A Multi-Algorithm Analysis of Three Iris Biometric Sensors

Ryan Connaughton, Amanda Sgroi, Kevin Bowyer, and Patrick Flynn

***Abstract*—The issue of interoperability between iris sensors is an important topic in large-scale and long-term applications of iris biometric systems. This work compares three commercially available iris sensors and three iris matching systems and investigates the impact of cross-sensor matching on system performance in comparison to single-sensor performance. Several factors which may impact single-sensor and cross-sensor performance are analyzed, including changes in the acquisition environment and differences in dilation ratio between iris images. The sensors are evaluated using three different iris matching algorithms, and conclusions are drawn regarding the interaction between the sensors and the matching algorithm in both the cross-sensor and single-sensor scenarios. Finally, the relative performances of the three sensors are compared.**

***Index Terms*—biometrics, iris recognition, sensor evaluation, interoperability**

I. Introduction

SENSOR technology for iris biometrics is evolving, with many companies releasing new sensors and improving existing sensors. This motivates an investigation of whether these sensors are interoperable. Several studies have investigated the interoperability of fingerprint, voice, and signature sensors [1],[2],[3],[4]. Additionally, some researchers have reported on sensor safety, illumination, and ease-of-use for iris recognition systems [5]. Other work has explored environmental and operational factors and their impact on biometric systems [6],[7],[8]. Nevertheless, few studies have been conducted to investigate the interoperability of iris sensors from varying manufacturers using multiple available matching algorithms.

This work explores sensor interoperability using three commercially available iris sensors by comparing results from single-sensor and cross-sensor experiments. Each of the three sensors examined in this study is used for iris enrollment and recognition in the field today. Additionally, three different algorithms are used to perform iris matching, which may provide a less biased context in which to compare sensors, as well as offer some insight into the effect of the algorithm on single-sensor and cross-sensor performance. Further, we investigate the impact of several external factors on sensor performance. We analyze the performance of each single-sensor and cross-sensor experiment to address several questions, including the following:

- How do subtle changes in environmental conditions and the order of sensors impact performance?
- How does pupil dilation affect sensor performance?
- How does cross-sensor performance between two sensors relate to the corresponding single-sensor performances?
- Is relative sensor performance consistent across matching algorithms?
- Can a relative sensor ranking be established?

The remainder of this paper is organized as follows. Section II discusses previous studies of the interoperability of iris biometric sensors. Section III describes each sensor, the experimental setup, the dataset, and the algorithms used to evaluate the three iris sensors. Various results produced by the three matching algorithms are shown and discussed in Section IV, and Section V concludes with general observations regarding the iris sensors and data comparisons.

II. Related Work

Since iris biometrics is increasingly becoming a large-scale application in which data is kept and used for long periods of time, the interoperability between iris sensors has become a recent topic of interest. Bowyer et al. investigated cross-sensor and cross-session comparisons using two iris sensors, the LG IrisAccess 4000 (LG 4000) and LG IrisAccess 2200 (LG 2200), and a single matching algorithm [9]. The authors found that the LG 2200 provided a less desirable match score distribution, which led to an even less desirable cross-sensor match score distribution. They concluded that if the LG 4000 was used to collect enrollment data and the LG 2200 was used to collect probe data, the system would achieve higher recognition rates than if the LG 2200 was used for enrollment and the LG 4000 was used to acquire probe data.

In 2005, International Biometrics Group (IBG) evaluated the performance of three of the most widely used iris acquisition and recognition systems at that time [10]. The authors evaluated each system through several criteria - false accept and reject rate, failure to enroll rate, failure to acquire rate, acquisition time, subject usability, and performance over time. The investigation showed that the sensor with the lowest failure to enroll rate had less robust matching over time than the sensor with a higher failure to enroll rate. The sensors were

This work is sponsored under IARPA BAA 09-02 through the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-10-2-0067. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing official policies, either expressed or implied, of IARPA, the Army Research Laboratory, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

R. Connaughton, A. Sgroi, K. Bowyer, and P. Flynn are with the Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN, 46556 USA e-mail: rconnaug,asgroi@nd.edu; kwb,flynn@cse.nd.edu

evaluated in both a single-sensor and cross-sensor context, but only one matching algorithm was used for the experiments.

In 2007, Authenti-Corp published an evaluation of three commercial sensors, though the sensors were not identified in the report [11]. The sensors were compared in both single-sensor and cross-sensor contexts using a single matching algorithm. The authors reported that one of the sensors performed best in a single-sensor context, while the other two sensors performed best in cross-sensor contexts. They conclude that further investigation is required to understand which image characteristics have the largest impact on sensor interoperability.

A survey report from 2009 presented the state of the art in third-party iris system evaluations [12]. The report summarized three third-party evaluations, including the IBG and Authenti-Corp reports. The third evaluation was the Iris Evaluation Challenge (ICE 2006) conducted by the National Institute of Standards and Technology (NIST), which evaluated three separate algorithms using data from a single sensor [13]. The survey noted that despite differences in hardware and software, all three studies reported false reject rates on the same order of magnitude when measured at a false accept rate of 0.001 [12].

Many factors can affect the accuracy of iris biometrics systems. These factors include, but are not limited to, pupil dilation [14], time lapse between enrollment and recognition [15], and contact lenses [16]. Some sensors may be more or less sensitive to these factors than others, according to the optics, illumination technique, hardware, and software of each sensor. In the IBG evaluation and other reports such as that performed by Du [5], the illumination schemes, sensor optics, and usability were evaluated. In our study we include an investigation of how the role of pupil dilation and algorithm selection affect interoperability of sensors.

Little previous work has been done on the effect of pupil dilation on sensor interoperability. Though the most obvious cause of pupil dilation is a low ambient light level, it has been shown that other factors such as age, medications, diseases, trauma, and stress can cause a change in pupil dilation. NIST has conducted several studies regarding the performance of iris biometrics given variations in pupil dilation [17]. To evaluate the effects of pupil dilation, Hollingsworth et al. considered the difference in dilation ratio between pairs of images [18]. Image pairs were grouped based on their difference in dilation ratios, and the performance of each group was examined. The authors showed that smaller differences in dilation ratio allows better performance. Because of this, recent research has been conducted regarding enrollment strategies based on pupil dilation [19].

The matching algorithm used to perform iris recognition also plays a large role in determining the performance of the system. In 2009, NIST released the IREX report comparing several algorithms using three iris datasets [17]. The IREX evaluation mostly measured algorithm performance with respect to run-time analysis and expense of computations. However, the authors also found that the choice in the iris recognition algorithm is more influential on the outcome given standardized iris images than in other biometrics, such as fingerprint.

In previous work, we compared single-sensor and cross-sensor results from four sessions of data in which changes in environmental conditions were minimized throughout all sessions [20]. These environmental conditions include distance between sensor and ambient light sources, distance between sensors, and height of sensor relative to subject. Experiments were conducted to analyze the relationship between cross-sensor and single-sensor performance, as well as the role the algorithm played in performance. As an extension of this work, we considered a second set of four sessions and analyze the impact of several other factors on system performance, including sensor order in the acquisition studio and dilation ratio in acquired images.

An important consideration when comparing sensor performance is the size of the dataset used to conduct the analysis. In the proceedings of Biometrics: Theory, Applications and Systems in 2010, experiments on iris biometrics were presented using datasets that ranged from 200 to 2000 iris images (25 to 264 unique subjects) [21],[22]. The experiments described in our cross-sensor study are conducted on a dataset of nearly 50000 iris images (630 unique subjects) and therefore represents a large improvement in the state of the art.

III. Methods

A. Sensors

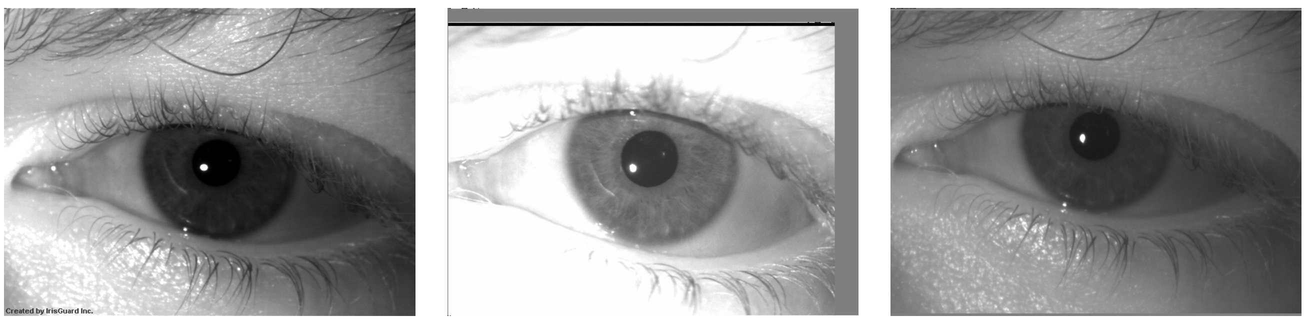
In this work, three commercially available iris sensors are compared in a cross-sensor and cross-session context to evaluate both the performance of each individual sensor and the interoperability between the sensors.

The first sensor used in this study is the **IrisGuard AD100**, referred to as the IrisGuard in this work [23],[24]. The IrisGuard is a dual eye iris capture sensor which employs direct and cross illumination through the use of two clusters of 6 near infrared (NIR) LED illuminators [25]. Subjects must be approximately 8 to 12 inches away from the camera for image capture to occur. In our acquisition process, 4 images were taken of each eye after one prompt of the camera. The firmware for this sensor calculates the amount of motion, measured by the level of activity of the eye, and determines whether each subject is using contact lenses or glasses and adjusts illumination accordingly [24].

The second sensor used is the **LG IrisAccess 4000**, referred to as the LG 4000 in this work, and it is also a dual capture iris sensor [26]. It makes use of two clusters of 12 NIR LED illuminators of varying wavelengths, which provide cross and direct illumination of both irises [27]. For acquisition to occur, a subject must be approximately 14 inches away with their eyes centered in a reflective acquisition window. In the first two acquisition sessions, three sets of images were captured and the subject was prompted by the camera after each capture. However, our capture software was upgraded after the second acquisition session such that four sets of images were captured after only one camera prompt, but no changes were made to the firmware or quality scoring.

The third sensor used in the study is the **LG TD100**, referred to as the TD100 in this work [28]. The TD100's intended

Fig. 1. Images of the three iris sensors used in this study. From left to right: the LG IrisAccess 4000, the IrisGuard AD100, and the LG TD100.



IrisGuard Image 04233d3014 LG 4000 Image 04233d3004 TD100 Image 04233d3010

Fig. 2. Images of the same iris taken by each of the three sensors during the same acquisition session. The LG 4000 image includes padding inserted by the sensor's firmware to maintain the 640x480 pixel image size. The contrast in each image appears different according to the illumination and hardware of each sensor.

use is as a handheld face and dual iris capture camera, but it can also be mounted on a tripod for stationary use. A subject's eyes must be positioned approximately 13 inches from the sensor and centered in a reflective window. Clusters of LEDs of multiple wavelengths of NIR illumination are used to illuminate the iris. This sensor does not prompt the subject, so an operator must provide verbal instructions to the subject, which results in a variable acquisition time. Three sets of images per subject were acquired using this sensor throughout the entirety of the acquisition sessions.

Though it is worth noting that two of the sensors, the LG 4000 and the TD100, are made by the same manufacturer, our cross-sensor evaluation still provides an analysis of three state of the art sensors. While other sensors could be included in future studies, these three sensors operate under similar acquisition conditions, including stationary subjects, near-infrared illumination, and tripod-mounted cameras.

B. Acquisition Process, Dataset, and Algorithms

For acquisitions, each sensor was used to collect left and right iris images for subjects over a span of several weeks under a human subjects protocol approved by the Notre Dame Human Subjects Institutional Review Board. Each sensor was mounted on a tripod and adjusted to each subject's height, and the tripods were placed in a row with equal spacing such that all three sensors were equidistant from the visible light sources in the acquisition studio. The controlled environment was designed to provide similar acquisition conditions for each sensor. Subjects approached the IrisGuard, LG 4000, and TD100 in succession with little time between each acquisition. Example images of the same iris acquired by each of the three sensors are shown in Figure 2. Images were acquired using the default settings for each sensor. Additionally, in all acquisition sessions, multiple images were acquired for each subject using each sensor, and the amount of time between successive image acquisitions varied from near-instantaneous, as with the IrisGuard, to several seconds, as with the TD100. All three sensors produced 640x480 pixel grayscale images.

The complete dataset was collected in eight acquisition sessions, which spanned a total of 21 weeks. Within this dataset, there are two subsets. The first subset contains four sessions (Sessions 1-4), which spanned over 12 weeks. During this time, the sensors were ordered such that subjects approached the IrisGuard, the LG 4000, and finally the TD100 sequentially. Additionally, the sensors were moved from one corner of the acquisition studio to another between Sessions 1 and 2, while attempting to maintain the same environmental conditions in the studio. Within Sessions 1-4, 23444 iris images were collected, spanning 510 unique subjects (1020 unique irises). Table I shows a detailed breakdown of the number of iris samples collected and subjects involved in Sessions 1-4 of the acquisitions.

The second subset also covers four sessions of acquisitions, which spanned over 9 weeks and took place after the entire first subset was collected. For this subset (Sessions 5-8), the sensors were rearranged such that the LG 4000 was approached first, followed by the IrisGuard, and finally TD100. In Sessions 5-8, though the order of the sensors was changed, the sensors were in the same locations as in Sessions 2-4. Sessions 5-8 contain a total of 25921 iris images from 476 unique subjects (952 unique irises). Table II shows the breakdown of samples and subjects from Sessions 5-8. The entire dataset (Sessions 1-8) thus consists of 630 unique subjects, or 1260 unique irises, and a total of 49365 iris images. Further, 310 of the total subjects were male and 320 were female, with ages ranging from 19 to 64 and an average age of 23. Also, 498 of the subjects reported their race to be ”White”, 32 reported ”Asian”, and the remaining reported more specific or unknown race.

In general, the number of images acquired within a particular session vary across sensors due to differences in acquisition software. Additionally, images for some subjects are missing for particular sensors due to operator errors.

Three iris matching algorithms - M1, M2, and M3 - were used to compare irises acquired from different acquisition sessions. One the three matchers is an in-house implementation, while the other two are commercially available.

IV. Results

In order to perform matching experiments using images from each of the three sensors, a preprocessing stage was required in which each of the three matching algorithms was used to extract an iris template from each of the original images. To create these templates, the irises were segmented and features extracted using techniques specific to each algorithm. In some cases, the algorithms were unable to produce templates for particular images. Table III shows the breakdown of failed segmentations for the dataset by sensor and algorithm. The failure to create a template can stem from many sources, such as occlusion, blur, and illumination. However, since we cannot obtain data about how the segmentations are performed, we cannot report the exact reason why a particular image failed. Summarizing, images from the TD100 produced the most segmentation failures for M1 and M2, while the LG 4000 produced the most failures for M3. Any image which failed to produce a template using a particular algorithm was omitted from all experiments using that matching algorithm. It should also be noted that successful template generation does not guarantee correct segmentation.

M1, M2, and M3 were each used to compare iris templates from different acquisition sessions. The following sections each describe the results from a different perspective; conclusions are drawn from the matching results with respect to particular factors. In Section IV-A, we first consider the impact of environmental conditions on each sensor. Section IV-B considers pupil dilation as a cause for variations in performance and shows how this affects the performance of single and cross-sensor experiments. Next, Section IV-C observes relationships between single and cross-sensor performance for each of the experiments. Section IV-D summarizes the role of the algorithm in determining sensor performance according to our results, and we finally conclude with Section IV-E, in which we determine whether a relative sensor ranking can be established based on the performance of each sensor in our experiments.

The matching results used in each discussion are divided based on the sensors used to acquire the images. Specifically, there were three single-sensor experiments:

- IrisGuard vs. IrisGaurd
- LG 4000 vs. LG 4000
- TD100 vs. TD100

Similarly, there were three cross-sensor experiments:

- IrisGuard vs. LG 4000
- IrisGuard vs. TD100
- LG 4000 vs. TD100

For example, the IrisGuard vs. IrisGuard experiment for Sessions 1-4 compares all pairs of images collected using the IrisGuard during Sessions 1-4, omitting image pairs that originated from the same session. Similarly, the IrisGuard vs. LG 4000 experiment for Sessions 1-4 compares all images collected using the IrisGuard to all images collected using the LG 4000 from Sessions 1-4, again omitting image pairs acquired during the same session. Each experiment was repeated using each of the three matching algorithms. Comparisons between image pairs originating from the same session were omitted from our results because same-session comparisons tend to produce unnaturally good performance and are not representative of practical biometric applications.

A. How Do Environmental Conditions Impact Performance?

In Sessions 1-4, changes in environmental conditions were minimized throughout all sessions. These environmental conditions include distance between sensor and ambient light sources, distance between sensors, and height of sensor relative to subject. To measure the effect of sensor order on performance, we also considered Sessions 5-8, in which the sensors were placed in the same locations but in a different order; we hypothesize that this reordering should not effect the relative performance of each experiment. In Sessions 1-4, subjects approached the IrisGuard, then the LG 4000, and finally the TD100, while in Sessions 5-8, the subjects approached the LG 4000, then the IrisGuard, and lastly the TD100.

Figure 3 shows ROC curves from the cross-sensor and single-sensor experiments using data from Sessions 1-4, 5-8, and 1-8 for all three matchers. The number of match and non-match comparisons used in each experiment varied depending on both the sensor and matching algorithm. In some cases, the matching algorithms filtered out comparisons which could not generate match scores above a particular confidence threshold; these comparisons are omitted from the results presented in this work. The total number of match and non-match comparisons used in each experiment are shown in Table IV. Table V shows a corresponding breakdown of match and non-match comparisons for Sessions 5-8. Further, on each curve in Figure 3, error bars are shown, which represent 95% confidence intervals calculated via bootstrapping. In particular, match and non-match scores were subsampled according to their score distributions, and an ROC curve was generated for each of 5000 bootstraps and used to create the error bars. If two error bars at a particular FAR do not overlap, then the

TABLE I
DETAILED ACQUISITION SUMMARY - SESSIONS 1-4

	Session 1	Session 2	Session 3	Session 4	Total
IrisGuard	2080 Samples 265 Subjects	1671 Samples 212 Subjects	2345 Samples 302 Subjects	2191 Samples 278 Subjects	8287 Samples 491 Subjects
LG 4000	1606 Samples 269 Subjects	1584 Samples 266 Subjects	2457 Samples 302 Subjects	2651 Samples 319 Subjects	8298 Samples 506 Subjects
TD100	1579 Samples 269 Subjects	1568 Samples 267 Subjects	1799 Samples 301 Subjects	1913 Samples 320 Subjects	6859 Samples 509 Subjects

TABLE II
DETAILED ACQUISITION SUMMARY - SESSIONS 5-8

	Session 5	Session 6	Session 7	Session 8	Total
IrisGuard	1939 Samples 247 Subjects	2131 Samples 269 Subjects	2624 Samples 332 Subjects	2568 Samples 324 Subjects	9262 Samples 463 Subjects
LG 4000	2273 Samples 278 Subjects	2366 Samples 297 Subjects	2658 Samples 334 Subjects	2600 Samples 327 Subjects	9897 Samples 476 Subjects
TD100	1642 Samples 274 Subjects	1775 Samples 296 Subjects	2002 Samples 334 Subjects	1343 Samples 224 Subjects	6762 Samples 456 Subjects

TABLE III
FAILED TEMPLATE GENERATIONS

		M1	M2	M3
IrisGuard	Sessions 1-4	0 Images	0 Images	2 Images
	Sessions 5-8	0 Images	3 Images	2 Images
	Total	0 Images	3 Images	4 Images
LG 4000	Sessions 1-4	0 Images	6 Images	66 Images
	Sessions 5-8	0 Images	2 Images	203 Images
	Total	0 Images	8 Images	269 Images
TD100	Sessions 1-4	0 Images	11 Images	20 Images
	Sessions 5-8	1 Image	12 Images	7 Images
	Total	1 Images	23 Images	27 Images

difference between the two curves at that FAR is statistically significant. However, if the error bars of two curves do overlap, statistical significance cannot be determined without further testing. This method is based upon the technique described by NIST [29]. Using these results, a relative ordering of sensor performance can be established by comparing true accept rate at a particular false accept rate.

By comparing the relative ordering of the sensor performances, we can see that there are variations between the results of each session group. For example, considering the single-sensor experiments using M1 on Sessions 1-4, we see that the TD100 vs TD100 experiment had the worst performance. For Sessions 5-8, however, the LG 4000 vs LG 4000 had the worst performance, though the difference between the TD100 vs TD100 and LG 4000 vs LG 4000 was not determined to be statistically significant for Sessions 5-8. The same reordering between Sessions 1-4 and Sessions 5-8 was observed using M3. Because all environmental conditions were sustained as best as possible between all sessions, it was hypothesized that the relative ordering of sensor performances would not change between Sessions 1-4 and Sessions 5-8. However, since this was not the case, further analysis is required to explore the causes of this reordering.

In order to determine a possible cause of this difference, we examined ROC curves generated by single sessions in order to see if one session had a particularly positive or negative effect on the overall performance. Namely, for each single-sensor and cross-sensor experiment, an ROC curve was generated for each session where a score contributed to the curve if one of the images which generated that score was acquired in that session. When this new separation of data was examined, an interesting trend appeared in several experiments for all matchers. Session 1 data performed significantly different than data from the other seven sessions.

Although all environmental conditions described were controlled between all sessions, one difference was noted in Session 1; the sensors were placed in a different corner of the acquisition studio. Thus although they were the same distance from ambient illuminants, it is possible the ambient lighting was varied, or that some other unknown environmental condition caused an unexpected variation in the data. In order to compare only data acquired in exactly the same location, ROC curves using data only from sessions 2-4 were generated and relative sensor performances were determined using these results. Comparing the relative ordering of performances between sessions 2-4 and Sessions 5-8, the ordering between the two session groupings appeared to be more closely related; yet, it appears that Session 1 did not account for all of the variation between Sessions 1-4 and 5-8, and therefore some other environmental conditions may have caused some variation in performance.

Based on these experiments, it appears that slight changes in environment can cause a change in both single-sensor and cross-sensor performance. Further, it appears that the ordering of sensors during acquisition had less of an effect than other environmental conditions, but it may still have had some influence on performance. Hence, it is important to explore the implications of changing an acquisition environment, even

TABLE IV
NUMBER OF MATCH AND NON-MATCH COMPARISONS IN EACH EXPERIMENT - SESSIONS 1-4

	M1	M2	M3
IrisGuard v IrisGuard	Match = 28,207 Non-Match = 12,786,533	Match = 28,207 Non-Match = 12,786,533	Match = 28,188 Non-Match = 12,772,638
LG 4000 v LG 4000	Match = 27,488 Non-Match = 12,648,502	Match = 27,453 Non-Match = 12,629,379	Match = 26,835 Non-Match = 12,330,790
TD100 v TD100	Match = 18,930 Non-Match = 8,780,601	Match = 18,903 Non-Match = 8,752,069	Match = 18,836 Non-Match = 8,714,118
IrisGuard v LG 4000	Match = 55,542 Non-Match = 25,548,861	Match = 55,494 Non-Match = 25,530,015	Match = 54,817 Non-Match = 25,199,025
IrisGuard v TD100	Match = 46,063 Non-Match = 21,216,720	Match = 46,031 Non-Match = 21,181,421	Match = 45,927 Non-Match = 21,123,435
LG 4000 v TD100	Match = 45,787 Non-Match = 21,156,602	Match = 45,721 Non-Match = 21,105,682	Match = 45,071 Non-Match = 20,848,573

TABLE V
NUMBER OF MATCH AND NON-MATCH COMPARISONS IN EACH EXPERIMENT - SESSIONS 5-8

	M1	M2	M3
IrisGuard v IrisGuard	Match = 38,848 Non-Match = 15,963,016	Match = 38,824 Non-Match = 15,952,339	Match = 38,819 Non-Match = 15,947,408
LG 4000 v LG 4000	Match = 43,844 Non-Match = 18,296,493	Match = 43,820 Non-Match = 18,288,986	Match = 41,234 Non-Match = 17,135,077
TD100 v TD100	Match = 20,920 Non-Match = 8,493,137	Match = 20,843 Non-Match = 8,465,585	Match = 20,837 Non-Match = 8,464,479
IrisGuard v LG 4000	Match = 82,229 Non-Match = 34,200,919	Match = 82,181 Non-Match = 34,182,565	Match = 79,592 Non-Match = 33,165,299
IrisGuard v TD100	Match = 57,215 Non-Match = 23,419,745	Match = 57,099 Non-Match = 23,373,157	Match = 57,081 Non-Match = 23,366,966
LG 4000 v TD100	Match = 60,757 Non-Match = 25,024,704	Match = 60,635 Non-Match = 24,978,771	Match = 58,684 Non-Match = 24,160,176

TABLE VI
NUMBER OF MATCH AND NON-MATCH COMPARISONS IN EACH EXPERIMENT - SESSIONS 1-8

	M1	M2	M3
IrisGuard v IrisGuard	Match = 148,802 Non-Match = 67,047,095	Match = 148,742 Non-Match = 67,023,993	Match = 111,207 Non-Match = 50,390,576
LG 4000 v LG 4000	Match = 157,789 Non-Match = 71,921,310	Match = 157,629 Non-Match = 71,856,781	Match = 109,456 Non-Match = 49,750,487
TD100 v TD100	Match = 88,725 Non-Match = 40,411,729	Match = 88,472 Non-Match = 40,281,028	Match = 64,882 Non-Match = 29,420,719
IrisGuard v LG 4000	Match = 305,143 Non-Match = 139,019,688	Match = 304,919 Non-Match = 138,934,027	Match = 219,858 Non-Match = 100,492,823
IrisGuard v TD100	Match = 229,379 Non-Match = 104,288,109	Match = 229,009 Non-Match = 104,099,125	Match = 134,525 Non-Match = 59,621,531
LG 4000 v TD100	Match = 237,025 Non-Match = 108,043,861	Match = 236,584 Non-Match = 107,819,975	Match = 132,706 Non-Match = 58,885,111

when the environment change may seem to be trivial.

B. How Does Pupil Dilation Affect Sensor Performance?

There are many image factors which may affect recognition performance, including pupil dilation, iris diameter, image focus, and occlusion. In our experiments, we analyze the effect of pupil dilation because dilation statistics were readily available from all three matching algorithms. Further, while it is known that pupil dilation has a large impact on iris biometric performance [18],[14], the effect of dilation on sensor interoperability is a relatively unexplored problem. It has been shown in the past that better performance is achieved when comparing eye images with highly constricted pupils than when comparing irises with highly dilated pupils [18]. Additionally, work such as Hollingsworth el al. suggests that performance increases as the *difference* between the dilation ratios of the images being compared decreases [14]. This section presents an analysis of the impact of the difference between dilation ratios of images in the context of our single and cross-sensor experiments.

We define the dilation ratio of image i to be

$$DR_i = \frac{Pupil\_Radius_i}{Iris\_Radius_i} \tag{1}$$

Using this notation, an image with a more constricted pupil would have a smaller dilation ratio. The dilatio ratio has a range between 0.0 and 1.0, though realistic values tend to fall between 0.2 and 0.6. Table VII shows statistics for the three sensors according to the segmentation routines of M1, M2, and M3. Summarizing, images taken using the IrisGuard generally had smaller dilation ratios than the other two sensor, and images taken using the LG 4000 generally had larger dilation ratios than the other two sensors. Further, this trend was observed in both Sessions 1-4 and Sessions 5-8, and thus

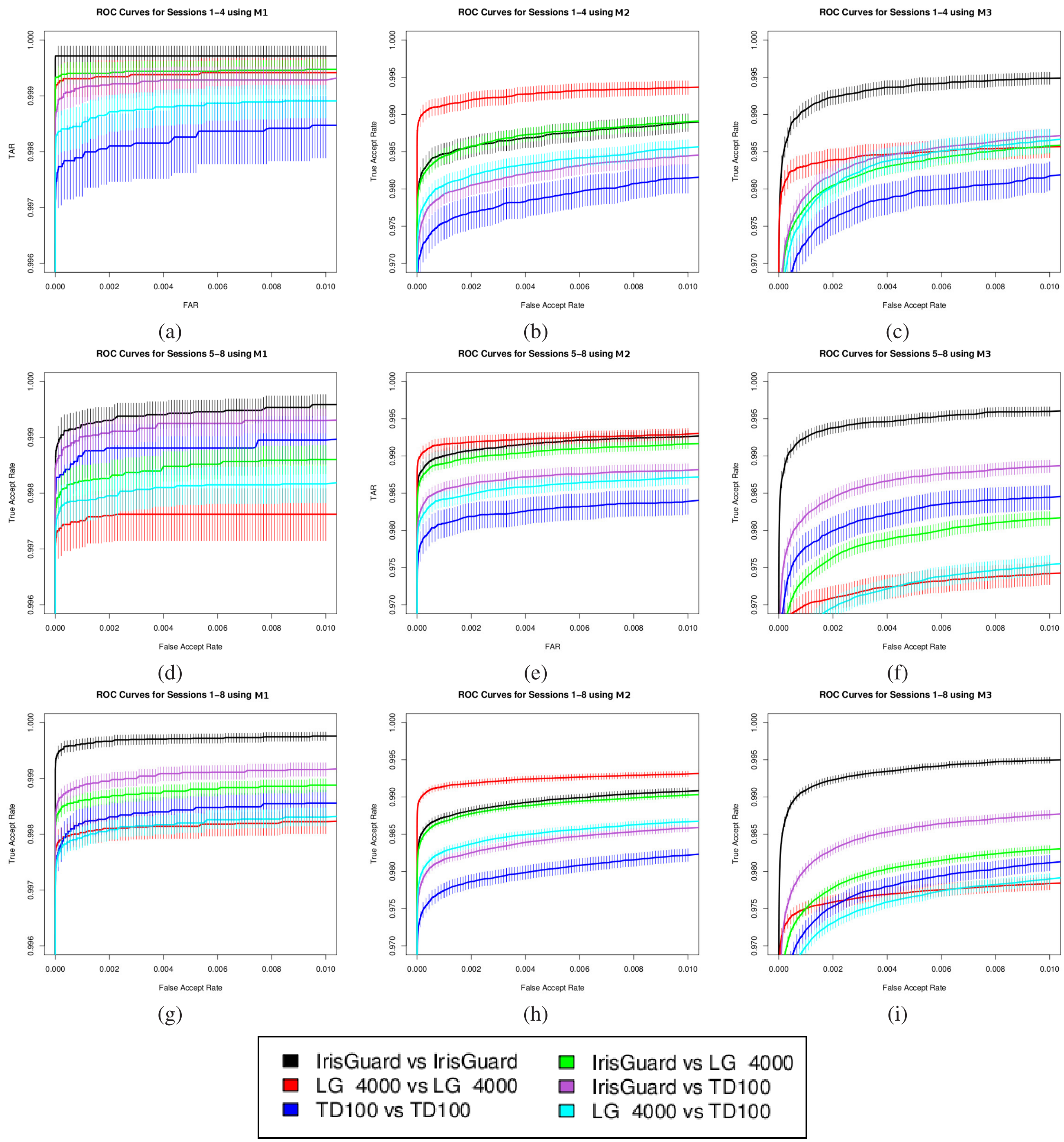


Fig. 3. ROC curves for Sessions 1-4 (top row), Sessions 5-8 (middle row), and Sessions 1-8 (bottom row) for all three matching algorithms. Note that the range of y-axis varies between figures.

the relative ordering of the sensors was not a major factor affecting the dilation ratio of the images acquired. It is likely that the use of visible light illuminants, in addition to NIR LEDs, contributed to the smaller dilation ratios observed using the IrisGuard.

The difference in dilation ratio (DDR) between image i and image j is defined as

$$DDR_{i,j} = abs(DR_i - DR_j) \tag{2}$$

The DDRs for all comparisons were computed and aggregated for each experiment in this study. To obtain each DDR, statistics from each software package was obtained regarding the size of the iris and pupil. The dilation ratio for each image was then calculated and compared to all other images in an experiment to obtain DRR's. As an example, Figure 4 shows the DDR distributions for all comparisons in the TD100 vs. TD100 (Sessions 1-4) experiment according to M2, separated by session. Because no single-session comparisons are used in this study, the curve for each session represents all com-

TABLE VII
AVERAGE DILATION RATIO FOR EACH SENSOR

	M1	M2	M3
IrisGuard	$\overline{DR}_{1-4}$ = 0.374 ($\sigma = 0.051$) $\overline{DR}_{5-8}$ = 0.382 ($\sigma = 0.046$) $\overline{DR}_{1-8}$ = 0.378 ($\sigma = 0.049$)	$\overline{DR}_{1-4}$ = 0.342 ($\sigma = 0.051$) $\overline{DR}_{5-8}$ = 0.350 ($\sigma = 0.049$) $\overline{DR}_{1-8}$ = 0.345 ($\sigma = 0.048$)	$\overline{DR}_{1-4}$ = 0.343 ($\sigma = 0.052$) $\overline{DR}_{5-8}$ = 0.351 ($\sigma = 0.047$) $\overline{DR}_{1-8}$ = 0.347 ($\sigma = 0.050$)
LG 4000	$\overline{DR}_{1-4}$ = 0.417 ($\sigma = 0.066$) $\overline{DR}_{5-8}$ = 0.414 ($\sigma = 0.063$) $\overline{DR}_{1-8}$ = 0.415 ($\sigma = 0.065$)	$\overline{DR}_{1-4}$ = 0.383 ($\sigma = 0.062$) $\overline{DR}_{5-8}$ = 0.393 ($\sigma = 0.062$) $\overline{DR}_{1-8}$ = 0.382 ($\sigma = 0.061$)	$\overline{DR}_{1-4}$ = 0.389 ($\sigma = 0.065$) $\overline{DR}_{5-8}$ = 0.383 ($\sigma = 0.066$) $\overline{DR}_{1-8}$ = 0.386 ($\sigma = 0.065$)
TD100	$\overline{DR}_{1-4}$ = 0.408 ($\sigma = 0.068$) $\overline{DR}_{5-8}$ = 0.414 ($\sigma = 0.061$) $\overline{DR}_{1-8}$ = 0.411 ($\sigma = 0.065$)	$\overline{DR}_{1-4}$ = 0.373 ($\sigma = 0.066$) $\overline{DR}_{5-8}$ = 0.388 ($\sigma = 0.063$) $\overline{DR}_{1-8}$ = 0.376 ($\sigma = 0.063$)	$\overline{DR}_{1-4}$ = 0.379 ($\sigma = 0.064$) $\overline{DR}_{5-8}$ = 0.384 ($\sigma = 0.059$) $\overline{DR}_{1-8}$ = 0.382 ($\sigma = 0.062$)

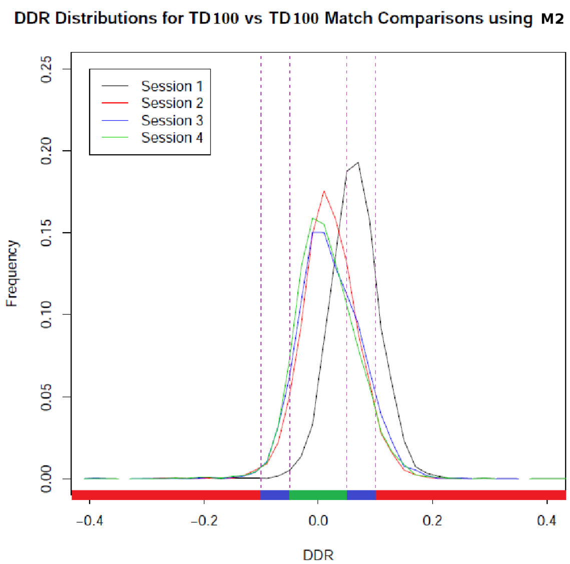


Fig. 4. Distributions of difference in dilation ratios for the TD100 vs TD100 (Sessions 1-4) experiment, separated by session. The x-axis is marked to indicate the DDR intervals. The region in green indicates DDR_Bin_1, the regions in yellow indicate DDR_Bin_2, and the regions in red indicate DDR_Bin_3.

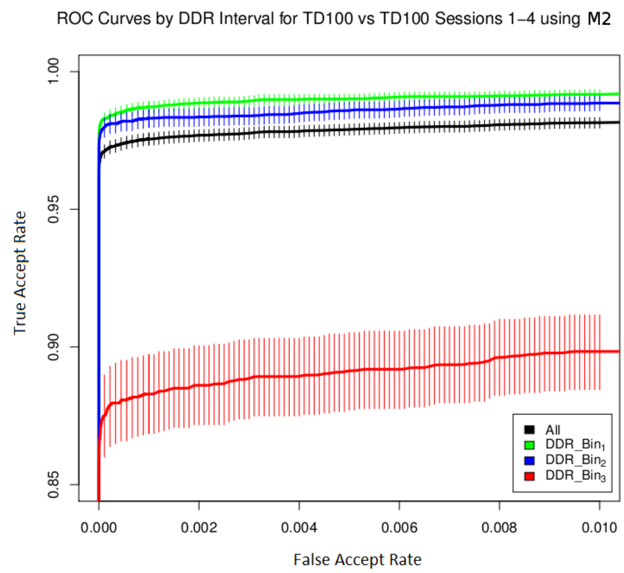


Fig. 5. ROC curves for the TD100 vs TD100 (Sessions 1-4) experiment, separated by DDR interval. The black curve indicates the combined performance of all DDR intervals.

parisons involving one image from that particular session. For example, the Session 1 curve includes all of the comparisons between Session 1 TD100 images and any other TD100 image from Sessions 2-4. As shown in Figure 4, the DDRs using images from Session 1 are dramatically different than the DDRs from the other sessions. This trend was observed in all experiments involving Session 1, but was less pronounced in experiments involving the LG 4000. Upon investigation, it was found that images from Session 1 from all three sensors had much smaller dilation ratios than images taken from the other seven sessions. As an example, the average dilation ratio for the TD100 images from Session 1 was 0.320, whereas the average dilation ratio for TD100 images from Sessions 2-8 was 0.383. This change in dilation ratio was less dramatic for the LG 4000; the average dilation ratio was 0.354 for Session 1 and 0.384 for Sessions 2-8. As mentioned in Section IV-A, the images in Session 1 were acquired with the sensors in a different location than the images acquired in Sessions 2-8, and this environment change is likely to be the cause of the change in dilation ratios between the sessions.

To analyze the impact of the DDR on system performance, match and non-match comparisons were separated into intervals based on the DDR between each image pair. The first DDR interval, DDR_Bin_1, included all comparisons for which $0 \leq abs(DDR) < 0.05$. DDR_Bin_2 included all comparisons for which $0.05 \leq abs(DDR) < 0.1$. The final DDR interval, DDR_Bin_3, included all comparisons for which $abs(DDR) \geq 0.1$. Figure 4 indicates the three DDR intervals along the x-axis. From Figure 4 it can be seen that Sessions 2-4 have DDR distributions centered within DDR_Bin_1, while the DDR distribution for Session 1 is centered within DDR_Bin_2.

ROC Curves were generated for each DDR interval for each cross-sensor and single-sensor experiment. Figure 5 shows theses ROC curves for the TD100 vs TD100 (Sessions 1-4). Summarizing, DDR_Bin_1 had the best performance, while DDR_Bin_3 by far had the worst performance. In the case of the TD100 vs TD100 experiments, we could not claim statistical significance between the performance of DDR_Bin_1 and DDR_Bin_2 at any FAR, but the relative ordering of performance of the three DDR intervals was consistent across all single-sensor and cross-sensor experiments using all three matchers. These results support the claim that lower DDRs yield higher performance in both single-sensor and cross-sensor contexts.

Figures 6a-f show the ROC curves for DDR_Bin_1 and DDR_Bin_2 using all three matching algorithms on Sessions 5-8. In all cases, performance decreased for DDR_Bin_2 when compared to DDR_Bin_1. We observed that the relative ordering of experiment performance between DDR intervals was mostly unchanged (aside from statistical noise) for all matchers and sensor combinations. The exception to this was the IrisGuard vs LG 4000 experiment using M2. In Figure 6c, the IrisGuard vs LG 4000 experiment was among the worst at an FAR of 0.001 for DDR_Bin_1. In Figure 6d, we see that this combination of sensors performs significantly better than the other sensor combinations at the same FAR, indicating that it was particularly robust to changes in DDR. This trend was not observed using the other matching algorithms. In Sessions 1-4 using M2 (not shown for space considerations), the IrisGuard vs LG 4000 was fairly robust to changes in DDR, although its performance difference from several other experiments was not statistically significant for DDR_Bin_1 and DDR_Bin_2. For both Sessions 1-4 and Sessions 5-8, the IrisGuard vs LG 4000 experiment performed significantly better than all other sensor combinations for DDR_Bin_3 as well. These results suggest that some combinations of algorithms and sensors are more robust to dilation ratio differences.

The environment changes between Session 1 and Session 2 caused a dramatic change in the dilation ratios of the images acquired using all three sensors, though the change was noticeably smaller for the LG 4000. This fact, along with the results shown in Figures 6a-f could account for some of the performance difference between Sessions 1-4 and Sessions 2-4 for all of the experiments; the inclusion of Session 1 results in an average increase in DDR, which negatively affects performance. Further, this has an insignificant impact on the LG 4000 vs LG 4000 experiments, as the change in average dilation ratio was smallest for this sensor. While the increased DDR resulting from the inclusion of Session 1 accounts for some of the performance difference between Sessions 1-4 and Session 2-4, it does not account for all of the change. Figure 7 shows the ROC curves for DDR_Bin_1 for M2 for Sessions 1-4. As can be seen, the LG 4000 vs LG 4000 experiment still performs significantly better than the other experiments even when DDR is taken into consideration. Further, removing Session 1 from the experiments still results in a significant increase in performance for all of the M2 experiments, excluding the LG 4000 vs LG 4000 experiment. This suggests that there is still some other unidentified effect resulting from the Session 1 environment change beyond change in dilation ratio.

C. What is the Relationship Between Single and Cross-Sensor Performance?

The results of these matching experiments can also be used to evaluate the interoperability of each of the three sensors and to investigate how the single-sensor performance relates to the cross-sensor performances.

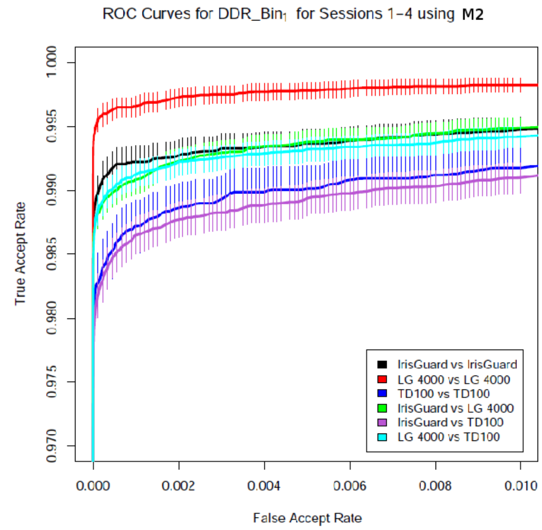


Fig. 7. ROC curves for DDR_Bin_1 using M2 on Sessions 1-4.

From the M1 results shown in Figures 3a, d, and g, it can be seen that single-sensor performance is an accurate predictor of the relative cross-sensor performance; in all session groups shown, the relative ordering of the cross-sensor experiments by performance was consistent with the relative ordering of the single-sensor experiments. For example, in Sessions 1-4, the IrisGuard vs IrisGuard experiment had the best performance of the single-sensor experiments (measured by TAR at FAR=0.001), followed by the LG 4000 vs LG 4000, and finally the TD100 vs TD100. Accordingly, the best cross-sensor performance was the IrisGuard vs LG 4000, followed by the IrisGuard vs TD100, and finally the LG 4000 vs TD100. Additionally, in almost all cases, the performance of each cross-sensor experiment was between the performance of the corresponding single-sensor experiments. For example, the performance of the IrisGuard vs LG 4000 experiment was between the performance of the IrisGuard vs IrisGuard and LG 4000 vs LG 4000 experiments. The exception to this was for Sessions 1-8, where the LG 4000 vs TD100 was worse than either single-sensor experiment, though the difference between the performances of the LG 4000 vs TD100 experiment and the LG 4000 vs LG 4000 experiment were not determined to be statistically significant.

Using the M2 results in Figures 3b, e, and h, we see that single-sensor performance predicts the relative ordering of cross-sensor performance for Sessions 1-4 and Sessions 1-8, but not for Sessions 5-8. In Sessions 5-8, the IrisGuard vs TD100 experiment performs better than the LG 4000 vs TD100 experiment despite the LG 4000 having better single-sensor performance than the IrisGuard, though the difference between the performance of the IrisGuard vs TD100 experiment and the LG 4000 vs TD100 experiment were not determined to be statistically significant. In all session groups considered, the IrisGuard vs TD100 performance and the LG 4000 vs TD100 performance were between the performances of the corresponding single-sensors. The performance of the IrisGuard vs LG 4000 experiment was always worse than both

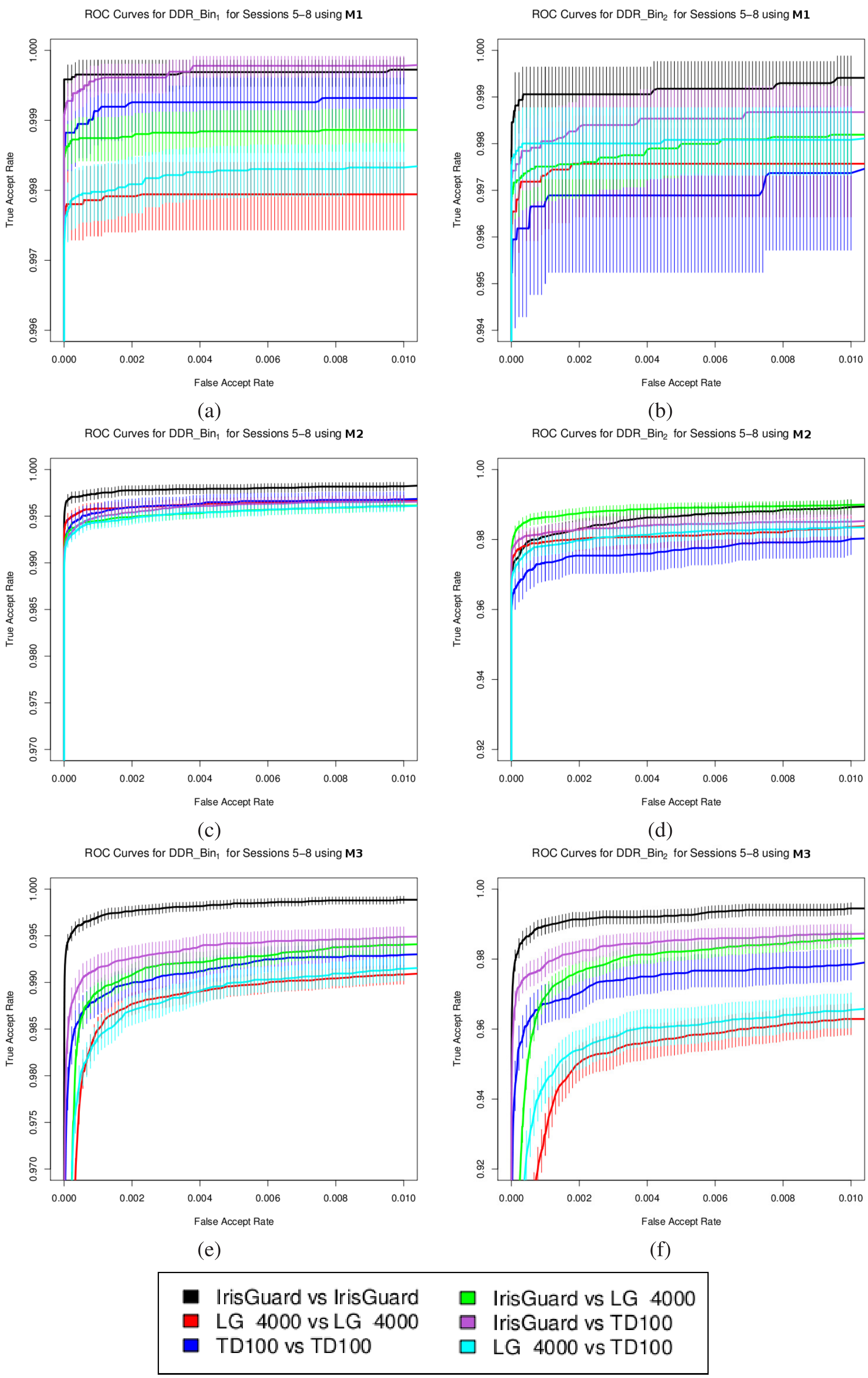


Fig. 6. ROC curves for DDR_Bin_1 and DDR_Bin_2 using all three matching algorithms on Sessions 5-8. Note that the range of y-axis varies between figures.

corresponding single-sensor experiments, though the difference between the IrisGuard vs LG 4000 performance and the IrisGuard vs IrisGuard experiment was not always significant.

Finally, from the results in Figures 3c, f, and i, our experiments showed that using M3, single-sensor performance did not predict relative cross-sensor performance for Sessions 1-4, where the IrisGuard vs TD100 experiment had the best cross-sensor performance, despite the TD100 vs TD100 having the worst single-sensor experiment. In Sessions 1-4, however, the difference between all three cross-sensor experiments were not determined to be statistically significant. In Sessions 5-8, single-sensor performance was an accurate predictor of cross-sensor performance. In Sessions 1-8, the IrisGuard vs TD100 experiment again had the best cross-sensor performance. In general, using M3, the relative ordering of single and cross-sensor performance was more sensitive to the FAR being considered when compared to the other matching algorithms, making it hard to derive strong conclusions from this matcher.

We also found that in no case did a cross-sensor experiment achieve higher performance than both of the corresponding single-sensor experiments.

D. What is the Role of the Matching Algorithm in Sensor Performance?

Another perspective from which to look at the results of these experiments can be observed by examining how the performance of each sensor was affected by the choice of matching algorithm. This section makes general observations about the performance of the three algorithms, and considers whether relative sensor performance was dependent on the matching algorithm.

From Figures 3a-i, it is clear that the M1 algorithm outperformed the other two matching algorithms on all six experiments for Sessions 1-4, 5-8 and 1-8. Using M1 for these experiments, the worst TAR at FAR = 0.001 was 0.9975, achieved by the LG 4000 vs LG 4000 experiment for Sessions 5-8. This TAR was still significantly better than the *best* performing experiment from the other two matchers, which was the IrisGuard vs IrisGuard experiment for Sessions 5-8 using M3 (TAR = 0.9925). The TAR's for all experiments are available in Table VIII. Beyond the obvious advantage of using M1 over the other two algorithms, it is difficult to make any strong conclusions about the general performance difference between M2 and M3.

Using Figures 3a-i and comparing the performance of each experiment across all three matching algorithms, it can be seen that while the IrisGuard vs IrisGuard experiment had the best performance in all session groups using M1 and M3, the LG 4000 vs LG 4000 experiment had the best performance in all session groups using M2. The difference between the LG 4000 vs LG 4000 and IrisGuard vs IrisGuard performances was statistically significant for Sessions 1-4 and Session 1-8, though the performances for these two experiments were much closer for Sessions 2-4 and 5-8. The reason the IrisGuard vs IrisGuard and LG 4000 vs LG 4000 performances are much closer for Sessions 2-4 and 5-8 is because the LG 4000 vs LG 4000 experiment using M2 was particularly unaffected by the environmental effect that impacted the Session 1 data. In fact, using M2, the performance of the LG 4000 vs LG 4000 experiment actually had slightly better performance for Sessions 1-4 than for Sessions 2-4. For M1 and M3, the Session 1-4 performance for the LG 4000 vs LG 4000 experiment was significantly lower than for Sessions 2-4. This suggests that M2 may be more robust to *small* fluctuations in the undetermined environment factor (as in the case of the LG 4000 data from Session 1), but is still affected by large variations. It is also possible that using M2 in combination with the LG 4000 inherently corrects for the environment condition. Regardless of whether or not the Session 1 data is considered in the study, it is clear that compared to the other sensors in a single-sensor context, the relative performance of the LG 4000 is better when using M2. In the cross-sensor context using M2, the IrisGuard vs LG 4000 experiment had the best performance for all session groups. The LG 4000 vs TD100 experiment outperformed the IrisGuard vs TD100 experiment only in session groups that included Session 1 data, though the difference between these two experiments was not determined to be significant.

As previously shown, both M1 and M3 yielded the best performance when using data collected with the IrisGuard. Beyond this trend, there is substantial reordering of the relative performances of the TD100 and LG 4000 depending on which session groups are considered. This indicates that for these matchers, the relative performance of the system is more heavily affected by environmental conditions than by the choice of sensor.

E. Can a Relative Sensor Ranking be Established?

Finally, the results of the experiments shown in this work can be used to measure the relative performance of each sensor. To evaluate the relative performance of each sensor, we compare the TAR of each experiment at an FAR of 0.001, while also considering the 95% confidence intervals generated by the previously described bootstrapping. Table VIII shows the TAR and error associated with each experiment for sessions 1-4, 5-8, and 1-8.

Using the values in Table VIII, the experiments were ordered by TAR to show relative ranking by session group, shown in Table IX. Summarizing, IrisGuard appears to have the best single-sensor results across all matchers, ranking first in both M1 and M3, and second in M2 for all sessions. The LG 4000 has varied performance among the matchers. Using M2, the LG 4000 vs LG 4000 has the best performance for all sessions. M1 and M3 rank LG 4000 vs LG 4000 highly for Sessions 1-4, but low for Sessions 5-8 and 1-8. The final same sensor experiment, TD100 vs TD100, ranks poorly in most instances; M2 ranks it at the lowest ranking for all sessions, whereas M1 and M3 ranked it well for session 5-8 but poorly otherwise. In general, cross-sensor experiment rankings showed little consistency. For M1, in Sessions 1-4 and 5-8 all cross-sensor rankings were between the corresponding single-sensors. This pattern was not seen for Sessions 1-8 using M1; the LG 4000 vs TD100 performed worse than both single-sensor experiments, though the difference may not have

TABLE VIII
TRUE ACCEPT RATES AT A FALSE ACCEPT RATE OF 0.001

		M1	M2	M3
IrisGuard vs IrisGuard	Sessions 1-4	0.9997 ± 0.0002	0.9847 ± 0.0014	0.9906 ± 0.0011
	Sessions 5-8	0.9992 ± 0.0003	0.9899 ± 0.0010	0.9925 ± 0.0008
	Sessions 1-8	0.9996 ± 0.0001	0.9872 ± 0.0006	0.9909 ± 0.0005
LG 4000 vs LG 4000	Sessions 1-4	0.9993 ± 0.0003	0.9912 ± 0.0011	0.9830 ± 0.0015
	Sessions 5-8	0.9975 ± 0.0005	0.9915 ± 0.0009	0.9699 ± 0.0016
	Sessions 1-8	0.9980 ± 0.0002	0.9914 ± 0.0004	0.9749 ± 0.0009
TD100 vs TD100	Sessions 1-4	0.9978 ± 0.0007	0.9755 ± 0.0023	0.9727 ± 0.0024
	Sessions 5-8	0.9986 ± 0.0005	0.9809 ± 0.0019	0.9777 ± 0.0020
	Sessions 1-8	0.9981 ± 0.0003	0.9774 ± 0.0010	0.9721 ± 0.0012
IrisGuard vs LG 4000	Sessions 1-4	0.9994 ± 0.0002	0.9844 ± 0.0011	0.9775 ± 0.0012
	Sessions 5-8	0.9982 ± 0.0003	0.9889 ± 0.0007	0.9736 ± 0.0011
	Sessions 1-8	0.9986 ± 0.0001	0.9867 ± 0.0004	0.9750 ± 0.0006
IrisGuard vs TD100	Sessions 1-4	0.9991 ± 0.0003	0.9793 ± 0.0014	0.9791 ± 0.0013
	Sessions 5-8	0.9990 ± 0.0003	0.9855 ± 0.0010	0.9820 ± 0.0011
	Sessions 1-8	0.9988 ± 0.0001	0.9816 ± 0.0006	0.9803 ± 0.0007
LG 4000 vs TD100	Sessions 1-4	0.9985 ± 0.0004	0.9805 ± 0.0013	0.9768 ± 0.0021
	Sessions 5-8	0.9978 ± 0.0004	0.9841 ± 0.0010	0.9665 ± 0.0014
	Sessions 1-8	0.9979 ± 0.0002	0.9826 ± 0.0005	0.9699 ± 0.0009

been significant. M2 appeared to have the most consistent ranking across session groups. For all sessions, ranks 1 through 3 remain consistent. Lastly, M3 shows the least amount of consistency between session rankings. Although IrisGuard vs IrisGuard appears in the first rank for all session groups, no other clear patterns are present.

Based on these results and rankings, it is difficult to establish a clear sensor ranking. Considering the rankings of all three matchers, it appears the IrisGuard would be the best sensor of the three, because the IrisGuard vs IrisGuard same sensor results were consistently among the best performing experiments, and cross-sensor results involving the IrisGuard often ranked better than the single-sensor results using the LG 4000 or TD100. There is significant disagreement between the matchers and session groups regarding whether the TD100 or the LG 4000 has better overall performance. The LG 4000 has the best performance using M2, but is sometimes out-performed by the TD100 using M1 and M3. From this, we conclude that while the IrisGuard appears to have the most robust performance across algorithms, the relative performances of the TD100 and LG 4000 using M1 or M3 are more affected by acquisition environment and data quality than sensor selection.

V. CONCLUSIONS

This work presents experiments which compare three commercially available iris sensors, the IrisGuard AD100, the LG IrisAccess 4000, and the LG TD100. The results are used to investigate the robustness of each sensor to changes in environment as well as difference in dilation ratio between image pairs. We also analyze the relationship between single and cross-sensor performance for these sensors, and observe the role of the matching algorithm on relative sensor performance.

Summarizing, we observed the following:

- Seemingly trivial changes in acquisition environment can have major impacts on performance. This was observed for all three sensors and all three algorithms, though the LG 4000 vs LG 4000 experiment using M2 appeared to be the most robust to changes in environment.
- Lower DDRs yield higher performance in both single-sensor and cross-sensor contexts.
- Some combinations of algorithms and sensors seem to be more robust to larger DDRs and particular changes in environment. From this we conclude that the relationship between matching algorithm, sensors, and acquisition environment should be considered when designing a system for uncontrolled acquisition environments.
- Relative cross-sensor performance was generally predicted by relative single-sensor performance when using M1 and M2, with rare exceptions. Similarly, cross-sensor performance was often between the performance of the corresponding single-sensor experiments when using M1 and M2, again with rare exceptions. In no instances did our cross-sensor experiments out-perform the corresponding single-sensor experiments. In general, results from M3 were inconclusive regarding the relationship between single-sensor and cross-sensor performance. From a practical perspective, this means that given an existing biometric system, introducing a higher quality sensor to be used for cross-sensor comparisons will generally result in performance that falls between the single-sensor performance of the new sensor and existing sensor, although in rare cases it may degrade performance rather than improve it.
- The IrisGuard appears to have the most robust performance across all of the algorithms and session groups tested in this work. When choosing between the LG 4000 and TD100 using M1 and M3, the relative performance of the system is more heavily affected by environmental conditions and data quality than by the choice of sensor.

In future work, we hope to further explore how specific changes in acquisition environment, such as ambient illumination levels, affect sensor performance. Additionally, we may replicate our analysis of the impact of dilation ratio difference to explore other image factors such as iris diameter, image focus, and occlusion.

TABLE IX
SENSOR RANKINGS BASED ON TRUE ACCEPT RATE AT A FALSE ACCEPT RATE OF 0.001

	Ranking	M1	M2	M3
1	Sessions 1-4	IrisGuard vs IrisGuard	LG 4000 vs LG 4000	IrisGuard vs IrisGuard
	Sessions 5-8	IrisGuard vs IrisGuard	LG 4000 vs LG 4000	IrisGuard vs IrisGuard
	Sessions 1-8	IrisGuard vs IrisGuard	LG 4000 vs LG 4000	IrisGuard vs IrisGuard
2	Sessions 1-4	IrisGuard vs LG 4000	IrisGuard vs IrisGuard	LG 4000 vs LG 4000
	Sessions 5-8	IrisGuard vs TD100	IrisGuard vs IrisGuard	IrisGuard vs TD100
	Sessions 1-8	IrisGuard vs TD100	IrisGuard vs IrisGuard	IrisGuard vs TD100
3	Sessions 1-4	LG 4000 vs LG 4000	IrisGuard vs LG 4000	IrisGuard vs TD100
	Sessions 5-8	TD100 vs TD100	IrisGuard vs LG 4000	TD100 vs TD100
	Sessions 1-8	IrisGuard vs LG 4000	IrisGuard vs LG 4000	IrisGuard vs LG 4000
4	Sessions 1-4	IrisGuard vs TD100	LG 4000 vs TD100	IrisGuard vs LG 4000
	Sessions 5-8	IrisGuard vs LG 4000	IrisGuard vs TD100	IrisGuard vs LG 4000
	Sessions 1-8	TD100 vs TD100	LG 4000 vs TD100	LG 4000 vs LG 4000
5	Sessions 1-4	LG 4000 vs TD100	IrisGuard vs TD100	LG 4000 vs TD100
	Sessions 5-8	LG 4000 vs TD100	LG 4000 vs TD100	LG 4000 vs LG 4000
	Sessions 1-8	LG 4000 vs LG 4000	IrisGuard vs TD100	TD100 vs TD100
6	Sessions 1-4	TD100 vs TD100	TD100 vs TD100	TD100 vs TD100
	Sessions 5-8	LG 4000 vs LG 4000	TD100 vs TD100	LG 4000 vs TD100
	Sessions 1-8	LG 4000 vs TD100	TD100 vs TD100	LG 4000 vs TD100

REFERENCES

[1] A. Ross and A. Jain, "Biometric sensor interoperability: A case study in fingerprints," in *Proceedings of the International ECCV Workshop on Biometric Authentication (BioAW)*, vol. 3087. Springer Publishers, May 2004, pp. 134–145.

[2] S. K. Modi, "Analysis of fingerprint sensor interoperability on system performance," PhD Thesis. Purdue University, August 2008.

[3] F. Alonso-Fernandez, J. Fierrex-Aguilar, and J. Ortega-Garcia, "Sensor interoperability and fusion in signature verification: A case study using tablet pc," in *Lecture Notes in Computer Science*, vol. 3781, 2005, pp. 180–187.

[4] A. Martin, M. Przybocki, G. Doddington, and D. Reynolds, "The nist speaker recognition evaluation - overview, methodology, systems, results, perspectives," pp. 225–254, 2000.

[5] Y. E. Du, "Review of iris recognition: Cameras, systems, and their applications," in *Sensor Review*. Emerald Group Publishing Limited, 2006, vol. 26, no. 1, pp. 66–69.

[6] M. Theofanos, B. Stanton, R. Micheals, and S. Orandi, "Biometric systematic uncertainty and the user," in *First IEEE International Conference on Biometrics: Theory, Applications, and Systems*, September 2007, pp. 1–6.

[7] F. Alonso-Ferdandez, J. Fierrez, and J. Ortega-Garcia, "Quality measures in biometric systems," in *IEEE Transactions on Security and Privacy*, vol. 99, December 2011.

[8] E. Kukula, M. Sutton, and S. Elliott, "The human-biometric-sensor interaction evaluation method: Biometric performance and usability measurements," in *IEEE Transactions on Instrumentation and Measurement*, vol. 59, March 2010, pp. 784–791.

[9] K. W. Bowyer, S. E. Baker, A. Hentz, K. Hollingsworth, T. Peters, and P. J. Flynn, "Factors that degrade the match distribution in iris biometrics," in *Identity in the Information Society*. Springer, 2010, vol. 2, no. 3, pp. 327–343.

[10] International Biometrics Group, "Independent testing of iris recognition technology," [online] http://www.biometricscatalog.org/itirt/itirt-FinalReport.pdf, May 2005.

[11] Authenti-Corp, "Final report: iris recognition study 2006 (IRIS06)," [online] http://www.nist.gov/customcf/get_pdf.cfm?pub_id=903606, 2006.

[12] E. Newton and P. Phillips, "Meta-analysis of third-party evaluations of iris recognition," in *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 39, January 2009, pp. 4–11.

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

[14] K. Hollingsworth, K. W. Bowyer, and P. J. Flynn, "Pupil dilaton degrades iris biometric performance," in *Computer Vision and Image Understanding*, vol. 113. Elsevier, 2009, pp. 150–157.

[15] S. E. Baker, K. W. Bowyer, and P. J. Flynn, "Empirical evidence for correct iris match score degradation with increased time-lapse between gallery and probe matches," in *International Conference on Biometrics*, June 2009, pp. 1170–1179.

[16] S. E. Baker, K. W. Bowyer, P. J. Flynn, and P. J. Phillips, "Degradation of iris recognition performance due to non-cosmetic prescription contact lenses," in *Computer Vision and Image Understanding*, vol. 114, September 2010, pp. 1030–1044.

[17] P. Grother, E. Tabassi, G. W. Quinn, and W. Salamon, "IREX 1: Performance of Iris Recognition Algorithms on Standard Images," [online] http://www.nist.gov/customcf/get_pdf.cfm?pub_id=903606, 2009.

[18] K. Hollingsworth, K. W. Bowyer, and P. Flynn, "The importance of small pupils: a study of how pupil dilation affects iris biometrics," in *IEEE Second International Conference on Biometrics: Theory, Applications, and Systems*, September 2008, pp. 1–6.

[19] E. Ortiz and K. Bowyer, "Dilation aware multi-image enrollment for iris biometrics," in *International Joint Conference on Biometrics*, October 2011.

[20] R. Connaughton, A. Sgroi, K. W. Bowyer, and P. Flynn, "A cross sensor evaluation of three commercial iris cameras for iris biometrics," in *IEEE Conference on Computer Vision and Pattern Recognitions Workshop on Biometrics*, June 2011, pp. 97–140.

[21] C. Rathgeb, A. Uhl, and P. Wild, "Incremental iris recognition: A single-algorithm serial fusion strategy to optimize time complexity," in *IEEE Fourth International Conference on Biometrics: Theory, Applications, and Systems*, October 2010, pp. 1–6.

[22] P. Johnson, P. Lopez-Meer, N. Sazonova, S. Schuckers, and F. Hua, "Quality in face and iris research ensemble (q-fire)," in *IEEE Fourth International Conference on Biometrics: Theory, Applications, and Systems*, October 2010, pp. 1–6.

[23] "IrisGuard Inc's IG-AD100 Dual-Eye Iris Recognition Camera Accomplishes Sweeping Results," [online] http://www.irisguard.com/uploads/IBG/ IG-AD100 Press Release.pdf, October 2009.

[24] IrisGuard, "IG-AD100 : Iris Camera System," [online] http://www.irisguard.com/uploads/AD100ProductSheet.pdf, 2008.

[25] G. Davies, "Lamp Standard Classification of an IrisGuard AD100 LED Device for IrisGuard," through Lucid Optical Services Ltd., August 2009.

[26] LG Electronics U.S.A., Inc , Iris Technology Division, "IrisAccess 4000 : next generation iris recognition system," [online] http://www.irisid.com/download/brochure/IrisID_IrisAccess4000.pdf.

[27] Korea Testing Laboratory, "IEC 60825-1 : Safety of Laser Products," Safety Report, September 2006.

[28] Iris ID Systems, Inc, "iCam TD100 : Iris Recognition and Face Camera," [online] http://www.irisid.com/download/brochure/IrisID_TD100.pdf, 2009.

[29] J. C. Wu, A. Martin, and R. N. Kacker, "Measures, uncertainties, and significance test in operational roc analysis," in *Journal of Research of the National Institute of Standards and Technology*, January 2011, vol. 15, pp. 517–537.