NISTIR 7440

Meta-Analysis of Third-Party Evaluations of Iris Recognition

Elaine M. Newton
P. Jonathon Phillips

NIST

National Institute of Standards and Technology
U.S. Department of Commerce

Meta-Analysis of Third-Party Evaluations of Iris Recognition

Elaine M. Newton
P. Jonathon Phillips

Image Group
Information Access Division
Information Technology Laboratory

August 2007

Meta-Analysis of Third-Party Evaluations of Iris Recognition

Elaine M. Newton and P. Jonathon Phillips

National Institute of Standards and Technology
Gaithersburg, MD 20899
e-mail: enewton@.nist.gov; jonathon@nist.gov

Abstract— **Iris recognition has long been widely regarded as a highly accurate biometric, despite the lack of independent, large-scale testing of its performance. Recently, however, three third-party evaluations of iris recognition were performed. This paper compares and contrasts the results of these independent evaluations. We find that despite differences in methods, hardware, and/or software, all three studies report error rates of the same order of magnitude: observed false non-match rates (FNMRs) from 0.0122 to 0.03847 at a false match rate (FMR) of 0.001. Further, the differences between the best performers' error rates are an order of magnitude smaller than the observed error rates.**

I. INTRODUCTION

Despite the prior lack of third-party testing of iris matching recognition, the conventional wisdom in the biometrics community has been that iris recognition is highly accurate – even the most accurate biometric. One of many examples of this belief is a comparative table in a seminal biometrics book, which ranks various types of biometrics' abilities based on the perception of three biometrics experts [1]. The table ranks the iris biometric as having "High" performance, along with DNA, fingerprint, and retina. (All others biometric examples listed had a medium or low ranking.) Another example of the conventional wisdom in the biometrics community is this statement from a biometric newsletter, which identifies itself as "the most established source of authoritative news, analysis, and surveys on the international biometrics market": "There is no denying that iris recognition is the most accurate biometric technology,…" [2].

Between May 2005 and March 2007, three major tests on iris recognition were released – the first of their kind. These tests were the Independent Testing of Iris Recognition Technology (ITIRT) conducted by the International Biometric Group (IBG) [3], the Iris RecognItion Study 2006 (IRIS06) conducted by Authenti-Corp (AC) [4], and the Iris Challenge Evaluation (ICE 2006) conducted by the National Institute of Standards and Technology (NIST) [5].

ICE 2006 was a technology evaluation that measured the performance of iris matching algorithms. A technology evaluation is an assessment of the performance of the underlying technology [6]. ITIRT and IRIS06 were scenario evaluations (but included other types of testing as well). Scenario evaluations assess how well a biometric technology meets the requirements, for a particular class of applications; e.g., verification performance for access control. ITIRT and IRIS06 measured sensor performance and the effect that iris images collected by the different sensors had on matcher performance. However, as a technology evaluation, ICE 2006 measured the performance of algorithms from three groups on the same set of iris images. This allowed for a direct comparison among the tested algorithms.

The scenario evaluation protocol in ITIRT and IRIS06 called for iris images to be collected from three sensors and matching to be performed by the same algorithm. (Note that ITIRT and IRIS06 used different algorithms.) Thus ITIRT and IRIS06 measured sensor effects on matcher performance.

This paper discusses these evaluations, their similarities and differences, and most importantly summarizes performance across the evaluations. To compare performance across evaluations, performance statistics selected for this meta-analysis took into account evaluation type, failure to enroll and failure to acquire, sensor quality software, and subject variability. Based on the selection criteria, across all three evaluations, reported false non-match rate (FNMR) at a false match rate (FMR) of 0.001 ranged from 0.0122 to 0.03847. At an FMR of 0.001, the range of FNMR for the best performers in each test was 0.0122 to 0.0175.

II. BACKGROUND

A. ITIRT Study

IBG's ITIRT study was funded by the US Department of Homeland Security (DHS) and began in July 2004. Final results were released in May 2005 [3].

IBG tested match rates, enrollment and acquisition failure rates, interoperability, and level of effort needed for transactions using three sensors: Panasonic BM-ET300, Oki Irispass, and LG 3000. Results of these tests include "cross-visit recognition," which is a one-to-one comparison of an enrollment iris template from an initial visit against the iris template captured during a second visit. Template matching software from Iridian performed matching tests on the collected biometric samples.

Of the 1224 subjects recruited for the study, 458 subjects made two visits. Of those, nearly 65% made their second visit 11 days to 20 days after their first visit, and 95% visited a second time 6 days to 30 days after their first visit. Only one subject visited more than 45 days after their initial visit.

B. IRIS06 Study

Authenti-Corp's study was funded jointly by the US Department of Justice, National Institute of Justice (NIJ) and the US DHS Transportation Security Administration (DHS/TSA) and kicked off in December 2005. The draft final report was released in March 2007 [4].

The IRIS06 study was a standards-based evaluation which conformed to a new International Organization for Standardization (ISO) testing standard [7], [8], an ISO iris image data format standard [9], and an American National Standards Institute (ANSI) InterNational Committee for Information Technology Standards (INCITS) API standard [10]. Authenti-corp also tested match rates, enrollment and acquisition failure rates, and level of effort needed for transactions using three sensors, but it did not identify the tested sensors, simply referring to them as Products A, B, and C. IRIS06 utilized a matching algorithm provided by Professor Daugman of the University of Cambridge.

IRIS06 also included interoperability testing and an experiment of performance when eyes are looking "off-axis." These results are unique to IRIS06 and are not considered here for comparison.

A total of 295 subjects participated in the study. Of these, 264 made two visits. The minimum, mean, and maximum time between the two visits were 14 days, 38 days, and 55 days, respectively. About 75% of the subjects made their second visit 34 days to 45 days after their first visit.

Authenti-corp provided the authors with a breakdown of how many subjects, iris samples, genuine comparison scores, and imposter comparison scores were used in testing for the results used in this analysis, as provided in Table I. The report provides performance results from both visits combined in the form of false non-match rates with upper and lower 95% confidence intervals (CIs).

C. ICE 2006

The ICE 2006 study conducted by NIST was funded jointly by the US DHS's Science and Technology Department and TSA, the US Director of National Intelligence's Information Technology Innovation Center, the US Federal Bureau of Investigation, the NIJ, and the Technical Support Working Group (TSWG). The study began in December 2003. The final report was released in March 2007 [5].

Part of a larger multi-biometric data collection, the ICE 2006 reported error rates for the left and right irises separately for three different matchers using the same images from a single sensor for data collection, at a single operating point – false accept rate (FAR) = 0.001. The matching algorithms tested were supplied by Sagem-Iridian (SG-2), Iritech (Irtch-2), and Cambridge (Cam-2). A modified LG EOU 2200 was used to collect iris images from 240 subjects. The LG EOU 2200 was modified so that the automatic quality check for capturing iris images was overridden. This allowed for up to two out of three captured images not to meet the built-in quality checks. There was a manual quality control step to cull images on the far end of low quality, "for example the eye was not visible at all due to the subject having turned their head." [5]

The collection protocol included inviting test subjects to return on a weekly basis. One-to-one matching tests were performed on iris images separated by at least one academic semester and no more than 17 months. A total of 240 subjects made at least two visits for the ICE 2006 study.

ICE 2006 divided its test set into 30 random test sets and reported results via boxplots for each algorithm, reporting a maximum, third quartile, median, first quartile, and minimum false reject rate (FRR).

III. METHODS

In this section, we discuss the rationale for choosing the points of comparison from each test, which includes issues of how quality, enrollment and acquisition, glasses, and timing between visits were handled.

Table I summarizes the sensors; algorithms; and the total number of subjects, biometric samples (both genuine and impostor), genuine comparison scores, and impostor comparison scores used in each of the tests reported in this analysis (in Tables IV and V). Table II summarizes key properties of the evaluations. The following evaluation properties were originally introduced in taxonomy of biometric applications by Wayman [11]: cooperative versus non-cooperative users, public versus private users, overt versus covert capture, attended versus unattended applications, habituated versus non-habituated, standard versus non-standard environment, and open versus closed systems. We have used some of these categories for comparison of the three studies in Table II, and we included categories to describe whether there was user training, the minimum number of successful samples required to enroll a subject, how many samples were required to meet a quality score threshold, the recognition mode used in the offline testing, the median or mean and the max time between collection of the first and 2nd samples, and whether samples were collected with or without glasses. These new categories will be further discussed in this section.

All studies used volunteer test subjects who were informed of the testing taking place. Each study used attendees to operate equipment to capture subjects' biometric samples. All studies also collected iris images indoors in a fixed environment within each study (but not necessarily the same across the studies).

While there is not enough information to discuss time to match irises across all three studies, none of the studies indicated an imposed timing constraint for offline matching experiments.

It is important to note that the Iridian and University of Cambridge algorithms are all based on John Daugman's work. In other words, all of the matching algorithms here except one – Iritech's algorithm in the ICE 2006 evaluation – have the same genesis.

A. Types of Errors

This paper discusses four types of errors, some of which were defined differently across studies. They are defined for the purposes of this paper as follows:

- The false match rate (FMR) is the rate at which a matching algorithm incorrectly determines that an impostor's biometric sample matches an enrolled sample.

- The false non-match rate (FNMR) is the rate at which a matching algorithm incorrectly fails to determine that a genuine sample matches an enrolled sample.

- The failure to enroll rate (FTE) is the rate at which a biometric system fails to enroll a subject's biometric sample.

- The failure to acquire rate (FTA) is the rate at which a biometric system fails to capture a subject's biometric sample for the purpose of recognition of the subject.

Note that neither the FMR nor the FNMR include FTE or FTA in their definitions. FMR and FNMR are strictly statistics of the capabilities of a matching algorithm. Further, the ICE 2006 study reports false accept rates (FARs) and false reject rates (FRRs) in its study, where the terms FMR and FNMR as defined above are more apropos given that FTE and FTA are not taken into account. This is further discussed in the Quality section below.

B. Comparison at FMR = 0.001

Given the number of samples in the studies as well as the data presented in the main body of the ICE 2006 report, we chose to compare these studies by comparing the FNMRs at the point that FMR is 0.001.

ICE 2006 did not take into account FTEs or FTAs. Hence, the results reported as false reject rates (FRRs) and false accept rates (FARs) are treated here to be the same as FNMR and FMR, respectively.

Numerical results for the FNMR figures in the boxplots at (FMR of 0.001) for ICE 2006 were made available for this paper. Results for the left and right iris were averaged together for the

purpose of comparison with the other studies, which did not report results separately for the left and right irises.

ITIRT's results are listed by Hamming distance. FNMR results were collected for cross-visit recognition where the FMR was nearly 0.001. When it was unclear which FMR to choose, we chose the higher FMR value (hence, the lower FNMR).

IRIS06 results are presented graphically in the report, but for our analysis, Authenti-corp provided the exact results in Figure 1 and listed in Table IV.

C. Quality

All three studies' sensors required enrollment images and second visit images to meet some minimum quality criteria. However, the criteria for enrollment and acquisition were controlled by the devices and not the testing labs except in two-thirds of the ICE 2006 images. The products in IRIS06 and the LG 3000 cameras in the ITIRT experiments have built in quality checks, and an Iridian module was added to the OKI Irispass and Panasonic BM-ET300 cameras in the ITIRT study that served a similar function. The ITIRT report noted that the LG 3000 sensors appeared to have lower criteria for enrollment than for recognition.

ICE 2006 was able to modify its LG EOU 2200 camera to circumvent the quality check for two-thirds of the images. This means that, even with a manual quality control check for extremely low quality images, images in the ICE 2006 data set are expected to be of lower quality than the other two studies. Iris samples that might have been rejected in the other two studies (and subsequently yield a failure to enroll or acquire statistic) would have been accepted in the ICE 2006 study, provided at least one sample out of a group of three passed the automatic quality checks.

In IRIS06, Authenti-corp did some post-analysis of images that did not meet their own quality criteria ("flagged" images) and showed additional analysis based on removing these flagged images.

Finally, there was no baseline quality algorithm applied across all the studies, and no study produced a distribution of the quality of their images. So more in-depth quality analysis is not possible.

D. Enrollment and Acquisition Protocols

ITIRT required two successful captures to enroll using the OKI Irispass or Panasonic BM-ET300 sensors and four using the LG 3000 sensors. In both ICE 2006 and IRIS06, one successfully captured iris image produced an enrollment template.

After enrollment in the ITIRT study, each subject had to make three recognition transactions with each sensor. A transaction consisted of three left and three right irises – producing up to 54 samples, less any failures to acquire.

For the ICE 2006 study, each subject on a visit (referred to as a "session") yielded two "shots" of three images for the left and for the right eye, for a total of 12 images. An acceptable shot had one

or more images that passed the LG 2200 camera's built-in quality checks, and all three images were saved. If none of the three images passed the built-in quality checks, then none of the images were saved.

For the IRIS06 data collection used for offline testing, most subjects made two visits, and both visits followed a similar protocol. Subjects were allowed up to three attempts to enroll at least one iris on each product. After an eye was successfully enrolled, no additional enrollment attempts were made. Then for recognition, exactly three attempts were allowed per sensor, regardless of success or failure of the attempts. This step was repeated after a short break.

E. Glasses

Subjects were asked to remove glasses for all interactions with the sensor for the ICE 2006 data collection. Subjects were asked to remove their glasses only for enrollment for both the ITIRT and IRIS06 study. Removal of glasses was left up to the subject for subsequent acquisitions in the ITIRT evaluation. The IRIS06 study, which collected three samples per subject for verification, had subjects who wore glasses remove their glasses for their third sample. Among the various analyses included in the IRIS06 report, the analysis in the paper uses the performance figures where images with eyeglasses were excluded from the set, using only the third attempt for subjects who wear eyeglasses. However, this conflates results with the benefit of better matching with the third attempt.

F. Differences in Time between Sample Acquisition

Timing between the first and second visit varied between studies, with IRIS06 having the shortest median time between visits while the IRIS06 average length between visits was twice that of ITIRT's. ICE 2006 had the longest minimum and maximum time between visits. In fact, the minimum time between visits in ICE 2006 is greater than mean or median of either of the other two studies. Both ITIRT and IRIS06 have a maximum time between visits being just over 50 days.

There is not enough information to discuss time to match irises across all three studies, but none of the studies indicated an imposed timing constraint for offline matching experiments.

IV. RESULTS

The results of this paper are presented graphically and numerically by comparing the observed FNMRs.

Fig. 1 shows the results of the three tests in order of when the studies began: oldest on the left (ICE 2006, ITIRT in the middle, and most recent on the right (IRIS06).

ICE 2006 results were presented in box plots. The horizontal line in the middle of the box is the median. The top and bottom of the box correspond to the 1st quartile (25th percentile) and 3rd quartile (75th percentile) values of the observations, respectively. The dashed lines above and

below the box, called "whiskers," end with a short horizontal line, which mark minimum and maximum data values. ITIRT results are single values. IRIS06 results give performance values with a range of estimated uncertainty in the form of 95% confidence intervals, computed using the logit beta-binomial method for the results used here in Table IV.

Collectively, these data points range from 0.00473 to 0.0465. The observed FNMR values range from 0.0122 (ICE 2006, SI-2) to 0.03847 (ITIRT, LG 3000).

Fig. 1 suggests two conclusions about the range of results observed over three evaluations. First, the difference among the best performers in each evaluation is an order of magnitude less than the observed performance. The best performer had an FNMR of approximately 0.01 at an FMR of 0.001, and the differences among the three best FNMRs were on the order of 0.001. Second, the range of FNMRs for all performers and the differences among all performance statistics was of the same order of magnitude.

These two suggested conclusions are examined using mean absolute difference. The absolute difference of x and y is $|x\text{-}y|$. The mean absolute difference between two sets of numbers x_i and y_j is $1/NM \sum_{ij} |(x_i - y_j)|$, where N and M are the length of x_i and y_j. Mean absolute difference provides a robust measure of the difference between two sets of numbers.

Table III provides a quantitative summary of the differences in performances plotted in Fig. 1. The first three rows in Table III summarize the performance difference between two evaluations; e.g., row one compares ICE 2006 and ITIRT. The first three columns of Table III give the absolute difference between each test-pair's lowest, median, and highest FNMRs, respectively; e.g., the first column reports the absolute differences between the best performer for each evaluation. The fourth column of Table III shows the mean absolute differences between all results for two evaluations. The last row of Table III reports the average statistic for each of the columns.

The mean of the absolute differences between the best performers (column labeled $\lceil \Delta \rceil$ of Lowest FNMRs) was 0.004. Since the mean performance of the best results from each evaluation is 0.01, this shows that the difference among best performers is an order of magnitude smaller than the observed performance value.

Over all three evaluations, the range of the observed FNMR statistics reported was 0.0122 to 0.03847. Since the the mean absolute differences across the three evaluations was 0.01, the observed performance statistics and the differences in performance are of the same order of magnitude.

V. CONCLUSION

One of the hallmarks of science is the repeatability of experiments. Here we have compared experiments from three independent sets of third-party testers. They independently designed tests, collected data from different populations, and conducted experiments on iris recognition systems. Despite the differences in collection efforts, sensors, matching algorithms, protocols, and other factors, these three tests produced consistent results and demonstrate repeatability.

The major differences between the evaluations for the results compared here include the level of the quality of images (including how eyeglasses were handled), the degree of habituation, time

between initial and subsequent visits, and the data subjects themselves. Yet, these variations are not reflected in the end results.

Because of the strong agreement among these tests, we conclude that these evaluations represent an accurate assessment of the state of the art in iris recognition as of Spring 2006. (Spring 2006 marks the last algorithm submission within this set of evaluations.)

As shown in Fig. 1, all of the observed FNMR values (at an FMR of about 0.001) fall roughly between 0.01 and 0.04. The two most similar configurations are also the bounds of the data: the ICE 2006 test of the LG EOU 2200 sensor and Sagem-Iridian-2 algorithm scoring a 0.0122 FNMR and the ITIRT test of the LG 3000 sensor and the Iridian KnoWho algorithm (v. 3.0).

There are two possible reasons why the results of these evaluations are so similar. First, all but one of the tested algorithms was based on the work of Professor John Daugman. Second, because there is a symbiotic relationship between the development of sensors and iris recognition algorithms, the dominance of the Daugman-based algorithms in the market place may also decrease variation in the output of different sensors.

The FNMRs examined here are all the same order of magnitude, and as observed in Fig. 1, there is a fair amount of overlap of the boxplots of ICE 2006 and the CI's of IRIS06. The mean differences show that the difference between each tests' data points is no greater than the order of magnitude of the FNMRs. Further, the differences between the best performers (i.e., the lowest FNMR scores) for all test-pairs is an order of magnitude less than the error rates.

APPENDIX

Table IV contains the data used to produce Fig. 1, as well as the location of the data in their respective reports. Table V shows data similar to Table IV but contains FNMRs when FMR = 0.0001.

ACKNOWLEDGMENT

REFERENCES

[1] A.K. Jain, R. Bolle, and S. Pankanti, "Introduction to biometrics," in *Biometrics: Personal Identification in Networked Society*, A.K. Jain, R. Bolle, and S. Pankanti, eds., Boston: Kluwer Academic, 1999, pp. 1-41.
[2] M. Lockie, "Comment," *Biometric Technology Today*, p. 12, November/Decmeber 2004.

[3] International Biometric Group. (2005, May). Independent testing of iris recognition technology (ITIRT) - Final Report. [Online].
Available: http://www.ibgweb.com/reports/public/ITIRT.html.

[4] Authenti-Corp. (2007, Mar. 31). Draft Final Report: Iris Recognition Study 2006 (IRIS06). version 0.40. [Online].
 Available:
 http://www.authenti-corp.com/iris06/report/IRIS06_draft_report_v0-40p_20070331.doc

[5] P. J. Phillips, *et. al.*(2007, Mar.) FRVT 2006 and ICE 2006 Large-Scale Results. National Institute of Standards and Technology. Gaithersburg, MD. Technical Report NISTIR 7408. [Online]. Available: http://iris.nist.gov/ice/FRVT2006andICE2006LargeScaleReport.pdf

[6] P. J. Phillips, A. Martin, C. L. Wilson, and M. Przybocki, "An introduction to evaluating biometric systems," *Computer*, vol. 33, pp. 56–63, 2000.

[7] *Information Technology – Biometric performance testing and reporting – Part 1: Principles and framework*, ISO/IEC 19795-1:2006.

[8] *Information Technology – Biometric performance testing and reporting – Part 2: Testing methodologies for technology and scenario evaluation*, ISO/IEC 19795-2:2007.

[9] *Information technology - Biometric data interchange formats - Part 6: Iris image data*, ISO/IEC 19794-6:2005.

[10]*Information technology - Biometrics Application Programming Interface (BioAPI) Specification*, ANSI/INCITS 358-2002, (Version 1.1).

[11]J. L. Wayman, *National biometric test center collected works,* National Biometric Test Center, San Jose, CA, 2000.

TABLE I
SUMMARY OF TEST SENSORS, ALGORITHMS, AND NUMBER OF BIOMETRIC SAMPLES USED IN EACH EVALUATION.

Evaluation	Sensor	Matching Algorithm	Totals for Tests in Fig. 1 (for left and right eyes): Subjects | Samples | Genuine Comparison Scores | Impostor Comparison Scores
ICE 2006	LG EOU 2200	Sagem-Iridian (SG-2)	240 | 59,558 | 3,085,351 | 562,301,273
		Iritech (Irtch-2),	
		Cambridge (Cam-2)	
ITIRT	Panasonic BM-ET300 (Pan)	Iridian's KnoWho OEM SDK v3.0	458 | 12,238 | 13,731 | 33,865,260
	OKI IRISPASS-WG (OKI)		458 | 12,587 | 15,047 | 36,597,230
	LG IrisAccess 3000 EOU & ROU (LG)		458 | 16,826 | 14,868 | 36,150,847
IRIS06	Product A	Daugman algorithm	285 | 4397 | 3845 | 1,070,834
	Product B		285 | 4467 | 3937 | 1,078,878
	Product C		284 | 4696 | 4214 | 1,095,724

TABLE II
COMPARISON OF KEY PROPERTIES OF THE THREE EVALUATIONS.

Properties	ICE 2006	ITIRT	IRIS06
Users	Cooperative, Compensated volunteer test subjects	Cooperative, Compensated volunteer test subjects	Cooperative, Compensated volunteer test subjects
Capture Mode	Overt	Overt	Overt
Operator Mode	Attended	Attended	Attended
User Training	No	Yes	No
Habituation	2-42 visits per subject	2 visits per subject	2 visits per subject
Environment	Indoors; Fixed	Indoors; Fixed	Indoors; Fixed
Min Number of Successful Sample(s) for Enrollment	1	2 for OKI & Pan, 4 for LG	1
Minimum Percentage Required to Meet Sensor Quality	33%	100%	100%
Recognition Mode	1:1 verification	1:1 verification	1:1 verification
Min, Median/Mean and Max Time b/t Collected Samples	40 days; N/A; 510 days	1 day; (16 to 20) days; 51 days	0 days; 0 days; 55 days
Glasses	Off	Off during enrollment; Optional for recognition	Off during enrollment; Off for recognition, here

TABLE III
ABSOLUTE AND MEAN DIFFERENCES BETWEEN PAIRS OF EVALUATIONS IN THIS PAPER.

Test-pair	|Δ| of Lowest FNMRs	|Δ| of Median FNMRs	|Δ| of Highest FNMRs	Absolute Mean Difference
ICE 2006 – ITIRT	0.0018	0.0104	0.0179	0.01
ICE 2006 – IRIS06	0.005	0.0117	0.0139	0.01
ITIRT – IRIS06	0.004	0.0014	0.0040	0.01
Mean of Differences	0.004	0.008	0.01	0.01

FIG. 1
FNMR RESULTS FROM THREE STUDIES (AT FMR = 0.001). ICE 2006 REPORTS RESULTS FOR ALGORITHMS ON BOXPLOTS; ITIRT REPORTS A SINGLE PERFORMANCE STATISTIC FOR EACH SENSOR; AND IRIS 06 REPORTS ESTIMATED FNMR WITH A 95% CONFIDENCE INTERVALS.

13
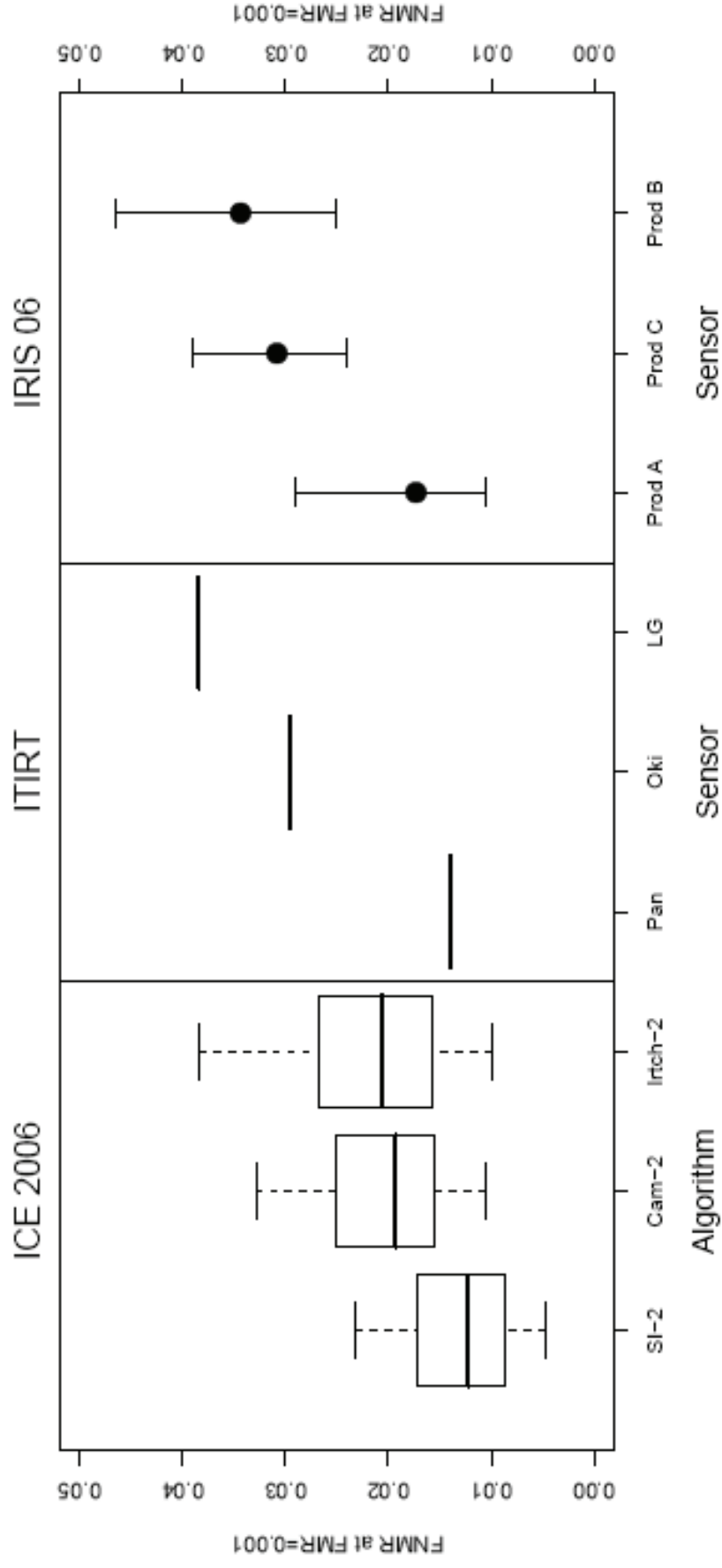
TABLE IV
REPORTED FNMR TEST RESULTS AT FMR = 0.001[1]

Test-Algorithm ID	Min FNMR	Lower 95% CI FNMR	1st Quartile FNMR	FNMR[2]	3rd Quartile FNMR	Upper 95% CI FNMR	Max FNMR	Location of Data
ICE 2006-SI-2 [5][3]	0.00473		0.00869	0.0122	0.0170		0.0231	p.11, Fig. 3
ICE 2006-Cam-2 [5][3]	0.0106		0.0155	0.0193	0.0248		0.0329	p. 11, Fig. 3
ICE 2006-Irtch-2 [5][3]	0.00993		0.0158	0.0206	0.0266		0.0384	p. 11, Fig. 3
ITIRT-Pan [3] @ FMR=0.0009959				0.01398				p. 92, Fig. 66
ITIRT-OKI [3] @ FMR=0.0010989				0.02964				p. 92, Fig. 66
ITIRT-LG [3] @ FMR=0.0010304				0.03847[4]				p. 92, Fig. 66
IRIS06-A [4]		0.0105	0.0175		0.0290			p. 125, Fig. 6-30, "Without Eyeglasses"
IRIS06-C [4]		0.0240	0.0310		0.0390			p. 125, Fig. 6-30, "Without Eyeglasses"
IRIS06-B [4]		0.0250	0.0345		0.0465			p. 125, Fig. 6-30, "Without Eyeglasses"

FNMR = False non-match rate, although the use of "false reject rate" (FRR) in ICE 2006 is similar to FNMR. FNMR results at FMR (or FMR) of 0.001 unless otherwise noted.
CI = Confidence Interval

TABLE V
REPORTED FNMR TEST RESULTS AT FMR = 0.0001[1]

Test-Algorithm ID	Min FNMR	Lower 95% CI FNMR	1st Quartile FNMR	FNMR[2]	3rd Quartile FNMR	Upper 95% CI FNMR	Max FNMR	Location of Data
ICE 2006-SI-2 [5][3]	0.00670		0.0110	0.0141	0.0222		0.0306	pp.54-55, Fig. 28-29
ICE 2006-Cam-2 [5][3]	0.0148		0.0212	0.0246	0.0350		0.0458	pp.54-55, Fig. 28-29
ICE 2006-Irtch-2 [5][3]	0.0136		0.0228	0.0267	0.0363		0.0586	pp.54-55, Fig. 28-29
ITIRT-Pan [3] @ FMR=0.0000996				0.01966				p. 92, Fig. 66
ITIRT-OKI [3] @ FMR=0.0001022				0.04154				p. 92, Fig. 66
ITIRT-LG [3] @ FMR=0.0001003				0.04600[4]				p. 92, Fig. 66
IRIS06-A [4]		0.0130	0.0200		0.0320			p. 125, Fig. 6-30, "Without Eyeglasses"
IRIS06-C [4]		0.0315	0.0390		0.0470			p. 125, Fig. 6-30, "Without Eyeglasses"
IRIS06-B [4]		0.0300	0.0400		0.0525			p. 125, Fig. 6-30, "Without Eyeglasses"

FNMR = False non-match rate, although the use of "false reject rate" (FRR) in ICE 2006 is similar to FNMR. FNMR results at FMR (or FMR) of 0.0001 unless otherwise noted.
CI = Confidence Interval

[1] The results are shown here as presented in their respective reports or to the authors directly. Any errors in computing significant figures or scoring are not considered in Tables IV or V.
[2] ICE 2006 reports a median score.
[3] ICE reports left and right iris results separately. The values presented here represent the combined averaged results.
[4] Note that this figure is from the main body of the report and not from Annex A, which uses a different imposter number than what is the main body of the report and follows the suggested method of analyzing LG's output (Option 4 as described in Section 5.4 of the ITIRT report).