IREX III

Supplement I: Failure Analysis

NIST Interagency Report 7853

G. W. Quinn and P. Grother

Information Access Division
National Institute of Standards and Technology

April 19, 2012

IREX III Supplement I: Failure Analysis

Contents

1 Introduction

Iris recognition has the potential to be extremely accurate, but like many other biometrics, it is highly dependent on the quality of the sample data. Occlusion of the iris, off-axis gazes, blurred images, and iris rotations are common problems that can make recognizing individuals more difficult. A fraction of the Iris Exchange (IREX) III dataset consisted of poor quality images that reduced overall performance in the evaluation. For single eye searches, the matching algorithms were never able to achieve an identification rate better than 98.3% at any reasonable decision threshold. It is fair to assume that some large fraction of those missed searches are the result of poor quality on the part of either the search image or its corresponding mate in the enrollment set.

The poor quality images often stem from systemic problems with the capture and data handling process. The purpose of this failure analysis is to identify common sources of poor sample quality and to provide best practice recommendations for how to improve the quality of captured samples. The analysis began with the construction of a special ”failure set” of images, which contained those images that were a common source of error for several matching algorithms. Strictly speaking, the failure set contained a set of image pairs, where each pair consisted of a search image and a corresponding mate that was used for enrollment. A thorough manual inspection was then performed on each image in order to identify specific cause(s) of failure. Section 2 describes the construction of the failure set and manual inspection process. Section 3 explores each of the common sources of poor sample quality and provides recommendations on how to adjust the capture protocol or data management practices to reduce the likelihood of their occurrence. Appendix A summarizes the best practice recommendations provided in the previous subsections.

2 Methodology

2.1 Construction of the Failure Set

Approximately 17,000 iris images were randomly selected and searched against an enrollment set of 1.6 million images of 1.6 million subjects. Each search image was known to have an enrolled mate, based on the ground-truth information that was available for the dataset. Matching was performed using 8 algorithms selected from the 8 better performing participants. The algorithms were the latest Class A submissions (i.e. the submissions that focused on speed) at the time that the failure set was built. The failure set consists of those searches that were missed at rank 20 by at least 4 of these matching algorithms. The failure set contains 1,013 pairs of images (1,013 search images with 1,013 corresponding mates), which comprise about 6.0% of the 17,017 searches performed. Although each search in the failure set was missed by at least 4 algorithms, most were still matched by at least one. Only 240 (or about 1.4%) of the searches were missed by all algorithms. Upon further inspection, nineteen (0.1%) of these were the result of ground truth errors.

2.2 Manual Inspection Process

Each of the images in the failure set was manually inspected by an individual that has extensive experience with studying the iris biometric. The inspector was tasked with identifying specific problems with each image (e.g. eyelid occlusion, compression artifacts, rotation, ground truth error). The inspector used specific criteria when possible, but at times it was necessary to use subjective judgement when deciding whether an image suffers from a particular problem. The criteria used by the inspector are documented later in this report. Interesting or unusual cases were loaded into a specialized investigational software suite called the IrisLab Toolset. Some of the figures in Section 3 are snapshots of the tool as it was operating on images (see Figure 1). The top row in each of the snapshots shows the unaltered iris images. The bottom row displays segmentation overlays, when available, generated by a commercial ”off-the-shelf” algorithm.

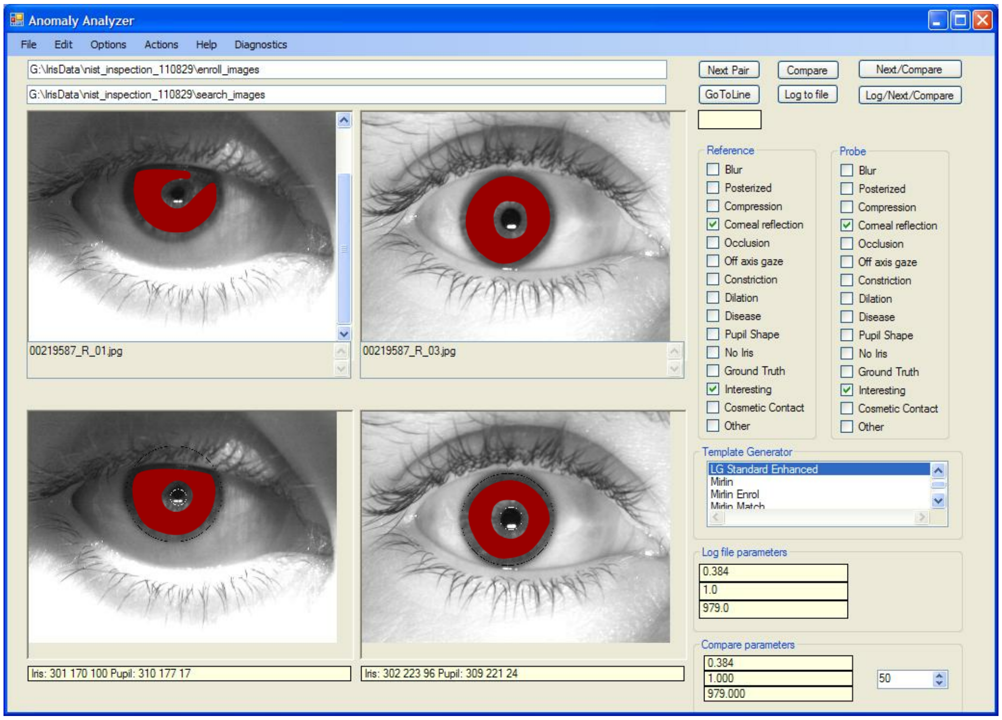


Figure 1: Example snapshot of the IrisLab Toolset.

2.3 A note on Recommendations

Many of our recommendations require additional involvement of the camera operator in the capture process. Ideally, the operator will receive training on how to identify poor quality images and how to avoid their capture. While some recommendations are common sense (e.g. keep the subject within the focal range of the camera), others require specific knowledge of how iris recognition works. Camera developers can assume some of this responsibility by implementing improvements to their capture devices. Examples would be to improve automated quality filtering or provide additional feedback to the operator when there is a problem with the scene (e.g. if the subject is standing too close). Since camera operators rarely receive extensive training in the collection of biometrics, instructions to the operators should be clear and brief.

3 Sources of Poor Image Quality

Quality problems can occur for several reasons. Sometimes presentation by the subject is poor (e.g. a closed eye); sometimes the camera is damaged (e.g. burnt out LEDs); and sometimes the capture environment introduces problems (e.g. reflections in the eye due to excessive ambient illumination). These examples occur at the capture stage, but other problems can occur during subsequent processing and storage of the images. For example, an image might be severely compressed or corrupted when it is transferred over a network. This section organizes the various causes of poor sample quality into three categories: 1) those attributable to poor presentation, 2) those that result from problems with the capture system or subsequent processing and storage, and 3) those that are due to unusual characteristics inherent in the individual (e.g. congenital iris defects). In reference to the ISO/IEC 29794-2 standard on biometric sample quality [7], the first two categories affect the *fidelity* of the sample, while the last relates to the *character* of the source from which the sample is derived. Some of the categorizations are ”soft” in the sense that reasonable people might disagree with respect to which category a particular image falls into. An image may also suffer from problems in multiple categories, in which case it will fall into more than one category.

Type	Problem	Number of Images	Number of (paired) Searches
Poor Presentation	Occluded Iris	236	214
	Rotated Iris	244	219
	Off-axis Gaze	39	37
	Specular Reflections	304	280
	Highly Dilated	44	38
	Highly Constricted	84	79
	Blurred	91	88
	Patterned Contact	6	6
	Not an Iris	4	4
System Problems	Ground Truth Bug	38	19
	Compression Artifacts	743	557
	Reduced Color Space	50	47
Poor Inherent Features	Abnormal Pupil Shape	45	36
	Total	1928	1624

Table 1: Number of times each problem was encountered during the manual inspection.

Number of Problems	Number of Images
0	505
1	1161
2	329
3	35
4	1
5	0

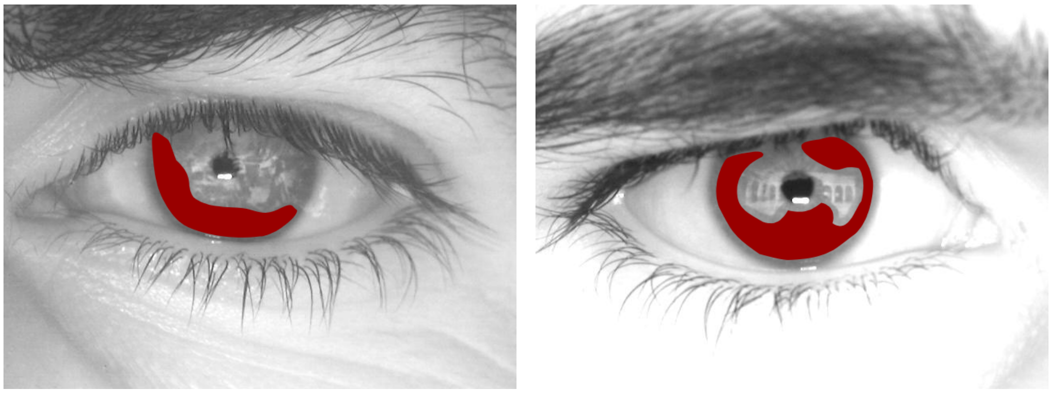
Table 2: Number of times an image had a specific number of problems.

Table 1 lists most of the problems identified during the manual inspection of the images as well as their frequency of occurrence. A detailed description of each of these problems is provided in later sections. The right-most column lists the number of searches where at least one of the paired images had the listed problem. Many of the images suffer from multiple problems, which is why the sum of the number of problems encountered is greater than the number of images. There were an average of 1.90 problems per image. The table shows that compression artifacts are the most common problem. More than half of the searches in the failure set involve at least one image that is highly compressed. The second most common problem is specular reflections (i.e. Purkinje images), which were conspicuous in 280 searches. Although specular reflections adversely affect recognition, by themselves they were typically not enough to cause a search to miss. Only 54 of the searches in the failure set were marked as having specular reflections as the sole problem.

Table 2 shows how often a search from the failure set suffers from a specific number of problems. If both the search image and the enrolled image suffer from the same problem, it was counted twice. Twenty-seven searches were not marked as having any problems, but this does not necessarily mean that the images involved in these searches were good quality. Problems such as burnt out LEDs and Bayer filter effects could not always be identified by the inspector and are therefore not represented in the table. 389 searches involved only one iris image that suffered from a marked problem while for 597 searches, both images were marked as having a problem. The fact that the greater dataset containing two million subjects did not contain any detected cases of aniridia (the absence of an iris) suggests that either the images, capture protocol, or sampling method, involved some type of filtering.

3.1 Poor Presentation

3.1.1 Reflective Glare



(a) (b)

Figure 2:

Specular reflections of the background scene could be seen in many of the iris images. These reflections, known as Purkinje images, are especially apparent when a large fraction of the light that is reflected off of the eye originates from the scene in front of the subject. An image of an eye can contain up to four Purkinje images (i.e. four kinds), where each image is created by light reflecting off of a different surface of the eye. The reflections that can be seen in the iris images are Purkinje images of the first kind because they are caused by light reflecting off of the anterior surface of the cornea. First Purkinje images are quite common in the overall dataset, although in most cases their presence did not appear to cause much difficulty for the automated matchers. The failure set contains 304 images involved in 280 searches where Purkinje images are very conspicuous (examples are shown in Figures 2(a) and 2(b)). For 160 of the searches, the presence of Purkinje images was the only problem identified with either image, suggesting that they may be the primary reason for the miss.

Purkinje images create false features on the surface of the iris that are difficult for automated matchers to distinguish from true iris features. The ideal way to avoid capturing an image with Purkinje artifacts is to ensure that the camera is the dominant source of illumination. Fluorescent bulbs and LED lamps emit less light in the infrared spectrum and are therefore less likely to introduce Purkinje images. In an outdoor environment the ambient illumination is rarely under control, and little can be done to avoid capturing images with Purkinje images. Since the reflections are typically of the scene in front of the subject, the camera operator can position the subject to face a uniform and featureless background. The operator can also position himself such that the sun is to his back to reduce the amount of solar illumination that reflects off of the scene and onto the eye (although admittedly these options may sometimes be mutually exclusive). Some iris cameras extend around the eye(s) to block ambient illumination.

Purkinje images sometimes reveal information about the environment in which the image was captured. For example, the reflections on the iris in Figure 2(b) reveal that the image was captured outdoors in front of an archway, and that the operator was using two hands to hold the camera. None of the Purkinje images in the data analyzed show enough detail to cause privacy concerns (e.g. they never show distinct facial features or large amounts of fingerprint area).

Recommendations: The iris camera should be the dominant source of illumination. Operators should be trained to reduce ambient sources of illumination when possible and to recognize the presence of specular reflections of the ambient environment. To minimize specular reflections in an outdoor environment, the camera operator should position himself such that his back is to the sun and the subject is facing a featureless scene.

(a) (b)

Figure 3: Examples of images with rotated irides.

3.1.2 Rotation

The failure set contains 244 images where the iris is highly rotated. The iris was deemed highly rotated if the rotation is greater than roughly 20 to 25 degrees according to a visual assessment. For 126 searches, rotation was the only problem identified with either image. High amounts of rotation occur when either the camera is not held fully upright or the subject is not positioned properly relative to the camera. Most of the iris recognition algorithms compare images over a limited range of possible rotations. If the difference in rotation for a mated comparison is beyond this range, then the iris images will fail to match. One way to solve this problem is to expand the range of rotational disparities that are tested, but doing so increases the probability that a non-mated comparison will result in a false match. Daugman [3] predicts the probability of a false match as

$$F_n(x) = 1 - [1 - F_1(x)]^n \tag{1}$$

where n represents the number of orientations tested, and $F_1(x)$ represents the probability of a false match for any single orientation at threshold x. When the false match rate is low, the relationship is approximately linear:

$$F_n(x) \approx nF_1(x) \tag{2}$$

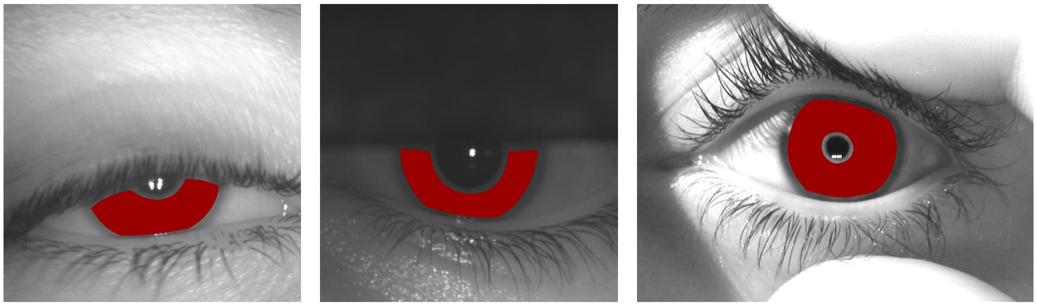
Varying the value for n introduces a trade-off, where higher values increase the probability that an individual will be recognized, but also increase the probability that a non-mated comparison will result in a false match. Higher values of n may also increase the time it takes to perform a comparison since more orientations are tested.

Most of the highly rotated iris images were captured by one-eye cameras. Two-eye cameras require the operator to hold the camera approximately parallel to the subject's face so that both eyes can be captured at once (this is more true for binocular style camera than single-lens ones). They also enable easier measurement of the rotation that does exist from an analysis of the orientation of the inter-pupilary line. Thus, a correction of the rotation is possible. This does not prevent the operator from holding the camera upside down, but it does appear to eliminate high amounts of rotation in properly acquired images. In any case, a rotated iris can be manually "unrotated" post-capture.

Recommendations: A single-eye camera is much more likely to capture a highly rotated iris than a dual eye camera – and introduces the potential for mislabeling the eyes (your right or mine). Hence, dual eye cameras are preferred, all other things being equal. For cases in which a single eye camera is used, rotation can be dealt with in various ways. In order of preference: 1) rotation can be minimized during the capture phase by ensuring that the camera is held upright and the subject's head is not tilted, 2) if the image is already captured and stored, the iris can be manually "unrotated", 3) the matching algorithm can be configured to compare iris templates over a more relaxed range of rotation disparities.

There is some discussion in the literature on automated methods for measuring eye rotation in an image [13], although none of the quality algorithms submitted to IREX II [19] attempted to measure it. If single eye cameras must be used, these methods should be explored.

3.1.3 Occlusion



(a) (b) (c)

Figure 4: Examples of occluded irides.

Occlusion is one of the most common problems, which was found in 236 images in the failure set. An iris was deemed highly occluded if the intrusion extends into the pupil region. For 108 of the searches, occlusion is the only problem identified in either image. The occlusion typically comes from the upper eyelid and is often accompanied by further occlusion and shadowing from the eyelashes. Other sources of occlusion are the lower eyelids, fingers, and parts of the iris camera (see Figure 4(b)). Hotspots (i.e. diffuse reflections) also occlude parts of the iris, but are not addressed in this section.

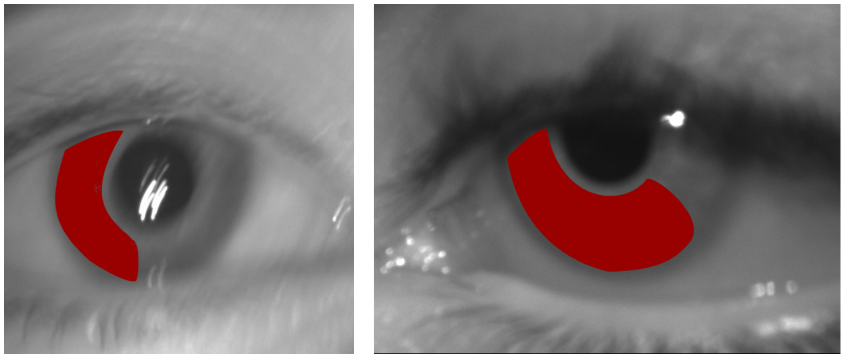
Occlusion introduces two difficulties for automated matching: 1) it makes correct segmentation of the iris more challenging, and 2) it obscures useful portions of the iris texture. The IREX I and IREX II reports both determined that high amounts of occlusion increase mated comparison scores, and sometimes decrease non-mated scores. Iris recognition algorithms typically mask out occluded parts of the iris before generating a template. Toward this end, several methods of occlusion detection have been proposed in the literature [4, 8, 15].

Although images with a high amount of occlusion are common in the failure set, the vast majority of images from the overall dataset contain very little occlusion. In many of the images, the eyes were held open by a pair of fingers (see Figure 4(c)). Most likely this was done to get the acquired image to pass the internal quality checks of the iris camera. Many iris cameras perform real-time quality assessment of the images they capture and will not return an image if a high amount of occlusion is detected.

Recommendations: Operators should be trained to instruct their subjects to open their eyes wide when their irides are captured. IREX II found that automated algorithms for measuring the amount of useable iris area have potential. If future cameras are designed with built-in occlusion detection, they could provide direction to operators to reacquire images that are highly occluded.

3.1.4 Blur

The failure set contains 91 images involved in 88 searches where the iris is noticeably blurred. Examples are shown in Figure 5. Blur occurs when either 1) the subject's eye is outside of the ideal focal range, or 2) the eye or the camera is moving when the image is acquired. While IQCE found that motion blur and defocus blur have similar effects on recognition accuracy (and thus do not require separate metrics in the ISO/IEC 29794-1 standard), it's still useful to distinguish between the two to determine the underlying cause. Motion blur is more likely to occur if the camera has a long exposure time. It can be difficult to visually distinguish motion blur from out-of-focus blur. One way of identifying motion blur is to see if the hotspots in the image appear as streaks, as is the case in Figure 5(a).



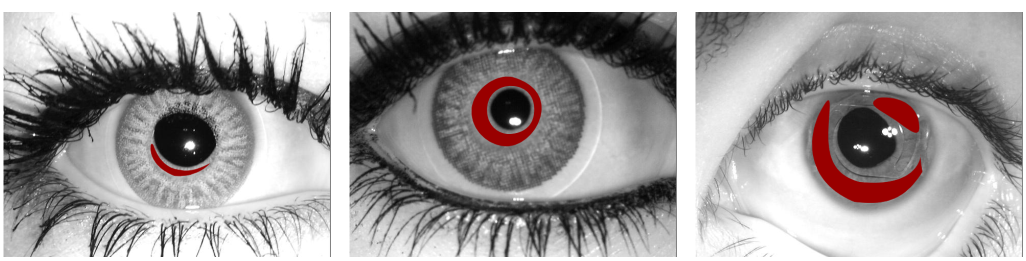
(a) (b)

Figure 5: Examples of blurry images from the failure set.

IREX II found that blur detrimentally affects recognition accuracy. One prior study [16] showed that recognition algorithms may be able to tolerate moderate amounts of blur in the absence of other problems. However, blur was rarely the only problem with the images in the failure set. For example, 59 of the 91 blurry images (almost two-thirds) contain an iris that is also highly occluded. Although a blurry image is easy to detect, little can be done to correct for the blur once the image has been captured and stored.

Recommendations: Operator training should include examples of blurred images so that they understand this issue. They should be instructed to hold the camera as steady as possible when capturing images and to keep the subject within the capture volume of the camera, which the operator must know. Operators should instruct subjects to stand still and look straight ahead during the capture. Several automated methods for measuring blur in images exist, many of which were evaluated as part of IREX II. There may be a benefit to incorporating automatic blur detection into iris cameras as a way to provide real-time feedback to the operator.

3.1.5 Contact Lenses



(a) (b) (c)

Figure 6: Examples of irides obscured by patterned contact lenses (a and b) and a crinkled transparent lens (c).

Six iris images in the failure set contain patterned contact lenses. In one case, a patterned contact is present in both the searched image and the enrolled image. The lenses were easy to spot because their patterns were printed in a dot-matrix format that is quite distinct from natural iris patterns (see Figures 6(a) and 6(b)). These fake patterns obscure the true features of the iris and make it extremely difficult, if not impossible, to accurately match the iris.

The problem of patterned contact lenses became apparent during the ICE 2006 Evaluation [14] when it was noticed that some of the images in the challenge sets contained similar dot-matrix styled patterned contacts. The ideal way to handle patterned contact lenses is to detect them before the iris is captured so that the subject can be instructed to remove them.

Daugman [3] suggests that contacts with the dot-matrix pattern can be detected by searching for areas of spurious energy in the Fourier plane that correspond to the sizes and periodicities of the dots. Lee and Park [9] propose a method of detecting fake iris patterns based on their 3D structure, but it requires the use of a specialized camera designed for the purpose. Since most patterned contacts are mass produced, it is not beyond possibility that two subjects wearing the same particular model of contact lens might, at some point, be compared to one another. No studies have been published on whether this ever results in a low comparison score (the one case in this study resulted in a high dissimilarity score). Any investigation into the issue should consider patterned contacts that are weighted to maintain a specific rotation (since patterns are typically not radially symmetric).

A patterned contact is likely to present a level of difficulty proportional to the amount of iris texture that it covers. An opaque lens that completely covers the iris makes proper identification impossible. Small parts of the true iris texture may become visible if the contact is displaced (as in Figure 6(a)) or if the pupil constricts beyond the pattern of the lens (as in Figure 6(b)). Many patterned contacts only occlude some of the underlying iris texture, which leaves some of the real texture still visible.

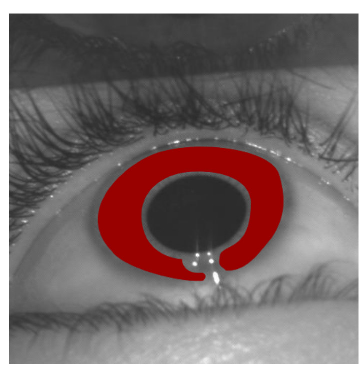
Transparent contacts were present in the iris dataset, but their effect on recognition accuracy did not appear to be significant compared to other sources of error. Only one comparison in the failure set involved transparent contact lenses, and the primary cause of failure was a high amount of motion blur in one of the images. There are specific situations where transparent contact lenses can introduce difficulties. Sometimes the boundary of the contact lens may make proper localization of the sclera more difficult. Soft contacts occasionally become crinkled or partially dislodged, especially when they are dry or immediately after the subject blinks. Figure 6(c) shows an example of a crinkled contact, although the image is not in the failure set. Baker et. al. [1] suggest that the various artifacts and optical distortions introduced by contacts can have a mild to severe effect on the Hamming distance, depending on the type of contact (hard, soft, etc.).

Tinted contact lenses are typically transparent in the infrared spectrum, where iris recognition systems work; hence they normally produce no more difficulty than a standard transparent lens.

Recommendations: Correctly worn transparent (clear or tinted) contact lenses are generally not problematic for iris recognition; as noted later, lenses that are turned over/crinkled can be problematic. Patterned/opaque contacts are a problem and should be removed before iris image capture. Many patterned/opaque contacts are reasonably obvious to a trained operator. In the short term, operators should be trained in the detection of such contacts. For the longer term, there are experimental methods proposed in the literature for automatically detecting patterned/opaque contacts. Automated detection of patterned contacts would relieve camera operators from the burden of having to detect them.

3.1.6 Off-Axis Gaze

(a)



(b)

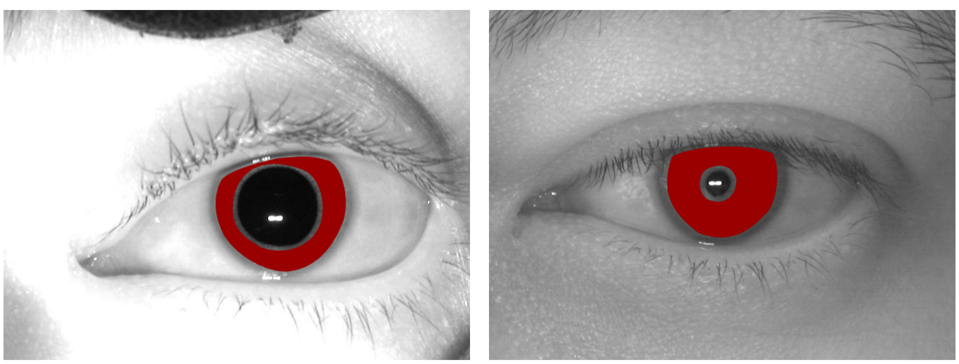
Figure 7: Examples of images with off-axis gazes.

The failure set contains 39 images where the eye's gaze is non-axial (see Figure 7). For this discussion, non-axial (sometimes referred to as ”off-axis” or ”non-frontal”) means that the optical axis and the line joining the camera and the eye differ considerably. Essentially, the subject was not looking at the camera at the time of capture. Images with off-axis gazes were rare in the overall dataset. In addition to instructions provided by the operators, the cameras may have had internal quality checks that forced a recapture if the gaze was off-axis. Recognition algorithms perform better on off-axis images when deformation models are used to process them into on-axis views [17, 11]. The correction models often assume the iris undergoes an affine transformation that is estimable from the gaze direction. In reality, the iris surface is subjected to non-linear distortions from a variety of sources (the cornea, barrel distortion from the photographic lens, changes in size with distance, etc). Furthermore, the iris surface is not entirely flat and can change appearance when viewed from different angles. The IREX II report found that recognition algorithms are unable to correct for all of these phenomena. Deformation models are useful for producing matchable templates from off-axis gazes, but the ideal solution is to capture an on-axis view to begin with.

The inspectors found that the failed matches for these cases were often due to an improper localization of the iris. Most methods of iris localization assume the iris is approximately annular in shape, but when viewed off-angle, the pupil and sclera boundaries become much more elliptical in appearance. Several methods of localizing the iris in off-axis images have been proposed [18, 21, 10].

Recommendations: The ideal option is proper field operator training. Camera operators should be trained to instruct the subjects to look straight ahead when capturing an image. Operators should also be trained to review the images. The IREX II evaluation found that automated algorithms for detecting off-axis gazes have potential. Consideration should be made into incorporating them into deployed systems – to provide real time feedback to the operators. For images that have already been collected, a human analyst can attempt a manual post-capture rectification of the off-axis image given the appropriate tools [12].

3.1.7 Dilation Change



(a) Searched Image (b) Enrolled Image

Figure 8: An example of a search in the failure set where the difference in pupil dilation is large.

The failure set contains 84 highly constricted pupils, 44 highly dilated pupils, and 8 searches where an iris with a highly dilated pupil was compared to an iris with a highly constricted one. In 5 of the 8 searches, the difference in dilation was the only problem identified with either image. The IREX I and IREX II evaluations found that large differences in dilation make recognition more difficult.

Although extremely dilated and constricted pupils can sometimes pose difficulties during iris segmentation, we suspect that the failures were more often due to the non-linear deformations that occur to the iris surface when the pupil size changes. Most iris matchers assume the deformations occur linearly along the radial direction of the iris. This is a fair approximation, but it does not account for all of the changes that occur. For example, when the pupil dilates and constricts, crypts often open and close in a way that introduces circumferential deformations.

Certain properties of the iris ensure that no deformation model can ever fully compensate for all of the deformations that occur. First, variations in pupil size change both the direction and shape of the stromal fibers that compose the iris, altering the way in which they overlap. Second, pupil dilation generates lateral pressure that causes some parts of the iris surface to fold over other parts, especially at circular contraction folds. Both of these phenomena change the apparent texture of the iris.

The amount of pupil dilation is commonly measured as a simple ratio [7]

$$D = R_p/R_i \tag{3}$$

where R_p and R_i are circular estimates of the pupil and limbus radii respectively. The authors of the IREX I evaluation report [5] proposed the following metric for measuring that amount of *dilation change* for a comparison

$$\Delta D = 1 - \left(\frac{R_i^{(2)}}{R_i^{(1)}}\right)\left(\frac{R_i^{(1)} - R_p^{(1)}}{R_i^{(2)} - R_p^{(2)}}\right) = 1 - \frac{1 - D^{(1)}}{1 - D^{(2)}} \tag{4}$$

where $D^{(1)} > D^{(2)}$ is assumed without loss of generality. The first form of the equation contains a pair of ratios. The first ratio is a scaling factor to compensate for possible differences in magnification. The second divides the annular width of the first iris by that of the second, which provides a direct measurement of the amount of radial stretching that occurs within the annular region. This may be a more accurate predictor of matching performance than simply taking the absolute value of the difference in dilation ratios between samples.

The dilation amount depends primarily on the pupillary light reflex. In outdoor environments where lighting cannot be controlled, little can be done to change the amount of pupil dilation. Iris images compare best when the pupils have similar amounts of dilation. Since it is difficult to predict what lighting conditions will be like during future captures, it is best to always try to perform the capture under normal lighting conditions.

Recommendations:

Operators should be trained to recognize the effect of lighting on pupil dilation: Dim ambient light will cause the pupil to dilate while excessive lighting will cause the pupil to constrict. The operators should be instructed to capture the iris under normal lighting whenever possible. If the subject's iris is extremely dilated or constricted due to a non-physiological cause (e.g. from certain drugs or opiates), then capture should be reserved for another day or time, if possible. For high-value cases, an expert human analyst may be able to adjust the algorithm parameters or segmentation boundaries to ensure that an accurate template is generated. Research into methods for modeling and compensating for changes in pupil dilation is ongoing [2]. As these methods become available, they should be tested in operational environments.

3.1.8 Small and Large Iris Radii

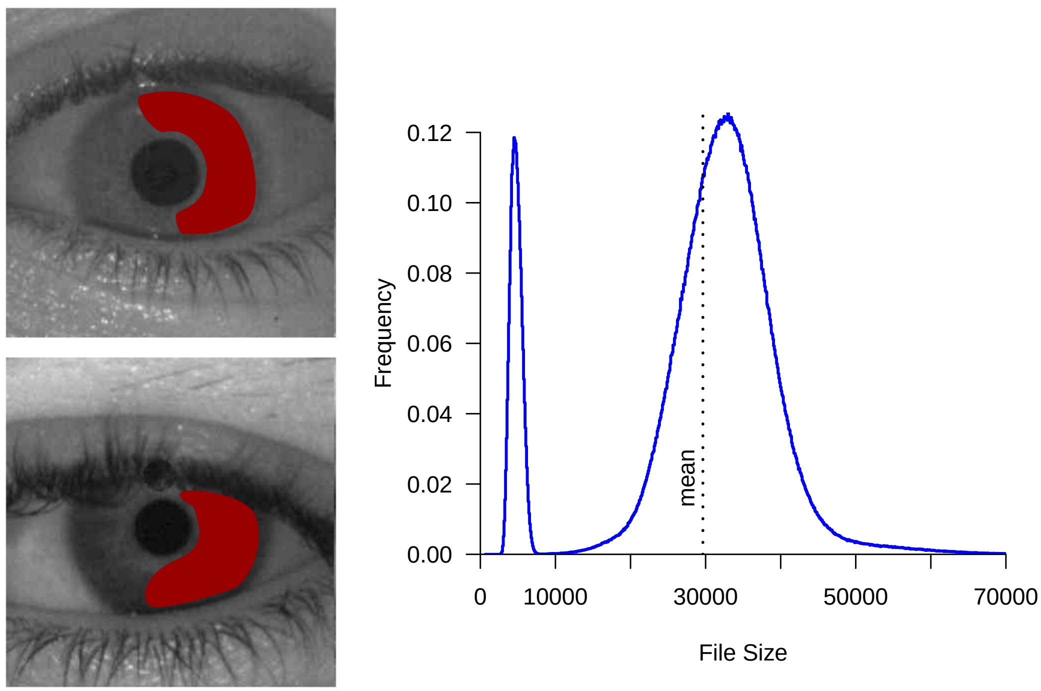
Some of the segmentation algorithms had difficulty localizing the sclera-iris (limbus) boundary in images where the iris was very large or very small. This was especially true for the images having pixel dimensions of 330-by-330, where the limbus diameter is typically around 60% of the image width. Note that this is not the same problem as localizing a partially cropped iris (which almost never occurred in the test dataset). The likely reason for the failed segmentations is that the algorithms were simply not calibrated to search for limbus radii outside of a certain range. The problem is easy to rectify by expanding the search range. Since very large and very small limbus radii are sometimes indicative of the subject standing too close to or far from the iris camera, the preferred solution may be to make an adjustment at capture time. The camera could provide feedback to the operator or subject indicating whether the subject's eye is at a proper distance from the camera.

Recommendations: Operators should be trained to know the capture volume of the camera – height, width, depth; and to understand that it is important to place the subject's eyes within that volume. Tests of iris segmentation algorithms on operationally relevant data would identify whether such algorithms are effective at providing useful feedback to operators.

On the back end, improvements in iris segmentation could enable operation over a larger range of radii.

3.2 System Problems

3.2.1 Lossy Image Compression



(a) (b)

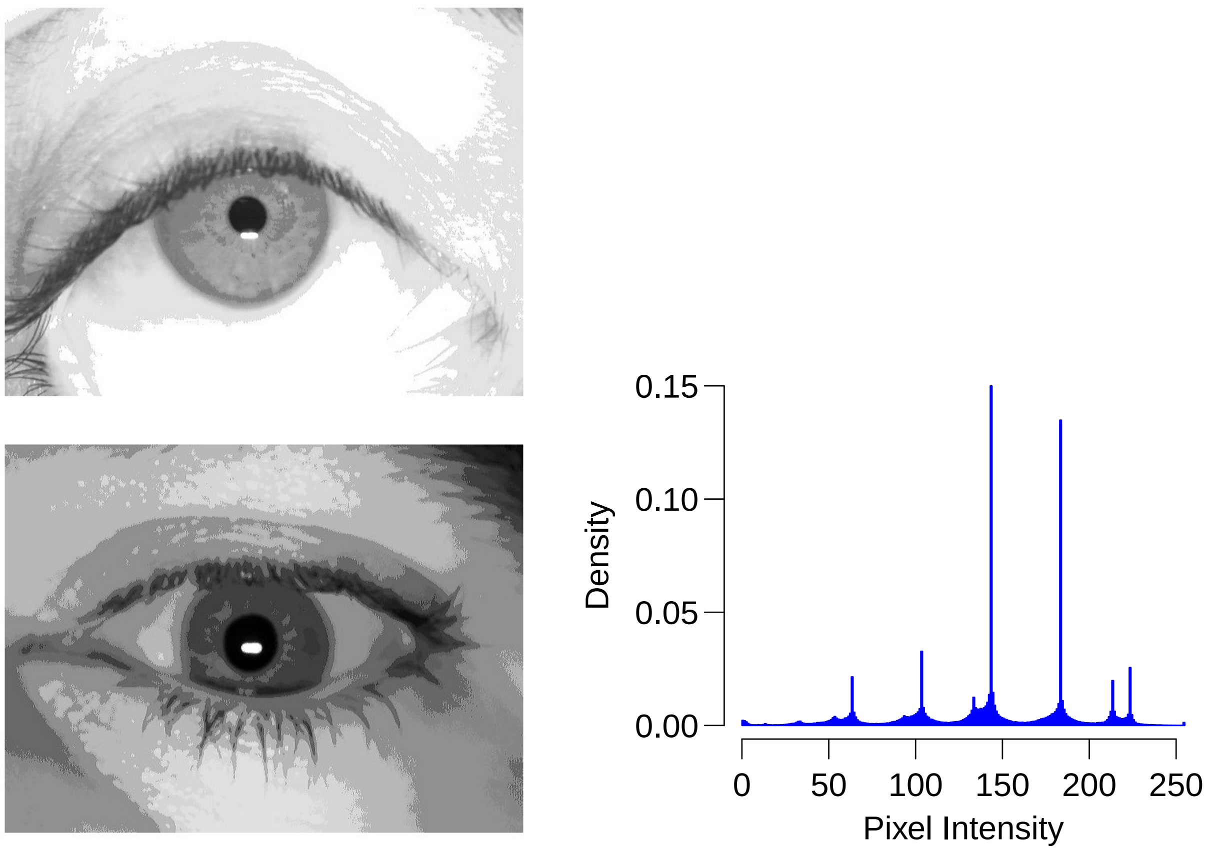
Figure 9: On the left are examples of highly compressed images. The graph on the right shown the distribution of file sizes for the overall dataset.

Lossy compression was applied to all of the images in the dataset (since they are stored as JPEGs), but only some of them were compressed enough to seriously degrade recognition performance. Figure 9(b) plots the histogram of file sizes for the approximately 4 million images in the test set. The smaller peak on the left corresponds to those images that were severely compressed, which comprise about 6% of the dataset. Examples of the highly compressed images are shown in Figure 9(a). The iris textures in these images suffer from severe compression artifacts. Note as well that the images were cropped such that the iris is always in the center of the image. The IREX I evaluation investigated the effect of image compression on recognition performance and determined that cropped JPEG images compressed to such small file sizes are unsuitable for automated matching. The images would have been suitable for matching had they been formatted in one of the standard compression formats defined in the ISO/IEC 19794-6:2011 standard [6]. Since the highly compressed images are unsuitable for operational use, they were excluded from the primary test set. They were, however, retained in the 17017 images used to generate the failure set.

The failure set contains 743 images involved in 557 searches that are highly compressed. Despite their poor quality, the matching algorithms were still able to correctly match them some of the time. Recognition performance over the highly compressed images is provided in the main IREX III evaluation report. The images show evidence of being processed in other ways as well. The iris radii are all very similar, suggesting the images may have been scaled. Specular hotspots from the iris camera are also masked out. The masking was typically performed inside the pupil region where it is unnoticeable, but in some cases perfectly circular dark spots can be seen over other areas in the images. The bottom image in Figure 9(a) shows an example of one of these dark spots situated directly above the pupil (although it may be difficult to see in paper print-outs). Interestingly, the dark spots do not consist of a single color as one would expect from a masking operation. The masking may, in fact, be the result of a contrast adjustment applied over the hotspots. The irregularities may also be artifacts introduced by compression.

Recommendations: If it is necessary to reduce the storage size of iris images, convert them to one of the compact formats defined in ISO/IEC 19794-6:2011. At this time, the standard does not recommend the removal of hotspots prior to compression.

3.2.2 Posterization (Quantization)



(a) (b)

Figure 10: On the left are examples of posterized images from the failure set (artifacts may not present well on paper printouts). The graph on the right shows the histogram of pixel intensities for the bottom iris image.

Posterization (also referred to as quantization and reduced color space) occurs when the number of distinct colors/tones represented in the image is reduced to a relatively small number, such that gradual changes in color tone appear as abrupt transitions from one color to another. This process was originally developed for creation of posters from photographs, whence comes the name. It is now a common option in digital image processing programs (such as PhotoShop, IrfanView, and Gimp). The presence of posterization in an image can be deliberate, or an unintended side effect of an image processing step such as lossy compression. The failure set contains 50 images involved in 47 searches that display obvious signs of posterization (see Figure 10(a) for examples). Fine details in these images are replaced with featureless areas of uniform color. The inspectors believe that posterization is the primary reason that these images failed to match.

Figure 10(b) shows the brightness intensity histogram for a posterized image. Other posterized images had similar histograms, but the peaks often differed in number and location. The peaks appear to ”bleed” into adjacent intensities due to additional compression artifacts introduced when the posterized images were converted to JPEGs. JPEG is a lossy compression format that changes the intensities of some pixels in order to store the image more compactly. The IREX II report proposes a metric referred to as GRAY SCALE SPREAD that seems appropriate for quantifying the amount of posterization in these images.

The images may have been posterized intentionally in an attempt to make them easier to compress. When fewer distinct shades of gray are used, images can be stored at lower bit-depths, thus requiring less space to encode. However, the images in the current dataset were always stored at 8 bits-per-pixel, regardless of how many distinct brightness intensities were actually used. Posterization also failed to make the images any easier to compress as JPEGs since it did not substantially reduce the amount of high frequency information in the images, which is expensive to encode. Guidelines for the efficient compression and storage of iris images are provided in ISO/IEC 19794-6:2011.

Once an iris image is posterized, the texture information contained in the iris that is lost cannot be recovered without capturing a new image.

Recommendations: Iris images should not be posterized. If it is necessary to reduce the storage requirements of iris images, convert them to one of the compact formats defined in ISO/IEC 19794-6:2011. Automated detection of posterized images is possible, as described in the IREX II report.

3.2.3 Ground Truth Errors

The inspector determined that for 19 of the searches, the search image and its corresponding enrolled mate were not of the same person’s iris. In other words, at least one of the paired images was assigned an incorrect subject identifier. The determination was made based on a visual inspection of the two irides and their surrounding ocular regions. If it was clear that the two irides are from different people, the inspector marked the pair as having a ”ground truth” error. These 19 cases comprise $19/17017 \approx 0.11\%$ of the searches in the failure set.

3.2.4 Bayer Filter Defect

The image in Figure 11 shows a dithering pattern that most likely resulted from a misbalance in the gains for the Bayer color channels. When a camera design for color images using the Bayer filter method is adapted for grayscale, the Bayer filter is left off the sensor and the Bayer channels (R,B,GB,GR) must be balanced. Failure to properly balance the channels will result in a noise pattern that repeats on a scale of 2 pixels. The effect occurs at a high spatial frequency in the underlying images, but we see it at a lower frequency in these images due to a Moire interaction between the sampling rate for the display and the underlying sensor pattern noise.

The pattern noise may be outside of the passband of the matching algorithms, in which case it would have little effect on matching performance. However, it IS noise and should be avoided. The presence of this problem in the image stream indicates a lack of attention to overall quality of the images. Although it is possible that this noise was introduced by other processes, the sensor is the most likely culprit. The prevalence of this problem in the failure set was not investigated during the manual inspection.

Recommendations: Images that appear to display Bayer artifacts should be provided to the camera vendor for analysis and remediation in the camera system. It is possible to post-process the images to mitigate this problem.

3.2.5 Low Contrast

Figure 12 shows a good quality search image (left) and an low quality enrollment image (right). The enrollment image suffers primarily from low contrast. The occlusion of the camera lens by the operators finger in the upper left may have contributed to the low contrast. The size of the pupil suggests that there is a fair amount of ambient illumination, but this

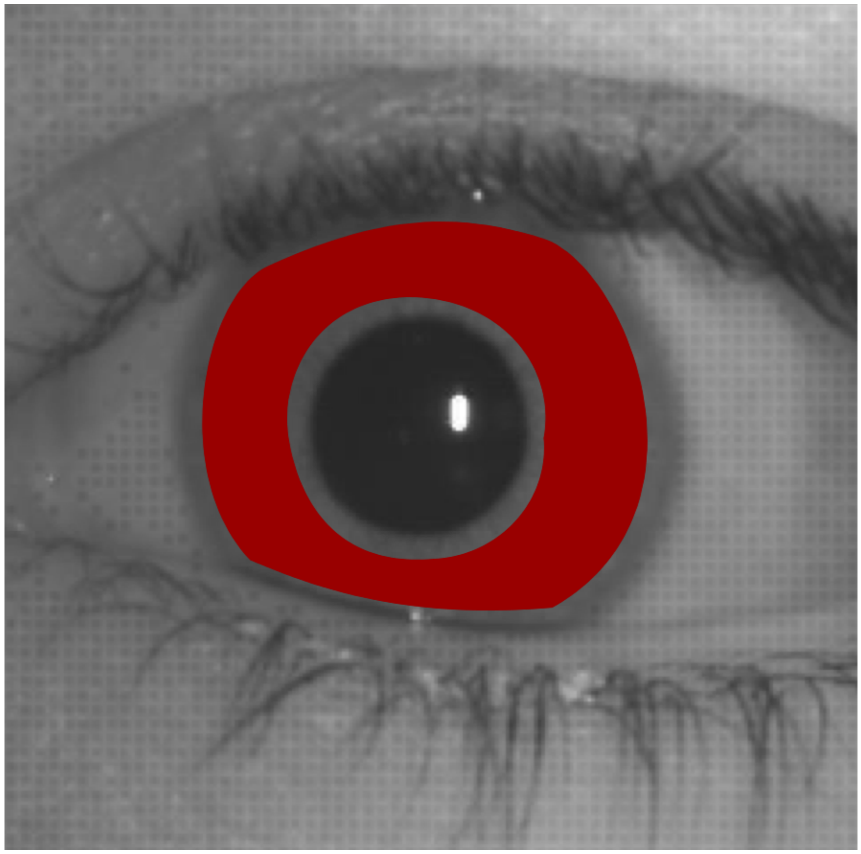


Figure 11: The dithering pattern in this image is likely introduced by a Bayer filter camera defect; see text for description.

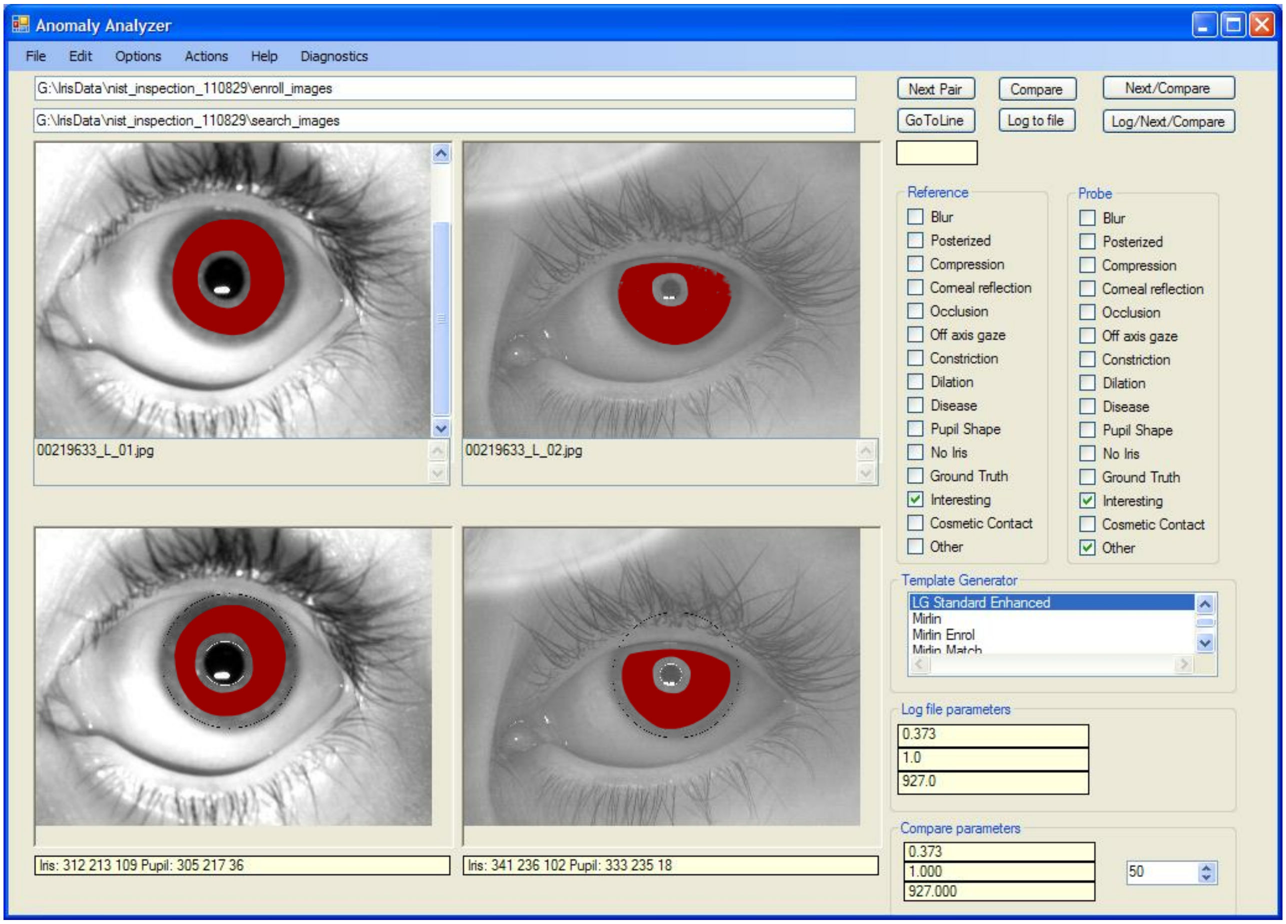
does not necessarily mean that there is a sufficient amount of near-IR for a good quality capture. The Hamming distance for this comparison is 0.37, which corresponds to a false match rate of about 1 in 600. This is a case where adjudication by a human operator might provide actionable intelligence.

Recommendations: In general, camera operators should be trained in good image acquisition practices. In this case, they should be trained to avoid capturing an image when their finger is partially blocking the camera lens. An automated algorithm may be able to detect this by measuring contrast. For cases in which there is no apparent reason for a low contrast image, an example should be provided to the camera vendor(s) for comment and explanation.

3.2.6 Broken LEDs

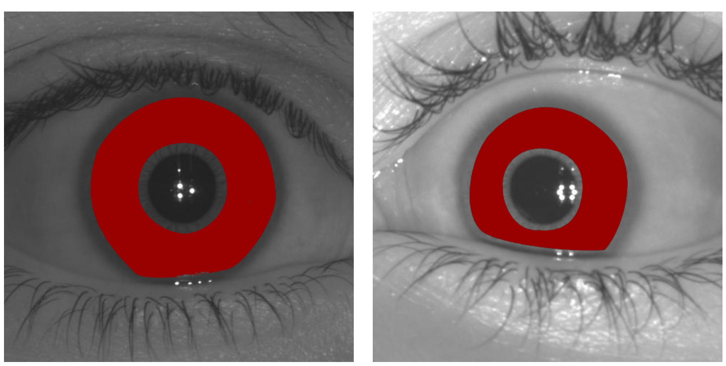
Figure 13 show two images captured by the same type of camera, but the left image has a lower number of light emitting diode (LED) reflections. Our inference is that some of the LEDs were burnt out when the left image was captured. Burnt out LEDs lead to insufficient illumination of the iris, which will make recognition more difficult. The prevalence of burnt out LEDs in the failure set was not investigated during the manual inspection and would be difficult to ascertain in some images. It may be possible to detect burn-out LEDs based on image properties. For example, ISO/IEC 29794-1 defines requirements for contrast and gray-scale utilization that may not be met by a camera operating with broken LEDs.

Recommendations: Camera operators should be trained in the basic characteristics of good quality iris images so that they can make rough, real-time assessment of the images they collect. In this case, the camera operators should be trained to recognize the nominal specularity pattern of their camera(s) and to realize when the specularity pattern is symptomatic of a problem.



(a)

Figure 12: Example of a case where the enrollment image is low contrast.



(a) (b)

Figure 13: Example of an iris captured by a camera with burnt out LEDs (left), and an iris capture by the same type of camera with functioning LEDs (right).

(a) (b)

Figure 14: Example images from the failure set where the pupil shape is abnormal.

3.3 Poor Inherent Features

3.3.1 Abnormal Pupil Shapes

The failure set contains 45 images involving 36 searches where the pupil's shape is abnormal (i.e. deviates significantly from circular). Examples of abnormal pupils are shown in Figure 14. Abnormal pupil shapes are not restricted to the failure set. However, an inspection of 1000 randomly selected images (not from the failure set) revealed only one where the pupil's shape is abnormal.

Occasionally, the abnormal shape was caused by a crinkled contact lens that created optical distortions. In all other cases, the pupil's coverage area appears to extend into regions normally occupied by the iris (as can be seen Figure 14). The abnormalities are most likely due to congenital defects such as Coloboma. (Watters [20] identifies several medical conditions that are likely to impact the performance of iris recognition.) Most of the time, the pupil shape was abnormal in only one of the subject's eyes (either the left or the right).

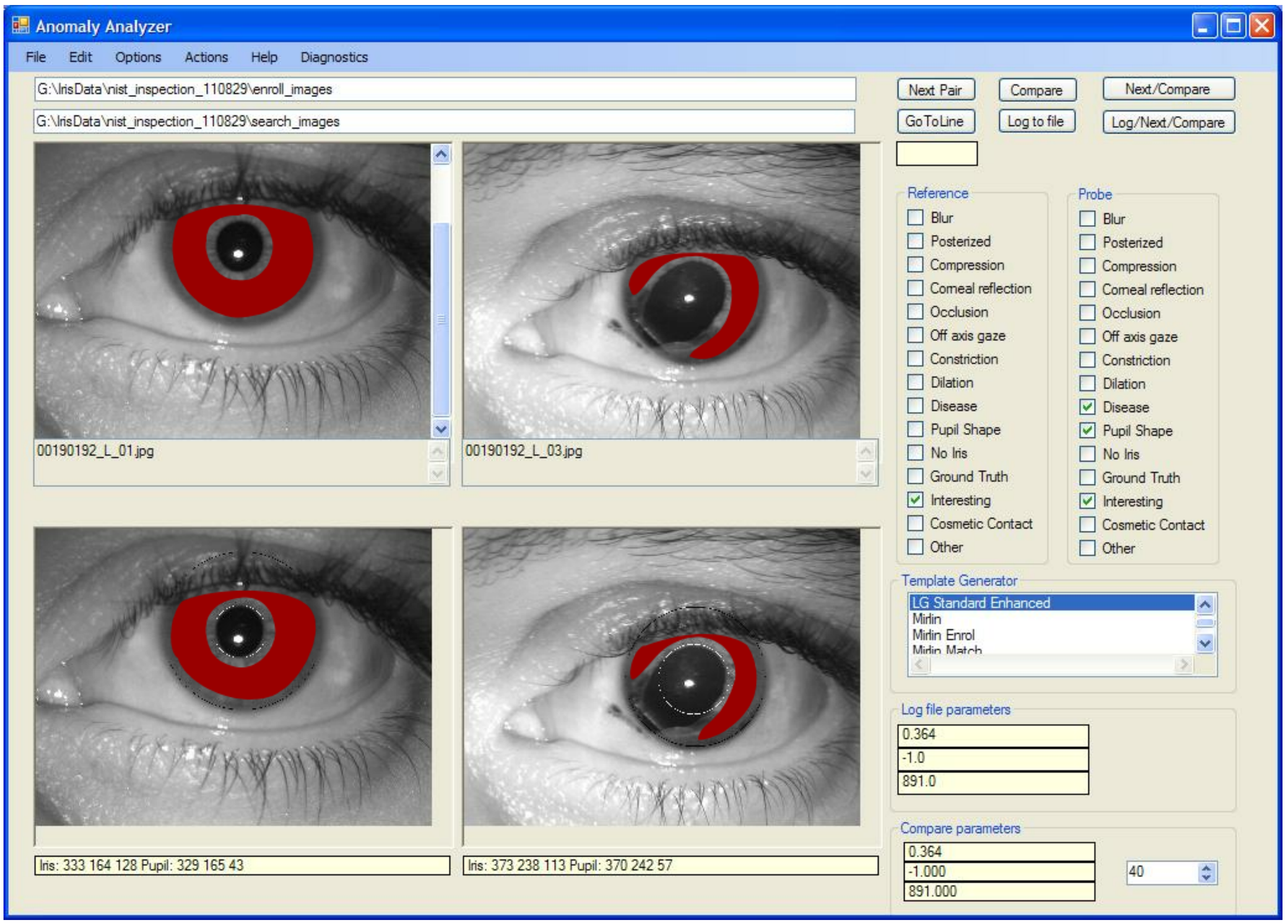
The irregular shape of the pupil can make it difficult to localize since most methods of boundary detection assume that the pupil is approximately circular in shape. Another obvious disadvantage to the abnormal pupil shape is that there tends to be a reduced amount of visible iris texture. It would be interesting to note how the iris texture changes in response to changes in pupil size for irregularly shaped pupils, but unfortunately the dilation amount varied little across different images of the same iris.

Recommendations: There is little that an operator can do if the subject's eye contains a chronic abnormality. For acute conditions (such as certain traumas) the subject can be instructed to return when the eye has recovered.

3.3.2 Disease and Trauma

The iris can suffer damage in response to diseases or blunt injuries to the eye. Additionally, the cornea can become temporarily or permanently scarred, which may also alter the appearance of the iris. Figure 15 shows an iris that appears to have suffered some type of trauma between captures. The dark blotting in lower left was enough to raise the Hamming distance between the before and after images to 0.36. This corresponds to a false match rate of about 1 in 2600. This is a case where adjudication by a human operator might provide actionable intelligence.

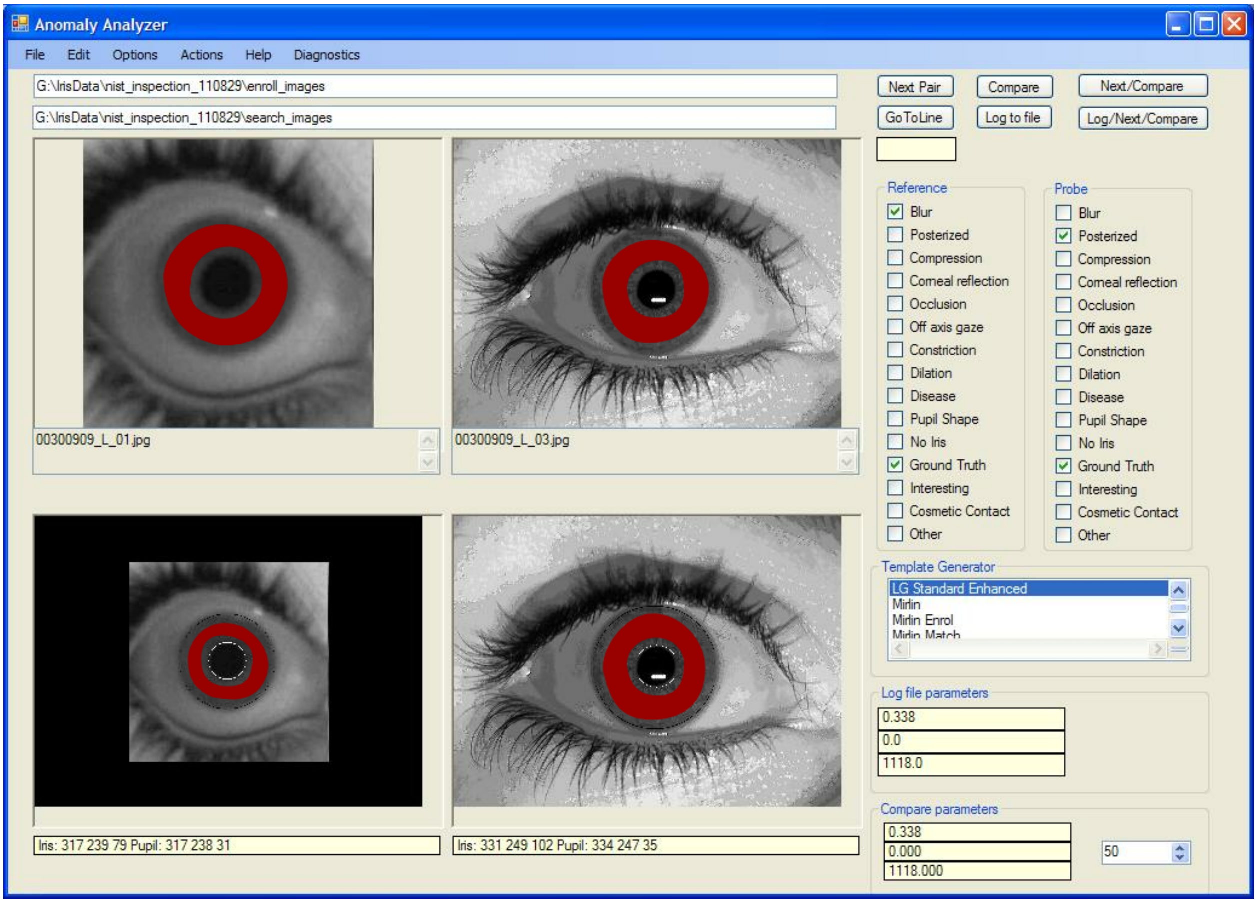
Recommendations: There is little that an operator can do with regard to disease or trauma of the subject's eye (other than to possibly delay iris capture until the eye has healed). Subsequent adjudication by a human analyst can be effective. With appropriate tools, a human analysis can at least mask out the areas of the iris that show visible signs of injury.



(a)

Figure 15: An example of an iris that suffered trauma between captures.

4 Poor Quality due to Multiple Problems



(a)

Figure 16: An example of a search that failed due to multiple problems.

While each of the previous sections focus on a single source of poor quality, this section addresses the problem of compounding factors that lead to failure. It was previously noted that nearly two-thirds of the blurry images in the failure set also have high amounts of occlusion. This suggests that blur exacerbates the detrimental effect of occlusion and vice versa. Figure 16 shows another example of a search that likely failed due to multiple problems with the images. Both the search image (left) and its enrolled mate (right) are mislabeled as left irides. In this case, two wrongs do make a right in-so-far as making a correct match is concerned. However, the images suffer from other problems. The search image contains noticeable compression artifacts and the enrolled mate is posterized. Nevertheless, the Hamming Distance between the two images is ∼0.34, just above a nominal cutoff of ∼0.33. The probability of a nonmated comparison producing a Hamming Distance at or below this value is ∼1:40000. How well the comparison would have faired in the absence of one or more of these problems is unknown. This is again a case where a human analyst could provide actionable intelligence.

5 Final Comments

Despite the quality problems in the operational dataset, the algorithms are usually robust enough to handle the images (see accuracy results in the main IREX III report). This is a mixed blessing, since greater tolerance of poor image quality can lead people to become less attentive to maintaining high standards of quality when handling image data. Poor quality images will inevitably lead to higher false non-match rates and a higher incidence of cases where a potentially dangerous individual is not recognized. Quality control procedures should be established to monitor the quality of the data streams and to ensure that prompt corrective action is taken when those streams depart significantly from defined norms.

- In general, the iris recognition algorithms work on the operational data provided for use in this study.
- The failure cases investigated in this report highlight areas where modifications to operational procedures would improve recognition performance.
- Some of the images from the operational dataset were originally stored in 24 bit RGB format. This is poor practice, especially if there are bandwidth and storage space restrictions, and considering that iris cameras typically capture images natively in 8 bits-per-pixel.
- If compression of iris images is necessary, use the compact formats defined in ISO/IEC 19794-6:2011 (and equivalently defined in ANSI/NIST ITL 1-2011). Images should not be posterized.
- If an iris image is cropped to the point that the medial canthus (the corner of the eye closest to the nose) is no longer visible, it will likely be impossible to determine whether the eye is a left or a right. Considering the prevalence of eye mislabelings in the dataset, we suggest that images not be cropped beyond the point at which the medial canthus is no longer visible. This may not be possible if storage limitations require that the images be stored in one of the compact formats defined in the ISO/IEC 19794-6 standard.

A Summary of Recommendations

This appendix consolidates the recommendations from each of the subsections in section 3.

1. **Contact lenses:** Correctly worn transparent (clear or tinted) contact lenses are generally not problematic for iris recognition; as noted later, lenses that are turned over/crinkled can be problematic. Patterned/opaque contacts are a problem and should be removed before iris image capture. Many patterned/opaque contacts are reasonably obvious to a trained operator. In the short term, operators should be trained in the detection of such contacts. For the longer term, there are experimental methods proposed in the literature for automatically detecting patterned/opaque contacts. Automated detection of patterned contacts would relieve camera operators from the burden of having to detect them.

2. **Occlusion** Operators should be trained to instruct their subjects to open their eyes wide when their irides are captured. IREX II found that automated algorithms for measuring the amount of useable iris area have potential. If future cameras are designed with built-in occlusion detection, they could provide direction to operators to reacquire images that are highly occluded.

3. **Rotation:** A single-eye camera is much more likely to capture a highly rotated iris than a dual eye camera – and introduces the potential for mislabeling the eyes (your right or mine). Hence, dual eye cameras are preferred, all other things being equal. For cases in which a single eye camera is used, rotation can be dealt with in various ways. In order of preference: 1) rotation can be minimized during the capture phase by ensuring that the camera is held upright and the subject's head is not tilted, 2) if the image is already captured and stored, the iris can be manually "unrotated", 3) the matching algorithm can be configured to compare iris templates over a more relaxed range of rotation disparities.

   There is some discussion in the literature on automated methods for measuring eye rotation in an image [13], although none of the quality algorithms submitted to IREX II attempted to measure it. If single eye cameras must be used, these methods should be explored.

4. **Off-axis gaze:** The ideal option is proper field operator training. Camera operators should be trained to instruct the subjects to look straight ahead when capturing an image. Operators should also be trained to review the images.

The IREX II evaluation found that automated algorithms for detecting off-axis gazes have potential. Consideration should be made into incorporating them into deployed systems – to provide real time feedback to the operators. For images that have already been collected, a human analyst can attempt a manual post-capture rectification of the off-axis image given the appropriate tools [12].

5. **Reflective Glare:** The LEDs on the camera should be the dominant source of illumination. Operators should be trained to re- duce ambient sources of illumination when possible and to recognize the presence of specular reflections of the ambient environment. To minimize specular reflections in an outdoor environment, the camera operator should position himself such that his back is to the sun and the subject is facing a featureless scene.

6. **Dilation Change:** Operators should be trained to recognize the effect of lighting on pupil dilation: Dim ambient light will cause the pupil to dilate while excessive lighting will cause the pupil to constrict. The operators should be instructed to capture the iris under normal lighting whenever possible. If the subject's iris is extremely dilated or constricted due to a non-physiological cause (e.g. from certain drugs or opiates), then capture should be reserved for another day or time, if possible. For high-value cases, an expert human analyst may be able to adjust the algorithm parameters or segmentation boundaries to ensure that an accurate template is generated. Research into methods for modeling and compensating for changes in pupil dilation is ongoing [2]. As these methods become available, they should be tested in operational environments.

7. **Blur:** Operator training should include examples of blurred images so that they understand this issue. They should be instructed to hold the camera as steady as possible when capturing images and to keep the subject within the capture volume of the camera, which the operator must know. Operators should instruct subjects to stand still and look straight ahead during the capture. Several automated methods for measuring blur in images exist, many of which were evaluated as part of IREX II. There may be a benefit to incorporating automatic blur detection into iris cameras as a way to provide real-time feedback to the operator.

8. **Small and large iris radii:** Operators should be trained to know the capture volume of the camera – height, width, depth; and to understand that it is important to place the subject's eyes within that volume. Tests of iris segmentation algorithms on operationally relevant data would identify whether such algorithms are effective at providing useful feedback to operators. On the back end, improvements in iris segmentation could enable operation over a larger range of radii.

9. **Lossy image compression:** If it is necessary to reduce the storage size of iris images, convert them to one of the compact formats defined in ISO/IEC 19794-6:2011. At this time, the standard does not recommend the removal of hotspots prior to compression.

10. **Posterization:** Iris images should not be posterized. If it is necessary to reduce the storage requirements of iris images, convert them to one of the compact formats defined in ISO/IEC 19794-6:2011. Automated detection of posterized images is possible, as described in the IREX II report.

11. **Bayer filter defect:** Images that appear to display Bayer artifacts should be provided to the camera vendor for analysis and remediation in the camera system. It is possible to post-process the images to mitigate this problem.

12. **Low contrast:** In general, camera operators should be trained in good image acquisition practices. In this case, they should be trained to avoid capturing an image when their finger is partially blocking the camera lens. An automated algorithm may be able to detect this by measuring contrast. For cases in which there is no apparent reason for a low contrast image, an example should be provided to the camera vendor(s) for comment and explanation.

13. **Broken LEDs:** Camera operators should be trained in the basic characteristics of good quality iris images so that they can make rough, real-time assessment of the images they collect. In this case, the camera operators should be trained to recognize the nominal specularity pattern of their camera(s) and to realize when the specularity pattern is symptomatic of a problem.

14. **Abnormal pupil shapes:** There is little that an operator can do if the subject's eye contains a physical abnormality. For acute conditions (such as certain traumas) the subject can be instructed to return when the eye has recovered.

15. **Disease and trauma:** There is little that an operator can do for chronic conditions. Subsequent adjudication by a human analyst can be effective. With appropriate tools, a human analysis can at least mask out the areas of the iris that show visible signs of injury.

References

[1] S. E. 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*, pages 1030–1044, 2010.

[2] A. Clark and A. A. Ross. Exploring nonlinear dynamics of iris deformation. In *Biometric Consortium Conference*, 2011.

[3] J. Daugman. Demodulation by complex-valued wavelets for stochastic pattern recognition. *International Journal of Wavelets, Multi-resolution and Information Processing*, 1:1–17, 2003.

[4] J. Daugman. The importance of being random: Statistical principles of iris recognition. *Pattern Recognition*, 36:279–291, 2003.

[5] P. Grother, G. Quinn, J. Matey, M. Ngan, W. Salamon, G. Fiumara, and C. Watson. IREX: Performance of Iris Identification Algorithms. Technical report, NIST, 2011.

[6] ISO. *ISO/IEC 19794-6 - Biometric Data Interchange Formats - Iris Image Data*. Geneva, Switzerland, 2005.

[7] ISO. *ISO/IEC 19794-6 - Biometric Sample Quality - Part 1: Framework*. Geneva, Switzerland, 2009.

[8] W. Kong and D. Zhang. Accurate iris segmentation based on novel reflection and eyelash detection model. In *Proceedings of the 2001 International Symposium on Intelligent Multimedia, Video and Speech Processing, 2001*, pages 263–266, 2002.

[9] E. Lee and K. Park. Fake iris detection based on 3d structure of iris pattern. *International Journal of Imaging Systems and Technology*, 20(2):162–166, 2010.

[10] P. Li, X. Liu, L. Xiao, and Q. Song. Robust and accurate iris segmentation in very noist iris images. In *image and Vision Computing*, volume 28, pages 246–253, Feb 2010.

[11] X. Li. Modeling intra-class variation for non-ideal iris recognition. In *In Springer LNCS 3832: Int. Conf. on Biometrics*, pages 419–427, 2006.

[12] J. R. Matey, D. Ackerman, J. Bergen, and M. Tinker. *Advances in Biometrics*, chapter 7. Springer Verlag, 2008.

[13] D. M. Monro and S. Rakshit. Rotation compensated human iris matching. In *Signal Processing Applications for Public Security and Forensics, 2007. SAFE '07. IEEE Workshop on*, pages 1 –4, april 2007.

[14] J. Phillips, T. Scruggs, A. O'Toole, P. Flynn, K. Bowyer, C. Schott, and M. Sharpe. FRVT 2006 and ICE 2006 large-scale results. Technical report, NIST, 2007.

[15] S. Ring and K. Bowyer. Detection of iris texture distortions by analyzing iris code matching results. In *Biometrics: Theory and Applications and Systems*, pages 1–6, September 2008.

[16] N. Sazonova, S. Schuckers, P. Johnson, P. Lopez-Meyer, E. Sazonov, and L. Hornak. Impact of out-of-focus blur on iris recognition. In *Proceedings of SPIE 8029*, April 2011.

[17] S. Schuckers, N. Schmid, A. Abhyankar, V. Dorairaj, C. Boyce, and L. Hornak. On techniques for angle compensation in non-ideal iris recognition. In *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, volume 37, pages 1176–1190, 2007.

[18] E. Sung, X. Chen, J. Zhu, and J. Yang. Towards non-cooperative iris recognition systems. In *ICARCV'02*, pages 990–995, 2002.

[19] E. Tabassi, P. Grother, and W. Salamon. IREX - IQCE performance of iris image quality assessment algorithms. Technical report, NIST, 2011.

[20] E. Watters. Iris pathologies medical conditions that may impact the performance of iris recognition systems. In *The Biometrics Consortium Conference (BCC)*, September 2011.

[21] J. Zuo and N. Schmid. On a methodoligy for robust segmentation of non-ideal iris images. In *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, volume 40, June 2010.