorithm can be expanded to quantify multiple scales with a few simple steps. The Prewitt filter that is convolved with the input image can be expanded to different sizes. A 5×5 Prewitt filter, for example, would look like Figure 4.3. Such a filter and larger ones serve to highlight longer lines that

```
N = # of neighbor pixels on the circle around center pixel
R = radius of the circle around center pixel
W = size of window around a pixel
L = location of the neighboring pixels
C = location of center pixel
LBP = value of the LBP operator

A = 2π / N
for i = 0 to N-1
   L[i] = [-R*sin(i*A) R*cos(i*A)]
end

LBP = 0
for i = 0 to N-1
   X = average intensity value of a window of size W around C
   Y = average intensity value of a window of size W around L[i]
   IF X > Y THEN D = 1 ELSE D = 0
   LBP = LBP + D*2^i
end
```

Figure 4.2: Pseudo-code for calculating the LBP operator at variable scales

1	1	1	1	1
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
-1	-1	-1	-1	-1

Figure 4.3: A 5×5 Prewitt filter.

are farther apart. Similar to the multiscale LBP algorithm, the modified HOG algorithm needs to quantify lines from pixels that represent that average value of nearby pixels. To accomplish this quantification, the image is first convolved with an averaging filter. An example of a 5×5 averaging filter can be seen in Figure 4.4. The pseudo-code for the proposed approach to compute HOG for an image is listed in Figure 4.5 with red text to signify differences from the original implementation.

4.3.3 LPQ

The LPQ descriptor quantifies phase information computed in a sliding window over the image. In this algorithm the four low-frequency coefficients of the Fourier transform on each window are decorrelated and uniformly quantized in an eight-dimensional space. A histogram of this data is used as the feature vector. This phase information is reported to be useful for classifying blurred texture images [36].

1/25	1/25	1/25	1/25	1/25
1/25	1/25	1/25	1/25	1/25
1/25	1/25	1/25	1/25	1/25
1/25	1/25	1/25	1/25	1/25
1/25	1/25	1/25	1/25	1/25

Figure 4.4: A 5×5 averaging filter.

```
G = gradient
I = input image
N = # of bins
W = size of filter
AVG = Averaging filter of size W
PRE = Prewitt filter of size W

I = I**AVG
G_x = I * PRE
G_y = I * PRE'
G_mag = sqrt(G_x^2 + G_y^2)
G_ang = ATAN2(G_y,G_x)

for i = 0 to N-1
   find all indexes (x) where floor(G_ang(x) / (2π/N)) = i
 HOG[x] = ∑ G_mag[x]
end
```

Figure 4.5: Pseudo-code for calculating the HOG operator at variable scales

In the same manner as the other LABFs, LPQ can be expanded to be multiscale. The sliding window from which the phase information is computed can be adjusted for multiple sizes. As a requirement for this adjustment, some preprocessing steps are applied to the input image and changes to other parameters are made. The pseudo-code for the proposed approach to compute LPQ for an image is listed in Figure 4.6, with red text to signify differences from the original implementation.

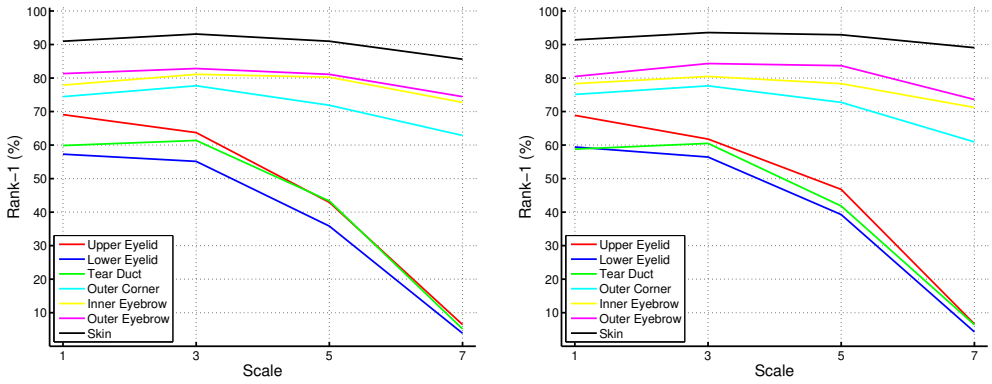
4.4 Results

Biometric experiments were conducted to quantify the performance of LABFs extracted from sub-regions of the periocular region at multiple scales. As the intention of the proposed approach is to extract features from multiple scales, it must first be determined if there are particular scales that are more discriminative than others with certain LABFs. From this, an appropriate and advantageous weighting can be selected for each scale. In the modified LBP algorithm (see Figure 4.2), the scale is represented by the R and W variables. In the modified HOG algorithm (see Figure 4.5), the scale is represented by the W variable. These scales can be any odd integer and their size is limited by the size of the input data. The Rank-1 recognition rates of 504 different experiments are given in Figures 4.7 - 4.12. Each experiment is performed in the same manner as other experiments in this dissertation with the changing variables displayed in the tables. Note that there are no values for LBP at a scale of 9 or 11. This absence of values is because the size of the images used in these experiments limits LBP to a scale of 7.

The best LBP performance in these experiments is seen in either Scale 1 or Scale 3. In non-top performing Scale 1 and Scale 3, the results show a marginal decrease in performance. There is a much more substantial drop in performance for Scales 5 and 7. The fact that LBP performs best at small scale values in these experiments is expected given that the original algorithm was developed to identify textures such as spots, line ends, and other small textures [33]. In the LBP experiments there is a strong relationship between the best performing scale and the sub-region of the periocular region across all four types of data: FRGC left eye, FRGC right eye, FERET left eye, and FERET right eye. The upper and lower eyelids perform best at Scale 1, while the remaining perform best at Scale 3. The one exception is the tear duct where the FRGC experiments show Scale 3 as having the best performance, while the FERET experiments indicate the highest Rank-1 results at Scale 1. A possible reason for the upper and lower eyelids performing well in Scale 1 is that the features extracted from these regions are the smallest. The lines that make up the upper and lower eyelid are distinct at the lowest scale. As you increase the radius of the circle that the samples are taken from and

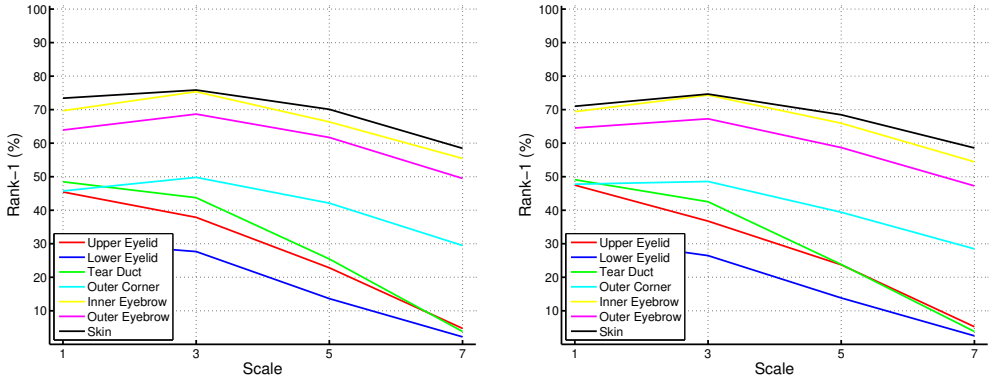
```
I = input image
W = size of window
R = radius of the window
ρ = 0.9
α = 1/W

x = array of length W with values -R to R
w0 = array of 1s of length W
w1 = e^0 + (−2∗π∗x∗α)i
w2 = complex conjugate of w1

// Compute the freqency response at 4 points
frequency response = 3D matrix of size (W,W,8)
filter response = (I ** w0') ** w1
frequency response(1) = real(filter response)
frequency response(2) = imag(filter response)
filter response = (I ** w1') ** w0
frequency response(3) = real(filter response)
frequency response(4) = imag(filter response)
filter response = (I ** w1') ** w1
frequency response(5) = real(filter response)
frequency response(6) = imag(filter response)
filter response = (I ** w1') ** w2
frequency response(7) = real(filter response)
frequency response(8) = imag(filter response)

// Decorrelation
C = matrix of covariances between each pixel position
q1 = w0'*w1
q2 = w1'*w0
q3 = w1'*w1
q4 = w1'*w2
M = [real(q1); imag(q1); real(q2); imag(q2);
    real(q3); imag(q3); real(q4); imag(q4);]
D = M*C*M'
A = [1.000007 1.000006 1.000005 1.000004 1.000003 1.000002 1.000001 1]
V = singular value decomposition(A*D*A)
freqResp = (V' * freqResp')'

for i = 1 to 8
    LPQ[i] = LPQ[i] + if(freqResp(i) > 0) (2^(i-1));
end

LPQ = 256 bin histogram of LPQ;
```

Figure 4.6: Pseudo-code for calculating the LPQ operator at variable scales



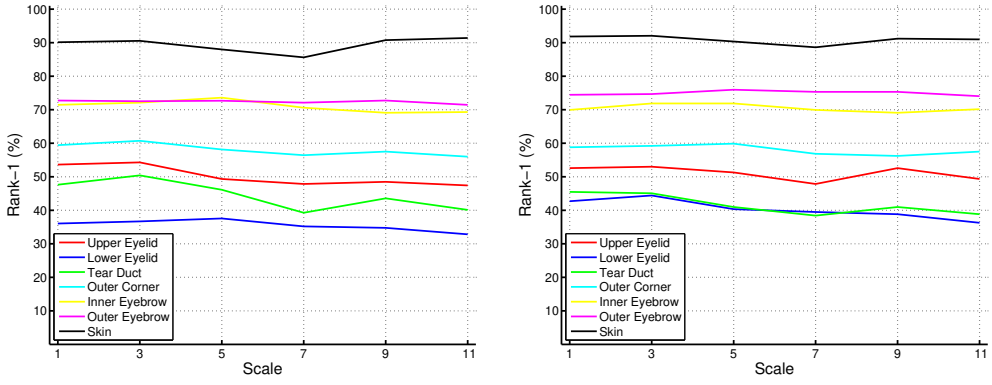
(a) (b)

Figure 4.7: Rank-1 results of LBP experiments using (a) left eye and (b) right eye FRGC images.



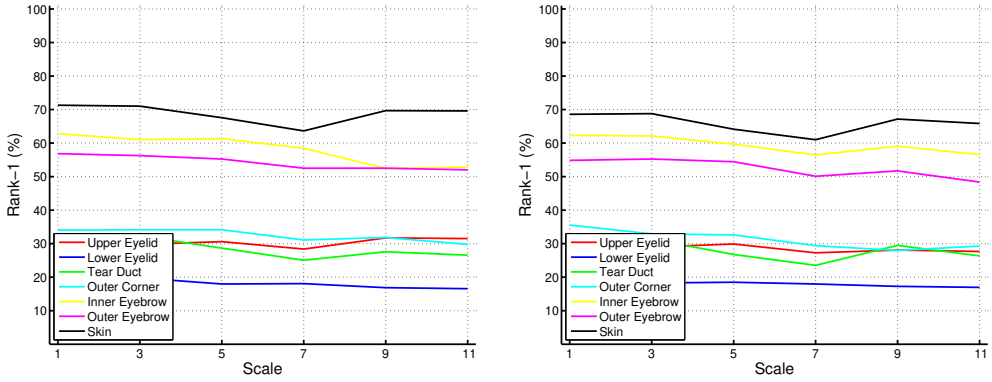
(a) (b)

Figure 4.8: Rank-1 results of LBP experiments using (a) left eye and (b) right eye FERET images.



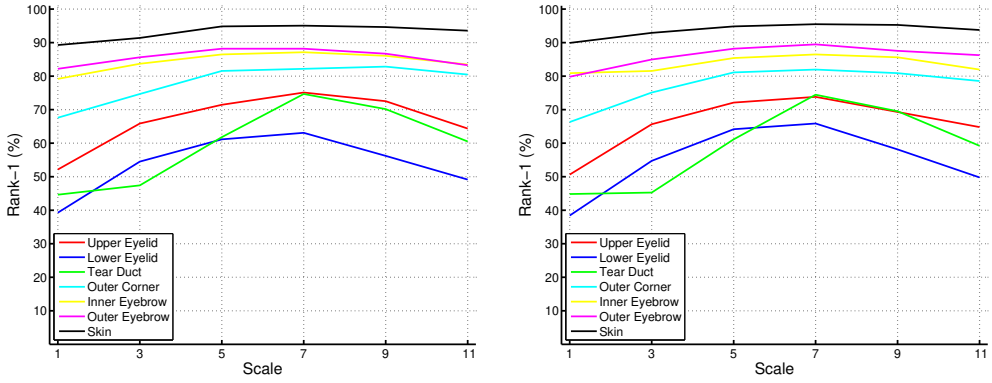
(a) (b)

Figure 4.9: Rank-1 results of HOG experiments using (a) left eye and (b) right eye FRGC images.



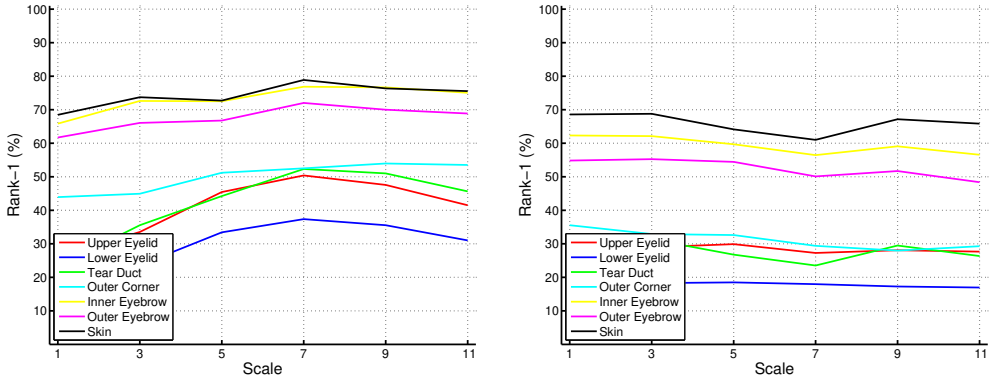
(a) (b)

Figure 4.10: Rank-1 results of HOG experiments using (a) left eye and (b) right eye FERET images.



(a) (b)

Figure 4.11: Rank-1 results of LPQ experiments using (a) left eye and (b) right eye FRGC images.



(a) (b)

Figure 4.12: Rank-1 results of LPQ experiments using (a) left eye and (b) right eye FERET images.

increase the averaging of the pixels in the window, the distinctiveness of the lines decrease. They come to a point at Scale 7 where they report single digit Rank-1 recognition rates. The other sub-regions maintain at least some measure of successful performance at that scale. The shapes, spots, and patterns quantified from the eye corners and the eyebrow are larger and can be seen in the performance increase from Scale 1 to Scale 3.

The performance of the modified HOG algorithm is interesting in that it does not appear to show a strong pattern between the different sources of data. A best performing experiment can be found in nearly every scale from Scale 1 to Scale 9 for at least one of the sub-regions of the periocular region. There are some sub-regions where the best performing scale is both Scale 1 and Scale 11, such as the skin from the FRGC images where the best performing scale is Scale 1 for left eye images and Scale 11 for left eye images. Unlike LBP, there is no steep drop in performance as you move away from the best performing scale. All scales perform similarly in HOG experiments. This observation can be explained by the properties of the modified HOG algorithm. The HOG algorithm evaluates gradients of the biometric image. Evaluating the image at different scales does not greatly influence the gradient.

The best performing LPQ scales appear more uniform than HOG. They are similar to the LBP results but have their best performance at higher scales. The performance metrics of experiments using LPQ do not decrease as quickly as the scales go further from the best performing scale. The best performing scale is Scale 7 for almost all sub-regions in both FRGC and FERET experiments. The high performance at high scales is not surprising in these experiments since LPQ has been compared to LBP for the purpose of recognizing blurred faces and performed extremely well in comparison, even in the base case [3].

Knowing that features taken from different scales produce different performance results, an investigation into whether a fusion of different scales could result in an increase in performance is conducted. To determine the optimal weighting for each scale, a weighted score level fusion is performed across all scales, where a score level fusion is defined as a weighted sum function over the computed distance matrices from the experiments for each scale. The weights applied to each distance matrix are a variable fraction where the sum of all weights equals 1. The optimum weight for each scale can be found by iterating over all possible scales. The optimum weights found with this data can be used as training for future experiments to test the final proposed algorithm with the current methods.

Table 4.1 shows the Rank-1 of experiments using a weighted fusion of multiple scales where the weights were determined by the method described above for experiments using images from the FRGC dataset. Table 4.2 lists the same results for FERET images. These results are comparable to the results from

	Left			Right		
Sub-region	LBP	HOG	LPQ	LBP	HOG	LPQ
upper eyelid	77.2532	59.6567	72.7468	76.6094	59.0129	71.4592
lower eyelid	71.0300	45.2790	62.4464	69.5279	48.4979	64.8069
tear duct	76.6094	57.7253	67.3820	77.6824	53.4335	69.3133
outer corner	85.8369	64.5923	81.1159	84.9785	63.7339	81.1159
inner eyebrow	83.4764	74.8927	87.9828	83.0472	74.2489	88.8412
outer eyebrow	84.7639	76.6094	89.4850	87.1245	78.7554	90.1288
skin	93.7768	91.6309	95.2790	94.2060	92.7039	95.9227

Table 4.1: Rank-1 results of fusion of scales experiments using FRGC images.

	Left			Right		
Sub-region	LBP	HOG	LPQ	LBP	HOG	LPQ
upper eyelid	55.4545	41.8182	47.3737	57.0707	38.0808	49.2929
lower eyelid	43.4343	27.1717	36.2626	42.1212	26.2626	34.9495
tear duct	63.2323	43.8384	55.3535	61.6162	44.5455	55.5556
outer corner	63.4343	42.9293	57.0707	60.2020	42.4242	54.8485
inner eyebrow	77.7778	66.0606	79.6970	77.4747	66.3636	80.0000
outer eyebrow	72.4242	61.7172	76.5657	72.2222	60.0000	75.6566
skin	78.9899	74.9495	80.2020	77.3737	72.3232	79.5960

Table 4.2: Rank-1 results of fusion of scales experiments using FERET images.

Tables 3.1 and 3.2. It can be seen that in every case the weighted fusion of multiple scales results in higher performance than when using the single smallest scale. The size of the difference in performance is related to the usefulness of features from multiple scales. The largest increase in performance using FRGC images was 29.8574 and the smallest was 5.5725. For FERET images it was 29.1962 and 4.9785.

The optimal scales and the amount they contributed to the results, shown in Table 4.3, can offer some insight into the different sub-regions of the periocular region. For instance, the optimal scales for LBP features extracted from the upper and lower eyelid are Scale 1 and Scale 3. The eyelids consist of small lines; the features extracted from the smallest scales are most discriminative. The tear duct has a more equal distribution between Scale 1 and Scale 3 when using LBP features. The outer corner of the eye has a higher contribution for Scale 1 features. However, more scales are involved. The results of experiments using LBP features extracted from the corners of the eye show an increase in contribution from higher scales. The eye corners do not have the small lines that are present in the eyelids, which is reflected in the resulting optimal scales. The eyebrows have similar levels of contribution among scales than the eye corners, but with slightly higher weights for the higher scales. Skin has similar weight results to the eyebrow and eye corners.

It has already been discussed how HOG features are invariant to changes in resolution and blur.

Sub-region	Scale 1	Scale 3	Scale 5	Scale 7
upper eyelid	60-80	20-40		
lower eyelid	60-80	20-40		
tear duct	50-55	45	0-5	0-5
outer corner	55-60	15-30	10	0-10
inner eyebrow	35-70	30-50	0-30	0-10
outer eyebrow	35-45	30-50	10-40	0-10
skin	25-75	15-40	5-20	0-10

Table 4.3: Contribution in percent of each scale to the results of the fusion of scales experiments for LBP features. A range is shown from the Left Eye FRGC, Right Eye FRGC, Left Eye FERET, and Right Eye FERET experiments.

This invariance to change is related to the idea of changes in scale, so it is expected that HOG features would be similarly discriminative regardless of the scales from which they were extracted. The optimal scales presented by HOG features extracted from different sub-regions of the periocular region do not offer the same level of insight as the LBP features. Almost all scales in all sets of data are found to contribute to the optimal weighted fusion of scales. In most cases all scales contribute equally. Despite no one scale being clearly more discriminative than another, the fusion of scales still produces much greater performance results than the baseline. The seemingly patternless contribution of the scales limits the effectiveness of this approach because the contribution of scales cannot be modeled.

The analysis of the LPQ features shares similarities with both LBP and HOG. The results are displayed in Table 4.4. Like HOG, most scales of LPQ features contribute to the optimal fusion of scales for each sub-region of the periocular region in most of the datasets. Like LBP, not all scales contribute equally and the scales that contribute the most and the degree to which they contribute change with the sub-region. The upper and lower eyelids have the highest contribution from the lower scales and the contribution becomes smaller as the scale increases. The contributions of the tear duct and the outer corner of the eye is where the results from LPQ features differ from the results using LBP features. Scale 3 has the highest weight. The eyebrow and skin sub-regions show a similar pattern to the eye corners.

4.5 Conclusion

The goals of Chapter 4 were to modify the LABF algorithms to extract features from multiple scales, to determine if the single small scale used in previous research provided the best performance in a typical biometric experiment, and to investigate if multiple scales used in conjunction could provide an increase in

Sub-region	Scale 1	Scale 3	Scale 5	Scale 7	Scale 9	Scale 11
upper eyelid	35-50	10-30	15-30	10	0-10	0-5
lower eyelid	25	30	30-35	0-10	0-5	0-5
tear duct	0-5	30-60	25-30	10-30	0-10	0
outer corner	20-45	5-10	10	10-15	10-15	10-15
inner eyebrow	0-25	25-45	10-15	25-45	0-15	0
outer eyebrow	15-40	15-35	5-15	15-35	0-15	0-15
skin	0-50	0-25	0-10	20-35	5-35	5

Table 4.4: Contribution in percent of each scale to the results of the fusion of scales experiments for LPQ features. A range is shown from the Left Eye FRGC, Right Eye FRGC, Left Eye FERET, and Right Eye FERET experiments.

performance. The modified LABF algorithms were used in experiments to show that the best performing scale was not always the smallest scale. Scale 1 was typically the best performing scale for LBP experiments, but Scale 1 and Scale 9 were similar in performance for HOG features and Scale 7 was typically the best for the LPQ experiments. Experiments that used a weighted fusion of scales were also performed and resulted in increased performance. The largest increase in Rank-1 using FRGC images was 29.8574% and the smallest was 5.5725%. For FERET images, it was 29.1962% and 4.9785%. The method developed in this chapter can be used in conjunction with the method developed in Chapter 3 to produce a novel algorithm for periocular feature extraction that incorporates the discriminative power of features from multiple scales.

Chapter 5

A Novel Method for Periocular Recognition

5.1 Introduction

The structure of the periocular region and how it influences the performance of an LABF-based periocular recognition system was examined in Chapter 3. The experiments presented in Chapter 3 suggest that some sub-regions of the periocular region are more discriminative than others. Using this knowledge, a novel block placement method for LABFs was proposed. When using only features extracted from each individual sub-region, the results of biometric experiments presented in this chapter suggest that LBP features are most discriminative in the upper eyelid, lower eyelid, tear duct, and outer corner, while LPQ features are most discriminative in the inner eyebrow, outer eyebrow, and skin. The experimentation of Chapter 3 could serve as preliminary or training data for a biometric system that considers the structure of the periocular region in determining recognition accuracy.

The performance of LABFs at multiple scales was examined in Chapter 4. Previous research looked at only a single, commonly the smallest, scale and discarded potentially useful information. The modifications made to the original LABF algorithms resulted in little increase in computation complexity, yet demonstrated a significant difference in performance in a basic biometric experiment. The method developed in Chapter 4 can be used in conjunction with the method developed in Chapter 3 to produce a novel algorithm for periocular feature extraction that incorporates the discriminative power of features from multiple scales.

These two areas of exploration are specific to the periocular region and can be used in conjunction to form the basis of a biometric feature extraction algorithm. This algorithm utilizes unique properties of the periocular region to provide a boost to recognition system performance over existing and more generalized methods.

The goals of this Chapter are as follows:

- Detail a novel feature extraction method for use with periocular region data that takes into consideration the structure of the periocular region as well as information from multiple scales.
- Perform experimentation to compare the performance of the new method to existing methods of periocular feature extraction.

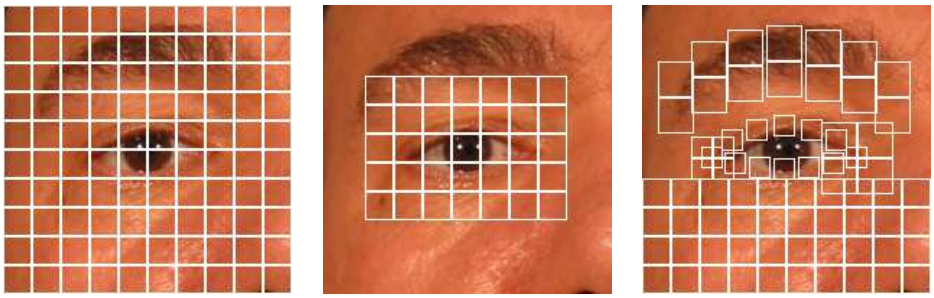
5.2 Data

The experiments reported in this chapter use the experimental datasets detailed in Chapter 2: the FRGC Experiment 1 subset and the FERET database. As explained before, these datasets are commonly used within the biometric research community to compare the relative performance of two methods. The new method detailed in this chapter can be compared to the basic methods presented in Chapter 2.

5.3 Method

A novel feature extraction method is presented in this chapter for use with periocular region data. This method is the logical fusion of the methods detailed in Chapter 3 and Chapter 4. Both methods have the potential to work together in such a way that the proposed method will offer an increase in performance over existing periocular feature extraction methods. As a reminder, the basic components of a biometric recognition algorithm are image preprocessing, feature extraction, feature comparison, and classification. Both elements of the proposed approach influence the feature extraction.

The first part of the proposed approach comes from the new block configuration method of Section 3.5. As a reminder, Figure 5.1 shows the proposed block arrangement beside the most common arrangements. Note that the blocks are not placed in a grid pattern like previous approaches. Instead, blocks are placed so that they correspond to physical sub-regions of the periocular region. These sub-regions of the periocular region are the upper eyelid, the lower eyelid, tear duct, outer corner, inner eyebrow, outer eyebrow, and the



(a) (b) (c)

Figure 5.1: Different models for block placement when using LABF: (a) Miller (b) Park (c) Proposed

skin. The suggested feature extraction method used varies based on the sub-region. Based on the results from Section 3.6, LBP features are extracted from the upper eyelid, lower eyelid, tear duct, and outer corner. LPQ features are extracted from the inner eyebrow, outer eyebrow, and skin.

The second part of the proposed approach involves using modified LBAFs to extract features from multiple scales. The details of the algorithms can be found in Section 4.3.1, Section 4.3.2, and Section 4.3.3. The experimental results presented in Section 4.4 used a weighted fusion of features from multiple scales to present the optimal results for those experiments. The experiments in this chapter will use the same scheme of weighted fusion. In this sense, the experiments of Section 4.4 serve as training data for this chapter's experiments.

5.4 Results

The experiments presented in this chapter follow the basic biometric experiment guidelines detailed in Section 2.3. Results are reported for experiments using the method proposed in Section 5.3 in comparison to the methods from Miller *et al.* [31] and Park *et al.* [38].

Table 5.1 shows the Rank-1 recognition rate, equal error rate, verification rate at 0.1% false accept rate, and D' for the three different LABF methods on FRGC Experiment 1 using both the Miller *et al.* and Park *et al.* methods. Table 5.2 shows the same performance metrics for the proposed method. Tables 5.3 and 5.4 show the same results for the FERET experiments, respectively. It can be seen that the proposed approach provides better performance results than any of the other approaches. The reasons to expect superior performance have been discussed in the previous chapters of this dissertation. The results shown here validate the theoretical assumptions given in previous chapters.

	Left Periocular				Right Periocular			
LBP								
	Rank-1	EER	VR at 0.1% FAR	D'	Rank-1	EER	VR at 0.1% FAR	D'
Miller	99.7068	8.8323	64.8841	2.7331	99.7005	8.2016	69.7151	2.4401
Park	98.6960	10.8464	46.2132	2.8250	98.8520	10.4823	48.6871	2.5137
HOG								
	Rank-1	EER	VR at 0.1% FAR	D'	Rank-1	EER	VR at 0.1% FAR	D'
Miller	99.6069	8.0829	69.6951	2.8350	99.6444	7.5245	72.2473	2.9378
Park	98.7834	10.3863	61.3708	2.5107	98.8208	9.9537	62.6322	2.5849
LPQ								
	Rank-1	EER	VR at 0.1% FAR	D'	Rank-1	EER	VR at 0.1% FAR	D'
Miller	99.7692	7.1183	75.9181	2.8654	99.7816	6.7227	76.6574	2.9439
Park	98.9706	11.0274	47.7062	2.4111	99.0454	10.5711	49.4899	2.4872

Table 5.1: Results of experiments using existing feature extraction methods on images from the FRGC Experiment 1 dataset

	Left Periocular				Right Periocular			
	Rank-1	EER	VR at 0.1% FAR	D'	Rank-1	EER	VR at 0.1% FAR	D'
Proposed	99.9321	6.9887	78.1764	2.9854	99.9798	6.6785	79.5894	3.1154

Table 5.2: Results of experiments using the proposed method on images from the FRGC Experiment 1 dataset

	Left Periocular				Right Periocular			
LBP								
	Rank-1	EER	VR at 0.1% FAR	D'	Rank-1	EER	VR at 0.1% FAR	D'
Miller	90.2020	4.9208	80.7071	3.2319	87.2727	5.2477	77.3737	3.2113
Park	72.7273	15.7522	47.4747	1.9316	72.8283	16.0677	45.2525	1.9190
HOG								
	Rank-1	EER	VR at 0.1% FAR	D'	Rank-1	EER	VR at 0.1% FAR	D'
Miller	87.1717	5.0597	80.6061	3.3462	86.6667	5.3547	78.6869	3.3305
Park	70.6061	9.1828	62.3232	2.7286	70.0000	9.4609	61.3131	2.6758
LPQ								
	Rank-1	EER	VR at 0.1% FAR	D'	Rank-1	EER	VR at 0.1% FAR	D'
Miller	92.0202	4.6486	83.9394	3.2062	92.2222	4.8796	83.4343	3.1487
Park	77.1717	15.5542	56.8687	1.9655	77.2727	16.2383	55.9596	1.9133

Table 5.3: Results of experiments using existing feature extraction methods on images from the FERET dataset

	Left Periocular				Right Periocular			
	Rank-1	EER	VR at 0.1% FAR	D'	Rank-1	EER	VR at 0.1% FAR	D'
Proposed	94.0667	4.0462	85.1113	3.5005	93.8887	4.3232	84.9190	3.8682

Table 5.4: Results of experiments using the proposed method on images from the FERET dataset

In experiments using both FRGC and FERET images, the proposed method produced a higher VR @ 0.1% FAR than either of the commonly used methods. The increase was between 1% and 3%. This performance comparison is given in regard to experiments that use periocular images only. A comparison to experiments that use face images would be an inappropriate comparison because the proposed method is intended to be used with periocular region images only. It is possible to conduct similar analysis on features extracted from the full face and produce a method similar to the one proposed in this chapter but such work falls outside of the scope of this dissertation. Also, we are not suggesting that the periocular region be used to replace the face when face data is available and collected in ideal settings.

5.5 Conclusions and Future Work

In Chapter 1, the periocular region was proposed as a useful biometric modality in response to the failure of traditional biometric modalities, such as face and iris, in non-ideal situations, e.g. the presence of occlusion or closed eyelids. The periocular region was seen as useful because it is captured alongside face and iris data during the standard biometric data collection process. The biometric system that saw a failure in the face or iris would immediately be able to use the periocular region to supplement the identification process.

Chapter 2 offered an examination of the periocular region by addressing three aspects of biometric recognition using the periocular region. First, common facial recognition data and common facial feature extraction algorithms were examined as they are implemented in a periocular based biometric system. The experimentation in Chapter 2 showed similar levels of performance between the face and periocular region for all of the features. Second, sub-regions of the face were examined to investigate if some produce more discriminative features than others. These experiments reinforced the claims from the literature review that the periocular region is the most discriminative region of the face. Third, environment-influenced concerns of biometric data were examined in the context of periocular recognition. These experiments showed that certain feature extraction methods stand up to non-ideal conditions better than others, and this information can be used to influence algorithms targeted at periocular recognition.

To further explore aspects of the periocular region that might influence the creation of a periocular specific biometric recognition algorithm, Chapter 3 looked at the structure of the periocular region. The experiments presented in this chapter suggest that the most discriminative sub-regions of the periocular region are near the eyebrows and eye corners. This observation was made by examining experimental results from

two different biometric datasets and three different feature extraction algorithms. The extent of the experimentation, which is large in comparison to any of the research mentioned in the literature review, supports the notion that the high performance of certain sub-regions of the periocular region is a universal trait of biometric images taken of the periocular region. A novel block placement method for LABFs was proposed in response to this discovery. This block placement method was based on the location of the structural sub-regions of the periocular region. When using only features extracted from each individual sub-region, the results of biometric experiments presented in this chapter suggested that LBP features are most discriminative in the upper eyelid, lower eyelid, tear duct, and outer corner, while LPQ features are most discriminative in the inner eyebrow, outer eyebrow, and skin.

One aspect of the use of LABFs in the literature is that they are typically used in the smallest scale possible. The work of Chapter 4 modifies the LABF methods to work in multiple scales and discovers that each algorithm has a different optimal performance scale. This experimentation was also performed using two biometric datasets and three feature extraction algorithms.

All of the knowledge of the periocular region that was discovered up to this point was used to produce a biometric recognition algorithm that is designed specifically to be used with the periocular region. The experimental results of that algorithm are presented in this chapter in comparison to two different methods in the literature and using two different biometric datasets. The theoretic and experimental efforts of this dissertation, in addition to producing a better understanding of the periocular region, lead to the creation of an algorithm that performed better at the task of recognition when presented with two different commonly used biometric datasets.

One shortcoming of the method developed in this dissertation is that a portion of the data used in the experiments of this chapter is used in the experiments of Chapter 3 and Chapter 4. So, data, from which the observations about the feature extraction methods that perform best in each sub-region of the periocular region, is used in part to conduct experiments that support the claim that the novel method for periocular recognition presented in this chapter produces better recognition performance than previous methods. The biometrics research community does not typically find this type of shortcoming to be significant. It is unavoidable in most cases because of the lack of high-quality, publicly available biometric data. There is no perfect biometric database and all research must work with what is available.

Some research will collect their own set of biometric data. These datasets are typically much smaller than either FERET or FRGC and not publicly available. Collecting our own set of biometric data for the experimentation of this dissertation was infeasible. Datasets such as the FRGC took over three years and

many researchers to compile. Datasets to the scale of FRGC are collected for the sake of collecting the data and not with specific scientific research in mind because of the time investment required.

Even though the data used to explore aspects of the periocular region and the data used to test the proposed method overlap, the observations made in each chapter come from using more than one dataset and using more than one feature extraction algorithm. Many of the observations are consistent across these variables which suggests that the observations would likely be made from experimentation with any periocular data. One suggested area of future work would be to test the proposed method on a new and larger set of periocular data.

This dissertation is focused on how the periocular region and the proposed method perform in the basic biometric experiments, e.g. frontal face images under controlled lighting conditions. This method could be tested in other, more difficult problems such as the aging of a subject. Aging is a difficult problem in biometrics research and it would be interesting to see how the proposed method could be used to address the problem of aging.

The observations about the types of features extracted from the periocular region could be used in many unforeseen ways. The observations and the methods used to generate them can be used to inform future research into all areas of biometrics, not just periocular recognition. The contributions of this dissertation to the state of biometrics research is much more significant than a method that provides improved performance in one biometric problem.

Appendices

Appendix A Abbreviations

CMC: Cumulative Match Characteristic

DET: Detection Error Tradeoff

MSD: Match Score Distribution

FRR: False Reject Rate

FAR: False Accept Rate

EER: Equal Error Rate

VR: Verification Rate

FERET: Facial Recognition Techology

FRGC: Facial Recognition Grand Challenge

PCA: Principle Component Analysis

LBP: Local Binary Patterns

HOG: Histogram of Oriented Gradients

LPQ: Local Phase Quantization

WLD: Webber Local Descriptor

SIFT: Scale Invariant Feature Transform

SURF: Speeded Up Robust Features

LABF: Local Appearance Based Features

Bibliography

[1] T. Ahonen, A. Hadid, and M. Pietikäinen. Face recognition with local binary pattern. In *Proceedings of the European Conference on Computer Vision*, pages 469–481, May 2004.

[2] T. Ahonen, A. Hadid, and M. Pietikäinen. Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12):2037–2041, 2006.

[3] T. Ahonen, E. Rahtu, V. Ojansivu, and J. Heikkilä. Recognition of blurred faces using local phase quantization. In *Proceedings of the International Conference on Pattern Recognition*, pages 1–4, December 2008.

[4] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. SURF: Speeded Up Robust Features. *Computer Vision and Image Understanding*, 110(3):346–359, 2008.

[5] S. Bharadwaj, H. Bhatt, M. Vatsa, and R. Singh. Periocular biometrics: When iris recognition fails. In *Proceedings of the IEEE International Conference on Biometrics: Theory, Applications, and Systems*, pages 1–6, September 2010.

[6] M. Bicego, A. Lagorio, E. Grosso, and M. Tistarelli. On the use of SIFT features for face authentication. In *Proceedings of the Conference on Computer Vision and Pattern Recognition Workshop*, page 35, June 2006.

[7] C. Chan, J. Kittler, and K. Messer. Multi-scale local binary pattern histograms for face recognition. In *Advances in Biometrics*, volume 4642 of *Lecture Notes in Computer Science*, pages 809–818. Springer, 2007.

[8] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikainen, X. Chen, and W. Gao. WLD: A robust local image descriptor. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1705–1720, September 2010.

[9] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 886–893, June 2005.

[10] J. Daugman. How iris recognition works. *IEEE Transactions on Circuits and Systems for Video Technology*, 16(1):21–30, 2004.

[11] O. Déniz, G. Bueno, J. Salido, and F. De la Torre. Face recognition using histograms of oriented gradients. *Pattern Recognition Letters*, 32(12):1598–1603, 2011.

[12] P. Dreuw, P. Steingrube, H. Hanselmann, and H. Ney. SURF-face: Face recognition under viewpoint consistency constraints. In *Proceedings of the British Machine Vision Conference*, pages 1–11, September 2009.

[13] G. Du, F. Su, and A. Cai. Face recognition using SURF features. In *Proceedings of the International Symposium on Multispectral Image Processing and Pattern Recognition*, October 2009.

[14] National Center for State Courts. Individual biometrics: Iris scan. http://ctl.ncsc.dni.us/biomet

[15] Francis Galton. *Finger Prints*. Macmillan, 1892.

[16] K. Hollingsworth, K. Bowyer, and P. Flynn. Identifying useful features for recognition in near-infrared periocular images. In *Proceedings of the IEEE International Conference on Biometrics: Theory Applications and Systems*, pages 1–8, September 2010.

[17] K. Hollingsworth, S. Darnell, P. Miller, D. Woodard, K. Bowyer, and P. Flynn. Human and machine performance on periocular biometrics under near-infrared light and visible light. *IEEE Transactions on Information Forensics and Security*, 7(2):588–601, April 2012.

[18] A. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):4–20, January 2004.

[19] A. Joshi, A. Gangwar, R. Sharma, and Z. Saquib. Periocular feature extraction based on LBP and DLDA. In *Advances in Computer Science, Engineering and Applications*, volume 166 of *Advances in Intelligent and Soft Computing*, pages 1023–1033. Springer, 2012.

[20] F. Juefei-Xu, K. Luu, M. Savvides, T. Bui, and C. Suen. Investigating age invariant face recognition based on periocular biometrics. In *Proceedings of the International Joint Conference on Biometrics*, pages 1–7, October 2011.

[21] F. Juefei-Xu and M. Savvides. Unconstrained periocular biometric acquisition and recognition using COTS PTZ camera for uncooperative and non-cooperative subjects. In *Proceedings of the IEEE Workshop on Applications of Computer Vision*, pages 201–208, January 2012.

[22] B. Kamgar-Parsi, B. Kamgar-Parsi, and A. Jain. Synthetic eyes. In *Proceedings of the 4th international conference on Audio- and video-based biometric person authentication*, pages 412–420, June 2003.

[23] D. Kisku, A. Rattani, E. Grosso, and M. Tistarelli. Face identification by SIFT-based complete graph topology. In *Proceedings of the IEEE Workshop on Automatic Identification Advanced Technologies*, pages 63–68, June 2007.

[24] Peter Komarinski. *Automated Fingerprint Identification Systems (AFIS)*. Academic Press, 2004.

[25] D. G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1150–1157, September 1999.

[26] J. Luo, Y. Ma, E. Takikawa, S. Lao, M. Kawade, and B. Lu. Person-specific SIFT features for face recognition. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 593–596, April 2007.

[27] J. Lyle, P. Miller, S. Pundlik, and D. Woodard. Soft biometric classification using periocular region features. In *Proceedings of the IEEE International Conference on Biometrics: Theory, Applications, and Systems*, pages 1–7, September 2010.

[28] J. Lyle, P. Miller, S. Pundlik, and D. Woodard. Soft biometric classification using local appearance periocular region features. *Pattern Recognition*, 45(11):3877–3885, 2012.

[29] J. Merkow, B. Jou, and M. Savvides. An exploration of gender identification using only the periocular region. In *Proceedings of the IEEE International Conference on Biometrics: Theory, Applications, and Systems*, pages 1–5, September 2010.

[30] P. Miller, J. Lyle, S. Pundlik, and D. Woodard. Performance evaluation of local appearance based periocular recognition. In *Proceedings of the IEEE International Conference on Biometrics: Theory, Applications, and Systems*, pages 1–6, September 2010.

[31] P. Miller, A. Rawls, S. Pundlik, and D. Woodard. Personal identification using periocular skin texture. In *Proceedings of the ACM Symposium on Applied Computing*, pages 1496–1500, March 2010.

[32] The National Institute of Standards and Technology (NIST). The face recognition grand challenge (FRGC). http://www.nist.gov/itl/iad/ig/frgc.cfm, 2011.

[33] T. Ojala, M. Pietikäinen, and T. Mäenpää. Gray scale and rotation invariant texture classification with local binary patterns. In *Proceedings of the European Conference on Computer Vision*, pages 404–420, Dublin, Ireland, June 2000.

[34] T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.

[35] V. Ojansivu and J. Heikkilä. Blur insensitive texture classification using local phase quantization. In *Proceedings of the International Conference on Image and Signal Processing*, pages 236–243, July 2008.

[36] V. Ojansivu and J. Heikkilä. Blur insensitive texture classification using local phase quantization. In *Image and Signal Processing*, volume 5099 of *Lecture Notes in Computer Science*, pages 236–243. Springer, 2008.

[37] U. Park, R. Jillela, A. Ross, and A. Jain. Periocular biometrics in the visible spectrum. *IEEE Transactions on Information Forensics and Security*, 6(1):96–106, March 2011.

[38] U. Park, A. Ross, and A.K. Jain. Periocular biometrics in the visible spectrum: A feasibility study. In *Proceedings of the IEEE International Conference on Biometrics: Theory, Applications, and Systems*, pages 1–6, September 2009.

[39] V. Pauca, M. Forkin, X. Xu, R. Plemmons, and A. Ross. Challenging ocular image recognition. *Sensing Technologies for Global Health, Military Medicine, Disaster Response, and Environmental Monitoring; and Biometric Technology for Human Identification VIII*, 8029(1):80291V, 2011.

[40] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 947–954, June 2005.

[41] P. Phillips, H. Wechsler, J. Huang, and P. Rauss. The FERET database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing*, 16:295–306, 1998.

[42] Henry T. F. Rhodes. *Alphonse Bertillon: Father of Scientific Detection*. George G. Harrap, 1956.

[43] Y. Rodriguez and S. Marcel. Face authentication using adapted local binary pattern histograms. In *Proceedings of the European Conference on Computer Vision*, pages 321–332, May 2006.

[44] A. Ross, R. Jillela, J. Smereka, V. Boddeti, B. VijayaKumar, R. Barnard, X. Hu, P. Pauca, and R. Plemmons. Matching highly non-ideal ocular images: An information fusion approach. In *Proceedings of the IAPR International Conference on Biometrics*, pages 446–453, March 2012.

[45] G. Santos and E. Hoyle. A fusion approach to unconstrained iris recognition. *Pattern Recognition Letters*, 33(8):984–990, 2012.

[46] K. Sato, S. Shah, and J. Aggarwal. Partial face recognition using radial basis function networks. In *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, pages 288–293, April 1998.

[47] M. Savvides, R. Abiantun, J. Heo, S. Park, C. Xie, and B. Vijayakumar. Partial holistic face recognition on FRGC-II data using support vector machine. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop*, page 48, June 2006.

[48] W. Schwartz, H. Guo, and L. Davis. A robust and scalable approach to face identification. In *Proceedings of the European Conference on Computer Vision*, pages 476–489, September 2010.

[49] Dr. Libor Spacek. Collection of facial images: Faces94. http://cswww.essex.ac.uk/mv/allfaces/faces94.html, 1994.

[50] T. Tan, X. Zhang, Z. Sun, and H. Zhang. Noisy iris image matching by using multiple cues. *Pattern Recognition Letters*, 33(8):970–977, June 2012.

[51] C. Teo, H. Neo, and A. Teoh. A study on partial face recognition of eye region. In *Proceedings of the International Conference on Machine Vision*, pages 46–49, December 2007.

[52] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neurosicence*, 3(1):71–86, 1991.

[53] D. Woodard, S. Pundlik, P. Miller, R. Jillela, and A. Ross. On the fusion of periocular and iris biometrics in non-ideal imagery. In *Proceedings of the IEEE International Conference on Pattern Recognition*, pages 201–204, August 2010.

[54] D. Woodard, S. Pundlik, P. Miller, and J. Lyle. Appearance-based periocular features in the context of face and non-ideal iris recognition. *Signal, Image and Video Processing*, 5(4):443–455, 2011.

[55] J. Xu, M. Cha, J. Heyman, S. Venugopalan, R. Abiantun, and M. Savvides. Robust local binary pattern feature sets for periocular biometric identification. In *Proceedings of the IEEE International Conference on Biometrics: Theory, Applications, and Systems*, pages 1–8, September 2010.

[56] L. Yunqi, J. Xutuan, S. Zhenxiang, C. Dongjie, and L. Qingmin. Face recognition method based on SURF feature. In *Proceedings of the International Symposium on Computer Network and Multimedia Technology*, pages 1–4, December 2009.

[57] G. Zhang, X. Huang, S. Li, Y. Wang, and X. Wu. Boosting local binary pattern (LBP)-based face recognition. In *Advances in Biometric Person Authentication*, volume 3338 of *Lecture Notes in Computer Science*, pages 179–186. Springer, 2005.