Template Aging in Iris Biometrics: Evidence of Increased False Reject Rate in ICE 2006

Sarah E. Baker, Kevin W. Bowyer, Patrick J. Flynn and P. Jonathon Phillips

Abstract Using a data set with approximately four years of elapsed time between the earliest and most recent images of an iris (23 subjects, 46 irises, 6,797 images), we investigate template aging for iris biometrics. We compare the match and non-match distributions for short-time-lapse image pairs, acquired with no more than 120 days of time lapse between them, to the distributions for long-time-lapse image pairs, with at least 1,200 days of time lapse. We find no substantial difference in the non-match, or impostor, distribution between the short-time-lapse and the long-time-lapse data. We do find a difference in the match, or authentic, distributions. For the image dataset and iris biometric systems used in this work, the false reject rate increases by about 50% or greater for the long-time-lapse data relative to the short-time-lapse data. The magnitude of the increase in the false reject rate varies with changes in the decision threshold, and with different matching algorithms. Our results demonstrate that iris biometrics is subject to a template aging effect.

1 Introduction

The term ”template aging” refers to degradation of biometric performance that occurs with increased time between the acquisition of an enrollment image and acqui-

Sarah E. Baker
University of Notre Dame, Notre Dame IN 46556 USA, e-mail: sbaker3@cse.nd.edu

Kevin W. Bowyer
University of Notre Dame, Notre Dame IN 46556 USA, e-mail: kwb@cse.nd.edu

Patrick J. Flynn
University of Notre Dame, Notre Dame IN 46556 USA, e-mail: flynn@cse.nd.edu

P. Jonathon Phillips
National Institute of Standards and Technology, Gaithersburg, MD 20899 USA e-mail: jonathon@nist.gov

sition of the image compared to the enrollment. Template aging effects are known to exist for biometrics such as face and fingerprint [7][28][31][19][27].

The iris biometrics community has long accepted the premise that the iris is ”essentially stable” throughout a person’s life, and that this means that template aging does not occur for iris biometrics. Daugman stated the core assumption this way - “As an internal (yet externally visible) organ of the eye, the iris is well protected and stable over time”[8]. This assumption is commonly repeated in research publications dealing with iris biometrics: “[the iris is] stable over an individual’s lifetime”[30], “[the iris is] essentially stable over a lifetime”[22], “the iris is highly stable over a person’s lifetime”[24]. The commercial iris biometrics literature explicitly connects this to the idea of lifetime enrollment - “only a single enrollment in a lifetime”[17].

Note that claims about stability of the iris texture and “lifetime enrollment” are never presented as dependent on the particular sensor, algorithm, length of time lapse or any other condition. They are presented as universal claims about iris biometrics in general. Thus a single counter-example is sufficient to disprove the universal claim.

It is well known in the medical literature that the eye and iris undergo a variety of changes with age [2][5][12][23][33][34]. Any of these effects could in principle alter details of the imaged iris texture. It is also possible that a tempate aging effect could be due to aging of the sensor, changes in how a person uses the biometric system, or other factors. The essential question for iris biometrics is - does the quality of a match between two images of the same iris change with increased time between the enrollment image and the image to be recognized? That is, does a template aging effect exist? We present results of the first systematic investigation of this question.

We use an image dataset involving 23 persons (46 irises) with approximately four years of time lapse between the earliest and latest images of a given iris. We consider image pairs in a short-time-lapse group, representing no more than 120 days of time lapse between the two images, and in a long-time-lapse group, representing at least 1,200 days of time lapse. We experiment with three iris biometric systems: our modification of the IrisBEE baseline matcher [26], Neurotechnology’s VeriEye system [32], and the Cam-2 submission to the Iris Challenge Evaluation 2006 [25]. We find that, for each of the three systems, there is no significant difference in the non-match, or “impostor”, distributions between the short-time-lapse and the long-time-lapse data. We also find that, for each of the three systems, the match distribution for the long-time-lapse data is different from that for the short-time-lapse data in a way that results in an increased false reject rate. Thus, we observe clear evidence of a template aging effect for iris biometrics.

2 Previous and Related Work

We do not know of *any* experimental study that supports the conclusion that template aging does not occur for iris biometrics. Claims about the stability of iris texture appear to be based on subjective human visual perception of iris texture in visible-light images of the iris. However, it has been shown that humans are able to perceive similarities in iris texture that do not result in closer iris biometric matches [15]. Thus human perception of the general iris texture pattern does not automatically or necessarily imply anything about iris biometric operation.

Gonzalez et al. [29] report an effect of time lapse on iris recognition that may initially seem similar to our results. However, Gonzalez et al. compare 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 here, we do not consider matches between images acquired in the same acquisition session, as we expect that this is not representative of a real-world biometric scenario. We expect that “same session” images will generally result in atypically good matches. Like Gonzalez et al., we do not find any significant difference in match scores for images with a few months time lapse. However, when considering a longer time lapse than that examined in Gonzalez et al., we do observe a statistically significant degradation in match scores.

This paper expands upon our initial results [4] in several ways. First, we have increased the number of subjects from 13 to 23 and the number of irises from 26 to 46. Second, in [4] 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 have tested the time-lapse effect on two additional iris biometric algorithms: Neurotechnology’s VeriEye [32] and the Cam-2 submission to the Iris Challenge Evaluation 2006 from the University of Cambridge [25]. We also test for various possible causes of match score degradation with increased time lapse. Finally, we present ROC curves for short-time-lapse and long-time-lapse matches for each of the three algorithms, and explicitly show the difference in the false reject rates.

3 Image Dataset and Algorithms

All of the iris images used in this study were acquired with the same LG 2200 iris imaging system [16], located in the same studio throughout the four years of image acquisition. The system had no hardware or software modifications during the four years. The LG 2200 model is now discontinued. However, current state-of-the-art iris imaging systems of course did not exist at the time that data acquisition for this experiment started. We are currently pursuing additional work with images acquired using a newer model sensor and initial results [9] are generally consistent with results of this study.

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. The image acquisition protocol was the same as that used in the Iris Challenge Evaluation (ICE) 2005 and 2006 [25][26]. However, it is important to note that while the protocol for the ICE acquisitions allowed for some images that did not pass the normal built-in quality control checks of the LG 2200 [25], *all images used in this study were manually screened for image quality*. Images of noticeably poor quality were excluded from this study; e.g., out-of-focus irises, major portions of the iris occluded, obvious interlace artifacts, etc., were all excluded. Also, images that resulted in a noticeably poor iris segmentation by the IrisBEE algorithm were excluded from the study. (The detailed segmentation was not available from the other systems.)

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 years. 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 subjects are Caucasian and seven are 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 selected for use in this study was 6,797.

We created two sets of image pairs, a short-time-lapse set and a long-time-lapse set. The short-time-lapse set consists of image pairs where the two images were acquired with no more than 120 days of time lapse between them. The average time lapse in this group is 44 days. The long-time-lapse set consists of image pairs acquired with no less than 1,200 days of time lapse. The average time lapse in this group is 1,405 days. A given iris image can participate in multiple short-time-lapse pairs and multiple long-time-lapse pairs.

4 Iris Matching Algorithms

To investigate the generality of any observed effects, three different iris biometric algorithms were included in the study. First, we used our own modified version of the IrisBEE system distributed through the ICE program [25]. This system represents an iris as a 240x10x2-bit iris code generated from the complex-valued responses of one-dimensional log-Gabor wavelet filters applied to the normalized iris image [20]. For the IrisBEE matcher, the output of matching two iris images is a fractional Hamming distance. The range of the fractional Hamming distance is $[0,1]$, with zero being a perfect match and 0.5 a random level of match. Second, we used the commercial VeriEye 2.2 Iris SDK from NeuroTechnology [32]. This system produces match scores on a different scale and with a different polarity than systems employing fractional Hamming distance. For the analysis in this paper, we negated the match scores so that lower scores represented better matches. The third system was the Cam-2 submission to the ICE 2006 from the University of Cambridge [25]. The output of the Cam-2 matcher is nominally a fractional Hamming distance. Thus we

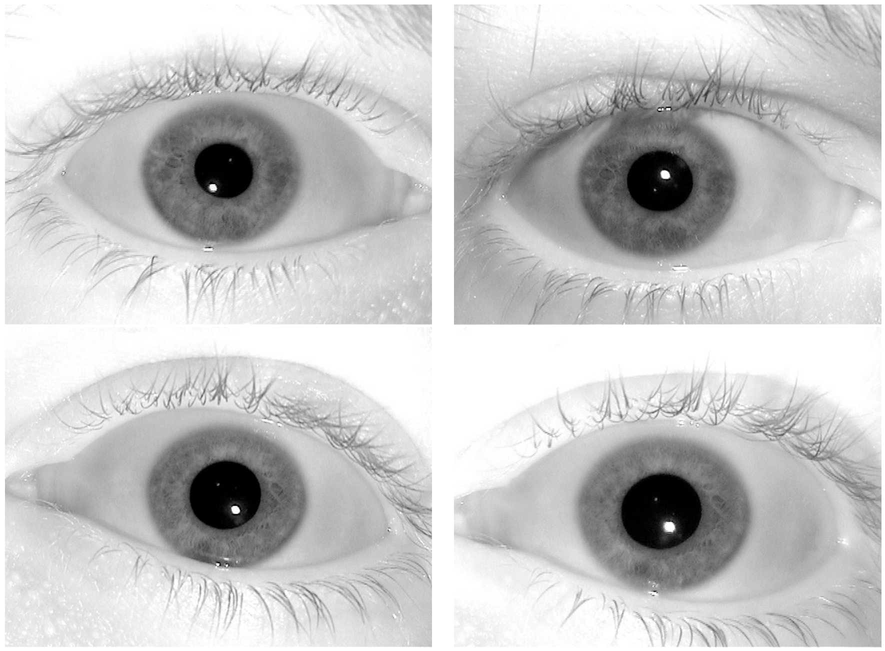

Fig. 1 Example iris images of a subject taken in 2004 and 2008 (subject 04233). Upper left: right iris from 2004; upper right: right iris from 2008; lower left: left iris from 2004; and lower right: left iris from 2008.

have used three different algorithms. One is based on a "baseline" source code that was made available to the research community, one is a readily available commercial product, and one was a best performer in the ICE 2006 results.

5 False Reject Rates for Short and Long Time Lapse

We computed the authentic and impostor distributions for each of the three algorithms. The impostor distributions showed no apparent difference between the short-time-lapse data and the long-time-lapse data. However, the authentic distributions for long-time-lapse data were shifted in the direction of the impostor distribution. For each of the three algorithms, the shift in the authentic distribution is such that it causes an increase in the False Reject Rate (FRR) for any practical choice of decision threshold.

Graphs that zoom in on the "tails" of the long-time-lapse and short-time-lapse authentic distributions for each algorithm are shown in Figure 2. These graphs show the tails of the distributions across a range of possible values for the decision threshold. Recall that for the IrisBEE and Cam-2 algorithms, a smaller value (of fractional

Hamming distance) represents a better match, while for the VeriEye algorithm a larger value of different units represents a better match.

This figure shows that for all three algorithms, across a broad range of possible threshold values, *the long-time-lapse authentic distribution has a higher false reject rate than the short-time-lapse authentic distribution*. The IrisBEE algorithm shows approximately 150% increase in the false reject rate across the range of decision thresholds, the VeriEye algorithm shows an approximately 70% increase, and the Cam-2 algorithm shows an approximately 40% increase. Thus we observe clear and consistent evidence of a template aging effect for each of three algorithms considered in this study.

6 Frequency of Authentic Distribution With Worse Mean Score

We also performed a one-sided sign test to check for statistical significance of the frequency, across the 46 irises, of the long-time-lapse authentic distribution having a worse mean match score than the short-time-lapse authentic distribution. A worse mean score is one closer to the impostor distribution. If time lapse has no effect, then we would expect that the long-time-lapse mean is worse for half of the irises and the short-time-lapse mean is worse for half. This is the null hypothesis for the test. The sign test does not make any distributional assumptions about the means of similarity scores. The one-sided test was selected because we are interested in the alternative hypothesis that the longer-time-lapse data has a larger mean score.

Table 1 Sign test for frequency of worse mean match score with longer time lapse.

Algorithm	No. irises	test statistic	p-value
IrisBEE	42	5.75	2.55×10^{-9}
VeriEye	41	5.46	2.20×10^{-8}
Cam-2	38	4.57	4.62×10^{-6}

The sign test results are presented in Table 1, including the test statistic, p-value, and number of irises for which the mean of the long-time match scores is worse than the mean of the short-time-lapse match scores ($\mu_L(i) > \mu_S(i)$). The results show that we can easily reject the null hypothesis for all three algorithms. The frequency of a worse match score occurring for the long-time-lapse is statistically significant. This indicates that the increased FRR seen in Figure 2 is not the result of a small number of unusual irises in the data set, but is characteristic of the data set in general.

Table 1 shows that for IrisBEE there are 42 of 46 irises for which the long-time-lapse mean HD is worse, for VeriEye there are 41 irises for which the long-time-lapse mean match score is worse, and for Cam-2 there are 38 irises for which the long-time-lapse mean HD is worse. One natural question is: how many of these irises are in common? The answers are presented in Table 2, which shows the number of irises in common. The last row reports that 34 irises have the time-lapse

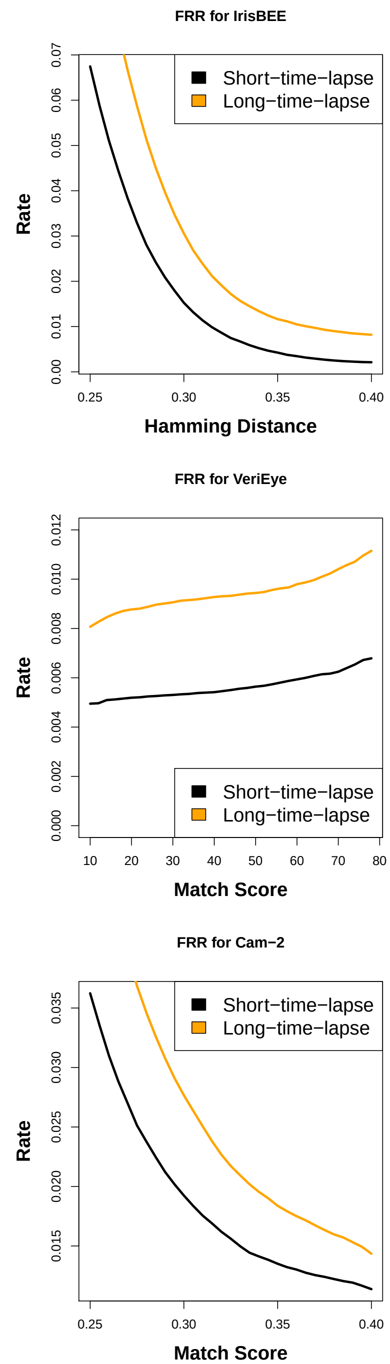

Fig. 2 Authentic distributions across a range of match scores, showing increased false reject rates.

Table 2 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 for all three algorithms.

Algorithms	N of 46 irises in common
IrisBEE-veriEye	38
IrisBEE-Cam2	35
VeriEye-Cam2	35
All three	34

effect for all three algorithms. A one-sided sign test for 34 of 46 irises showing an effect across all three algorithms produces a test statistic of 3.391 with a p-value of 8.207×10^{-4}. Thus, even if we use the criteria that all three algorithms must agree on the movement of the means, the null hypothesis is rejected.

7 Possible Causes of an Increased False Reject Rate

We considered a variety of factors that could conceivably contribute to causing the observed result. For example, it is known that the presence of contact lenses can adversely affect match quality [3]. If the short-time-lapse data contained image pairs where a subject did not wear contact lenses and the long-time-lapse data contained image pairs where the same subject was wore contacts, this could conceivably cause an increased FRR for long-time-lapse relative to short-time-lapse. Similarly, if a person was wearing the same type of contacts in short-time-lapse image pairs, but a different type in long-time-lapse image pairs, this could conceivably cause an increased FRR.

We manually checked for the presence of contact lenses in all images included in this study. We found that each subject in this study either wore contacts for all acquisition sessions, or did not wear contacts to any acquisition session. Also, for the subjects who wore contacts, none appear to have changed the type of contacts worn. Thus we conclude that the wearing of contact lenses is not an appreciable factor in our observed results.

Hollingsworth et al. [13] showed that the degree of the pupil dilation, and the difference in pupil dilation between two images, can affect the match distribution. We performed an analysis of the changes in pupil dilation and its possible effect on the difference between long-time-lapse and short-time-lapse data.

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-lapse 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-lapse change in the pupil to iris ratio and the average long-time-lapse change in the pupil to iris ratio, denoted by $\rho_L(i) - \rho_S(i)$. For the IrisBEE algorithm, we created a scatter plot of the change in the pupil-to-iris ratio between long-time-lapse and short-time-lapse match pairs and change in match score between long-time-lapse and short-time-lapse. 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 increase in false reject rate could be attributed to a 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 pairs than for short-time match pairs. In turn this implies that match scores should degrade. However, our analysis shows minimal correlation between $\mu_L(i) - \mu_S(i)$ versus $\rho_L(i) - \rho_S(i)$. Thus we conclude changes in pupil dilation are not an appreciable factor in our observed result.

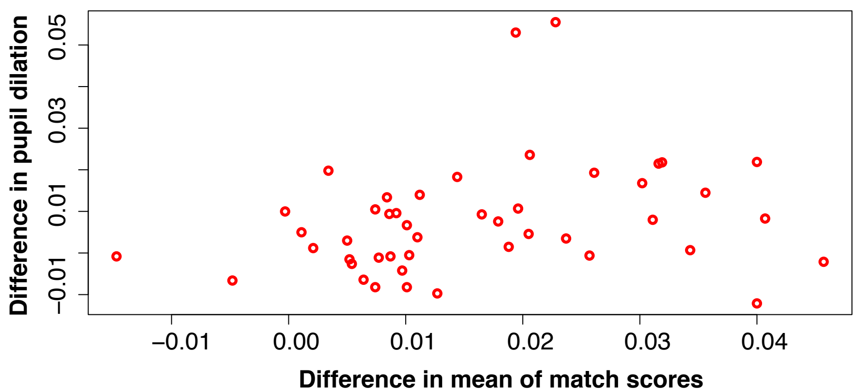

Fig. 3 Scatterplot of the change in match score between long-time and short-time lapse for each iris versus the change in the pupil to iris ratio between long-time and short-time lapse match pairs ($\mu_L(i) - \mu_S(i)$ versus $\rho_L(i) - \rho_S(i)$). The horizontal axis is the change in mean match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

The percentage of an iris that is occluded can affect iris matching performance [10]. The more of the iris that is observable, the better the expected performance. Thus one possible factor contributing to the observed increase in the false reject rate is that the percentage of the iris that is observable decreased in the long-time-lapse data relative to the short-time-lapse data.

In the IrisBEE algorithm [25], the fraction of the iris that is visible is indicated by the fraction of the iris code bits that are marked in the iris code mask as representing non-occluded portions of the iris. To determine if there is a change over time in the fraction 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 bits 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 bits marked as non-occluded and time. The resulting Kendall's correlation coefficient is -0.131. This indicates that there is no substantial correlation between number of bits marked as non-occluded and elapsed time. Thus, we conclude that change in the amount of iris occluded does not account for the increase in the false reject rate observed in our results.

The iris images in the time-lapse study were collected with the same LG 2200 sensor [16]. It is conceivable that the sensor properties of the LG 2200 could have changed over time in such a way as to cause an increased false reject rate in the long-time-lapse data. To test for this, 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 collected in session one and session three, both sessions using the original sensor. 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 compared to image pairs collected with the original sensor. The average match score for image pairs collected with the original sensor is 0.215; the average match score for image pairs collected with the two different sensors was was 0.217. Figure 4 shows a histogram for the match and non-match distributions for both within and between sensor comparisons. Based on this analysis, we conclude that a sensor aging effect cannot account for the increase in false reject rate that is seen in our results.

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. [26, 25]). It is conceivable that if there were more matches between images acquired with the same LED in the short-time-lapse group, and more matches between images acquired with different LEDs in the long-time-lapse group, that this could result in an increased false reject rate for the long-time-lapse group.

We grouped the matches into those in which the two images were taken with the same LED and those in which the two images were taken with different LEDs. For both groups, we observed an increased false reject rate of about 50% across all feasible decision threshold values for the long-time-lapse data over the short-time-

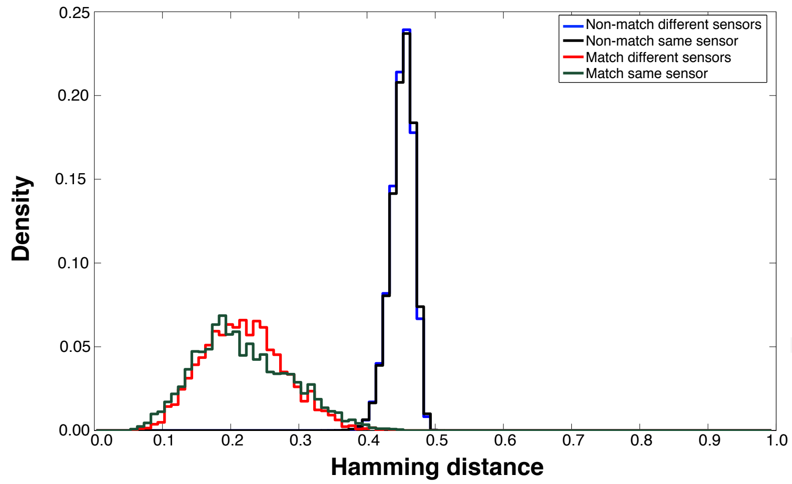

Fig. 4 The match and non-match distributions for the within and between sensors experiments. The match and non-match distributions are for the Hamming distance from the IrisBEE algorithm. The mean Hamming distance for match scores collected with the same sensor is 0.2153 and for match scores collected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

lapse data. Thus we conclude that variations in the particular LED illuminating the images is not the cause of the increased false reject rate seen in our results.

8 Conclusions and Discussion

For three different matching algorithms, and across the range of practical decision threshold values for each matching algorithm, we found that the false reject rate increases with longer time lapse between enrollment and verification. This is seen clearly in the difference in the tails of the authentic distributions. Also, the frequency of irises with a worse mean match score for long-time-lapse compared to short-time-lapse is statistically significant. Thus our experimental results show clear and consistent evidence of a template aging effect for iris biometrics. The magnitude of the template aging effect varies between algorithms, with the value of the decision threshold, and other factors.

We were able to test for a variety of factors that could potentially contribute to observing an increased false reject rate with increased time lapse. We concluded that factors such as varying pupil dilation, wearing of contact lenses, differences in amount of iris occluded, and sensor aging are not an appreciable factor in our experimental results.

It is possible that the template aging effect observed in our experimental results is caused by normal aging of the eye. One well-known example of age-related change in the normal eye involves pupil size. Winn et al. studied factors affecting light-adapted pupil size and found that "of the factors investigated, only chronological age had a significant effect on the size of the pupil"[33]. They concluded "the results of this study are consistent with previous reports suggesting that pupil size becomes smaller in an almost linear manner with increasing age" [33]. The iris, of course, controls the pupil size, and so this change in average pupil size reflects a change in the functioning of the iris tissue. As the Merck Manual of Geriatrics describes it, "The iris comprises two sets of muscles that work together to regulate pupillary size and reaction to light. With aging, these muscles weaken and the pupil becomes smaller (more miotic), reacts more sluggishly to light, and dilates more slowly in the dark" [23].

There are also age-related changes in the melanocytes, the cells that produce melanin, in the iris. Eye color is largely determined by the melanocytes in the anterior layer of the iris. For some segments of the population, aging can lead to a noticeable change in the melanocytes, and so the eye color. Bito et al. report that "Most individuals had stable eye color after early childhood. However, there was a subpopulation of white subjects with eye color changes past childhood. Approximately 17% of twins and 11% of mothers experienced a change in eye color of 2 U or more. [...] Thus, eye color, and hence, iridial pigmentation, seems to change in some individuals during later years" [5]. They found that the changes in eye color were more similar for identical twins than fraternal twins, indicating a genetic link to this particular element of aging. One element of melanocyte aging can, in rare cases, lead to a cancer. "The melanocytes in the iris are constantly exposed to UV radiation, and this leads to the malignant transformation of these cells to form a specific type of malignant tumor, the uveal melanoma" [12].

Also connected with the melanocytes, iris freckles and nevi can arise in the iris, and can grow over time. "Iris freckles are the most common iris tumors found in children as well as adults. They are collections of benign, but abnormal melanocytes that vary in size and shape. Although congenital, they tend to become more prominently pigmented with age. Iris freckles are clusters of normal melanocytes and have no malignant potential. Nevi efface the iris architecture and may cause clinical structural alterations ..." [34].

In addition, it is known that the cornea undergoes age-related changes. "The shape and aberrations of the cornea change with age. It is well known that the radius of curvature slightly decreases with age, and the asphericity also changes. On average, the cornea becomes more spherical with age and, as a consequence, spherical aberrations tend to increase" [1]. The iris is imaged through the cornea, thus, corneal changes may affect iris images.

Small, incremental changes in imaged iris texture over time should be considered normal, as "... age related changes take place in all ocular tissues of the human eye ..." [2]. The relevant question for iris biometrics is the time scale at which normal aging has an appreciable effect on the biometric template computed from the imaged iris texture. To underscore this point, we quote from the Flom and Safir iris recognition patent [11] - “The basic, significant features of the iris remain extremely stable and do not change over a period of many years. Even features which do develop over time, such as the atrophic areas discussed above, usually develop rather slowly, so than an updated iris image will permit identification for a substantial length of time”. In this quote, it is clear that Flom and Safir anticipated the possibility that small, incremental changes in iris texture could potentially result in the need for an “updated image” and re-enrollment of the iris template. One interpretation of our results is that they confirm that the possibility that Flom and Safir envisioned is in fact true.

In an attempt to identify the regions of the iris that changed, degrading the match quality, 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, if the template aging effect is due to normal aging of the eye, humans may not be able to easily perceive the subtle changes that are involved.

Much additional research remains to be done in the area of template aging for iris biometrics. While we have experimentally observed a template aging effect, and have ruled out several factors as primary causes of the observed effect, we have not conclusively identified a primary cause of the observed template aging. It is important to understand the cause of the observed template aging effect, so that techniques can be developed to mitigate the effect. It would also be valuable to know whether or not iris biometric template aging is constant across different demographic groups, and whether it occurs at a faster or slower rate as a person ages. Studies that collect new and larger data sets, involve a larger pool of subjects, different sensors, a longer time period, and / or a sample of subjects that represent a greater range of demographics would all be important.

Acknowledgements 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. P. Artal, “Chapter 3: Aging effects on the optics of the eye,” in *Age-Related Changes of the Human Eye*, Carlo A.P. Cavallotti and Lucianon Cerulli, editors, Humanan Press, Totowa, NJ, 2008.

2. D. Atchison, E. Markwell, S. Kasthurirangan, J. Pope, G. Smith, P. Swann. “Age-related changes in optical and biometric characteristics of emmetropic eyes,” *Journal of Vision* 8(4):29, 1-20, 2008.
3. S. Baker, A. Hentz, K.W. Bowyer and P.J. Flynn, “Degradation of iris recognition performance due to non-cosmetic prescription contact lenses”, *Computer Vision and Image Understanding* 114 (9), 1030-1044. Sept. 2010.
4. 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.
5. L. Bito, A. Matheny, K. Cruickshanks, D. Nondahl, O. Carino, “Eye color changes past early childhood, ” *Archives of Ophthalmology*, 115, 659-663, May 1997.
6. K.W. Bowyer, K. Hollingsworth, and P. Flynn. “Image Understanding for Iris Biometrics: A Survey.” *Computer Vision and Image Understanding*, 110(2):281-307, 2008.
7. J.W. Carls, R. Raines, M. Grimaila, S. Rogers, “Biometric enhancements: Template aging error score analysis”, *8th IEEE Int’l Conference on Automatic Face and Gesture Recognition*, 2008. FG ’08. Sept 2008, 1-8.
8. J. Daugman, “How Iris Recognition Works,” *IEEE Transactions On Circuits and Systems for Video Technology,* 14(1):21-30, 2004.
9. S. Fenker, K.W. Bowyer, “Experimental evidence of a template aging effect in iris biometrics,” *IIEEE Computer Society Workshop on Applications of Computer Vision.* January 2011..
10. J. Daugman, “New Methods in Iris Recognition,” *IEEE Transactions On Systems, Man, and Cybernetics.* 37(5):1167-1175, Oct 2007.
11. L. Flom, A. Safir, “Iris Recognition Systems,” U.S. Patent No. 4641394, 1987.
12. D. Hu, “Photobiology of the Uveal Tract,” *Photobiological Sciences Online*, Kendric C. Smith, editor, http://www.photobiology.info/.
13. K. Hollingsworth, K.W. Bowyer, P.J. Flynn, “Pupil Dilation Degrades Iris Biometric Performance,” *Computer Vision and Image Understanding,* 113(1): Jan 2009.
14. K. Hollingsworth, K.W. Bowyer, P.J. Flynn, “The Best Bits in an Iris Code,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(6): June 2009.
15. Hollingsworth, K.W. Bowyer, P.J. Flynn, “Similarity of Iris Texture Between Identical Twins,” *Computer Vision and Pattern Recognition Biometrics Workshop*, June 2010.
16. LG. http://www.lgiris.com/, accessed April 2009.
17. LGE Iris Tech Win In India Redefines Biometric Scalability. http://www.findbiometrics.com/article/115, accessed April 2009
18. 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.
19. A. Lanitis, “A survey of the effects of aging on biometric identity verification”, *Int’l Journal of Biometrics*, 2(1): 34-62 2010.
20. X. Liu, K.W. Bowyer, P. Flynn. “Experiments with an improved iris segmentation algorithm,” In *Proc. Fourth IEEE Workshop on Automatic Identification Technologies*”, 118-123, Oct 2005.
21. X. Liu. “Optimizations in Iris Recognition.” PhD Dissertation, University of Notre Dame, 2006.
22. 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.
23. *The Merck Manual of Geriatrics, 3rd edition*, Mark H. Beers, editor, Chapter 126: Aging and the Eye, http://www.merck.com/mkgr/mmg/sec15/sec15.jsp.
24. D. Monro, S. Rakshit, D. Zhang. “DCT-Based Iris Recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):586-595, April 2007.
25. 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*, 32: 831-846, 2010.

26. 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.
27. P.J. Phillips, P. Grother, R. Michaels, D. Blackburn, E. Tabassi, M. Bone, "Face Recognition Vendor Test 2002: Overview and Summary", 4 March 2000.
28. J. Ryu, J. Jang, H. Kim, "Analysis of Effect of Fingerprint Sample Quality in Template Aging", *NIST Biometric Quality Workshop II*, Nov 7-8, 2007.
29. 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.
30. 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.
31. U. Uludag, A. Ross, A. Jain, "Biometric Template Selection and Update: a Case Study in Fingerprints", *Pattern Recognition*, 37: 1533-1542 2004.
32. VeriEye Iris Recognition Technology. http://www.neurotechnology.com/verieye.html, accessed November 2008.
33. B. Winn, D. Whitaker, D. Elliot, N. Phillips, "Factors Affecting Light-adapted Pupil Size in Normal Human Subjects", *Investigative Ophthalmology and Visual Science*, 35(3): 1132-1137 1994.
34. K. Wright, P. Spiegel, *Pediatric Ophthalmology and Strabismus*, Springer-Verlag, New York, 2003, page 438.