Factors that degrade the match distribution in iris biometrics

**Kevin W. Bowyer · Sarah E. Baker ·
Amanda Hentz · Karen Hollingsworth ·
Tanya Peters · Patrick J. Flynn**

Abstract We consider three "accepted truths" about iris biometrics, involving pupil dilation, contact lenses and template aging. We also consider a relatively ignored issue that may arise in system interoperability. Experimental results from our laboratory demonstrate that the three accepted truths are not entirely true, and also that interoperability can involve subtle performance degradation. All four of these problems affect primarily the stability of the match, or authentic, distribution of template comparison scores rather than the non-match, or imposter, distribution of scores. In this sense, these results confirm the security of iris biometrics in an identity verification scenario. We consider how these problems affect the usability and security of iris biometrics in large-scale applications, and suggest possible remedies.

K. W. Bowyer (✉) · S. E. Baker · A. Hentz · K. Hollingsworth · T. Peters · P. J. Flynn
Department of Computer Science & Engineering, University of Notre Dame, Notre Dame,
IN 46530, USA
e-mail: kwb@cse.nd.edu

K. Hollingsworth
e-mail: kholling@cse.nd.edu

T. Peters
e-mail: tpeters@cse.nd.edu

P. J. Flynn
e-mail: flynn@nd.edu

Introduction

Iris biometrics offers the potential for highly accurate identity verification (Daugman 1993, 2000, 2004). Iris biometrics research has expanded greatly in recent years (Bowyer et al. 2008). The number of companies offering iris biometrics systems has also expanded, and competing technical approaches are represented in commercial products. Experience with the use of iris biometrics in some relatively large applications, for example in the United Arab Emirates, has been reported as quite positive (Daugman 2006). These various factors have led to consideration of iris biometrics for use in what might be termed nation-sized, all-citizens applications (Mansfield and Rejman-Greene 2003). A natural question to ask is—What unexpected difficulties might potentially arise in using iris biometrics in an application meant to serve all citizens of a nation?

This paper considers three widely accepted "truths" about iris biometrics and presents experimental evidence that each is in fact false. These accepted truths involve pupil dilation, contact lenses and template aging. We also consider a typically ignored problem that can arise from a requirement for system interoperability. Each of these four problems affects primarily or solely the stability of the authentic distribution, referred to as the "match distribution" in the remainder of the paper.

The paper is organized as follows. "Overview of iris biometrics technology" gives a brief overview of iris biometrics technology. The reader already familiar with iris biometrics may be able to skip this section. "Effects of pupil dilation on iris biometrics accuracy" considers the effect of varying pupil dilation on the accuracy of iris biometrics. "Effects of contact lenses on iris biometrics accuracy" examines how wearing normal, prescription contact lenses can affect iris biometrics performance. "Template aging" examines the issue of significant time lapse between enrollment and verification, or "template aging." "Interoperability between systems" discusses an issue that can arise from sensor interoperability requirements. Lastly, "Discussion" discusses the pattern of results in the previous sections, and suggests possible means of dealing with the problems that are identified.

Overview of iris biometrics technology

The iris is the colored portion of the eye that surrounds the pupil (Fig. 1). Iris biometrics uses a representation of the texture pattern of the iris as a means to verify a person's identity. The predominant technical approach to iris biometrics is what we will call the Daugman-style approach. Daugman (1993) created this approach in the early 1990's and has introduced numerous improvements since then (e.g., Daugman 2007). Thus the phrase "Daugman-style approach" actually refers to a family of distinct but closely related algorithms. We use the term "Daugman-style" here to refer to an approach in which the segmented region is "unwrapped" to a standard-size rectangular frame, the results of analyzing the iris texture are represented as a binary code, and the difference between two binary iris codes is computed as a fractional Hamming distance.
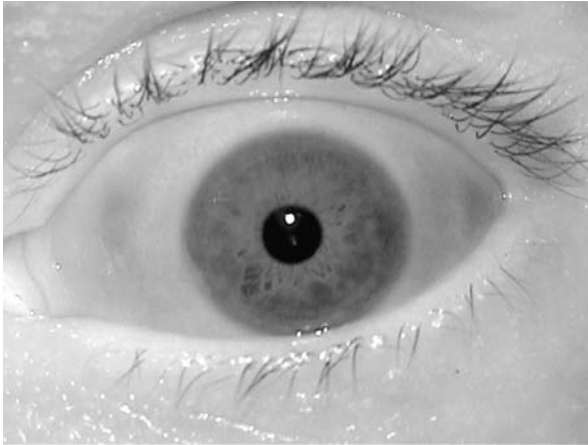
Fig. 1 Example image acquired using an LG 2200 iris biometrics system. The highlight in the otherwise dark pupil region is due to the near-infrared illuminator used by the system. The iris is the textured, "donut-shaped" region surrounding the pupil. Iris biometrics uses a representation of the pattern of texture in the iris to create an identifier intended to be specific to that particular iris

The first step in processing an iris image is to find the boundaries of the iris region. Then the iris region is "unwrapped" from its annular appearance in the original image onto a standard-size rectangular frame. A texture filter is then applied at each of a fixed set of locations in the rectangular frame. The texture filter is typically a Gabor or log-Gabor function, but researchers have explored a variety of other texture filters as well (Bowyer et al. 2008). The result of each texture filter application is a complex number. Each complex-valued result is quantized to two bits of information, based on the signs of the real and the imaginary parts of the complex filter response. The collection of bits for the texture filter results then creates a binary "iris code." The iris code is effectively a high-level abstraction of the appearance of the iris texture. Two iris codes are matched by computing the fractional Hamming distance between them; that is, the fraction of the iris code bits in which the two codes disagree. If the fractional Hamming distance is less than a specified threshold, the images are declared to be a match; that is, to represent the same iris. If the distance is greater than the threshold, then the images are declared to be a non-match; that is, to represent different irises.

The effects of potential problems that may occur with iris biometrics can often usefully be understood in terms of the changes in the match, or authentic, distribution and the non-match, or imposter, distribution. Figure 2 shows the match and non-match distributions found using an iris image dataset acquired in our laboratory and the version of the open-source IrisBEE software used in our laboratory (Liu et al. 2005; Phillips et al. 2009). The IrisBEE software is a particular Daugman-style implementation used as a baseline in the Iris Challenge Evaluation (Phillips et al. 2009).

The match distribution represents the observed fractional Hamming distances between iris codes obtained from pairs of images of the same iris. The non-match distribution represents fractional Hamming distances from comparing iris codes for
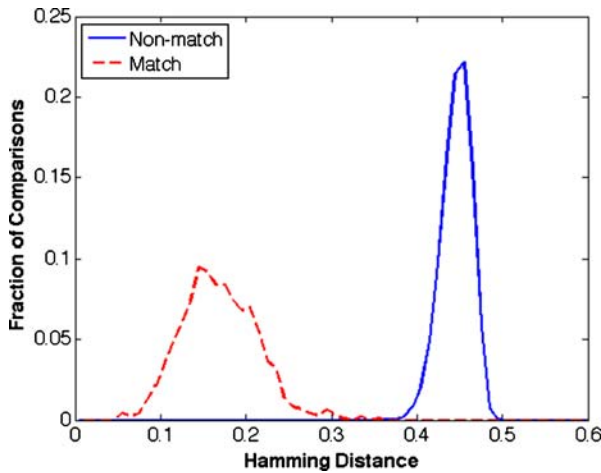
Fig. 2 Example match (Authentic) and non-match (Imposter) distributions. This data results from matching that includes searching a range of possible rotations to align the irises, and so the non-match distribution is not centered on 0.5. The "tail" of the match distribution toward larger values of fractional Hamming distance is generally due to segmentation inaccuracies or image artifacts of various kinds

images of different irises. The non-match distribution depicted in Fig. 2 has a "tail" toward lower Hamming distance values. This tail arises as a result of allowing for some difference in orientation of the head, and so of the iris, between two images. The iris code for the iris to be recognized is matched against the enrolled iris code multiple times, shifting the iris code to be recognized through a range of possible rotation values, and keeping the best (lowest) Hamming distance result. This causes the non-match distribution to have a degree of skew toward lower Hamming distance values. The match distribution depicted in Fig. 2 has a "tail" toward higher Hamming distance values. We have examined iris matches represented in this tail and found that the higher Hamming distance values can often be attributed to segmentation inaccuracies or image artifacts.

In an identity verification scenario, a person's claimed identity is verified by comparing the iris code from a current image to the iris code from their enrollment image. If the fractional Hamming distance is less than a specified threshold, the images are declared to be a match; that is, to represent the same iris. If the value is greater than the threshold, then the images are declared to be a non-match; that is, to represent different irises. In this scenario, the two types of errors that the system can make are (1) to falsely declare a match, committing a "false match" or "false accept" error, and (2) to falsely declare a non-match, committing a "false non-match" or "false reject" error. The greater the separation between the match and the non-match distributions, the more powerful is the biometric system, because a smaller fraction of either distribution would fall on the wrong side of the decision threshold.

Daugman has shown that a binomial distribution with a certain number of degrees of freedom agrees well with the non-match distribution obtained from a body of experimental data available to him and using his own iris biometrics implementation (Daugman 1993). A less well-designed implementation of a Daugman-style approach would not necessarily have the same number of degrees of freedom.

Assuming that the fitted binomial distribution is an appropriate model for the data, a threshold can be chosen for the fractional Hamming distance such that only 1 in 1.2 million of the non-match distances fall below the threshold. Thus, under this assumed model, the chance of a false match error using the selected threshold is 1 in 1.2 million. However, the probability of the false non-match error when using this threshold is not specified, and it depends on the shape and placement of the match distribution. It is essential that the false non-match rate fall in an acceptable range for the particular application at hand. As Daugman (1993) has observed, "It is important to note immediately the uselessness of either error rate statistic alone in characterizing performance."

Two additional points are worth making about the binomial model of the non-match distribution and the resulting "1 in 1.2 million" estimation of the false match error rate. First, the binomial model does not directly apply to non-match distributions of the sort depicted in Fig. 2, which result from optimizing over a range of possible rotations in the iris code. However, iris biometric systems that capture both eyes in the same image and use the line between the iris centers to normalize the orientation of the iris code can potentially avoid this step. Second, a skeptical view might say that the appropriateness of the binomial model, especially in the extreme tails of the distribution, needs further validation. However, larger-scale results have appeared that do not show the model to be inappropriate (Daugman 2006).

Effects of pupil dilation on iris biometrics accuracy

"Even though the visible portion of the iris changes as a function of pupil dilation, this does not adversely affect authentication." (Weaver 2006)
"Variations in pupil size do not interfere with the randomness or uniqueness of iris patterns." (EC 2005)

As the above quotes illustrate, one element of accepted truth about iris biometrics is that a varying degree of pupil dilation between images does not degrade iris biometric performance. At first glance, this assertion may seem plausible because in the typical approach the segmented iris region is transformed into standard rectangular size. However, recent work in our laboratory shows that varying pupil dilation can in fact degrade iris biometric performance (Hollingsworth et al. 2008, 2009b).

In this experiment, 18 subjects had iris images acquired on several different days with normal indoor room lighting levels in the studio, and also with the lights off. Having the lights off allowed us to observe a degree of pupil dilation without the use of eye drops or sunglasses. For the results presented here, we use a dataset of 1,263 iris images (Hollingsworth et al. 2009b), acquired using an LG 2200 iris imaging system (LG 2009a), and processed with our version of the IrisBEE software (Liu et al. 2005; Phillips et al. 2009). For each image, the pupil dilation ratio was computed using the results of segmenting the iris region using circular boundaries. The pupil dilation ratio is simply the radius of the pupil divided by the radius of the iris. This results in a value that can vary between 0 and 1, with low values representing a

constricted pupil and high values representing a dilated pupil. See Fig. 3 for example images representing low and high values of pupil dilation. A pupil dilation ratio near 0.2 appears as a strongly constricted, or small, pupil, and a pupil dilation ratio near 0.7 appears as a highly dilated, or large, pupil.

The fractional Hamming distances for matches between two iris images were then divided into three groups, according to the difference in the pupil dilation ratio between the two images. If the difference in pupil dilation ratio between the two images is less than or equal to 0.1, it is considered to be small. If the difference is greater than 0.1 and less than or equal to 0.2, it is considered to be medium. And if the difference is greater than 0.2 and less than or equal to 0.3, it is considered to be large. There were too few image pairs with a difference greater than 0.3 to generate a useful distribution. The match and the non-match distributions were computed separately for the three groups. The non-match distribution did not change substantially with a change in the difference in the pupil dilation ratio. However, as shown in Fig. 4, the mean of the match distribution increased consistently from low to medium to high difference in pupil dilation. The results of this experiment can be interpreted as saying that, in an identity verification scenario, for a given decision threshold, the larger the difference in pupil dilation between the time of enrollment and the time that a person attempts to verify their identity, the greater the chance that the person will experience a false non-match.

In a different experiment, the match and non-match distributions were computed separately for matches between pairs of images that both had small dilation, that both had medium dilation and that both had large dilation. Thus this experiment looks at the effect of the degree of pupil dilation when both images have similar dilation, rather than the effect of differences in dilation between the two images. Across the three groups in this experiment, we found that the mean of the match distribution increases with increasing dilation, and that the mean of the non-match distribution decreases slightly with increasing dilation. These results can be interpreted as saying that, for a given fixed decision threshold, increasing the dilation in both images increases the chances of both false match and false non-match outcomes. Of course, if the pupil dilation ratio in the enrollment image can be controlled, then this effect can be avoided.
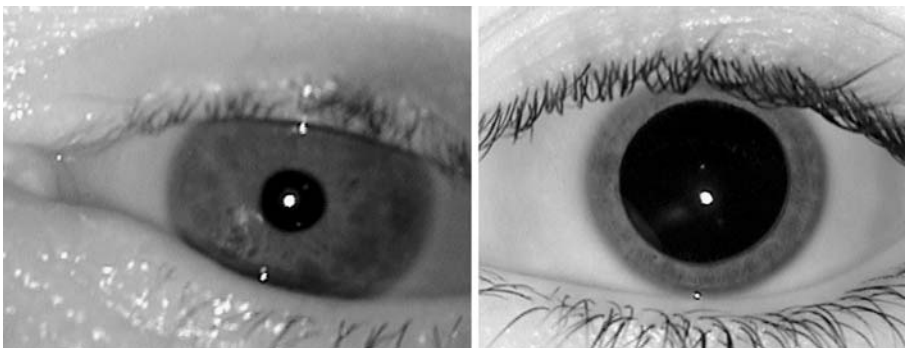

Fig. 3 Example irises with small and large pupil dilation ratio. The image on the left has a pupil dilation ratio of about 0.25 and the image on the right has a pupil dilation ratio of about 0.7. These images are of different irises
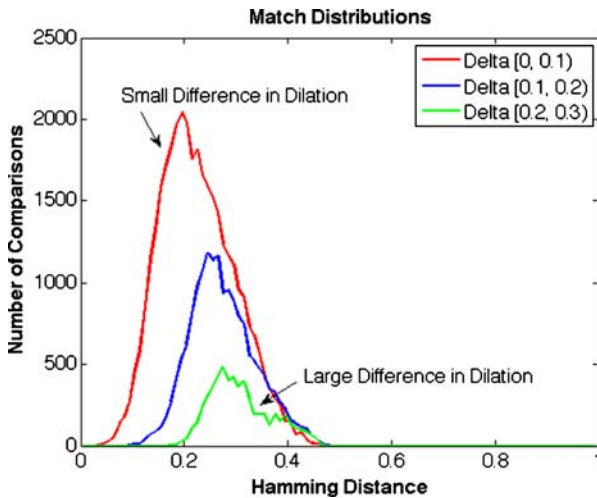
Fig. 4 Match distributions for small, medium and large difference in pupil dilation ratio between the two images. This figure is adapted from one originally published in (Hollingsworth et al. 2009b), "Pupil dilation degrades iris biometric performance", *Computer Vision and Image Understanding*, 113 (1), January 2009, pp. 150–157

While the ambient lighting level is typically the first factor thought of as causing pupil constriction or dilation, there are a number of other factors that also affect pupil dilation. It is known that the average degree of normal dilation decreases with age. There are drugs that can cause the pupil to constrict or dilate. A person's emotional state can cause the pupil to constrict or dilate. Even certain types of perceptual events cause involuntary pupil dilation. Einhauser et al. (2008) describe experiments in which a subject is exposed to an ambiguous visual or auditory stimulus and pupil dilation coincides with the moment of the subject's resolving the stimulus. Given all the factors that can affect pupil dilation, it seems that a simple strategy of a constant lighting level is not necessarily sufficient to ensure a given degree of pupil dilation.

It is likely reasonable in most scenarios to assume that one can control the degree of pupil dilation represented in the enrollment image. In this case, a relatively constricted pupil would be favored over a relatively dilated pupil. To our knowledge, no current commercial system stores the degree of pupil dilation as a parameter associated with an iris code. However, pupil dilation ratio is a parameter is a pending proposed ISO / IEC standard (ISO 2009). The degree of dilation associated with the two iris codes may be a useful element in determining the confidence in a match decision.

Effects of contact lenses on iris biometrics accuracy

"Successful identification can be made through eyeglasses and contact lenses..." (Negin et al. 2000)

"Eyeglasses and contact lenses do not reduce accuracy as long as the iris is clearly visible." (Weaver 2006)

"Iris recognition efficacy is rarely impeded by glasses or contact lenses." (Wiki 2009)

The above quotes are representative of the accepted truth that normal, as opposed to "cosmetic," contact lenses do not affect the accuracy of iris biometrics. However, recent results from our laboratory indicate that wearing standard prescription contact lenses can in fact degrade the accuracy of iris biometrics (Baker et al. 2009b). This example illustrates how segments of the population may experience different levels of user-friendliness if iris biometrics is deployed in a nation-scale application.

This set of experiments uses iris images from 48 subjects (96 irises) wearing contact lenses and 64 different subjects (128 irises) without contact lenses and not wearing glasses. For the 96 irises with contact lenses, we have approximately 2900 total images, and for the 128 irises without contact lenses, we have approximately 3700 total images. All images were visually inspected to confirm acceptable image quality and the presence or absence of a contact lens. These images were selected from the NS-IRIS-0405 dataset (Bowyer and Flynn 2009), a superset of the Iris Challenge Evaluation dataset (Phillips et al. 2009). Segmentation and matching was again performed using our own version of the IrisBEE software.

The match and non-match distributions were computed separately for the with-contacts set of images and the without-contacts set of images. Approximately 126,000 match scores went into the empirical match distribution for non-contacts and 69,000 into the match distribution for contacts. A total of 3.6 million matches went into the non-match distribution for with-contacts, and a total of 6.9 million went into the non-match distribution for without-contacts. We find that non-match distributions are nearly identical between the with-contacts group and the without-contacts group, but that the match distribution changes significantly between the groups, as shown in Fig. 5.

The percentage of the match distribution falling above a threshold of 0.32 was calculated for each distribution. For persons not wearing contacts, 0.27% of the distribution fell above 0.32. Oversimplifying a bit, this implies that the average person who does not wear contacts would experience about a 1 in 360 chance of a false reject when attempting to verify their identity. For the contact-wearers distribution, 5.64% of the distribution fell above 0.32. Thus, for the particular set
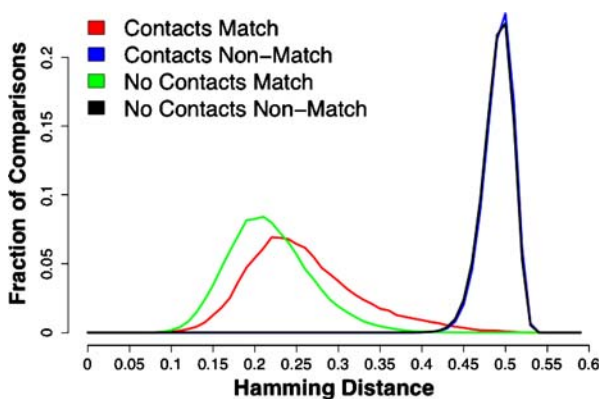

Fig. 5 Match and non-match distributions for with-contacts and without-contacts image sets. The match distribution changes in a way that indicates that contact lens wearers will experience a higher rate of false-non-match outcomes

of people and contact lenses in this study, contact lens wearers have an approximately 20 times greater likelihood of experiencing a false reject than those not wearing contacts. The number of subjects in this current study is too small, and the variety of normal prescription lenses is too large, to allow any confident projection of the size of this effect in the general population. Additional research on larger datasets is needed. Also, it is likely that different iris biometric systems may be affected to a different degree.

Contact lenses present a wide variety of effects in the iris image, ranging from visually obvious to relatively subtle. Rigid gas-permeable lenses, sometimes called "hard lenses," present obvious image artifacts. Cosmetic contact lenses also present obvious image artifacts. See part (a) of Fig. 6 for examples of these kinds of effects. However, the results in Fig. 5 do not include such lenses; they include only normal prescription lenses. Examples of normal prescription lenses are shown in part (b) of Fig. 6.

In a report on the feasibility of biometrics-enabled entitlement programs, Mansfield and Rejman-Greene (2003) noted the need to be concerned with glasses and contacts at the time of enrollment—"Removal of spectacles and designer contact lenses is clearly desirable." Our results suggest that it may be desirable to remove normal prescription contacts as well as cosmetic contacts, and to remove them at the time of verification as well as at the time of enrollment. However, while this may be desirable from a technical point of view, it may not a practical requirement for many application scenarios. It is likely more effective on average for contact lens wearers who experience a false non-accept outcome to simply repeat the process, or to use an alternate modality, than to ask all contact lens wearers to remove their contacts.

(a) Images with a cosmetic lens (left and middle) and a GP, or "hard," lens (right)

(b) Images with normal prescription contacts; see lettering and circle artifacts

Fig. 6 Examples of image artifacts due to various types of contact lenses

Template aging

> "... [the iris] is believed to be stable throughout life (barring accidents and surgical operations)" (EC 2005)
> "... the iris is highly stable over a person's lifetime ..." (Monro et al. 2007)
> "... a key advantage of iris recognition is its stability, or template longevity as, barring trauma, a single enrollment can last a lifetime." (Thornton et al. 2007)
> "... [the iris is] essentially stable over a lifetime" (Miyazawa et al. 2008)

Another "accepted truth" about iris biometrics is that "a single enrollment can last a lifetime." This expresses the idea that once a person is enrolled in the system, normal aging does not result in any noticeable change in the biometric template, and therefore a person would not normally need to be re-enrolled. Enrollment is naturally a more complicated step than recognition, because it is at enrollment that the person would prove their identity in an independent manner. Re-enrollment represents an added cost in administering an iris biometrics system (Mansfield and Rejman-Greene 2003), as well as an inconvenience to the user.

We have been acquiring iris image datasets at the University of Notre Dame since the spring of 2004. A small number of subjects have now had images acquired multiple times over a period of 4 years. This iris image dataset allows us to perform an experimental test of the stability of the iris biometric template over time (Baker et al. 2009a). For 23 persons, or 46 irises, we have a number of images collected from the spring of 2004 through 2008. The subjects ranged in age from 22 to 56 at the end of the 4-year period. Sixteen are male and seven are female, and sixteen are Caucasian and seven are Asian. The gender breakdown does not follow the ethnicity breakdown; the similar numbers are a coincidence.

The images used in this study were all acquired using the same LG 2200 iris imaging system. The segmentation and matching was done with our modified version of the IrisBEE software. For each iris, we compared the average fractional Hamming distance for matches made between images taken on different days but with less than 100 days time lapse, with the average fractional Hamming distance for matches made between images with at least 1,000 days time lapse. If there is no effective aging of the iris biometric template, then the average fractional Hamming distance for long-time-lapse matches should be greater than that for short-time-lapse matches only approximately 50% of the time. However, we found that 43 of the 46 irises had a larger average Hamming distance for the long-time-lapse matches. This is a statistically significant result (sign test, $p<0.0001$).

Another way of considering the results is that if there is no effective aging of the iris biometric template, then the mean (long-time-lapse average—short-time-lapse average) difference across irises should not be significantly different from zero. However, we found a mean difference of 0.019, which is statistically significantly greater than zero (paired t test, $p<0.00001$).

These results can be interpreted as saying that the user of an iris biometric system will experience an increase in the false non-match rate with increasing time lapse from enrollment. Our results indicate that the increase is relatively minor for a 4-year time lapse, arising from only about a 0.02 increase in average Hamming distance. Because our results showing a template aging effect for iris biometrics have been

controversial, additional research by other research groups is desirable, preferably for larger datasets and representing even longer time lapse.

Interoperability between systems

As biometrics are used in larger-scale applications and over longer periods of time, it becomes increasingly likely that some persons will be enrolled in the system using one model of sensor and then have their identity verified with a different model of sensor. The obvious interoperability issue is that the two systems must use the same representation of iris texture, or be able to convert between representations. But even for systems that use the same algorithmic approach to generating the iris code, there is still an issue of whether or not performance is degraded with cross-sensor matching.

An example of this issue arises in the context of a technology upgrade. People have been enrolled in the biometric application using the older model of sensor, and need to have their identity verified using the newer model. To investigate what can happen in this context, we acquired a set of iris images using both an LG 2200 system and an LG 4000 system (LG 2009b) for the same set of subjects. The LG 2200 system was once state-of-the-art but is now a discontinued model. The LG 4000 is a current state-of-the-art commercial system. (See Fig. 7.) The two systems are technically interoperable in the sense that iris codes generated by one system can be readily matched against iris codes generated by the other system.

Three potentially important differences between the LG 2200 and LG 4000 sensors are (1) the location of the near-infrared illumination relative to the eye, (2) the field of view, and (3) the camera technology. The LG 2200 cycles the illumination between three LEDs, positioned to the top, the lower left, and the lower right of the eye. The LG 4000 uses LEDs at fixed positions to the side of the eye. Thus the geometric relation between the illuminator and the iris is different between the systems. Also, the LG 2200 acquires an image of one iris at a time, whereas the LG 4000 acquires images of both irises at the same time. This makes it possible, in principle, to create the iris code for a standard orientation of the eye, as mentioned at the end of "Overview of iris biometrics technology". Lastly, the LG 4000 acquires

Fig. 7 LG 2200 (*left*) and LG 4000 (*right*) iris sensors. The 2200 system has near-IR LEDs at the top, lower left and lower right, and images one iris at a time. The 4000 has near-IR LEDs on the left and the right, and can image both irises at once

digital images whereas the LG 2200 digitizes an image from an analog video signal. Thus the LG 4000 will be less susceptible to image artifacts such as interlacing.

In Fig. 8, we show experimental match and non-match distributions for images taken by the LG 2200 and the LG 4000 for the same set of irises. The images used meet the normal built-in quality checks of the respective sensor. In the results for the LG 2200, we only use images in which the iris is illuminated by either the left or the right LED, and not the top LED. We have found that this results in better recognition performance for the image datasets acquired in our laboratory. The image analysis and matching were again done with our modified IrisBEE software. A simple contrast enhancement step was applied uniformly to all LG 2200 and LG 4000 images before processing with our IrisBEE software. (Subjectively, this contrast enhancement changes the LG 4000 images more than the LG 2200 images, due to the LG 2200 having a built-in contrast enhancement step.) Even with the restriction on the illuminator to improve the LG 2200 results, note that the LG 4000 shows better recognition performance than the LG 2200. This is evident in a decidability index, or d', of 4.97 for the LG 4000 data, versus 4.27 for the LG 2200 data. The d' statistic is a measure of the separation between the match and non-match distributions, with larger values representing greater separation and therefore greater power for recognition. The distributions in this case clearly have some deviations from Gaussian, and so it is not entirely appropriate to use the d' metric to summarize the separation of the match and non-match distributions. However, for the case of comparing performance on the same sets of subjects across different sensors, we might still use d' to indicate relative changes.

This experiment realistically models the expectations of a technology upgrade scenario. The newer sensor, the LG 4000, represents a measurable improvement over the older sensor, the LG 2200. The complication occurs in cross-sensor operation. What sort of performance can be expected when persons who enrolled in the biometric application using the old sensor attempt to verify their identity using the new sensor?

To investigate this, we create the match and non-match distributions for matching the LG 4000 images against the LG 2200 images. These results appear in part (c) of Fig. 8. It is perhaps not unexpected that the performance is poorer than if enrollment and verification were both done using the LG 4000. However, the performance is also poorer than if enrollment and verification were both done using the LG 2200! In other words, the cross-sensor performance yields a (small) step backward in performance rather than a step forward.

The cross-sensor performance degradation occurs primarily due to a shift in the match distribution, rather than the non-match distribution. Assuming other parameters of the application are kept constant, we can think of this as meaning that the person attempting to verify their identity using the LG 4000 based on an enrollment with the LG 2200 will experience a higher rate of false non-matches than if they had also enrolled with the LG 4000. That is, they will have to re-try their identity verification more often in the cross-sensor scenario.
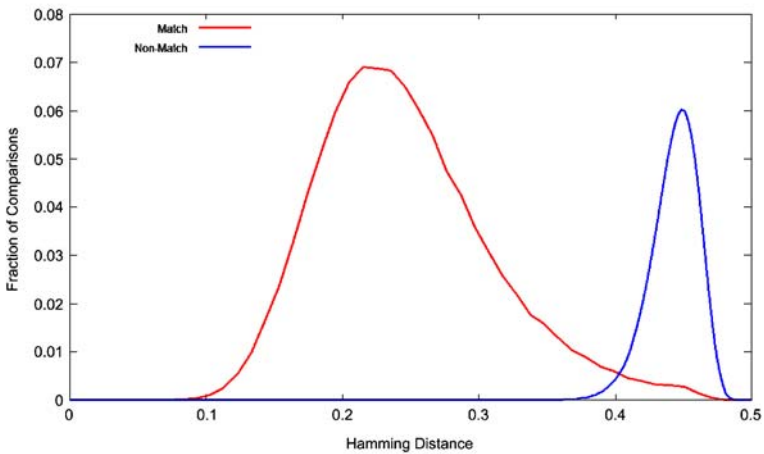
Discussion

Our results show that iris biometric performance can be degraded by varying pupil dilation, by wearing non-cosmetic prescription contact lenses, by time lapse between
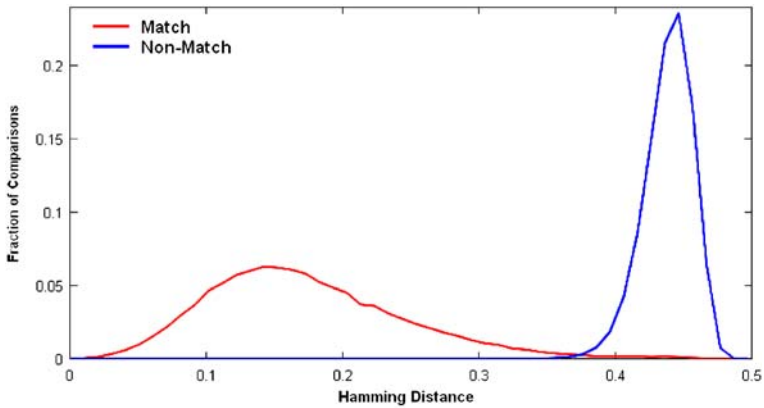
a Match and non-match distributions for the LG 4000.
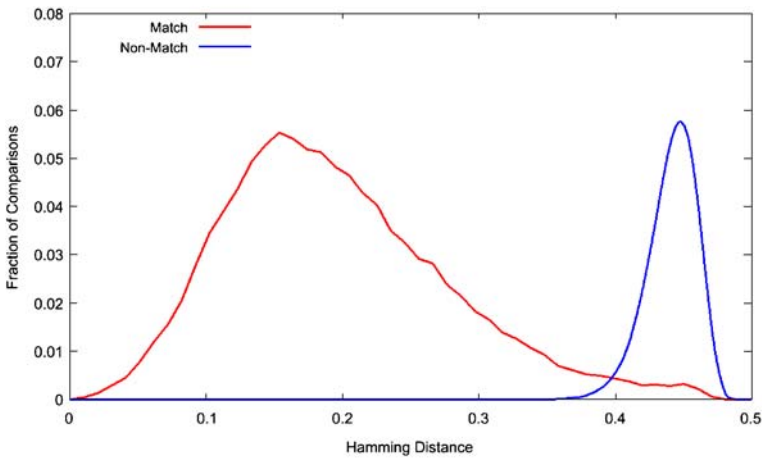
b Match and non-match distributions for the LG 2200.

c Match and non-match distributions for cross-sensor matching.

Fig. 8 Match and non-match distributions for same-sensor and cross-sensor matching scenarios

enrollment and verification, and by cross-sensor operation. However, the most surprising theme running through these experiments is perhaps not that factors exist that can degrade iris biometrics performance, but that all of the factors identified here affect primarily the match distribution. For an identity verification scenario, this is important. Oversimplifying, it means that while the user-friendliness of the system can degrade due to various factors, the security of the system is not affected. For a "watch list" scenario, however, our results suggest various means that a wanted person might use to try to evade detection. In either case, there is the possibility that future research could substantially alleviate the problems that we have identified.

Limitations to our experimental results

There are potential limitations regarding the experimental results presented here. One is that the results are based on relatively small data sets. Results obtained on larger data sets would naturally allow greater confidence. Another potential limitation is that the results are obtained using one particular iris biometric implementation, IrisBEE. The advantage of this particular implementation is that the source code is available and so it is not a "black box". But it is important to determine if the same basic pattern of results holds for a different implementation. We are currently using the commercial VeriEye software (Neurotechnology 2009) to run experiments parallel to some of those presented here. It is also true that most of the image datasets used in the experiments described here were acquired with an LG 2200 system, which the manufacturer has replaced with a newer model. Our experience with other iris image acquisition systems does not suggest that any of the effects documented here are dependent on the image acquisition system.

Stability of the non-match distribution

Each of the factors considered in this paper degrades iris biometrics performance by changing the match distribution. The non-match distribution, on the other hand, appears to be stable in the face of the factors that we have considered here. The one instance that we identified in which the non-match distribution is degraded is when a person's pupil is highly dilated in both their enrollment image and their image to be recognized. Since in most applications the degree of pupil dilation at enrollment can be controlled to some degree, the practical significance of this instance seems limited. Thus our results can be seen as supporting the premise that an iris biometrics system can be set to operate at a fixed threshold of a "one in 1.2 million chance of a false match." The chances of a false non-match outcome may, of course, shift according to the various factors explored here.

It is perhaps worth noting that this conceptual analysis of the chances of a false match outcome assumes a "zero-effort imposter" (Bolle et al. 2004). That is, the estimate of the false match rate is based on a random selection of an imposter, and on the imposter not actively modifying their biometric in an attempt to defeat the system. It is not meant to model the situation in which a best imposter is selected from among a set of possible imposters, or in which an imposter actively attempts to modify their biometric in some way.

Watch list scenario versus verification scenario

We have assumed an identity verification scenario for most of the discussion in this paper. In a watch list scenario, a "watch list" is constructed for a list of people who should be denied access, and perhaps detained if they are detected. The two types of errors that can occur are a false positive and a false negative. A false positive occurs when a person who is not on the watch list is said to match someone on the watch list and so is falsely identified as a wanted person. A false negative occurs when someone who is on the watch list, and so should be detained, is not detected by the system. For a watch list scenario, the factors identified in this paper suggest ways that a wanted person might try to fool the system into a false negative result and so evade detection. That is, someone who is on a watch list and wants to evade detection increases their probability of evading detection by having a large difference in pupil dilation, by wearing contacts, by a longer time lapse since enrollment, or by a cross-sensor matching.

The enrollment stage of an identity verification system has some similarity to a watch list scenario. In order to detect attempted multiple enrollments, each new enrollment can be compared against all existing enrollments. In a sense, this check for a match to an already-existing enrollment is like a watch list scenario. However, the situation for multiple-enrollment fraud is more complicated than the situation for evading detection in a watch list scenario. To commit multiple-enrollment fraud, the person perpetrating the fraud would presumably want to be able to present as any selected one of the multiple enrollments. Our results do not suggest that pupil dilation, wearing normal prescription contacts, template aging or sensor interoperability can be well exploited for this purpose.

Future research topics

For each of the problems documented in this paper, there are lines of research that could potentially reduce or resolve the problem. We briefly suggest some of these possibilities.

The problem of varying pupil dilation can potentially be addressed in several ways. One possibility is that researchers will develop more sophisticated models for transforming the segmented iris region into a standard size frame. The current popular "rubber sheet" model uses linear interpolation in the radial direction to normalize unwrapped iris size, and it is possible that some higher-order or adaptive interpolation scheme would result better performance when matching irises with a large difference in dilation ratios. There is some initial research in this direction (Thornton et al. 2007; Wei et al. 2007). In the interim, until such algorithms can be developed and validated, it makes sense to record the pupil dilation factor as meta-data associated with an iris code (ISO 2009). This would make it possible to know when a match is attempted between irises with a large difference in pupil dilation ratio.

Current iris biometrics systems are able to detect some forms of cosmetic contact lenses (Daugman 2003). It may be possible to develop image analysis algorithms to detect the presence of normal prescription contact lenses, and / or to develop a means of detecting the distortions caused by contact lenses.

The effects of template aging can potentially be managed by setting up an appropriate schedule for re-enrollment. Therefore it should not be a barrier in principle to application of iris biometrics. It is also possible for the iris biometric

system to know the time lapse since enrollment when a verification attempt is made, and this can be factored into a certainty measure for the decision. In addition, it is conceivable that further research will suggest a model for how iris biometric templates age, and / or suggest a method of dealing more directly with the problem. For example, it is conceivable that the bits of the iris code that age are spatially related to the bits of the iris code that are masked out as being inconsistent (Hollingsworth et al. 2009a). If this was the case, then a transformation of the mask for inconsistent bits might be able to locate those bits that are likely to change from aging. Template aging is known to occur with other biometrics, such as face (Phillips et al. 2003), but we are not aware of any study that looks at the magnitude of template aging effects across different biometric modalities.

One method of addressing the problem of cross-sensor operation is to track the sensor from which the iris code is generated and take the pair of sensors represented in a match into account. To the degree that the degradation in cross-sensor operation is caused by variation in the illuminator design across systems, it may be possible to address this through standards.

Other population-level issues

There are various medical conditions that affect iris biometrics in some way that will need to be dealt with in an application that serves all citizens of a country. For example, a person with the condition of aniridia (not having an iris) would most likely not be able to use an iris biometrics system. Mansfield and Rejman-Greene (2003) suggest that "... perhaps 1 in 10,000 people do not have an iris that can be used for iris recognition." Also, there is one study that suggests that at least some people having cataract surgery would need to be re-enrolled after the surgery (Roizenblatt et al. 2004).

References

Baker S, Bowyer KW, Flynn PJ. Empirical evidence for correct iris match score degradation with increased time lapse between gallery and probe images. Lect Notes Comput Sci. 2009a;5558:1170–9.
Baker S, Bowyer KW, Flynn PJ. Contact lenses: handle with care for iris recognition. Proceedings of the Third Int Conf on Biometrics: Theory, Applications and Systems (BTAS 09), September 2009b.
Bolle RM, Connell JH, Pankanti S, Ratha N, Senior AW. Guide to biometrics, Springer, 2004.
Bowyer KW, Hollingsworth K, Flynn PJ. Image understanding for iris biometrics: a survey. Comput Vis Image Underst. 2008;110(2):281–307.
Bowyer KW, Flynn PJ. The ND-IRIS-0405 iris image dataset University of Notre Dame Technical Report. at: http://www.nd.edu/~cvrl/CVRL/Data_Sets.html, 2009 (accessed Nov. 29, 2009).

Daugman J. High confidence visual recognition of persons by a test of statistical independence. IEEE Trans Pattern Anal Mach Intell. 1993;15(11):1148–61.

Daugman J. Biometric decision landscapes, UCAM-CL-TR #482. At: http://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-482.pdf, 2000 (accessed Nov. 29, 2009).

Daugman J. Demodulation by complex-valued wavelets for stochastic pattern recognition. Int J of Wavelets, Multiresolution Inform Process. 2003;1(1):1–17.

Daugman J. How iris recognition works. IEEE Trans Circuits Syst Video Technol. 2004;14(1):21–30.

Daugman J. Probing the uniqueness and randomness of IrisCodes: results from 200 billion iris pair comparisons. Proceedings of the IEEE. 2006;94(11):1927–35.

Daugman J. New methods in iris recognition. IEEE Trans Syst Man, Cybern B. 2007;37(5):1167–75.

EC. Biometrics at the frontiers: assessing the impact on society. European Commission Joint Research Centre (DG JRC), Institute for Prospective Technological Studies. At: http://www.jrc.es, 2005.

Einhauser W, Stout J, Koch C, Carter C. Pupil dilation reflects perceptual selection and predicts subsequent stability in perceptual rivalry. Proc Natl Acad Sci. 2008;105(5):1704–9.

Hollingsworth K, Bowyer KW, Flynn PJ. The importance of small pupils: a study of how pupil dilation degrades iris biometric performance. Proceedings of the Second IEEE Int Conf on Biometrics: Theory, Applications and Systems (BTAS 08). At doi:10.1109/BTAS.2008.4699341, 2008.

Hollingsworth K, Bowyer KW, Flynn PJ. The best bits in an iris code. IEEE Trans Pattern Anal Mach Intell. 2009a;31(6):964–73.

Hollingsworth K, Bowyer KW, Flynn PJ. Pupil dilation degrades iris biometric performance. Comput Vis Image Underst. 2009b;113(1):150–7.

ISO. ISO / IEC 19794-6 Iris data interchange standard. At: http://webstore.iec.ch/preview/info_isoiec19794-6%7Bed1.0%7Den.pdf, 2009 (accessed Nov. 29, 2009).

LG. LG Iris products and solutions: LG 2200. At: http://www.lgiris.com/ps/products/previousmodels/irisaccess2200.htm, 2009a (accessed Nov. 29, 2009).

LG. LG Iris products and solutions: LG 4000. At: http://www.lgiris.com/ps/products/irisaccess4000.htm, 2009b (accessed Nov. 29, 2009).

Liu X, Bowyer KW, Flynn PJ. Experiments with an improved iris segmentation algorithm. Fourth IEEE Workshop on Automatic Identification Technologies (AutoID). October 2005, pp. 118–123.

Mansfield T, Rejman-Greene M. Feasibility study on the use of biometrics in an entitlement scheme. UK National Physical Laboratory technical report. At: http://dematerialisedid.com/PDFs/feasibility_study031111_v2.pdf, 2003 (accessed Nov. 29, 2009).

Miyazawa K, Ito K, Aoki T, Kobayashi K, Nakajima H. An effective approach for iris recognition using phase-based image matching. IEEE Trans Pattern Anal Mach Intell. 2008;30(10):1741–56.

Monro DM, Rakshit S, Zhang D. DCT-based iris recognition. IEEE Trans Pattern Anal Mach Intell. 2007;29(4):586–95.

Negin M, Chmielewski TA, Salganicoff M, von Seelen UM, Venetainer PL, Zhang GG. An iris biometric system for public and personal use. IEEE Computer. 2000;33(2):70–5.

Neurotechnology, VeriEye SDK, November 2009. At: http://www.neurotechnology.com/verieye.html (accessed Nov. 29, 2009).

Phillips PJ, Grother P, Michaels RJ, Blackburn DM, Tabassi E, Bone M. The Face Recognition Vendor Test 2002 Evaluation Report. March 2003. At: http://www.frvt.org/DLs/FRVT_2002_Evaluation_Report.pdf (accessed Nov. 29, 2009).

Phillips PJ, Scruggs WT, O'Toole AJ, Flynn PJ, Bowyer KW, Schott CL et al. FRVT 2006 and ICE 2006 large-scale experimental results. IEEE Trans on Pattern Analysis and Machine Intelligence. 2009; to appear. doi:10.1109/TPAMI.2009.59.

Roizenblatt R, Schor P, Dante F, Roizenblatt J, Belfort R. Iris recognition as a biometric method after cataract surgery. Biomed Eng Online. 2004;3(1):2.

Thornton J, Savvides M, Vijaya Kumar BVK. A Bayesian approach to deformed pattern matching of images. IEEE Trans Pattern Anal Mach Intell. 2007;29(4):596–606.

Weaver AC. How things work: biometric authentication. IEEE Computer. 2006;39(2):96–7.

Wei Z, Tan T, Sun Z. Nonlinear iris deformation correction based on Gaussian model. Lect Notes Comput Sci. 2007;4642:780–9.

Wiki. Iris recognition. In Wikipedia. At: http://en.wikipedia.org/wiki/Iris_recognition, 2009 (accessed Nov. 29, 2009).