Empirical Evidence for Iris Match Score Degradation with Time Lapse in ICE 2006

Sarah E. Baker,

Kevin W. Bowyer, *Fellow, IEEE,*

Patrick J. Flynn, *Senior Member, IEEE*,

and P. Jonathon Phillips, *Senior Member, IEEE*

sbaker3, kwb, flynn@cse.nd.edu; jonathon@nist.gov

SEB, KWB and PJF are with the University of Notre Dame. PJP is with the National Institute of Standards and Technology.

Abstract

It is widely assumed that, barring traumatic injury to the eye or eye surgery, a person's iris does not change over time. This implies that iris recognition performance does not degrade as time increases after initial enrollment- also known as the template-aging problem. We present results of the first study to investigate the validity of this assumption. We explore the effects of time lapse since enrollment on iris biometric match scores using a data set with four years time lapse between the earliest and most recent images of an iris (23 subjects, 46 irises, 6,814 total images). Experimental results are reported for three iris recognition algorithms. We find that the empirical mean of the distribution for similarity scores between iris images of the same person changes over four years–in biometric parlance, the match or genuine distribution changes. The change is statistically significant when comparing the mean of pairs of images taken within 120 days ($\sim$4 months) and over 1200 days ($\sim$3.25 years) for the three algorithms in our study. The mean of the match distribution changes so that the expected performance would degrade due to an increase in the false reject rate. Our results suggest that iris biometric templates undergo aging.

Index Terms

iris biometrics, enrollment template, template aging, time-lapse, match distribution stability

I. Introduction

The iris biometrics research community has broadly accepted the premise that the appearance of the iris is highly stable throughout most of a person's life. Daugman stated the assumption this way–"As an internal (yet externally visible) organ of the eye, the iris is well protected and stable over time" [1]. The assumption is repeated in similar form in numerous academic references: "[the iris is] stable over an individual's lifetime" [4], "the iris is highly stable over a person's lifetime" [6], "[the iris is] essentially stable over a lifetime" [5]. Descriptions in commercial literature are even more bold: "only a single enrollment in a lifetime" [3]. With the exception of a preliminary paper by the authors [20], we are not aware of any studies that test the validity of this assertion.

In a typical verification application, a person has an iris image acquired at the time of enrollment into a biometric system. At a later point in time, the person will present their iris to the system along with a claimed identity. A biometric system will verify the identity claim by comparing the enrolled iris image with the new iris image. If the two irises are determined to be sufficiently similar, the claimed identity will be accepted.

The heart of an iris recognition system is a matching algorithm that compares two iris images. After comparing two iris images, a matching algorithm produces a similarity score. A similarity score measures how likely both iris images are from the same person. A similarity score between two irises of the same person is called a match score and a similarity score between two irises of different people is called a non-match score. If the "stable over a lifetime" assertion is true, then the distribution of match scores from pairs of iris images with longer time lapse should not be different from that of image pairs with shorter time lapse.

We investigate the effect of lapsed time on the performance of three algorithms: our modification of the irisBEE baseline matcher [12], Neurotechnology's VeriEye [10], and the Cam-2 submission to the Iris Challenge Evaluation 2006 from the University of Cambridge [9]. For the three algorithms, the effect of lapsed time is measured by the difference in the mean match score of iris image pairs taken with at most 120 days of time lapse and with at least 1200 days of time lapse. For all three algorithms, the differences in the means of the match scores from the longer time lapse and shorter time lapse groups are statistically significant. The movement of the means in longer time-lapse is towards the mean of the non-match distribution. The results show that the match distribution is not stable over time and suggests that increased time since enrollment results in an increased false reject rate.

There are potentially other factors that could produce the drift in the means between the short and long time lapsed iris pairs. We have either controlled for a number of factors that could influence performance or specifically tested for the influence of factors on performance. We conclude that none of the factors considered are able to account for our observed results. The factors taken into account are the presence of contact lenses, changes in pupil dilation, changes in the amount of the iris occluded, changes in the acquisition characteristics of the iris sensor due to aging, and differences in the light-emitting diode illuminating the iris.

II. Previous and Related Work

This paper expands upon our initial results in several ways [20]. First, we have increased the number of subjects from 13 to 23 and the number of unique irises from 26 to 46. Second, in the previous paper we only considered images from spring 2004 and spring 2008 and the matches within one semester and matches across the four years. In this work we now consider all images acquired from 2004 through 2008 and have set two time thresholds in defining our short-time-lapse and long-time-lapse matches. Third, we present an additional analysis with subsets of each set of time-lapse match scores so that each image contributes to only one match score in the short-time-lapse matches and one match score in the long-time-lapse matches. Finally, we have tested the time-lapse effect on two additional algorithms: Neurotechnology's VeriEye [10] and the Cam-2 submission to the Iris Challenge Evaluation 2006 from the University of Cambridge [9]. We also describe additional testing done to eliminate other possible causes of match score degradation.

Gonzalez et al., [13] report an effect of time separation on iris recognition that may initially seem similar to our previous and current results. However, Gonzalez et al., is based on comparing matches between images acquired at the same acquisition session with those acquired with at most three months time lapse. They report a better match statistic for images from the same session than for those across sessions. However, they show little change in match statistics when comparing matches with short time lapses, between two weeks and three months. In our results presented in this paper, we do not consider matches between images acquired in the same acquisition session, as we expect they would not be representative of a real-world biometric scenario. Like Gonzalez et al. we do not find any significant difference in match scores for images with a few months time lapse. However, looking at a longer time lapse than that examined in Gonzalez et al. we do observe a statistically significant degradation in match scores.

III. Time-lapse Experiment

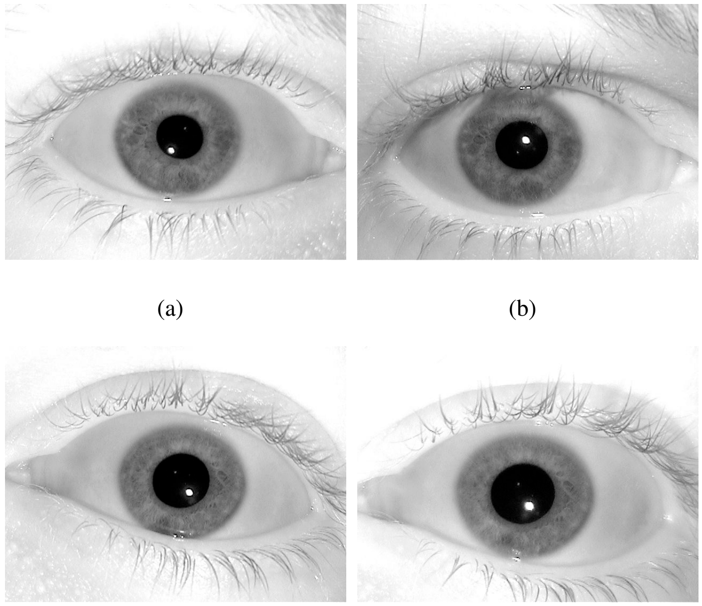
This section describes the iris image dataset, the iris recognition algorithms, the method of analyzing results, and results of the time-lapse experiment.

A. Image Dataset and Algorithms

The image acquisition protocol was the same as that used for acquiring images used in the Iris Challenge Evaluations 2005 and 2006 (ICE) [9], [12]. All of the iris images were acquired with the same LG 2200 iris imaging system with no hardware or software modifications during the four years of image acquisition [2]. Also, image acquisition was performed in the same studio throughout the four years, with the same ambient indoor lighting.

Image acquisition sessions were held at multiple times in each academic semester across the four years. At a given acquisition session, for a given subject, six images were acquired of each eye. As described in the ICE 2005 and 2006 image acquisition protocol, some of the images acquired may not pass the normal built-in quality control checks of the LG 2200 [9], [12]. We visually inspected all images considered for use in this study, and rejected any of noticeably low quality, e.g., out of focus irises and significant portions of the iris were occluded. Also excluded from the study were images for which the baseline irisBEE algorithm produced a noticeably poor segmentation of the iris region.

A total of 23 persons participated in data acquisitions from 2004 through 2008; see Figure 1 for examples of iris images. There are images from both irises of the 23 subjects over the four year period. Subject age ranges from 22 to 56 years old at the end of the four-year period. Sixteen subjects are male and seven are female. Sixteen of the subjects are Caucasian and seven Asian. The repeated sixteen by seven breakdown is a coincidence; the ethnicity division does not follow the gender division. None of the subjects wore glasses for any of the data acquisition. Five subjects wore contact lenses at all acquisition sessions, and eighteen subjects did not wear contact lenses at any acquisition session. The total number of iris images in this study was 6,814.



(a) (b)

(c) (d)

Fig. 1. Example iris images of a subject taken in 2004 and 2008 (subject 04233). (a) right iris from 2004, (b) right iris from 2008, (c) left iris from 2004, and (d) left iris from 2008.

To guard against the possibility of observed time-lapse effects being specific to one particular system, three different iris matching algorithms were included in the study. First, we used our own modified version of the publicly available irisBEE system [9]. This system represents an iris as a 240x20x2-bit iris code generated from the complex-valued responses of one-dimensional log-Gabor wavelets [7]. For the irisBEE matcher, the output of matching two iris images is a (fractional) Hamming distance. In iris recognition, the range of a Hamming distance is $[0, 1]$ with 0 being a perfect match. Second, we used the commercial VeriEye Iris SDK from NeuroTechnology [10]. This system produces match scores on a different scale and with a different polarity than systems based on a Daugman-like iris code. For the analysis in this paper, we flipped the polarity of the match scores. The third system was the Cam-2 submission to the ICE 2006 from the University of Cambridge [9]. The output of the Cam-2 matcher is nominally a Hamming distance.

B. Experimental Method

In the analysis, we consider all 46 irises as independent and perform analysis by comparing the change in the match score distribution over the four year time period. For each iris, we created two sets of iris image pairs. The first set is the short-time lapse set. This set consists of iris image pairs where both iris images were acquired within 120 days. The second set is the long-time lapse set. This set consists of iris image pairs where the iris images were acquired at least 1,200 days apart. An iris image can be in multiple short-time lapse pairs and multiple long-time lapse pairs. Formally, let $S(i)$ be the set of short-time iris pairs for iris i and $L(i)$ be the long-time iris pairs.

For an iris image pair, all three algorithms in this study compute a similarity score. For each algorithm and each iris, we computed the average similarity score for the short-time iris pairs. Also, we computed the average similarity score for the long-time iris pairs. Formally, for an algorithm, $\mu_S(i)$ is the mean similarity score over the set of iris image pairs in $S(i)$ and $\mu_L(i)$ is the mean similarity score over the set of iris image pairs in $L(i)$.

The problem of determining if there is a time-lapse effect is formulated as a paired test of the means of the short-time lapse and long-time lapse similarity scores. For each iris, either the mean of the long-time match scores is greater than or less than the mean short-time match scores ($\mu_L(i) > \mu_S(i)$ or $\mu_L(i) < \mu_S(i)$, ignoring ties). If there is not a time-lapse effect, then we would expect for about half of the iris pairs that the mean of the long-time match scores is greater than the mean short-time match scores. This is the null hypothesis for our study. If, on the other hand, the number of irises for which the mean of the long-time match scores is greater than the mean of the short-time match scores ($\mu_L(i) > \mu_S(i)$) becomes sufficiently large, we can reject the null hypothesis. Rejecting the null hypothesis indicates that there is a time-lapse effect. To formally make this determination, we use the non-parametric one-sided sign test. The one-sided test was selected because we are testing for degradation in performance. The advantage of the one-sided sign test is that it does not make any distributional assumptions about the similarity scores or the means of similarity scores.

C. Experimental Results

The core of the analysis for the time-lapse experiment is the difference between the means of the long-time match scores and the means of the short-time match scores. Figure 2 shows a histogram of the difference in the means for the long-time and short-time match scores, $\mu_L(i) - \mu_S(i)$, for each of the three algorithms in this study.

To test if the mean of the long-time lapse match scores is greater than the mean of the short-time lapse match, we perform a one-sided sign test. This test is performed for each algorithm. The results of the sign test are presented in Table I. Table I includes the test statistic, p-value, and number of irises for which the mean of the long-time match scores is greater than the mean of the short-time match scores ($\mu_L(i) > \mu_S(i)$). The results clearly show that we can reject the null hypothesis for all three algorithms and that the means of the long-time match scores are greater than the means of the short time lapse match scores.

TABLE I

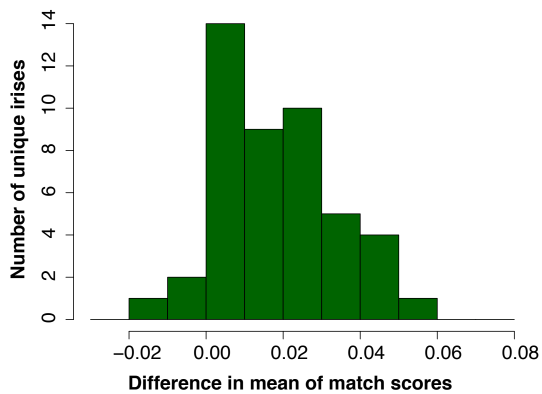
ANALYSIS WITH ONE-SIDED SIGN TEST.

Algorithm	No. irises	test statistic	p-value
irisBEE	43	6.0451	2.311×10^{-10}
VeriEye	40	5.1605	3.103×10^{-7}
Cam-2	38	4.5707	9.2477×10^{-6}

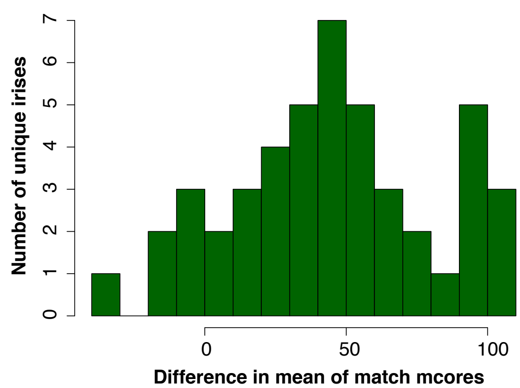
One natural question is: what is the overlap in the irises for which the mean of the long-time match scores is greater than the mean for the short-time match scores? The answers are presented in Table II which shows the overlap. The last row reports that 35 irises have the time-lapse effect for all three algorithms. In this experiment, a high hurdle for acceptance of a lapsed-time effect is that the effect must be observed for all three algorithms. Performing the one-sided sign test on the 35 irises produces a test statistic of 3.686 with a p-value of 5.3559×10^{-4}. Even if we use the criteria that all three algorithms must agree on the movement of the means, the null hypothesis is still rejected and we still conclude that there is a time-lapse effect.

D. Summary

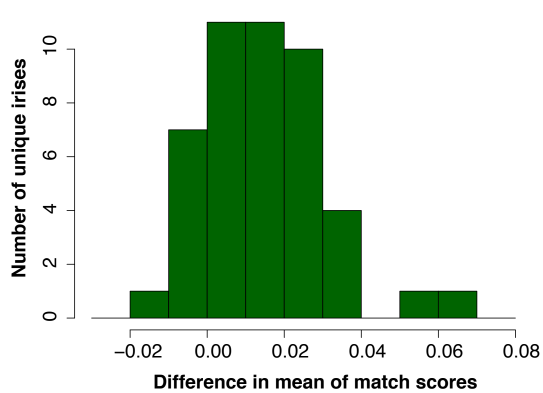
The essential summary of the experimental results is that we find statistically significant evidence that the match score for two images of the same iris degrades with increasing time lapse. This is true whether the results are considered in terms of the frequency of a degraded average match score across the set of irises, or are considered in terms of the mean of the



(a)



(b)



(c)

Fig. 2. Histogram of the differences of the mean match scores between long-time and short-time-lapse match scores, $\mu_L(i) - \mu_S(i)$. The horizontal is the long-time mean minus the short-time mean and the vertical axis is the number of uniques irises. (a) is the histogram for irisBEE, (b) is the histogram for VeriEye, and (c) is the histogram for Cam-2.

TABLE II

OVERLAP IN NUMBER OF IRISES FOR WHICH THE MEAN OF THE LONG-TIME MATCH SCORES IS GREATER THAN THE MEAN FOR THE SHORT-TIME MATCH SCORES. THE OVERLAP IS REPORTED FOR ALL COMBINATIONS OF THE THREE ALGORITHMS AND THE OVERLAP FOR ALL THREE ALGORITHMS.

Algorithms	No. overlapping irises (out of 46 total)
irisBEE-veriEye	38
irisBEE-Cam2	37
VeriEye-Cam2	36
All three	35

distribution of time-lapse differences being greater than zero. The result also holds true across the three algorithms in this study.

IV. OTHER POSSIBLE EXPLANATIONS

There are a number of other possible explanations for the degradation of match scores observed in Section III. In this section will examine six possible alternative explanations.

A. Contact Lenses

It is known that the presence of contact lenses can adversely affect match quality [17], [18]. There are two possible ways that contact lenses can affect performance. The first is if there exist match pairs where in one image a subject is wearing contact lenses and in the second contact lenses are not being worn. We performed a manual, retrospective check for contact lenses in all images included in this study. We found that either a subject always wore contracts or they never wore contacts. A similar possible cause of degradation in match scores is for a subject to switch the type of contact lenses worn, e.g., from hard to soft contact lenses. For the subjects who wore contacts, none appear to have changed the type of contacts worn.

Second, for subjects wearing contact lenses, there could be a greater rate of degradation in performance over time than for subjects who did not wear contact lenses. Five subjects (10 irises) wore contact lenses and 18 subjects (36 irises) did not wear them. We repeated the analysis in Section III for the wearing and not wearing contact lenses conditions. For the wearing contact lenses condition we found statistically significant evidence that the match score for two images of the same iris degrades with increasing time lapse for all three algorithms (irisBEE p-value = 0.00098, VeriEye p-value = 0.011, Cam-2 p-value = 0.00098). For the non-wearing contact lenses condition we found statistically significant evidence that the match score for two images of the same iris degrades with increasing time lapse for all three algorithms (irisBEE p-value = 1.14×10^{-7}, VeriEye p-value = 6.46×10^{-6}, Cam-2 p-value = 0.00060). Based on these results, we conclude that contact lenses did not cause the observed time-lapse effect.

B. Pupil Dilation

The degree of dilation of a pupil and the area an iris occupies in an image are intertwined. The more dilated the pupil, the smaller the iris and hence the fewer the pixels on the iris in an image; for an image where the pupil is not dilated, the greater the number of pixels on the iris. Hollingsworth et al. [15] showed that the size of the pupil effects the match distribution. The largest degradation in match scores was observed when comparing an image of a non-dilated pupil and an image of a dilated pupil.

To discount this effect, we performed an analysis of the changes in pupil dilation and its possible effect on the difference between long-time and short-time match scores. The first step in the analysis was to compute the ratio of the pupil diameter to the iris diameter for each image. The second step was to compute the difference in the pupil to iris ratio for the iris images in each match pair. Then, for each subject, we computed the average change in the pupil to iris ratio over all short-time match pairs. We denote this by $\rho_S(i)$. Similarly, we computed the average change in the pupil to iris ratio for all long-time match pairs, denoted by $\rho_L(i)$. Then for each iris, we computed the difference between the average short-time change in the pupil to iris ratio and the average long-time change in the pupil to iris ratio, denoted by $\rho_L(i) - \rho_S(i)$. For the irisBEE algorithm, we graphed a scatter plot of the change in the pupil to iris ratio between long-time and short-time lapse match pairs and change in match score between long-time and short time lapse, see Figure 3. Formally, Figure 3 is a scatter plot of $\mu_L(i) - \mu_S(i)$ versus $\rho_L(i) - \rho_S(i)$. The corresponding Kendall correlation coefficient is 0.217.

If the observed time lapse effect in Section III could be attributed to the change in pupil dilation, then $\mu_L(i) - \mu_S(i)$ versus $\rho_L(i) - \rho_S(i)$ would be substantially correlated. If $|\rho_L(i)| > |\rho_S(i)|$, then there is a greater difference in diameters of the pupils for long-time match match pairs than for short-time match pairs. In turn this implies that match scores should degrade

(Hamming distance increase for irisBEE). However, our analysis shows minimal correlation between $\mu_L(i) - \mu_S(i)$ versus $\rho_L(i) - \rho_S(i)$. Therefore, changes in pupil dilation do not explain our observed degradation match scores over time.

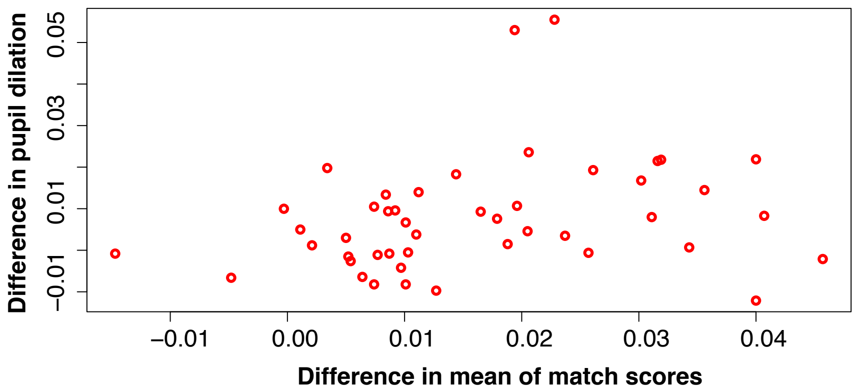


Fig. 3. Scatterplot of the change in match score between long-time and short-time lapse for each iris versus the change in the pupil to iris ratio between long-time and short-time lapse match pairs ($\mu_L(i) - \mu_S(i)$ versus $\rho_L(i) - \rho_S(i)$). The horizontal axis is the change in mean match scores for the long-time and short time lapse iris pairs. The vertical axis is the change in the average short-time change in the pupil to iris ratio and the average long-time change in the pupil to iris ratio. Each red circle is an iris.

C. Iris Occlusion

The percentage of an iris that is observable can affect performance [14]. Typical causes of occlusion are eyelids, eyelashes, and specular reflections. The more of the iris that is observable, the better the expected performance. One possible explanation for the observed time-lapse effect is that the percentage of the iris that is observable changed over time.

One of the components of the irisBEE algorithm takes a segmented iris and warps the iris into a rectangle [1]. All irises are warped to the same sized rectangle. Then a mask is created were the iris is occluded. Formally, the mask is the same size rectangle as the warped iris and the mask marks pixels where the iris is not obscured. We measure the amount an iris is visible by the number of pixels in the mask marked as not occluded by our modified version of irisBEE.

To measure if there is a change in the portion of the iris that is occluded, we divided the time period over which the data was collected for this study into 30 day intervals. We computed the average number of pixels marked as non-occluded in the mask for all images collected in each 30 day interval. We then computed Kendall's correlation coefficient between the average number of pixels marked as non-occluded and time. The resulting Kendall's correlation coefficient is -0.1328. If the percentage of the iris that is visible does not vary over time, then number of pixels marked as non-occluded and time should not be correlated. Our results show that there is no substantial correlation between number of pixels marked as non-occluded and elapsed-time.

D. Sensor Aging

The iris images in the time-lapse study were collected with the same LG 2200 sensor [2]. Over this time, it is conceivable that the sensor properties of the LG 2200 could have changed in such a way to degrade match score. To test for sensor degradation, in the Fall 2008 we collected iris images with a second rarely-used LG 2200 camera. We collected approximately 3000 images from 77 subjects (154 irises) who attended three separate acquisition sessions (labeled "session one," "session two," and "section three"). There was approximately two weeks elapsed time between each session. During sessions one and three, iris images were collected with the original camera; during session two the iris images were collected with the second rarely-used camera.

The first step in our sensor aging analysis was to compute the match and non-match score distributions between iris images collect in session one and session three. In sessions one and three the iris images were collected with the original sensor. The match and non-match scores were generated by the modified irisBEE algorithm. The second step was to compute the match and non-match score distributions between iris images collected in session one and session two. In session two, the images were collected with the second rarely-used sensor. If the sensor age affects match quality, we would expect a significant degradation in match scores between images collected from the two different sensors than images collected with the original sensor.

The average match score for image pairs collected with the original sensor is 0.2153; the average match score for image pairs collected with the two different sensors was was 0.2167. Figure 4 shows a histogram for the match and non-match distributions for both within and between sensor comparisons. We observe that the differences in the mean match scores, match distributions, and non-match distributions are not significant. Therefore, we conclude that sensor aging does not account for the degradation in match scores that we observed.

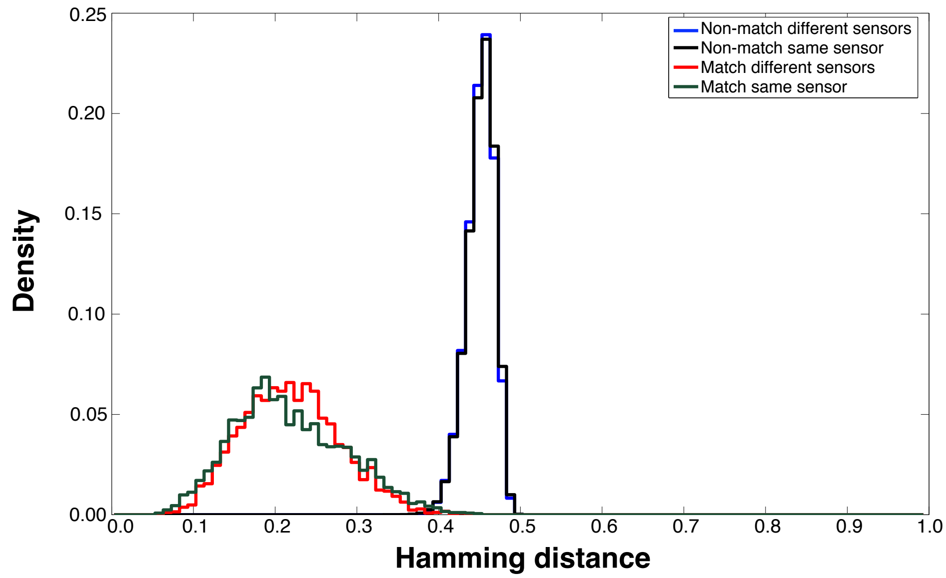


Fig. 4. The match and non-match distributions for the within and between sensors experiments. The match and non-match distributions are for the Hamming distance from the irisBEE algorithm. The mean Hamming distance for match scores collected with the same sensor is 0.2153 and for match scores collected with difference sensors is 0.2167. The mean Hamming distance for non-match scores collected with the same sensor is 0.4483 and for non-match scores collected with difference sensors is 0.4478.

E. Images Participating In Multiple Matches

In the time-lapse analysis results in Section III, each iris image was included in all possible short-time and long-time image pairs. A different experimental method would allow each image to participate in only one short-time image pair and one long-time image pair. This substantially decreases the total number of match scores in the analysis. However, it reduces the influence that any one image may have on the overall results.

For the analysis in this section, the set of short-time and long-time image pairs was randomly sampled so that each iris was only in one short-time and long-time image pair. The remainder of the statistical analysis was the same as in Section III. For the irisBEE algorithm, for 43 out of the 46 irises the long-time average match score was worse than the short-time average match score (p-value = 2.311×10^{-10}). For the VeriEye algorithm, for 44 out of the 46 irises the long-time average match score was worse than the short-time average match score (p-value = 1.538×10^{-11}). The results of this analysis clearly show that the image pair sampling method did not cause the observed time-lapse effect.

F. Illuminant Effects

The LG 2200 camera actively illuminates the iris using three infrared light emitting diodes (LED) positioned on the left, right, and top of the sensor. When acquiring images, the camera is designed to take three images, one with each LED. In commercial applications, the camera will save the best quality image and discard the other two. For our acquisitions, the system had the capability to save all three images (for a detailed explanation see Phillips et al. [9], [12]).

We repeated the analysis in Section III except that we restricted our analysis to two cases. In the first case we only looked at iris pairs illuminated by the same LED. For the irisBEE algorithm, for 43 out of the 46 irises the long-time average match score was worse than the short-time average match score (p-value = 2.311×10^{-10}). In the second case we only looked at iris pairs illuminated by different LEDs. For the irisBEE algorithm, for 43 out of the 46 irises the long-time average match score was worse than the short-time average match score (p-value = 2.311×10^{-10}). The results of this analysis clearly shows that the LED that illuminated an iris does not cause the time-lapse effect.

V. Discussion and Future Work

The results presented here run counter to conventional wisdom about iris biometrics. Our is the only work that we know of that actually test the “one enrollment for a lifetime” concept. Our results are based on images of 46 irises collected from 23 subjects over a four year period. Within this four year lapse we found statistically significant changes in the match score distributions for three iris recognition algorithms. While we observe a statistically significant change in the distribution of match scores, we did not observe a change in the non-match distrubtions. This implies that template aging primarily increases the false reject rate, and does not effect the false accept rate.

In our analysis we controlled for six other factors that could have potentially caused the observed time-lapse effect. We did not find evidence that any of the six factors could explain the reported time-lapse effect. The effects examined were presence of contact lenses; change in pupil dilation; occlusion of the irises; changes in the sensor characteristics due to sensor aging; the methodology for selecting iris image pairs; and the LED illuminating the iris.

This study and the preceding conference paper are the first studies to explicitly check the degradation of performance of iris recognition systems with samples collected over a multi-year time period. As with all scientific studies, our results should be investigated and replicated by other research groups. We recommend that studies include larger data sets, a large pool of subjects, different sensors, a longer time period, and a subjects that represent a greater range of demographics.

In an attempt to identify the cause of the degradation in match scores, we visually examined the iris images. Visual examination of the iris image pairs with the poorest match scores for the irisBEE algorithm revealed no drastic or obvious changes in the irises or their textures. This suggests that the iris template aging effect is based on subtle differences. The original claims of stability appear to be based on visual inspection. However, humans may not be able to observe the subtle changes that cause iris template aging.

To attempt to identify potential sources of these subtle changes, we examined changes in the iris code templates created by the irisBEE algorithm. In the irisBEE algorithm, iris images are compared by computing the Hamming distance between two iris code templates. Each bit in the iris code corresponds to a pixel in the iris. Therefore, it is possible to identify those regions of the iris that are responsible for the degradation in match scores. For each subject, we computed the iris codes for a set of images from Spring 2004. We then determined the consistency of each bit in the iris code for the short-time-lapse matches. Figures 5(b) and 6(b) show the areas in the iris that are the least consistent for short-time-lapse matches. Similarly, we compared the iris codes for each image from the Spring 2004 set with the iris code from each image of a set of Spring 2008 images and determined the consistency measure of each bit in the long-time-lapse matches. Figures 5(c) and 6(c) show the areas in the iris that are the least consistent for long-time-lapse matches. We observe that these localized regions are similar for both short and long-time-lapse matches. When computing the consistency measures for the two time-lapses, we only consider bits that are not masked for either occlusion or fragile bit masking in at least 50% of the images. For each bit in the aligned iris codes, we then compute the difference in the consistency measures between the long-time-lapse matches and the short-time-lapse matches. Figures 5(d) and 6(d) show the areas of the iris that experience the most change in consistency between the two time-lapse matches. The areas of the iris that undergo the largest change in consistency are generally found around the areas of the iris that were not consistent in the short-

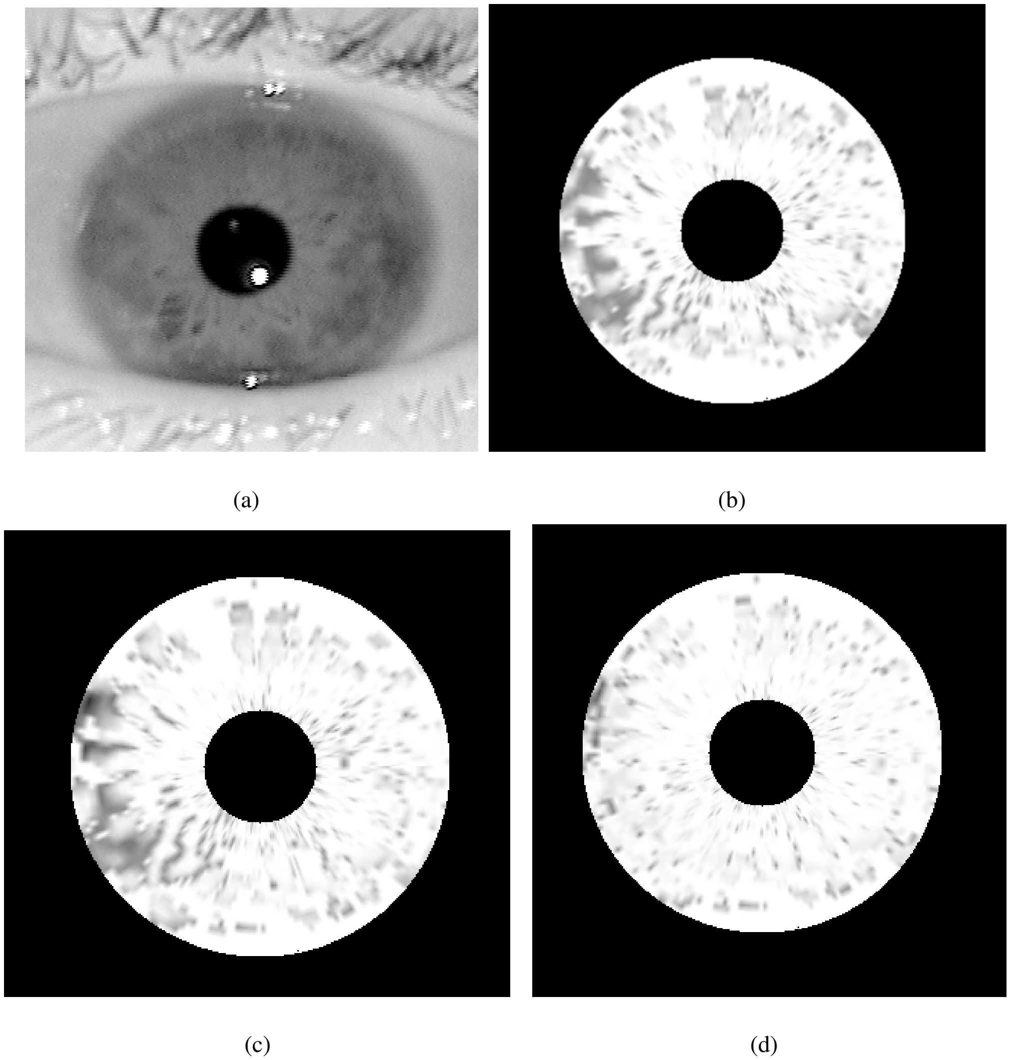
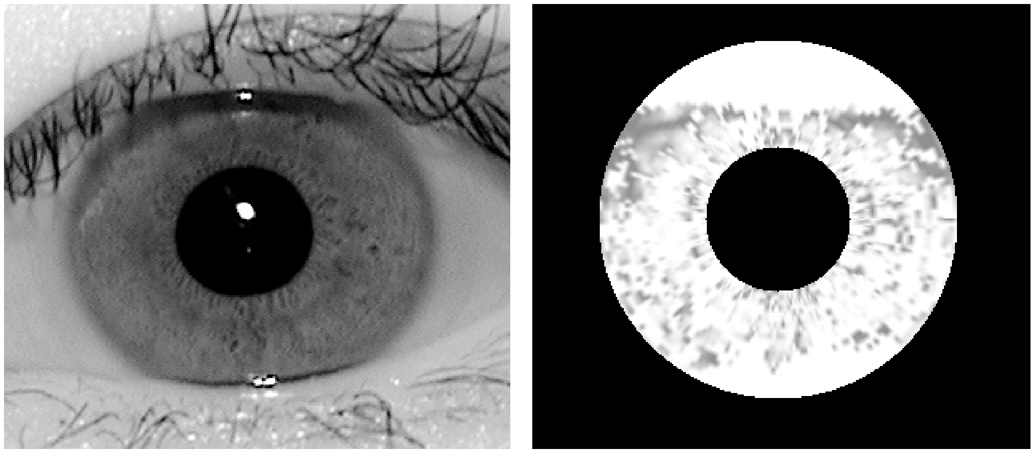


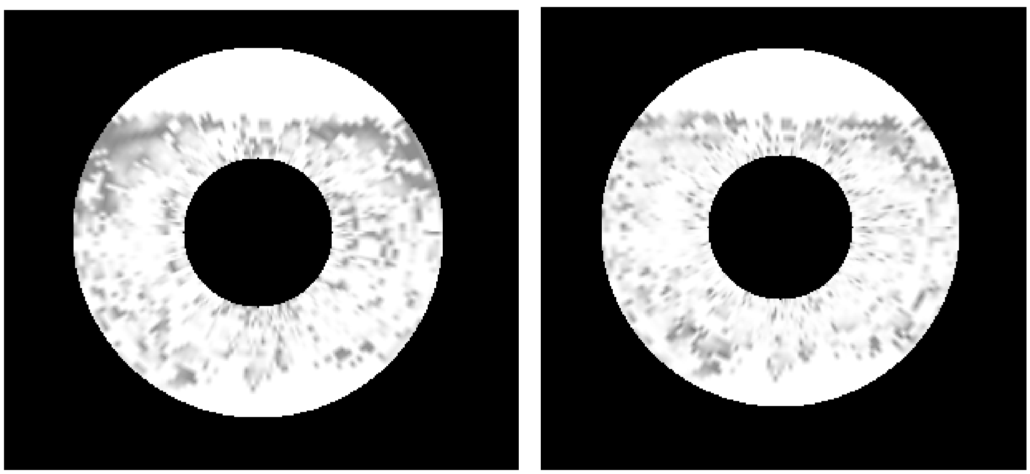
Fig. 5. Example of the difference between short-time and long-time fragile bits. (a) example iris image (left iris subject 02463), (b) fragile bits for short-time-lapse matches, (c) fragile bits for long-time-lapse matches, (d) difference in fragile bits between short-time and long-time-lapse matches.

time-lapse and the long-time-lapse matches. We conclude that these regions are changing slightly over time and are causing a change in the inconsistent bits over longer periods of time. While Figures 5 and 6 illustrate this result for two subjects, we repeated the analysis for all subjects and obtained similar results.

For three iris recognition algorithms, our results show that the match distribution changes as templates age. These results suggest that iris templates undergo aging. For other biometrics, this is an expected phenomenon. Visual inspection of the iris images did not yield obvious changes in



(a) (b)



(c) (d)

Fig. 6. Example of the difference between short-time and long-time fragile bits. (a) example iris image (left iris subject 04629), (b) fragile bits for short-time-lapse matches, (c) fragile bits for long-time-lapse matches, (d) difference in fragile bits between short-time and long-time-lapse matches.

the irises that could have caused the change in the match distribution. This suggests that changes in the iris images are subtle and could not be detected by visual inspection. The effect of subtle changes has broader implications for iris recognition. The general assumption with irises is that factors that affect performance would be noticeable and observable by humans; however, if the changes are not noticeable or observable, then selecting factors that effect iris performance by humans is problematic. Clearly, experiments on a larger number of subjects are required. The experiments should be designed to quantify the effect of iris template aging on performance, and the ability to model performance as a function of elapsed time after initial enrollment. The results of such experiment would provide a solid basis for evaluating the lifetime enrollment concept. Such an evaluation would allow for appropriate protocols for iris re-enrollment and enhance the performance iris recognition systems,

VI. Acknowledgement

SEB, KWB, and PJF were supported by the National Science Foundation under grant CNS01-30839, by the Central Intelligence Agency, by the Intelligence Advanced Research Projects Activity and by the Technical Support Working Group under US Army contract W91CRB-08-C-0093. PJP acknowledges the support of the the Biometric Task Force, the Department of Homeland Security's Directorate for Science and Technology, the Intelligence Advanced Research Projects Activity (IARPA), the Federal Bureau of Investigation (FBI), and the Technical Support Working Group (TSWG). The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of our sponsors. The identification of any commercial product or trade name does not imply endorsement or recommendation by the authors, the University of Notre Dame, or the National Institute of Standards and Technology.

References

[1] J. Daugman, "How Iris Recognition Works," *IEEE Transactions On Circuits and Systems for Video Technology,* 14(1):21-30, 2004.

[2] LG. http://www.lgiris.com/, accessed April 2009.

[3] LGE Iris Tech Win In India Redefines Biometric Scalability. http://www.findbiometrics.com/article/115, accessed April 2009

[4] J. Thornton, M. Savvides, V. Kumar. "A Bayesian Approach to Deformed Pattern Matching of Iris Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):596-606, April 2007.

[5] K. Miyazawa, K. Ito, T. Aoki, K. Kobayashi, H. Nakajima. "An Effective Approach for Iris Recognition Using Phase-Based Image Matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(10):1741-1756, Oct. 2008.

[6] D. Monro, S. Rakshit, D. Zhang. "DCT-Based Iris Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):586-595, April 2007.

[7] X. Liu, K. Bowyer, P. Flynn. "Experiments with an improved iris segmentation algorithm," In *Proc. Fourth IEEE Workshop on Automatic Identification Technologies*", 118-123, Oct 2005.

[8] X. Liu. "Optimizations in Iris Recognition." PhD Dissertation, University of Notre Dame, 2006.

[9] P. Phillips, W. Scruggs, A. O'Toole, P. Flynn, K. Bowyer, C. Schott, M. Sharpe, "FRVT 2006 and ICE 2006 Large-Scale Experimental Results," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, in press.

[10] VeriEye Iris Recognition Technology. http://www.neurotechnology.com/verieye.html, accessed November 2008.

[11] K. Hollingsworth, K. Bowyer, P. Flynn, "The Best Bits in an Iris Code," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(6): June 2009.

[12] P. J. Phillips, K. Bowyer, P. Flynn, X. Liu, T. Scruggs "The Iris Challenge Evaluation 2005" In *Proc. Second IEEE Conference on Biometrics: Theory, Applications, and Systems*. Sept. 2008.

[13] P. Tome-Gonzalez, F. Alonso-Fernandez, J. Ortega-Garcia, "On the Effects of Time Variability in Iris Recognition" In *Proc. Second IEEE Conference on Biometrics: Theory, Applications and Systems*. Sept. 2008.

[14] J. Daugman, "New Methods in Iris Recognition," *IEEE Transactions On Systems, Man, and Cybernetics.* 37(5):1167-1175, Oct 2007.

[15] K. Hollingsworth, K. Bowyer, P. Flynn, "Pupil Dilation Degrades Iris Biometric Performance," *Computer Vision and Image Understanding,* 113(1): Jan 2009.

[16] K. Bowyer, K. Hollingsworth, and P. Flynn. "Image Understanding for Iris Biometrics: A Survey." *Computer Vision and Image Understanding*, 110(2):281-307, 2008.

[17] S. Ring and K. Bowyer, "Detection of Iris Texture Distortions by Analyzing Iris Code Matching Results" In *Proc. Second IEEE Conference on Biometrics: Theory, Applications, and Systems*. Sept. 2008.

[18] S. Baker, A. Hentz, K.W. Bowyer and P.J. Flynn, "Contact Lenses: Handle With Care for Iris Recognition" In *Proc. Third IEEE Conference on Biometrics: Theory, Applications, and Systems*. Sept. 2009.

[19] N. Kalka, J. Zuo, A. Schmid, B. Cukic, "Image Quality Assessment for Iris Biometrics," In *Proc. Biometric Technology for Human Identification III*, 6202(1) 2006.

[20] S. Baker, K. Bowyer, P. Flynn, "Empirical Evidence for Correct Iris Match Score Degradation with Increased Time-Lapse Between Gallery and Probe Matches." In *Proc. Third International Conference on Biometrics*. 1170-1179, 2009.