Human and Machine Performance on Periocular Biometrics Under Near-Infrared Light and Visible Light

Karen P. Hollingsworth, Shelby S. Darnell, Philip E. Miller,

Damon L. Woodard, Kevin W. Bowyer, and Patrick J. Flynn

Abstract

Periocular biometrics is the recognition of individuals based on the appearance of the region around the eye. Periocular recognition may be useful in applications where it is difficult to obtain a clear picture of an iris for iris biometrics, or a complete picture of a face for face biometrics. Previous periocular research has used either visible-light or near-infrared light images, but no prior research has directly compared the two illuminations using images with similar resolution. We conducted an experiment in which volunteers were asked to compare pairs of periocular images. Some pairs showed images taken in visible light, and some showed images taken in near-infrared light. Participants labeled each pair as belonging to the same person or to different people. Untrained participants with limited viewing times correctly classified visible-light image pairs with 88% accuracy, and near-infrared images with 79% accuracy. For comparison, we presented pairs of iris images from the same subjects. In addition, we investigated differences between performance on light and dark eyes and relative helpfulness of various features in the periocular region under different illuminations. We calculated performance of three computer algorithms on the periocular images. Performance for humans and computers was similar.

Index Terms

ocular biometrics, iris recognition, periocular recognition, visible light, near-infrared light.

Hollingsworth, Bowyer, and Flynn are with the Computer Science and Engineering Department, University of Notre Dame, Notre Dame, IN 46556. Darnell, Miller, and Woodard are with the School of Computing, Clemson University, Clemson, SC 29634.

I. INTRODUCTION

The periocular region is the part of the face surrounding the eye. The term "periocular" comes from the prefix, "peri-", meaning "around or about", and the root word, "ocular", meaning "of or relating to the eye". In biometrics, the term has been applied to both a small region including the eye, eyelids, and lashes, or a larger region including the eyebrow [1]. The term has also been applied to regions that include the iris [2] and to regions that mask the eye, using only skin and eyebrow information around the eye [3]. Irises and faces have both been studied extensively as biometric characteristics [4], [5], but the periocular region has only recently received attention as a biometric characteristic. Some initial work has suggested that the periocular region may be as discriminating by itself as the face as a whole [3], [6].

Since periocular recognition is such a new field in the area of biometrics, there remain a number of open questions regarding this biometric characteristic. For instance, face recognition traditionally uses visible light but iris recognition uses near-infrared light; which (if any) light spectrum should be preferred for periocular recognition? Which lighting condition reveals more discriminatory information in the periocular region? Which features in the region are most useful in the two different modalities?

In such a new field of research, there is no single automatic algorithm generally accepted as the "best" for the periocular region. Therefore, in order to investigate questions about different lighting for the periocular region, we turned to the human visual system. Other computational vision problems have benefitted from a good understanding of the human visual system. In a recent book chapter, O'Toole [7] says, "Collaborative interactions between computational and psychological approaches to face recognition have offered numerous insights into the kinds of face representations capable of supporting the many tasks humans accomplish with faces" [7]. Sinha et al. [8] describe numerous basic findings from the study of human face recognition that have direct implications for the design of computational systems. Their report says "The only system that [works] well in the face of [challenges like sensor noise, viewing distance, and illumination] is the human visual system. It makes eminent sense, therefore, to attempt to understand the strategies this biological system employs, as a first step towards eventually translating them into machine-based algorithms" [8].

Studying how the human visual system performs on near-infrared and visible-light images has three different purposes. First, it provides a baseline for judging performance of forensic human examiners who might at some point need to compare periocular images. Second, it aids in the development of automatic recognition algorithms by providing insights about features in the region that contain identifying

information. Third, it sets a baseline for judging performance of automatic recognition algorithms.

We displayed pairs of images to human volunteers and asked them to determine whether the pair showed two images from the same person or from two different people. Our experiments were designed to answer questions such as the following:

- How well can humans identify people from the periocular region, and how does performance change for visible-light (VL) images versus near-infrared (NIR) images?
- Does performance in a given condition depend on whether the images showed light-eyed subjects or dark-eyed subjects?
- Are individual subjects that are difficult to identify in one modality also difficult in other modalities?
- Which periocular features are most discriminatory under near-infrared light?
- Which periocular features are most discriminatory under visible light?
- How well can humans identify subjects using the iris region?

We used three computer algorithms to analyze the same periocular images. The algorithms used local binary patterns (LBPs), histograms of oriented gradients (HOGs), and scale-invariant feature transform (SIFT). We employed similar analysis on the machine-generated results as we had on the human performance experiments.

Our experiments are presented as follows. Section II summarizes prior research in periocular biometrics. Section III describes our data set and preprocessing. Our experimental methods are described in Section IV. Results are given in Section V. Section VI summarizes our findings and presents recommendations for future work.

II. RELATED WORK

Research in periocular biometrics can be divided into two categories, namely image classification and image recognition. The first category includes papers that extract various features in an attempt to classify an image based on gender, ethnicity, or whether the image shows a left or right eye. The second category includes papers that focus specifically on recognition. Some work combines periocular recognition with other types of recognition to improve performance.

A. Periocular Feature Extraction for Image Classification

A classifier that determines whether an eye image contains a left or right eye is a valuable tool for detecting errors in labeled data. Abiantun and Savvides [9] first used the locations of the pupil and iris centers to differentiate between right and left eyes. The pupil is often located to the nasal side of the iris

rather than being directly in the center. Subsequently, they evaluated five different methods for detecting the tear duct in an iris image: (1) Adaboost algorithm with Haar-like features, (2) Adaboost with a mix of Haar-like and Gabor features, (3) support vector machines, (4) linear discriminant analyisis, and (5) principal component analysis. Their tear-duct detector using boosted Haar-like features correctly classified 179 of 199 images where their first method had failed.

Bhat and Savvides [10] used active shape models (ASMs) to fit the shape of the eye and predict whether an eye is a right or left eye. They trained two different ASMs: one for right eyes, and one for left eyes. They ran both ASMs on each image, and evaluated the fit of each using Optimal Trade-off Synthetic Discriminant Filters. On poorly-centered iris images, their classifier correctly classified 62.5% of left eyes correctly and 67.7% of right eyes. On centered images with consistent scale, they classified 91% of left eyes correctly and 89% of right eyes correctly.

The papers aimed at predicting race and gender from the periocular region use markedly different approaches. Li et al. [11] used eyelashes for ethnic classification. They observed that Asian eyelashes tend to be more straight and vertically oriented than Caucasian eyelashes. To extract eyelash feature information, they used active shape models to locate the eyelids. Next, they identified nine image patches along each eyelid. They applied uni-directional edge filters to detect the direction of the eyelashes in each image patch. After obtaining feature vectors, they used a nearest neighbor classifier to determine whether each image showed an Asian or a Caucasian eye, and achieved a 93% correct classification rate.

Merkow et al. [12] used grayscale pixel intensities and local binary patterns (LBPs) for gender classification on 936 images (468 male and 468 female) downloaded from the web. They tested three types of classification algorithms: (1) linear discriminant analysis (LDA) with a rank-1 nearest neighbor classifier, (2) principal component analysis (PCA) with a rank-1 nearest neighbor classifier, and (3) linear support vector machines (SVMs). For both types of features and multiple-sized regions, LDA and SVMs outperformed PCA. On a periocular region containing the eyes, eyebrows, and the bridge of the nose, PCA performance was near 70%, LDA achieved about 80% accuracy, and SVMs achieved 84.9% accuracy.

Lyle et al. [13] performed both gender and ethnicity classification using 4232 periocular images of 404 subjects, cropped from the Face Recognition Grand Challenge (FRGC) [14] data. They also used grayscale pixel intensities and local binary patterns as features, but utilized a non-linear SVM as a classifier. They achieved 93% gender classification accuracy when using periocular images taken under controlled lighting. They classified images as Asian and Non-asian and achieved about 91% accuracy.

B. Periocular Recognition Performance

Park et al. [2] presented a feasibility study for the use of periocular biometrics. The authors implemented two methods for analyzing the periocular region. In their "global method", they used the center of the limbus circle as an anchor point. They defined a grid around the iris and computed gradient orientation histograms and local binary patterns (LBPs) for each point in the grid. They quantized each feature into eight distinct values to build an eight-bin histogram, and then used Euclidean distance to evaluate a match. Their "local method" involved detecting key points using a SIFT matcher. They collected a database of 899 high-resolution visible-light face images from 30 subjects. A face matcher obtained 100% rank-one recognition for these images, and their matcher that used only the periocular region yielded 77% correct rank-one recognition. Their initial work has been extended and published as a journal paper [1].

Miller et al. [15] also used LBPs to analyze the periocular region. They used visible-light face images from the FRGC data [14] and the Facial Recognition Technology (FERET) data [16]. The periocular region was extracted from the face images using the provided eye center coordinates. They extracted the LBP histogram from each block in the image and used city block distance to compare the information from two images. They achieved 89.76% rank-one recognition on the FRGC data, and 74.07% on the FERET data. Adams et al. [17] extended Miller's work by training a genetic algorithm to select the subset of features that would be best for recognition. The use of the genetic algorithm increased accuracy from 89.76% to 92.16% on FRGC. On FERET, the accuracy increased from 74.04% to 85.06%.

Woodard et al. [3] added color histogram matching to LBP features. They tested their method on subsets of images from FRGC 2.0, experiment 1. After testing multiple distance metrics, they found that city block metric is best for comparing LBP features and that Bhattacharya coefficient performs best for color features. They applied their same algorithms to the whole face for comparison. On the whole face, adding color to LBP texture never improved performance. On the periocular region, adding color helped on the subsets of images using different expressions. The highest performance score of 98.3% rank-1 recognition came from same-session comparisons. Different session comparisons gave 91% rank-1 recognition.

Miller et al. [6] analyzed how blur, resolution, and lighting affect periocular recognition using an LBP algorithm. At extreme values of blur, periocular recognition performed significantly better than face recognition. The results for down-sampling the resolution were similar. Rank-1 recognition rates for both face and periocular matching under uncontrolled lighting (using FRGC 2.0 experiment 4 images) were very poor, indicating that local binary patterns are not well suited to recognizing subjects under

uncontrolled lighting. Miller et al. also explored different color channels and found that using the green color channel gave the highest recognition rate. Experiments indicated that the red color channel contained significantly less discriminative texture.

Xu et al. [18] tested several feature extraction techniques on the challenging FRGC 2.0 experiment 4 data. They tried Walsh Masks, Laws' Masks, Discrete Cosine Transform, Discrete Wavelet Transform, Force Field Transform, Gabor Filters, Laplacian of Gaussian Filters, LBPs, SIFT features, and SURF features. Feature sets were compared using various distance metrics. Fusing many of these feature types with local binary patterns improved verification performance. They achieved the best performance using Local Walsh-Transform binary pattern encoding.

Bharadwaj et al. [19] proposed a periocular recognition algorithm to use on the UBIRIS v2 low-quality, visible-light iris data set. These images are particularly challenging for iris recognition because they are taken in visible light at distances between four and eight meters. Bharadwaj et al. suggested using both global and local features. Their global descriptors use second order statistics evaluating the spatial arrangement of components in the image. These descriptors test for attributes such as the distribution of edges in the horizontal and vertical directions or the size of prominent objects in the image. They use circular local binary patterns for local features. By combining global and local features in both left and right eye regions, they obtained a rank-1 identification accuracy of 73.65%.

While the above papers focused exclusively on periocular recognition, Woodard et al. [20] fused information from a periocular recognition algorithm with information from an iris recognition algorithm. Using the MBGC near-infrared portal data the authors analyzed the iris texture using a traditional Daugman-like algorithm [21], and they analyzed the periocular texture using LBPs. Periocular identification performed better than iris identification, and the fusion of the two modalities performed best.

Hollingsworth et al. [22] investigated human performance on a periocular recognition task. They found that when presented with unlabeled pairs of near-infrared periocular images, humans could correctly determine whether the two images were from the same person or from different people with 92% accuracy. They further found that eyelashes, tear ducts, eye shape, and eyelids were helpful when determining whether two images matched.

This paper differs from previous research in that it is the first to compare verification performance on the same subjects' images presented under four different conditions, namely both near-infrared and visible light, and both iris and periocular regions. This paper is also the first to present both human and machine performance results on the same periocular images. Furthermore, no other paper has investigated human verification performance on visible-light periocular images, or analyzed what features are helpful

to human participants under both lighting conditions.

III. Data

One of our goals of this experiment was to test whether near-infrared or visible light was better for the periocular region. We used a set of 640x480 near-infrared eye images acquired with an LG2200 EOU camera. We also acquired a set of 2613x3900 visible-light face images of the same subjects using a Nikon D80 CCD camera. All images were collected under a human subjects research protocol approved by the Institutional Review Board at the University of Notre Dame. The images were collected from cooperative users using controlled lighting. The average interpupillary distance was 1142 pixels. From such images, we manually marked the eye centers and cropped 640x480 visible-light periocular images from the face, centered at the marked point. The resulting visible-light images showed approximately the same amount of face region as the near-infrared LG2200 images.

We obtained images from 210 subjects, plus images from an additional 8 subjects to use in examples prior to the experiment. We planned to show pairs of images to volunteers and ask volunteers to determine whether two images in a pair were from the same person or from different people. Consequently, we assigned 70 of our subjects for whom we had eye images to be used in "match" trials, and 140 of our subjects to be used in "nonmatch" trials. Subjects were randomly assigned to "match" or "nonmatch" pools, with two exceptions. First, the 66 subjects who only came to one acquisition session were necessarily nonmatch subjects because we did not have two different images of them. Second, we required an equal number of light-eyed match and non-match queries, and similarly an equal number of dark-eyed queries in the two sets. To determine whether a subject was "light-eyed" or "dark-eyed", we manually examined each image and labeled blue, green, and hazel eyes as "light", and brown eyes as "dark."

After dividing the subjects into match and nonmatch pools, nonmatch subjects were paired so that two subjects in a pair had the same gender and race. In addition, similar subjects were paired as follows. For all possible pairs of images, we computed a difference score based on eye color, presence of makeup, dilation ratio, percent eye occlusion, eyelashes, and contact lenses. Let C_i be the eye color for the ith image, where 1 = blue eyes, 2 = green, 3 = hazel, 4 = light brown, and 5 = brown. Let M_i represent the amount of makeup for the ith image, where lack of makeup = 1, light makeup = 2, and heavy makeup = 3. L_i represents the presence of contact lenses, with 1 = no contacts and 2 = contacts. E_i represents the length of eyelashes, with 1 = short eyelashes, 2 = medium eyelashes, and 3 = long eyelashes. The dilation ratio D_i is the ratio of pupil radius to iris radius. The percent occlusion P_i is the percent of iris area covered by eyelashes or eyelids. Let d_M represent the maximum difference between two dilation ratios in
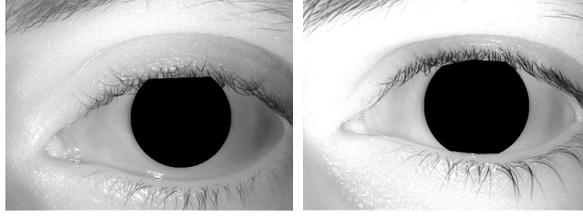
Fig. 1: This example nonmatch query shows subjects 05503 and 05458 using near-infrared periocular modality.

our image set and p_M represent the maximum difference between percent occlusion values. A difference score $\delta_{VL}(i,j)$ between the ith and j^{th} visible-light images in the set was defined to be

$$
\begin{aligned}
\delta_{VL}(i,j) &= |C_i - C_j| + |M_i - M_j| + \frac{1}{d_M}|D_i - D_j| + \\
&\quad \frac{1}{p_M}|P_i - P_j| + \frac{1}{2}|L_i - L_j| + \frac{1}{4}|E_i - E_j|.
\end{aligned}
$$

A difference score $\delta_{NIR}(i,j)$ between the ith and j^{th} near-infrared images was defined in the same way. In computing these difference scores, the differences in contact lenses and eyelash length were weighted less heavily because of the difficultly in accurately and objectively assigning those labels. To make nonmatch queries, we paired subjects i and j with smallest overall scores $\delta(i,j) = \delta_{VL}(i,j) + \delta_{NIR}(i,j)$.

For both match and nonmatch queries, we used images taken at least a week apart so that no query would show images from the same session. Additionally, we randomly chose whether to show two left eyes or two right eyes for the query. Ultimately, from 210 subjects, we had created 140 subject-disjoint query pairs, where half were match queries and half were nonmatch queries.

To separate the influence of the periocular region from the iris region in our experiment, we manually segmented the iris region in each image, and created two different images: one showing only the iris, and a second showing only the periocular region. As a result, we could display each query in one of four ways: periocular near-infrared, periocular visible light, iris near-infrared, and iris visible light (See Figures 1, 2, 3, and 4).

IV. METHOD

A. Human Experiments

We conducted our human experiments using a protocol approved by the Human Subjects Institutional Review Board at the University of Notre Dame. Volunteers were solicited from the University of Notre Dame student and staff population. They were offered $10 to participate in the experiment, and an
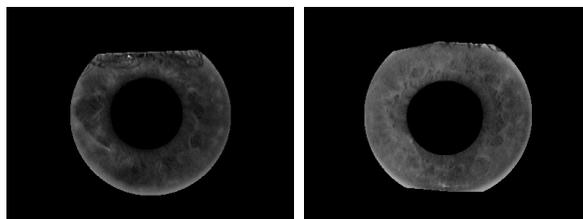
Fig. 2: This example nonmatch query shows subjects 05503 and 05458 using near-infrared iris modality.

Fig. 3: This example nonmatch query shows subjects 05503 and 05458 using visible-light periocular modality.

Fig. 4: This example nonmatch query shows subjects 05503 and 05458 using visible-light iris modality.

additional $5 if they scored in the top 50% of respondents. To the best of our knowledge, none of the volunteers had participated in any similar experiments before. Fifty-six people participated in the experiment. The demographic information of our subjects is shown in Table I.

We had four different sets of queries, based on periocular or iris content and near-infrared (NIR) or visible light (VL) illumination:

- Periocular NIR
- Iris NIR
- Periocular VL
- Iris VL

Each set of queries contained 140 pairs of images. To increase the number of people viewing each query set, we asked each volunteer to respond to two sets. The four sets could be paired in six different ways,

Table I: Demographic Information of Volunteers

Total number of volunteers	56
Number of males	23
Number of females	33
Age 18-19	22
Age 20-21	20
Age 22-30	10
Age 31-40	2
Age > 40	2

so we had the following six tasks:

1) Classify pairs of periocular NIR and iris NIR,

2) Classify pairs of periocular NIR and periocular VL,

3) Classify pairs of periocular NIR and iris VL,

4) Classify pairs of iris NIR and periocular VL,

5) Classify pairs of iris NIR and iris VL, and

6) Classify pairs of periocular VL and iris VL.

We randomized the order in which we presented the sets, so that in task one, for example, half of participants saw the periocular NIR images first, and half saw the iris NIR images first.

We wrote custom software to display our queries to our participants. The program displayed each pair of images for three seconds (Figure 5). After three seconds, the display would change, and the user would be prompted to respond whether the two images in the pair were from the same person or from two different people (Figure 6). The five possible responses were:

- Same person (certain)

- Same person (likely)

- Don't know

- Different people (likely)

- Different people (certain)

In our report in this paper, we have used the terms "match" and "same person" interchangeably. Also, the terms "nonmatch" and "different people" have been used interchangably.

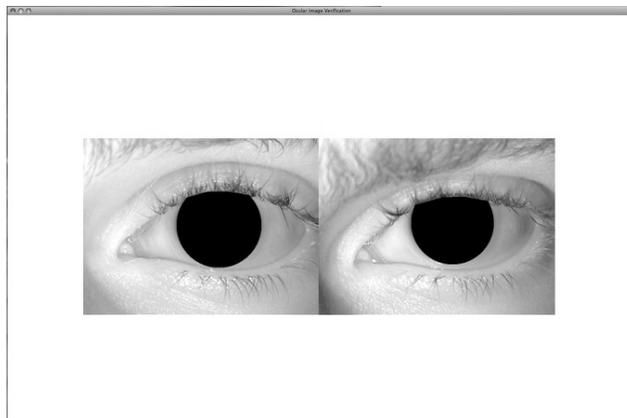We ran the experiment over the course of one week. Volunteers signed up for a time to participate, and

Fig. 5: Our software displayed a pair of images like the pair shown above for three seconds. After the three seconds, the images were hidden and users could respond whether the two images were same or different.

Fig. 6: After displaying a pair of images for three seconds, our software hid the images and prompted users to state whether the two images were same or different.

would come to our studio, register, and view example images and a practice question. Next they viewed the 140 queries of their first type of assigned images. After viewing those image pairs, if the images had shown the periocular region, the computer program would prompt them to rate the value of different features in making their decisions. The features that the volunteers were asked to rate were "Eye shape", "Tear Duct", "Outer Corner", "Eyelashes", "Skin", "Eyebrow", "Eyelid", "Color", "Blood vessels", and "Other". Next the volunteers viewed 140 queries of their second type of assigned images. Again, if the second type of images were of the periocular region, they would be prompted to rate the helpfulness of the features in the image. One entire participant session lasted approximately 40 minutes.

B. Machine Experiments on Periocular Images

We ran three automated experiments using appearance features in the periocular region. Appearance features can be extracted from input images using algorithms that create representations based on local or global information. The most obvious global representation approach, correlation, computes a difference between two images, on a pixel by pixel basis, for comparison. Shortcomings of this method include a large feature vector representation and high sensitivity to unaligned data. Well-known evolutions of this very basic global approach are Principal Component Analysis (PCA) and Independent Component Analysis (ICA).

Our work uses algorithms employing local information. One advantage of these types of algorithms is that they are capable of creating very small feature vectors that represent images. A drawback is that they do not retain any spatial information of the 2D image. Local approaches such as local binary patterns (LBPs), histogram of oriented gradients (HOGs), and Scale-invariant Feature Transform (SIFT) allow for rotation invariant texture representations of local facial features. LBPs, as introduced by Ojala et al. in [23], detect and encode texture representations like spots, line ends, edges, corners and other distinct patterns. HOGs, as shown in [24] by Dalal and Triggs, classify edge directions and other local texture features from the gradient angle and magnitude to encode a rotation invariant texture representation. SIFT, first introduced by David Lowe in [25], was originally used in more general computer vision applications but has since been shown to be useful for biometric problems. LBPs, HOGs, and SIFT have been shown in various works [2], [15], [26]–[28] to encode feature representations that perform well for matching. We tested all three approaches as follows to extract features and compare two input images.

Let I be a preprocessed near-infrared or visible-light image that is divided into N non-overlapping rectangular blocks of M pixels each. The overall feature representation of the image is given by a vector $\mathcal{T}(I) = \left\{ T^{(1)}, \ldots, T^{(N)} \right\}$, where $T^{(1)}, \ldots, T^{(N)}$ are the feature vectors computed from the N blocks. For matching two feature representations, a distance function $D\left(\overline{\mathcal{T}}(I_1), \overline{\mathcal{T}}(I_2) \right)$ is used to compare the features for two images (I_1 and I_2). The city block metric (also known as Manhattan or L1 distance), $\sum_{j=1}^{n} \left| \overline{\mathcal{T}}(I_1)_j - \overline{\mathcal{T}}(I_2)_j \right|$, was chosen to compute distance. Here, n is the length of the feature vectors being compared.

1) Local Binary Patterns: A local binary patterns is a representation 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 x_i. Then the LBP score at this location is given by

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

where $s(.)$ is the sign operator and x_j is a pixel on the circle around x_i. The occurrences of each LBP score are accumulated into a 59 bin histogram ($size(T^{(i)}) = P(P-1) + 3$, where $P = 8$). The total feature vector length of an image is the product of the number of blocks and the size of each block's histogram. See [15] for a detailed explanation of LBP within our local appearance-based approach.

2) Histogram of Oriented Gradients: We use a modified HOG algorithm for extracting features from the periocular region. The HOG algorithm begins by computing the gradient of the image. In this work we used a Prewitt convolution kernel for computing the gradient. The gradient magnitude, G_{mag}, and gradient angle, G_{ang}, were computed from the image gradient, G_x in the horizontal direction and G_y in the vertical direction. In the next step, each pixel location was used to increment orientation bins. The orientation bins represent evenly spaced segments of a circle. In this work we used 12 bins, so each bin corresponds to a 30-degree segments of the circle. For each pixel P, the orientation bin corresponding to $G_{ang}(P)$ was incremented by $G_{mag}(P)$. The feature vector length for each block was 12 elements and the feature vector length of an image was the product of the number of blocks and the number of orientation bins for each block.

3) Scale-invariant Feature Transform: In this work, we use a publically available SIFT implementation to extract features [29]. SIFT features, known as keypoints, are minima and maxima of a difference of Gaussians function applied to an image and its image pyramid (successive resampling of an image). Difference of Gaussians (DOG) is intuitively found by subtracting a wide Gaussian from a narrow Gaussian

$$DOG = \frac{1}{2\pi\sigma^2}e^{-(u^2+v^2)/(2\sigma^2)} - \frac{1}{2\pi K^2\sigma^2}e^{-(u^2+v^2)/(2K^2\sigma^2)} \tag{2}$$

for some value K. Each keypoint maintains information about its location, scale, and orientation.

4) Stratification of Thresholds: To compare the algorithmic results with those from the human participants, we used the same five recognition options ("certain, match" to "certain, nonmatch"). First, from the set of 70 matches and 70 nonmatches in a query set, we calculated distances between image pairs. Next, we defined the "certain, match" threshold (for machine recognition) to be a distance less than the minimum of all nonmatch pairs. We set the "certain, nonmatch" threshold to be a distance greater than the maximum of all match pairs. We found the threshold that gave the best performance when determining similarity, and used that value for the upper bound of the third recognition option, "don't know." The upper bound for the "likely, match" recognition was the the mean between the "certain, match" and

"Don't know" thresholds. Lastly, the "likely, nonmatch" recognition spanned values between the "don't know" and "certain, nonmatch" thresholds.

C. Machine Experiments on Iris Images

We conducted an automated matching experiment using our enhanced version of the IrisBEE iris biometrics system made available through the ICE program [30]. The software unwraps the iris region to a rectangular normalized image. Next, it applies log-Gabor filters to a grid of points on the normalized image, and creates a binary iris code using the phase of the filter results. The result of a comparison between two iris codes is reported as a fractional Hamming distance.

V. ANALYSIS

A. Periocular performance under different illuminations

In our human experiment, we found that untrained humans with limited viewing times could correctly determine whether two eye images showed the same person or different people with performance significantly higher than random guessing. The average score for the periocular tests, including both lighting conditions, was 83.6% (standard deviation 6.75%). This is a noticeable drop from one of our previous similar experiments [22], where we found that humans were 92% accurate. However, in the previous experiment, nonmatch images were randomly paired together, and viewers had unlimited viewing time for examining the images. In this experiment, we paired similar images together using information about gender, race, eye color, percent occlusion, dilation, make-up, and contacts. Even on this more challenging test, humans still were capable of verifying identity.

Looking at the two periocular tests separately, the average score on the near-infrared periocular test was 78.8% (standard deviation 5.70%), and the average score on the visible-light periocular test was 88.4% (standard deviation 3.42%). ROC curves for these tests are shown in Figure 7. We conducted a two-sample t-test to test the null hypothesis that the scores on the two modalities come from distributions of equal mean. We found statistically significant evidence to reject the null hypothesis. The difference is statistically significant at the $\alpha = 0.05$ level (p-value less than 10^{-4}). This result suggests that visible light may be a better modality for periocular recognition than near-infrared light.

Our subjects had no prior training in viewing these types of images. However, there seems to be some learning taking place as the test progresses. On the near-infrared periocular test, the average score was 77.9% on the first half of the test and 79.6% on the second half of the test. On the visible light periocular test, the average score was 86.1% on the first half of the test and 90.8% on the second half of
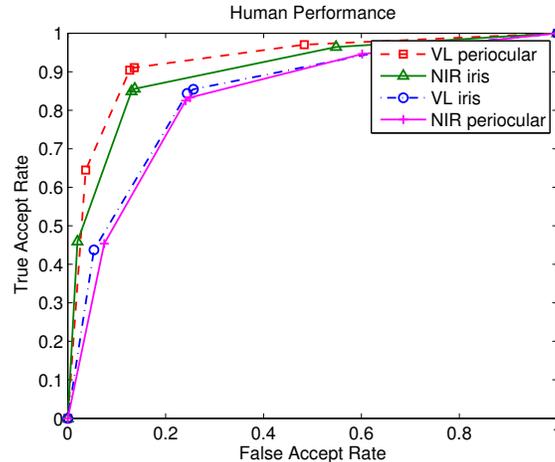
Fig. 7: This ROC curve shows the human verification performance on both the periocular and iris regions. For the periocular region, performance was higher on the visible light images. For the iris region, performance was higher on the near-infrared images.

the test. We computed one-tailed t-tests to check whether the scores on the second half were statistically significantly higher than the scores on the first half. For the near-infrared modality, the resulting p-value was 9.53×10^{-2}. Thus there is insufficient evidence to conclusively show that the subjects learned over the course of that test. However, for the visible light modality, the resulting p-value was 2.58×10^{-5}. We conclude that subjects did learn over the course of that test.

Human experiments and computer experiments using the periocular region yielded similar results. The machine performance for the periocular tests, averaged across three algorithms and both lighting conditions, was 83.6%, which is within 1% of human subject performance. The scores for the periocular visible-light experiments were 87.1% using LBP, 90.0% using HOG, and 94.3% using SIFT. For near-infrared the scores were 74.3% using LBP, 79.3% using HOG, and 76.4% using SIFT. ROC curves for these experiments are found in Figure 8. These results also suggest that visible light may be the better modality for periocular recognition.

There are some fundamental differences in features acquired under the different illuminations. First, human melanin pigment absorbs different amounts of light in visible light versus near-infrared light. The visible-light images show melanin-related differences that do not appear in near-infrared images. Second, the sclera and skin are less reflective in longer wavelengths [31]. To account for that, the gain settings on near-infrared cameras will be higher. In order to capture good iris images, the amount of near-infrared light emitted by the LG2200 iris camera sometimes causes the image of skin regions around the eye to
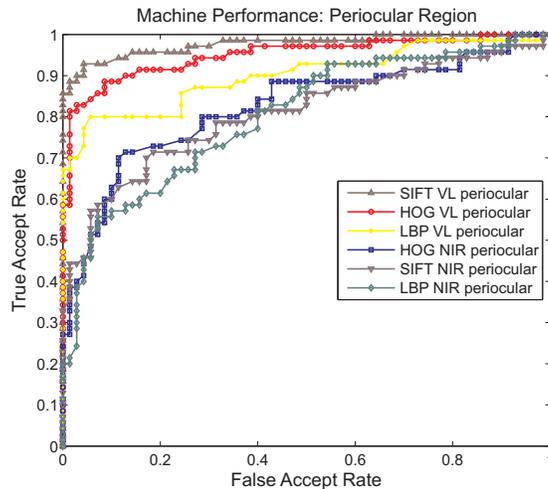
Fig. 8: This ROC curve shows machine verification performance on the periocular region. For all machine algorithms tested, performance was higher on the visible light images.

be saturated in a few of the images, thus hiding potentially useful skin feature information. It may be possible that other periocular algorithms will perform better on near-infrared images, but for humans and the machine algorithms tested, visible-light illumination is preferable.

B. Iris performance under different illuminations

In our human experiment, the average score on the near-infrared iris test was 85.6% (standard deviation 5.33%), and the average score on the visible-light iris test was 79.3% (standard deviation 5.81%). Therefore, our participants clearly found irises easier to differentiate when imaged with near-infrared illumination. This difference is statistically significant (p-value 1.0×10^{-4}). ROC curves for the human experiment iris tests are shown in Figure 7 on the same axes as the periocular ROC curves.

We compared the results from the first and second halves of the tests to see if subjects were learning during the iris tests. On the near-infrared iris test, the average score was 84.3% on the first half of the test and 86.9% on the second half of the test. This improvement was statistically significant (p-value = 1.67×10^{-2}). On the visible light iris test, the average score was 76.8% on the first half of the test and 81.9% on the second half of the test. This improvement was also statistically significant (p-value = 5.63×10^{-4}).

Computer performance on the iris queries surpassed human performance. The performance averaged across both modalities was 95.4%, which is about 13% better than human performance. The performance
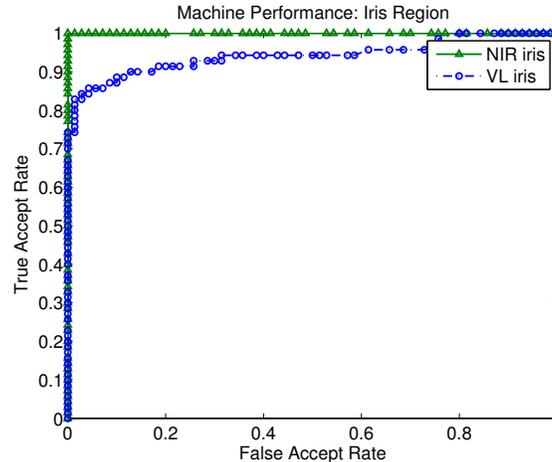
Fig. 9: This ROC curve shows machine performance on the iris region. Performance was higher on the near-infrared images.

of the machine algorithm was 100% on the near-infrared iris queries and 90.7% on the visible light iris queries. While the performance is higher, the computer results suggest the same conclusion as the human results, namely that near-infrared light is a better modality for iris recognition. ROC curves for the machine experiment are shown in Figure 9.

As mentioned in Section V-A, the absorption spectrum of human melanin pigment is different under visible and near-infrared illumination [31]. Automated iris recognition algorithms use near-infrared light to image stromal features of the iris that are hidden when imaged in visible light. From a subjective point of view, the authors found that judging pairs of dark brown irises imaged in visible light was a particularly difficult task. Thus our results agree with previous work in the literature that suggests that near-infrared light is more valuable for iris recognition than visible light [21].

C. Does performance depend on whether subjects are light-eyed or dark-eyed?

Out of the 140 subject-disjoint queries in our experiment, 50 of the queries contained subjects with dark eyes. The other 90 queries contained subjects with light eyes (blue, green, or hazel). We divided all results into sets based on whether the query showed dark eyes or light eyes, and again considered the performance.

1) Human performance on light and dark eyed subjects: The average scores for the human experiment, divided by light and dark-eyed subjects, are reported in Table II, along with standard deviations. We conducted paired t-tests to determine whether the differences between light and dark eyes were statistically

Table II: Human Performance

Average Scores on Light and Dark Eyes (standard deviation)

Modality	Average Score on Light-eyed Queries	Average Score on Dark-eyed Queries	P-value
NIR Iris	84% (5.7%)	88% (6.3%)	7.8×10^{-4}
VL Iris	84% (6.0%)	72% (8.5%)	less than 10^{-4}
NIR Peri	78% (6.5%)	81% (6.4%)	6.5×10^{-3}
VL Peri	88% (4.4%)	89% (4.1%)	7.4×10^{-1}

significant for each modality. The corresponding p-values are also given in Table II. ROC curves showing performance on visible light images are shown in Figure 10.

The largest difference between light and dark eyes occurred for the visible-light iris modality. In this case, performance on light eyes was 12% better than performance on dark eyes, indicating that dark irises are particularly difficult to match when imaged in visible light. As mentioned previously, stromal features of the iris are hidden when heavily pigmented irises are imaged in visible light.

The differences between light and dark eyes in near-infrared modalities were also statistically significant, although the differences were much smaller and actually represented only five or six query pairs in the set of 140 queries.

2) Machine performance on light and dark eyed subjects for the periocular region: In all periocular experiments, the iris was masked, so any discrepancies in results between dark-eyed and light-eyed subjects would be due to features correlated with eye color (e.g. ethnicity), but not directly due to eye color itself.

Table III gives the scores for the periocular machine recognition experiments divided by light-eyed and dark-eyed subjects. ROC curves for the experiments using visible-light images are found in Figure 11. The table and ROC curves show that the machine algorithms achieved better performance on the dark-eyed subjects than the light-eyed subjects. However, many of the differences are small (three to five percent on 140 queries). Further experimentation on a larger set of images would reveal whether this trend is a general phenomenon or unique to this set of subjects.

3) Machine performance on light and dark eyed subjects for the iris region: Table IV gives the scores for the iris machine recognition experiments divided by light-eyed and dark-eyed subjects. Figure 12 shows the ROC curves for the visible light images, separated by light and dark-eyed subjects. These results present the same conclusion as the analysis of the human experiment; namely that in visible light,
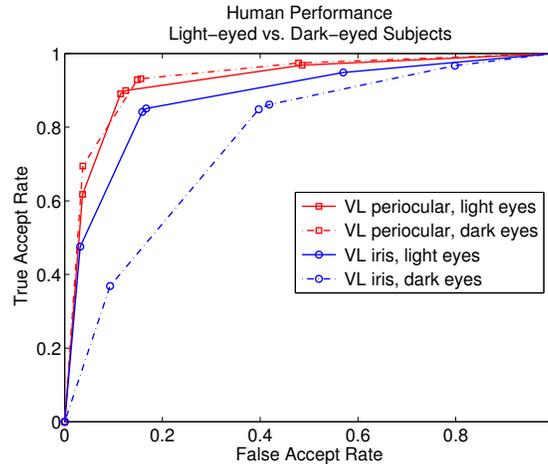
Fig. 10: Dark irises are particularly difficult for humans to match when imaged in visible light. The performance on light irises was 12% better than performance on dark irises for visible light iris images. In all other modalities (NIR iris, VL periocular, and NIR periocular), humans found dark eyes were easier to compare, although the difference was not always statistically significant.

Table III: Machine Performance: Periocular Region

Scores on Light and Dark Eyes

Modalities	Score on Light-eyed Queries	Score on Dark-eyed Queries
HOG NIR Peri	77%	86%
LBP NIR Peri	73%	84%
SIFT NIR Peri	76%	80%
HOG VL Peri	89%	94%
LBP VL Peri	86%	90%
SIFT VL Peri	93%	96%

dark-eyed subjects are harder to recognize.

In summary, for visible-light iris images, dark eyes are difficult to recognize, whether by humans or by machine algorithms. The other modalities showed some differences between light and dark-eyed subjects, but the differences were not as large, and we suggest that larger data sets would be necessary to determine whether any such trends are true of the general population.
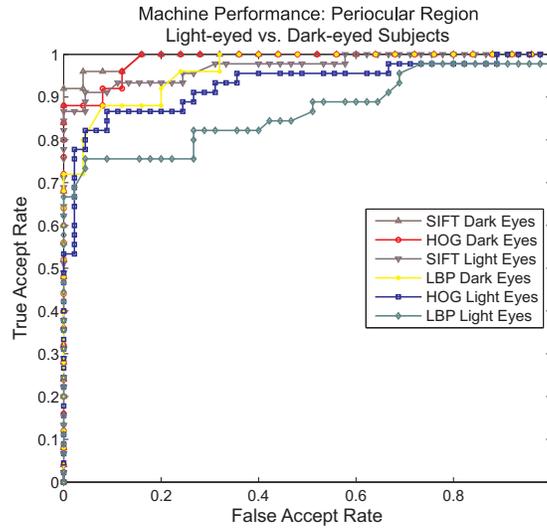
Fig. 11: For this data set, machine recognition finds that comparing periocular regions from subjects with dark eyes is easier than comparing periocular regions from subjects with light eyes. However, many of the differences are small, and we do not have sufficient information to know whether this trend would extend to other data sets.

Table IV: Machine Performance: Iris Region
Scores on Light and Dark Eyes

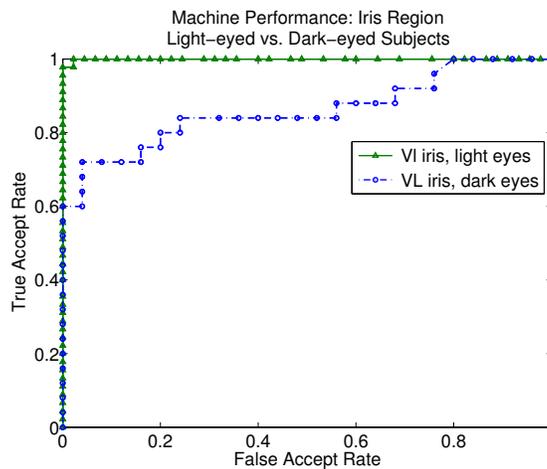| Modalities	Score on Light-eyed Queries	Score on Dark-eyed Queries
NIR Iris	100%	100%
Vl Iris	99%	84%

Fig. 12: Dark irises are difficult for machines to recognize when imaged in visible light. This result parallels results from both our human experiment and prior research [21].

Fig. 13: There was one visible-light periocular query misclassified by all three machine algorithms. This query was the fourth hardest query for human volunteers; only 64% (18 of 28) of human participants responded correctly to this query. These are images 05337d65 and 05337d66.

D. Difficult image pairs

1) Difficult periocular queries: There was one visible-light periocular match query misclassified by all three machine algorithms, LBP, HOG, and SIFT. This query was the fourth hardest query for human volunteers. It was classified correctly by only 64% of participants. This query is shown in Figure 13.

There were eight near-infrared periocular match queries misclassified by all the machine algorithms. Three of these were also extra difficult for humans, with less than 60% of humans classifying them correctly. One of these eight was the second-most misclassified by humans in this set (Figure 14); only 36% of participants responded correctly for that query. On the other hand, one of the eight queries difficult for machines was classified correctly by 93% of human volunteers (Figure 15). In Figure 14, one image of the query pair showed the eye more open than the other image. We speculate that in general, a difference in openness of the eye would be misleading to both humans and machines. An interesting feature to note in Figure 15 is that the eyebrow is visible in one of the two images but not in the other. We suspect that in general, humans are more able to deal with this type of discrepancy than the current machine algorithms. A possible improvement for machine algorithms would be the ability to transform an image to account for changes in expression such as raised eyebrows.

All of the nonmatch queries were correctly classified as not matching by at least one of the three machine algorithms, LBP, HOG, or SIFT.

2) Difficult iris queries: There were thirteen visible-light iris match queries misclassified by the machine algorithm. Less than fifty percent of humans responded correctly for one of these queries (Figure 16). On the other hand, over ninety percent of humans responded correctly for four of the queries. There was one visible-light iris nonmatch query misclassified by the machine algorithm. However, it was relatively easy for human participants; 93% of participants correctly classified this pair as a nonmatch.

Fig. 14: This query was difficult for both machine algorithms and for humans. Only 36% (10 of 28) of participants correctly labeled this pair as a match. These are images 05525d37 and 05525d53.
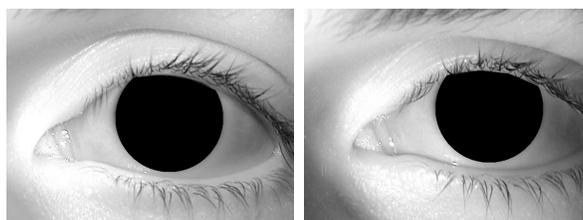

Fig. 15: This query was difficult for machine algorithms but relatively easy for humans. Ninety-three percent of human volunteers (26 of 28) correctly determined that these two images were from the same person. These are images 05341d166 and 05341d107.

Fig. 16: This query was challenging for both humans and machines. Only 46% of humans (13 of 28) correctly determined that these images were from the same person. Previous research has shown that a difference in pupil dilation can degrade iris recognition performance [32]. The difference in dilation in this query may have misled both humans and machines. These are images 04233d1615 and 04233d1607.

All near-infrared iris queries were classified correctly by the machine algorithm.

In general, we believe that some characteristics, such as difference in pupil dilation, cause difficulties for both humans and machines. Other characteristics affect either humans or machines. For instance, any color information in the visible-light iris images would benefit human participants but would be unused by the machine algorithm which relies solely on iris texture.
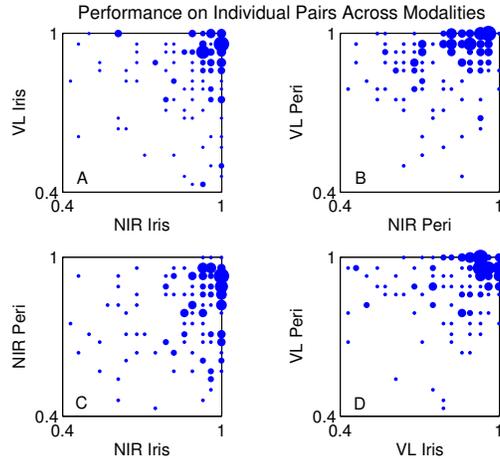
Fig. 17: In the human experiments, there was little evidence of correlation for pairs of subjects across different modalities. This result suggests that a "difficult" subject to match in one modality may be easier in other modalities.

E. Are difficult image pairs in one modality also difficult in other modalities?

We wanted to test whether queries that were difficult in one modality were also difficult in other modalities. That is, for a match pair, was the imaged subject difficult to identify as a "match" in both near-infrared and visible light? For a nonmatch pair, were the two imaged subjects difficult to differentiate in both near-infrared and visible light? Similarly, did performance on iris images correlate with performance on periocular images?

For each pair of images, we computed the percent of times the pair was correctly classified in our human experiment. To compare accuracy on near-infrared irises with accuracy on visible-light irises, we made a scatter plot of the corresponding classification rates for each image pair. Figure 17a shows the result, with near-infrared accuracy on the x-axis, and visible-light accuracy on the y-axis. The points are clustered slightly more on the right of the graph than on the top, reflecting the fact that performance on near-infrared irises was better than performance on visible-light irises. However, there is no evidence of correlation across the modalities. In fact, the correlation coefficient comparing the classification rates of the two modalities is 0.070, which suggests almost no correlation at all.

We made similar scatterplots comparing near-infrared periocular performance with visible-light periocular performance (Figure 17b), near-infrared iris with near-infrared periocular performance (Figure 17c), and visible-light iris with visible-light periocular performance (Figure 17d). The correlation coefficients are given in Table V.

Table V: Human Experiment
Correlation Coefficients Across Modalities

Comparison	Correlation Coefficient
NIR Iris vs. VL Iris	0.07
NIR Peri vs. VL Peri	0.33
NIR Iris vs. NIR Peri	0.21
VL Iris vs. VL Peri	0.19

The cross-modality comparison with the largest correlation coefficient was the comparison between near-infrared periocular and visible-light periocular performance, which yielded a correlation coefficient of 0.33. In contrast, the correlation between near-infrared iris images and visible-light iris images was only 0.07. This discrepancy implies that the periocular region is more similar under the two different types of wavelengths than the iris region is.

The periocular region showed some similarity when imaged under two different types of light; however, a correlation coefficient of 0.33 is still far below a perfect correlation of 1.0. This result suggests that no one individual is inherently difficult to verify, and no two individuals are inherently difficult to differentiate. Prior research agrees with this conclusion. Yager and Dunstone [33] have said that "evidence is mounting that there are few people who are destined to have difficulty with biometric authentication due to an inherent 'unmatchability,'" and that "when errors are common for a particular individual, they can usually be addressed using improved enrollment/capture processes and robust matching algorithms." Furthermore, these results offer good news to designers of bi-modal systems, because they may be able to compensate for poor scores in one modality by considering images in another modality.

In the human experiment, we had multiple participants evaluating each image pair. Thus it was possible to consider two illuminations and compare the average number of times a query pair was classified correctly under each illumination. We used these percentages to make a judgment on whether a difficult image pair in near-infrared light was also difficult in visible light. A machine algorithm provides a single similarity score instead of a set of judgments, so we employed a different approach to answer this question for the machine algorithms.

Using the periocular machine algorithms, we computed the rank score of each pair. Then we computed the correlation coefficient of ranks of the near-infrared images with the visible light images. The correlation coefficients are given in Table VI. The correlation coefficients for these algorithms are higher than

Table VI: Periocular Machine Algorithms

Correlation Coefficients of Ranks Across Modalities

Comparison	Correlation Coefficient
NIR Peri vs. VL Peri using LBP	0.55
NIR Peri vs. VL Peri using HOG	0.60
NIR Peri vs. VL Peri using SIFT	0.53

the correlations in the human experiment. A test performing rank-level fusion across both illuminations and three feature representations (LBP, HOG and SIFT) yields a recognition rate of 92.1%, 2% less than the best single modality. This suggests that for the algorithmic matching of the periocular region, difficult images are difficult across modalities.

One explanation of this phenomenon relates to the feature extraction strategies used. The machine algorithms are using the same feature extraction techniques for both near-infrared and visible light images. In contrast, humans tend to look at different features in different modalities. Since the periocular machine algorithms employ the same feature extraction techniques for analyzing both near-infrared periocular images and visible-light periocular images, the difficulty of analyzing a given subject pair is similar, even across the two modalities.

A similar analysis of the iris machine algorithms is not meaningful because the iris algorithm is achieving 100% correct recognition on the near-infrared images.

F. Which features are useful in near-infrared light?

Different wavelengths of light can reveal different features of the object being imaged. For example, iris recognition uses near-infrared light to reveal iris texture that may not be seen in visible light. We anticipated that features in the periocular region would have differing levels of importance when imaged under different lighting conditions.

As mentioned in section IV, we asked volunteers to rate the helpfulness of different features in the image. Each feature could be rated as "very helpful", "helpful", or "not helpful". At the conclusion of the experiment, we tallied the results, and sorted the features based on the number of times the volunteers rated the features as "very helpful". Figure 18 shows how the features were ranked for near-infrared periocular images. The ranking from most helpful to least helpful was (1) eyelashes, (2) eyelid, (3) tear duct, (4) eye shape, (5) eyebrow, (6) outer corner, (7) skin, (8) color, (9) blood vessels, and (10) other.

We conducted one previous experiment which likewise presented pairs of near-infrared images to volunteers and asked for feedback about helpful features [22]. While this prior experiment paired images randomly rather than pairing images with similar metadata, it is still interesting to compare the results. In both the current and the previous experiments, eyelashes were the most helpful feature considered. Since eyelashes have twice been rated highly on two separate human-viewer experiments, we can confidently assert that humans use eyelashes extensively for this type of recognition task. The next three features – eyelid, tear duct, and eye shape – were not ranked in exactly the same order, but were still ranked in the top four helpful features. Eyebrow, outer corner of the eye, and skin were rated fifth, sixth, and seventh respectively in both this experiment and in the previous experiment. The low rank for eyebrows is likely due to the fact that our images did not include the whole eyebrow, and some images did not show any eyebrow at all. Research on images showing a larger periocular region has shown that eyebrows are especially useful for recognition [2], [8]. If we conducted an experiment using a larger periocular region, we expect that eyebrows would be ranked higher in this list of features.

"Color" and "blood vessels" were included in the list of features for this experiment because we expected those features to be useful for the visible-light modality and we wanted to use the same features list in both conditions. Unsurprisingly, these two features were not considered very helpful for the near-infrared images in this experiment.

Users who wished to mention additional features could enter responses in a textbox. The only other features entered were "mascara", "freckles", and "shape of blacked-out iris". Our previous experiment required testers to rate features after every image pair instead of once at the end of a set of pairs, and thus there were a larger variety of inputs. Other useful features mentioned in that experiment were clusters of eyelashes, single "stray" eyelashes, eyelash density, eyelash direction, eyelash length, eyelash intensity, unusual eye shape, slant of eyes, amount the eye was open, contacts, and make-up.

G. Which features are useful in visible light?

Figure 19 shows how testers ranked features in the visible-light periocular images. The ranking from most helpful to least helpful was (1) blood vessels, (2) skin, (3) eye shape, (4) eyelashes, (5) eyelid, (6) eyebrow, (7) tear duct, (8) color, (9) other, and (10) outer corner. Notice that the contrast between feature rankings is not as large for the visible-light modality as in the near-infrared modality. In the near-infrared images, 22 of the 28 subjects rated eyelashes as "very helpful" while only 15 thought the next feature – eyelid – was "very helpful". In contrast, for visible-light images, blood vessels, skin, and eye shape all received 15 or 16 "very helpful" votes.
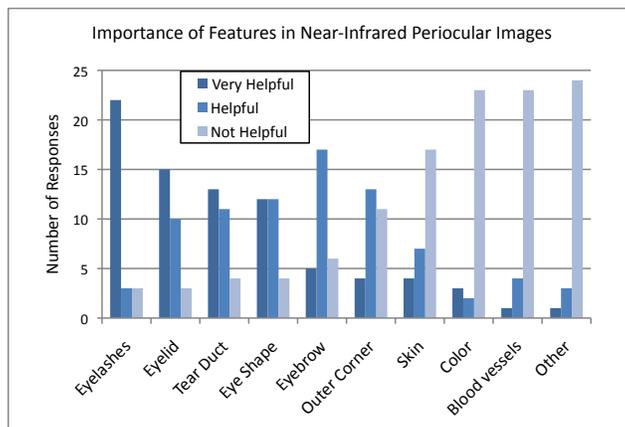
Fig. 18: Volunteers ranked the features that they found most useful in determining whether two images came from the same person. Eyelashes were considered the most useful feature in the near-infrared periocular images. Eyelids, tear duct, and eye shape were also valuable.
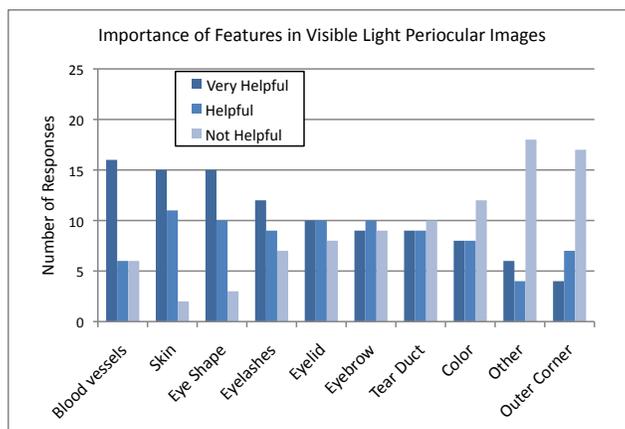

Fig. 19: Blood vessels and skin, which were not useful in the near-infrared images, were particularly useful in images taken with visible light.

Blood vessels in the sclera are clearly more apparent in the visible-light images than in the near-infrared images. Also, discriminating features of the skin were more visible to our human volunteers when imaged in visible light. Note that it may be possible to design a near-infrared camera that reveals more discriminating skin information than the near-infrared cameras we used for this experiment, which were designed for iris imaging. Color ranked lower than we expected, but this may be because the images in this experiment were specifically paired to match similar skin colors together.

VI. CONCLUSION

We analyzed human verification performance on periocular and iris images under two different lighting conditions. We found that participants scored higher on visible-light periocular images than near-infrared periocular images. Specifically, average performance on visible-light periocular images was 88.4% while average performance on near-infrared periocular images was 78.8%. The average score on the near-infrared iris test was 85.6%, and the average score on the visible-light iris test was 79.3%. We anticipate that scores on all modalities could be improved if participants were trained, or allowed longer time to view queries. Machine performance on visible-light periocular images was 94.3% using scale-invariant feature transform (SIFT), and on near-infrared periocular images it was 76.4%. Both the human and machine results suggest that visible light may be a better modality for periocular recognition than near-infrared light, at least for the type of data used in this paper.

We compared performance on light eyes and dark eyes for all images. In all periocular experiments, the iris was masked, so any discrepancies in results would be due to features correlated with eye color, but not directly due to eye color itself. For both human and machine periocular tests, performance on dark-eyed subjects was better than performance on light-eyed subjects. Further research would have to be conducted to see whether this pattern extends to other data sets. We compared performance on light eyes and dark eyes for irises as well. For human performance on the visible-light iris modality, the average score on queries showing light eyes was 12% higher than the average score on queries showing dark eyes. This comparison emphasizes the value of using near-infrared light for iris recognition on dark irises.

We investigated whether a "difficult" subject in one modality was also difficult in other modalities. In the human experiments, we considered the percent of times each pair was classified correctly, and then computed correlation coefficients for these performances across modalities. The correlation coefficient for iris images across modalities was only 0.07, while the correlation coefficient for periocular images across modalities was a higher value of 0.33. In the machine experiments, we used rank-level fusion to combine information across modalities for the periocular images, but failed to see any performance improvement; thus this analysis suggests high correlation across illuminations for the periocular region. We draw two conclusions from these results. First, the periocular region seems more similar when imaged under the two wavelengths of light than the iris region. Second, machine performance was more highly correlated across the two modalities than human performance, probably because the machine algorithms used identical feature extraction methods for both lighting conditions while humans could choose which features they found most discriminating. Future machine algorithms might be improved by tuning the

feature extraction techniques separately for the two lighting conditions.

In near-infrared light, humans found eyelashes, eyelid, tear duct, and eye shape to be the most useful periocular features. In contrast, when considering visible light, humans found blood vessels and skin to be slightly more helpful than eye shape and eyelashes. This result again suggests that computer algorithms for periocular recognition could benefit from using different feature extraction techniques for the two different lighting conditions.

Based on all of our experiments, we recommend that future research in periocular biometrics should focus on using visible-light periocular images instead of near-infrared images in order to be most successful. To make improvements in near-infrared periocular biometrics, researchers will be required to investigate the use of other periocular feature types in addition to texture based features. Shaped based features are a possible option which could be combined with texture based features to improve overall system performance.

ACKNOWLEDGMENT

REFERENCES

[1] U. Park, R. R. Jillela, A. Ross, and A. K. Jain, "Periocular biometrics in the visible spectrum," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 1, pp. 96–106, Mar 2011.

[2] U. Park, A. Ross, and A. K. Jain, "Periocular biometrics in the visible spectrum: A feasibility study," in *IEEE Int. Conf. on Biometrics: Theory, Applications, and Systems*, Sept 2009, pp. 1–6.

[3] D. L. Woodard, S. J. Pundlik, J. R. Lyle, and P. E. Miller, "Periocular region appearance cues for biometric identification," in *IEEE Computer Vision and Pattern Recognition Biometrics Workshop*, Jun 2010, pp. 162–169.

[4] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, 2003.

[5] K. W. Bowyer, K. P. Hollingsworth, and P. J. Flynn, "Image understanding for iris biometrics: A survey," *Computer Vision and Image Understanding*, vol. 110, no. 2, pp. 281–307, 2008.

[6] P. E. Miller, J. R. Lyle, S. J. Pundlik, and D. L. Woodard, "Performance evaluation of local appearance based periocular recognition," in *IEEE Int. Conf. on Biometrics: Theory, Applications, and Systems*, Sept 2010, pp. 1–6.

[7] A. Calder and G. Rhodes, Eds., *Handbook of Face Perception*. Oxford University Press, 2010, ch. Cognitive and Computational Approaches to Face Perception by O'Toole, in press.

[8] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell, "Face recognition by humans: Nineteen results all computer vision researchers should know about," *Proceedings of the IEEE*, vol. 94, no. 11, pp. 1948–1962, Nov 2006.

[9] R. Abiantun and M. Savvides, "Tear-duct detector for identifying left versus right iris images," in *IEEE Applied Imagery Pattern Recognition Workshop*, 2008, pp. 1–4.

[10] S. Bhat and M. Savvides, "Evaluating active shape models for eye-shape classification," in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, 2008, pp. 5228–5231.

[11] Y. Li, M. Savvides, and T. Chen, "Investigating useful and distinguishing features around the eyelash region," in *IEEE Applied Imagery Pattern Recognition Workshop*, 2008, pp. 1–6.

[12] J. Merkow, B. Jou, and M. Savvides, "An exploration of gender identification using only the periocular region," in *IEEE Int. Conf. on Biometrics: Theory, Applications, and Systems*, Sept 2010, pp. 1–5.

[13] J. R. Lyle, P. E. Miller, S. J. Pundlik, and D. L. Woodard, "Soft biometric classification using periocular region features," in *IEEE Int. Conf. on Biometrics: Theory, Applications, and Systems*, Sept 2010, pp. 1–7.

[14] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the Face Recognition Grand Challenge," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2005, pp. 1–8.

[15] P. Miller, A. Rawls, S. Pundlik, and D. Woodard, "Personal identification using periocular skin texture," in *ACM Symposium on Applied Computing*, 2010, pp. 1496–1500.

[16] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, Oct 2000.

[17] J. Adams, D. L. Woodard, G. Dozier, P. Miller, K. Bryant, and G. Glenn, "Genetic-based type II feature extraction for periocular biometric recognition: Less is more," in *IEEE Int. Conf. on Pattern Recognition*, 2010, pp. 205–208.

[18] J. Xu, M. Cha, J. L. Heyman, S. Venugopalan, R. Abiantun, and M. Savvides, "Robust local binary pattern feature sets for periocular biometric identification," in *IEEE Int. Conf. on Biometrics: Theory, Applications, and Systems*, Sept 2010, pp. 1–8.

[19] S. Bharadwaj, H. S. Bhatt, M. Vatsa, and R. Singh, "Periocular biometrics: When iris recognition fails," in *IEEE Int. Conf. on Biometrics: Theory, Applications, and Systems*, Sept 2010, pp. 1–6.

[20] D. L. Woodard, S. Pundlik, P. Miller, R. Jillela, and A. Ross, "On the fusion of periocular and iris biometrics in non-ideal imagery," in *IEEE Int. Conf. on Pattern Recognition*, 2010, pp. 201–204.

[21] J. Daugman, "How iris recognition works," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 21–30, 2004.

[22] K. Hollingsworth, K. W. Bowyer, and P. J. Flynn, "Identifying useful features for recognition in near-infrared periocular images," in *IEEE Int. Conf. on Biometrics: Theory, Applications, and Systems*, Sept 2010, pp. 1–8.

[23] 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*, vol. 24, no. 7, pp. 971–987, 2002.

[24] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2005, pp. 886–893.

[25] D. G. Lowe, "Object recognition from local scale-invariant features," in *The Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999*, 1999.

[26] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.

[27] S. Z. Li, R. Chu, M. Ao, L. Zhang, and R. He, "Highly accurate and fast face recognition using near infrared images," *Proc. Advances in Biometrics, Int. Conf. (ICB), Lecture Notes in Computer Science 3832, Springer*, pp. 151–158, 2006.

[28] Z. Sun, T. Tan, and X. Qiu, "Graph matching iris image blocks with local binary pattern," *Proc. Advances in Biometrics, Int. Conf. (ICB), Lecture Notes in Computer Science 3832, Springer*, pp. 366–373, 2006.

[29] VLFeat, 2011, http://www.vlfeat.org/index.html.

[30] P. J. Phillips, W. T. Scruggs, A. J. O'Toole, P. J. Flynn, K. Bowyer, C. L. Schott, and M. Sharpe, "FRVT 2006 and ICE 2006 large-scale experimental results," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 831–846, May 2010.

[31] J. Daugman, "Absorption spectrum of melanin," http://www.cl.cam.ac.uk/~jgd1000/melanin.html, accessed July 2009.

[32] K. P. Hollingsworth, K. W. Bowyer, and P. J. Flynn, "Pupil dilation degrades iris biometric performance," *Computer Vision and Image Understanding*, vol. 113, no. 1, pp. 150–157, 2009.

[33] N. Yager and T. Dunstone, "The biometric menagerie," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 2, pp. 220–230, 2010.

Karen P. Hollingsworth received B.S. degrees in computational math and math education from Utah State University in 2004. She completed graduate school at the University of Notre Dame in 2010 with M.S. and Ph.D. degrees in computer science. She worked with advisors Dr. Kevin Bowyer and Dr. Patrick Flynn in investigating performance improvements to iris biometrics.

Dr. Hollingsworth is an Intelligence Community Postdoctoral Fellow and a Research Assistant Professor at the University of Notre Dame. She is currently studying periocular biometrics.

Shelby Solomon Darnell is a Ph.D. student of computer science in the Human Centered Computing Division at Clemson University. Shelby S. Darnell became a member of IEEE in 2008. He obtained a computer engineering bachelors in 2003 and software engineering masters in 2006, both from Auburn University in Auburn, Alabama USA.

He has had several research assistant and teaching assistantships in graduate school. He is currently a research assistant for Juan E. Gilbert, Ph.D. for an NSF funded project. He has worked for a defense contractor, L3 Communications Mission Integration Systems, in Greenville, TX for more than three years with the responsibilities of a Software and Junior Project Engineer. His research interests include Biometrics, Human Centered Computing, Computational Intelligence and Image Processing.

Philip E. Miller received his M.S. degree in computer science from Clemson University in 2010. He is currently a Ph.D. candidate in the School of Computing at Clemson University. This author has been a member of IEEE since 2009.

Damon L. Woodard (M' 00, SM' 2011) received a B.S. in computer science and computer information systems from Tulane University in 1997, a M.E. in computer science and engineering from Penn State University in 1999, and a Ph.D. in computer science and engineering from the University of Notre Dame in 2005.

From 2004 - 2006, he was an Intelligence Community (IC) Postdoctoral Fellow. In 2006, Dr. Woodard joined the faculty of Clemson University and is currently an Assistant Professor within the Human-Centered Computing (HCC) Division in the School of Computing. His research interests include biometrics, pattern recognition, and computational intelligence.

Dr. Woodard is a member of the Association for Computing Machinery and the National Society of Black Engineers.

Kevin W. Bowyer currently serves as Schubmehl-Prein Professor and Chair of the Department of Computer Science and Engineering at the University of Notre Dame. His broad research interests are in computer vision, pattern recognition and data mining. Particular contributions in biometrics include algorithms for improved accuracy in iris biometrics, face recognition using three-dimensional shape, 2D and 3D ear biometrics, advances in multi-modal biometrics, and support of the government's Face Recognition Grand Challenge, Iris Challenge Evaluation, Face Recognition Vendor Test 2006 and Multiple Biometric Grand Challenge programs. His paper "Face Recognition Technology: Security Versus Privacy," published in IEEE Technology and Society, was recognized with an "Award of Excellence" from the Society for Technical Communication in 2005. Professor Bowyer served as General Chair of the IEEE International Conference on Biometrics Theory, Applications and Systems in 2007, 2008 and 2009, and as General Chair of the International Joint Conference on Biometrics in 2011.

Professor Bowyer earned his Ph.D. in computer science from Duke University. While on the faculty at the University of South Florida, he won three teaching awards, received a Distinguished Faculty Award for his mentoring work with under-represented students in the McNair Scholars Program, and was awarded a sequence of five National Science Foundation site grants for Research Experiences for Undergraduates. Also during this time, Professor Bowyer served as Editor-In-Chief of the IEEE Transactions on Pattern Analysis and Machine Intelligence, recognized as the premier journal in its areas of coverage, and was elected as an IEEE Fellow for his research in object recognition.

Patrick J. Flynn is Professor of Computer Science & Engineering and Concurrent Professor of Electrical Engineering at the University of Notre Dame. He received the B.S. in Electrical Engineering (1985), the M.S. in Computer Science (1986), and the Ph.D. in Computer Science (1990) from Michigan State University, East Lansing. He has also held faculty positions at Washington State University and Ohio State University.

His research interests include computer vision, biometrics, and image processing. Dr. Flynn is a Senior Member of IEEE, a Fellow of IAPR, an Associate Editor of IEEE Trans. on Information Forensics and Security and IEEE Trans. on Image Processing. He is a past Associate Editor and Associate Editor-in-Chief of IEEE Trans. on PAMI, and a past associate editor of Pattern Recognition and Pattern Recognition Letters.