
IREX I

Performance of Iris Recognition Algorithms on Standard Images

NIST Interagency Report 7629

P. Grother, E. Tabassi, G. W. Quinn, W. Salamon

Information Access Division
National Institute of Standards and Technology



September 22, 2009

ACKNOWLEDGMENTS

The authors would like to thank the Department of Homeland Security's U. S. VISIT Office for supporting this work. In addition, the authors are grateful to the Department of Homeland Security's Science and Technology Directorate for supporting the ongoing standardization aspects of this activity. We extend our thanks to the U.S. Department of Defense and the University of Bath for providing iris imagery to the project.

Finally, thanks go IREX participants for their contributions toward the ISO/IEC 19794-6 standard and the development of the IREX test specification¹.

DISCLAIMER

Specific hardware and software products identified in this report were used in order to perform the evaluations described in this document. In no case does identification of any commercial product, trade name, or vendor, imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the products and equipment identified are necessarily the best available for the purpose.

¹The formal CONOPS and API specification is available at http://iris.nist.gov/irex/irex_api.pdf.

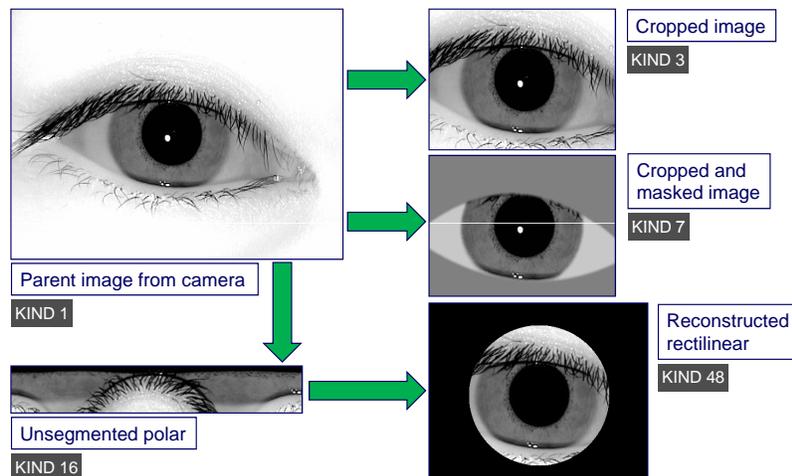
EXECUTIVE SUMMARY

The Iris Exchange (IREX) was initiated by NIST in late 2007 to support interoperable exchange of iris imagery in high performance biometric applications. The first activity in the program, the IREX I evaluation, was conducted in cooperation with the iris recognition industry to develop and test standard image formats, and to demonstrate that iris recognition algorithms can maintain their accuracy and interoperability with compact images. Standard formats are needed in federated applications in which iris data is exchanged between interoperating systems. Compact size is a current and vital requirement for applications in which imagery is passed across bandwidth-limited networks, or stored on identity credentials.

IREX I was initiated to give quantitative support to the revision of the ISO/IEC 19794-6 and ANSI/NIST TYPE 17 standards, and to form a multi-provider marketplace around those standards. As the largest independently administered test of iris recognition technology to date, IREX I includes a formal evaluation of the state-of-the-art of iris recognition algorithms from the following providers:

- | | | | | |
|-----------------------|----------------|-------------------------|----------------|---------|
| Cambridge University | Cogent Systems | Crossmatch Technologies | Honeywell | Iritech |
| L1 Identity Solutions | LG | Neurotechnology | Retica Systems | Sagem |

Recognition algorithms from these organizations were evaluated in a three stage process. First, algorithms were applied to convert raw images from contemporary iris cameras to the standardized iris images (i.e. IREX records) depicted here:



Preparation of these records requires various detection, localization, cropping, sampling and masking operations. These operations are non-trivial. They precede the second stage of processing in which features are extracted from standard images to form a template. The IREX records are not iris templates; instead they are specialized interoperable images designed for efficient storage. Templates contain proprietary “black box” feature representations. Their content is non-standard, non-interoperable and not suitable for cross-agency exchange of iris data. The last stage, recognition, involves matching of templates to produce comparison scores. The role of standardized images is depicted as follows.



A = SAGEM	B = COGENT	C = CROSSMATCH	D = CAMBRIDGE	E = L1	x1 = PRIMARY
F = RETICA	G = LG	H = HONEYWELL	I = IRITECH	J = NEUROTECHNOLOGY	x2 = SECONDARY
KIND 1 = RAW 640x480		KIND 3 = CROP		KIND 7 = CROP+MASK	
KIND 16 = CONCENTRIC POLAR					

The primary impacts of IREX are listed below. These are followed, in the next section, by an extensive, technical, summary of the results.

- ▷ IREX has advanced iris recognition toward the level of technical maturity and interoperability of fingerprint biometrics and has affirmed the potential for using iris biometrics as a second modality for large-scale identity management applications. This result will support storage of iris biometrics on identity credentials such as the United States Government's PIV cards in support of Homeland Security Presidential Directive (HSPD) 12. In addition it will directly support the interoperability goals of Homeland Security Presidential Directives 6, 11 and 24.
- ▷ IREX quantified the core algorithmic capability of nineteen recent iris recognition software implementations from ten organizations. This represents an order of magnitude expansion in the number of providers over the last half decade.
- ▷ IREX required participating organizations to implement the image formats proposed for the ISO standard. The result is that each provider now has off-the-shelf, or readily portable, software to support creation, validation, and recognition of standard images.
- ▷ IREX I complements considerable activity in the area of iris camera development. This has occurred particularly in the stand-off capture (where iris images are acquired at a few meters) and mobile device arenas. These, and other, cameras are technically capable of producing images in conformance to formal standards. IREX recommends that users should require cameras to do so.
- ▷ Standard iris image records with size of approximately thirty kilobytes can be produced for large-scale identification applications. This represents a factor of ten reduction in size over the images captured using contemporary cameras.
- ▷ Standard iris image records with sizes around three kilobytes can be produced that are suitable for one-to-one authentication applications. This factor of one hundred reduction in size over the images captured using contemporary cameras makes the images suitable for storage on "smart card" credentials.
- ▷ There is an industry-wide accuracy versus speed tradespace: Large-scale identity management applications benefit from fast algorithms; Forensic and unconstrained-capture applications should leverage newer, more computationally expensive, iris recognition algorithms.

The authors are available to discuss and brief this report.

A = SAGEM	B = COGENT	C = CROSSMATCH	D = CAMBRIDGE	E = L1	x1 = PRIMARY
F = RETICA	G = LG	H = HONEYWELL	I = IRITECH	J = NEUROTECHNOLOGY	x2 = SECONDARY
KIND 1 = RAW 640x480	KIND 3 = CROP	KIND 7 = CROP+MASK	KIND 16 = CONCENTRIC POLAR		

TECHNICAL SUMMARY

The significant results of the test are listed by subject-matter area below. These should be weighed in light of the caveats presented on page 13. Vendor comments on their IREX algorithms are included in the accompanying appendices². These are accompanied, separately, by free vendor comments on IREX itself.

IMPACT ON THE MARKETPLACE

- ▷ In parallel with the revision of the ISO/IEC 19794-6 iris image interchange standard, the IREX study attracted ten organizations into implementing the standard. This entailed significant effort on the part of the providers with respect to the production of syntactically correct code and with respect to algorithmic functionality. The result is that each provider now has off-the-shelf, or readily portable, software to support creation, validation, and recognition of standard images.

Sec. 4
- ▷ IREX quantified the core algorithmic capability of nineteen recent³ iris recognition software implementations from ten organizations. This represents an order of magnitude expansion in the number of providers over the last half decade. There are at least two other commercial providers whose algorithms were not openly submitted to IREX. The availability of standards-compliant implementations from these providers is not known.

Sec. 4
- ▷ The compact interoperable formats tested here are amenable to lossless compression to as little as 20 kilobytes. Lossless compression preserves imaging detail and ensures that iris recognition accuracy for large scale one-to-many applications is not compromised.

Sec. 8.7
- ▷ Used with lossy compression, compact interoperable images occupy as little as two kilobytes. This makes them suitable for storage on ISO/IEC 7816 integrated circuit “smart card” identification tokens. In comparison to the other biometric interchange records currently used on such credentials, the IREX record is somewhat larger than standard fingerprint minutia templates (approx. 300 bytes⁴), but smaller than standard fingerprint images (typically 6 to 10 kilobytes⁵) and e-Passport face images (from about 15 to 20 kilobytes⁶).

Sec. 8
- ▷ Compressed iris image sizes are similar to the template sizes measured for many algorithms submitted to the IREX evaluation. Such images can be matched without loss of recognition accuracy. There are no standards for iris recognition templates, and they are not interoperable. Their size advantage is small. Standard iris images should be exchanged instead.

Sec. 7.5
- ▷ The ISO/IEC 19794-6 standard is application-neutral. Standard images are suitable for iris recognition applications embedding large-scale one-to-many identification searches (watchlist, de-duplication, fraud detection), one-to-many token-less verification claims, and one-to-one verification claims with a credential.
- ▷ By executing a wide ranging study and reporting detailed results, IREX is expected to influence technical development of iris recognition algorithms in unforeseen ways.

INTEROPERABLE IMAGE FORMATS

²The appendices include: optional provider-supplied text describing each specific recognition algorithm; comments from the participants on the IREX activity; and detailed technical information for each specific IREX implementation. The appendices may be downloaded from http://iris.nist.gov/irex/irex_appendices.pdf.

³The implementations were sent to NIST in mid February 2009.

⁴A INCITS 378 minutia template from a flat index-finger impression containing 38 minutiae encoded in 6 byte format.

⁵A typical ISO/IEC 19794-4 single finger record, acquired at 500 pixels per inch and WSQ compressed at 15:1.

⁶A color ISO/IEC 19794-5 token face image, interocular distance 90 pixels, 15:1 compression and chrominance sub-sampling.

A = SAGEM	B = COGENT	C = CROSSMATCH	D = CAMBRIDGE	E = L1	x1 = PRIMARY
F = RETICA	G = LG	H = HONEYWELL	I = IRITECH	J = NEUROTECHNOLOGY	x2 = SECONDARY
KIND 1 = RAW 640x480	KIND 3 = CROP	KIND 7 = CROP+MASK	KIND 16 = CONCENTRIC POLAR		

- ▷ For applications without size, transport, or communications-related throughput constraints the uncropped uncompressed KIND 1⁷ rectilinear record may be used. Such images, of size 640 x 480 pixels, can be losslessly compressed using the ISO/IEC 15948 Portable Network Graphics compression algorithm (PNG) by about a factor of two. The resulting records occupy a median size of 150 kilobytes.

Sec. 8.7
- ▷ If the acquisition process can include a coarse iris detection operation, then the centered crop-only KIND 3 record almost always gives fewer false rejection errors than its un-cropped, uncompressed, and unconstrained KIND 1 parent. With lossless compression, KIND 3 instances require 50-80 kilobytes of storage. Instances may be further compressed using the lossy JPEG2000 algorithm. The crop-only KIND 3 image format should be retained in the ISO/IEC 19794-6 standard.

Sec. 8.2
- ▷ The cropped-and-masked KIND 7 image format proposed for the ISO/IEC 19794-6 standard should be retained and advanced as the primary format for the exchange of compact iris images smaller than 3KB. At larger sizes or lower compression ratios, the KIND 3 format should be preferred: it is more easily and safely instantiated. The false rejection performance for some implementations exceeds that of the KIND 1 parent. The cropped-and-masked KIND 7 format is particularly amenable to lossless compression. This allows iris records to be produced in the 20-40 kilobyte range. This format should usually only be used in conjunction with the JPEG2000 and PNG compression algorithms.

Sec. 8.2
- ▷ The KIND 16 unsegmented polar format proposed for the ISO/IEC 19794-6 standard should be rejected. The recognition error rates associated with the format are much larger than those attainable with rectilinear KIND 3 and KIND 7 records. This is true natively, when a single provider prepares and matches the records, and in the interoperable case also, when different providers do so.

Sec. 8.2
- ▷ For some images, lossless compression will not be able to achieve a specific target size, and JPEG2000 should be applied at a specific targeted bit rate. For images below about 20KB, lossy JPEG2000 compression will usually be needed.

Sec. 8
- ▷ Using false rejection error as a metric, the cropping operation used in preparation of KIND 3 and KIND 7 records should extend to no closer than 0.6 iris radii from the iris in the horizontal direction, and 0.2 radii in the vertical direction.

Sec. 8.5
- ▷ Iris cameras should not internally apply compression to iris images, unless they are manufactured for a dedicated, profiled, application in which standardized compressed iris images are produced. An exception is lossless compression.
- ▷ There is a large academic literature addressing the iris localization problem, and a rich diversity of algorithms can be employed to effect the detection and localization steps necessary to instantiate KIND 7 and KIND 3 records. Detection and localization are non-trivial operations and are influential on recognition accuracy. Users should evaluate implementations accordingly.

Sec. 1

RECOGNITION ACCURACY

- ▷ False rejection performance depends on the following (in decreasing order of importance): the recognition algorithm, the particular image dataset, the standard image format, and on the amount of compression applied. The observed error rate variations span at least an order of magnitude.

Sec. 7.3.1

⁷For a visual description of the image KINDS evaluated in IREX, refer to Figure 3 on page 18.

A = SAGEM	B = COGENT	C = CROSSMATCH	D = CAMBRIDGE	E = L1	x1 = PRIMARY
F = RETICA	G = LG	H = HONEYWELL	I = IRITECH	J = NEUROTECHNOLOGY	x2 = SECONDARY
KIND 1 = RAW 640x480	KIND 3 = CROP	KIND 7 = CROP+MASK	KIND 16 = CONCENTRIC POLAR		

- ▷ False match rate (FMR) calibration curves are computed. These confirm the low false match rates published for iris recognition algorithms at specific operational thresholds. By comparing images from two occupationally and geographically separated populations, of combined size $O(10^4)$, the FMR calibration is based on $O(10^9)$ comparisons. While this population size extends prior independent studies, larger operational corpora should be leveraged to refine FMR performance estimates and give statistical significance to FMR measurements on national-scale identification searches.

Sec. 7.4
- ▷ The stability of the impostor distribution is measured and reported. At a specific threshold, the false match rate depends on the dataset. This applies across all formats and compression conditions, including uncompressed and unprocessed iris images. In addition, for most iris recognition algorithms, false match rates vary under compression. When severe compression is applied to damage the iris texture, some algorithms maintain low FMR; others do not.

Sec. 7.4

INTEROPERABLE ACCURACY

- ▷ Given a standard for iris images, there are two separate tasks for iris recognition: Generation of the standard image, and matching of standard images (matching is implemented via by proprietary templates). These tasks will generally be executed by different providers' algorithms. The implementations that most accurately match standard IREX records are not generally the implementations that prepare the most matchable IREX records. That is, the error rates associated with the initial detection of the sclera-iris and iris-pupil boundaries are distinct from, but smaller than, the error rates associated with the end-stage fine-grained localization, feature extraction and matching.

Sec. 8.6
- ▷ The interoperability of standardized iris images in IREX is better than that reported for standard fingerprint minutiae templates: There, the best accuracy was observed when the same provider generated and matched the enrollment and verification templates. In IREX, this native-mode bias is small. Instead the iris recognition algorithm (fine-grained localization, feature extraction, and matching) is most influential on outcome. Fingerprint minutia interoperability is degraded by idiosyncratic (i.e. algorithm-specific) minutia detection and selection[8, 25, 23]. For iris, the standard interchange medium is image data; for fingerprint minutiae, it is (x, y, θ) point data and the relative interoperability of the two is a product of the difficulty of consistently and uniformly instantiating the semantic content of the respective standards.

Sec. 8.6

COMPUTATIONAL EFFICIENCY

- ▷ The IREX test plan explicitly stated that algorithm timing estimates would be reported. Further, IREX encouraged the submission of slow-but-accurate vs. fast-but-less-accurate implementations within a generous timing budget. The result, across the nineteen IREX algorithms, is that there are two orders of magnitude for the time needed to prepare an IREX record and to generate a template from an IREX record, and *three* orders of magnitude for the time needed to execute a one-to-one match.

Sec. 7.6
- ▷ The more computationally expensive algorithms give fewer recognition errors. This accuracy benefit applies on both the segmentation side, and the matching side. While the results demonstrate the existence of an industry-wide accuracy vs. time tradespace, not all providers demonstrate such a tradeoff between their primary and secondary SDKs.

Sec. 7.6.3

A = SAGEM	B = COGENT	C = CROSSMATCH	D = CAMBRIDGE	E = L1	$x1$ = PRIMARY
F = RETICA	G = LG	H = HONEYWELL	I = IRITECH	J = NEUROTECHNOLOGY	$x2$ = SECONDARY
KIND 1 = RAW 640x480	KIND 3 = CROP	KIND 7 = CROP+MASK	KIND 16 = CONCENTRIC POLAR		

Role	Recommended		Target Record Size								
	Format	Compressor	2KB	4KB	8KB	16KB	32KB	64KB	128KB	256KB	307KB
All	KIND 1	Uncompressed									█
All	KIND 3	Uncompressed							█	█	
All	KIND 7	Uncompressed							█	█	
All	KIND 3	PNG Lossless						█			
All	KIND 7	PNG Lossless				█	█				
1:N	KIND 3	JPEG 2000 Lossy			█	█					
1:N	KIND 7	JPEG 2000 Lossy			█	█					
1:1	KIND 3	JPEG 2000 Lossy	█	█	█	█					
1:1	KIND 7	JPEG 2000 Lossy	█								

Figure 1: Application-specific recommendations on compression and format. The horizontal axis shows target file size in kilobytes on a logarithmic scale.

- Sec. 7.6.4
Sec. 7.6
Sec. 7.6.2
Sec. 7.6.1
Sec. 7.6.1
Sec. 7.6.4

SELECTION OF COMPRESSION ALGORITHMS

- Sec. 8.2

- ▷ The IREX study has confirmed the findings of previous studies, namely that compression gives low, graduated, increases in false rejection errors. Prior work, published in the academic literature, has often considered single non-commercial algorithms running on smaller datasets. Sec. 8.1.1
- ▷ Lossy compression algorithm such as JPEG and JPEG2000 unrecoverably damage iris images. This has an adverse effect on false non-match error rates, and, for some iris recognition algorithms, on false match rates too. This latter aspect contraindicates application of lossy compression to images used in one-to-many searches. Sec. 8.1.3
- ▷ Compression should be applied at the minimum level needed to attain a storage or bandwidth requirement. The primary operational target variable is the size of the compressed image. This may be set directly using the JPEG2000 algorithm. Default guidance is given in Figure 1. However, implementers should quantify compression damage in terms of bits per pixel, and this will depend on iris radius - large irises should be compressed more lightly. Sec. 8.6
- ▷ The ISO/IEC 10918 JPEG compression algorithm should be deprecated. The presence of Discrete Cosine Transform blocking artifacts produces elevated false match rates. This recommendation applies particularly to compact iris images, but also to the cases where JPEG encoding is being used solely as a convenient container. It is recommended that the lossless PNG compressor be used in such cases because it is too easy to invoke JPEG with adverse parameters. Sec. 8.1

EFFECTS OF DILATION, OCCLUSION, CENTER DISPLACEMENT

- ▷ Higher amounts of pupillary dilation increase image-specific false non-match and false-match rates. The effect diminishes when the enrollment images are JPEG2000 compressed. Sec. 8.9.2
- ▷ Images stored in KIND 16 format with constricted pupils produce higher image-specific false match rates. Perhaps high amounts of constriction make the features more difficult to localize. Sec. 8.9.2
- ▷ The difference between the dilations present in two iris images affects their comparison score. In particular, large disparities in dilation elevate false rejection error rates. False match rates are not changed. The magnitude of the effect is comparable to that of eyelid occlusion. Sec. 8.9.2
- ▷ The effect of dilation change on recognition accuracy diminishes when the enrollment images are JPEG2000 compressed. Sec. 8.9.2
- ▷ Changes in dilation tend to be smaller for intra-person comparisons than for inter-person comparisons. Under controlled illumination conditions, the amount of dilation may serve as an ancillary discriminating factor. Sec. 8.9.2
- ▷ Large amounts of eyelid occlusion increased the probability of a non-match for most algorithms, and increased the probability of a false match for some algorithms. Sec. 8.9.1
- ▷ Some IREX algorithms exhibit a small adverse dependence on the displacement of the pupil and limbus centers. Specifically, genuine matching scores are elevated. Sec. 8.9.3

IMAGE QUALITY ASSESSMENT

A = SAGEM	B = COGENT	C = CROSSMATCH	D = CAMBRIDGE	E = L1	x1 = PRIMARY
F = RETICA	G = LG	H = HONEYWELL	I = IRITECH	J = NEUROTECHNOLOGY	x2 = SECONDARY
KIND 1 = RAW 640x480		KIND 3 = CROP		KIND 7 = CROP+MASK	KIND 16 = CONCENTRIC POLAR

- ▷ Three SDKs reported iris image quality scores on the standard range of [0-100]. A quality score, computed during the production of a KIND 3 record, is considered effective if they are quantitatively indicative of matching performance. Quality scores are considered interoperable if they are effective and uniformly interpretable. Two of three quality assessment algorithms, D1 and G1, are effective at assigning higher quality values to images with lower image-specific false match and false non-match rates. SDK D1 exhibits a larger stratification of the median quality scores assigned to four accuracy-categorized partitions of the ICE dataset. Sec. 8.8.1
- ▷ When images are excluded from testing on the basis of poor image quality assessments by SDKs G1 and D1, the false non-match rate of almost all matching algorithms improves. The best gain in accuracy is achieved for SDKs A1, I1, I2, G2, H2, H1. Exclusion of images judged to have low quality by SDK A1 does not appreciably improve observed false non-match rates. Sec. 8.8.2
- ▷ Quality scores are not interoperable: The distributions of the D1 and G1 quality scores are different and relate to different image error rates. This lack of interoperability implies a need for calibration. Sec. 8.8.2
- ▷ The IREX test plan did not require image quality assessment. While there is an increasing consensus that image quality estimates have greatest utility if they they are indicative of recognition accuracy, the IREX test plan did not mandate a semantic meaning for image quality. The new ISO/IEC 29794-6 iris image quality standard is intended to improve this situation. Academic and commercial research, and IREX activities will support this project. Sec. 8.8.2

EXPANDED TEST METHODS AND METRICS

- ▷ Detection error tradeoff characteristics are included that present lines joining points of fixed threshold. These reveal changes in *both* FNMR and FMR at a fixed threshold. These expose dependencies of FMR on dataset and compression parameters. Sec. 7.3
- ▷ Within-dataset and cross-dataset impostor comparisons were conducted. The latter guarantee the integrity of the impostor status of all image pairs. Matching of samples from different cameras is an implied aspect of applications based on standardized images. Sec. 7.4
- ▷ Conditional false non-match rates were defined and used to quantify errors induced by a change to a dataset (e.g., compression). This offers a more precise approach to failure analysis when the covariate is under the control of the experimenter. Sec. 8.2
- ▷ Template sizes were reported. Sec. 7.5
- ▷ Processing times were reported for IREX record preparation, template generation and one-to-one comparison. The computational cost vs. accuracy tradespace was documented. Sec. 7.6
- ▷ Image-specific false match and false non-match error rates were defined by inheriting concepts from the biometric zoo. These metrics support failure mode analyses by allowing association of a covariate (e.g., dilation) with a matching error rate without having to consider the covariate of a comparison image. Image-specific error rates are useful in detection of ground truth errors in datasets. Sec. 6.4
- ▷ Image-specific error rates were used to create four partitions of the ICE dataset: These are termed CLEAR ICE, BLUE GOATS, BLUE WOLVES, and BLACK ICE. The latter consists of those images that have pathological error rates on all SDKs. On request NIST will release the image partitions to interested parties. Sec. 8.10

A = SAGEM	B = COGENT	C = CROSSMATCH	D = CAMBRIDGE	E = L1	$x1 = \text{PRIMARY}$
F = RETICA	G = LG	H = HONEYWELL	I = IRITECH	J = NEUROTECHNOLOGY	$x2 = \text{SECONDARY}$
KIND 1 = RAW 640x480		KIND 3 = CROP		KIND 7 = CROP+MASK	KIND 16 = CONCENTRIC POLAR

- ▷ The IREX study defined a “C” API to support the various investigations. This API was implemented successfully by all the IREX participants. It is published⁸, freely available, and easily re-usable by other testing programs. It could be used with its full IREX record functionality, or in a more generic raster-template-match mode.

NEXT STEPS AND SUPPORT FOR DEVELOPERS

- ▷ NIST will consider requests for additional quantitative feedback in support of algorithm development. As part of this process NIST will invite IREX participants to inspect some problematic images.
- ▷ NIST invites comment on what further work is needed in support of standardization of iris image interoperability. Further, NIST solicits input on how the IREX umbrella program might be extended to support iris image interoperability. Comments and inquiries are welcome via IREX@NIST.GOV.
- ▷ NIST will initiate a second activity under the IREX umbrella. The project, the *Iris Quality Evaluation and Calibration* (IQEC), focuses on evaluation and calibration of iris quality scores. IQEC is a large scale evaluation of iris quality scores and will commence in Fall 2009. IQEC aims to evaluate the effectiveness of image quality assessment algorithms IQAA in predicting recognition accuracy of particular comparison algorithms (from the supplier of the IQAA), and of others’ algorithms. Given the IREX result that quality scores are not immediately interoperable, IQEC will establish a calibration procedure of IQAAs.

⁸See http://iris.nist.gov/irex/irex_api.pdf