



Facial Image Quality Improvement and Face Recognition Study Final Report

December 7, 2009



**Homeland
Security**

*United States Visitor and Immigrant
Status Indicator Technology
(US-VISIT) Program*

Contents

Executive Summary	5
1 Document References	11
2 Introduction.....	12
2.1 Background.....	12
2.2 Business Requirements.....	13
2.2.1 Current Facial Image Capture Process	13
2.2.2 Proposed Facial Image Capture Process	14
2.2.3 Requirements.....	14
3 Baseline Quality of Current POE Images	15
3.1 Overview.....	15
3.2 Manual Quality Assessment	15
3.2.1 Cropping.....	16
3.2.1.1 Effect on automated matching	16
3.2.2 Intensity Saturation	17
3.2.2.1 Effect on automated matching	17
3.2.3 Head Pose.....	18
3.2.3.1 Effect on automated matching	18
3.3 Variation in Image Quality at Different Ports of Entry	19
3.3.1 Results	19
4 Test Phases.....	23
4.1 Camera Pre-Assessment	23
4.2 Image Quality Assessment Software	25
4.3 Phase I Testing.....	26
4.3.1 Test Setup.....	27
4.3.2 Test Administration.....	27
4.4 Phase II	28
4.4.1 Test Setup.....	29
4.4.2 Test Administration.....	29
5 Results.....	32
5.1 Camera Pre-Assessment	32
5.2 Phase I.....	34
5.2.1 Image Collection	35
5.2.2 Automated Quality Assessment	35
5.2.2.1 Comparison of Commercial Quality Assessment Products	35
5.2.2.2 Performance of Commercial Image Quality Assessment Products	38
5.2.2.3 Failure to Enroll	39
5.2.3 Conclusions and Down-select	40
5.2.4 Analysis of Phase I Image Quality Scores	40
5.2.4.1 Visual Determination	40
5.2.4.2 Image Quality as a Predictor of FR Performance	41
5.3 Phase II	47
5.3.1 Image Collection	47
5.3.2 Acquisition Rates	48
5.3.3 Reliability of Automated Quality Assessment/False Accept	49

5.3.4	Failure to Acquire/False Rejection.....	51
5.3.5	FR Performance.....	53
5.3.5.1	Performance at Different Timeouts.....	54
5.3.5.2	User Selected Image.....	55
5.3.6	Timing Results	56
5.3.6.1	Eye Finding Time.....	56
5.3.6.2	Quality Assessment Time	56
5.3.6.3	Operational Frame Rate	56
5.3.7	Phase II Quality Metric Distributions Compared to Three Other Datasets.....	56
6	Conclusions.....	62
6.1	Camera Pre-assessment.....	62
6.2	Phase I.....	63
6.3	Phase II	63
6.3.1	Speed of Quality Test.....	63
6.3.2	Reliability of Quality Test.....	64
6.3.3	Failure to Acquire.....	64
6.3.4	Face Recognition.....	64
6.3.5	Comparison of Quality of Test Images with POE Images	64
6.4	Recommendations.....	64
7	NIST Interagency Report 7540, Assessing Face Acquisition.....	65
8	Acronyms and Abbreviations	67
	Appendix A: Summary of Camera Characteristics.....	1
	Appendix B: Summary of Camera Measurements	B-1

The following are available as separate attachments to this report:

- Attachment 1: Baseline Quality of US-VISIT POE Facial Images
- Attachment 2: Camera Pre-Assessment

Figures

Figure 2-1: a) US-VISIT POE image and b) e-Passport image	13
Figure 3-1: Effect of manually encoding cropping on matching performance.....	17
Figure 3-2: Effect of manually labeling saturation on matching performance.....	18
Figure 3-3: Effect of manually labeling aberrant poses on matching performance.....	19
Figure 3-4: Variation in eye detection confidence across POEs.....	20
Figure 3-5: Variation in face shadow measures across POEs.....	20
Figure 3-6 - Variation in background brightness measures across POEs.....	21
Figure 3-7: Variation in the face centering measure across POEs.....	21
Figure 3-8: - Variation in the background consistency measure across POEs	22
Figure 3-9: Variation in the yaw pose estimate across POEs	22
Figure 4-1: Image test patterns: a) Kodak Q13 grayscale test pattern; b) GretagMacbeth ColorChecker with reference map; c) ISO 16067-1 with slant edge regions of interest	24
Figure 4-2: Phase I GUI.....	28
Figure 4-3: Phase II FIQIFRS Application GUI.....	30
Figure 5-1: CanonG9 image with enlarged eye captured at (a) 4000 x 3000 pixels and (b) 1600 x 1200 pixels.....	34
Figure 5-2: Logitech QuickCam Pro 5000 (a) and 9000 (b) images	34
Figure 5-3: Reference images	38
Figure 5-4: Vendor C failure-to-enroll rates.....	39
Figure 5-5: Vendor A failure-to-enroll rates.....	39
Figure 5-6: a) Compliant image that passes quality test; b) Non-compliant image that fails quality test.....	41
Figure 5-7: FR scores by camera; demonstrates a statistically significant difference ($p < 0.0001$) in FR performance between the two tested cameras across lighting scenarios.	42
Figure 5-8: Overall (Canon and Logitech 9000) FR by lighting scenario; statistically significant difference ($p < 0.0001$) in performance for each of the lighting scenarios.....	43
Figure 5-9: Canon FR scores by lighting scenario.....	43
Figure 5-10: Logitech 9000 FR scores by lighting scenario.....	44
Figure 5-11: Predicting FR from quality metrics (Canon, all lighting scenarios); illustrates the ability of linear model for Canon to predict FR performance across all lighting scenarios.	45
Figure 5-12: Predicting FR from quality metrics (Canon, ambient); linear model predicts FR performance with very high degree of accuracy ($rSq = 0.9$).	46
Figure 5-13: FR prediction model vs. quality metric (Canon, ambient).....	46
Figure 5-14: Acquisition rates versus time for (a) the Logitech Camera and (b) the Canon camera. In this context, the acquisition rate refers to the fraction of participants for whom there is at least one available image that passes the quality checks.	49
Figure 5-15: Compliance rate of automatically passing images and those selected by operator .	50
Figure 5-16: Percentage of automatically-selected images deemed non-compliant by human....	50
Figure 5-17: Percentage of cases where at least one frame passed the quality test (and was also deemed compliant by a human) within 10 seconds	51
Figure 5-18: Example cases for which software could not find a passing image.....	52
Figure 5-19: FTA cases - where no images passed the quality test.....	53
Figure 5-20: DET curves for images captured using quality-in-the-loop and for the first frame images. (a) Logitech Camera. (b) Canon Camera.....	54

Figure 5-21: Simulation of various timeout periods for (a) Logitech, and (b) Canon.....	55
Figure 5-22: Performance comparison of operator-selected image versus the first image that passed all the quality checks	55
Figure 5-23: Phase II Eye Distances Compared to 3 Other Datasets	57
Figure 5-24: Phase II Sharpness Metric Compared to 3 Other Datasets	58
Figure 5-25: Phase II Mouth Open Metric Compared to 3 Other Datasets	58
Figure 5-26: Phase II Deviation from Frontal Pose Metric Compared to 3 Other Datasets.....	59
Figure 5-27: Phase II Eye Gaze (Looking at Camera) Metric Compared to 3 Other Datasets.....	59
Figure 5-28: Phase II Eyes Open Metric Compared to 3 Other Datasets	60
Figure 5-29: Phase II Head Roll Metric Compared to 3 Other Datasets	60
Figure 5-30: Phase II Brightness Exposure Metric Compared to 3 Other Datasets	61
Figure 5-31: Phase II Uniform Lighting Metric Compared to 3 Other Datasets	61

Tables

Table ES-1: Findings	6
Table ES-2. Technical Recommendations.....	7
Table 3-1: POE image non-frontal pose categorizations	18
Table 4-1: Cameras tested.....	23
Table 4-2: Phase I Lighting Scenarios	26
Table 5-1: Capture Dimensions, Frame Rate, Compression, and Field of View.....	33
Table 5-2: Example images captured during Phase I.....	36
Table 5-3: Presence of desired metrics in commercial image QA products.....	38
Table 5-4: Example images captured during Phase II	47
Table 5-5: Eye distances	48

Executive Summary

The Facial Image Quality Improvement and Face Recognition Study (FIQIFRS) project was initiated in February 2007 to investigate technology for improving the quality of face images captured at United States (U.S.) ports of entry (POEs). The project's goal was to bring US-VISIT face images into compliance with standards defined in the *Registry of U.S. Government Recommended Biometric Standards*¹ and to improve quality sufficiently to ensure accurate recognition by both humans and computer systems while minimizing operational impacts and allowing for technology maturation. The project was a Technology Assessment that involved laboratory testing and development of a proof-of-concept application; it did not include integration or deployment to POEs. Baseline image quality was established through analysis of operational POE images by the National Institute of Standards and Technology (NIST).

The project team investigated hardware (camera) and software (face finding and image quality assessment) approaches to ensure capture of compliant images in as automated and expeditious a manner as possible. The camera types evaluated included the currently deployed webcam, higher resolution webcams, a video camera with pan-tilt-zoom, a wide dynamic range camera, and a digital still camera. Image quality characteristics of each camera were profiled objectively using a series of test targets and associated image analysis software. Facial image collection then occurred in two phases. Phase I occurred in April 2008; and Phase II in September 2008. For Phase I, cameras from six different categories were selected for evaluation, and images were collected from a small volunteer population in a simulated POE environment under several different lighting conditions. A custom interface was developed to control the cameras and capture images. The images were analyzed retrospectively with image quality assessment (QA) software. Image quality thresholds for the next phase of testing were established based on the analysis of Phase I data.

Phase II testing involved integrating two cameras with a commercially-available image QA software product to automate image capture by performing real-time QA. This 'quality-in-the-loop' application was tested on a larger volunteer population in a simulated POE environment under several lighting scenarios. All images that met the quality thresholds were displayed to the tester, and were saved for subsequent analysis. If no images met the thresholds, the tester captured a snapshot manually, similar to current operational practice.

Image collection was conducted during a mock inspection interview. Volunteers were instructed to stand at a fixed distance (70 cm) from each camera. Although this 'arms length' distance is somewhat greater than the camera-traveler distance in current operations, it represents the minimum distance specified in the face recognition (FR) standard [5] and most closely reflects the capture conditions and lane widths at POEs. The cameras were positioned at a fixed height (60 inches) and remained so throughout the test session.

The facial image data collected during Phase I and Phase II was analyzed for image quality and FR match performance. FR performance was assessed by matching reference images captured separately from each volunteer to the images captured during the mock inspection interview.

¹ <http://www.biometrics.gov/Standards/StandardsRegistry.pdf>

Table ES-1 below contains significant findings from the project, followed by the working group's recommendations in Table ES-2.

Table ES-1: Findings

Image Quality Assessment (QA) Software	Capture Performance—at 0.25 seconds, the processing speed with quality-in-the-loop can support real-time operations.
	Image QA Software: <ul style="list-style-type: none"> • QA software can be used to acquire a standards-compliant face image that is suitable for both human and automated FR. • Commercial image QA products differ widely due to the lack of standardization, as each product measures different quality factors over different value ranges. • QA software is fallible; hence, no perfect cut-off threshold could be found for determining compliance/non-compliance. A consistent set of QA metrics that predicted FR match scores was not found using regression analysis; therefore, quality metric thresholds had to be determined visually
	Failure Rates <ul style="list-style-type: none"> • Lighting impacted image enrollment rates. The overall failure-to-acquire rate (those cases where no images in a test session met QA thresholds) was 16.7%. Ambient lighting, which is most similar to POE lighting conditions, had more failures than other lighting scenarios (with the exception of back lighting). • Failures to acquire are believed to be the result of the QA software failing to recognize good-quality face images and to the lack of vendor algorithm training with dark-skinned individuals.
Face Recognition	FR scores were higher for the digital still camera than for the webcam.
	Images that met quality thresholds had better match scores than the first image captured at the start of each Phase II test session (captured prior to executing quality-in-the-loop), demonstrating that integrating QA into the capture process can produce face images that are more suitable for automated FR.
	This report recommends that automated FR be deployed in DHS operations only after the performance and quality gains demonstrated in this study can be demonstrated in at least a close-to-operational POE field trial. If elevated quality can be achieved, FR <u>might</u> supplement fingerprint-based verification processes.
User Interface	The interface for quality-in-the-loop prompted the user to select the best image from up to four that met quality thresholds. Results showed that user selection did not improve FR match performance; however, the user-selected images had a higher standards-compliance rate (as determined by a human reviewer) than the other images that met the QA thresholds.
Camera Evaluation	The digital still camera produced images that were visually superior to the other cameras examined; the newer webcams tested were superior to the webcam in use when the study was initiated.
	Mounting and operating the camera in portrait mode and at a fixed height allowed for capture of a greater range of heights without camera repositioning.
	The webcams were difficult to mount on standard tripods.
Other	Eye Glasses – test administrators did not consistently ask that people wearing glasses remove their glasses during image capture. On some occasions, presence of eye glasses resulted in failures to acquire. This supports current operating procedure of asking travelers to remove their glasses prior to facial image capture.

Recommendations

More study will likely be needed before a strategy for integrating improved facial image quality with inspection operations can be developed. The following are some preliminary observations for training and design of the facial image capture interface.

- Efforts should be made to communicate the importance of image quality to Customs and Border Protection (CBP) Officers and to provide information about basic image quality standards— i.e., eyes open, mouth closed, face full frontal, traveler arms' length from camera. Consideration should be given to initiating these communication efforts prior to the integration of image quality and face-finding software in the inspection process.
- No changes are necessary to the sequence of basic tasks associated with biometric capture.
- Changes to the user interface should minimize the input required from CBP Officers.
- US-VISIT and CBP should continue to work together to determine the optimal points for user input in the design of the interface. These may include an action to initiate image capture and an action to indicate that image capture has been completed.

Table ES-2 contains recommendations for future integration of image QA into the inspection process.

Table ES-2. Technical Recommendations

Camera Hardware Specification	The study supports CBP's decision to adopt the Logitech 9000 camera. This choice affords superior optical performance of the camera and adequate frame rate. The camera should be used in a portrait format (where the height is the longest dimension) to accommodate variation in visitor height, and at the resolution specified in this report (1600 x 1200).
	While contemporary digital point-and-shoot cameras offer superior optical performance, they should not be used because the elevated resolution is not needed for the current intended use of the images (manual confirmation of identity or automated facial recognition) and produces slower operation. In addition, the point-and-shoot cameras are likely to be more frequently stolen, their interface is proprietary, and their power feed and data cables are separate.
	CBP has experienced difficulty acquiring sufficient quantities of replacement Logitech webcams as the cameras reach end-of-life. Newer models tend to use different drivers and mounting options than prior models. CBP should aim for a modular software interface (e.g., a standardized Application Programming Interface) to minimize the operational impact of camera end-of-life.
	The optical performance of future cameras should be validated against current baselines. A summary of this procedure is given in Attachment 2 (Camera Pre-Assessment). Optical performance here is a generic phrase to reference numeric values of frequency response, uniformity, linearity, and distortion.
Physical Infrastructure in POEs	The gooseneck mounting of face cameras should be eliminated and replaced with a fixed-mount camera. The camera should operate in a wide field of view and portrait mode to support a wide range of visitor heights. The field of view should be set to be sufficient to image a person in a wheelchair. To handle the extremely tall exceptions the operator might instruct the visitor to bend at the knee and look at the camera. Extremely short visitors should be directed to a handicapped access lane.
	The camera should be placed such that the visitor-camera distance is about 0.7m. CBP practice should respect this approximately "arms length" specification.

Physical Infrastructure in POEs (cont'd)	The floor of the POE lanes should be equipped with "yellow footprint" ² stickers to indicate where the visitor should stand. The footprints guide the visitor in both their longitudinal (along-the-lane) and lateral (near-far from the Officer) positions. The footprints also guide visitors on their orientation with respect to the Officer.
	Modification of the POE lighting environment was explicitly out of the scope of this study. However, it is recommended that further study be conducted if physical infrastructure upgrades of POEs can include modification to the lighting.
	A considerable variation between the quality of images has been observed across POEs. This arises because of varying local environmental conditions (e.g. lighting) and possibly operating procedures. Thus we recommend that installation of new cameras be accompanied by an immediate review of the installation followed by an interval of performance monitoring.
Client Side Software Specifications	Automated face image QA software should be included in the CBP client application. The face image quality software should report eye coordinates to support the recommendation for the Token Frontal image type. Source code for the US-VISIT face improvement capture harness is available. To improve eye finding speed, once the initial eye locations have been determined by the software, eye finding for subsequent frames could be restricted to a narrow region based on location of the eyes from the previous frame.
	The project configured the face image QA software and established a set of quality thresholds. In making a determination of acceptable face image quality, the CBP client should be configured to compute an appropriate set of quality metrics, and to compare those against pre-calibrated thresholds.
	In cases where the quality analysis software fails to render a verdict (e.g., because quality criteria were not met) the image with the best pose angle estimate should be used. If no such image is available then the inspector should issue an explicit instruction for the visitor to look at the camera and initiate a manual shot (as is the existing practice circa 2008). The CBP client should support manual override. In such cases, the CBP client might display an "oval overlay area" in which the visitor's head should appear.
Client side Graphical User Interface (GUI) specifications	Careful design of the CBP client user interface is important. The automatic quality-in-the-loop checks are effective at excluding images of poor quality. This resulted in the presentation only of good images to the inspector, and it is, therefore, recommended that the CBP client: <ul style="list-style-type: none"> • Should not require the inspector to manually select an image. • Should indicate "ready" as soon as one acceptable quality image has been acquired. • Should allow a manual override and restart.
Standards Compliance	All face images retained in the Automated Biometric Identification System (IDENT) or other US-VISIT systems should be formatted in the binary format defined in ISO/IEC 19794-5:2005 (FACESTD). The records should claim "basic" conformance (ref clause 6.4.3 of the standard). The standard is recommended in the National Science and Technology Council's <i>Registry of USG Recommended Biometric Standards</i> .
	All face images should be cropped and rotated (in the roll direction) to have the fixed "Token Frontal" geometry defined in FACESTD. This specification requires estimation of the eye positions, which can be determined by QA software. This format is used for ePassport images.
	Given extant bandwidth constraints to IDENT, the face image data should be

² Ref [16], NISTIR 7540, Assessing Face Acquisition.

Standards Compliance (cont'd)	compressed with JPEG 2000. The compressed data resides in the FACESTD data record. Open source software (ref JasPer Project) is available for JPEG 2000 compression. Examples of the records may be sent to NIST for an informal test of conformance.																				
	<p>The target eye distance for US-VISIT images should be 90 pixels (normative requirement for FACESTD Full Frontal face images). This eye distance corresponds to a width and height of 360 and 480, respectively, using the equations below.</p> <table border="1"> <thead> <tr> <th>Feature or Parameter</th> <th>Value (FACESTD Table 14)</th> <th>Recommended US-VISIT Value</th> </tr> </thead> <tbody> <tr> <td>Image Width</td> <td>W</td> <td>360</td> </tr> <tr> <td>Image Height</td> <td>W/0.75</td> <td>480</td> </tr> <tr> <td>Y coordinate of Eyes</td> <td>0.6 * W</td> <td>216</td> </tr> <tr> <td>X coordinate of First (right) Eye</td> <td>(0.375 * W) - 1</td> <td>134</td> </tr> <tr> <td>X coordinate of Second (left) Eye</td> <td>(0.625 * W) - 1</td> <td>224</td> </tr> <tr> <td>Width from eye to eye (exclusive³)</td> <td>0.25 * W</td> <td>90</td> </tr> </tbody> </table>	Feature or Parameter	Value (FACESTD Table 14)	Recommended US-VISIT Value	Image Width	W	360	Image Height	W/0.75	480	Y coordinate of Eyes	0.6 * W	216	X coordinate of First (right) Eye	(0.375 * W) - 1	134	X coordinate of Second (left) Eye	(0.625 * W) - 1	224	Width from eye to eye (exclusive ³)	0.25 * W
Feature or Parameter	Value (FACESTD Table 14)	Recommended US-VISIT Value																			
Image Width	W	360																			
Image Height	W/0.75	480																			
Y coordinate of Eyes	0.6 * W	216																			
X coordinate of First (right) Eye	(0.375 * W) - 1	134																			
X coordinate of Second (left) Eye	(0.625 * W) - 1	224																			
Width from eye to eye (exclusive ³)	0.25 * W	90																			
Image Specifications	The images should be compressed with JPEG 2000, with a compression ratio less than 20:1.																				
	Exchangeable image file format (Exif) information should be stored in the image's header to keep track of camera models and other metadata tags, such as color profiles and camera settings.																				
Integration into IDENT	Ensure that the IDENT eXchange Messages (IXM) specification supports encapsulation of FACESTD face images.																				
Future Tests	We recommend that a field study in a POE be conducted. Such a field study should use a two computer configuration. The first runs the unmodified operational CBP client. The second implements the quality-in-the-loop acquisition. The default POE camera and the field study camera should be identical (Logitech 9000), and mounted in very close proximity (inches). The system should be tested with volunteers first, and, after appropriate trials, it should proceed with real travelers.																				

³ The standard incorrectly says “*inclusive*”.

Disclaimer

Specific hardware and software products identified in this report and its attachments were used to perform the technology assessment described in the report. In no case does identification of any commercial product, trade name, or vendor, imply recommendation or endorsement by DHS/US-VISIT, its contractors, or the National Institute of Standards and Technology. As a result of this technology assessment, DHS made no commitment to purchase any software or service offered by the vendors. The results reported were produced in experiments designed and conducted by DHS/US-VISIT to develop a proof-of-concept application, and should not be construed as an evaluation of the products, nor as representative of the vendors' maximum-effort or full capabilities.

1 Document References

- [1] American National Standards Institute/ InterNational Committee for Information Technology Standards (ANSI/INCITS) 385-2004, American National Standard – Information Technology - Face Recognition Format for Data Interchange.
- [2] Approaches to Face Image Capture at US-VISIT Ports of Entry, November 2007, L. Nadel, in Proceedings of the Second NIST Biometric Quality Workshop, <http://biometrics.nist.gov/quality/workshop07/presentations.html>.
- [3] Appendix F, Integrated Automated Fingerprint Identification System (IAFIS) Image Quality Specifications, of the Electronic Biometric Transmission Specification (EBTS), Ver. 8.1, 11/19/2008, www.fbibiospecs.org.
- [4] International Civil Aviation Organization (ICAO) Document 9303, Machine Readable Travel Documents, Part 1, Machine Readable Passports, Volume 2, Specifications for Electronically Enabled Passports (E-Passports) with Biometric Identification Capability, Sixth edition, 2005.
- [5] International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) 19794-5:2005 — Information technology — Biometric data interchange formats — Part 5: Face image data (FACESTD).
- [6] ISO 13407:1999, Human-centred design processes for interactive systems.
- [7] E-Passport Mock Port of Entry Test Findings, November 29-December 2, 2004.
- [8] Homeland Security Adopts Facial Recognition Standard, DHS Press Release, October 27, 2004.
- [9] NISTIR 7540, Assessing Face Acquisition, Mary Theofanos, Brian Stanton, Charles Sheppard, Ross Michaels, John Libert, Shahram Orandi, NIST. http://zing.ncsl.nist.gov/biiousa/docs/face_IR-7540.pdf.
- [10] Registry of (U.S. Government) USG Recommended Biometric Standards, www.biometrics.gov/standards.
- [11] Attachment 1: Facial Image Quality Improvement and Face Recognition Study— Baseline Quality of US-VISIT POE Facial Images.
- [12] Attachment 2: Facial Image Quality Improvement and Face Recognition Study— Camera Pre-Assessment.

2 Introduction

The Facial Image Quality Improvement and Face Recognition Study (FIQIFRS) project was defined and conducted as a proof-of-concept of image quality-in-the-loop using commercially available cameras and image QA software, not a formal technology test. As such, it examined representative cameras from several different categories; it did not include an exhaustive study of different camera types. Likewise, a small set of available facial image QA products was identified and examined to demonstrate and evaluate the quality in-the-loop concept. This report documents the activities performed for the FIQIFRS project, including the test environment, camera assessment, image collection, analysis, and FR performance. It includes the results of baseline analysis of operational images, and findings and recommendations of the project team.

2.1 Background

The face images currently captured at U.S. ports of entry (POEs) do not conform either to the national FR standard⁴ [1] adopted by DHS or to FACESTD [5], the international standard specified in the *Registry of USG Recommended Biometric Standards* [10] and in the standard for ePassports [4], and are of relatively low quality, typically insufficient for use with automated FR systems. Customs and Border Protection (CBP) workstations use Logitech webcams for the acquisition of face images at POEs. The earlier Logitech models installed at POEs (QuickCam Pro 4000 and 5000) have low spatial resolution, and due to their low-quality optics and sensor, the image quality obtainable from these cameras is limited. The more recently deployed webcam, the Logitech QuickCam Pro 9000, was one of the cameras evaluated for the FIQIFRS project. It has higher resolution and a better sensor capable of acquiring good quality images, as described in this report. The webcams are manually positioned and operated by CBP Officers, and the lack of controls over pose and head size in the image, combined with non-uniformity in lighting and background, and variable contrast and brightness, substantially decreases the potential accuracy of FR when applied to the images captured.

Since the webcams installed at POEs lack a universal threaded hole for mounting, they are mounted on a gooseneck, which provides flexibility for CBP Officers to position the camera for each visitor. There are drawbacks to these mounts, however. The camera mounting frequently springs back when aimed at a significant yaw angle, which requires the Officer to hold the camera in place. Latency in actual image capture is such that the image captured may be significantly different (i.e., traveler has had a chance to move head or close eyes) than what the Officer intended to capture. However, if the Officer did not continue to hold the camera in place, it would spring back and the captured image would be skewed. *Note: These mounts have subsequently been replaced with mounts that permit only forward and backward movement to accommodate traveler height, which should alleviate the need to hold the camera in place during image capture.*

Inspection Officers have received training and instructions on capture of fingerprints, and the fingerprint client software assesses and provides feedback to the officer on fingerprint quality.

⁴ ANSI INCITS 385–2004

For facial images, however, Officers are instructed only to capture the facial photograph — there are no requirements regarding facial expression, centering or size of the face within the frame, or pose angle (i.e., conformance to best practices for taking photographs for facial images), no quality assessment of the image, and no feedback to the operator to indicate that the facial image did or did not meet any image quality standards.

To make the images more useful for visual confirmation of identity, to support interagency exchange of facial images (i.e., as required in Homeland Security Presidential Directive-24), and to enable future use of automated FR as a facilitator biometric, the resolution and quality of the images needs to be improved. As examples of image quality, Figure 2-1a depicts an example POE image, and Figure 2-1b shows an ICAO Document 9303 [4] compliant e-Passport image.

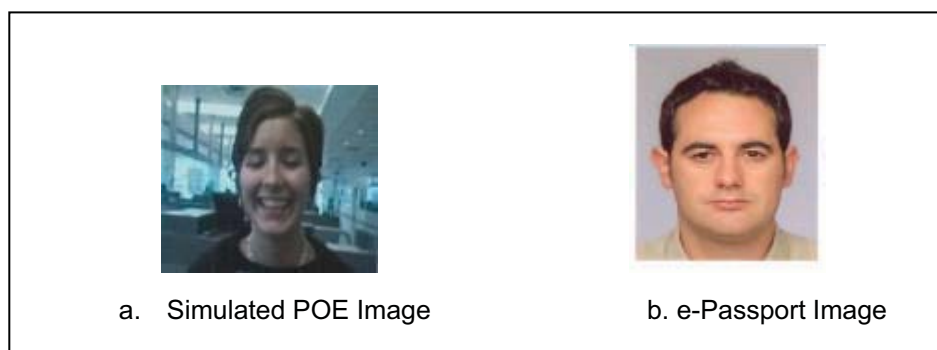


Figure 2-1: a) US-VISIT POE image and b) e-Passport image

The US-VISIT Face WG embarked upon the FIQIFRS project to investigate technology for improving the quality of face images captured at POEs. The FIQIFRS study examined changes to software and cameras to automatically capture and assess facial images to ensure they meet minimum thresholds for image quality. The metrics used to assess image quality are based on requirements defined in ISO/IEC 19794-5 and ANSI/INCITS 385-2004. The goal of this project was to bring US-VISIT face images into compliance with standards and to improve image quality sufficiently to ensure accurate recognition by both humans and computer-based systems, while minimizing operational impacts and allowing for technology maturation.

2.2 Business Requirements

Business requirements for the study were developed based upon examining the current and proposed context under which facial images are captured during the inspection process. The following sections describe the current and proposed business use cases for image capture.

2.2.1 Current Facial Image Capture Process

Facial images of US-VISIT in-scope visitors are captured during the biometric capture portion of the inspection process. The IDENT client application first prompts the officer to collect fingerprints from the visitor. Following fingerprint capture, the live video stream from the Logitech webcam is displayed, the officer positions the Logitech webcam, instructs the visitor to look at the camera, and presses a key (e.g., Enter), to capture a single facial image from the video stream; the application then disconnects from the camera. If the officer deems the captured image to be acceptable, the officer selects the “save” option to save the image. The photo capture time is estimated at 5 seconds based on data collected during the deployment of ten-fingerprint scanners.

2.2.2 Proposed Facial Image Capture Process

The automated facial image capture process (quality-in-the-loop) involves integration of a commercially available image QA software development kit (SDK) with the cameras. The FIQIFRS application captures a series of facial images within a specified timeframe, and each image is analyzed by image QA software to determine if it meets minimum thresholds for metrics that are known to impact face recognition (e.g., minimum resolution determined by distance between eyes, eyes open/closed, frontal pose). Images that pass the minimum thresholds are displayed to the operator, and the operator selects the image he/she determines to be the best and saves it. If no images pass the quality thresholds, the operator either restarts the video capture and image QA sequence or takes a single still image, similar to the current operating procedure.

2.2.3 Requirements

The business requirements for automated image capture and image QA were developed by US-VISIT and CBP as changes to the current manual process. The requirements are defined as part of the facial image capture process.

1. Following collection of fingerprints, the CBP Officer instructs the traveler to position self toward the camera for facial image capture, and initiates the automated image capture process. Requirements for positioning and capture are:
 - a. Automated image capture will use the fact that each individual is anchored to a specific field of view when presenting their fingers for capture to the fingerprint capture device to maximize correct orientation of the facial image.
 - b. Automated image capture should require minimal or no adjustment of camera angle to capture facial image in the correct orientation.
 - c. Automated image capture should minimize CBP Officer involvement to the greatest extent possible.
 - d. Camera should require no lighting (strobes, flashes, etc.) that is deemed obtrusive to the traveler or Officer.
 - e. Camera system should be able to adapt to the current system as a replacement for current camera system.
2. The system examines the facial photo and determines if the image meets acceptable quality:
 - a. Quality of the facial image should use quality requirements stated in DHS standards as criteria.
 - b. Quality of facial image should be sufficient for both human and automated facial recognition.
3. If no facial photo meets acceptable quality, the system repeats steps 1 & 2.
4. In the event the system is unable to capture an image of acceptable quality, the system shall offer the Officer an opportunity to override the quality check and save the image.
5. If photo does meet acceptable quality, the system saves the facial image.
 - a. System processing time for positioning, image capture, quality determination, any needed recapture and/or subject repositioning, and processing time to save the image should not exceed current processing time.

3 Baseline Quality of Current POE Images⁵

To establish the baseline quality of POE images, US-VISIT provided a set of operational POE images to the National Institute of Standards and Technology (NIST) for analysis. The POE images were collected between 2004 and 2007 at various ports of entry. The set of image data included the corresponding Department of State (DOS) BioVISA (BVA) images collected during visa application. NIST analyzed the data by manually assessing the frequency at which certain defects occurred, and then used a FR product to determine the impact of those defects on one-to-one matching performance. This characterization established a baseline for face image quality as it currently exists against which images collected as part of the FIQIFRS project could be compared.

3.1 Overview

NIST initially examined the face images in April 2004 and found that the images were generally of poor quality and should not ordinarily be used in automated FR processes. Analysis of more recently captured images showed no improvement in FR performance despite a camera upgrade and minor changes to the capture protocol over that time.

NIST performed a manual inspection of several thousand POE images to identify the frequency and impact on FR performance of specific defects (e.g. poor lighting, non-frontal faces, face cropping). An automated image QA tool was used to compare POE images to images from other face databases, such as the Face Recognition Technology (FERET) database. The QA tool indicated that POE images suffer from several problems, including non-centered faces, blurry faces, non-frontal head poses, and poorly illuminated faces. The following sections present results of NIST's manual QA of the images and analysis of the variation in image quality at different POEs.

3.2 Manual Quality Assessment

In early 2008, NIST conducted a more systematic survey of the POE images. This involved manual inspection of 20,000 images and application of a graphical image categorization tool to label images presenting certain defects.

Inspecting the images manually is reliable in the sense that a human observer is capable of identifying a particular problem even in the presence of other problems, and can distinguish between failure modes. For example, if the facial region is saturated and also cropped at the left eye, a human observer will note both defects, whereas an automated QA tool is likely to not find the face at all and report nothing.

A drawback of the approach is that it is subjective. Thus, when categorizing saturation, there is an inherent judgment to be made in distinguishing a bright image from a saturated one.

⁵ The full report on NIST's baseline assessment of image quality is available as Attachment 1 [11] to this report.

The manual inspection focused on three specific defects that are known to exist in the set of POE images: cropped faces, over-exposed faces, and non-frontal head poses. Together these defects strongly degrade the ability of contemporary FR engines to identify or verify the subjects appearing in the images. The following subsections describe the criteria for each of the defects and present results of the manual inspection.

3.2.1 Cropping

Images were manually inspected for cropped faces and each image was labeled as either cropped or un-cropped. An image was considered cropped if any part of the face, from chin to mid-forehead, or from ear to ear, was not present in the image. The determination was made regardless of whether other problems were present, such as a non-frontal head pose or poor lighting.

The manual inspection identified cropped faces in 11.1% of the total set of images. Cropped faces always occurred for one of three reasons:

- The camera was not pointing at the subject, and part of the face was outside of the camera's field of view.
- The subject was standing too close, and the camera's field of view was not wide enough to capture the entire face at once. The camera operator typically compensated by positioning the camera such that the chin was clipped off the bottom of the image. Small distances between the subject and the camera also cause lens distortion (i.e., the fish-eye effect).
- In rare cases, the subject's face was partially obscured by an object in the foreground, such as a suitcase or the back of a baby's head.

3.2.1.1 Effect on automated matching

The effect that cropped images have on matching performance was evaluated using an archived commercial matcher (produced circa 2005). The facial template generator was unable to find a face in 55% of the cropped images. This type of error, known as a failure to enroll, notionally causes the generation of a blank template that always gives a low score when matched.

Figure 3-1 plots verification performance for:

- the entire set of POE images matched against BVA images.
- the subset of images that were identified as un-cropped.

The improvement in performance is small suggesting that cropping is not a significant contributor to the large observed error rates⁶. In addition the improvement manifests itself more at higher false match rates (above 0.01). This indicates that cropping inhibits the initial ability to find the face in the image.

⁶ The error rate TMR = 0.55 at FMR = 0.01 is very much inferior to passport-against-passport matching, or to high-resolution-still to high resolution still matching, for which the error rates can exceed TMR = 0.95 at FMR = 0.01.

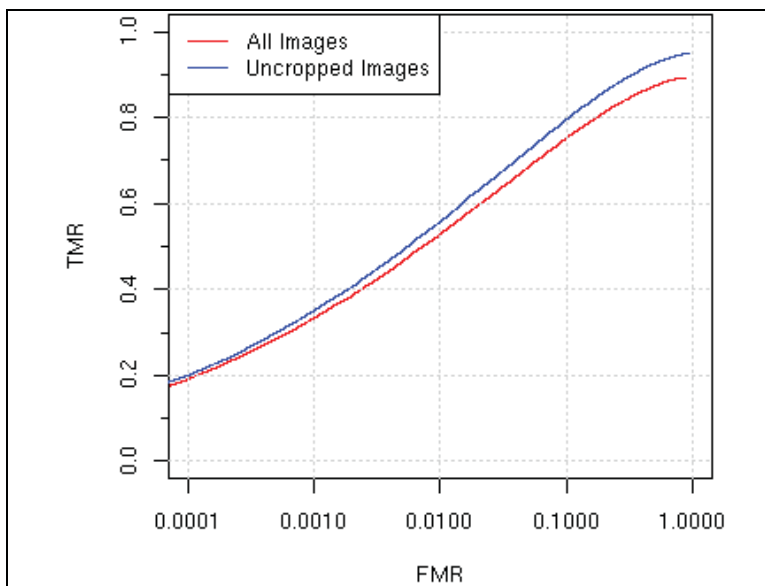


Figure 3-1: Effect of manually encoding cropping on matching performance

3.2.2 Intensity Saturation

Images were manually inspected for faces that were over-exposed to light at the time of capture. These faces contain excessively bright areas, sometimes referred to as hotspots. The pixels within a hotspot, expressed as red-green-blue (RGB) triplets, will have maximum intensities for all three of the colors ($R = G = B = 255$, if 255 is the maximum color). An image is labeled as saturated if hotspots are clearly visible in any region of the subject's face, from chin to eyebrows, or from ear to ear. Note this excludes certain parts of the face, such as the left or right flank of the nose and the forehead. Hotspots were so prevalent in these parts that not excluding them would have led to nearly every POE image being labeled as saturated. If a face was over-exposed to light, but not to the point of saturation, it was not labeled as saturated.

The manual inspection identified 18.4% of the total set of images as saturated.

3.2.2.1 Effect on automated matching

Figure 3-2 shows Receiver Operating Characteristic (ROC) curves for the images labeled saturated and not. The result, that there is essentially no difference, is perhaps surprising in that fully saturated pixels (i.e. regions at value 255) do not appear to undermine the FR process beyond the other defects present in the images (pose, resolution etc). Nevertheless, saturation may remain problematic once other problems are remedied.

This negative result is also included here so that such analyses are not repeated without a specific motivation.

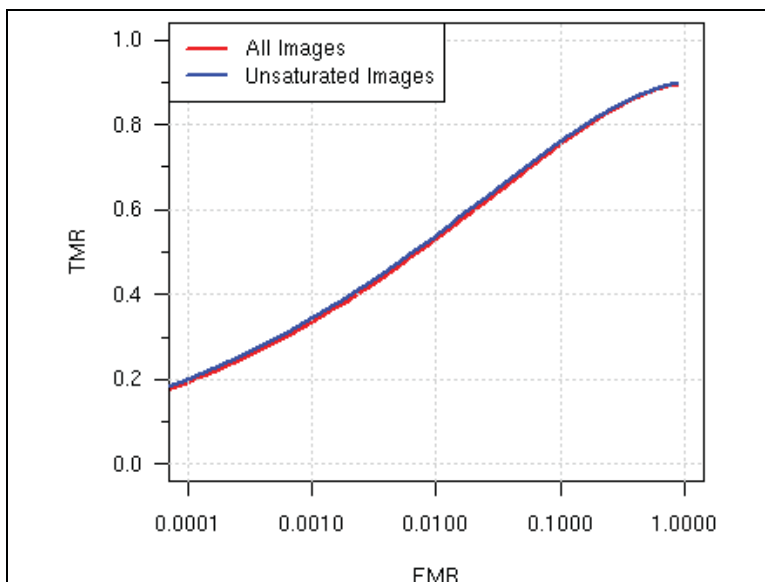


Figure 3-2: Effect of manually labeling saturation on matching performance

3.2.3 Head Pose

Images were manually inspected for non-frontal head poses. The images were assigned one of the following categories:

Code	Description	Fraction (estimated over 10000 images)
FR	Fully frontal, or very close (pitch and yaw are within roughly 5 degrees)	48.87%
PF	Partially Frontal. Not fully frontal, but not catastrophically off (5-15 degrees of yaw or pitch)	49.51%
NF	Non-frontal. Off by a lot.	1.62%

Table 3-1: POE image non-frontal pose categorizations

For each image, its category was assigned by visual inspection by a NIST staff member. This process is clearly not quantitative and any given image might be adjudicated differently by a different judge. Note that head pose is not the same as gaze direction, as a subject can be looking at the camera but still not be fully frontal. This circumstance may arise from an instruction (implied or explicit) from the officer to the traveler to "look at the camera" but for which the response is to adjust only the gaze. This might also be due to synchronization. The ideal response for FR should be to orient the head toward the camera.

3.2.3.1 Effect on automated matching

Figure 3-3 is included to support the assertion that FR will remain difficult in POE primary inspections without improved control of the head pose in relation to the camera. Note however that even for fully frontal images, the recognition process is poor (TMR = 0.6, FMR = 0.01). This

is a consequence of poor resolution and illumination, non-ideal compression, and image quality inadequacies in the accompanying BVA images collected by DOS.

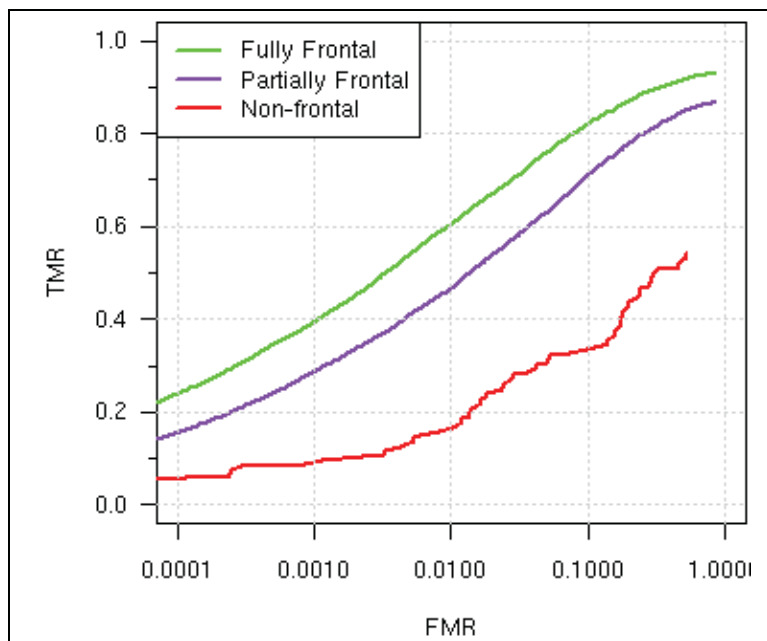


Figure 3-3: Effect of manually labeling aberrant poses on matching performance

3.3 Variation in Image Quality at Different Ports of Entry

The 2004 set of POE images was accompanied by the POE identifiers for each image. This supports comparison of image properties between border crossing sites. For the analysis conducted in 2008, the 18 busiest POEs were selected. For each, 3000 images were randomly sampled and an image QA SDK was applied to quantify quality.

Figure 3-4 – Figure 3-9 show the resulting score distributions for six quality metrics by POE. Five of the metrics (Face Shadow, Face Centering, Background Consistency, Background Brightness, and Face Brightness) were selected because they measure what NIST regarded as the most significant quality problems with the images. In addition, eye confidence was selected because it provides a general measure of the quality of an image (the eye confidence is expected to be lower for poorer quality images⁷). This variable was used to establish an ordering of the POEs. That is, from left to right the POEs appear in increasing order of the median eye confidence.

3.3.1 Results

Figure 3-4 shows that eye finding is significantly easier in the images collected at some POEs than others. This holds only to the extent that the variable reported, eye confidence, is a reliable

⁷ Successful automated facial recognition is critically dependent on the accurate and consistent localization of landmarks, primarily the eye centers.

indicator of actual detection of the eyes. In any case, it provides a consistent ordering for the POEs indicated in the remaining plots.

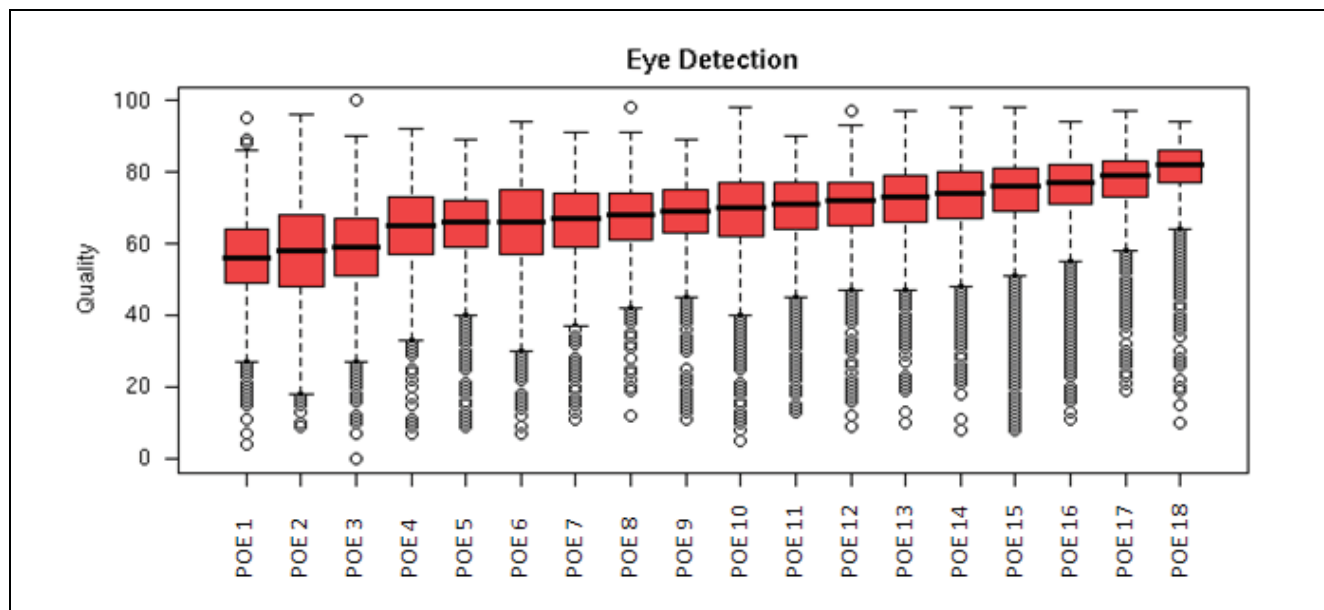


Figure 3-4: Variation in eye detection confidence across POEs

The trend of Figure 3-5 shows a positive correlation of reported face shadow with the eye detection confidence. This would be expected since shadows inhibit accurate localization of the eyes. The best and worst POEs are the same as for eye confidence. The variance of the distributions of the face shadow quantity is larger than for eye confidence. This may be due to the fact that a small pose variation can produce shadows while not affecting eye detection.

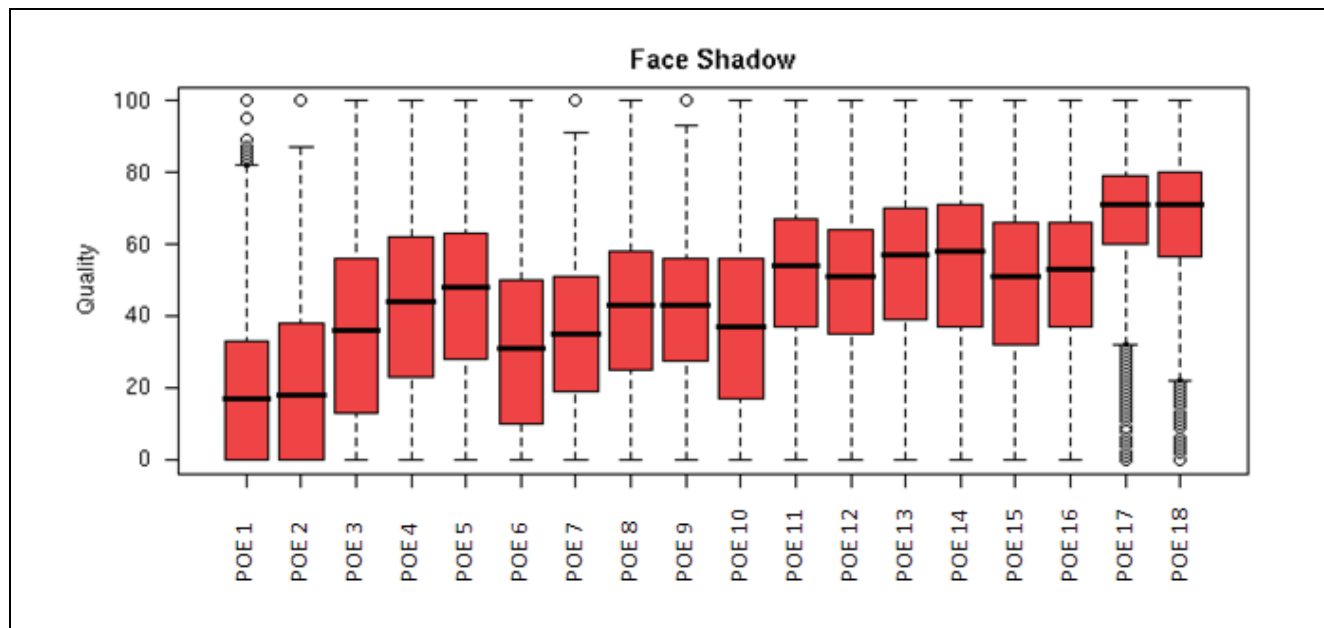


Figure 3-5: Variation in face shadow measures across POEs

Figure 3-6 shows a more varied picture. While a trend is present, the low variance but wide variation means that the background brightness measure is a characteristic of the POE itself. It is less correlated with eye detection confidence because the ability to detect eyes is not related to the background brightness if the face itself is properly exposed.

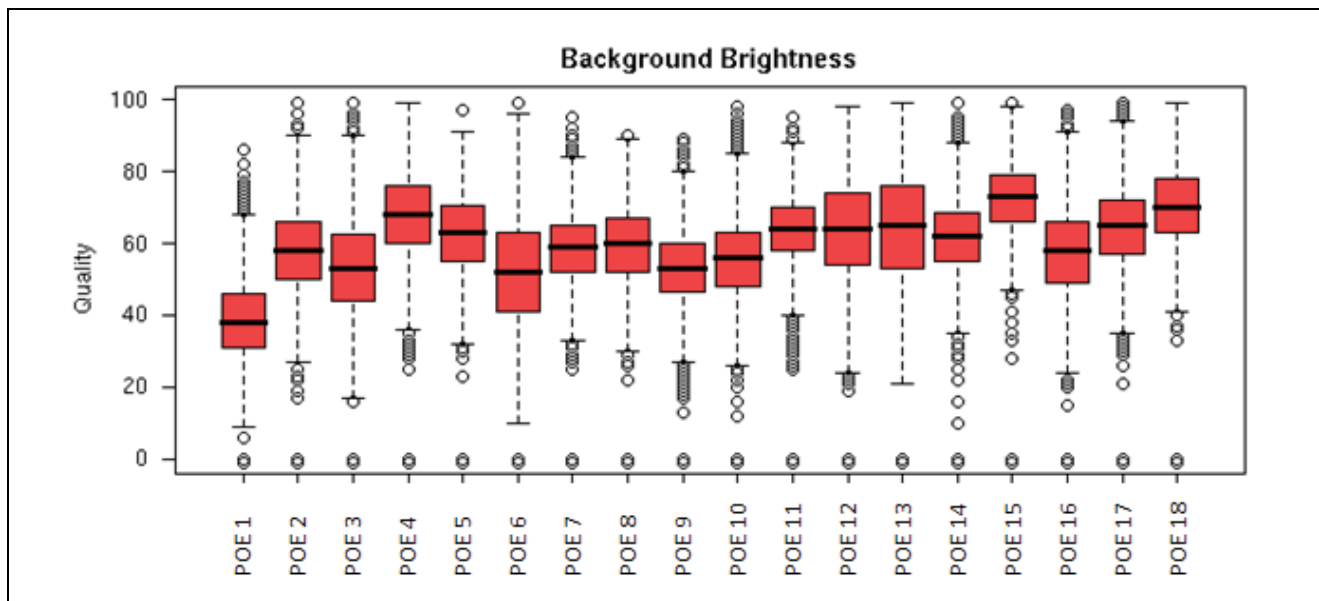


Figure 3-6 - Variation in background brightness measures across POEs

In comparison to the other variables, Figure 3-7 shows considerable consistency across POEs. The value in question, face centering, should be a property of the way CBP Officers aim the cameras, and of how travelers respond to the instruction. In addition, eye detection is largely independent of how well centered the faces are (as long as the face is not cropped).

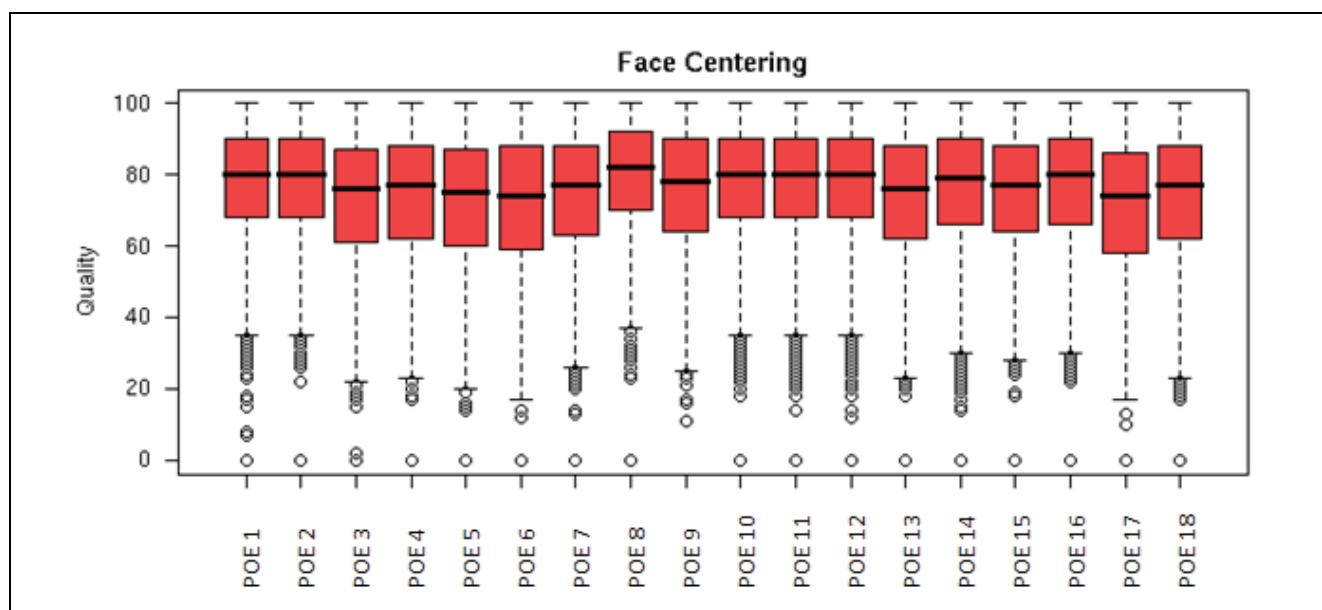


Figure 3-7: Variation in the face centering measure across POEs

Figure 3-8 shows poor background consistency. This is entirely consistent with the unconstrained nature of POEs.

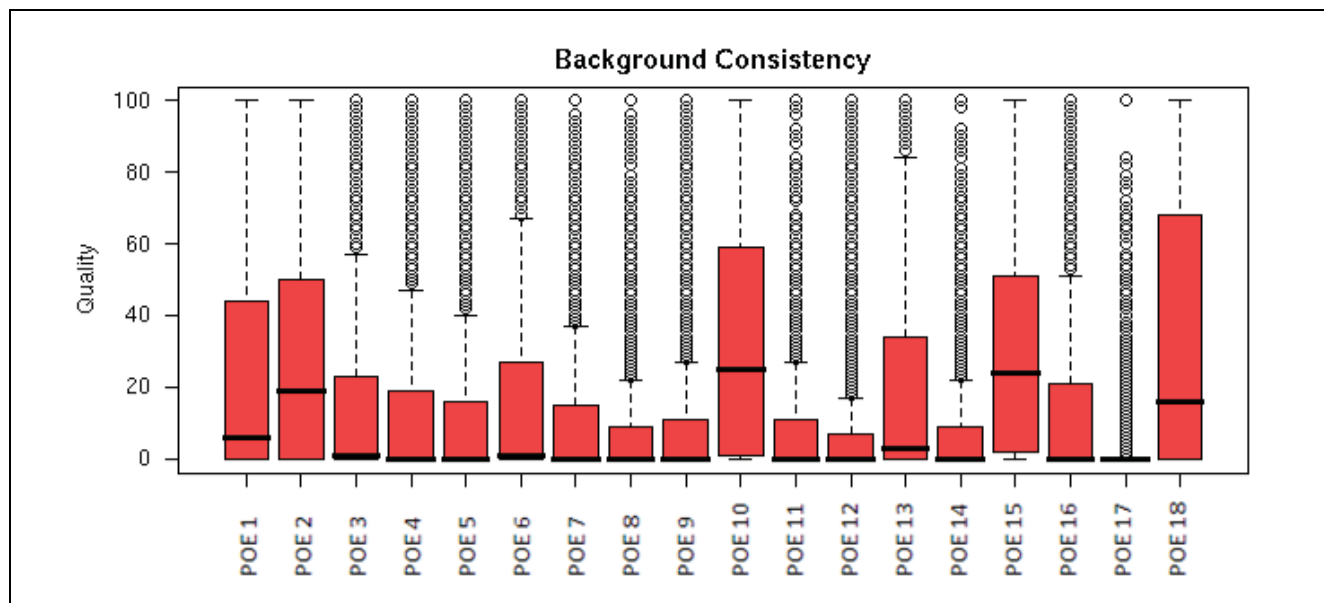


Figure 3-8: - Variation in the background consistency measure across POEs

Finally, Figure 3-9 shows that the yaw estimate is uniform across POEs. This is consistent with the known observation that many images are non-frontal and that the effect is largely a result of the precise officer-traveler interaction and synchronization. Interestingly there is a small negative bias toward the yaw angle. This would imply that subjects have a tendency toward one side or another.

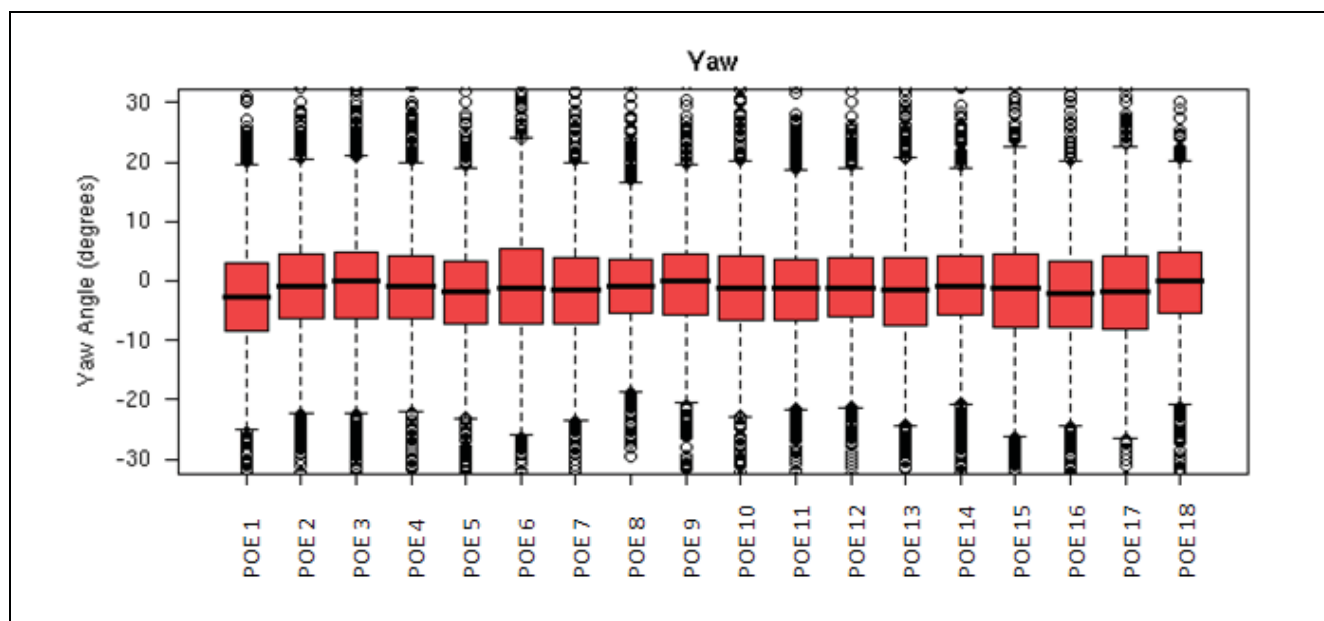


Figure 3-9: Variation in the yaw pose estimate across POEs

4 Test Phases


The FIQIFRS project included investigation of hardware and software approaches to facilitate acquiring compliant images in as expeditious a manner as possible and with minimal CBP Officer involvement. Prior to collecting images, a market/technology survey was conducted to identify available cameras and image quality software suitable for POE environments. Representative cameras and software products were selected for evaluation and testing on a volunteer population in a simulated POE environment under several different lighting conditions. A subset of these products was then integrated with image QA software.

The FIQIFRS application was developed in two phases. Phase I employed a custom interface to control and capture images from six different types of cameras. The images were analyzed retrospectively with QA software, and the cameras and QA software were down-selected for integration in Phase II. The Phase II application automated the image capture process by performing real-time QA on facial images (referred to as ‘quality-in-the-loop’). The FIQIFRS solutions were assessed with respect to quality improvement, performance, speed, operator training, and operator and traveler usability⁸. The following sections describe the test phases.

4.1 Camera Pre-Assessment

A market/technology survey was conducted to identify available cameras and sensors from various categories for use in POE environments. The WG identified desired features, determined product categories, identified and surveyed commercial-off-the-shelf (COTS) products in each category and reviewed their specifications, selected categories from which to draw products, and procured representative products from those categories. The camera categories examined were: webcam, digital still camera, digital camera in video mode, industrial video, PTZ video, wide dynamic range video, and smart camera. The cameras depicted in Table 4-1 were selected and evaluated objectively with respect to FACESTD compliance by imaging test targets (Figure 4-1) in an optimal laboratory environment and measuring their characteristics. Example face images were captured and assessed visually. Full details of this evaluation are documented in Attachment 2 [12] to this report.

Table 4-1: Cameras tested

Camera	Picture	Mount	Size (WxHxD, in.)	Connection	Software
Logitech QuickCam Pro 5000		Monitor clip; no threaded hole	2.5 x 2.5 x 2.5 (without clip)	USB	drivers

⁸ ISO 13407:1999 defines usability as “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use.”

Camera	Picture	Mount	Size (WxHxD, in.)	Connection	Software
Logitech QuickCam Pro 9000		Monitor clip; no threaded hole	3.5 x 1.5 x 1.5 (without clip)	USB	drivers
Logitech QuickCam Orbit AF		Flat base; no threaded hole	3.25 x 4.25 x 3.25	USB	drivers
Sony EVID70		threaded hole	5.25 x 5.75 x 5.75	S-video, requires capture card; power adaptor; VISCA RS-232 camera control	software for camera settings
Canon G9		threaded hole	4.19 x 2.83 x 1.67	USB; power adaptor or battery	SDK
Wide dynamic range	Not pictured	threaded hole	1.7 x 1.8 x 2.75	BNC connection, requires capture card; power adaptor, USB for camera settings	SDK

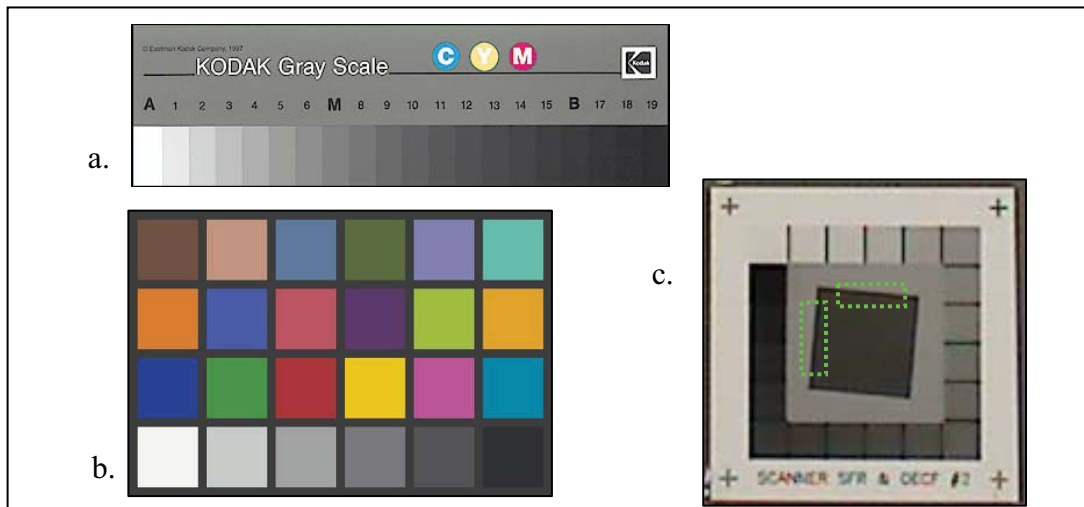


Figure 4-1: Image test patterns: a) Kodak Q13 grayscale test pattern; b) GretagMacbeth ColorChecker with reference map; c) ISO 16067-1 with slant edge regions of interest

4.2 Image Quality Assessment Software

A market survey of image quality software was conducted to identify desired features, identify and survey COTS products, and review product specifications. The following basic features were required of facial image QA software:

- Microsoft Windows XP environment.
- Integration with .Net based test software.
- Support C/C++ or Java or .Net C# programming languages.
- Integration with camera SDKs to capture frames either via native software or Microsoft DirectShow Framework.
- Process a single image as input in (near) real time and provide a method for offline batch processing of image files.

In order to be a useful tool for assessing the acceptability of input images, the values returned by the quality metrics should correlate with subjective perceptions of quality and/or with the accuracy of an FR system. The quality metrics should also measure the degree of compliance of face images with published face image interchange standards, such as ANSI INCITS 385-2004 and FACESTD. The face standard's specifications for the Full Frontal Face Image Type were used for the metrics, with inheritance of the requirements from the Basic and Frontal Image Types. The assignment of values to these metrics should reflect the degree of compliance to the normative requirements in the standard and/or the best practices in FACESTD, informative Annex A. The metrics that should be calculated by an image QA tool, and the relevant sections from FACESTD are:

- **Dynamic Range in Face** – intensity density in the facial region, should be at least 7 bits of intensity variation (at least 128 unique values) in face after conversion to grayscale (FACESTD 7.4.2).
- **Eyes Closed/Obstructed** – measured as a percentage, value should reflect degree of obstruction of eyes due to eyeglass rims, tint, or glare, bangs, eye patches, head clothing, or eyes closed (e.g., 100 percent obstructed if both eyes are closed; 50 percent obstructed if one eye is obstructed) (FACESTD 7.2.3, 7.2.11).
- **Color Balance** – must reflect natural colors with respect to expected skin tones. This value can be affected by inappropriate white balancing or red-eye (FACESTD 7.3.4).
- **Lighting Uniformity on Face** – measured as a percentage, value should measure symmetry as affected by shadows or hot spots on the face (FACESTD 7.2.7-7.2.10).
- **Background Uniformity** – measured as a percentage, value should measure symmetry and consistency as affected by shadows on the background, textured backdrops, or extraneous objects in the background. (FACESTD A.2.4.3).
- **Head Size** – head width to image width ratio should be between 5:7 and 1:2 (FACESTD 8.3.4, 8.3.5, A.3.2.2).
- **Centering** – horizontal and vertical position of face (FACESTD 8.3.2, 8.3.3).
- **Distance Between Eyes** – measured in pixels, should be at least 90 pixels between eyes (FACESTD 8.4.1) and preferably 120 pixels (FACESTD A.3.1.1).

- **Focus** – measures sharpness and resolution of the facial area. Depth of focus must maintain at least 2 mm per-pixel-resolution and preferably 1 mm per-pixel throughout the face. Image should not be overly sharpened. (FACESTD 7.3.3, A.2.5).
- **Rotation (yaw)** – value should measure deviation from frontal in degrees, compliance requirement is $<\pm 5^\circ$ (FACESTD 7.2.2).
- **Tilt (roll)** – value should measure deviation from frontal in degrees, compliance requirement is $<\pm 5^\circ$ (FACESTD 7.2.2).
- **Confidence in Face** – measures the confidence of the eye finding and the confidence that the object is a face.
- **Brightness Exposure/ Contrast:**
 - low score if too dark or too bright, exposure measured in RGB values.
 - gradations in skin texture should be visible, no saturation on the face (FACESTD 7.3.2).

Although there is currently no standard or device certification for face image capture devices, such a standard does exist for fingerprint capture devices. Electronic Biometric Transmission Specification (EBTS) Appendix F⁹ specifies that a fingerprint scanner “must be capable of producing images that exhibit good geometric fidelity, sharpness, detail rendition, gray-level uniformity, and gray-scale dynamic range, with low noise characteristics.” One of the objectives of the FIQIFRS project was to identify equivalent quality metrics for face images and determine how they could be measured. Use of an overall image quality measure, or selection of a subset of metrics to apply to facial images, should deliver standards-compliant images.

Three COTS image QA software products were identified for the project: Aware PreFace, CryptoMetrics VisProAnalyzer, and Cognitec FaceVACS-SDK. Two of the COTS products were used to analyze images for POE baseline image quality and for offline analysis of images collected during Phase I testing. For Phase II, a single QA product was identified for integration with the selected cameras.

4.3 Phase I Testing

During the first phase of FIQIFRS testing, facial images of a small volunteer population (13 individuals) were captured with the six cameras identified in Table 4-1 under the distinct lighting conditions described in Table 4-2.

Table 4-2: Phase I Lighting Scenarios

Imaging Scenario Description	Description of Simulation
Ideal	Capture of FACESTD compliant face images with digital still Canon G9 camera; no flash; two 500 Watt incandescent diffused lamps positioned at approximately 45° to the camera-to-subject line; plain background
Ambient	Overhead fluorescent lights on

⁹ Appendix F, IAFIS Image Quality Specifications, of the Electronic Biometric Transmission Specification (EBTS), Ver. 8, 9/24/2007, prepared by FBI.

Overhead	Overhead fluorescent lights on; shop light immediately above subject turned on
Dim (Natural Light)	Overhead fluorescent lights turned off; blinds ¾ closed
Side Lighting	Overhead fluorescent lights on; side lighting simulated with one 500 Watt incandescent diffused lamp situated to the right of subject
Back Lighting	Blinds opened; overhead fluorescent lights on; cameras/inspection station moved so that windows are behind the subject

Three test stations were set up: 1) volunteer registration; 2) capture of reference images under ideal lighting conditions; and, 3) simulated inspection station.

4.3.1 Test Setup

All cameras except the prototype wide dynamic range camera were mounted at a height of 60 inches (152 cm) on desktop tripods. The wide dynamic range camera was installed in a custom-built housing provided by Vendor C, which included a processor for performing real-time image QA. This “smart camera” system was positioned so that the camera lens was 60 inches high. Tape markings (in lieu of footprints¹⁰) were placed on the floor at a distance of 28 inches (70 cm) from the camera to position volunteers at a fixed distance from the cameras. A camera-to-subject distance of 70 centimeters was chosen as it represents the minimum distance specified in FACESTD (sub-clause B.2.2.2) and most closely reflects the capture conditions and lane widths at POEs.

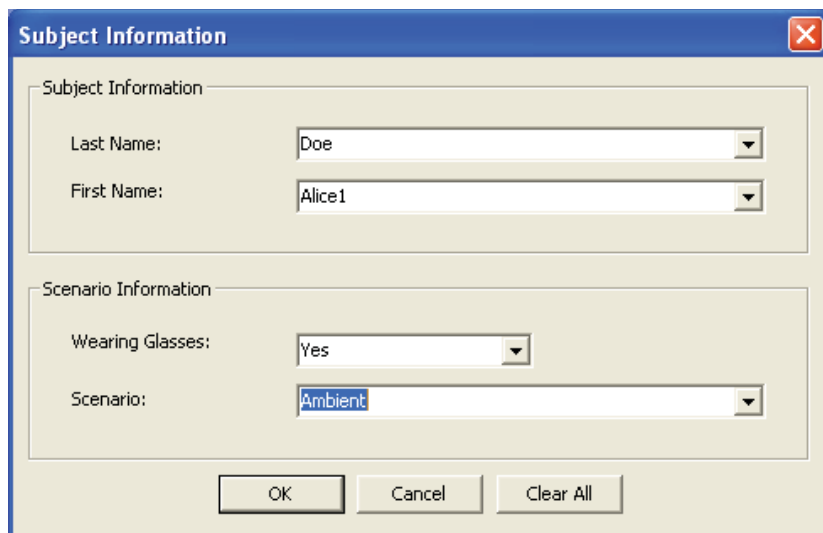
The FIQIFRS software was installed on government furnished equipment (GFE) at the simulated inspection station, and all cameras were connected to, and operated from, one workstation for Phase I testing. The Canon camera was installed and programmed to collect photographs in portrait mode (where the longest dimension is the height as opposed to the width); all other cameras were installed to capture photographs in their default landscape mode.

The webcams, Sony pan-tilt-zoom (PTZ) video camera, and the wide dynamic range camera were operated from a custom-built application called WebCamDemo. The Canon camera was operated from a modified version of the Canon sample program, RelCtrl.

4.3.2 Test Administration

Testing was administered by members of the Technical WG, who simulated the border inspection interview and facial image capture process. A reference image was captured with a Canon G9 camera against a plain background as described for the Ideal scenario in Table 4-2. For the reference image, the Canon was set to manual mode, with 1/60 second shutter speed, F5.6 aperture, and 200 ISO. For the mock inspection, volunteers participated as ‘travelers’ going through primary inspection and biometric facial image capture. The test administrator selected the camera from the application, and entered information about the test subject and the test scenario using the simple GUI shown in Figure 4-2.

¹⁰ Ref [16], NISTIR 7540, Assessing Face Acquisition.



The screenshot shows a 'Subject Information' dialog box. It has a blue title bar with a close button. The dialog is divided into two sections: 'Subject Information' and 'Scenario Information'. In the 'Subject Information' section, there are two dropdown menus: 'Last Name' with the value 'Doe' and 'First Name' with the value 'Alice1'. In the 'Scenario Information' section, there are two dropdown menus: 'Wearing Glasses' with the value 'Yes' and 'Scenario' with the value 'Ambient'. At the bottom of the dialog, there are three buttons: 'OK', 'Cancel', and 'Clear All'.

Figure 4-2: Phase I GUI

After asking some inspection interview questions, the test administrator instructed the volunteer to position him or herself for image capture, and initiated ten (10) seconds of image capture. After approximately two to three seconds, the test administrator continued the interview while the volunteer subject remained in place. Throughout testing, the position and location of the cameras remained fixed (operators did not move the cameras for different subject heights).

The test sessions were conducted by lighting scenario over a two-day period. Images were collected from all cameras under the ambient, bright overhead and dim (natural) scenarios on Day 1; and cameras and lighting were moved for the second day of testing under the side lighting and back lighting scenarios. For each test event, the test scenario, camera, camera exposure settings, and subject identifier were saved in an eXtensible Markup Language (XML) file associated with the image data. On completion of testing, the data was encrypted and transferred to an external hard drive for analysis. The images were analyzed retrospectively to determine the feasibility of using automated QA software to select standards-compliant images from video sequences.

4.4 Phase II

Phase II testing involved integration of an image QA SDK into the FIQIFRS application for real-time image quality analysis, and a down selection of cameras for testing. Two cameras were selected for integration with the image QA SDK based on results of the camera pre-assessment and Phase I analysis. The Phase I analysis showed little difference between images collected under the ambient and overhead lighting conditions, so the overhead scenario was not included in Phase II. The back lighting scenario generated a high percentage of failures to enroll; therefore the scenario was tested on only two participants in Phase II. Image collection with quality-in-the-loop was conducted on all volunteers under the ideal, ambient, dim, and side lighting scenarios described in Table 4-2. As with Phase I, testing was conducted in the simulated POE lab, using the test stations for volunteer registration and signing of consent forms, capture of reference images, and simulated inspection station.

4.4.1 Test Setup

The Canon G9 and Logitech QuickCam Pro 9000 were selected for integration with image QA software for Phase II testing. Both cameras were installed on tripods at a height of 60 inches; and both were operated in portrait mode for Phase II testing (the G9 was also used in portrait mode for Phase I testing). As in Phase I, tape markings were placed on the floor to position volunteers at a fixed distance of 28 inches (70 cm) from the camera. The FIQIFRS Phase II application integrated the Canon SDK, webcam drivers, and the image QA SDK. Quality metrics and thresholds for this quality-in-the-loop application were selected based on analysis of Phase I image data. Settings for the different lighting scenarios were changed programmatically for the Canon camera. To simplify and speed up the test process, two GFE workstations were used in Phase II testing; the Canon camera was connected to one GFE and the webcam to the other. Each test session involved thirty seconds of image capture at a resolution of 1200 x 1600 pixels.

Volunteers for testing were solicited from US-VISIT government and contractor personnel. Volunteers were provided with a consent form describing the test process, and were instructed to bring their signed consent form on the day of testing.

4.4.2 Test Administration

Testing was administered by members of the Technical WG, who simulated the border inspection interview and facial image capture process. Volunteers were first directed to the registration station, where the consent form was collected, and an identifier was assigned to the subject. Demographic information (height, eye color, presence of glasses) was collected by the test administrator, and volunteers were then sent to the reference image station for collection of the enrollment image.

Test volunteers participated as travelers going through the U.S. border inspection process with biometric facial image capture. The FIQIFRS application analyzed facial images in real time, and displayed up to four cropped token images that passed the image QA thresholds to the test administrator. If no image met the thresholds, the test administrator captured a single manual snapshot. Test administrators did not request volunteers to remove their glasses; however, on some occasions, if none of the images met the thresholds, the administrators captured a manual snapshot, asked the volunteer to remove their glasses, and then re-initiated the image capture session. Volunteers remained in place, looking at the camera, for the duration of each test scenario (combination of lighting condition and camera).

The Phase II application GUI is depicted in Figure 4-3. The large window in the upper left displays a real-time image of the capture scene. The four smaller windows below show the token faces for the image frames that passed the quality test. Real-time image QA started when the test administrator clicked on the "Capture" button. If a particular frame failed the quality test, text was displayed above the token images describing the specific cause of failure (e.g. "Face not Frontal"). The test administrator could then use this information to provide appropriate instruction to the participant. For an image to be acquired and displayed to the user, it had to pass the quality test. Capture terminated when either four images were acquired, or at a thirty-second timeout. Upon completion of real-time QA, the test administrator was prompted to click on the token face that appeared to be the best quality. If no token images were available for selection, the test administrator captured a manual snapshot by clicking on the "Single Shot" button. Afterward, a

pop-up window appeared to let the test administrator know that capture was complete for the given participant and lighting scenario.

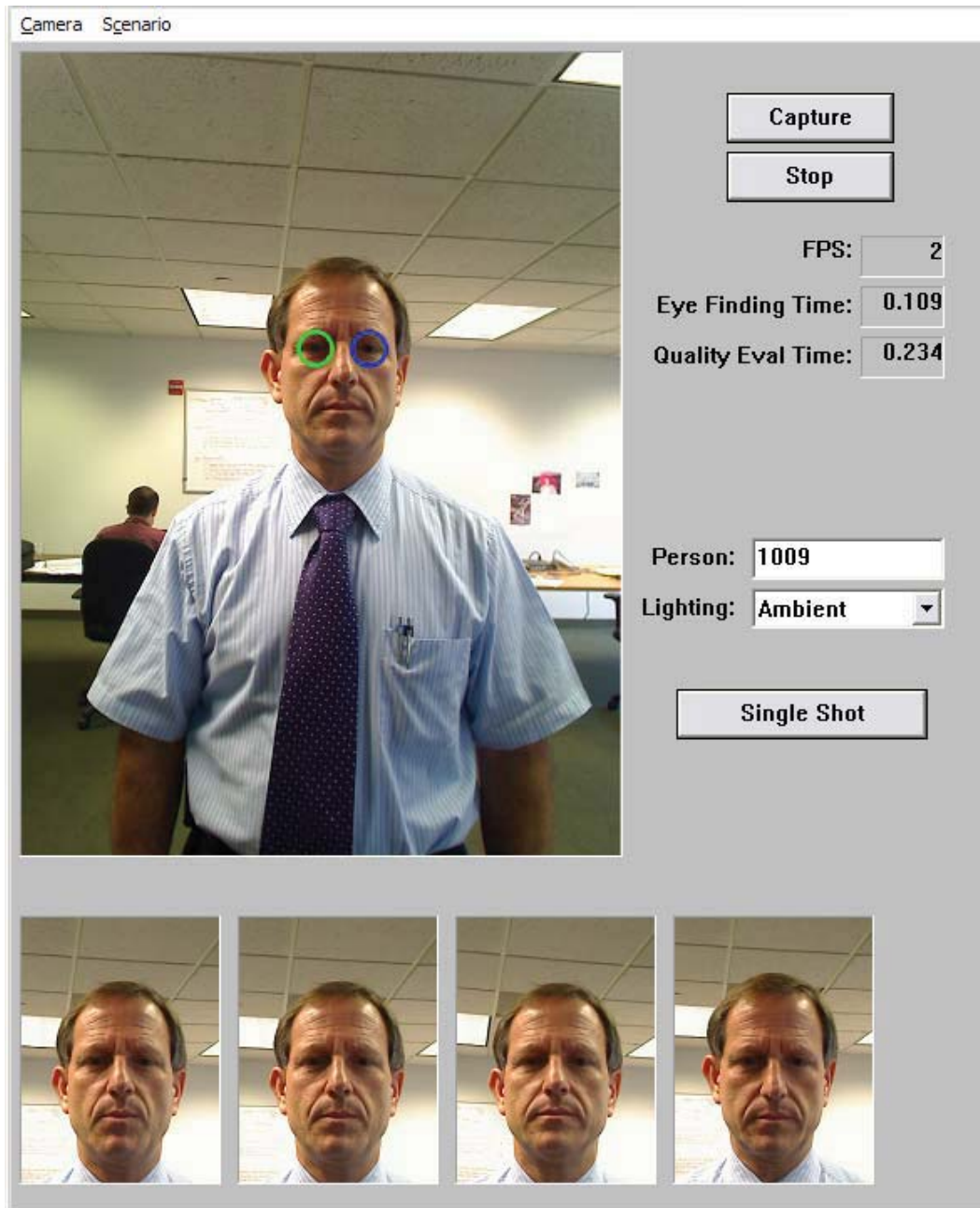


Figure 4-3: Phase II FIQIFRS Application GUI

Testing was conducted over a two-week period. The test scenario, camera, camera exposure settings, and subject identifier were saved in an XML file associated with the image data and log files. At the end of each week, all collected data was encrypted and stored on an external hard drive for later analysis. At the end of the collection period, all participant information and commercial software was removed from the computers used for collection.

After testing was complete, the ability of quality-in-the-loop to improve the capture process was evaluated in the following ways:

- 1) Face image quality: Quality-in-the-loop should improve the quality of captured face images. Since higher quality images are expected to perform better when matched, a commercial FR algorithm was used to quantify the improvement offered by quality-in-the-loop over the traditional capture method. Operational POE images were also matched for reference.
- 2) Capture time: Quality-in-the-loop should not substantially increase the image capture time. Timing information was recorded for each capture session. In addition to determining how long it took to acquire good quality images, the timing information was used to measure the speed at which the software performed eye finding and QA.
- 3) Accuracy of quality software: Images were manually inspected to determine when and how often the QA software performed an incorrect assessment of a face image. Two types of errors are possible: 1) a *false rejection*, when the QA software rejects a good-quality image, and 2) a *false acceptance*, when the QA software accepts a bad-quality image.

5 Results

As described in Section 4, the FIQIFRS project examined the characteristics of several classes of COTS cameras, collected facial images with various cameras in a simulated POE environment under distinct lighting conditions, and then down-selected to a smaller number of cameras for integration with image QA software. Through the course of the FIQIFRS project, face images of volunteers were captured and analyzed in two phases. This section describes the results of the evaluation of the cameras, QA software, and images captured during testing. The following test metrics were collected and reported for the project:

- Image conformance to FACESTD.
- Image quality metrics for each image analyzed.
- Inter eye distance of at least 90 pixels.
- Error rates (e.g., failure to acquire).
- Time to calculate image quality metrics for each image.
- Overall processing time (image capture plus image QA).
- Ease of use of image capture/camera control interface.
- Range of subject heights captured without moving the camera .

5.1 Camera Pre-Assessment

The cameras depicted in Table 4-1 were evaluated for a number of factors that impact image quality and usability in a POE. The size of the camera, as well as the type of mount and connections affect a camera's suitability to the POE environment. The size of the camera should be compact, so as not to obscure the CBP Officer's view of the traveler. Because the space in POE lanes is limited, it is desirable to minimize the number of camera connector cables and external hardware required to operate the camera. Another important camera characteristic is the provision, by its manufacturer, of software for image capture and for setting camera parameters (e.g., exposure, zoom). Sample images captured with the cameras and camera characteristics are described below.

Table 5-1 summarizes the capture dimensions, frame rates, compression ratios, and fields of view for each camera, with the highest values in green and the lowest in red. The Canon G9 allowed for the highest capture dimensions (12 mega-pixels) and inter-eye distance. The Logitech QuickCam Pro 9000 and Orbit AF had the next highest capture dimensions (2 mega-pixels) and an inter-eye distance that complies with the FACESTD required inter-eye distance of 90 pixels¹¹.

¹¹ FACESTD, 8.4.1 – Resolution (Normative)

Table 5-1: Capture Dimensions, Frame Rate, Compression, and Field of View

Camera	Capture Dimensions (px., WxH)	Inter-eye distance (px.)	Frame Rate (frames per sec.)	Com-pression	Sampling Frequency (mm/px.)	Field of View Size (in., WxH)	Field of View Area (in.2)	Head Lengths
Canon Powershot G9	4000 x 3000*	325	N/A	Normal, Fine, Super-Fine, RAW	0.2	31.5 x 23.6*	744	3.5*
	3264 x 2448*	258			0.24	30.7 x 23*	706.1	3.4*
	2592 x 1944*	207			0.3	30.6 x 22.9*	700.7	3.4*
	1600 x 1200*	127			0.49	30.9 x 23.2*	716.9	3.4*
	640 x 480*	51			1.23	31 x 23.2*	719.2	3.4*
Sony EVID70	640 x 480	58	30	~4:1	0.98	24.7 x 18.5	457.3	2.1
Logitech Quickcam Pro 5000	640 x 480	55	30	~11:1	1.04	26.2 x 19.6	514.8	2.2
Logitech Quickcam Pro 9000	1600 x 1200*	110	30	~13:1	0.56	35.3 x 26.5*	933.3	3.9*
Logitech Quickcam Orbit AF	1600 x 1200	111	30	~15:1	0.52	32.7 x 24.6	804.7	2.7
Prototype wide dynamic range camera	640 x 480	82	30	~3.7:1	0.687	17.3 x 13	224.8	1.4

Example images of a human face were captured from the selected cameras in an ideal test capture environment. Sample images captured with the Canon G9 (Figure 5-1) and the Logitech QuickCam Pro 5000 and 9000 (Figure 5-2a-b) are shown below.

♦ When operated in portrait mode

* Can also be operated in portrait mode, in which case the width and height values are reversed.

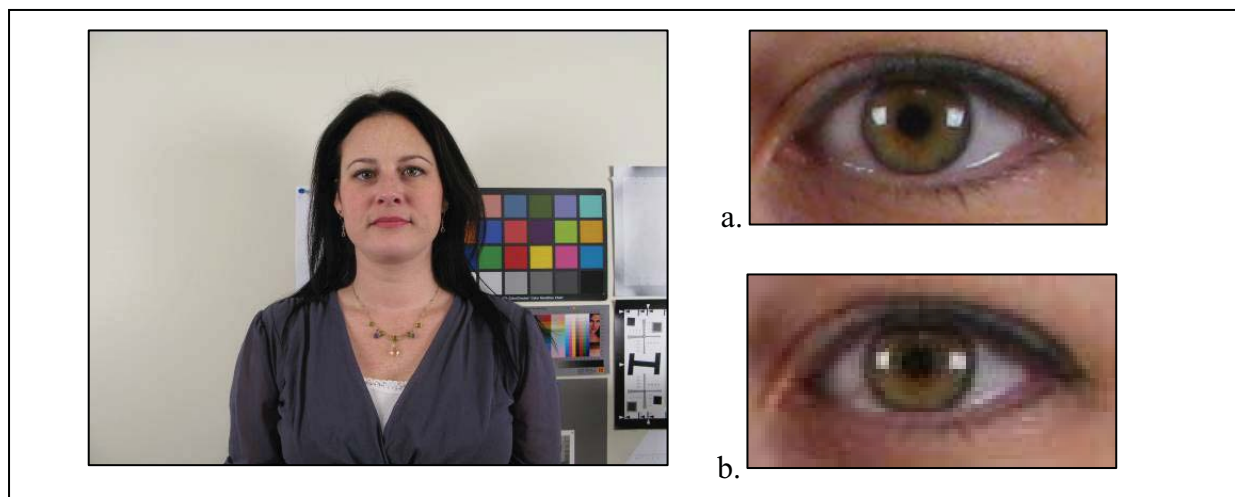


Figure 5-1: CanonG9 image with enlarged eye captured at (a) 4000 x 3000 pixels and (b) 1600 x 1200 pixels

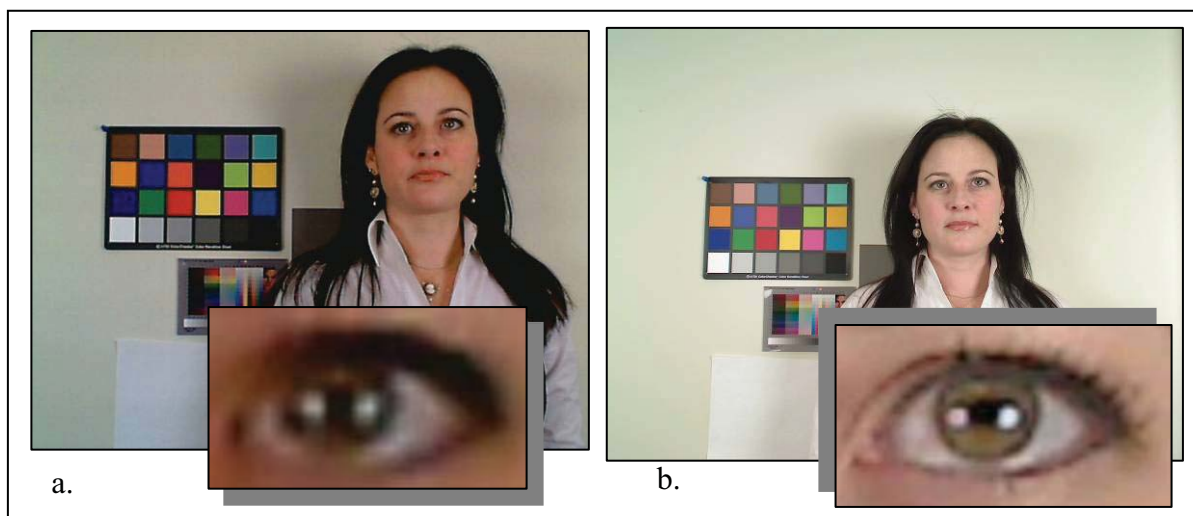


Figure 5-2: Logitech QuickCam Pro 5000 (a) and 9000 (b) images

The full camera pre-assessment report, which describes testing for these factors as well as tests to measure the cameras' geometric accuracy, spatial uniformity, depth of field, tonal response, color accuracy, and spatial resolution, is available as Attachment 2 to this report.

5.2 Phase I

During the first phase of the FIQIFRS testing, still images and video of a small volunteer population were captured with the six cameras listed in Table 4-1 in a simulated POE environment under six distinct lighting conditions (ideal, ambient, dim, side lighting, back lighting, and overhead lighting) using a custom-developed interface. Testing was conducted as explained in Section 4.2.

5.2.1 Image Collection

Each “event” (i.e., combination of subject, camera, and lighting scenario) consisted of 10 seconds of video (at a minimum of five frames per second) or multiple still images (for the digital still camera). A total of approximately 30,000 images were captured during Phase I, examples of which are shown in Table 5-2 for each camera and scenario.

In addition to these simulated POE images, an ISO-compliant reference, or *control*, image of each subject was captured in an ideal environment to be used as the mate against which the test images were matched with FR (see example reference images in Figure 5-3).

There were 13 participants; nine (9) males and four (4) females. Of the participants, nine (70%) were Caucasian, two (15%) were African American, and two (15%) were Asian. The self reported height ranged from 63 inches (160 cm) to 76 inches (193 cm).

5.2.2 Automated Quality Assessment
















The images captured during Phase I testing were analyzed offline with automated face image QA products. The performance of the products’ individual quality metrics (e.g., pose angle, eyes open) were determined by manual inspection and histogram analysis. Those metrics that were deemed to correctly predict standards compliance and correlate with human perception were selected and thresholds for the metrics were established (by visual inspection) for incorporation into the Phase II FIQIFRS quality-in-the-loop application. This set of metrics and thresholds served as the criteria for automatically selecting a compliant image from an image capture sequence in Phase II.

5.2.2.1 Comparison of Commercial Quality Assessment Products

As noted in Section 4.2, three COTS QA software products were considered for evaluation in this test phase. The metrics, range of values, and vendor-recommended thresholds provided by each vendor were compared. Unfortunately, it is difficult to compare the accuracy of these products due to the lack of standardization, as each product measures different quality factors over different value ranges. Published metrics and factors for face images with respect to compliance with the ANSI INCITS 385-2004 or FACESTD standards are needed.

Table 5-3 contains a checklist indicating the presence of the WG’s desired metrics, which were selected from the list in Section 4.2. It should be noted that metrics pertaining to illumination and background were not employed, because neither the lighting nor the background in the POE environment could be changed.

Table 5-2: Example images captured during Phase I

	Ambient	Dim	Side-lit	Overhead	Back-lit
Logitech QuickCam Pro 5000					
Logitech QuickCam Pro 9000					
Logitech QuickCam Orbit AF					
















	Ambient	Dim	Side-lit	Overhead	Back-lit
Sony EVID70					
Canon G9					
Prototype Wide dynamic Range Camera					



Figure 5-3: Reference images

Table 5-3: Presence of desired metrics in commercial image QA products

Desired Metric	Vendor A	Vendor B	Vendor C
<i>Dynamic Range in Face</i>	✓	✓	
<i>Eyes Closed/Obstructed</i>	✓		✓
<i>Color balance</i>	✓		✓
<i>Focus</i>	✓	✓	✓
<i>Rotation (yaw)</i>	✓ ¹²	✓	✓
<i>Tilt (roll)</i>	✓	✓	✓
<i>Confidence in Face</i>	✓		✓
<i>Brightness/Contrast</i>	✓	✓	✓
<i>Overall Quality</i>	✓		

5.2.2.2 Performance of Commercial Image Quality Assessment Products

The SDKs from Vendor A and Vendor C were executed on the images collected during Phase I testing. Note that the aim of the project was to test and evaluate the concept of quality-in-the-loop using representative cameras and image QA software; it was not a thorough examination of all available cameras or image QA products. The two image QA SDKs used during testing were selected as representative products that measured the desired metrics.

One means for assessing the performance of quality metrics is to examine the distribution of the values they return. For most metrics that provide continuous values, a histogram of the values output for representative images should exhibit a smooth Gaussian-like distribution. Such a distribution would enable a fine-tuned selection of quality cutoff thresholds. Ideally, there would be a fairly clear separation between the values measured for low-quality and high-quality images.

¹² Not output in degrees.

5.2.2.3 Failure to Enroll

Failure to enroll (FTE), in this context, is defined as a case where the QA software could not find eyes and thus, did not evaluate quality metrics for an image. The percentages of images that Vendor C and Vendor A failed to enroll are shown in Figure 5-4 and Figure 5-5.

Vendor A enrolled more images overall (85%) than did Vendor C (72.3%). Back lighting caused the highest FTE rates for both Vendors, with the Canon and wide dynamic range being less sensitive to back lighting than the other cameras. Dim lighting (natural light from windows, with fluorescent lights turned off) had the lowest FTE rates with both Vendor C and Vendor A, perhaps due to the absence of artificial lights. Unexpectedly, side lighting did not adversely impact Vendor C’s enrollment rates, and overhead lighting improved enrollment rates over ambient lighting. With Vendor A, the Logitech webcams experienced the highest FTE rates (especially the Orbit), while the Canon had the lowest FTE rates.

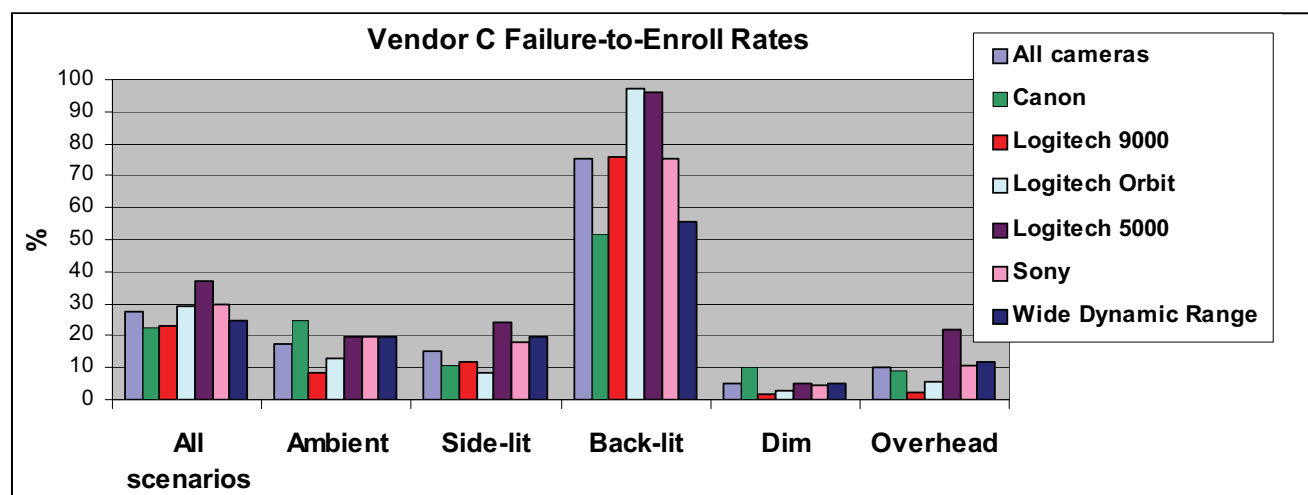


Figure 5-4: Vendor C failure-to-enroll rates

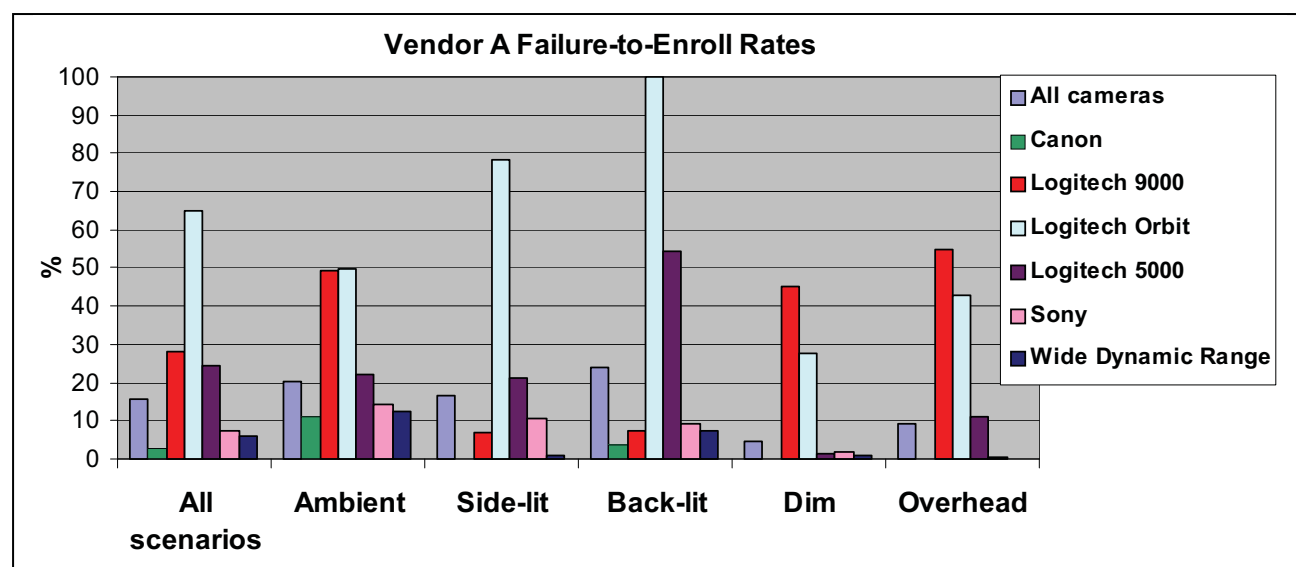


Figure 5-5: Vendor A failure-to-enroll rates

5.2.3 Conclusions and Down-select

The results of QA analysis (including FTE rates) and usability issues determined which cameras were down-selected for use in Phase II testing. During Phase I testing, volunteers who were 72 inches or taller had to crouch down to fit within the field of view for the wide dynamic range and Sony; the tallest participant (76 inches) also had to crouch to fit within the Logitech QuickCam Pro 5000's field of view. The chin of the 63-inch-tall participant was cropped in the wide dynamic range's field of view.

The Logitech QuickCam Orbit AF was eliminated due to its high FTE rates and focusing problems that were observed during testing. The Logitech QuickCam Pro 5000 was a discontinued model; it was only used for baseline comparisons because it was deployed at POEs. The wide dynamic range was eliminated due to its low resolution, small field of view, and manual focus. The Sony was eliminated due to the inability to take advantage of its PTZ functionality in addition to its low capture resolution. Therefore, the WG down-selected the cameras to the Canon G9 and Logitech QuickCam Pro 9000 for the next test phase.

As noted earlier, a single vendor's image QA product was selected for integration with the Canon and Logitech webcams.

5.2.4 Analysis of Phase I Image Quality Scores

Vendor A's QA values on the Phase I test set were analyzed to devise a quality test for use in Phase II testing.

5.2.4.1 Visual Determination

A web-based visualization software tool was created for viewing Phase I images, sortable by meta-data in ascending or descending order. The following metrics were selected to use during Phase II for quality-in-the-loop and thresholded based on visual assessment using the visualization tool: eyes open, eye gaze frontal, eyes tinted, pose angle roll, deviation from frontal pose, sharpness, mouth closed, eye height, and eye distance.

If an image scored below the threshold for any of the metrics, it failed the quality test. The objective of the quality test was to pass standards-compliant images, such as the one shown in Figure 5-6a, and to fail images with visible quality defects (e.g., out of focus, head tilting down), such as the one shown in Figure 5-6b.

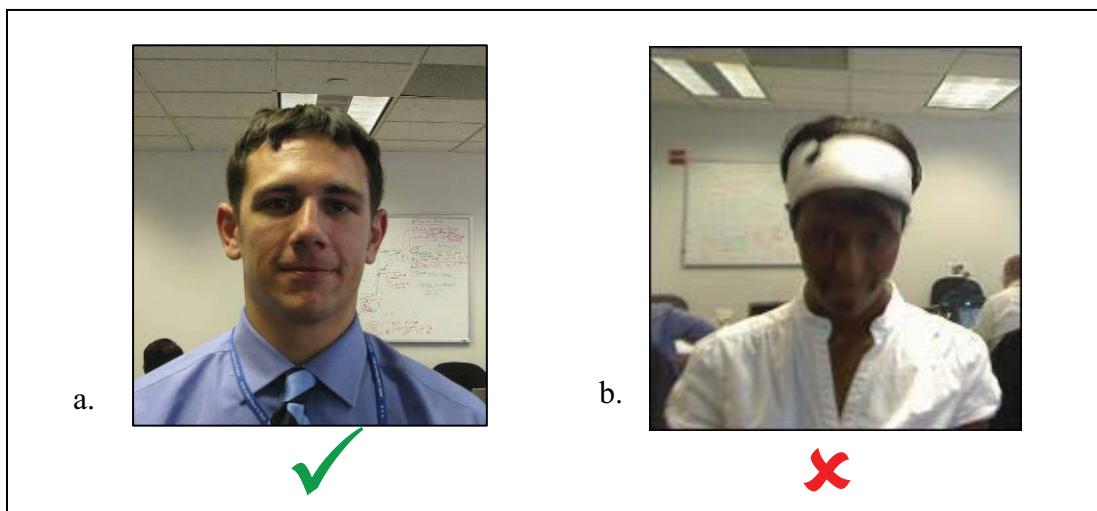


Figure 5-6: a) Compliant image that passes quality test; b) Non-compliant image that fails quality test

5.2.4.2 Image Quality as a Predictor of FR Performance

The FR scores of the Logitech 9000 and Canon test images compared to the reference images were used to augment the threshold analysis of the image QA software. A correlation analysis was conducted to determine which of the metrics (if any) predict FR score. The correlation¹³ between quality metric values and image-pair FR scores is a measure of the predictability of the quality metric for the FR score and can be calculated from analyses of sample measurements.

Multivariate linear regression was used to predict the software product's overall quality metric. Using correlation analysis, it was determined that the overall quality metric reported by the software did not adequately predict FR performance. As a result, the individual quality metrics were analyzed to determine if a subset would better predict FR performance.

A factor analysis was used to select a subset of the metrics, which were then used to develop a regression model for predicting FR performance. This model was refined to improve its performance for each lighting scenario. Finally, the overall and per lighting scenario models were evaluated. The main results are summarized as follows:

1. Figure 5-7 demonstrates that there was a statistically significant difference ($p < 0.0001$) in FR performance between the two tested cameras across all lighting scenarios. FR scores were higher for the Canon than for the Logitech 9000.

¹³ Correlation is a measure of the degree of the relationship between two variables. Correlation coefficients may range from negative 1 to positive 1. Values ranging from -0.1 to +0.1 are often referred to as trivial or no correlation (i.e., knowing one variable does not assist in predicting the other variable). A positive correlation coefficient indicates that the dependent variable increases as the independent variable increases. A negative correlation indicates that the dependent variable decreases as the independent variable increases. Correlation coefficients above 0.5 or below -0.5 are usually considered to represent meaningful correlation. In part because the correlation coefficient can be either negative or positive, the square of the coefficient – R^2 – is often reported instead.

2. Figure 5-8 demonstrates that there was a statistically significant difference ($p < 0.0001$) in FR performance for each of the lighting scenarios, with dim (or natural light) performing best, followed by overhead, ambient, sidelight, and backlight.
3. Figure 5-9 and Figure 5-10 show the FR performance of each camera across the five lighting scenarios. Note that the Canon performed significantly ($p < 0.0001$) different in three of the five scenarios, while the Logitech performed significantly ($p < 0.0001$) different across all five scenarios.
4. Figure 5-11 shows the ability of the model for the Canon camera to predict FR performance across all lighting scenarios, while Figure 5-12 shows that, for the ambient lighting scenario, it is possible to predict FR with very high accuracy ($rSq = 0.9$).
5. Figure 5-13 shows the significant result that using the Canon camera under the ambient lighting scenario, it is possible to predict FR performance with a correlation of 0.9.

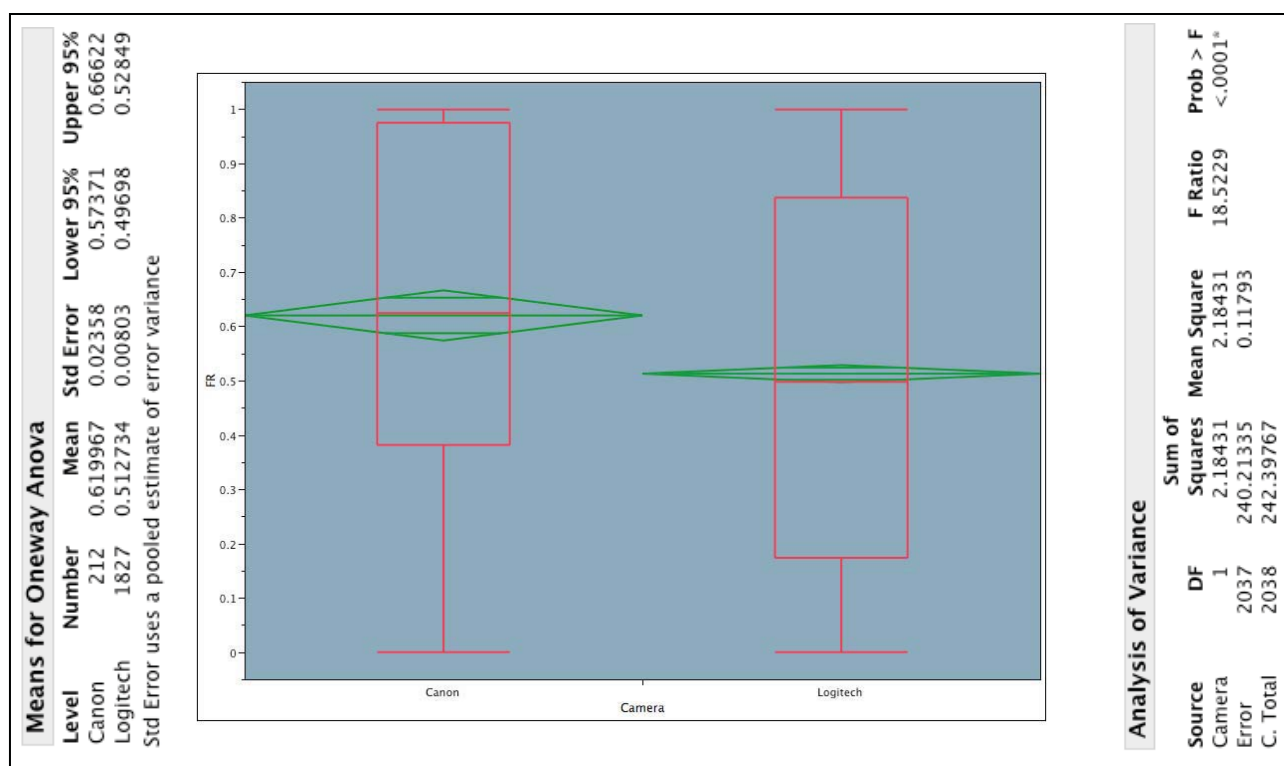


Figure 5-7: FR scores by camera; demonstrates a statistically significant difference ($p < 0.0001$) in FR performance between the two tested cameras across lighting scenarios.

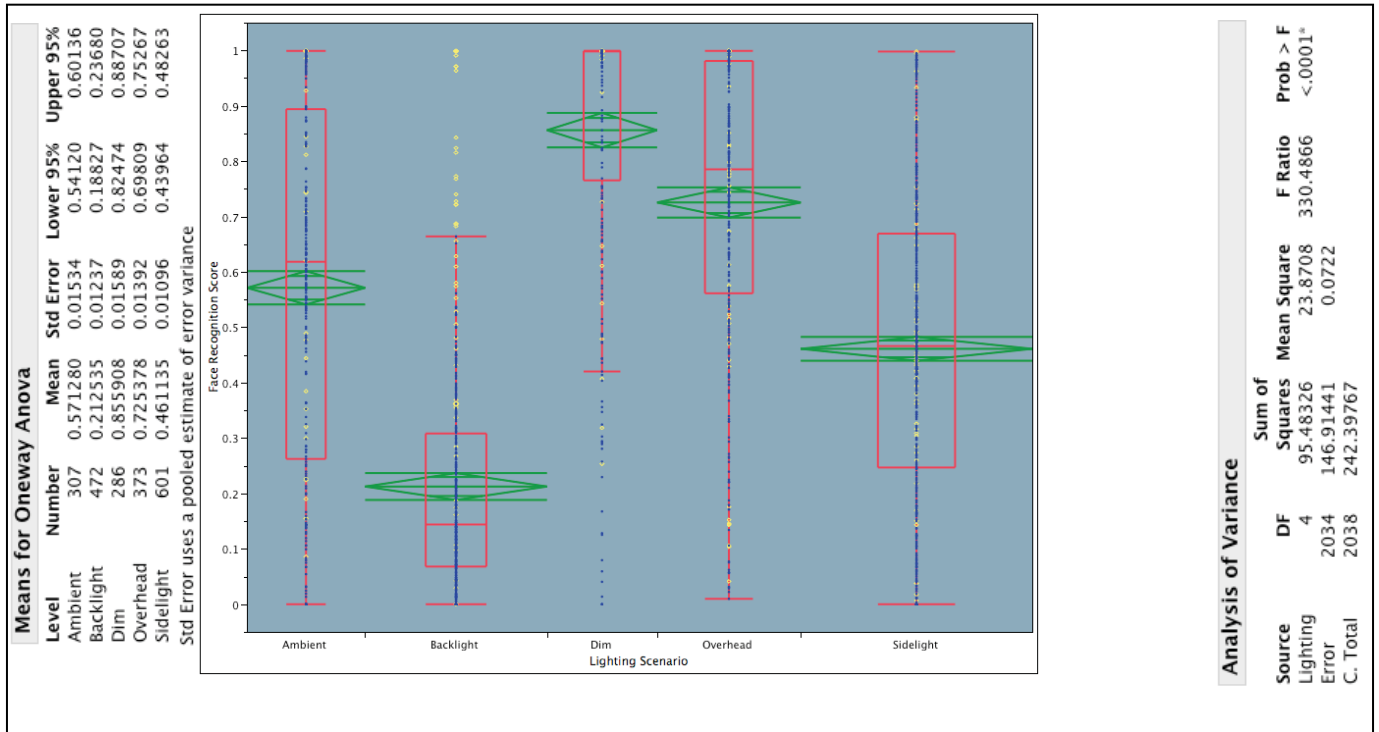


Figure 5-8: Overall (Canon and Logitech 9000) FR by lighting scenario; statistically significant difference ($p < 0.0001$) in performance ($p < 0.0001$) in performance for each of the lighting scenarios

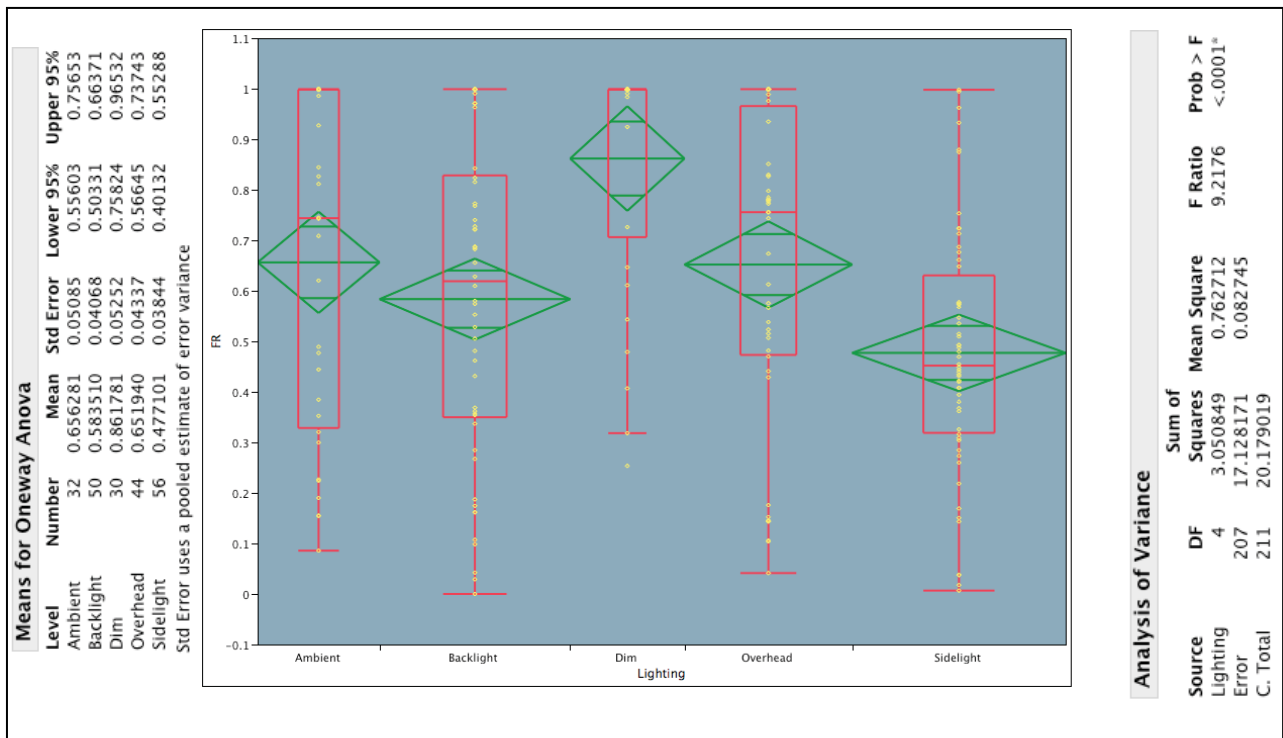


Figure 5-9: Canon FR scores by lighting scenario

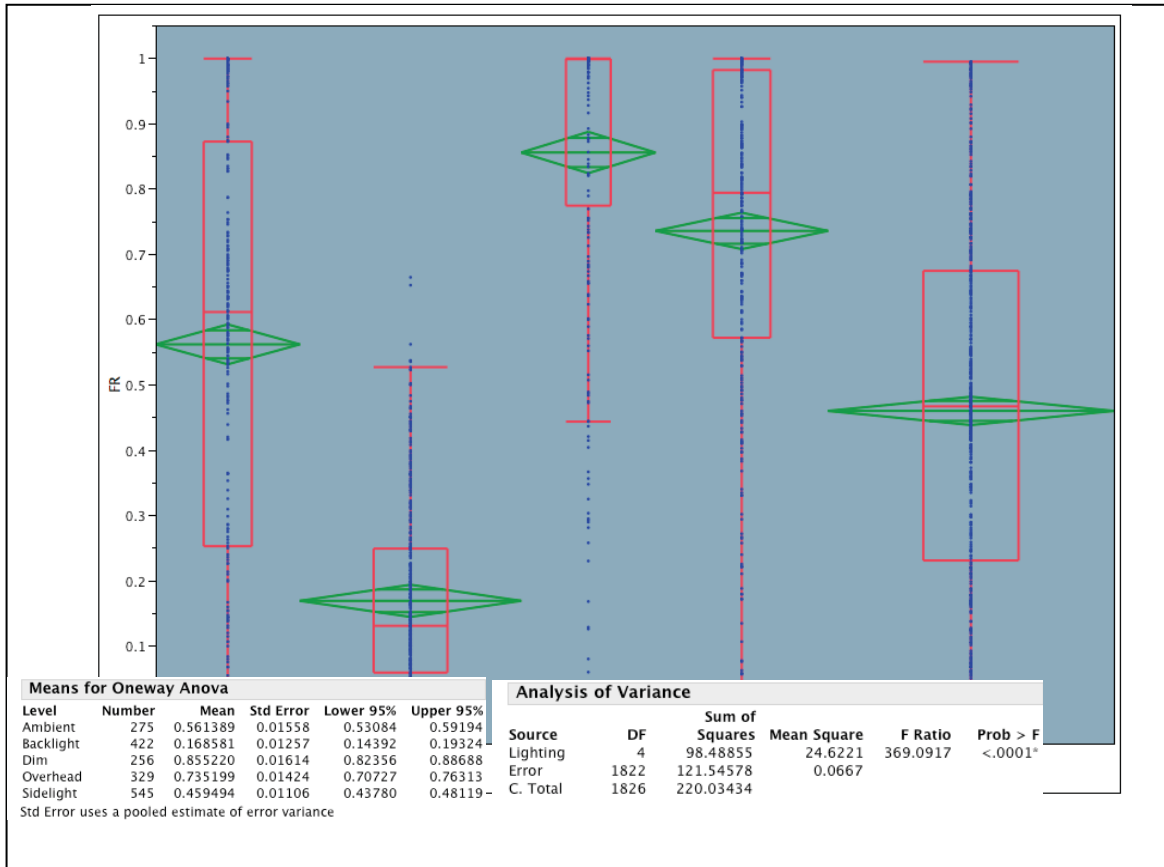


Figure 5-10: Logitech 9000 FR scores by lighting scenario

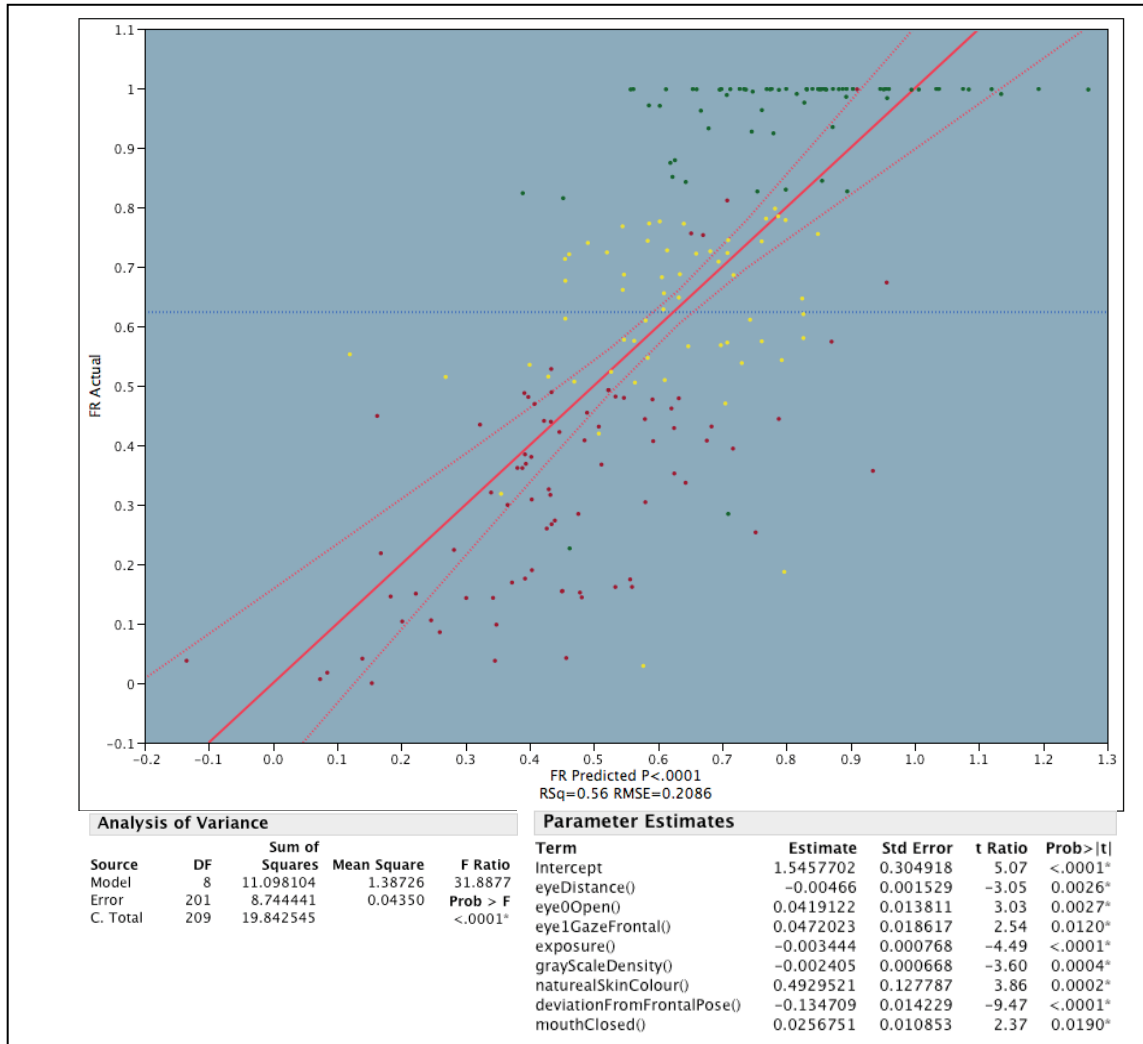


Figure 5-11: Predicting FR from quality metrics (Canon, all lighting scenarios); illustrates the ability of linear model for Canon to predict FR performance across all lighting scenarios.

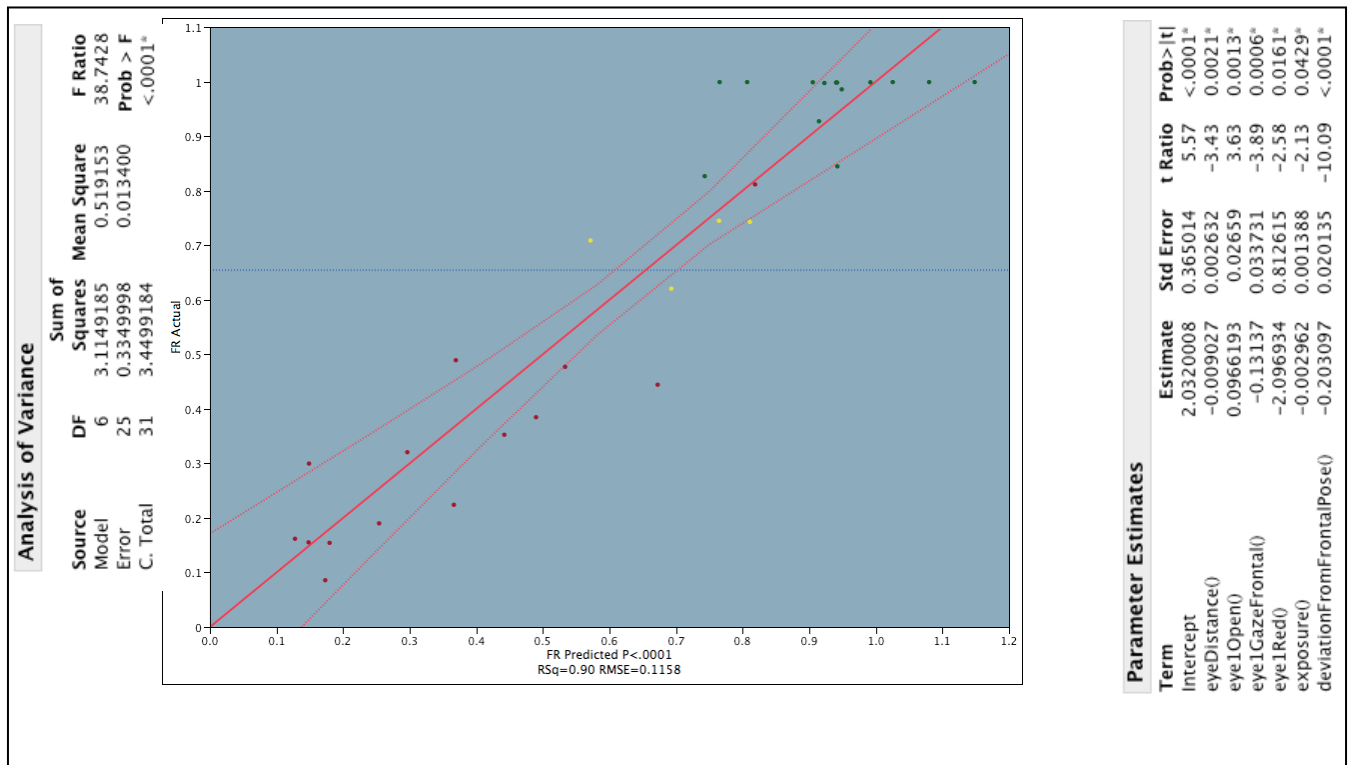


Figure 5-12: Predicting FR from quality metrics (Canon, ambient); linear model predicts FR performance with very high degree of accuracy (rSq = 0.9).

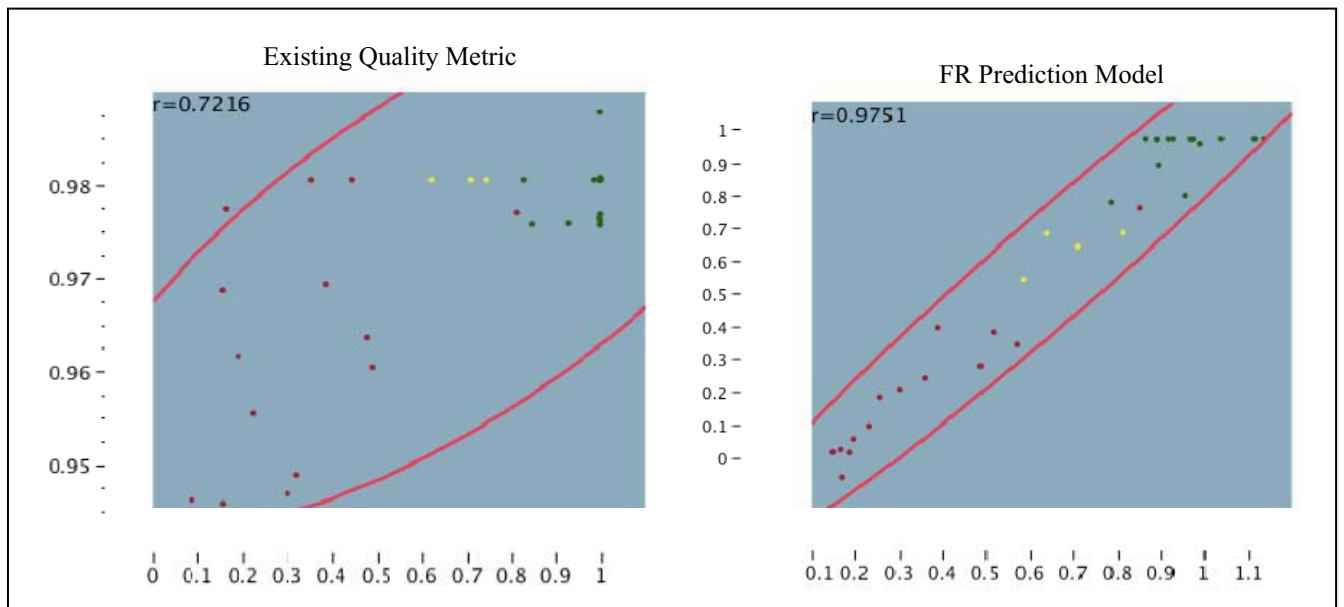


Figure 5-13: FR prediction model vs. quality metric (Canon, ambient)

5.3 Phase II

During Phase II of FIQIFRS testing, facial images of 63 volunteers were captured with the down-selected cameras (Logitech 9000 and Canon G9) in four simulated POE lighting scenarios (ideal, ambient, dim, and side lighting). The FIQIFRS application was modified to incorporate automated QA software in-the-loop to select standards-compliant images from video sequences. In each test case (subject/camera/lighting scenario combination), each frame captured by the camera was evaluated by the quality test described in Section 5.2.4.1 in real time. The process terminated when four images passed the quality test or after a 30 second timeout. If no images passed the quality test, the operator captured a manual snapshot, similar to current operations.

5.3.1 Image Collection

A total of 6,442 images were captured during Phase II, examples of which are shown in Table 5-4 for each camera and scenario. In addition to the simulated POE images, an ISO-compliant reference, or *control*, image of each subject was captured in an ideal environment to be used as the mate against which the test images are matched with FR (see example reference images in Figure 5-3).

The participants consisted of 63 adults. There were 39 males and 24 females. Of the 63 participants, 47 (75%) were Caucasian, 6 (10%) were African American, 3 (5%) were Asian, and 7 (11%) were of other ethnicities. The self-reported height of the participants ranged from 60 inches (152 cm) to 75 inches (191 cm). The eye distances (in pixels) are shown in Table 5-5.

Table 5-4: Example images captured during Phase II




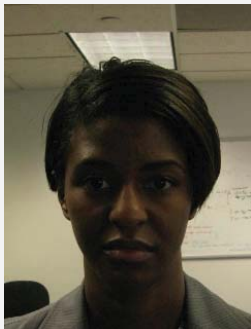


	Ambient	Dim	Side-lit
Logitech QuickCam Pro 9000			
Canon G9			

Table 5-5: Eye distances

Camera	Average	Minimum	Maximum
Canon G9	133.4	109.1	166.8
Logitech 9000	110	90.7	139.9

5.3.2 Acquisition Rates

The capture process should take as little time as possible to maintain current US-VISIT throughput levels. While the capture process could occur concurrently with other tasks such as fingerprinting, capturing a face image quickly has the additional benefit of reducing workload for the operator and maintaining transparency to the visitor.

The *acquisition rate* refers to the fraction of participants for whom there is at least one available image that passes the quality checks. Figure 5-14 shows the acquisition rates for both cameras as a function of time. The timer started when the operator clicked on the capture button and continued for up to 30 seconds. The Logitech camera was able to immediately capture an image that passed the quality checks for between 25 and 40 percent of the participants, depending on the lighting scenario. For the ambient and side-lit scenarios, the Logitech camera was able to acquire a compliant image for 70% of the participants after 10 seconds had elapsed. The acquisition rate was slightly higher for the dim scenario. Results were similar for the Canon camera.

Figure 5-14 shows a diminishing return as time elapses. In general, the acquisition rate improved most rapidly within the first 10 seconds of capturing and made only minor improvements beyond 10 seconds. For example, the acquisition rate improved by only 8% from 10 to 30 seconds for the Logitech camera under ambient lighting conditions. This result is typical of the other lighting scenarios and the Canon camera.

In addition, the participants were generally cooperative and followed operator instructions. This would indicate that the failures to acquire are the result of the QA software failing to recognize good-quality face images.

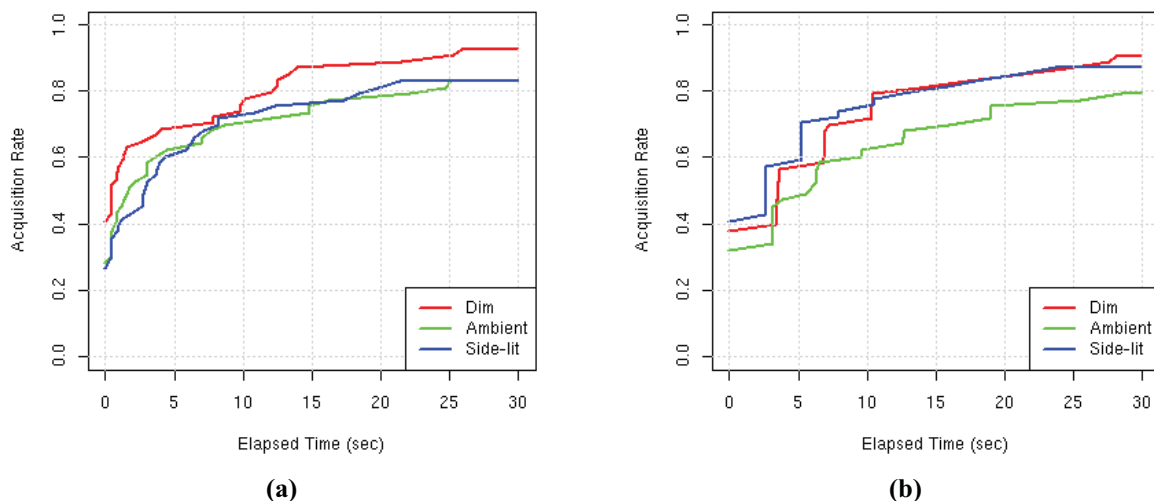


Figure 5-14: Acquisition rates versus time for (a) the Logitech Camera and (b) the Canon camera. In this context, the acquisition rate refers to the fraction of participants for whom there is at least one available image that passes the quality checks.

5.3.3 Reliability of Automated Quality Assessment/False Accept

In an effort to measure the reliability of the quality test, 100% of the images that passed the automatic quality test were visually inspected for compliance to FACESTD. Images were evaluated with respect to: pose, eyes open, eyes tinted, roll, sharpness, mouth closed, and expression. Lighting and background factors were not considered during this manual “ground-truthing”, because the quality test did not assess these factors (since they are intrinsic to the POE environment and cannot be changed). This human inspection determined that the overall compliance rate of the quality test was 88.6%. The compliance rate of the images that the operator selected among the up to four passing images for each case was higher (91.6%) than the images that the operator did not select (87.4%), as graphed in Figure 5-15. Figure 5-16 shows the percentages of images that passed the automatic quality test, but were deemed non-compliant during human review based on the various factors.

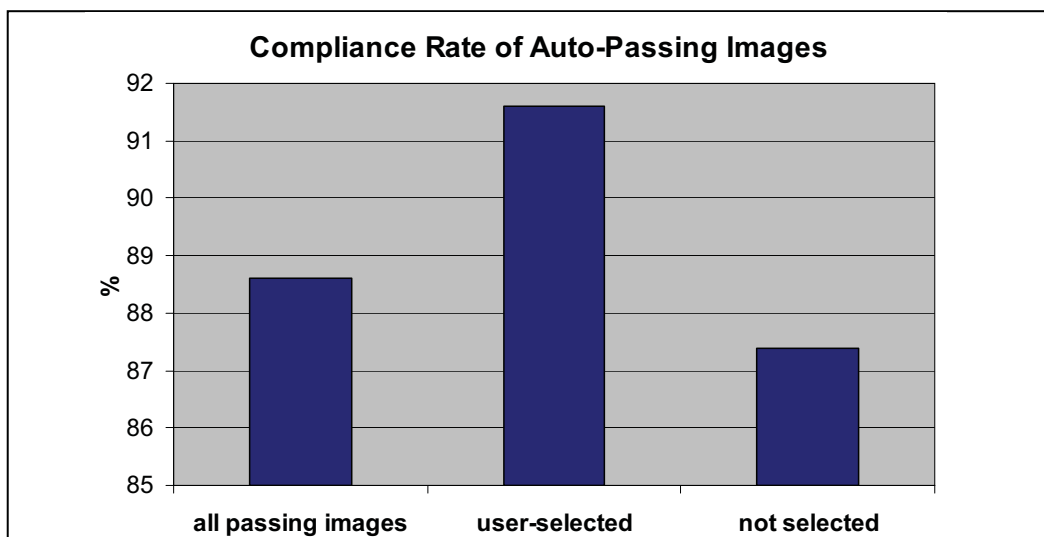


Figure 5-15: Compliance rate of automatically passing images and those selected by operator

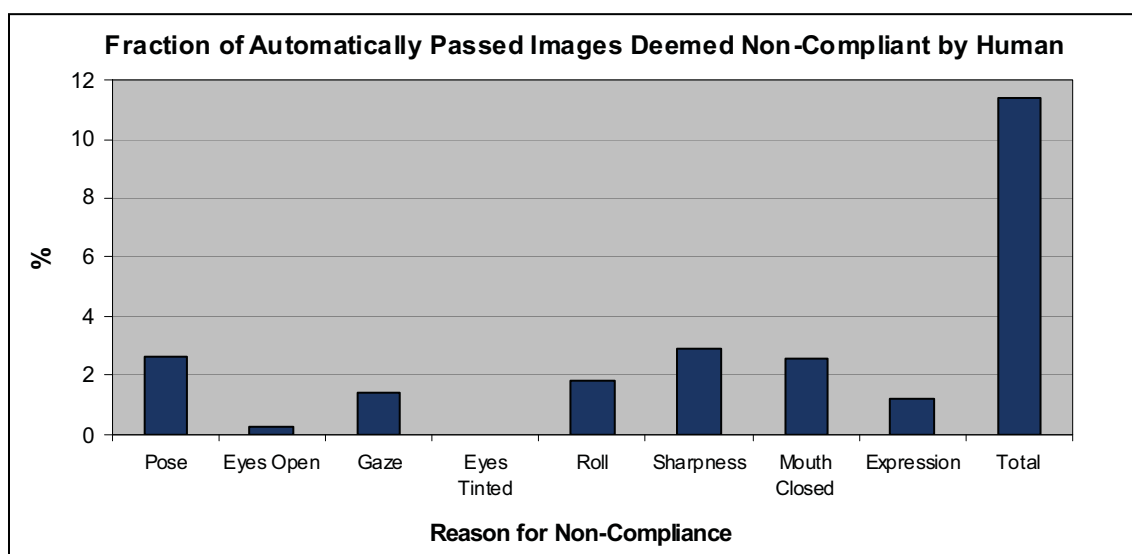


Figure 5-16: Percentage of automatically-selected images deemed non-compliant by human

The acquisition rates described in Section 5.3.2 were combined with the reliability measurements to produce the Compliant Acquisition Rate, which is defined in this context as the fraction of cases for which the quality test found at least one passable image, which was also deemed compliant by a human, within 10 seconds. The Compliant Acquisition Rates are graphed in Figure 5-17. The Logitech 9000 had a slightly higher Compliant Acquisition Rate (66.5%) than did the Canon (61.3%). The side lighting scenario had the highest Compliant Acquisition Rates, followed by dim, and then ambient.

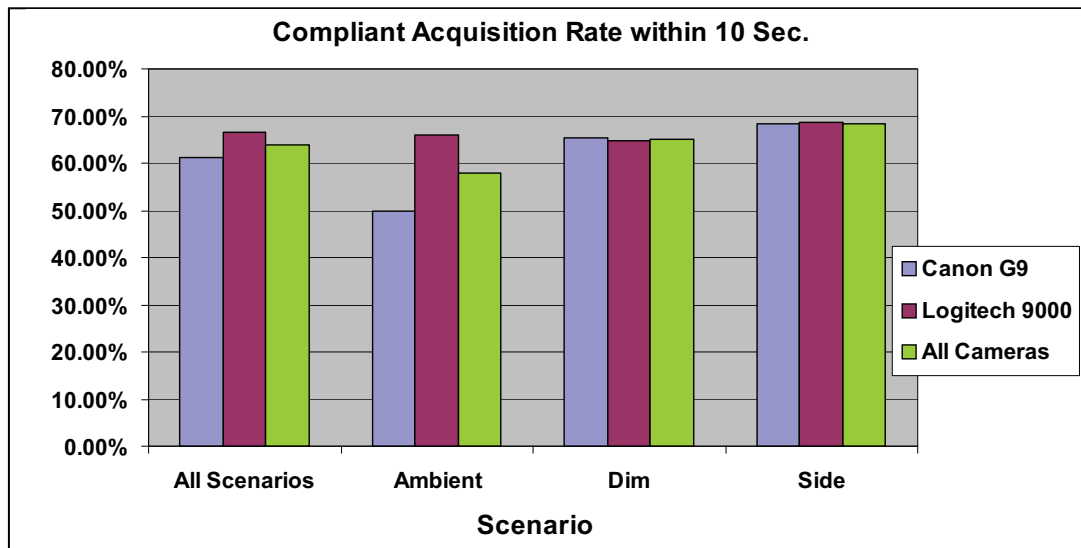


Figure 5-17: Percentage of cases where at least one frame passed the quality test (and was also deemed compliant by a human) within 10 seconds

5.3.4 Failure to Acquire/False Rejection

Failure to acquire (FTA), in the context of the FIQIFRS project, is defined as a case (camera/subject/lighting combination) where none of the frames passed the quality-in-the-loop test, and hence, an image was not acquired. This error is different from Phase I's FTE (Section 5.2.2.3), in which case the image QA software could not find a face in a previously captured image. In Phase II's FTA cases, the operator captured a manual snapshot, examples of which are shown in Figure 5-18.

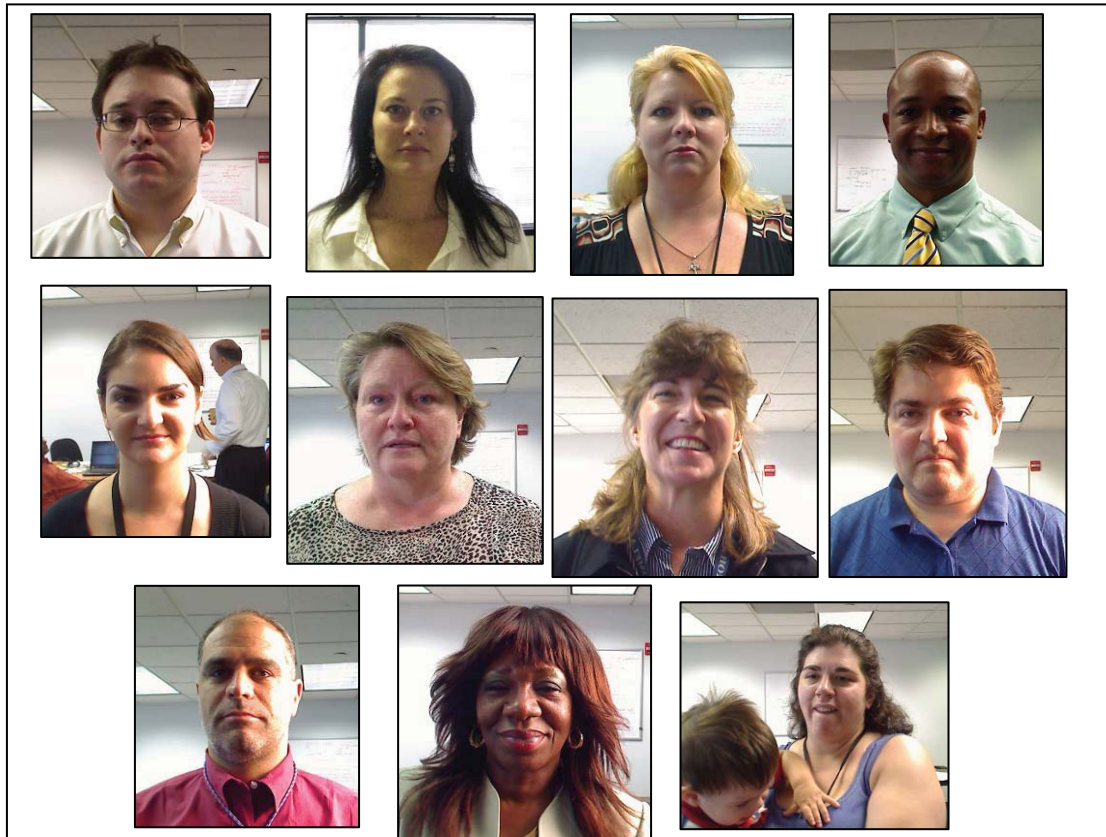


Figure 5-18: Example cases for which software could not find a passing image

The overall FTA rate was 16.7%. The percentages of FTA cases for each combination of camera and lighting scenario are graphed in Figure 5-19. The graph shows that both cameras had very similar FTA rates, and that ambient lighting had more FTAs than the other lighting scenarios, perhaps due to the overhead fluorescent lights causing shadows on the face. Such shadows may have been reduced in the side lighting scenario due to the light shone on the side of the face. The lowest FTA rate occurred with the dim scenario, probably because it was not subject to artificial light. 46.6% of the test subjects failed to be acquired in at least one case (camera/subject/lighting combination), indicating that the failures were generally not subject dependent.

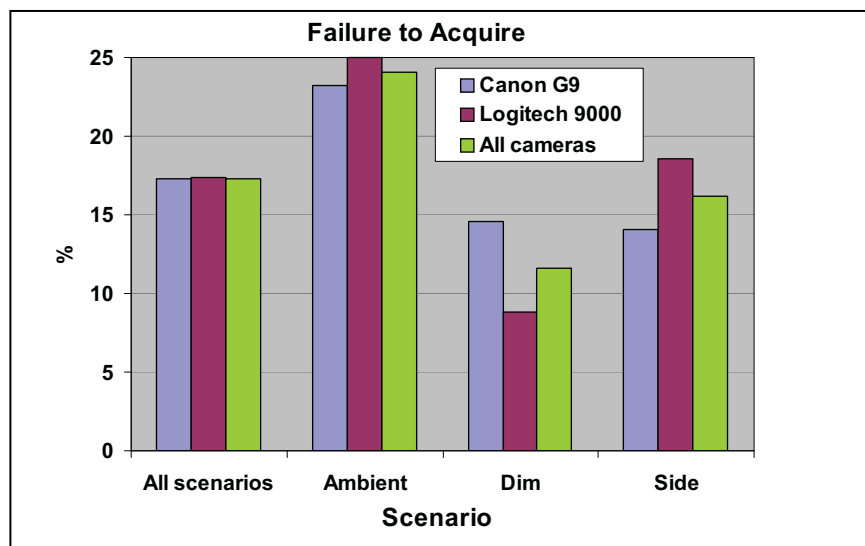


Figure 5-19: FTA cases - where no images passed the quality test

However, the quality software did have difficulty recognizing good-quality face images for some participants. These participants were generally cooperative and presented good-quality face images to the QA software, yet none of the images passed the quality checks. In general, the quality software reported that the participants were either not facing the camera (i.e. they failed the pose test), their mouths were open, or they were not looking at the camera (i.e. they failed the eye gaze test). In fact, the participants were often facing forward and staring directly at the camera for a significant portion of the capture period.

5.3.5 FR Performance

Ideally, image quality should correlate both with human assessment and automated matching performance. If images captured using quality-in-the-loop do represent better quality images, they should produce lower error rates when matched. A FR algorithm was used to compare samples acquired with quality-in-the-loop to the reference images collected for each subject. In addition, the first image frame for each capture session was also saved (regardless of its quality) and compared to the reference samples. This first frame image was captured immediately after the capture button was pressed and is intended to represent the baseline scenario of capturing in the traditional fashion (i.e. without employing QA).

Figure 5-20 shows Detection Error Trade-off (DET) curves for images captured using the quality-in-the-loop method with a 30 second timeout and for the first frame images. In general, the lower the curve in the figure, the better the matching performance. Both graphs show lower error rates for the images captured using the quality-in-the-loop method. At a given false match rate (FMR), the false non-match rate (FNMR) for the images captured using quality-in-the-loop is at least half that of the first frame images. This is strong evidence that integrating QA into the capture process can produce significantly improved face images.

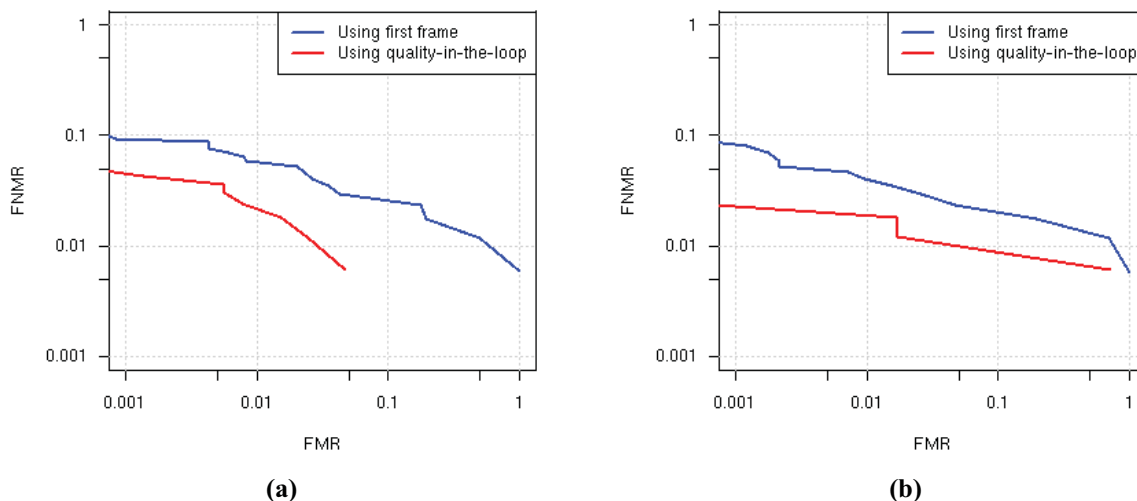


Figure 5-20: DET curves for images captured using quality-in-the-loop and for the first frame images. (a) Logitech Camera. (b) Canon Camera

5.3.5.1 Performance at Different Timeouts

Since the benefit of integrating quality into the capture process appears to diminish the longer the camera continues to capture images, it may be advantageous to terminate the capture process prior to the full 30 second timeout. Figure 5-21 shows performance curves for simulations of various timeout periods. If an image that passes the quality checks was not available for the participant within the given time frame, the manually-captured image was used instead. If a manually-captured image was not available, the first frame image was used as a reasonable representation of a manual capture. The Logitech camera appears to benefit little for timeout periods longer than 10 seconds. The Canon camera, on the other hand, continues to show improvements for timeout periods beyond 10 seconds. This is likely due to the slower frame rate of the Canon camera.

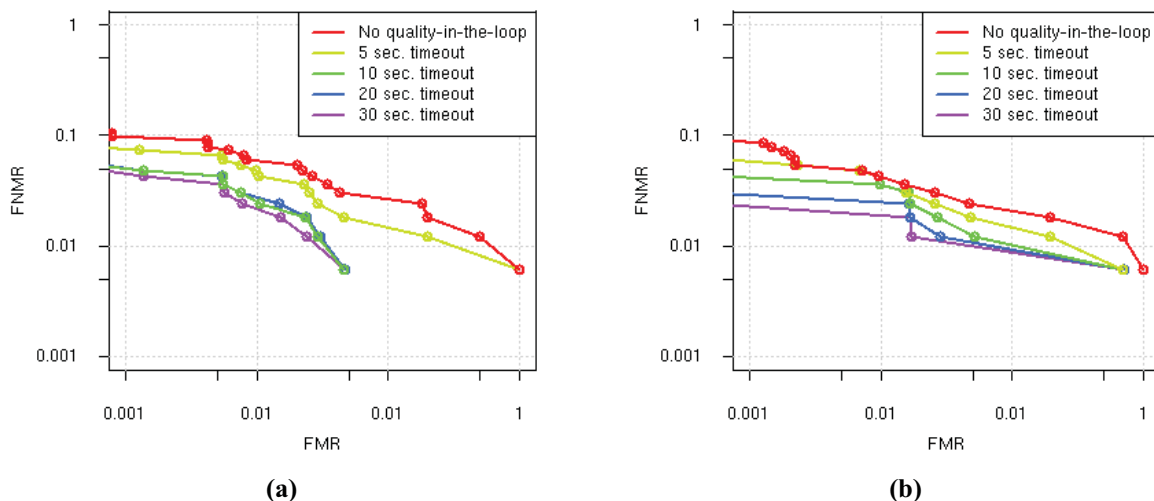


Figure 5-21: Simulation of various timeout periods for (a) Logitech, and (b) Canon

5.3.5.2 User Selected Image

For each capture session, the operator was tasked with selecting the face image that appeared to be the best from the ones that passed the quality checks. The software application would capture images until up to four face images passed the quality test, or a 30 second timeout occurred. Thus, the operator had up to four images from which to choose.

Figure 5-22 shows DET curves for the user-selected images and for the first image that passed the quality checks for each capture session. The three lighting scenarios for each camera were aggregated in order to produce a more robust curve. The figure shows that user selection did not improve matching performance. This is corroborated by statements from the operators who claimed that the images to choose from tended to all look the same. Furthermore, the results seem to indicate that it may be feasible to terminate the capture process as soon as the first image that passes the quality checks is acquired. The Officer would be given an opportunity to reject an unacceptable image and initiate a recapture if necessary.

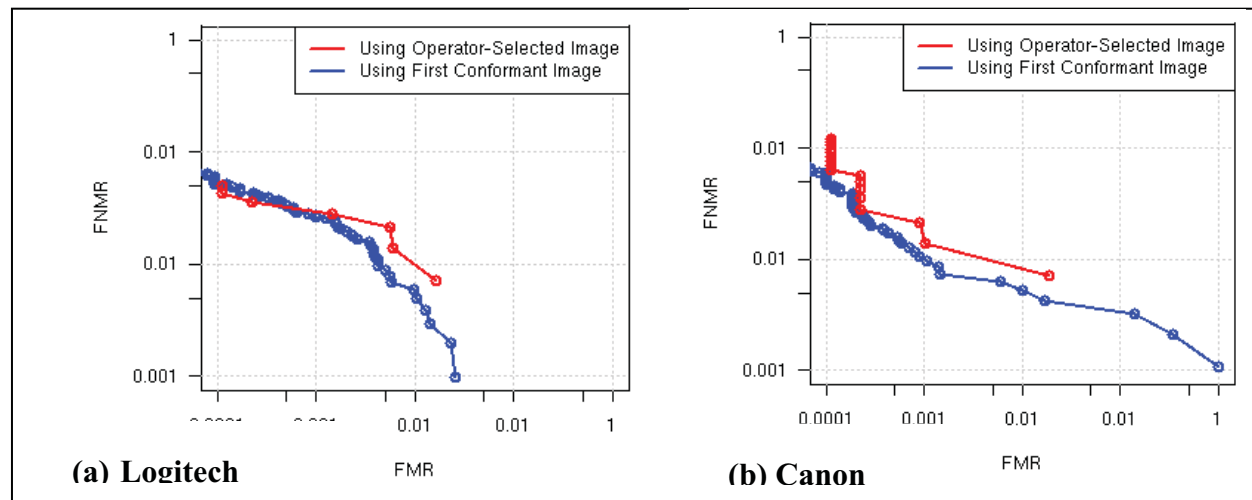


Figure 5-22: Performance comparison of operator-selected image versus the first image that passed all the quality checks

5.3.6 Timing Results

Automated QA should be fast enough to process images at a rate close to the frame-rate of the capture device, which is typically on the order of several frames per second. The Logitech camera has the ability to capture two megapixel images at a maximum rate of five frames per second using Logitech's QuickCam driver. The Canon camera has a lower frame rate due to limitations of the interface software. The Canon drivers require the camera to first store the image on its on-board memory card before it can be uploaded to the computer via a Universal Serial Bus (USB) connection. The additional steps involved with the image transfer limited the capture rate of the Canon camera to approximately one image every 2.5 seconds.

5.3.6.1 Eye Finding Time

The location of the participant's eyes within the image had to be determined before automated QA could be performed. The image QA software performs eye finding by first locating the face, and then finding the eyes within the face. Face and eye finding together took an average of 0.08 seconds to perform for the Logitech camera and 0.09 seconds for the Canon camera. These times remained relatively fixed and did not change appreciably for the different lighting scenarios or for different participants.

Future implementations of the Phase II application could improve the speed of eye finding by taking advantage of the fact that eye locations vary little from one frame to the next. Thus, once the initial eye locations are determined, eye finding could be restricted to a narrow region within the image based on where the eyes were located in the previous image frame.

5.3.6.2 Quality Assessment Time

Automated QA took an average of 0.16 seconds to perform per image for the Logitech camera, and 0.16 seconds for the Canon camera. As with eye finding, this time remained relatively fixed across lighting scenarios and for the different participants.

5.3.6.3 Operational Frame Rate

Eye finding and QA together took an average of approximately 0.25 seconds to complete. This would suggest a potential frame rate of four per second, but operationally a lower frame rate was observed. The limitations of the Canon interface software restricted its frame rate to about one image every 2.5 seconds. For the Logitech camera, an average of 0.45 seconds elapsed between successive frames, which corresponds to a frame rate of slightly more than two per second. This lower frame rate is due to overhead from the software application and could be improved in future implementations.

5.3.7 Phase II Quality Metric Distributions Compared to Three Other Datasets

This section uses the QA software to compare Phase II images to other face datasets, including operational POE images. Since each quality metric addresses a different aspect of face image quality, each metric offers a different way to compare the datasets.

Phase II focused on those metrics that are likely to change during the duration of the capture process. For example, a participant's head pose and eye gaze are likely to change, but environmental conditions such as lighting are expected to remain relatively unchanged. This

section compares the datasets with respect to these metrics, as well as a few others that were produced by the QA software.

The QA software was used to compare Phase II images to the following datasets:

- POE: Operational POE images collected in 2007.
- Usability Study: Images collected for use in the NIST “Assessing Face Acquisition” study (Section 7). The high-resolution images (1944 X 2592 pixels) were collected in a mock POE environment, but with specific usability enhancements. The usability enhancements were shown to improve quality over current operational images.
- Face Recognition Technology (FERET): Images from the well-known FERET¹⁴ database. The FERET images were captured in a controlled environment (e.g., by a professional photographer using studio lights) and are of relatively good quality. For the purposes of this comparison, they serve as the “gold standard” for image quality.

Each of the figures below compares the datasets with respect to a different quality metric. When a dotted line is present, it represents the threshold used by the Phase II FIQIFRS application during face image collection. In the figures, the Phase II images are separated by lighting scenario and camera. Each of these distributions was generated using one image per subject, which was either the user-selected image that passed the quality test, or the manual snapshot. Not all Phase II images meet the threshold requirements, because some of them had to be captured manually.

The Phase II FIQIFRS application required a distance between the eyes of at least 90 pixels. As shown in Figure 5-23, the POE images have a small *eye distance* because they were captured at a lower resolution. The eye distances for the Logitech camera are slightly lower than for the Canon camera because the Logitech camera uses a Wide-Angle-Zoom lens.

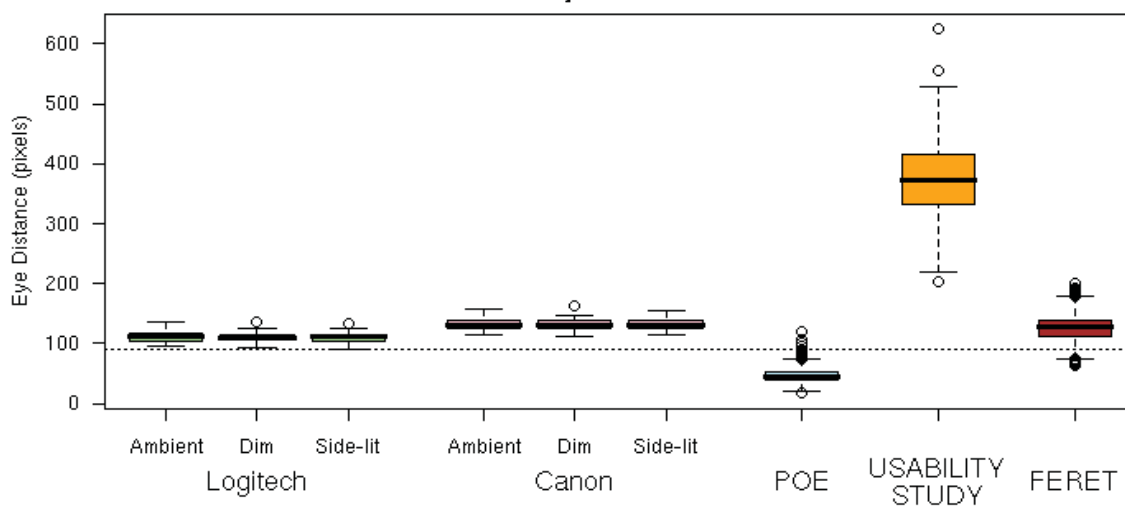


Figure 5-23: Phase II Eye Distances Compared to 3 Other Datasets

¹⁴ www.frvt.org/FERET/

The Phase II application rejected face images that had a *sharpness* value below 0.013. Blur can be caused by an out-of-focus camera, but it can also be caused by motion blur, due to either the person moving or the camera shaking. It is often difficult for a camera to focus in a poorly lit environment. As shown in the graph in Figure 5-24, about half of the POE images were below the threshold for sharpness, which is consistent with visual assessment of the POE images.

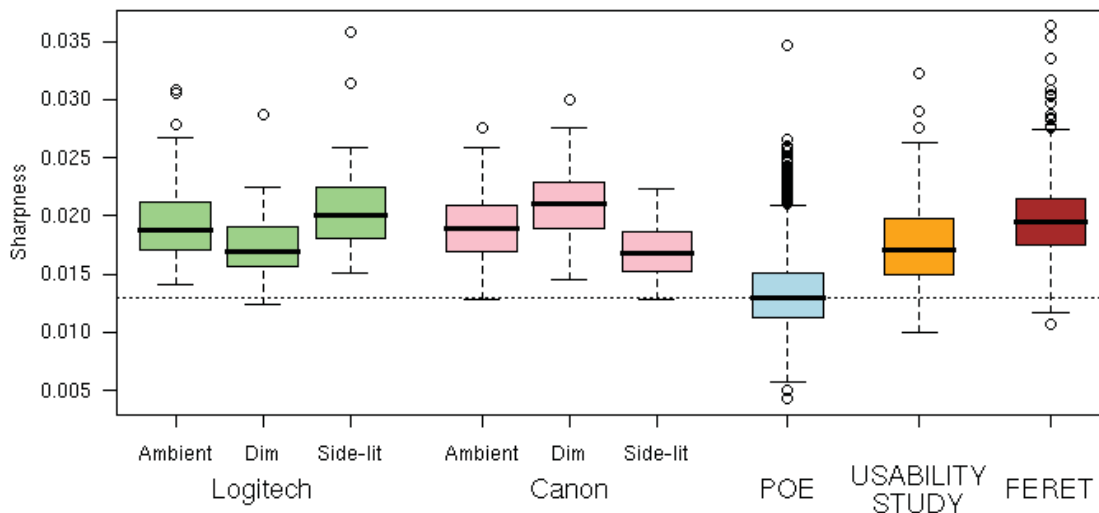


Figure 5-24: Phase II Sharpness Metric Compared to 3 Other Datasets

The neutral expression defined in FACESTD specifies mouth closed. The Phase II FIQIFRS application only accepted faces when the *mouth-closed* value was at least 0.7. Graphed in Figure 5-25, the mean value for the POE images is above the threshold; however, there are many outliers, possibly due to the travelers speaking while their photos were being taken.

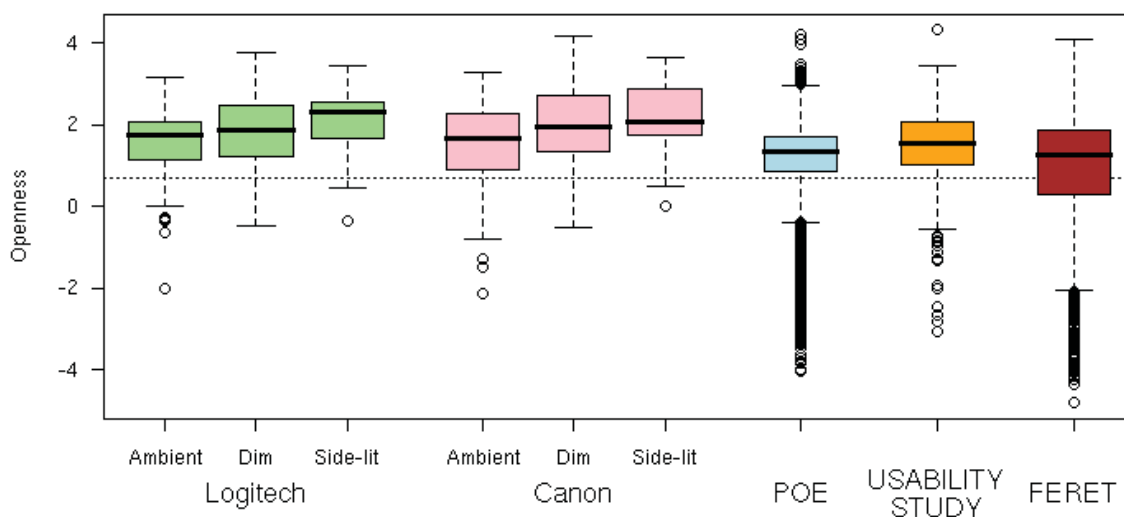


Figure 5-25: Phase II Mouth Open Metric Compared to 3 Other Datasets

While higher *deviation from frontal pose* values generally indicate greater pose deviation, this metric is only intended to be used as a boolean test of compliance to FACESTD. The Phase II

application rejected images with a *deviation from frontal pose* value greater than zero, which seemed to allow roughly seven degrees of pose deviation in either the yaw or pitch directions. Figure 5-26 shows that POE images failed the pose test most often.

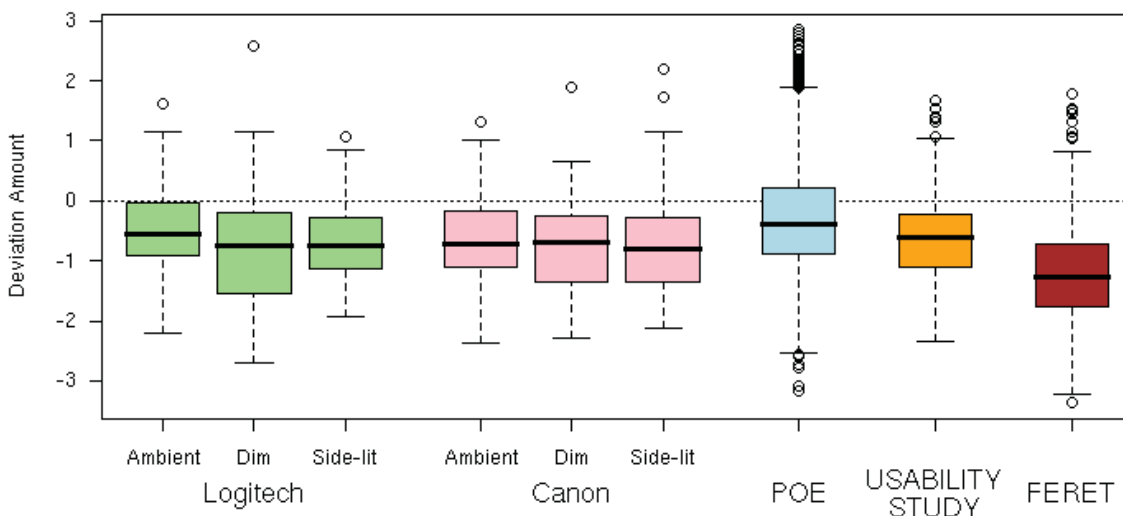


Figure 5-26: Phase II Deviation from Frontal Pose Metric Compared to 3 Other Datasets

Low *eye gaze* values indicate that the person is not looking at the camera. Figure 5-27 shows that *eye gaze* is fairly uniform across the datasets.

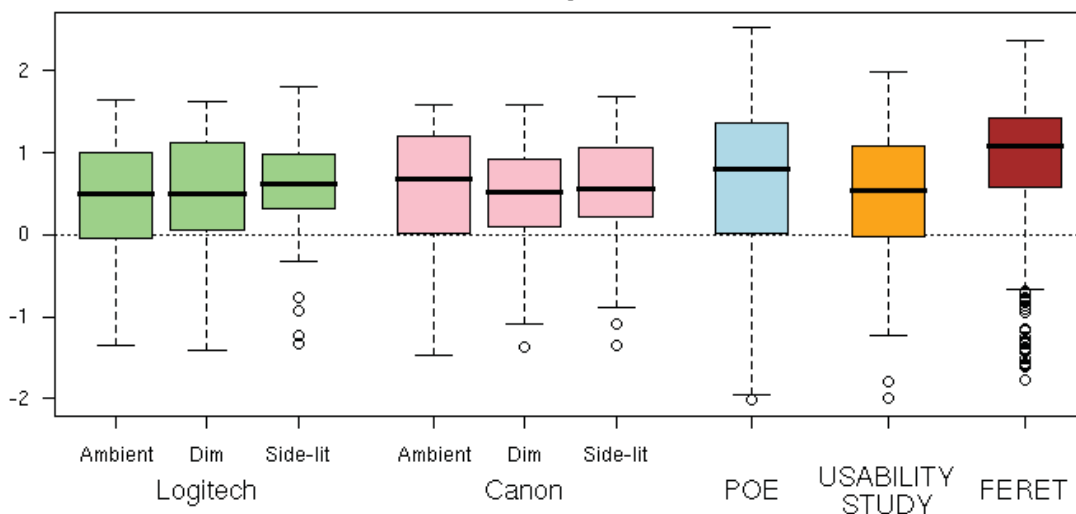


Figure 5-27: Phase II Eye Gaze (Looking at Camera) Metric Compared to 3 Other Datasets

The *eyes open* metric indicates the clarity and openness of the eye. Figure 5-28 shows that the POE images had the highest standard deviation in *eyes open* values, indicating that the eye was obstructed in many of the POE images.

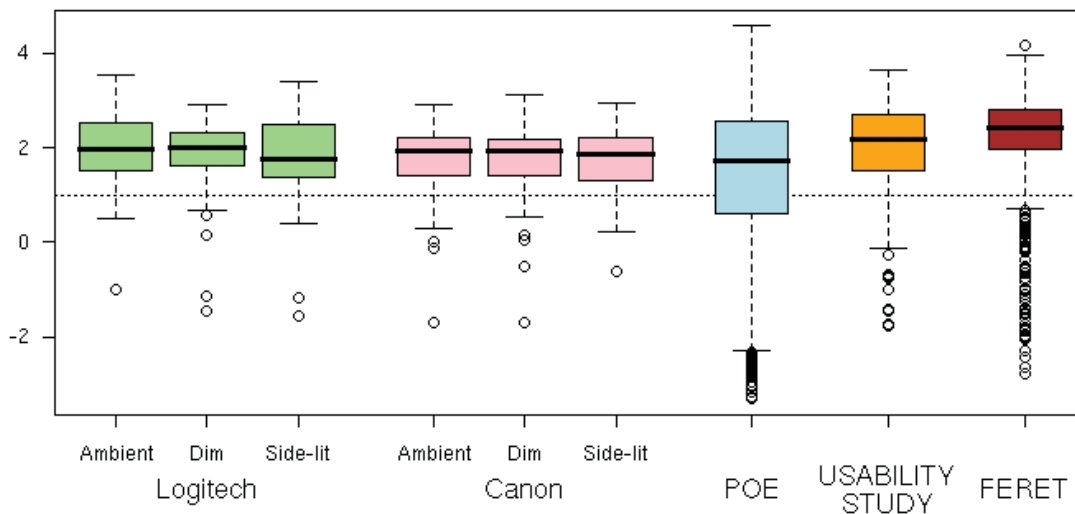


Figure 5-28: Phase II Eyes Open Metric Compared to 3 Other Datasets

Head *roll* occurs when the head in the image is tilted toward the shoulders. The roll can be computed from the eye coordinates and can be corrected post-capture. A value of zero corresponds to an unrotated head. As shown in Figure 5-29, the POE images displayed the largest variation in *roll* values, with many outliers not within the thresholded region. The reason that there are some Phase II outliers above the upper threshold is that the WG mistakenly neglected to impose the upper threshold as a result of lack of information in the vendor documentation. It was determined after analysis of the results that an upper and lower threshold must be imposed.

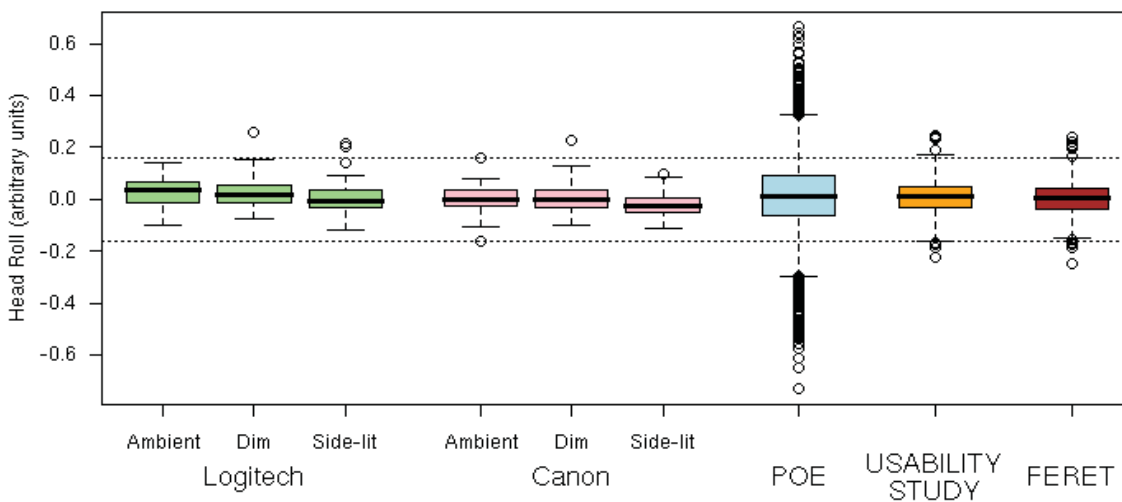


Figure 5-29: Phase II Head Roll Metric Compared to 3 Other Datasets

Exposure refers to the degree of illumination of the face. Ideal values were not provided in the vendor’s documentation. The aperture and shutter speed on the Canon were fixed for each lighting scenario while the Logitech 9000 camera was in auto-exposure mode, which allowed the camera to adjust the exposure amount to the scene. Again, the POE images displayed the highest variation in *exposure* values (Figure 5-30).

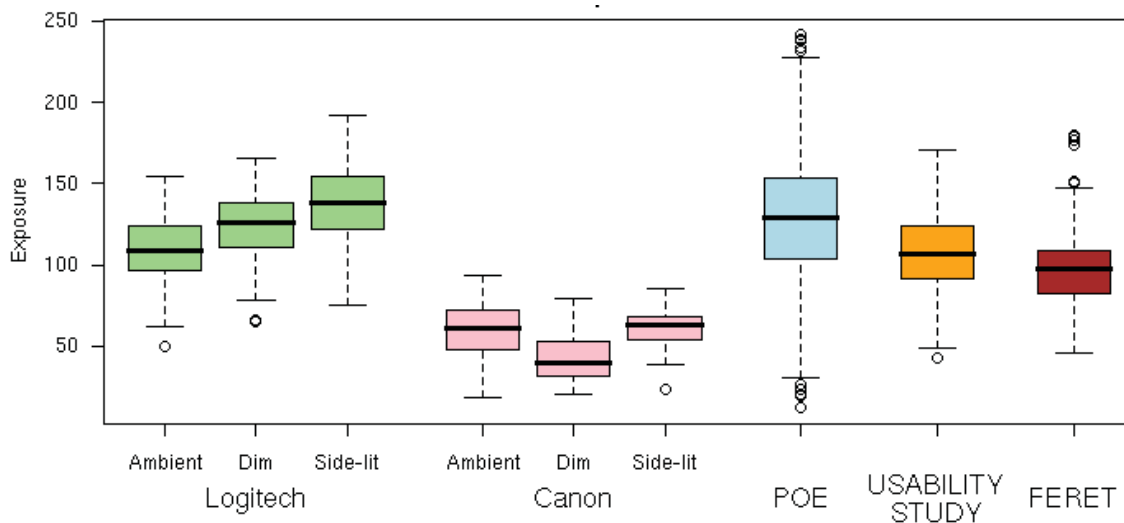


Figure 5-30: Phase II Brightness Exposure Metric Compared to 3 Other Datasets

Higher values for the *deviation from uniform lighting* metric indicate non-uniformity of lighting across the face. The quality software estimated that the Phase II images captured under dim lighting conditions had less uniformity than the ones captured under ambient lighting conditions (Figure 5-31), which is somewhat surprising, since the ambient lighting scenario consisted of overhead fluorescent lights which cast shadows on the face. The auto-exposure functionality of the Logitech camera appears to partially mitigate the disparity in lighting across the three scenarios. Fixed aperture and shutter settings were used on the Canon in lieu of the Canon’s auto-exposure mode. POE images appear to have relatively uniform lighting, although there are several outliers. The mean value for the POE images is very close to that of the Phase II ambient scenario, which may indicate that the ambient scenario was a good model of an actual POE.

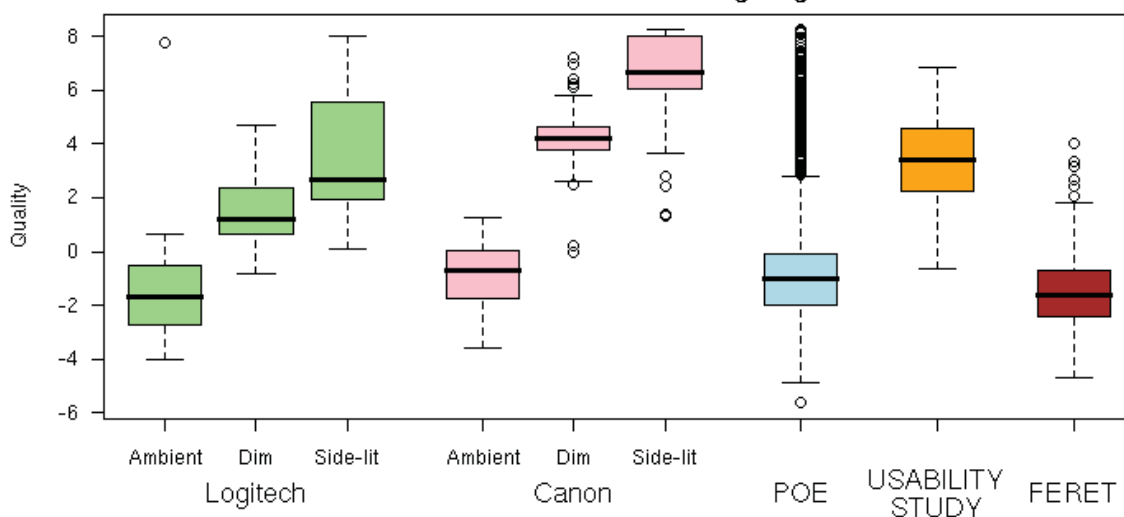


Figure 5-31: Phase II Uniform Lighting Metric Compared to 3 Other Datasets

6 Conclusions

6.1 Camera Pre-assessment

Representative cameras from six categories were evaluated for suitability for face capture in US-VISIT POEs. A desirable camera would be mountable on a universal mount, compact, and easy to use (e.g., auto-focus and auto-exposure), and would have high capture dimensions, fine spatial resolution, little noise, low compression, accurate color fidelity, and good light sensitivity and spatial uniformity. The characteristics of the tested cameras are summarized in Appendix A:. The images output from the cameras were evaluated with respect to certain metrics, and the measurements are reported in Appendix B:. Example images and the full conclusions of the camera pre-assessment are contained in Attachment 2 [12] to this Report.

Following is an overview of conclusions from the camera pre-assessment:

- The Canon G9 image was visually superior to the images from other cameras in its color and fine detail, even at its medium capture dimensions.
- The newer Logitech models tested were superior to the webcam currently in use at the POEs (Logitech QuickCam Pro 5000) in resolution, focus, and color fidelity.
- The Logitech webcams were difficult to mount as procured. A custom mount would be necessary for use in the POE environment.
- The Logitech webcams have fixed compression levels and do not permit manual white balancing.
- Logitech QuickCam Pro 9000 and Orbit AF had similar output results.
- Logitech QuickCam Pro 5000, Sony EVID70, and wide dynamic range camera produced images with insufficient eye distances.
- While the smallest cameras in the study were the Logitech webcams, all of the cameras were a reasonable size for the POE environment.
- *Field of view* - when operated in portrait mode, the Logitech QuickCam Pro 9000 and Canon G9 had the largest fields of view, followed by the Orbit AF in landscape mode.
- *Geometric accuracy* - all of the cameras exhibited less than one percent of distortion, which is probably negligible for FR.
- *Spatial uniformity* - for the most part, the spatial uniformity results followed a trend - the larger the field of view of the camera, the less uniform was the intensity. This reduction in spatial uniformity is probably a fair trade-off for the significant benefit of having a large field of view.
- *Depth of field* - the only tested camera that had adequate depth of field is the Canon G9 (at capture dimensions of 2592 x 1944 pixels and higher).
- *Noise* - the cameras with the highest (best) SNR were the Logitech webcams, and the lowest (worst) SNR was from the Canon G9. The lower amount of noise in the webcam images may be due to the inherently smaller number of pixels per patch and by the JPEG compression, which reduces the variability in the patches.

- *Spatial resolution* - all of the tested cameras exhibited overshooting in their edge profiles and aliasing beyond the Nyquist frequency. The Canon G9 had superior spatial resolution, as evidenced by its much higher Nyquist frequency and shorter 10-90% rise.

6.2 Phase I

The conclusions of Phase I were as follows:

- The Logitech Quickcam Orbit AF exhibited high FTE rates and focusing problems.
- The Canon G9 should be operated at a medium resolution (1600 x 1200) in burst mode (where multiple still images are captured in a series).
- COTS QA products differ widely due to the lack of standardization, as each product measures different quality factors over different value ranges.
- Using correlation analysis, it was determined that the overall quality metric reported by the image QA software did not adequately predict FR performance.
- A consistent set of image QA metrics that predicted FR score was not found using regression analysis; therefore, quality metric thresholds were determined visually.
- FR scores were higher for the Canon than for the Logitech 9000.
- FR scores were highest under the dim lighting scenario, followed by overhead, ambient, side light, and back light.
- Face images acquired in the back lit scenario were visually inferior to the other scenarios and also resulted in the highest FTE rates with QA software.
- Unexpectedly, side lighting did not adversely impact Vendor C's image enrollment rates, and overhead lighting improved enrollment rates over ambient lighting.
- As expected, images captured with the Logitech 5000 were visually inferior to those from other cameras.
- Although they were not visually superior, the images under the dim lighting scenario had the lowest FTE rates, perhaps due to the absence of artificial lights.

6.3 Phase II

The conclusion of Phase II was that image quality software can, in fact, be used to acquire a standards-compliant face image of a subject that is suitable for both human and automated FR.

6.3.1 Speed of Quality Test

Face and eye finding together took an average of 0.08 seconds to perform for the Logitech camera and 0.09 seconds for the Canon camera. Automated QA took an average of 0.16 seconds to perform per image. Eye finding and QA together took an average of approximately 0.25 seconds per frame to complete. Future implementations could improve the speed of eye finding by restricting the search region in subsequent frames after finding the eye locations in the first frame.

6.3.2 Reliability of Quality Test

The overall compliance rate¹⁵ of the quality test was determined to be 88.6%. The compliance rate of the images that the operator selected among the up to four passing images for each case was higher (91.6%) than the images that the operator did not select (87.4%), although not by a significant amount. In general, the acquisition rate¹⁶ improved most rapidly within the first 10 seconds of capturing and made only minor improvements beyond 10 seconds. The Logitech 9000 had a slightly higher Compliant Acquisition Rate¹⁷ (66.5%) than did the Canon (61.3%). The side lighting scenario had the highest Compliant Acquisition Rates, followed by dim, and then ambient.

6.3.3 Failure to Acquire

The overall FTA¹⁸ rate was 16.7%. Both cameras (Canon and Logitech 9000) had very similar FTA rates, and ambient lighting had more FTAs than the other lighting scenarios. 46.6% of the test subjects failed to be acquired in at least one case (camera/subject/lighting combination). FTAs are believed to be the result of the QA software failing to recognize good-quality face images and to the vendor's lack of algorithm training with dark-skinned individuals.

6.3.4 Face Recognition

At a given FMR, the FNMR for the images captured using quality-in-the-loop is at least half that of the first frame images. FR performance proved that user selection of one of four images that passed the quality test did not improve matching performance. The Logitech camera appears to benefit little for timeout periods longer than ten (10) seconds. The Canon camera, on the other hand, continues to show improvements for timeout periods beyond ten (10) seconds.

6.3.5 Comparison of Quality of Test Images with POE Images

The same QA product was executed on the test images and operational POE images circa 2007. In the case of most quality metrics, the mean values for the POE images were above the thresholds; however, the POE images showed a higher standard deviation and variance around the mean, indicating that many POE images would not have passed the quality test, especially in the *eyes open*, *deviation from frontal pose*, and *sharpness* metrics. Almost none of the POE images had the recommended eye distance of 90 pixels.

6.4 Recommendations

Recommendations are listed in Table ES-2 of the Executive Summary, and are not repeated here

¹⁵ Fraction of images that passed the automatic quality test that was deemed compliant by a human.

¹⁶ Fraction of participants for which there is at least one available image that passes the quality checks.

¹⁷ Fraction of cases for which the quality test found at least one passable image, which was also deemed compliant by a human, within 10 seconds.

¹⁸ Failure to acquire (FTA), in this context, is defined as a case (camera/subject/lighting combination) where none of the frames passed the quality test.

7 NIST Interagency Report 7540, Assessing Face Acquisition

In addition to NIST's participation as biometric technical experts for this project, NIST performed usability and human factors research to improve facial image capture. The research was based on observations from visiting Dulles Airport and testing of an image overlay with camera operators and paid volunteer subjects. Although it was not part of the FIQIFRS testing, some of the recommendations and findings have been included in the recommendations contained in this report. The following is excerpted from the Report's Executive Summary¹⁹.

US-VISIT requested that the biometrics usability team at NIST examine the current US-VISIT face image collection process to identify any usability and human factors that may improve the existing face image capture process. As such this study did not address other technologies or technology solutions. This report presents the results of a study that examined five usability and human factors enhancements to the current US-VISIT collection process:

- 1. The camera should resemble a traditional camera;*
- 2. The camera should click when the picture is taken to provide feedback to the traveler that the picture is being taken;*
- 3. The camera should be used in portrait mode;*
- 4. The operator should be facing the traveler and the monitor while positioning the camera; and,*
- 5. Provide some marking on the floor (such as footprints) to indicate to the traveler where to stand for the photograph.*

The study was conducted as follows: first we visited and observed a representative operational setting (Dulles Airport) in order to understand the primary users and the context of use. Based on these observations we identified the 5 usability and human factors enhancements enumerated above that may improve the face image capture process. A usability study was designed that mimicked the operational process but incorporated the 5 enhancements and face images were collected from 300 participants. A visual inspection evaluation methodology based on an image overlay was used to quantify the various characteristics of face imagery based on the face image standards. Results from the visual inspection process compared favorably with preliminary automated face image quality metrics under development.

The FIQIFRS project team has the following comments on the recommended usability and human factors enhancements:

1. "Camera should resemble a traditional camera". This recommendation, while ideal, is considered less important than the other factors identified in the main recommendations of this report.
2. "The camera should click". This recommendation, while ideal, is appropriate for single frame capture scenarios. This contrasts with the multiple frame capture paradigm advanced in the main body of the report.
3. "The camera should operate in portrait mode". This recommendation appears in the main body.

¹⁹ The complete report is at http://zing.ncsl.nist.gov/biousa/docs/face_IR-7540.pdf

4. "The operator should be facing the traveler". This recommendation corrects situations in which the operator had been required to lean and turn to adjust the existing POE gooseneck camera. With the recommendation to produce a fixed-mount camera, this recommendation is met.
5. "Provide a marking on the floor". This recommendation has been made in the main body of the report.

8 Acronyms and Abbreviations

ANSI	American National Standards Institute
BVA	BioVisa
CBP	Customs and Border Protection
COTS	Commercial-Off-The-Shelf
DET	Detection Error Trade-off
DHS	Department of Homeland Security
DOS	Department of State
EBTS	Electronic Biometric Transmission Specification
EXIF	Exchangeable image file format
FACESTD	ISO/IEC 19794-5:2005, Information Technology — Biometric data interchange formats — Part 5: Face image data
FERET	Face Recognition Technology
FIQIFRS	Facial Image Quality Improvement and Face Recognition Study
FMR	False Match Rate
FNMNR	False Non-Match Rate
FR	Face Recognition
FTA	Failure to Acquire
FTE	Failure to Enroll
GFE	Government Furnished Equipment
GUI	Graphical User Interface
IAA	InterAgency Agreement
ICAO	International Civil Aviation Organization
IEC	International Electrotechnical Commission
INCITS	InterNational Committee for Information Technology Standards
ISO	International Organization for Standardization
IXM	IDENT eXchange Messages
JPEG	Joint Photographic Experts Group
MTF	Modulation Transfer Function
NIST	National Institute of Standards and Technology
POE	Port of Entry
PTZ	Pan-Tilt-Zoom
QA	Quality Assessment
RGB	Red-Green-Blue
ROC	Receiver Operating Characteristic
SBA	Smart Border Alliance
SDK	Software Development Kit
SFR	Spatial Frequency Response
SMIA	Standard Mobile Imaging Architecture

SNR	Signal-to-Noise Ratio
UDM	US-VISIT Delivery Methodology
USB	Universal Serial Bus
USG	United States Government
US-VISIT	United States Visitor and Immigrant Status Indicator Technology
WG	Working Group
XML	eXtensible Markup Language

Appendix A: Summary of Camera Characteristics

Camera Mount	Logitech Quickcam Pro 5000	Logitech Quickcam Pro 9000	Logitech Quickcam AF	Sony EVID70	Canon G9	Wide Dynamic Range
Size (WxHxD, in.)	2.5 x 2.5 x 2.5 (without clip)	3.5 x 1.5 x 1.5 (without clip)	3.25 x 4.25 x 3.25	5.25 x 5.75 x 5.75	4.19 x 2.83 x 1.67	Threaded hole 1.7 x 1.8 x 2.75
Connections	USB drivers	USB drivers	USB drivers	S-video, requires capture card; power software for camera settings	USB; power adaptor or battery SDK	BNC port, requires capture SDK
Software	drivers	drivers	drivers	software for camera settings	SDK	SDK
Capture dimensions	640 x 480	Still: 1600 x 1200 Video: 960 x 720	1600 x 1200	640 x 480	still: 4000 x 3000, 3264 x 2448, 2592 x 1944, 1600 x 1200, 640 x 480; Video: 1024 x 768, 640 x 480, 320 x 240	640 x 480
Compression	fixed; ~11:1	fixed; ~13:1	fixed; ~15:1	~4:1	Normal, Fine, SuperFine, RAW	~3.7:1
Field of View (in.)	26.2 x 19.6	35.3 X 26.5	32.7 x 24.6	24.7 x 18.5	31.5 x 23.6	17.3 x 13
Frame Rate	30 fps	30 fps	30 fps	1 to 1/10,000 s	30 fps	30 fps
Face detection	yes	yes	yes	no	yes (on camera for focusing)	no
PTZ	no	no	Mechanical P,T	Mechanical PTZ; Pan angle: -170° to +170° (max. pan speed: 100°/s); Tilt angle: -30° to +90° (max. tilt speed: 90°/s)	Mechanical zoom	no
Zoom	Digital zoom	Digital zoom	Digital zoom	18x Optical, 12x Digital	6x Optical / 4x Digital	none
Optics	Fixed focus	Carl Zeiss® lens Autofocus system	Carl Zeiss® lens Autofocus system	18x zoom, f=4.1 mm (wide) to 73.8 mm (tele), F1.4 to F3.0	7.4-44.4mm f/2.8-4.8	8mm, F1.4
Output format	JPEG, WMV	JPEG, WMV	JPEG, WMV	AVI	RAW, JPEG, AVI	AVI
Sensor	High quality VGA	CMOS sensor	CMOS sensor	1/4-type EXview HAD CCD	1/1.7 inch type CCD	2/3 inch
Cost	\$60	\$100	\$130	\$900	\$500	?

Appendix B: Summary of Camera Measurements

Camera	Logitech Quickcam Pro 5000	Logitech Quickcam Pro 9000	Logitech Quickcam AF	Sony EVID70	Canon G9	Wide Dynamic Range
Light Sensitivity (min lux)	not available	not available	not available	1 lx	ISO 1600	0.8 lux
Inter-eye distance (px.)	55	110	111	58	325	82
Geometric Accuracy	Distortion = -0.802%	Distortion = -0.21%	Distortion = -0.00359%	Distortion = -0.606%	Distortion = -0.588%	Distortion = -0.0439%
Spatial Uniformity	14% variation	18.3% variation	23.7% variation	18.4% variation	21.5% variation	8.9% variation
Depth of Field	fail	fail	fail	fail	pass, 9.2 in.	fail
Tonal Response (gamma)	0.938	0.664	0.54	0.797	0.631	0.7
Noise (average SNR)	137.7	198.19	199.22	121.87	80.04	113.25
Color Fidelity (ΔE)	19	8.94	7.44	9.11	10.6	14.9
Spatial Resolution (MTF)	10-90% rise= 1.3mm; Nyquist= 0.48 cy/mm; MTF(50)= 0.35	10-90% rise= 0.61 mm; Nyquist= 0.9 cy/mm; MTF(50)= 0.65	10-90% rise= 0.47 mm; Nyquist= 1.03 cy/mm; MTF(50)= 0.72	10-90% rise= 1.5 mm; Nyquist= 0.51 cy/mm; MTF(50)= 0.31	10-90% rise= 0.27 mm; Nyquist= 2.47 cy/mm; MTF(50)= 1.75	10-90% rise= 1.6 mm; Nyquist= 0.8 cy/mm; MTF(50)= 0.4

Attachment 1: Facial Image Quality Improvement and Face Recognition Study

Baseline Quality of US-VISIT POE Facial Images

NIST Deliverable to DHS US-VISIT
Face Image Quality Improvement Project

Version 2

Patrick Grother and George Quinn

February 6, 2009

Information Access Division
National Institute of Standards and Technology



Summary

February 6, 2009

This document is intended to be a complete summarization of the properties of US VIST POE images collected between 2004 and the present. This document serves as a basis against which the performance of potential and future US-VISIT facial image capture systems can be gauged.

As a baseline, performance is established in terms of the measured quality of the captured images. The analyses herein are based on IDENT images collected between 2003 and 2007. These images are of poor quality and will offer high verification or identification error rates. Newer images (collected in 2006 and 2007) do not appear to offer any improvement over older images (2004) despite minor changes in the capture protocol. The current images should not ordinarily be used in automated facial recognition processes.

This document may be revised. Particularly supplemental analyses may be conducted on the extant corpus of POE images in response to ongoing investigation of the new cameras tested in this project.

Version History

February 6, 2009	Incorporated results for new 2007 POE images.
April 29, 2008	Added material summarizing quality across 20 POEs.
April 14, 2008	Added yaw and quality metric sections. Added matching results for manual labeled strata
Mar 9, 2008	Version 0.1 text including 2004 assessment and latest VisPRO results Added criteria for manual inspection, and results.
Feb 26, 2008	Boxplots for VisPRO on five datasets
Feb 15, 2008	Info on first VisPRO runs

Table of Contents

1 Overview..... 1

2 Scope 1

3 Prior work 1

4 Manual Quality Assessment 2

 4.1 Overview..... 2

 4.2 Cropping 2

 4.2.1 Effect on automated matching 2

 4.3 Intensity Saturation 3

 4.3.1 Effect on automated matching 3

 4.4 Head Pose 4

 4.4.1 Effect on automated matching 4

5 Automated Quality Assessment 5

 5.1 Approach 5

 5.2 The VisPRO Image Quality Analysis Tool 5

 5.3 Image Databases 6

 5.3.1 POE Databases 7

 5.3.2 BVA Database..... 7

 5.3.3 Mexican Border Images 8

 5.3.4 FERET Database..... 8

 5.4 Results 8

 5.5 Processing Time 10

 5.6 Yaw Estimates..... 12

 5.7 Image Quality at Different Ports of Entry..... 13

 5.7.1 Results 14

6 Objectives for Next Generation Cameras..... 17

7 References..... 17

List of Figures

Figure 1 - Effect of manually encoding cropping on matching performance 3

Figure 2 - Effect of manually labeling saturation on matching performance 4

Figure 3 - Effect of manually labeling aberrant poses on matching performance 5

Figure 4 - VisPro Quality metrics..... 9

Figure 5 - VisPro Quality metrics (continued) 10

Figure 6 - Processing time for each of the quality metrics for 2004 POE images..... 11

Figure 7 - Processing time for each of the quality metrics for BVA images..... 11

Figure 8 - Processing time per image for each dataset..... 12

Figure 9 - VisPRO Yaw angle estimates for the FERET b-series image sets..... 13

Figure 10 - Variation in eye detection confidence across POEs..... 14

Figure 11 - Variation in face shadow measures across POEs..... 14

Figure 12 - Variation in background brightness measures across POEs 15

Figure 13 - Variation in the face centering measure across POEs 15

Figure 14 - Variation in the background consistency measure across POEs 16

Figure 15 - Variation in the yaw pose estimate across POEs 16

List of Tables

Table 1 - POE image non-frontal pose categorizations..... 4
Table 2 - Size and eye-detection rates for the five face databases 7
Table 3 - Overview of FERET B-series sets. 12

1 Overview

In 2003, DHS deployed face cameras in primary and secondary inspection processes for U.S. Ports of Entry. These are being used to collect a facial image of visitors to the United States. As such the images constitute a second biometric modality supplementing two fingerprints as the biometric data submitted to the IDENT system. While the current images are not suitable for automated facial recognition, and cannot usefully augment the fingerprints for US-VISIT identity management purposes, they do have some utility for manual resolution of exceptional events.

Future facial capture processes promise to yield face images capable of supporting automated multimodal biometric operations.

2 Scope

A sample of the POE images was provided to NIST in 2004 and again in 2008. These images form the sole basis for the analyses of this report which is intended to establish a baseline for facial image quality as it exists in POEs in the period 2003-2008 against which future improvements to cameras and acquisition procedures can be gauged. The mechanisms for doing this are

- manual categorization of images
- application of an automated face quality assessment implementation

The following assumptions are made

- There has been no systematic change in the collection procedures since the sample used here (mid 2007) was collected.

3 Prior work

NIST first examined the US-VISIT face images in April 2004 and delivered an analysis of the properties of the POE images to US-VISIT [1]. The report noted:

- Images were sized at 240x240 pixels (i.e. 0.06 megapixels)

- Images were compressed at 15.7 to 1 to a file size of 11KB.

- Subjects were imaged with a mean inter-eye distance of 49 pixels, with 95% of them below 74 pixels (machine determined over 1.52 million images)

- A commercial face recognition engine failed to find the eyes in 10.6% of images.

- The engine found two faces in about 0.05% of cases

- A cursory manual inspection of a few thousand images showed that the subject was frontal to the camera in only about 5% of images, and only about 30% of subjects were within 10 degrees of frontal.

- About 5% have some part of the face cropped (out of picture)

- About 1% have blur (usually motion artifacts)

The report recommended against use of automated face recognition with the POE images. This reflected geometric problems (in order: pose, size, cropping, fish-eye effects) and photometric problems (in order: compression, cluttered backgrounds, saturation). These findings were later aired publicly [2].

4 Manual Quality Assessment

4.1 Overview

In early 2008, NIST conducted a more systematic survey of the 2004 POE images. This involved manual inspection of 20,000 images by two NIST staff. This involved application of a graphical image categorization tool to label images presenting certain defects.

Inspecting the images manually is reliable in the sense that a human observer is capable of identifying a particular problem even in presence of other problems, and can distinguish between failure modes. For example, if the facial region is saturated and also cropped at the left eye, an automated quality assessment tool is likely to not find the face at all and report nothing.

A drawback of the approach is that it is subjective. Thus, when categorizing saturation, there is an inherent judgment to be made in distinguishing a bright image from a saturated one.

The manual inspection focused on three specific defects that are known to exist in the set of POE images: cropped faces, over-exposed faces, and non-frontal head poses. Together these defects strongly degrade the ability of contemporary face recognition engines to identify or verify the subjects appearing in the images.

The following subsections describe the criteria for each of the defects and present results of the manual inspection.

4.2 Cropping

Images were manually inspected for cropped faces. Each image was labeled as either cropped, or un-cropped. An image is considered cropped if any part of the face, from chin to mid-forehead, or from ear to ear, is not present in the image. The determination was made regardless of whether other problems were present, such as a non-frontal head pose or poor lighting.

The manual inspection identified 1,775 cropped faces (11.1% of the total set of images). Cropped faces always occurred for one of three reasons:

The camera was not pointing at the subject, and part of the face was outside of the camera's field of view.

The subject was standing too close, and the camera's field of view was not wide enough to capture the entire face at once. The camera operator typically compensated by positioning the camera such that the chin was clipped off the bottom of the image. Small distances between the subject and the camera also cause lens distortion (i.e. the fish-eye effect).

In rare cases, the subject's face was partially obscured by an object in the foreground, such as a suitcase or the back of a baby's head.

4.2.1 Effect on automated matching

The effect that cropped images have on matching performance was evaluated using an archived commercial matcher (produced circa 2005). The facial template generator was unable to find a face in 989 of the 1,775 cropped images. This type of error, known as a failure to enroll, notionally causes the generation of a blank template that always gives a low score when matched.

Figure 1 plots verification performance for

the entire set of POE images matched against BVA images

the subset of images that were identified as un-cropped.

The improvement in performance is small suggesting that cropping is not a significant contributor to the large observed error rates¹. In addition the improvement manifests itself more at higher false match rates (above 0.01). This indicates that cropping inhibits the initial ability to find the face in the image.

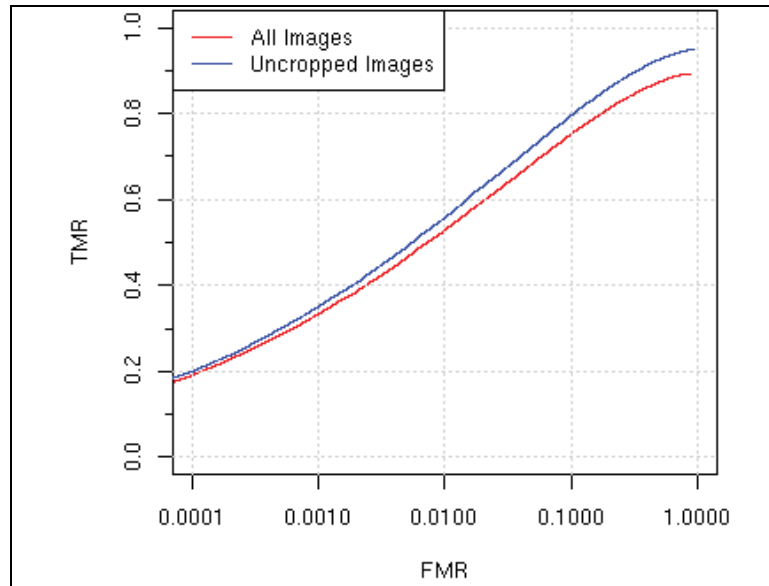


Figure 1 - Effect of manually encoding cropping on matching performance

4.3 Intensity Saturation

Images were manually inspected for faces that were over-exposed to light at the time of capture. These faces contain excessively bright areas, sometimes referred to as hotspots. The pixels within a hotspot, expressed as RGB triplets, will have maximum intensities for all three of the colors ($R = G = B = 255$, if 255 is the maximum color). An image is labeled as saturated if hotspots are clearly visible in any region of the subject's face, from chin to eyebrows, or from ear to ear. Note this excludes certain parts of the face, such as the left or right flank of the nose and the forehead. Hotspots were so prevalent in these parts that not excluding them would have led to nearly every POE image being labeled as saturated. If a face was over-exposed to light, but not to the point of saturation, it was not labeled as saturated.

The manual inspection identified 3,673 saturated images (18.4% of the total set of images).

4.3.1 Effect on automated matching

Figure 2 shows ROCs for the images labeled saturated and not. The result, that there is essentially no difference, is perhaps a surprising result in that fully saturated pixels (i.e. regions at value 255) do not appear to undermine the face recognition process beyond the other defects present in the images (pose, resolution etc). Nevertheless, saturation may remain problematic once other problems are remedied.

This negative result is also included here so that such analyses should not be repeated without a specific motivation.

¹ The error rate $TMR = 0.55$ at $FMR = 0.01$ is very much inferior to passport-against-passport matching, or to high-resolution-still to high resolution still matching, for which the error rates can exceed $TMR = 0.95$ at $FMR = 0.01$.

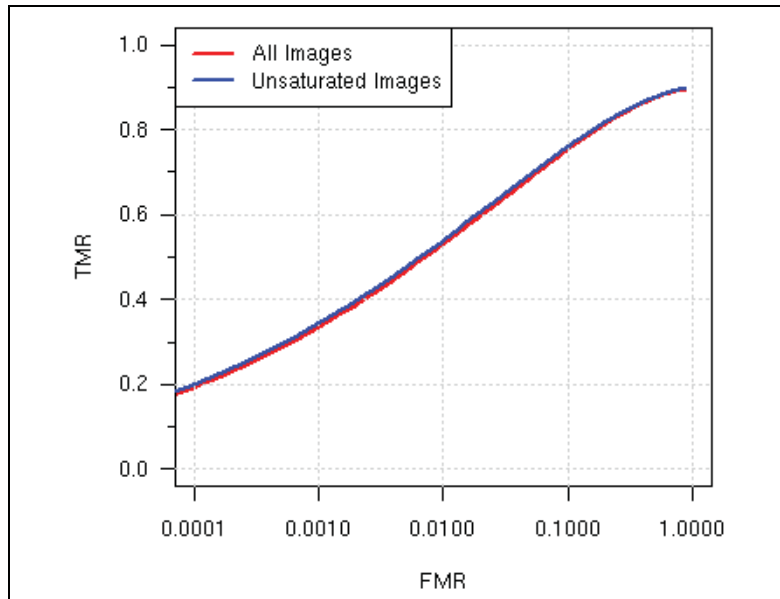


Figure 2 - Effect of manually labeling saturation on matching performance

4.4 Head Pose

Images were manually inspected for non-frontal head poses. The images were assigned one of the following categories:

Code	Description	Fraction (estimated over 10000 images)
FR	Fully frontal, or very close (pitch and yaw are within roughly 5 degrees)	48.87%
PF	Partially Frontal. Not fully frontal, but not catastrophically off (maybe 5-15 degrees of yaw or pitch)	49.51%
NF	Non-frontal. Off by a lot.	1.62%

Table 1 - POE image non-frontal pose categorizations

For each image, its category is assigned by visual inspection by a NIST staff member. This process is clearly not quantitative and any given image might be adjudicated differently by a different judge. Note that head pose is not the same as gaze direction, as a subject can be looking at the camera but still not be fully frontal. This circumstance may arise from an instruction (implied or explicit) from the officer to the traveler to "look at the camera" but for which the response is to adjust only the gaze. This might be due to synchronization also. The ideal response for face recognition should be to orient the head toward the camera.

4.4.1 Effect on automated matching

Figure 3 is included to support the assertion that face recognition will remain difficult in POE primary inspections without improved control of the head pose in relation to the camera. Note however that even for fully frontal images, the recognition process is poor (FMR = 0.01, TMR = 0.6). This is a consequence of poor resolution and illumination, non-ideal compression and of image quality inadequacies in the accompanying BioVisa images (BVA) collected by DOS.

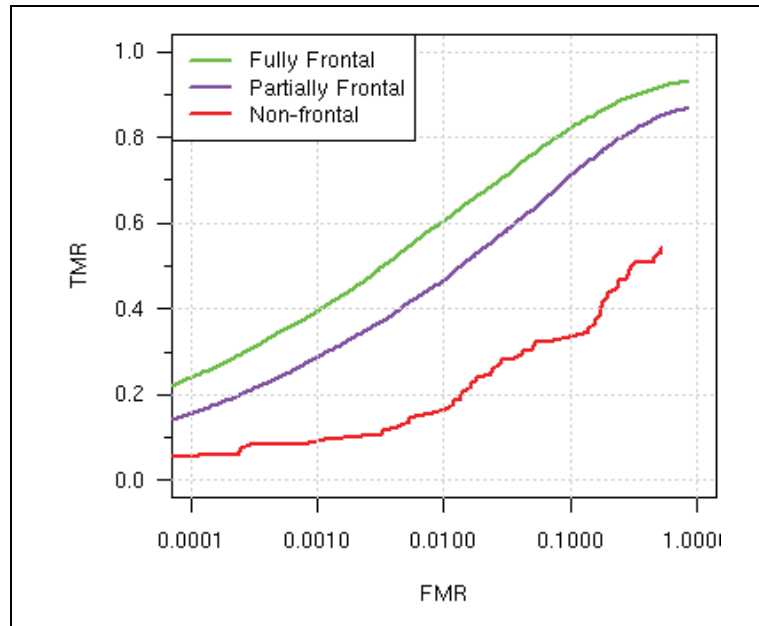


Figure 3 - Effect of manually labeling aberrant poses on matching performance

5 Automated Quality Assessment

5.1 Approach

This section documents the application of a commercial image quality analysis tool to five image databases. This allows new images collected from candidate next generation US-VISIT cameras to be compared subject to the same analyses and compared with the quality statistics reported here.

5.2 The VisPRO Image Quality Analysis Tool

The VisPro SDK does not produce an overall quality value per sample. Rather, it rates the suitability of an image for automated matching using a series of quality metrics. Each metric focuses on a particular feature of the image that is known to affect matching performance. Dividing the quality analysis in this way allows one to identify specific weaknesses in an image. Most of the metrics rate the face image with a score on the range from 1 to 100, with 100 being the best quality. Note this method of scoring can be counter-intuitive in some cases. For example, one metric focuses on the amount of shadow that is present on the face. Initially, one might assume a higher score implies more shadow, but since less shadow is preferable, a face with fewer shadows will receive a higher score.

The VisPro SDK computes the following quality metrics:

Brightness: Measures the brightness of the face in the image. Low values indicate the face is dimly light, while very high values indicate the face is over-exposed to light. A middle range of values is ideal - code from the vendor flagged values between 30 and 80 as acceptable.

Eye Distance: Computes the pixel distance between the eye coordinates. Generally, a greater eye distance is preferred, since FRVT 2006 [3] demonstrated that automated face matchers perform better on higher resolution images.

Eye Shadow: Measures how much shadows in the eye sockets are a problem for the current image. Higher values indicate eye shadows are less of a problem. Eye shadows can be caused by lighting originating from above.

Face Shadow: Measures how much face shadows are a problem for the current image. Higher values indicate face shadows are less of a problem. Face shadows occur if lighting is greater on just one side of the face.

Eye Detection: Describes how confident the VisPro SDK is that the eyes were correctly located. Higher confidences are expected for better quality face images.

Sharpness: Measures the sharpness of the face in the image. A high score indicates the image is sharp and in-focus. Blurry faces can be caused by an out-of-focus camera, or from motion blur.

Contrast: Measures the color contrast of the face. Low values indicate poor contrast. Images with low contrast can appear gray or “foggy”. Too much contrast can also be bad. Values between 50 and 90 are considered acceptable.

Color Balance: Determines if there is a red or blue color bias in the image. A score is output for both colors. A score of zero indicates no bias, while a positive score indicates a bias for that color. Some cameras, such as the Logitech 4000 and 5000 series, white balance on the brightest object in the scene. If the color white is not prominent in the scene, the color balance may be biased.

Background Brightness: Provides a measurement of the brightness of the background. Darker backgrounds without directly visible light sources are preferred. For this reason, brighter backgrounds receive lower values.

Background Consistency: Measures the consistency of the background in the image. Plain backgrounds will receive high scores, while “busy” backgrounds will receive low scores. Consistent backgrounds are preferred.

Background Shadow: Examines the image for visible background shadows. Ideally, no shadows will be present, and the score will be high.

Roll: Measures the roll angle of the face. Deviations in the roll angle can be corrected post-capture, but they can make automated eye finding more difficult.

Yaw: Measures the yaw angle of the face (i.e. left/right movement). Faces looking left receive positive values.

5.3 Image Databases

Five existing databases were used for the initial evaluation. POE images were separated to determine the effect (if any) of recent changes in the capture protocol. The other databases are included to elucidate the natural variation of the various properties.

The following subsections describe the five databases that were used with the VisPRO tool. Table 2 provides an overview, including the important eye detection rate statistics. Typically, the automated eye finder fails to find the eyes if the person’s face is not clearly visible (e.g. it is occluded, non-frontal, or cropped off the image).

Image set	Source	Subset or selection	Number of images / people	Eyes Detection Count	Eyes Detection Rate (%)	Character
POE 2004	DHS US-VISIT	Mates of BVA	10,000 / 10,000	9,414	94.1	Webcam, unconstrained environment often non-frontal, low resolution, poor compression.
POE 2 2006 & 2007	DHS US-VISIT	Random	3,000 / 3,000	2773	92.4	
BVA	DOS	Random	10,000 / 10,000	9,944	99.4	Geometric properties are improved over POE. Frontal pose. Photograph of paper printed photograph. Low resolution, poor compression.
FRVT Mexican NIV	DOS	Random	20,000 / 10,000	19,985	99.8	Geometric properties are improved over POE. Extensively documented in FRVT 2002/2006 [3]. Frontal, no glasses, 70 pixels eye-to-eye. Medium-poor compression.

Image set	Source	Subset or selection	Number of images / people	Eyes Detection Count	Eyes Detection Rate (%)	Character
FERET	NIST Publicly available	Frontal FA and FB series	1,986 / 992	1,981	99.7	The best data set, both geometric and photometric properties are better than POE. Lab research set. Frontal, some lighting variation. Medium-good quality. 120 pixels eye to eye. Uncompressed.

Table 2 - Size and eye-detection rates for the five face databases

5.3.1 POE Databases

POE images were collected at United States ports of entry as part of the US-VISIT program. Of the five databases, the two containing POE images might be characterized as the most problematic for face recognition. The eyes were not found in 5.9% of the 2004 images and 7.6% of the 2006 & 2007 images, due in large part to cropped faces and significant pose deviations. Figure 4 and Figure 5 graph the score distributions for each of the quality metrics for each of the datasets. The graphs highlight many of the problems with the POE images. Some of the notable conclusions are:

Yaw has large variance. The first and third quantiles approximately line up with the ± 5 degree marks, indicating that about half the faces have more than 5 degrees of yaw.

Eye distances are very small compared to the other databases. The median eye distance is only about 50 pixels. FRVT 2006 [3] demonstrated that automated face matchers are more accurate when the eye distance is significantly greater than this.

POE images score poorly for the *Background Consistency* metric. The background is very cluttered in most of the images and frequently includes partially visible faces of other people waiting in line.

Background Brightness values are poorer for POE images. In addition to being cluttered, the background frequently contains lights. Background lights cause the camera to automatically reduce its exposure time, resulting in a dimly lit face. It can also cause the face in the foreground to appear hazy.

POE images score poorly for the *Face Centering* metric. This accurately reflects the fact that POE faces are frequently un-centered, usually as a result of the camera operator not correctly pointing the camera at the individual.

POE images score poorer for the Sharpness metric. Many of the images are blurry, possible due to the head or camera moving at the time of capture.

In addition, metric distributions do not differ appreciably for the POE 2004 and POE 2006 & 2007 databases.

5.3.2 BVA Database

BVA images were collected at Consulates for persons applying for a U.S VISA. These images are somewhat controlled, but still suffer from a variety of problems. The background is always plain, and the face is always centered in the image, but facial expression and lighting still vary somewhat. The most significant problem is that the face texture often appears granular and pixilated. This might occur if the images were heavily compressed.

Figure 4 reveals that BVA images differ from the other images in the following ways:

Color Balance is poor for these images. The graph suggests that the images tend to have a red and blue color bias.

Variations in Yaw are extremely minute.

For most of the metrics, the BVA and Mexican Border Image databases produce very similar distributions, suggesting that the two databases are comparable in terms of overall quality. Compared to the POE database, the BVA database frequently produces a more favorable distribution. This is expected, since BVA images tend to be much better quality.

5.3.3 Mexican Border Images

Mexican border images were provided from the U.S Department of State’s Mexican non-immigrant VISA archive [3]. The images are of good quality and were gathered in a consistent manner. The face is always fully frontal, well-illuminated, and centered in the image. Furthermore, subjects are never wearing glasses. Figure 4—Figure 5 reveals no significant problems with these images. The median eye distance is only about 75 pixels, which might limit the usefulness of the images for automated matching.

5.3.4 FERET Database

The FERET database [4] consists of images collected in a controlled environment. They are generally regarded as good quality, although there are a few characteristics of the images that might reduce matching accuracy. The facial expression sometimes varies for multiple instances of a subject, and some of the subjects are wearing glasses. Compared to the other databases, FERET images have a large eye distance.

Some notable conclusions for FERET images are:

The Eye Detection confidence is high despite the presence of glasses on some of the faces. Glasses may be less of a problem when the eye distance is large.

Background Consistency is highest for FERET images. In truth, the background is fairly consistent for BVA, FRVT and FERET images.

FERET images score well for the Eye Shadow and Face Shadow metrics, suggesting illumination is very uniform across the face.

Brightness and Contrast are lowest for FERET images.

5.4 Results

The following figures give the distribution of the VisPRO quality metric variables for images from each of the five indicated databases.

<p style="text-align: center;">Background Brightness</p>	<p style="text-align: center;">Background Consistency</p>	<p style="text-align: center;">Background Shadow</p>
<p>High variance ⇒ poor.</p> <p>This result is not relevant in the POE context because the images are collected in POEs. The metric is intended for passport or visa applications in which background should be uniform.</p>	<p>Low values ⇒ poor.</p> <p>This result is expected because the FERET used a standard mugshot background. Similarly, because the visa images too are required to be conformant to a DOS specification. The POE images are very poor.</p>	<p>High values ⇒ poor.</p> <p>This result is not relevant in the POE context because the images are collected in POEs. The metric is intended for passport or visa applications in which shadows should not occur on the background.</p>

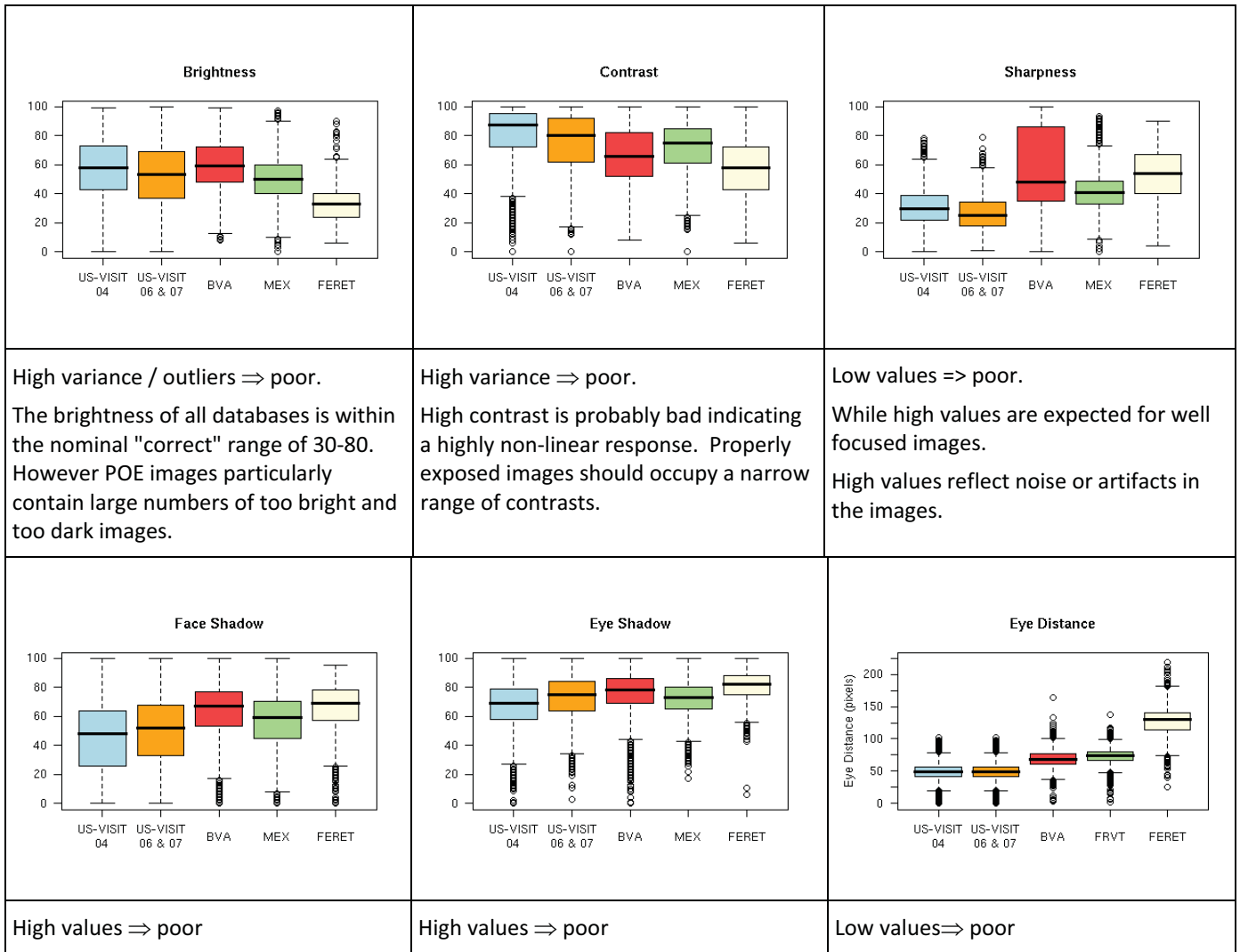


Figure 4 - VisPro Quality metrics

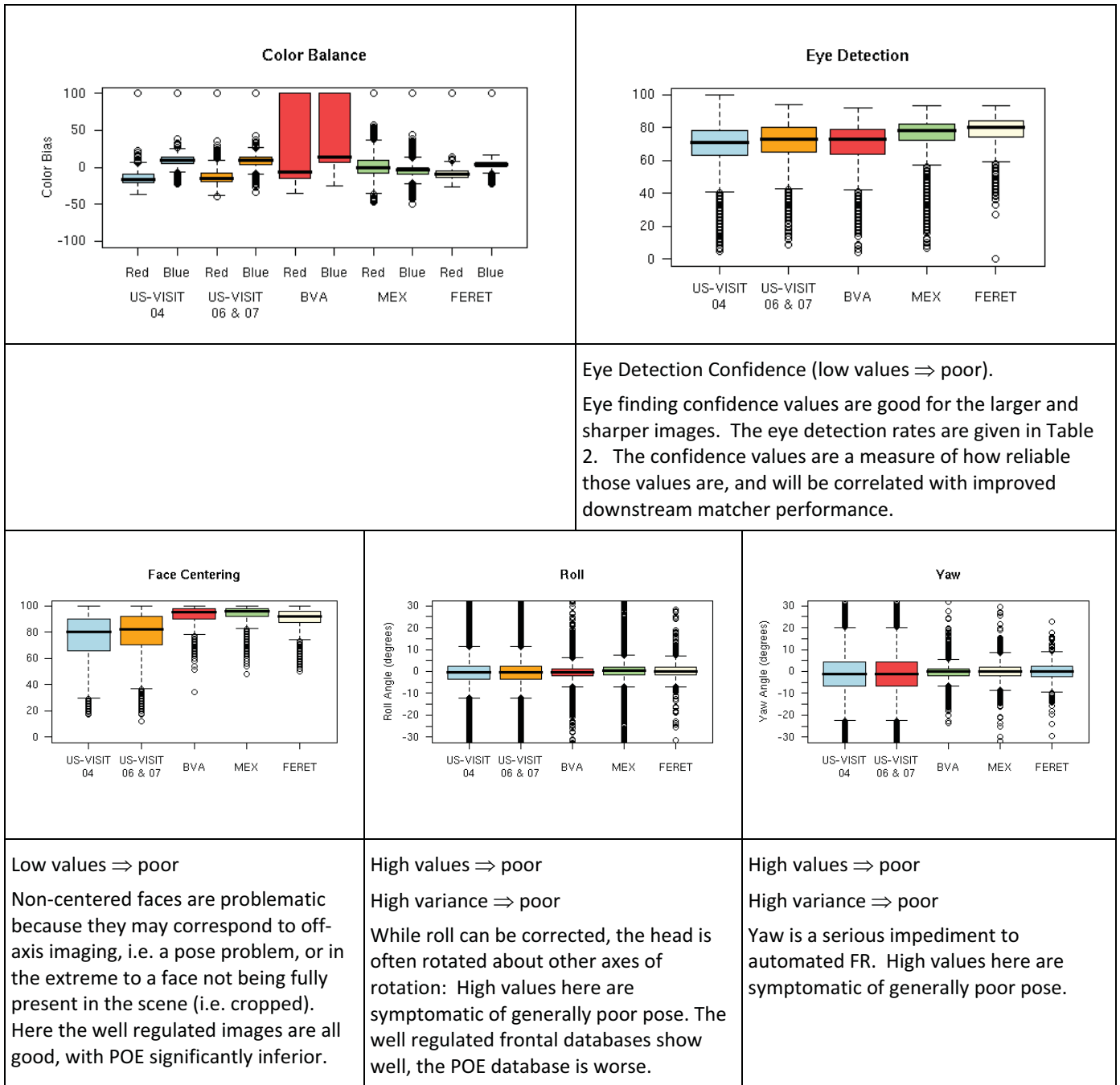


Figure 5 - VisPro Quality metrics (continued)

5.5 Processing Time

The time to call each of the quality metric functions was recorded for 2004 POE images. The results are shown in Figure 6 — Figure 7 Note the different vertical axis for the two figures. The VisPro SDK was run on a Dell PowerEdge 1950 with dual 3.0GHz processors and 4GB of memory. Most functions returned a value almost instantaneously (at

about the timing resolution), indicating the actual computation of those values must have occurred previously, during the eye finding. Computation of the Background Consistency and Background Shadow appear to be the exceptions. Interestingly, it took significantly longer to compute these values for 2004 POE images, which tend to have much “busier” backgrounds.

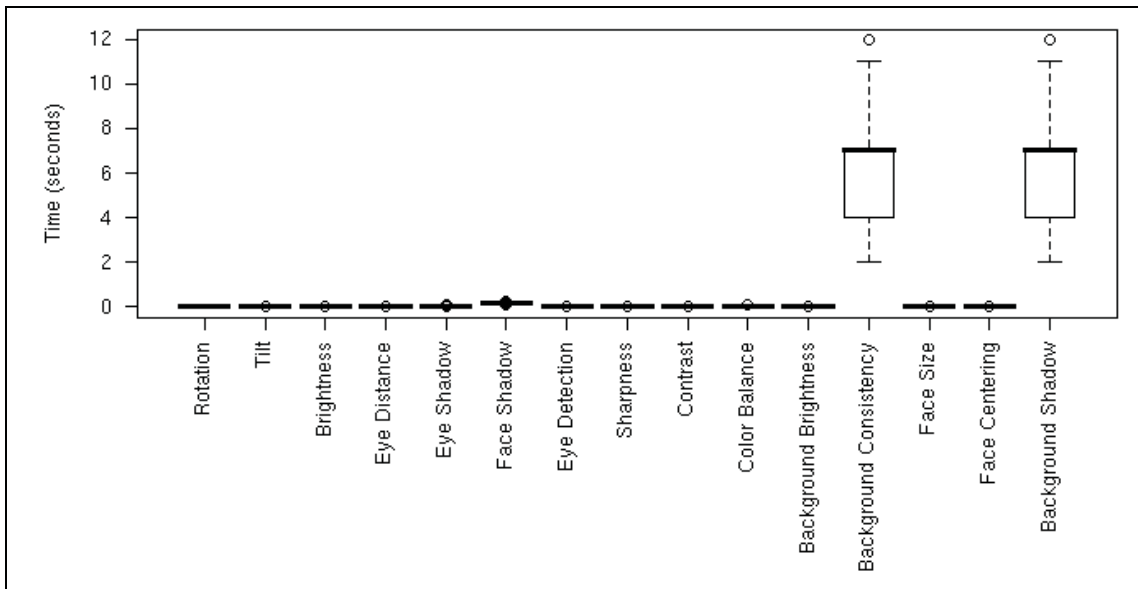


Figure 6 - Processing time for each of the quality metrics for 2004 POE images.

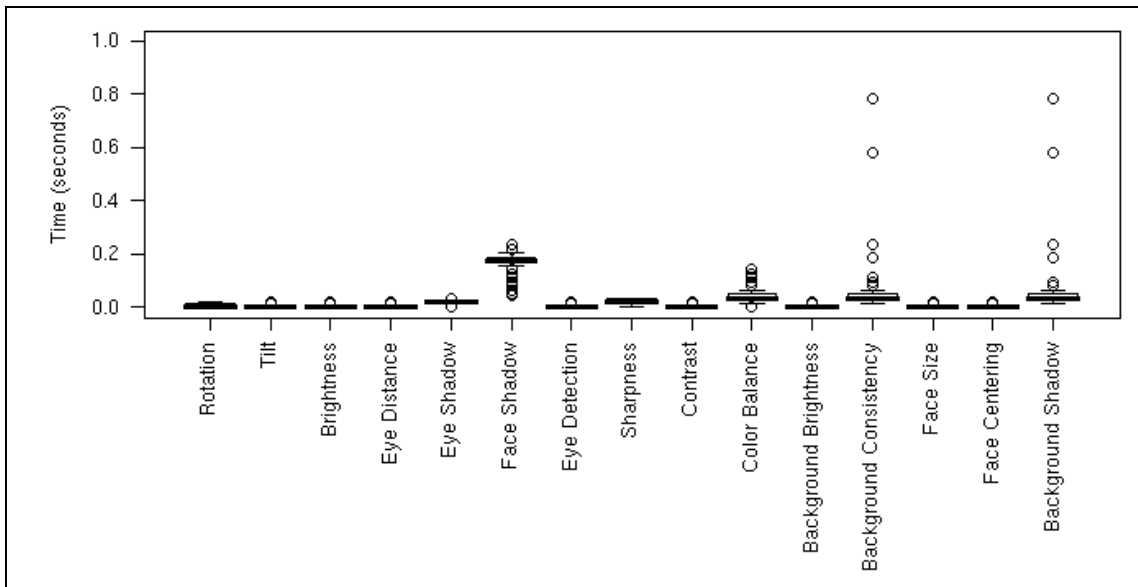


Figure 7 - Processing time for each of the quality metrics for BVA images.

The time it took to call the eye finding function was also recorded. Since it is highly probable that eye finding occurs concurrently with computation of most of the quality metrics, we refer to this time as the image processing time. Figure 8 shows the distribution of image processing times for each dataset. The times appear to correlate loosely with image size. The pixel dimensions are 240x240 for POE images, 252x300 for BVA images, 252x300 for FRVT Mexican images, and 512x768 for FERET images.

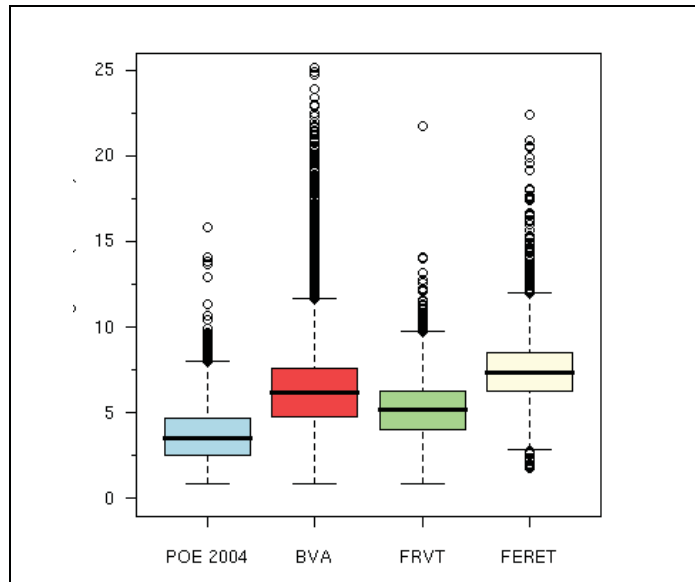


Figure 8 - Processing time per image for each dataset.

5.6 Yaw Estimates

The accuracy of VisPro’s yaw measurements is evaluated by comparing the estimated yaw to the actual yaw for FERET B-series images. The B-series images are intentionally captured at several predetermined yaw angles and were originally used in FRVT 2000 to investigate the effect of pose angle on face recognition performance. Accurate pose estimates are important not only for quality assessment, but also for performing pose correction. One method of correcting the pose after capture is to fit a 3D morphological model to the face to generate a fully frontal view from the off-angle face. Knowledge of the precise head pose can assist with the morphing process.

Table 3 outlines the properties of the FERET B-series sets. Some eye finding statistics are also provided. Each B-series set was generated with the same 200 subjects, and all images of a subject were collected during the same session. The B-series images were only included in the earlier grayscale release of the FERET Database. These images are similar to the color FERET images, except they are grayscale and half the size (256x384). The images were collected by instructing the subject to face straight ahead while the camera was positioned at different points to the subject’s left or right. A less accurate method that is sometimes employed is to position the camera directly in front of the subject, and have the subject look at different points on the wall. This could lead to unintended variations in the face pose, since people do not always look where they are facing.

Set	Yaw Angle (positive ⇒ subject faces left of camera)	Number of images / people	Eyes Detection Rate
BA	0	200 / 200	200 (100%)
BB	+60	200 / 200	191 (95.5%)
BC	+40	200 / 200	196 (98.0%)
BD	+25	200 / 200	198 (99.0%)
BE	+15	200 / 200	200 (100%)
BF	-15	200 / 200	200 (100%)
BG	-25	200 / 200	200 (100%)
BH	-40	200 / 200	199 (99.5%)
BI	-60	200 / 200	198 (99.0%)

Table 3 - Overview of FERET B-series sets.

Figure 9 shows the distribution of computed yaw angles for each set. It serves as a calibration in that it plots estimates vs. ground truth. The median estimated yaw for sets BB through BJ is only about half the actual yaw, indicating the VisPro SDK has a propensity to underestimate the angle by a factor of 2. NIST does not believe this is an error in the FERET metadata, which has been very widely studied. Multiplying the computed yaw angle by a constant could mitigate the problem, since the estimated yaw appears to have a direct linear relationship to the actual yaw.

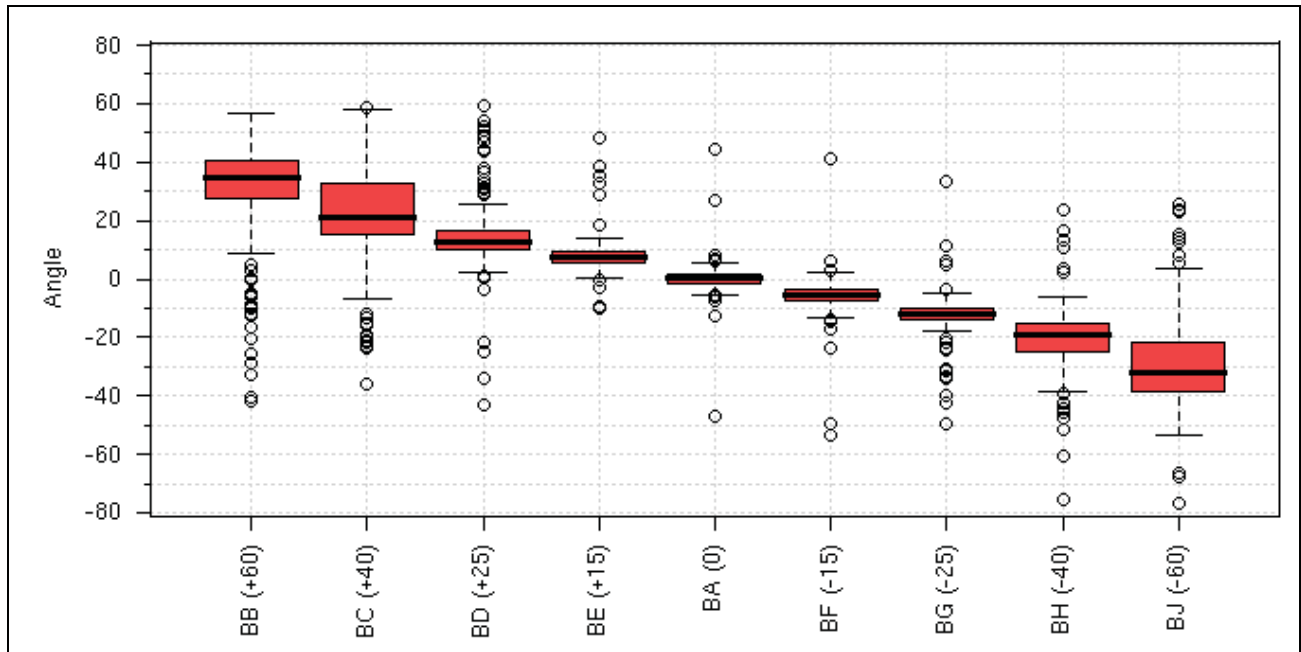


Figure 9 - VisPRO Yaw angle estimates for the FERET b-series image sets.

5.7 Image Quality at Different Ports of Entry

The 2004 set of POE images was accompanied by the port of entry identifiers for each image. This supports comparison of image properties between border crossing sites. For the current analysis we selected the eighteen busiest POEs which primarily are the busier terminals in the nation's largest airports. For each, we randomly sampled 3000 images and applied the VisPro SDK to quantify quality.

Figure 10 shows the resulting score distributions for six quality metrics by port of entry. Five of the metrics (Face Shadow, Face Centering, Background Consistency, Background Brightness, and Face Brightness) were selected because they measure what we regard to be the most significant quality problems with the images. In addition, eye confidence was selected because it provides a general measure of the quality of an image (the eye confidence is expected to be lower for poorer quality images²). We use this variable to establish an ordering of the POEs. That is, from left to right the POEs appear in increasing order of the median eye confidence.

² Successful automated facial recognition is critically dependent on the accurate and consistent localization of landmarks, primarily the eye centers.

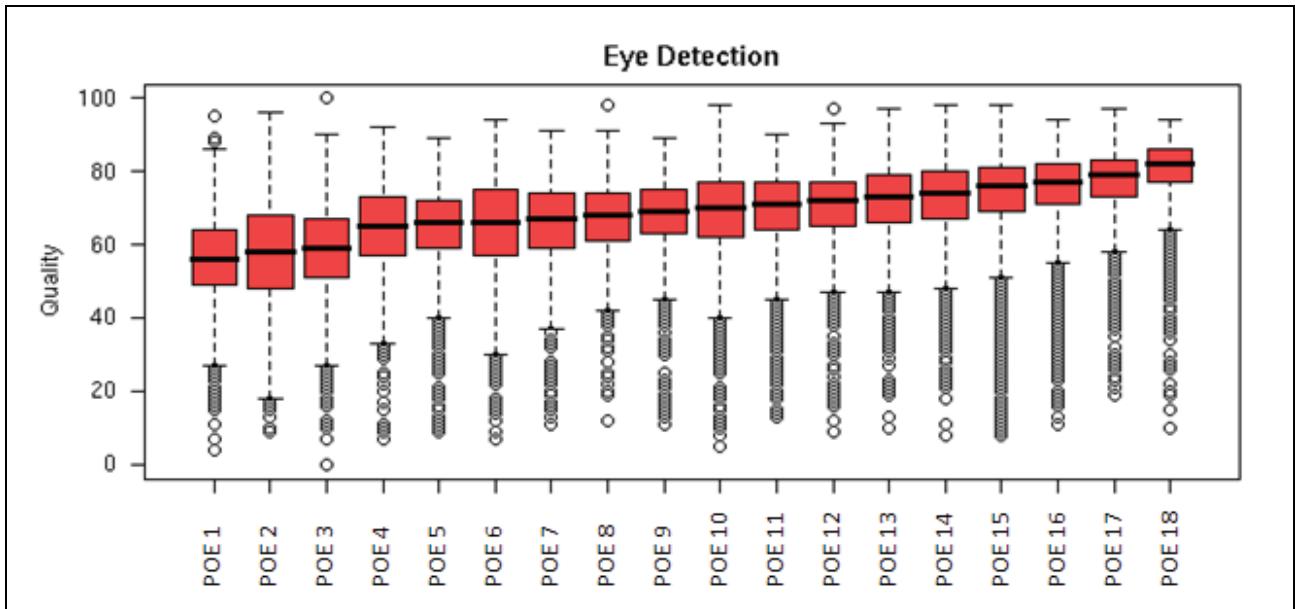


Figure 10 - Variation in eye detection confidence across POEs

5.7.1 Results

Figure 10 shows that eye finding is significantly easier in the images of some ports of entry (e.g. POE18) than for others (e.g. POE1). This holds only to the extent that the variable reported, eye confidence, is a reliable indicator of actual detection of the eyes. In any case, it gives a consistent ordering for the POEs indicated in the remaining plots.

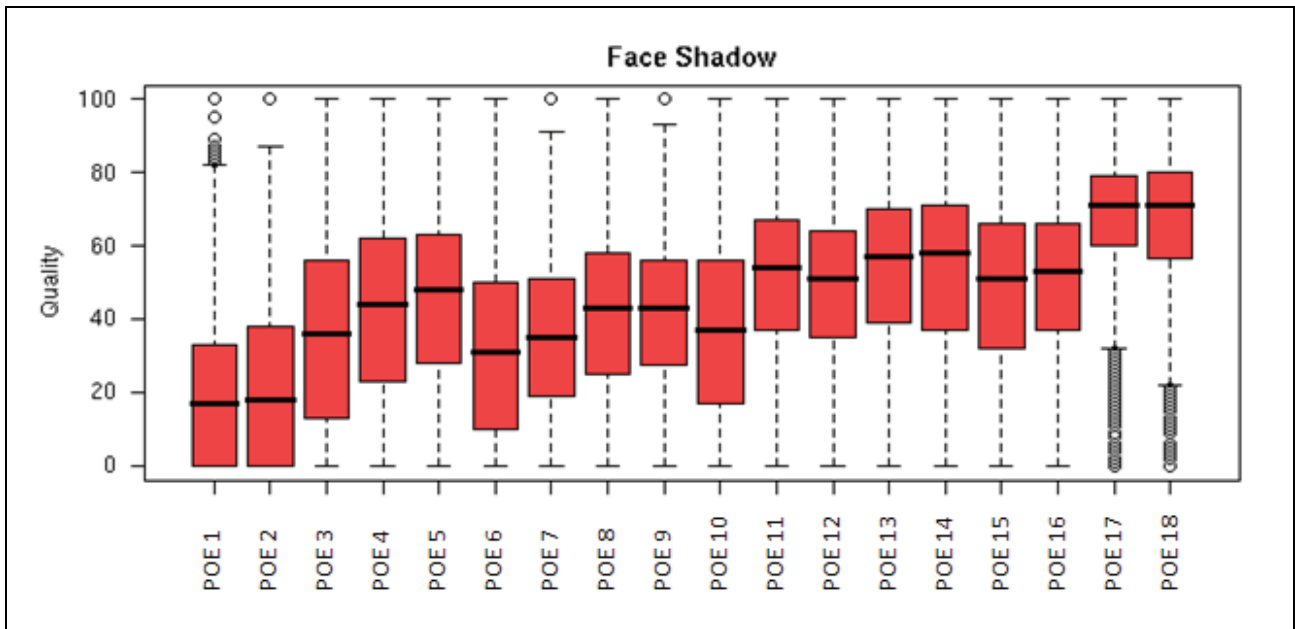


Figure 11 - Variation in face shadow measures across POEs

The trend of Figure 11 shows a positive correlation of reported face shadow with the eye detection confidence. This would be expected since shadows inhibit accurate localization of the eyes. The best and worst POEs are the same as for eye confidence. The variance of the distributions of the face shadow quantity is larger than for eye detection. This may well be due to the fact that a small pose variation can produce shadows while not affecting eye detection.

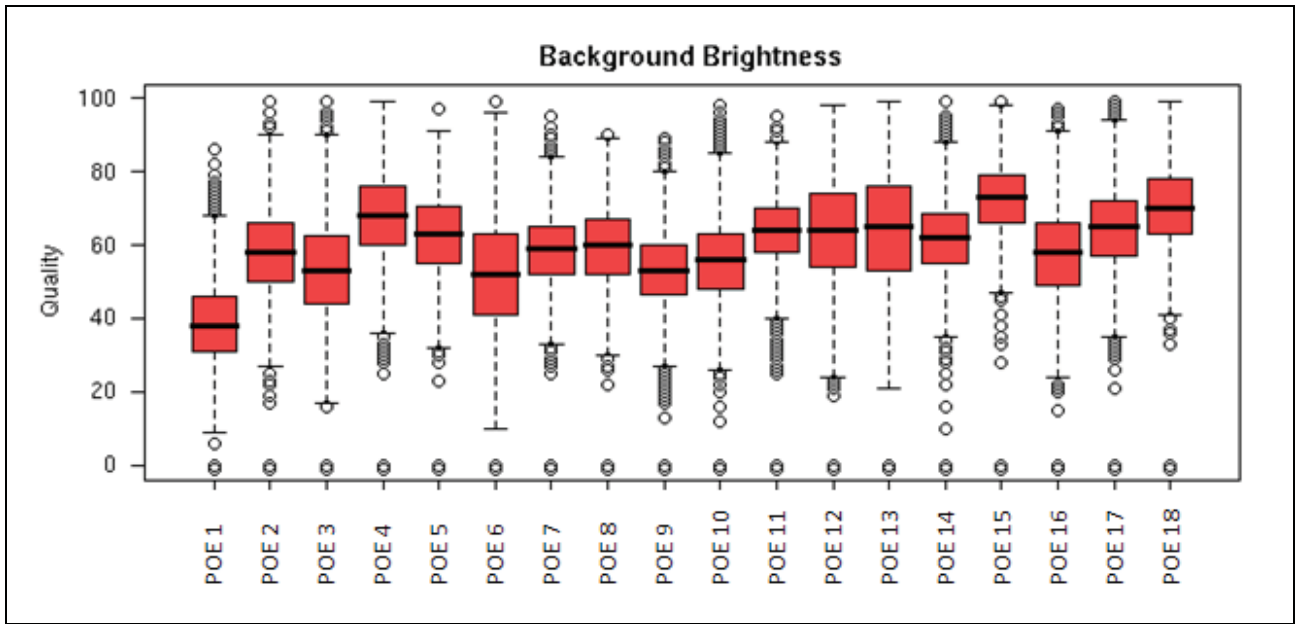


Figure 12 - Variation in background brightness measures across POEs

Figure 12 shows a more varied picture. While a trend is present, the low variance but wide variation means that the background brightness measure is a characteristic of the POE itself. It is less correlated with eye detection confidence because the ability to detect eyes is not related to the background brightness if the face itself is properly exposed.

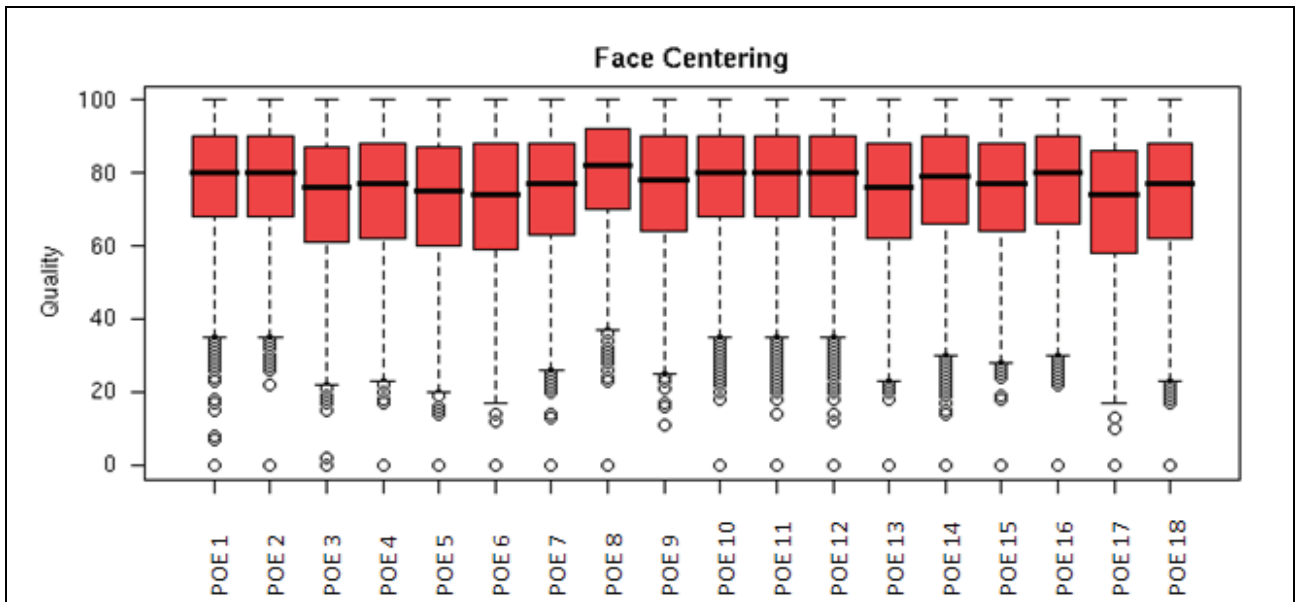


Figure 13 - Variation in the face centering measure across POEs

In comparison to the other variables, Figure 13 shows considerable consistency across POEs. The value in question, face centering, should be a property of the way CBP officers aim the cameras, and of how travelers respond to the instruction. In addition, eye detection is largely independent of how well centered the faces are (as long as the face is not cropped).

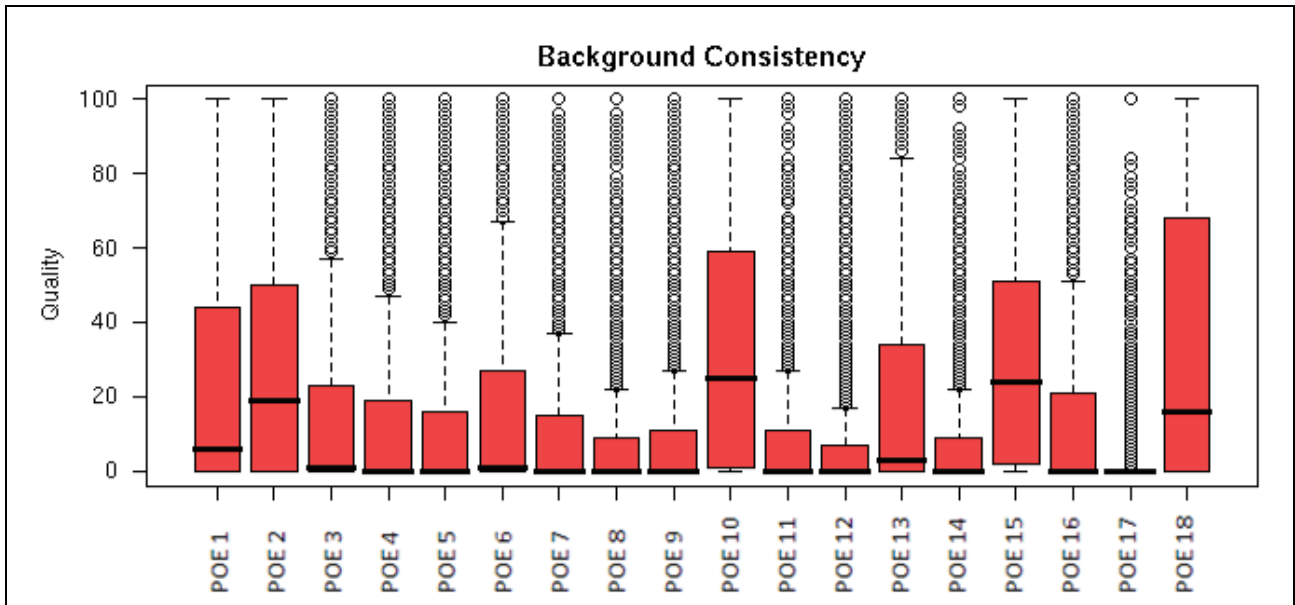


Figure 14 - Variation in the background consistency measure across POEs

Figure 14 shows poor background consistency. This is entirely consistent with the unconstrained nature of POEs.

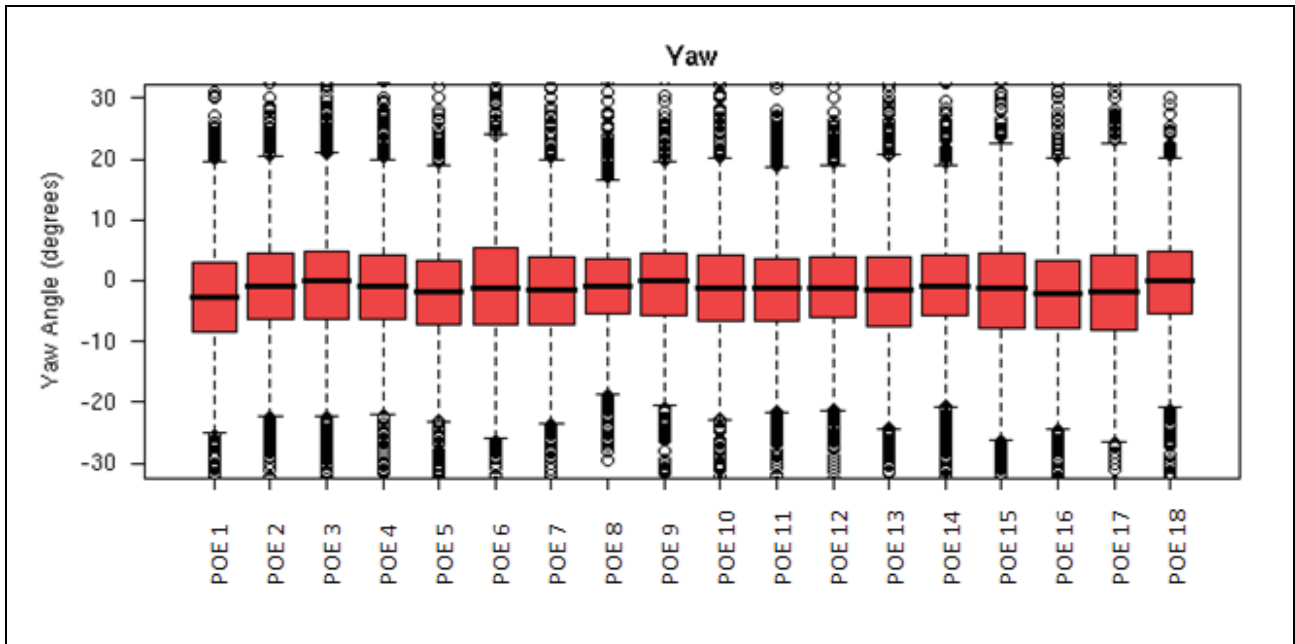


Figure 15 - Variation in the yaw pose estimate across POEs

Finally Figure 15 shows that the yaw estimate is uniform across POEs. This is consistent with the known observation that many images are non-frontal and that the effect is largely a result of the precise officer-traveler interaction and synchronization. Interestingly there is a small negative bias toward the yaw angle. This would imply that subjects have a tendency toward one side or another.

6 Objectives for Next Generation Cameras

The next generation of face cameras in US-VISIT deployment should achieve superior performance. This can be measured in terms of the metrics indicated in this report. In particular, for the VisPRO metrics:

The eye confidence values should improve;

The face shadow values should improve;

The face centering values should improve;

The background brightness might reasonably remain unchanged (absent architectural changes to POEs);

The background consistency might reasonably remain unchanged (again, absent renovation of the POEs);

The *variance* in yaw angle should decrease from that reported here.

7 References

[1] Patrick Grother, NIST, *Overview of DHS POE Images*, April 2004, PowerPoint deliverable to DHS US-VISIT.

[2] Larry Nadel, *Approaches to Face Image Capture at US-VISIT Ports of Entry*, November 2007, in Proceedings of the Second NIST Biometric Quality Workshop <http://biometrics.nist.gov/quality/workshop07/presentations.html>

[3] Phillips Jonathan, Scruggs Todd, O-Toole Alice, Flynn Patrick, Bowyer Kevin, Schott Cathy, Sharpe Matthew, NIST, *FRVT 2006 and ICE 2006 Large-Scale Results*. NIST 7408, <http://www.frvt.org/FRVT2006/docs/FRVT2006andICE2006LargeScaleReport.pdf>


[4] P. Jonathon Phillips, Harry Wechsler, Jeffrey Huang, Patrick J. Rauss, *The FERET database and evaluation procedure for face-recognition algorithms*. Image Vision Comput. 16(5): 295-306 (1998).

Attachment 2: Facial Image Quality Improvement and Face Recognition Study—Camera Pre-Assessment

The US-VISIT Facial Image Quality Improvement and Face Recognition Study (FIQIFRS) project included investigation of hardware and software approaches to facilitate acquiring compliant images in as expeditious a manner as possible and with minimal Customs and Border Protection (CBP) Officer involvement in positioning the camera and making image quality determinations. The initial step for the project involved conducting a market survey to identify available cameras and sensors from various categories for use in port of entry (POE) environments. The working group (WG) identified desired features, determined product categories, identified and surveyed commercial-off-the-shelf (COTS) products in each category and reviewed their specifications, selected categories from which to draw products, and procured representative products from those categories. The camera categories examined were: webcam, digital still camera, digital camera in video mode, industrial video, pan-tilt-zoom (PTZ) video, wide dynamic range video, and smart camera. The cameras depicted in Table 1 were selected and evaluated with respect to ISO/IEC 19794-5:2005, Information Technology — Biometric data interchange formats — Part 5: Face image data (FACESTD) compliance by imaging faces and test targets (Figure 1) in an optimal environment and measuring their characteristics. This document describes the methodology followed for the camera pre-assessment.

Table 1: Cameras tested

Camera	Picture	Mount	Size (WxHxD, in.)	Connection	Software
Logitech QuickCam Pro 5000		Monitor clip; no threaded hole	2.5 x 2.5 x 2.5 (without clip)	USB	Drivers
Logitech QuickCam Pro 9000		Monitor clip; no threaded hole	3.5 x 1.5 x 1.5 (without clip)	USB	Drivers
Logitech QuickCam Orbit AF		Flat base; no threaded hole	3.25 x 4.25 x 3.25	USB	drivers
Sony EVID70		threaded hole	5.25 x 5.75 x 5.75	S-video, requires capture card; power adaptor; VISCA RS-232 camera control	software for camera settings

Camera	Picture	Mount	Size (WxHxD, in.)	Connection	Software
Canon G9		threaded hole	4.19 x 2.83 x 1.67	USB; power adaptor or battery	SDK
Wide dynamic range	Not pictured	threaded hole	1.7 x 1.8 x 2.75	BNC connection, requires capture card; power adaptor, USB for camera settings	SDK

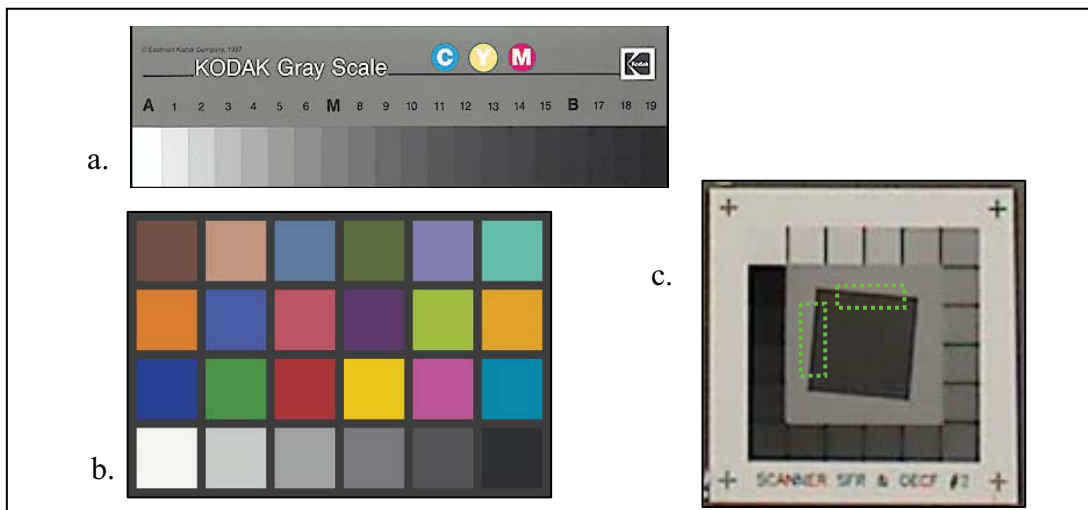


Figure 1: Image test patterns: a) Kodak Q13 grayscale test pattern; b) GretagMacbeth ColorChecker with reference map; c) ISO 16067-1 with slant edge regions of interest

1 Capture Environment

Although the POE environments are largely uncontrolled with variable lighting, backgrounds, and subject distances and heights, the capture environment for the camera pre-assessment was modeled to be as close to optimal as possible in order to evaluate cameras objectively. Each camera's highest sampling frequency and lowest compression ratios were employed. In order to accommodate the largest variation in subject heights, the cameras were operated at their maximum wide angle (lowest focal length, no zoom employed).

The background was a plain, off-white wall. Illumination consisted of two 500 Watt floodlights (1000 Watt total) with softbox diffusers (see Figure 2a), positioned at approximately 45 degree angles on either side of the camera-to-subject line.

Due to the narrow width of some POE lanes, the typical camera-to-subject distance is less than ideal (sometimes only a few inches). According to ISO 19794-5 Amendment 1: *Conditions for taking photographs for face image data*, sub-clause B.2.1.1,

"the camera-to-subject distance should be within the range of 1.2 to 2.5 m. Arranging the lighting without creating shadows will likely be difficult if the camera is placed any closer to the subject."

For the capture environment, a camera-to-subject distance of 70 cm was chosen, because it is the minimum distance specified in ISO 19794-5, sub-clause B.2.2.2, *Example Configuration for a Photo Booth* (see Figure 2b).

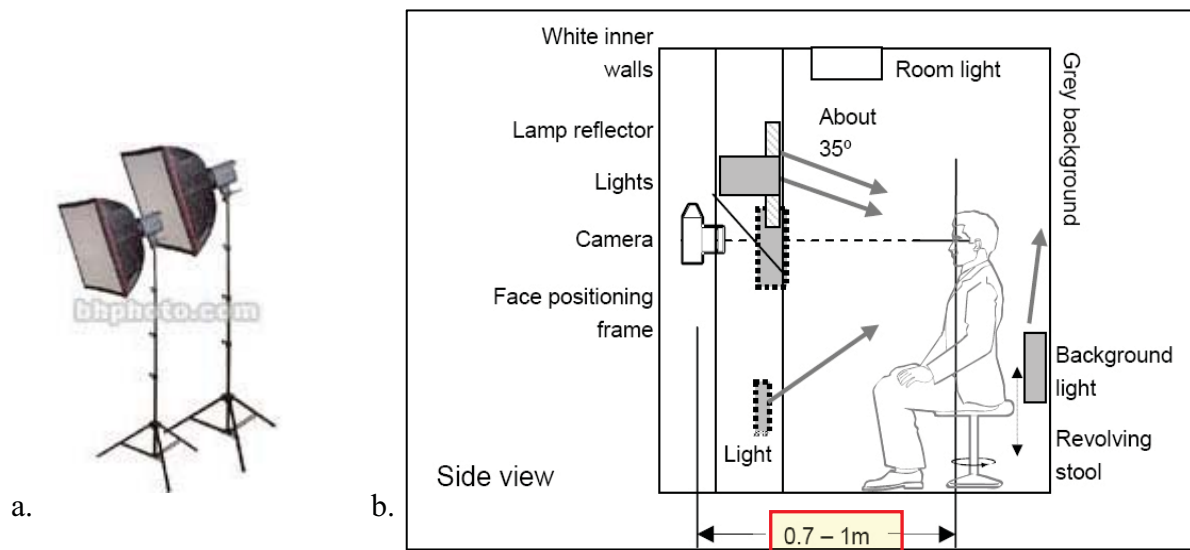


Figure 2a) Lamps, and b) ISO 19794-5, sub-clause B.2.2.2, Example Configuration for a Photo Booth

2 Assessment Results and Findings

The cameras depicted in Table 1 were evaluated for a number of factors that impact image quality and usability in a POE.

2.1 Physical Characteristics

Suitability of a type of camera for POEs depends on convenience and ease of use as well as image quality, which must be adequate for face recognition (FR) by humans and automated systems. The size of the camera, as well as the type of mount and connections, affect a camera's suitability to the POE environment. The camera must be mountable on a universal camera mount, such as a tripod or pole with a screw. The camera must be theft-proof, and hence, it should not have any removable parts. The size of the camera should be compact, so as not to obscure the CBP inspector's view of the traveler. Because the space in POE inspection stations is limited, it is desirable to minimize the number of camera connector cables and external hardware required to operate the camera. Another important camera characteristic is the provision, by its manufacturer, of software for image capture and for setting camera parameters (e.g., exposure, zoom).

2.1.1 Size

The size of the cameras evaluated for the project, as well as the type of mount and connections are listed in Table 1. The smallest cameras in the study were the Logitech QuickCam Pro 5000 and 9000 webcams, while the Sony EVID70 was the largest camera. None of the cameras were considered too large for the POE environment.

2.1.2 Mount

The Canon, wide dynamic range, and Sony cameras all provided a threaded hole for mounting on a standard tripod or other camera mount. The Logitech QuickCam Pro 5000 and 9000 webcams were the most difficult to mount, because they were designed to clip atop a monitor, and, therefore, do not have a threaded hole. The Logitech QuickCam Orbit AF does not have a threaded hole; however, its flat base was easily glued onto a wooden block into which a threaded hole was created for mounting onto a tripod. Unfortunately, the globe is removable from the base, and hence may be vulnerable to theft.

2.1.3 Connection

It is desirable to minimize the number of camera connector cables and external hardware in the POE environment for ease of use and space conservation. The Logitech webcams require only a USB connection from the computer to the camera. The rest of the cameras required power adaptors as well as camera-to-computer connectors. In addition, the Sony EVID70 and the wide dynamic range camera required the installation of capture card hardware on the computer in order to save video or video frames. A third cable was required for the wide dynamic range camera and Sony EVID70, for remote adjustment of camera settings, although the wide dynamic range camera did have on-camera buttons with the same functionality.

2.1.4 Software

Software drivers were provided by Logitech for the webcams, enabling video and frame capture as well as limited adjustment of camera settings. Digital point-and-shoot cameras require software in order to capture images remotely with a computer (rather than saving the images to the camera's memory card). Canon provided a software development kit (SDK) for the G9 camera, which consisted of a set of application programming interfaces, dynamic-link libraries, and static link libraries that provide an interface for accessing the camera, changing camera settings, and retrieving data from the camera.

A video capture card and its accompanying software were purchased and installed to save the analog output from the video cameras. Software and cables were also provided with the video cameras. The wide dynamic range video camera had SDK support, but the Sony did not.

2.2 Visual Assessment

Example images of a human face were captured from the selected cameras in an ideal test capture environment. These examples are shown in Figure 3 - Figure 6 along with enlarged regions of the subject's eye for observing detail. The Canon G9 image is visually superior to the images from other cameras in its color and fine detail, even at its medium capture dimensions, as evidenced by the enlarged eyes in Figure 3a-d. The Sony EVID70 (Figure 4a), Logitech QuickCam Pro 5000 (Figure 4b), and wide dynamic range camera (Figure 6) images exhibited much lower resolution, as similar fine details of the eye cannot be discerned. The Logitech QuickCam Pro 5000 and wide dynamic range images contain visible Joint Photographic Experts Group (JPEG) compression artifacts (blockiness). Although the images from the Logitech QuickCam Pro 9000 (Figure 5a) and Orbit AF (Figure 5b) cameras (which use the same sensor) appear slightly washed out, they have fairly good resolution and color fidelity. The image from the wide dynamic range camera appears to have very poor color fidelity.



Figure 3: Example CanonG9 image with enlarged eye captured at (a) 4000 x 3000 pixels, (b) 3264 x 2448 pixels, (c) 2592 x 1944 pixels, and (d) 1600 x 1200 pixels

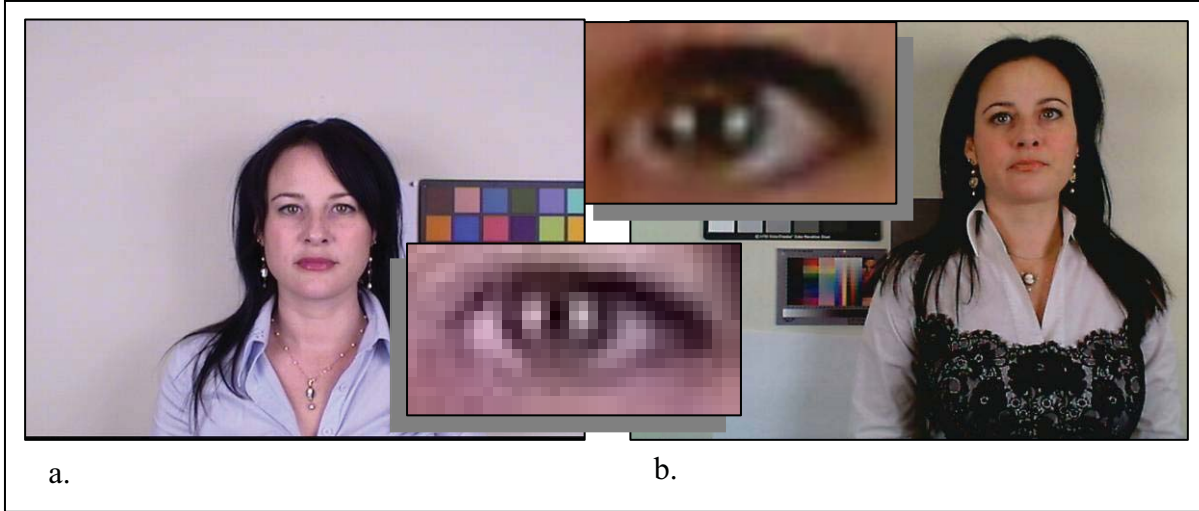


Figure 4: Example Sony EVID70 (a) and Logitech QuickCam Pro 5000 (b) images

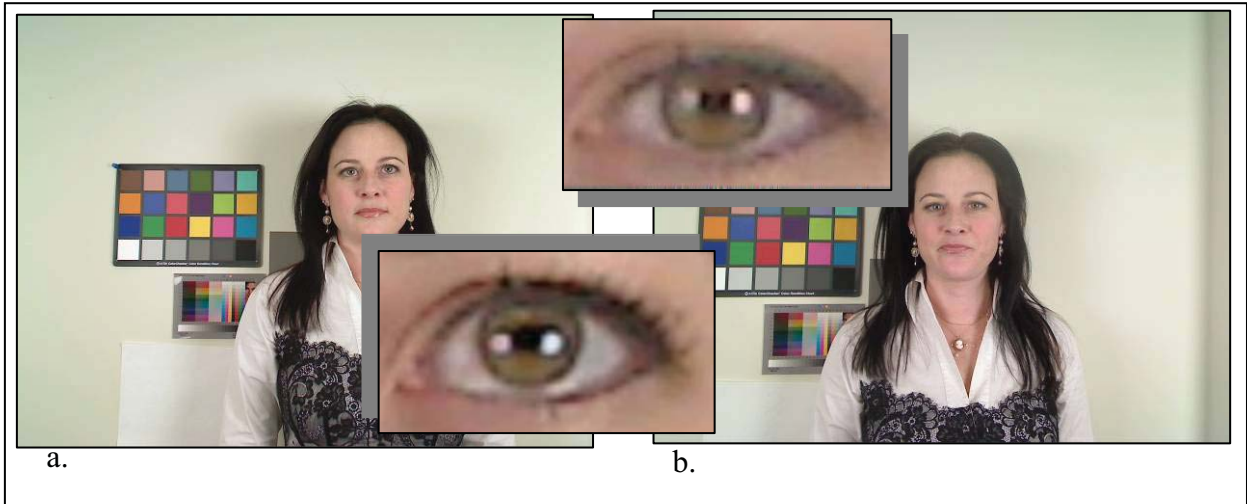


Figure 5: Example Logitech QuickCam Pro 9000 (a) and Orbit AF (b) images

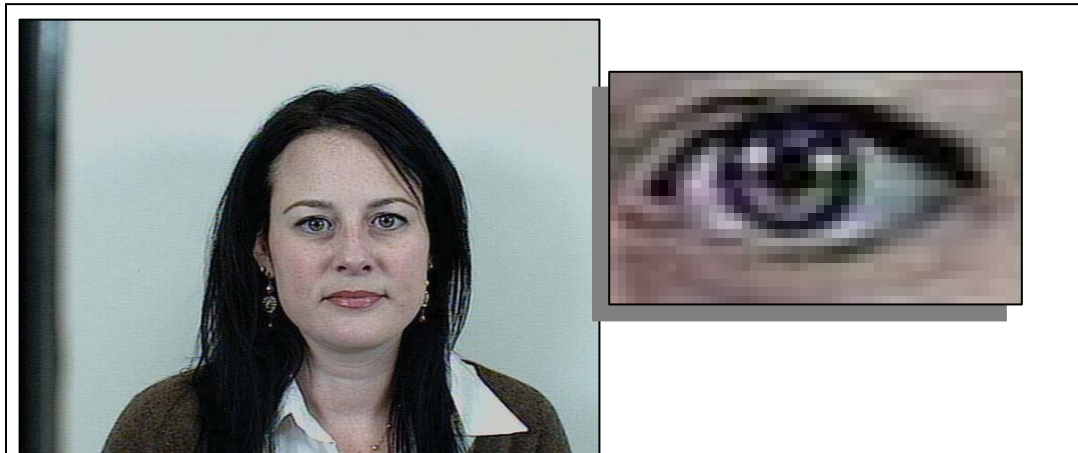


Figure 6: Example prototype wide dynamic range camera image (note: smaller field of view)

2.3 Capture Dimensions

Table 2 summarizes the capture dimensions, frame rates, compression ratios, and fields of view for each camera, with the highest values in green and the lowest in red. The Canon G9 allowed for the highest capture dimensions (12 megapixels) and inter-eye distance, as well as the finest sampling frequency. The Logitech QuickCam Pro 9000 and Orbit AF had the next highest capture dimensions (2 megapixels) and a sufficiently high inter-eye distance. All other test cameras failed to comply with the FACESTD required inter-eye distance of 90 pixels¹.

2.4 Compression

FACESTD, A.3.3, Photo Resolution (Informative) specifies that a 416 x 536 pixel face image scanned at 300 pixels per inch should have a file size no smaller than 11 KB. The corresponding compression ratio of ~60:1 is quite high, given that newer FR systems use skin texture, and such high ratios typically produce JPEG blockiness. A more appropriate compression ratio would be ~20:1. At this ratio, no compression artifacts should be visible.

The Canon G9 camera had four levels of compression ranging from raw (uncompressed) to normal (moderate compression). One shortcoming of the Logitech webcams was that the compression level is fixed, and the webcams had the highest compression ratios found in this evaluation. All cameras evaluated had an acceptable compression ratio of less than 20:1.

Table 2: Capture Dimensions, Frame Rate, Compression, and Field of View

Camera	Capture Dimensions (px., WxH)	Inter-eye distance (px.)	Frame Rate (frames per sec.)	Com-pression	Sampling Frequency (mm/px.)	Field of View Size (in., WxH)	Field of View Area (in.2)	Head Lengths
Canon Powershot G9	4000 x 3000*	325	N/A	Normal, Fine, Super-Fine, RAW	0.2	31.5 x 23.6*	744	3.5*
	3264 x 2448*	258			0.24	30.7 x 23*	706.1	3.4*
	2592 x 1944*	207			0.3	30.6 x 22.9*	700.7	3.4*
	1600 x 1200*	127			0.49	30.9 x 23.2*	716.9	3.4*
	640 x 480*	51			1.23	31 x 23.2*	719.2	3.4*
Sony EVID70	640 x 480	58	30	~4:1	0.98	24.7 x 18.5	457.3	2.1
Logitech QuickCam Pro 5000	640 x 480	55	30	~11:1	1.04	26.2 x 19.6	514.8	2.2
Logitech QuickCam Pro 9000	1600 x 1200*	110	30	~13:1	0.56	35.3 x 26.5*	933.3	3.9*
Logitech QuickCam Orbit AF	1600 x 1200	111	30	~15:1	0.52	32.7 x 24.6	804.7	2.7

¹ FACESTD, 8.4.1 – Resolution (Normative)

* Head lengths when operated in portrait mode

* Can also be operated in portrait mode, in which case the width and height values are reversed.

Camera	Capture Dimensions (px., WxH)	Inter-eye distance (px.)	Frame Rate (frames per sec.)	Com-pression	Sampling Frequency (mm/px.)	Field of View Size (in., WxH)	Field of View Area (in.2)	Head Lengths
Prototype Wide Dynamic Range	640 x 480	82	30	~3.7:1	0.687	17.3 x 13	224.8	1.4

2.5 Field of View

A camera's field of view depends on the lens, the camera-to-subject distance, and the zoom level. The field of view determines the variation of subject heights that can be accommodated without adjusting the camera height. The fields of view in Table 2 were measured for the tested cameras with the lenses at their widest angle available (no zoom), and at a distance of 70 cm from the subject. The Canon and Logitech 9000 were the only cameras in this test that could be operated in portrait mode (where the longest dimension is the height as opposed to the width), which is advantageous for POEs due to the wide variation in subject heights. When operated in portrait mode, the Logitech 9000 and the Canon G9 had the largest vertical fields of view, with heights of approximately 3.9 and 3.5 head-lengths², respectively, followed by, in descending order, the Orbit AF, Logitech 5000, and Sony EVID70, with the wide dynamic range camera having the smallest field of view, with a height of only 1.4 head-lengths. It should be noted that the lens, which affects the field of view, is interchangeable on the wide dynamic range camera, but not on any of the other cameras evaluated herein. The relative fields of view for each camera are diagrammed in Figure 7 (with the Canon G9 and Logitech 9000 in portrait mode).

² A head length is defined as the number of heads that can fit in height of the image, where the head height is assumed to be nine inches from chin to crown.

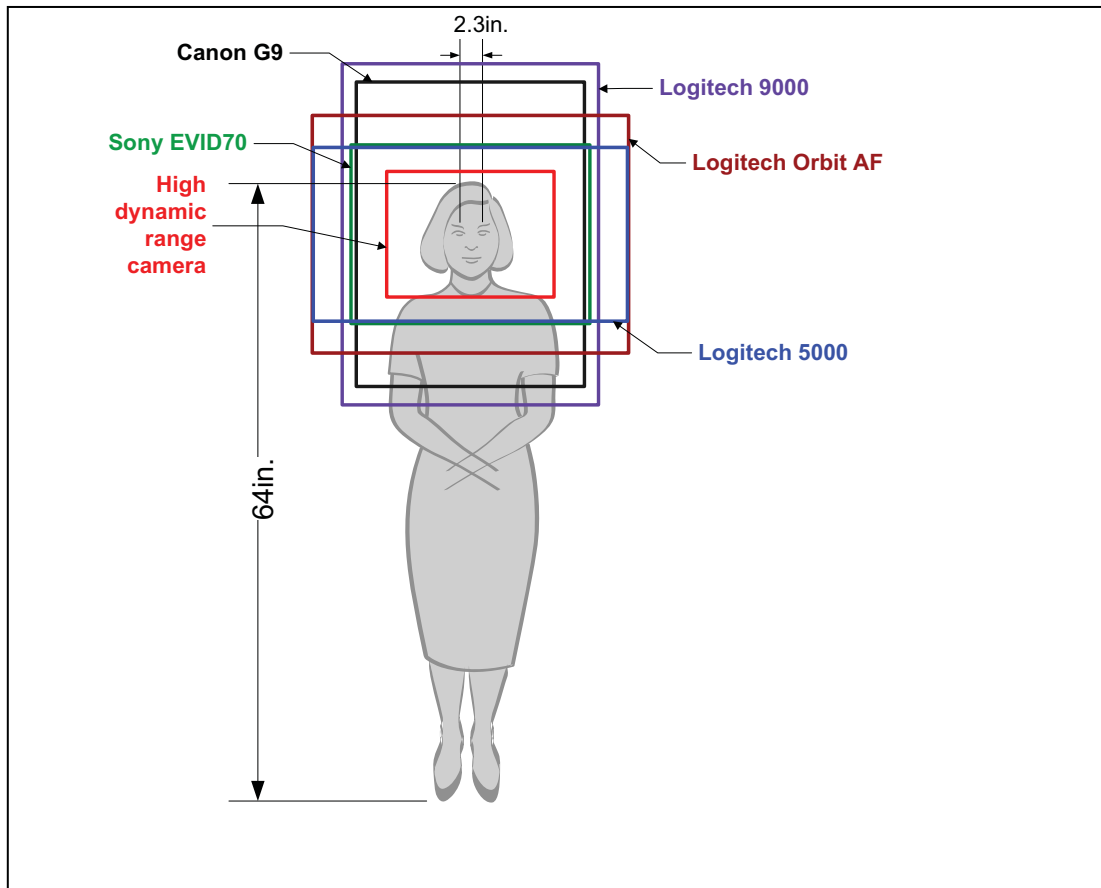


Figure 7: Cameras' fields of view

2.6 Geometric Accuracy

In order to measure geometric accuracy, which refers to the amount of radial lens distortion, a square grid pattern was photographed with each camera and run through Imatest's "Distortion" module. Radial lens distortion is an aberration that causes straight lines to curve. It tends to be most serious in extreme wide angle, telephoto, and zoom lenses; wide-angle lenses are used in POE environments because of the need to accommodate varying subject heights. Figure 8—Figure 13 show each camera's image of the "Distortion" grid, with arrows illustrating the change in radius when the distortion is corrected.

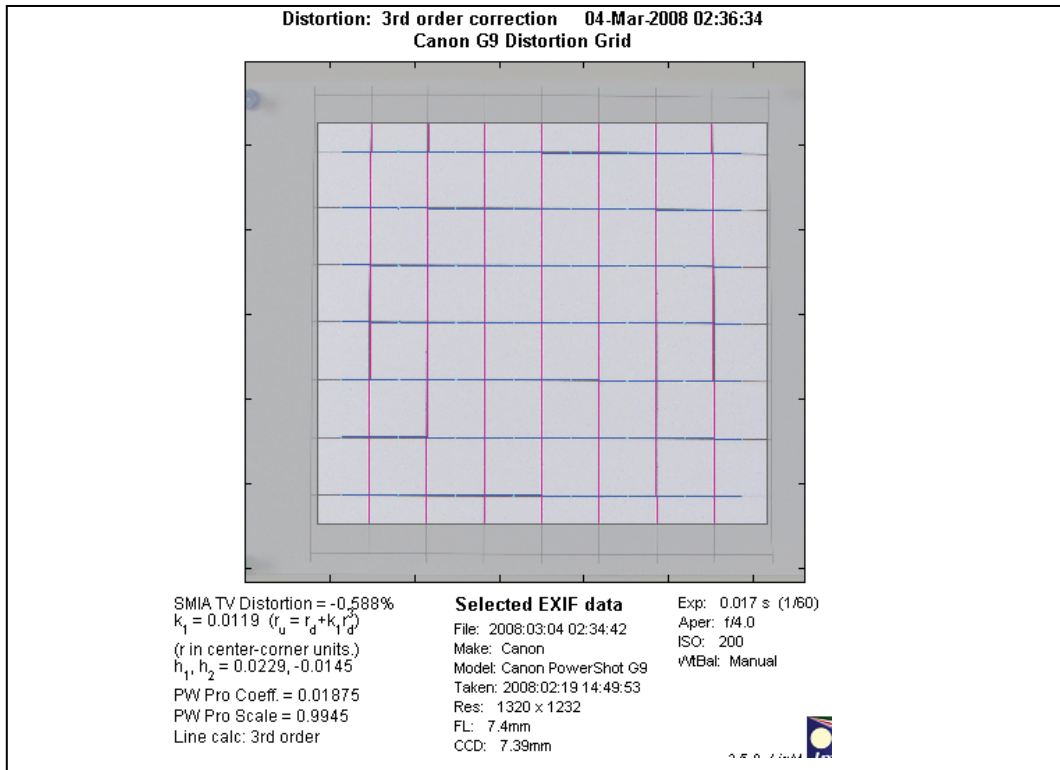


Figure 8: Canon G9 distortion results

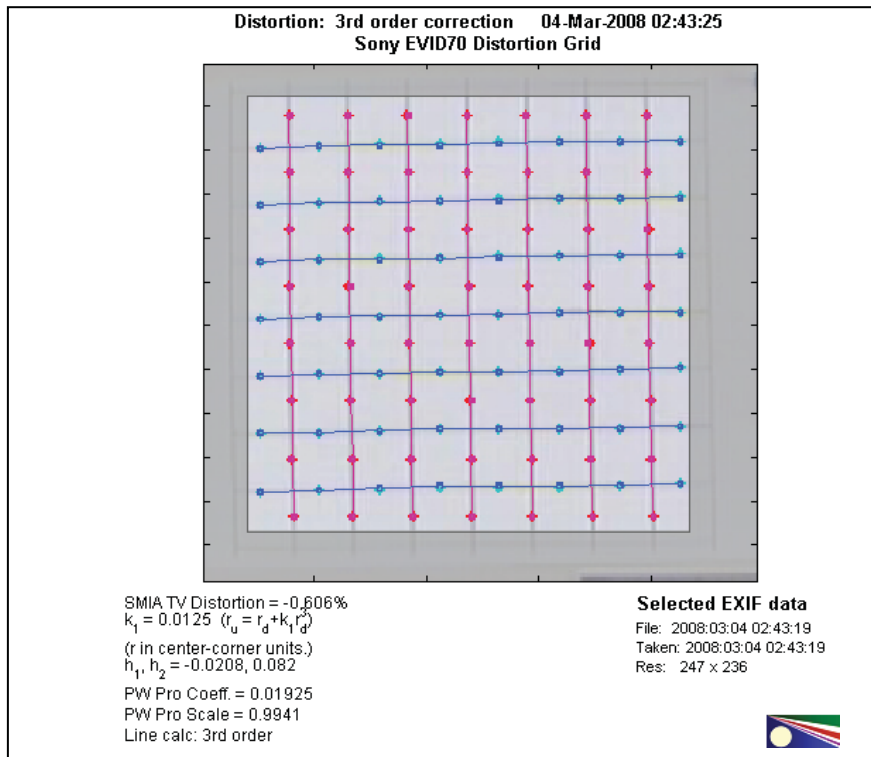


Figure 9: Sony EVID70 distortion results

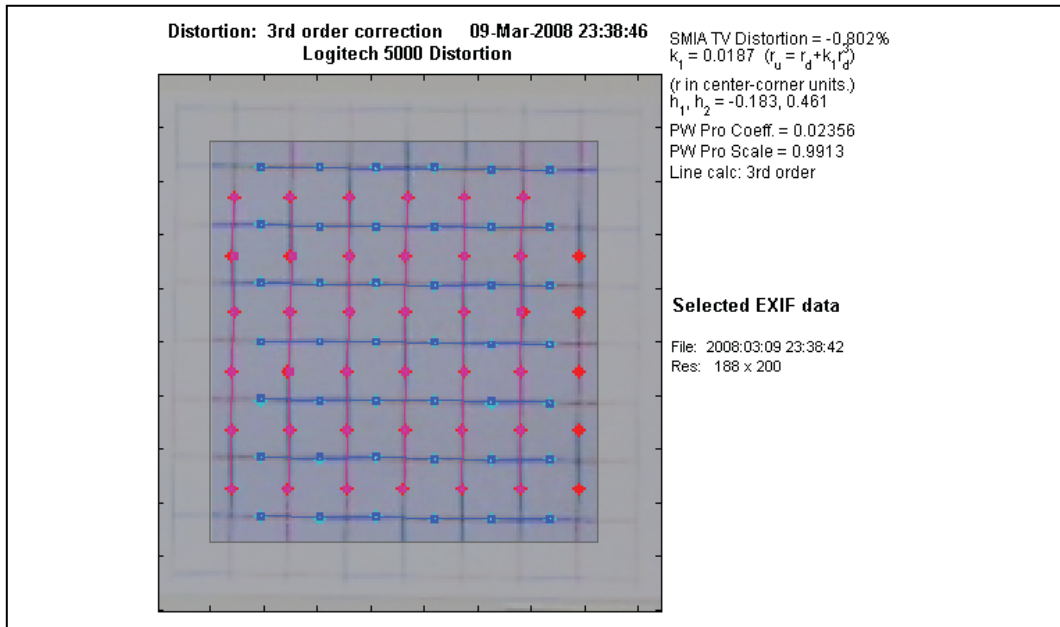


Figure 10: Logitech QuickCam Pro 5000 distortion results

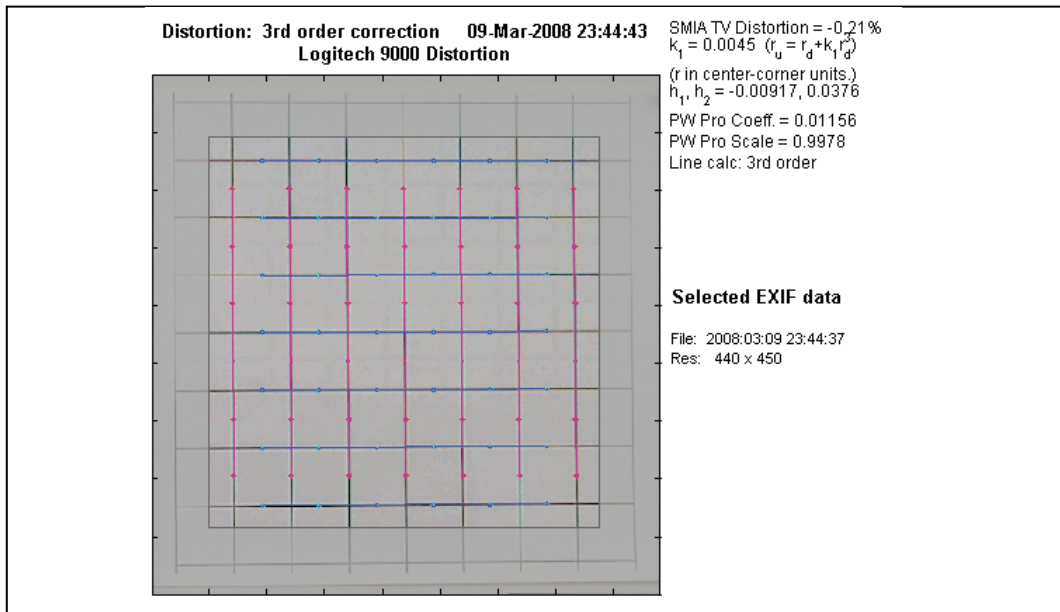


Figure 11: Logitech QuickCam Pro 9000 distortion results

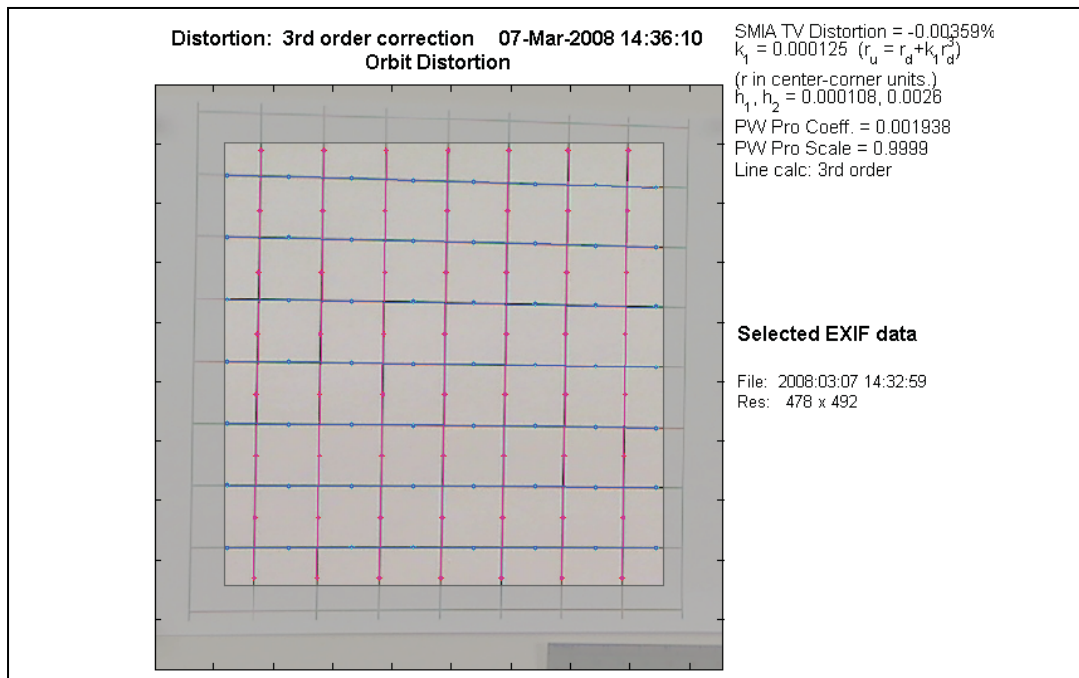


Figure 12: Logitech QuickCam Orbit AF distortion results

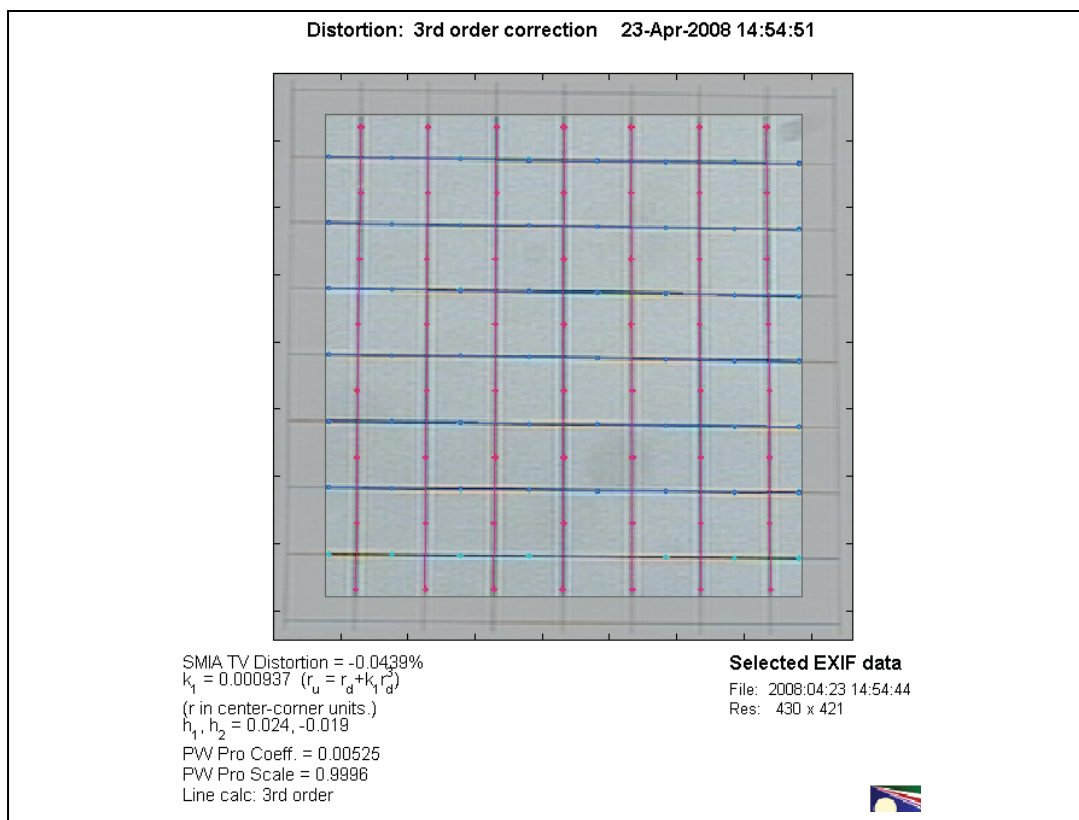


Figure 13: Prototype wide dynamic range camera distortion results

The Standard Mobile Imaging Architecture (SMIA) TV Distortion values for the various cameras were recorded in Table 3, with the best camera's value in green and the worst value in

red. The camera that produced the least amount of distortion was the Logitech QuickCam Orbit AF. All of the cameras exhibited less than one percent of distortion, which is probably negligible for FR.

Table 3: Geometric Accuracy

Camera	(SMIA) TV Distortion (%)
Canon Powershot G9	-0.588
Sony EVID70	-0.606
Logitech QuickCam Pro 5000	-0.802
Logitech QuickCam Pro 9000	-0.21
Logitech QuickCam Orbit AF	-0.00359
Prototype wide dynamic range camera	-0.0439%

2.7 Spatial Uniformity

Spatial uniformity was measured by determining the light falloff. Some vignetting is expected, especially when the lens of the camera is wide open and when supplemental light sources are used, as was the case in these experiments.

To measure spatial uniformity, the plain wall was photographed with each camera and analyzed with Imatest's "Light Falloff" module, which generated the channel contour plots in Figure 14—Figure 19. Deviation from the expected contour shapes could be explained by noise and non-uniformity of the sensor response.

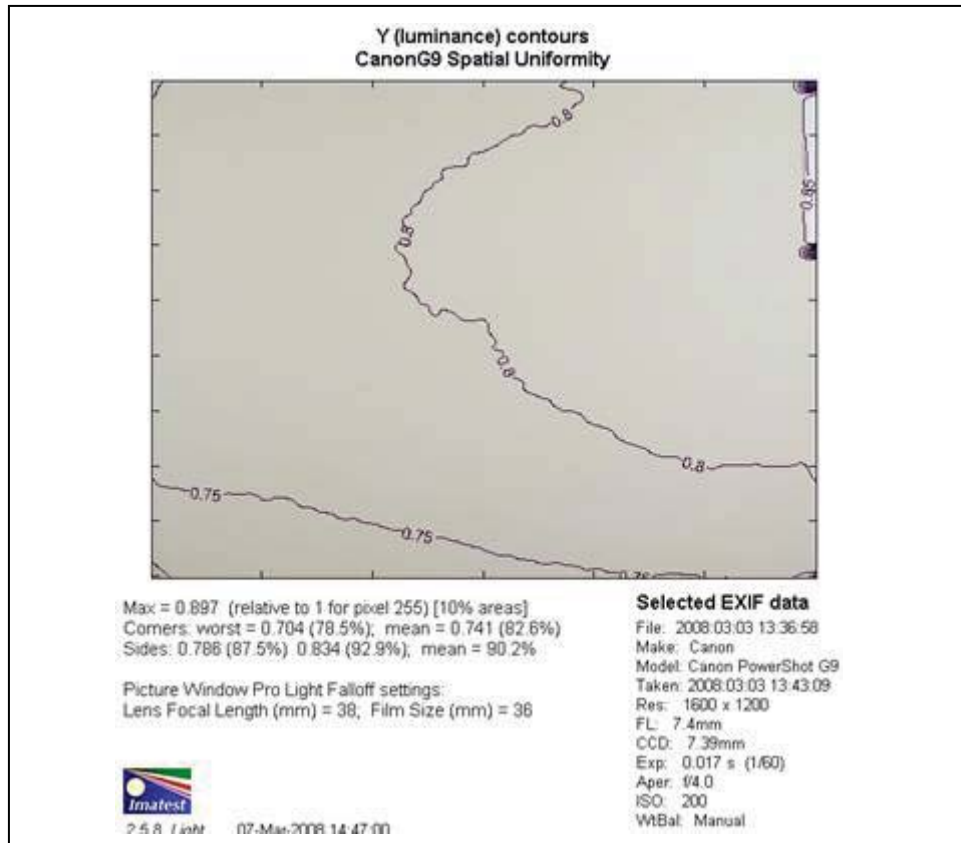


Figure 14: Canon G9 spatial uniformity results

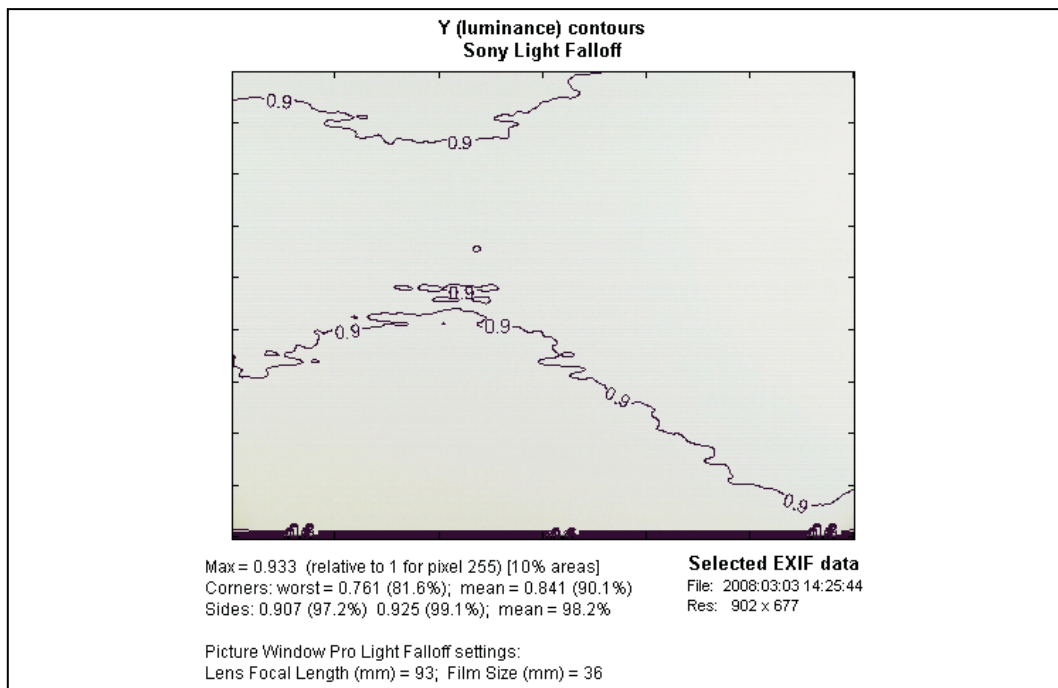


Figure 15: Sony EVID70 spatial uniformity results

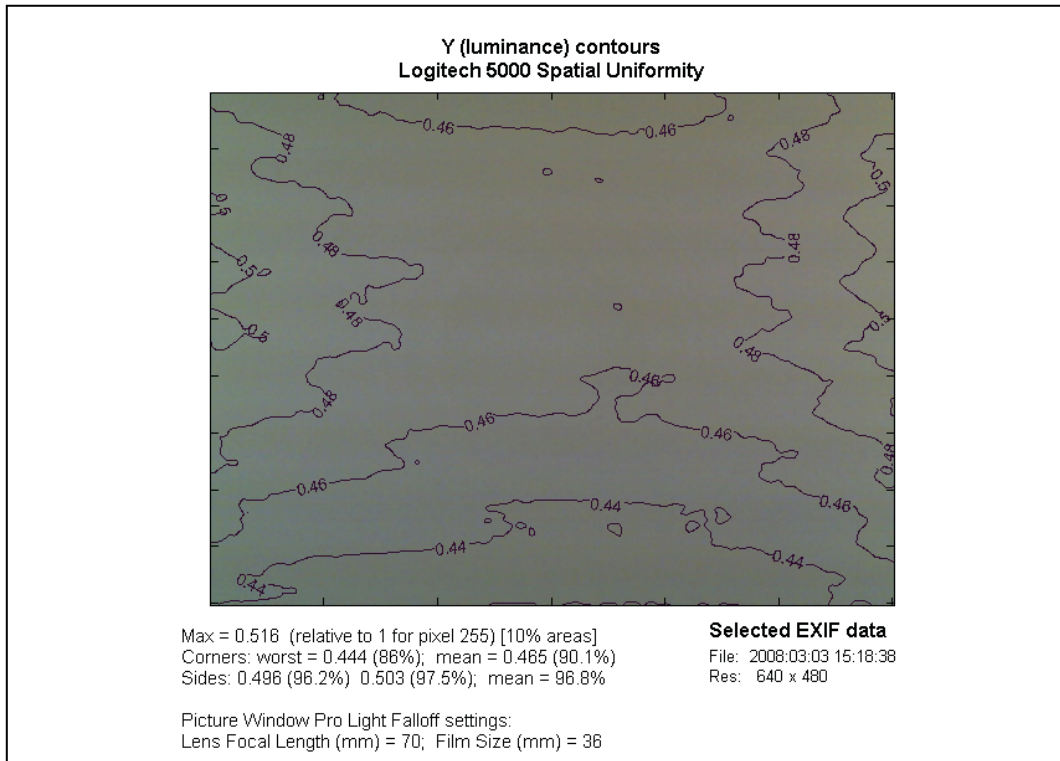


Figure 16: Logitech QuickCam Pro 5000 spatial uniformity results

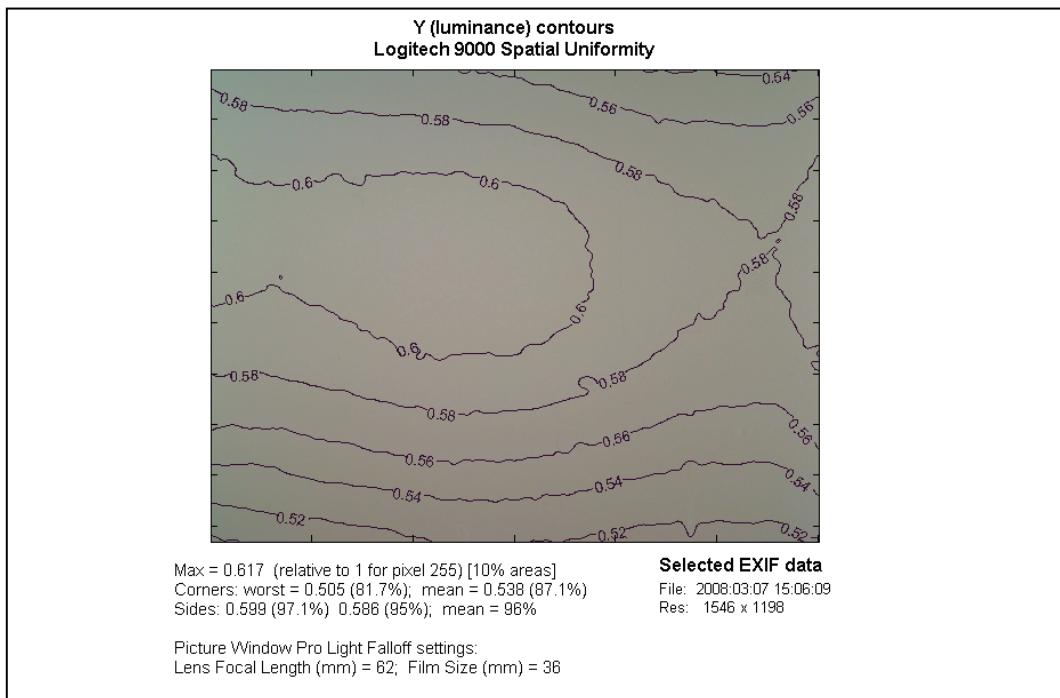


Figure 17: Logitech QuickCam Pro 9000 spatial uniformity results

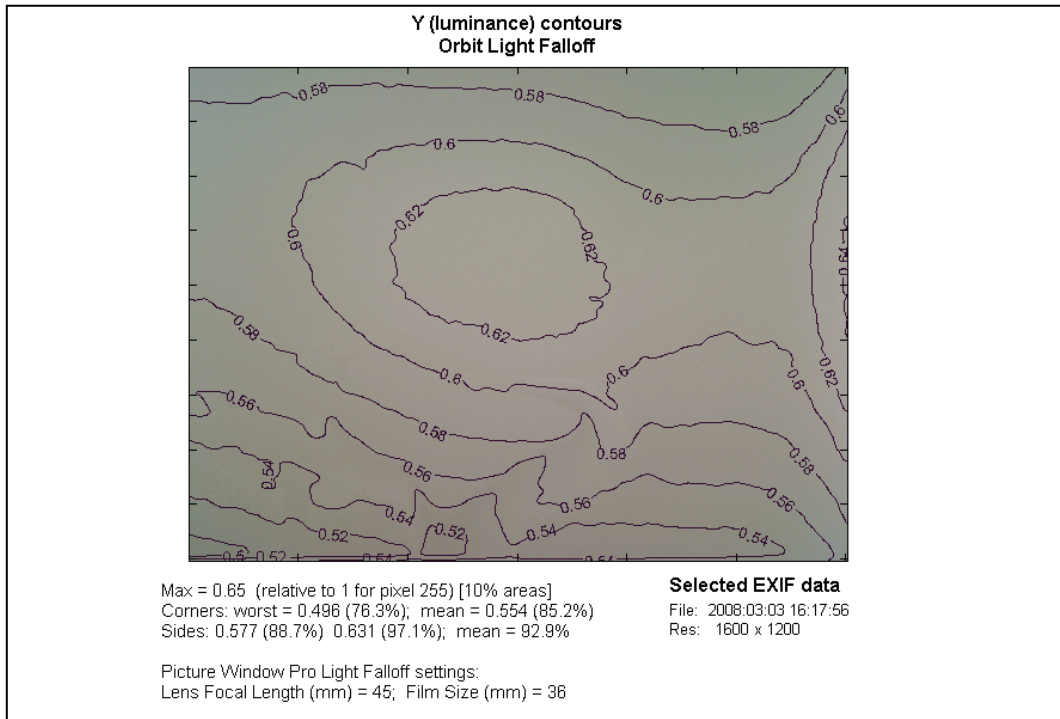


Figure 18: Logitech QuickCam Orbit AF spatial uniformity results

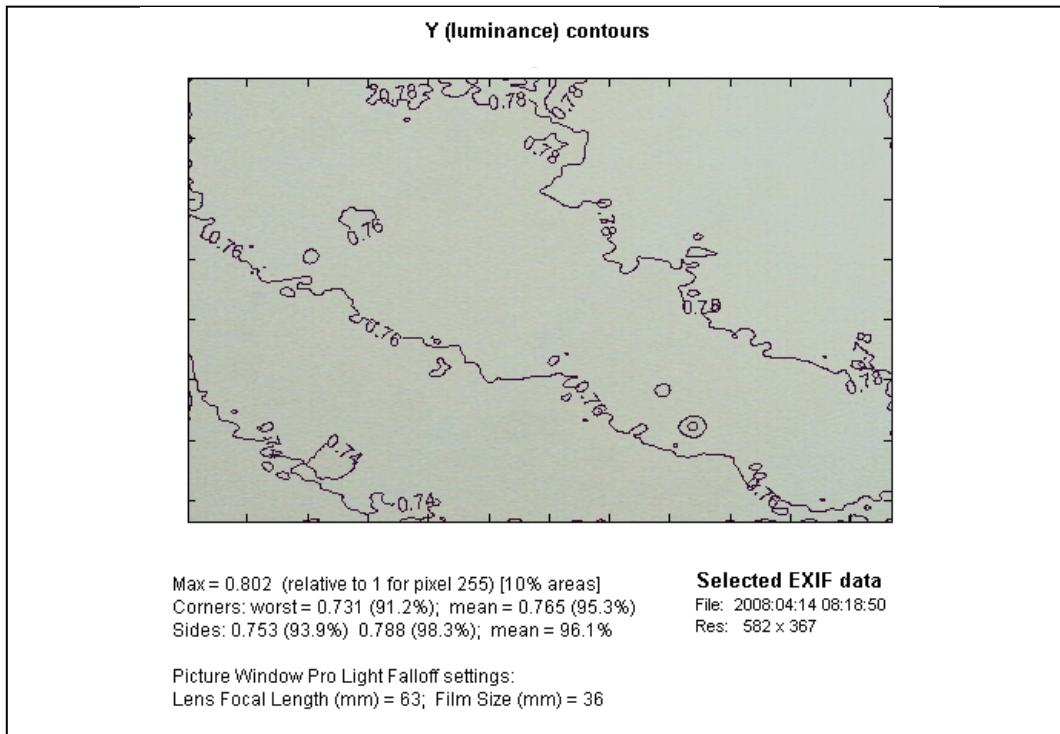


Figure 19: Prototype wide dynamic range camera spatial uniformity results

Table 4 quantifies spatial uniformity in terms of the percent difference between the maximum and minimum intensity values, with the best camera's value in green and the worst value in red. The most uniform image was captured with the wide dynamic range camera, and the least uniform was captured with the Logitech QuickCam Orbit AF. For the most part, these results followed a trend—the larger the field of view³ of the camera, the less uniform the intensity. These results were expected, because the light incident on the wall from the angled side lamps is more uniform in a concentrated, central area.

Table 4: Spatial Uniformity Results

Camera	% Difference Between Maximum and Minimum Intensities
Canon Powershot G9	21.5
Sony EVID70	18.4
Logitech QuickCam Pro 5000	14
Logitech QuickCam Pro 9000	18.3
Logitech QuickCam Orbit AF	23.7
Prototype wide dynamic range camera	8.9

2.8 Depth of Field

The depth of field, or distance range that is in focus, suggested for FR is such that "the individual millimeter markings of rulers placed on the subject's nose and ear facing the camera can be seen simultaneously in a captured test image."⁴ For the purposes of this evaluation, the distance between a subject's nose and ear is assumed to be five inches.

The depth of field was calculated for the Canon G9 using the formula in Figure 20. The Canon G9 has a 1/1.7-inch type Charge Coupled Device (16.275 mm diagonal, 13.02mm width) and capture dimensions of 4000 x 3000 pixels, from which the diameter of the circle of confusion (c), or pixel pitch, was calculated to be 0.003255 mm. The camera was operated at a focal length (f) of 7.4 mm, a camera-to-subject distance (s) of 70 cm, and an F-stop (F) of 4. The Canon G9's depth of field was 233 mm or 9.2 inches, which exceeds the ISO suggested depth of field of five inches.

³ Fields of view are listed in the second-to-last column in Table 2.

⁴ FACESTD A.2.5 - Focus and depth of field (Informative).

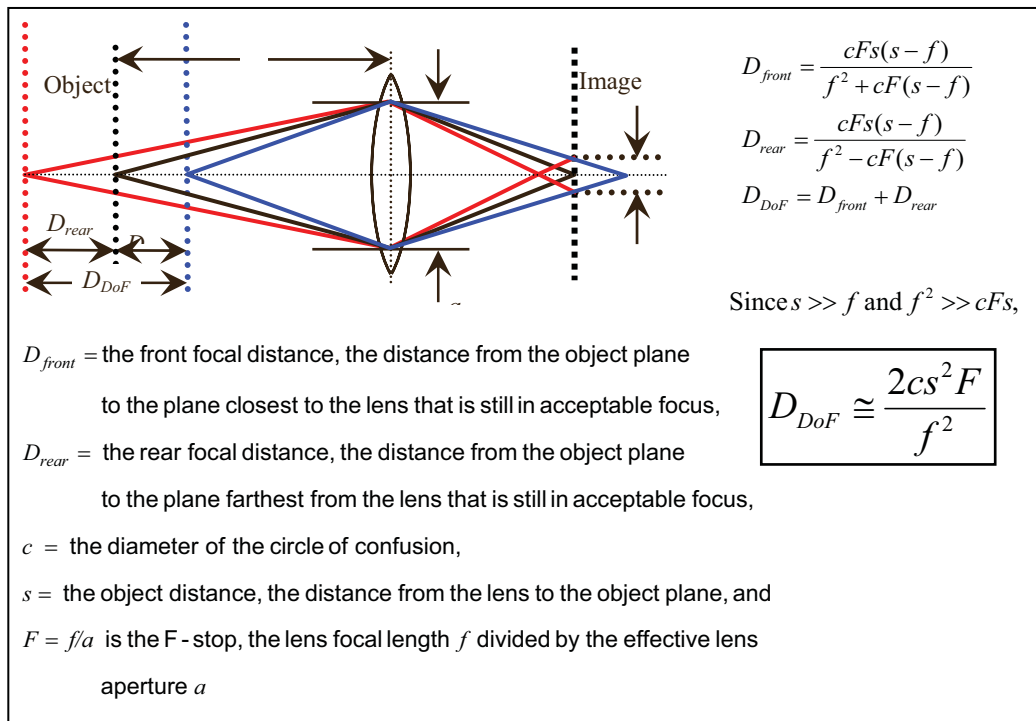


Figure 20: Depth of field calculation

Depth of field conformance was assessed for each camera by viewing, in the photographs in Figure 21—Figure 26, the millimeter demarcations on rulers placed five inches apart. In the Canon G9's images in Figure 21, both rulers' millimeter demarcations were discernible at the Canon's three highest capture dimensions (4000 x 3000 pixels, 3264 x 2448 pixels, and 2592 x 1944 pixels), which corroborates the depth of field calculation above. Neither ruler's millimeter demarcations were discernible at capture dimensions of 1600 x 1200 pixels or 640 x 480 pixels. In the images captured by all other tested cameras (Figure 22—Figure 26), neither ruler's millimeter demarcations were discernible. Therefore, the only tested camera that had adequate depth of field was the Canon G9 (at capture dimensions of 2592 x 1944 pixels and higher).

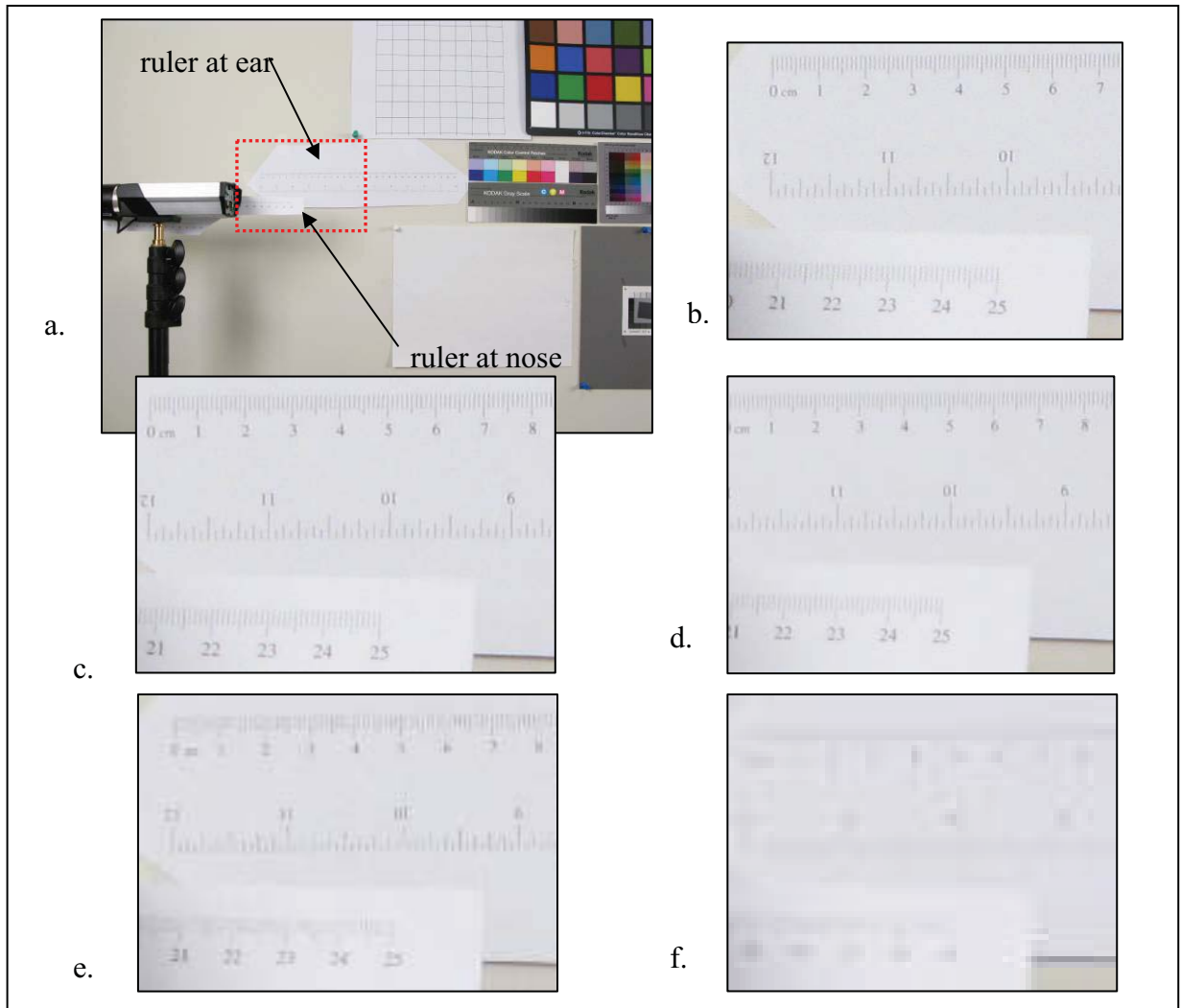


Figure 21: Canon G9 depth of field results – (b) region in red box cropped from an original image of 4000 x 3000 pixels, (c) 3264 x 2448 pixels, (d) 2592 x 1944 pixels, (e) 1600 x 1200 pixels, and (f) 640 x 480 pixels

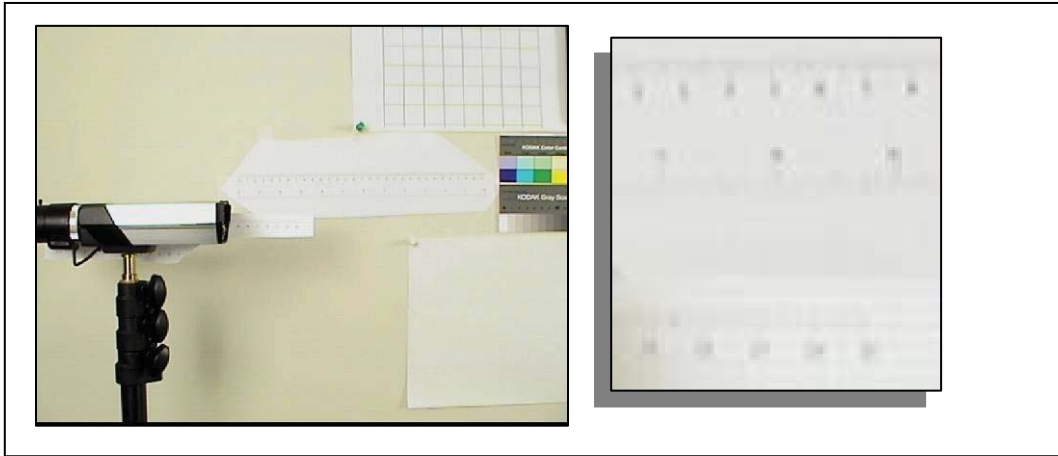


Figure 22: Sony EVID70 depth of field results

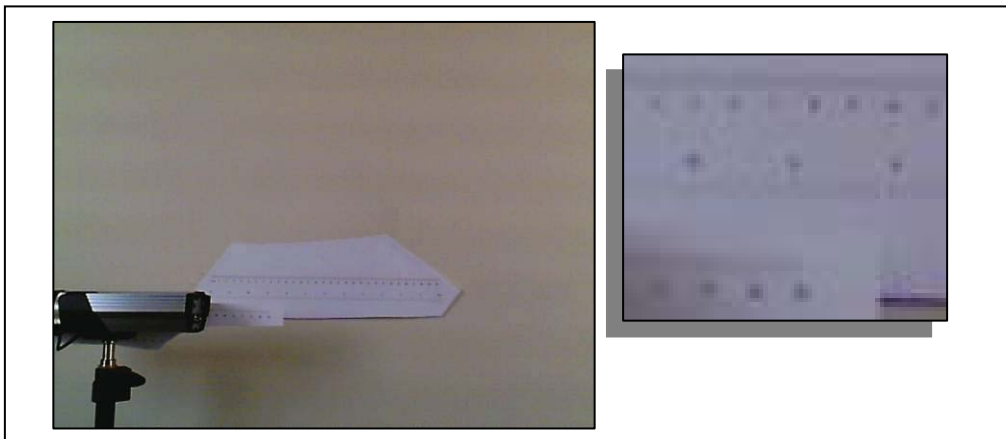


Figure 23: Logitech QuickCam Pro 5000 depth of field results

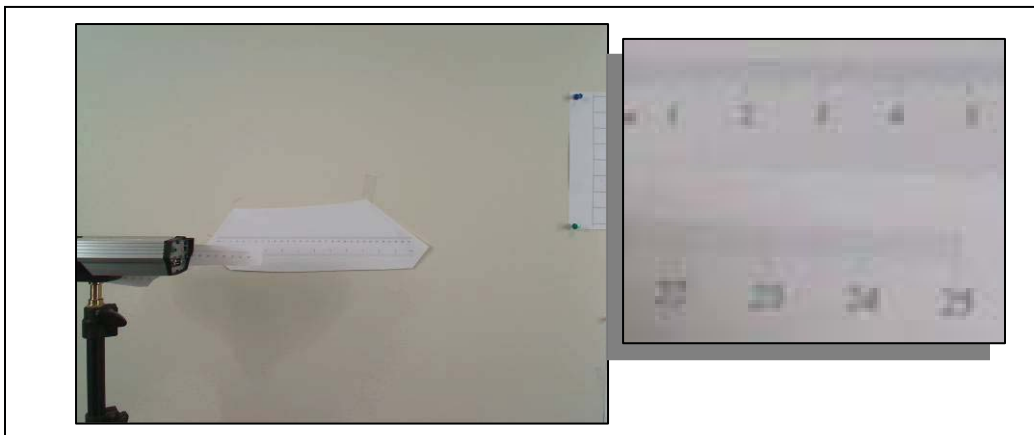


Figure 24: Logitech QuickCam Pro 9000 depth of field results

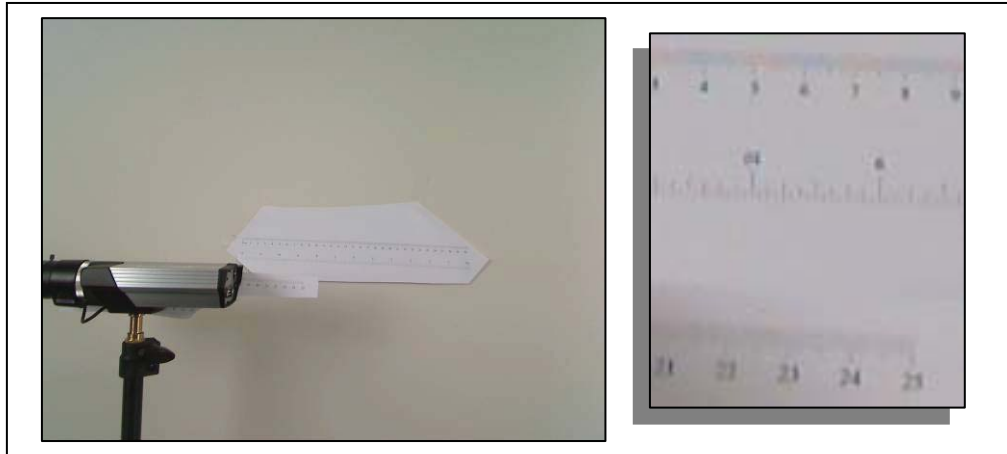


Figure 25: Logitech QuickCam Orbit AF depth of field results

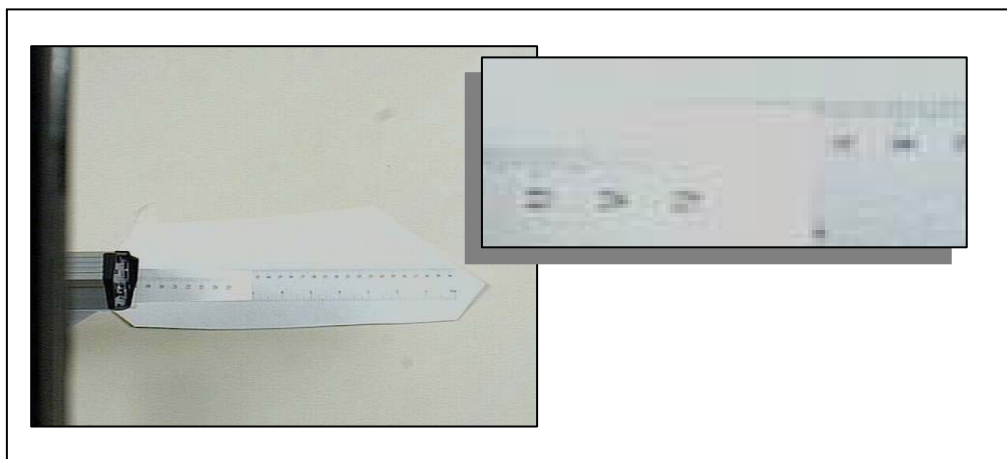


Figure 26: Prototype wide dynamic range camera depth of field results

2.9 Tonal Response

The tonal response curve defines how tonal values in a raw image are mapped nonlinearly to intensities. Tonal response was measured by photographing the Kodak Q13 test pattern (Figure 1a) and analyzing the image with Imatest's "Stepchart" module. Figure 27—Figure 32 show plots of the average density of the grayscale patches of the Kodak Q13 chart (black curve) and first and second order density fits (dashed blue and green curves) for each camera. The horizontal axis is the distance along the target. The exposure accuracy is indicated by the maximum density value. The graph should resemble a step curve, and all twenty of the gray zones should be detected by the camera.

The Canon G9 graph (Figure 27) most closely resembles a step curve. Edge sharpening is evident in the Sony EVID70's response (Figure 28) and in the wide dynamic range camera's response (Figure 32), as evidenced by bumps in the steps. The Logitech QuickCam Pro 5000 had the lowest maximum density, indicating that its image was too dark.

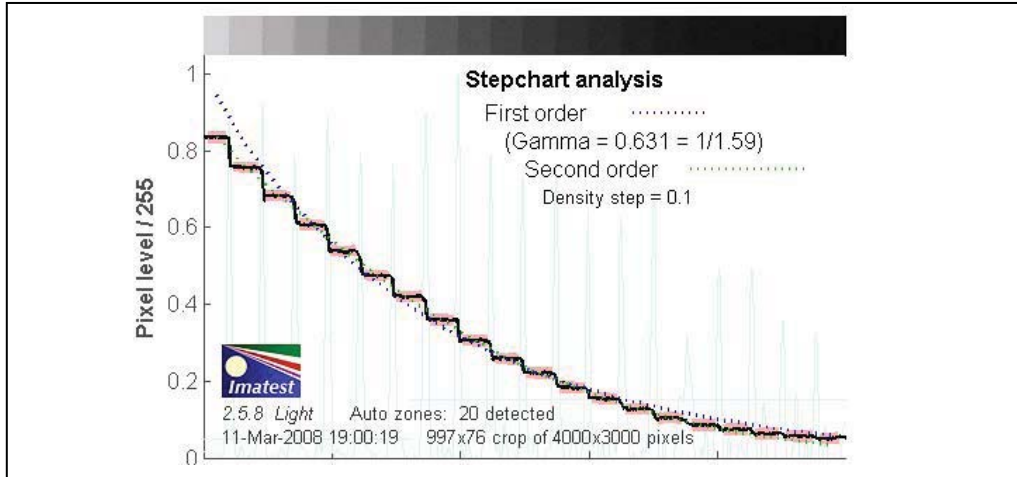


Figure 27: Canon G9 tonal response

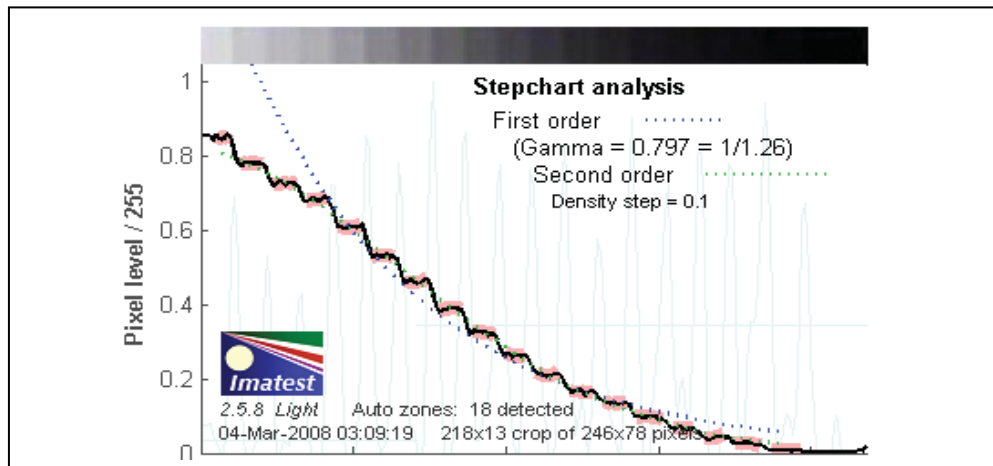


Figure 28: Sony EVID70 tonal response

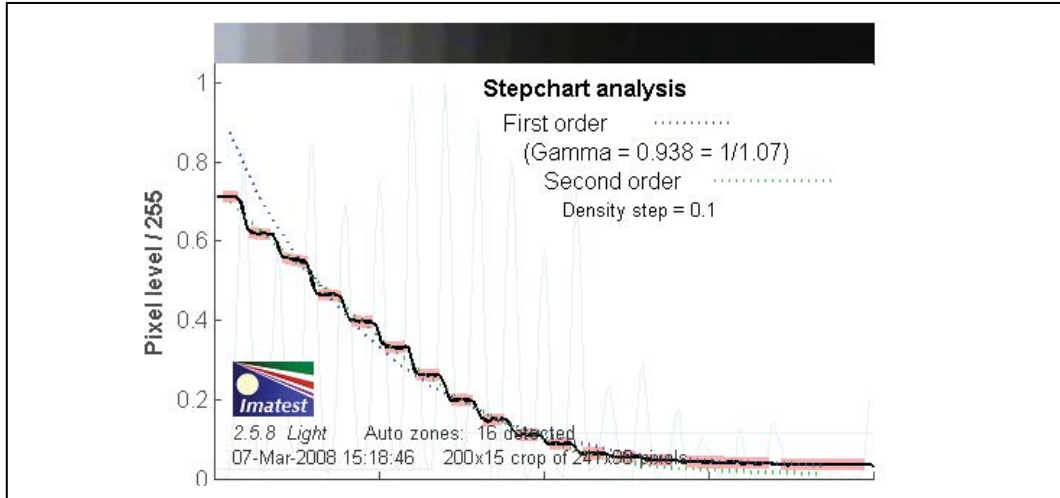


Figure 29: Logitech QuickCam Pro 5000 tonal response

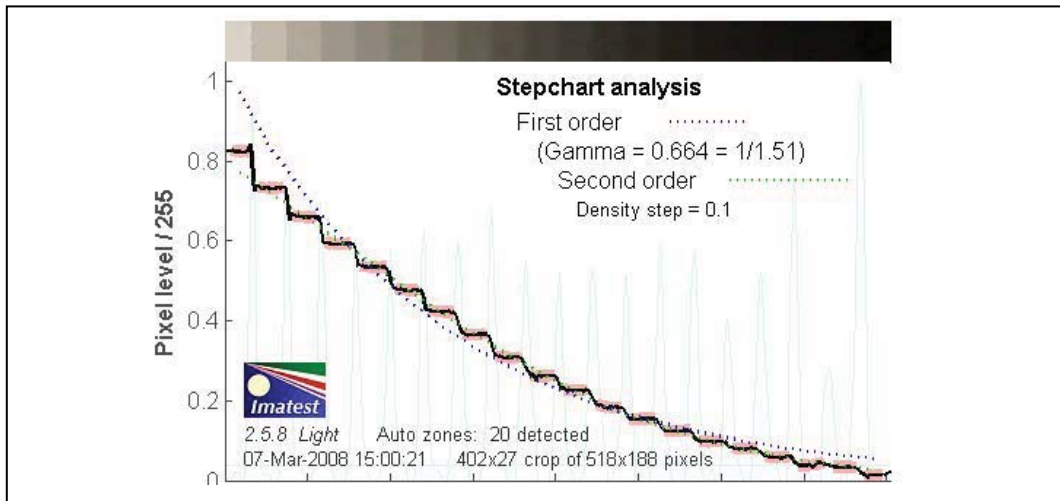


Figure 30: Logitech QuickCam Pro 9000 tonal response

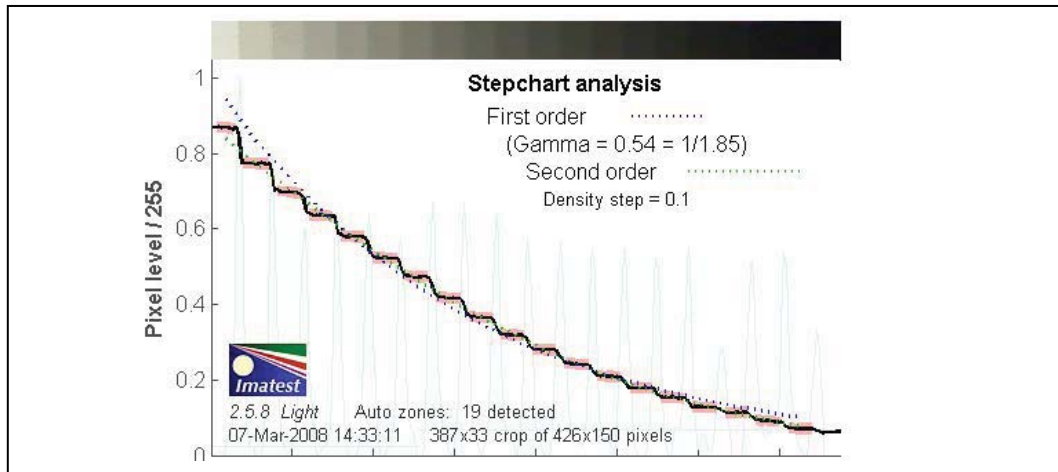


Figure 31: Logitech QuickCam Orbit AF tonal response

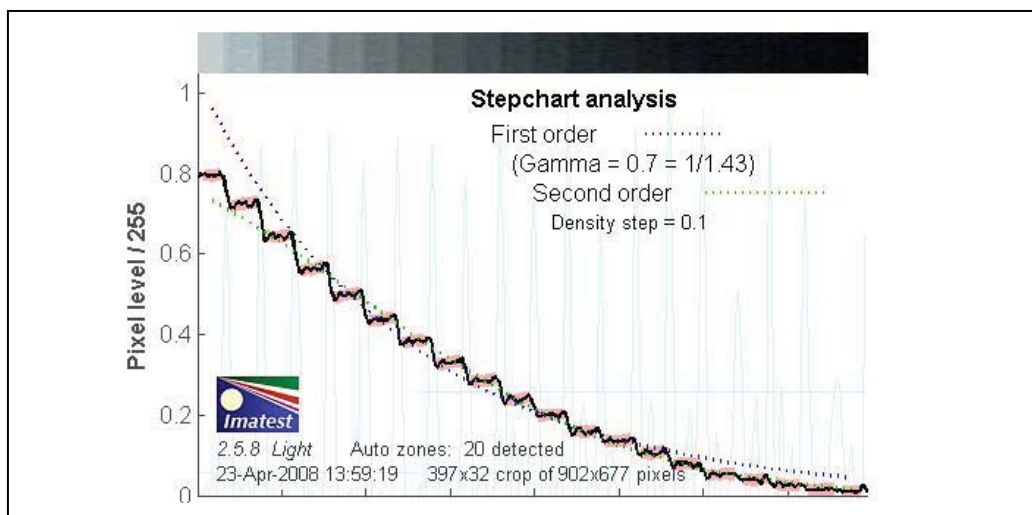


Figure 32: Prototype wide dynamic range camera tonal response

The number of zones detected by each camera is summarized in Table 5. The Logitech QuickCam Pro 5000 detected the fewest number of zones. The Canon, wide dynamic range, and Logitech 9000 detected all 20 zones.

The gammas of the cameras in this study, which were estimated by Imatest's "Stepchart" module, are listed in Table 5. The Logitech QuickCam Orbit AF's gamma of 0.54 is closest to the optimal sRGB gamma of 0.45, while the Logitech QuickCam Pro 5000's gamma of 0.938 is the least desirable.

Table 5: Detection of gray zones

Camera	# Zones Detected	Gamma
Canon Powershot G9	20	0.631
Sony EVID70	18	0.797
Logitech QuickCam Pro 5000	16	0.938
Logitech QuickCam Pro 9000	20	0.664
Logitech QuickCam Orbit AF	19	0.54
Prototype wide dynamic range camera	20	0.7

2.10 Noise

Noise, or variation in pixel level, was measured by photographing the GretagMacbeth ColorChecker® (Figure 1b) and analyzing the image with Imatest's "Colorcheck" module. In Figure 33—Figure 38, the signal-to-noise ratio (SNR) (in pixel intensity) for each patch: R, G, B, (Red, Green, Blue) and Y (luminance) is plotted versus the gray zone.

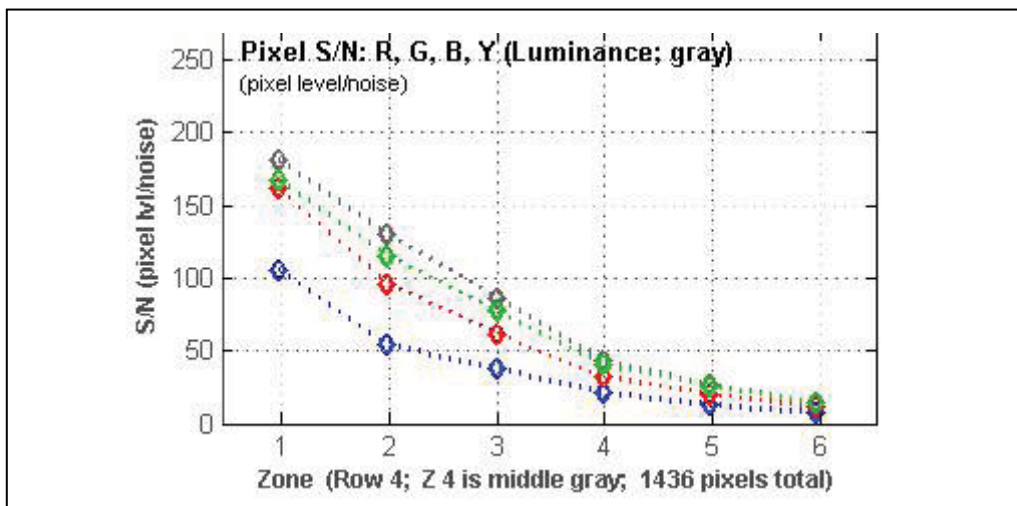


Figure 33: Canon G9 SNR results

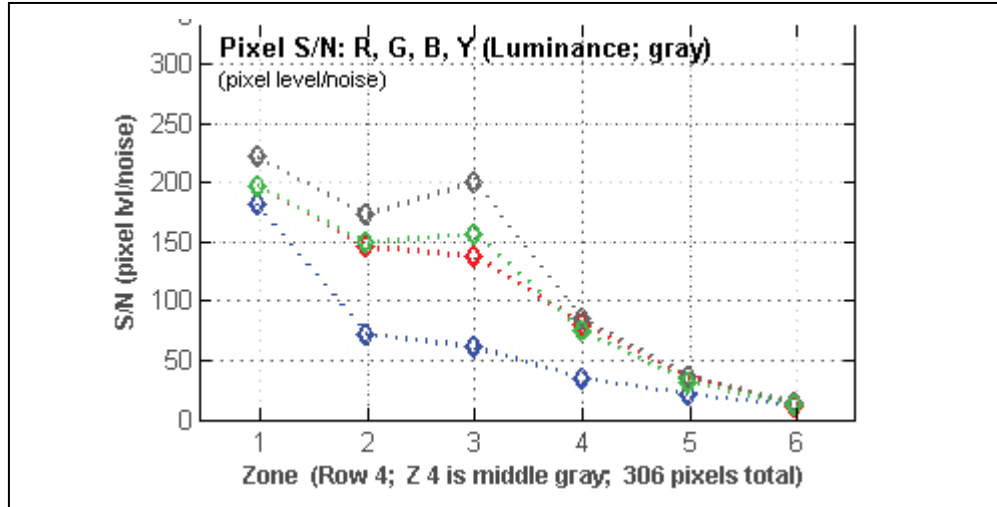


Figure 34: Sony EVID70 SNR results

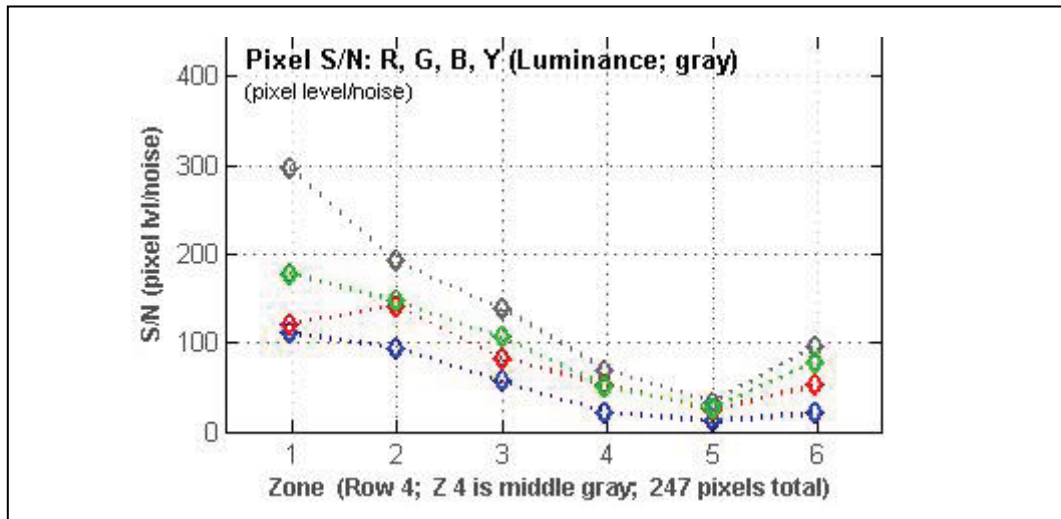


Figure 35: Logitech QuickCam Pro 5000 SNR results

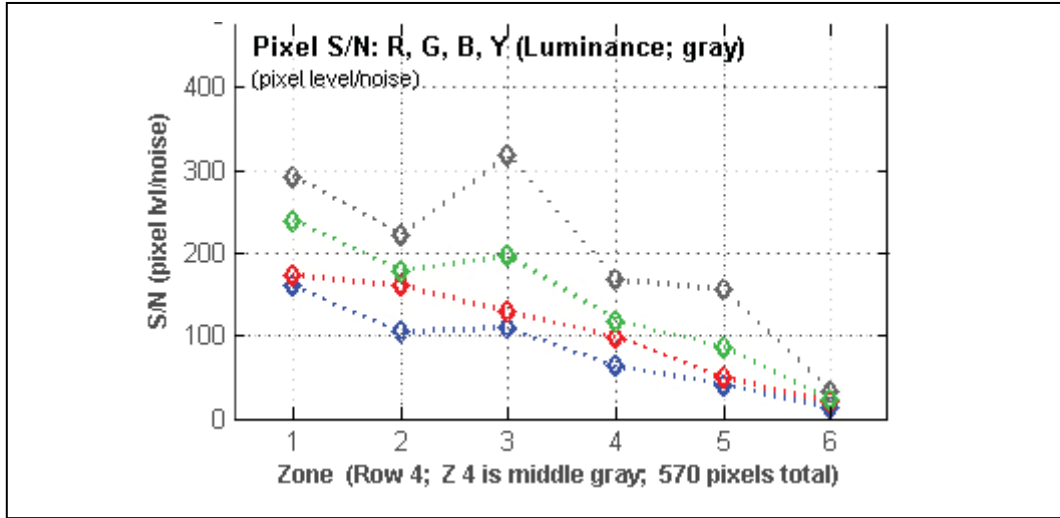


Figure 36: Logitech QuickCam Pro 9000 SNR results

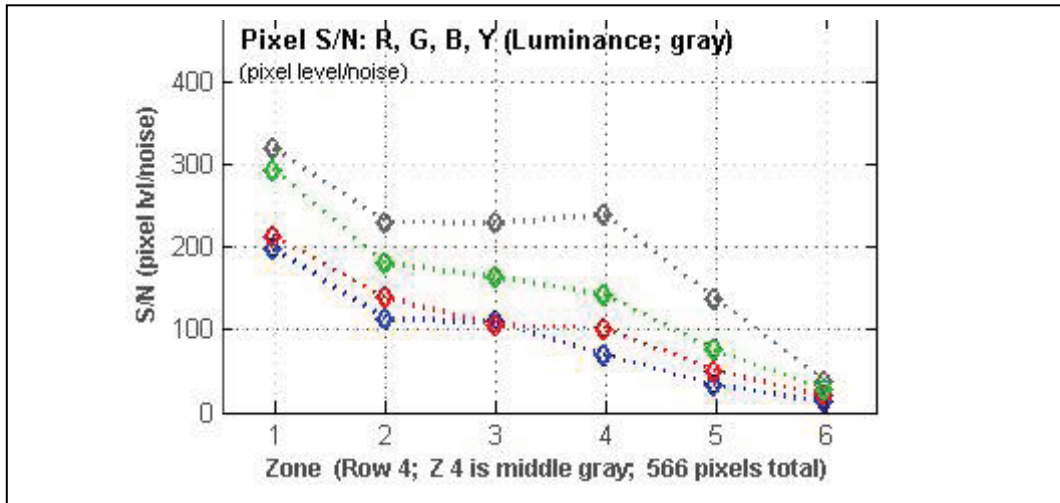


Figure 37: Logitech QuickCam Orbit AF SNR results

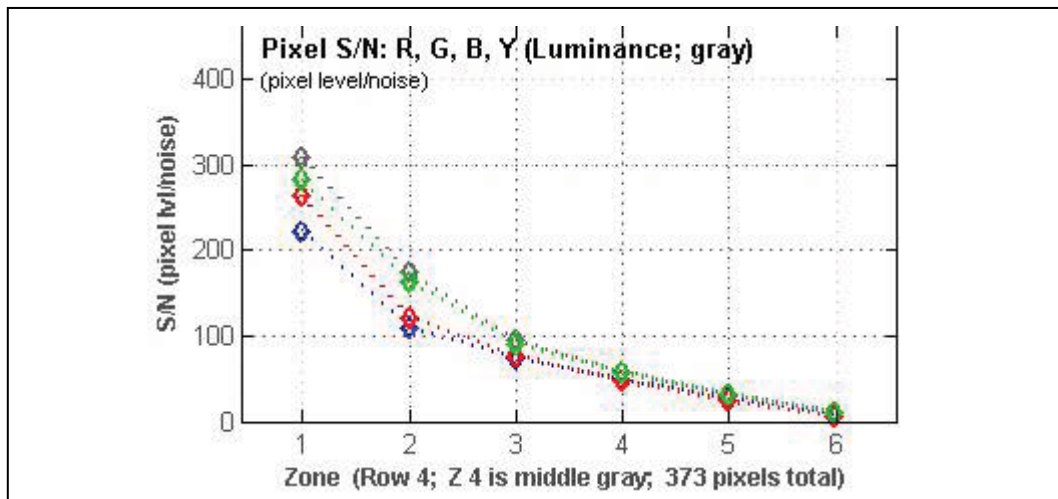


Figure 38: Prototype wide dynamic range camera SNR results

The SNR results (in the luminance channel) for all of the cameras are combined in Figure 39 for comparison. Each camera's average SNR (over all of the gray zones) in the luminance channel is listed in Table 6.

The cameras with the highest (best) SNR were the Logitech webcams, and the lowest (worst) SNR was achieved with the Canon G9. The lower amount of noise in the webcam images may be due to the inherently smaller number of pixels per patch and to the JPEG compression, which reduces the variability in the patches.

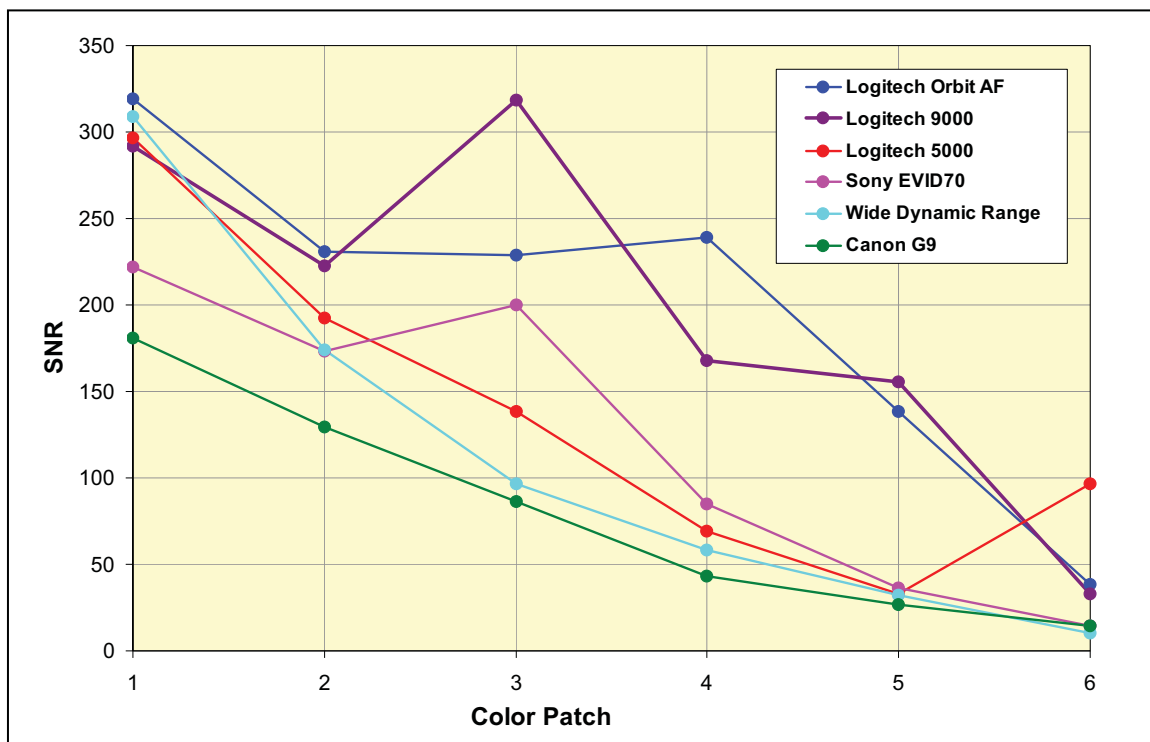


Figure 39: Combined SNR results (luminance channel)

Table 6: Average SNR

Camera	Average SNR (in luminance channel, over 6 gray zones)
Canon Powershot G9	80.04
Sony EVID70	121.87
Logitech QuickCam Pro 5000	137.7
Logitech QuickCam Pro 9000	198.19
Logitech QuickCam Orbit AF	199.22
Wide dynamic range camera	113.25

2.11 Color Accuracy

The color accuracy of the various cameras was tested by photographing the GretagMacbeth™ ColorChecker® chart (Figure 1b) and analyzing the images with Imatest's "Colorcheck" module. The Canon Powershot G9 and the Sony EVID70 were white balanced manually prior to imaging the test charts. To manually white balance a camera, an image of white paper is captured, and the camera uses it as the reference color. A shortcoming of the Logitech webcams and the wide dynamic range camera was that they do not permit manual white balancing. While automatic white balance settings are usually acceptable, they can be incorrect when a strong color dominates the scene.

Table 7 reports, for each camera, the mean color error, ΔE (the difference between the measured and ideal, or reference, values), and the color error for the dark skin and light skin patches, the accuracy of which is important for human recognition of faces. The best value is shown in green and the worst value is in red in each column. The Canon G9, Logitech QuickCam Pro 9000, Logitech QuickCam Orbit AF, and Sony EVID70 had reasonable mean color errors, while those for the Logitech QuickCam Pro 5000 and the wide dynamic range camera were poor.

Figure 40 – Figure 43 show the color chart as photographed by each camera (outer boxes) compared to the ideal values (inner boxes, see reference diagram in Figure 40b).

Table 7: Color Error

Camera	ΔE mean	ΔE dark skin	ΔE light skin
Canon Powershot G9	10.6	12.57	11.73
Sony EVID70	9.11	9.51	10.77
Logitech QuickCam Pro 5000	19	20.69	12.36
Logitech QuickCam Pro 9000	8.94	10.65	4.2
Logitech QuickCam Orbit AF	7.44	9.92	2.35
Wide dynamic range camera	14.9	4.88	17.26

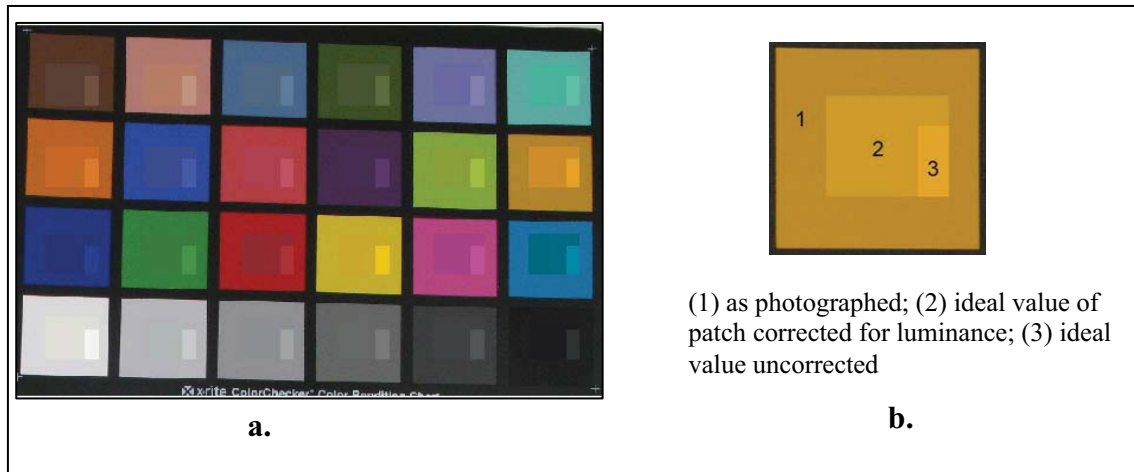


Figure 40: Canon G9 colorchart results (a) and reference diagram (b)

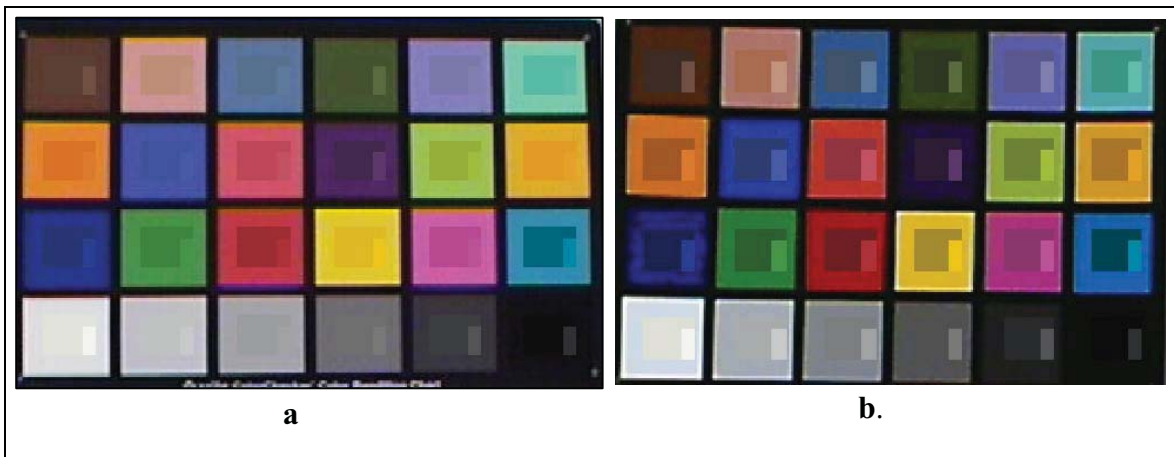


Figure 41: Sony EVID70 (a) and Logitech QuickCam Pro 5000 (b) colorchart results

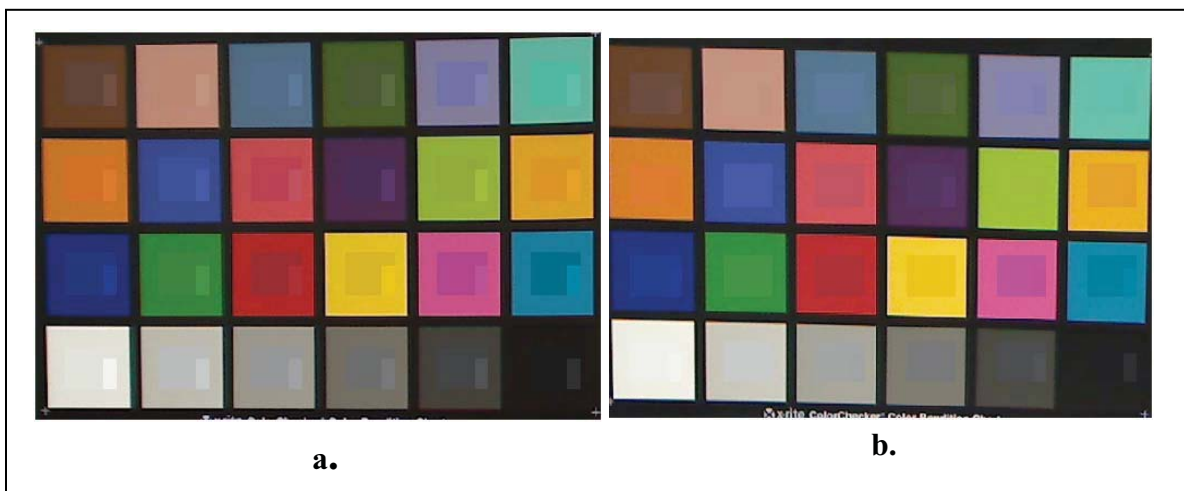


Figure 42: Logitech QuickCam a) Pro 9000 and b) Orbit AF colorchart results



Figure 43: Prototype wide dynamic range camera colorchart results

2.12 Spatial Resolution

The measure of fineness of detail that can be discerned in an image, or the relative contrast at a given spatial frequency (output contrast/input contrast), is called the Modulation Transfer Function (MTF) or Spatial Frequency Response (SFR). The SFR or MTF was measured by photographing the ISO 16067-1 slant edge test pattern (Figure 1c) and analyzing the slant edge region of interest with Imatest's "SFR" module. Indicators of image sharpness are the spatial frequencies where MTF is 50% of its low frequency value (MTF50) and the Nyquist frequency, above which the sensor response exhibits aliasing. The ideal response would have high MTF below the Nyquist frequency and low MTF at and above it. The MTF and edge profile, a measure of sharpness in the spatial domain, for each test camera are shown in Figure 44 – Figure 49. For comparison, the SFRs for all cameras were plotted on the same graph in Figure 50. Similarly, Figure 51 contains each camera's edge profile on the same graph. The Canon G9, Logitech QuickCam Orbit AF, and Logitech QuickCam Pro 9000 MTFs exhibited oversharpening, as annotated in Figure 50.

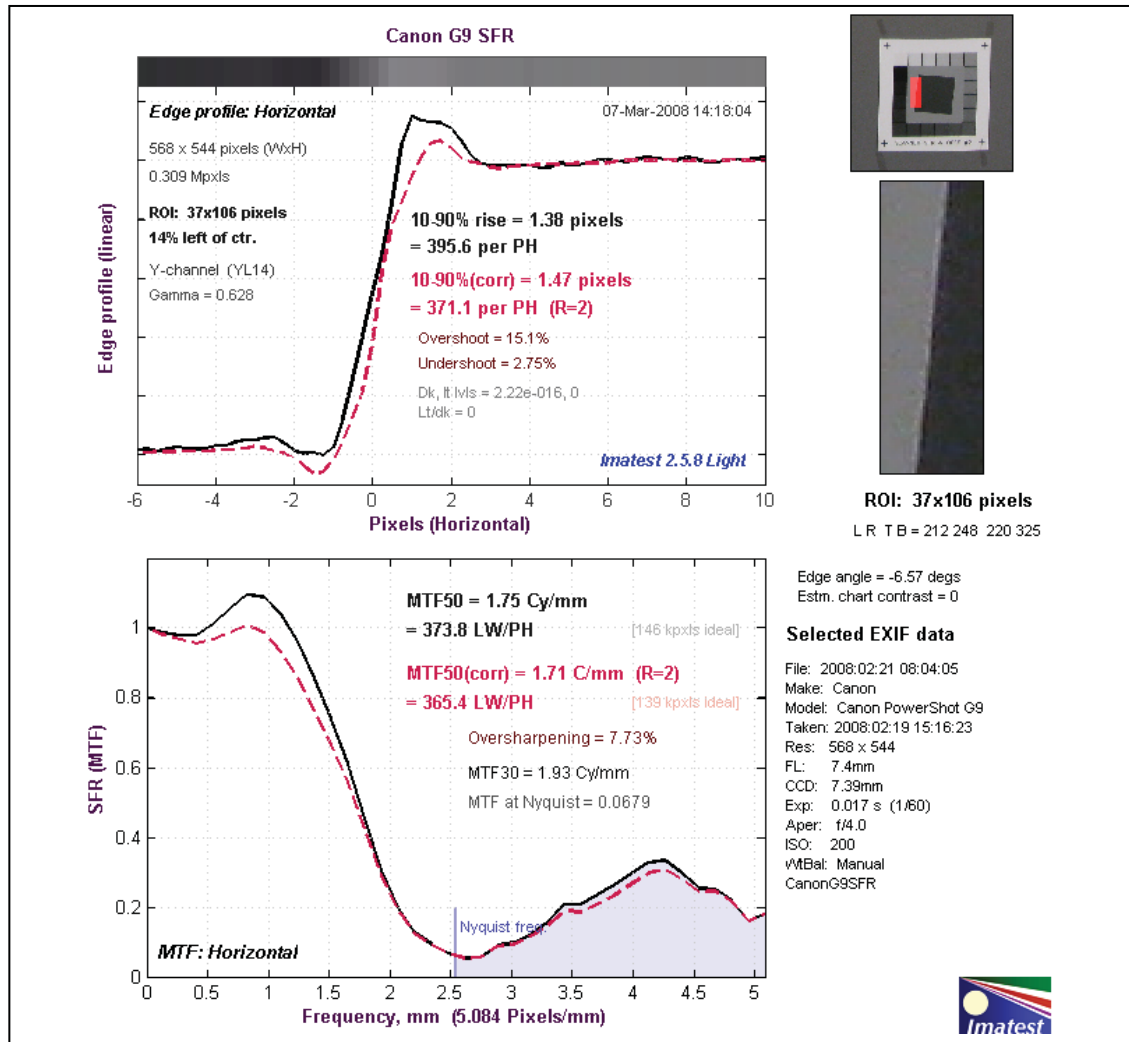


Figure 44: Canon Powershot G9 SFR and edge profile

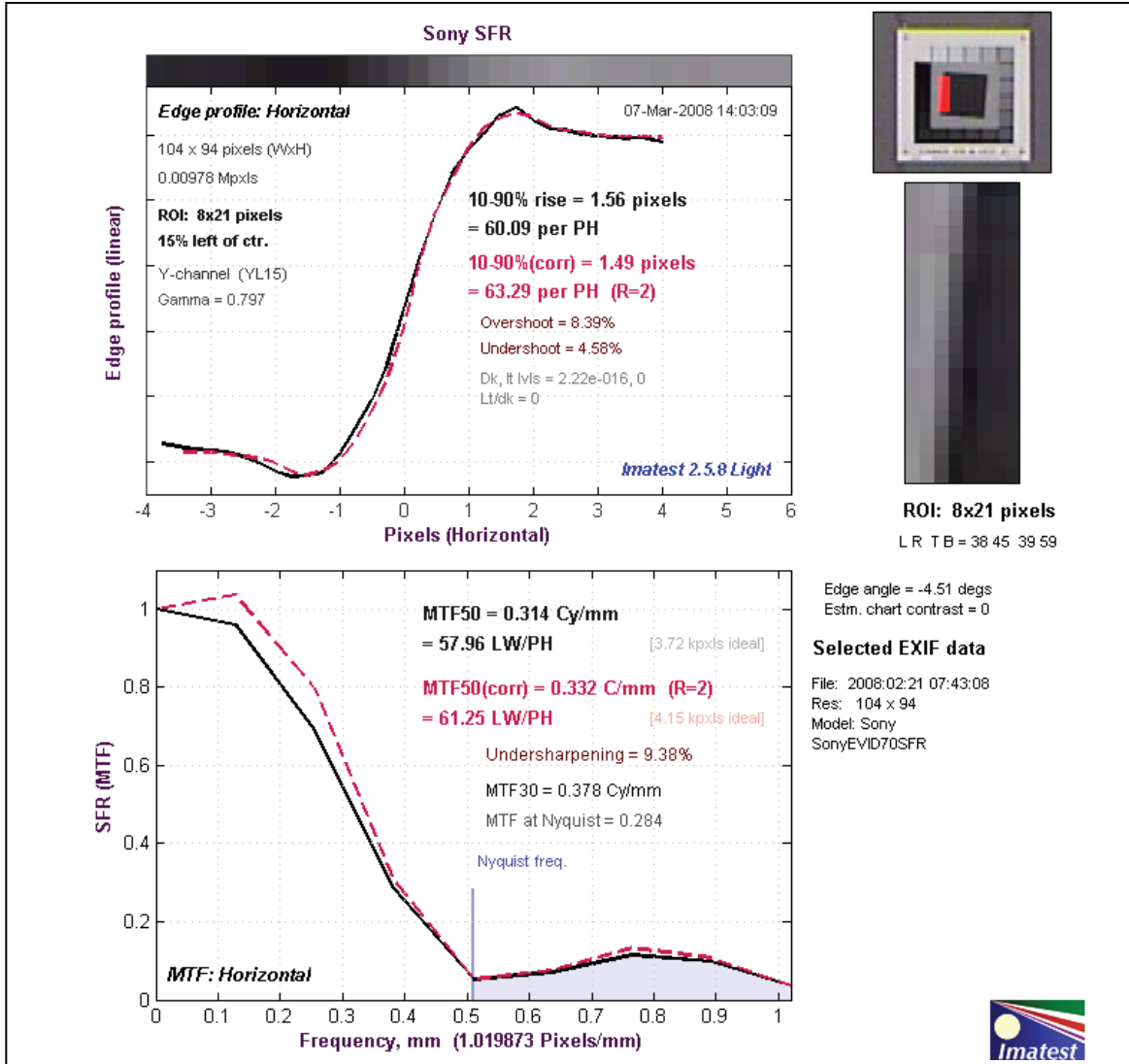


Figure 45: Sony EVID70 SFR and edge profile

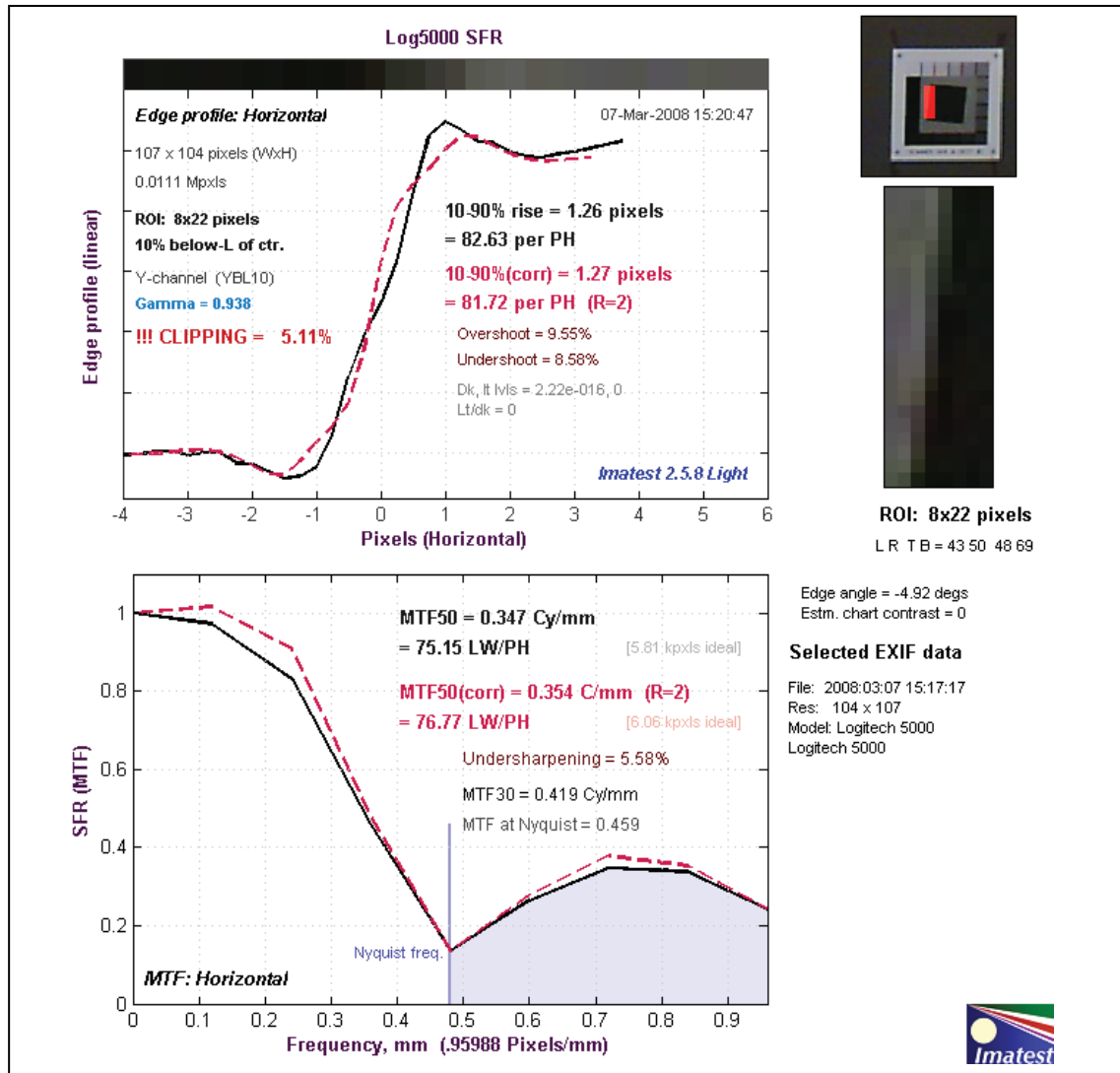


Figure 46: Logitech QuickCam Pro 5000 SFR and edge profile

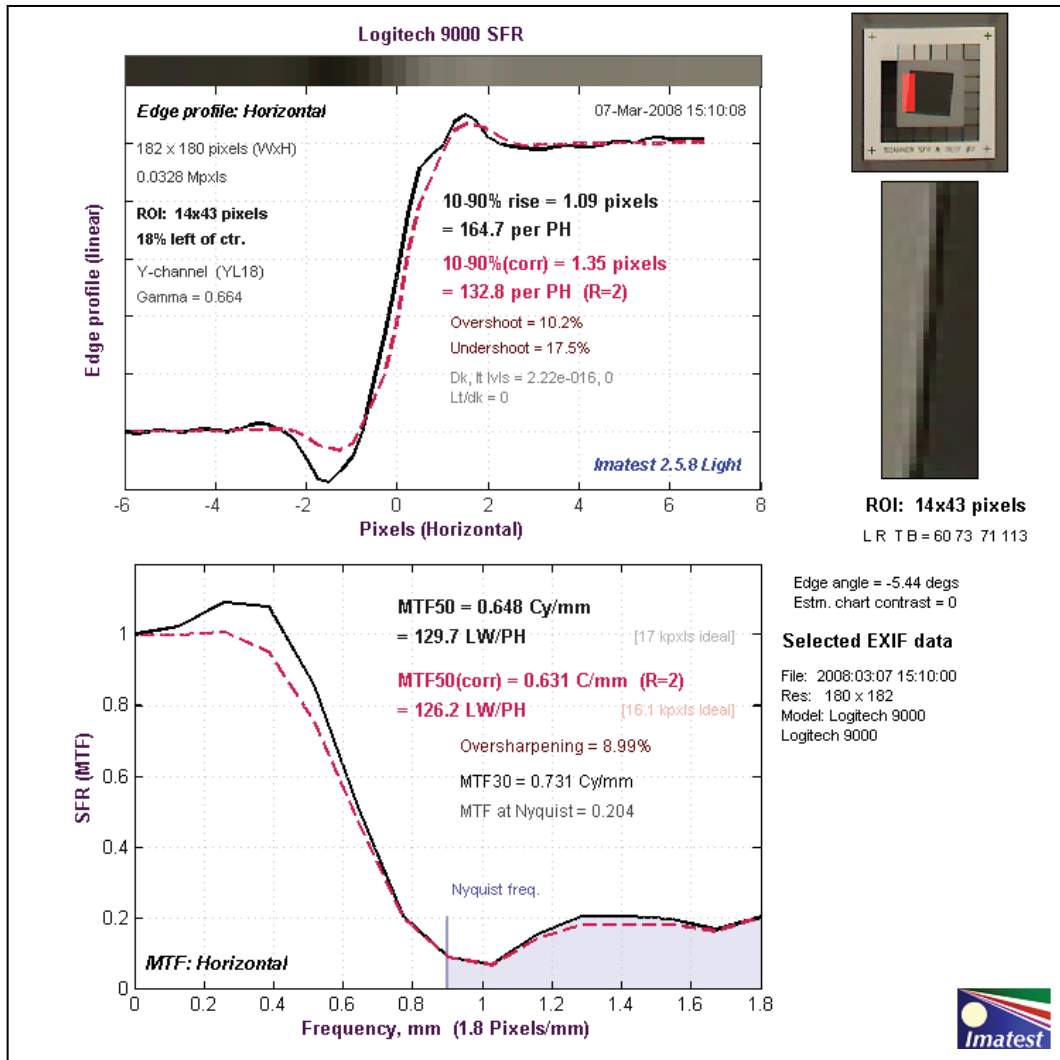


Figure 47: Logitech QuickCam Pro 9000 SFR and edge profile

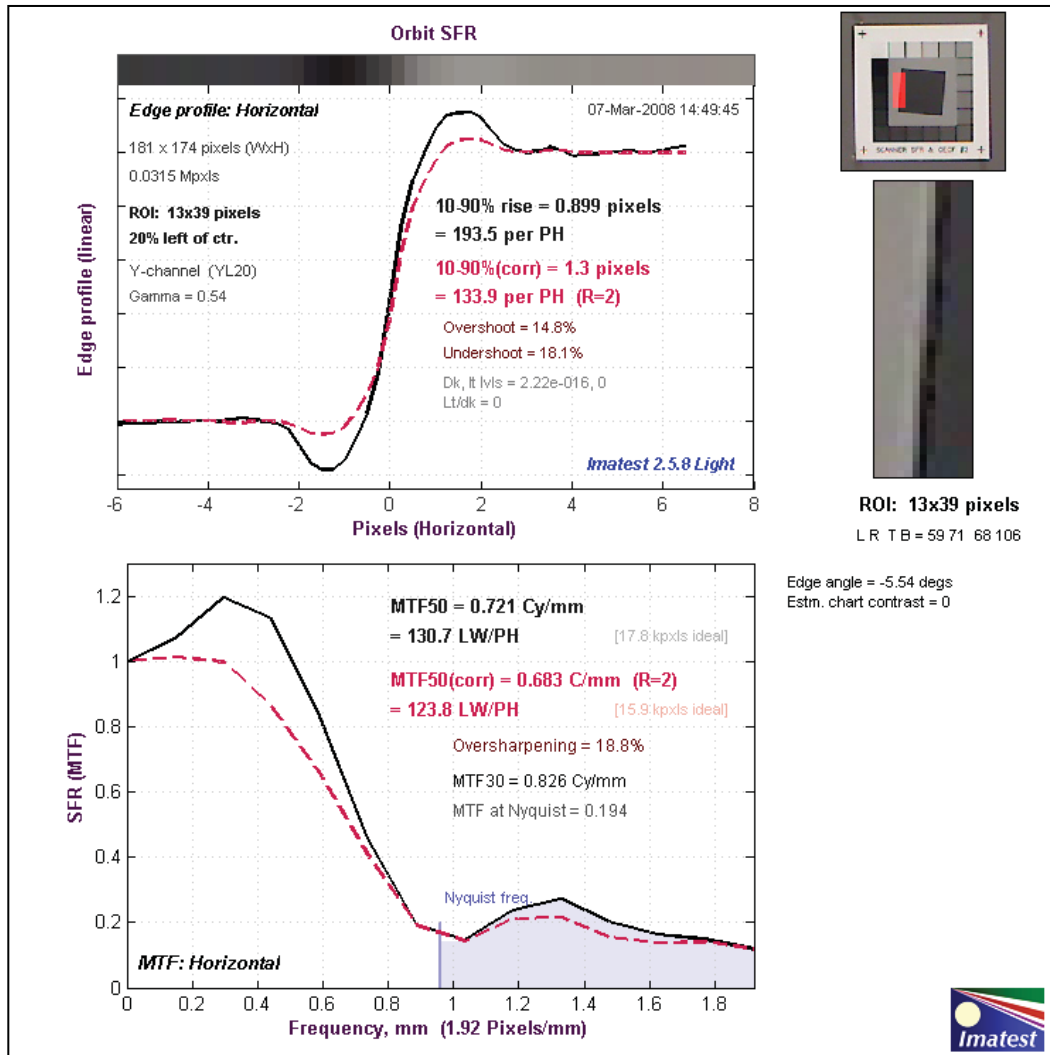


Figure 48: Logitech QuickCam Orbit AF SFR and edge profile

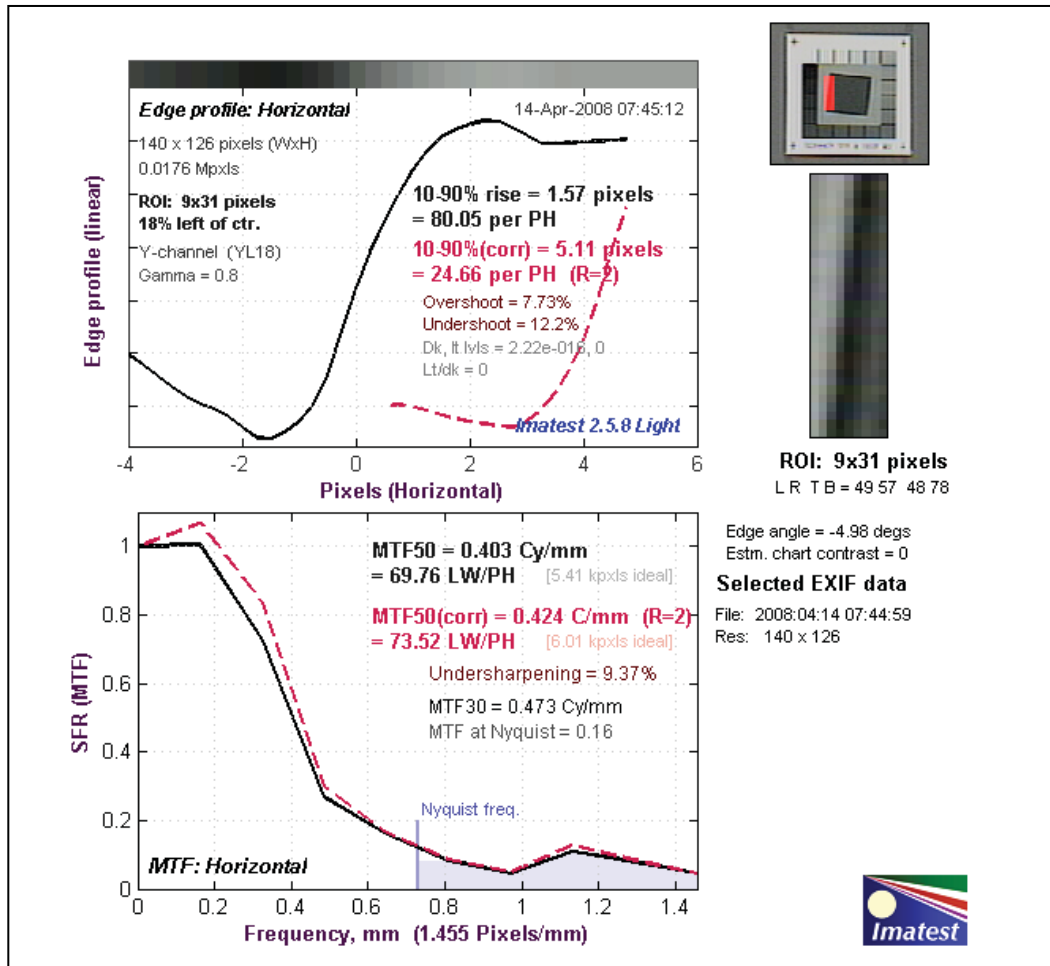


Figure 49: Prototype wide dynamic range camera SFR and edge profile

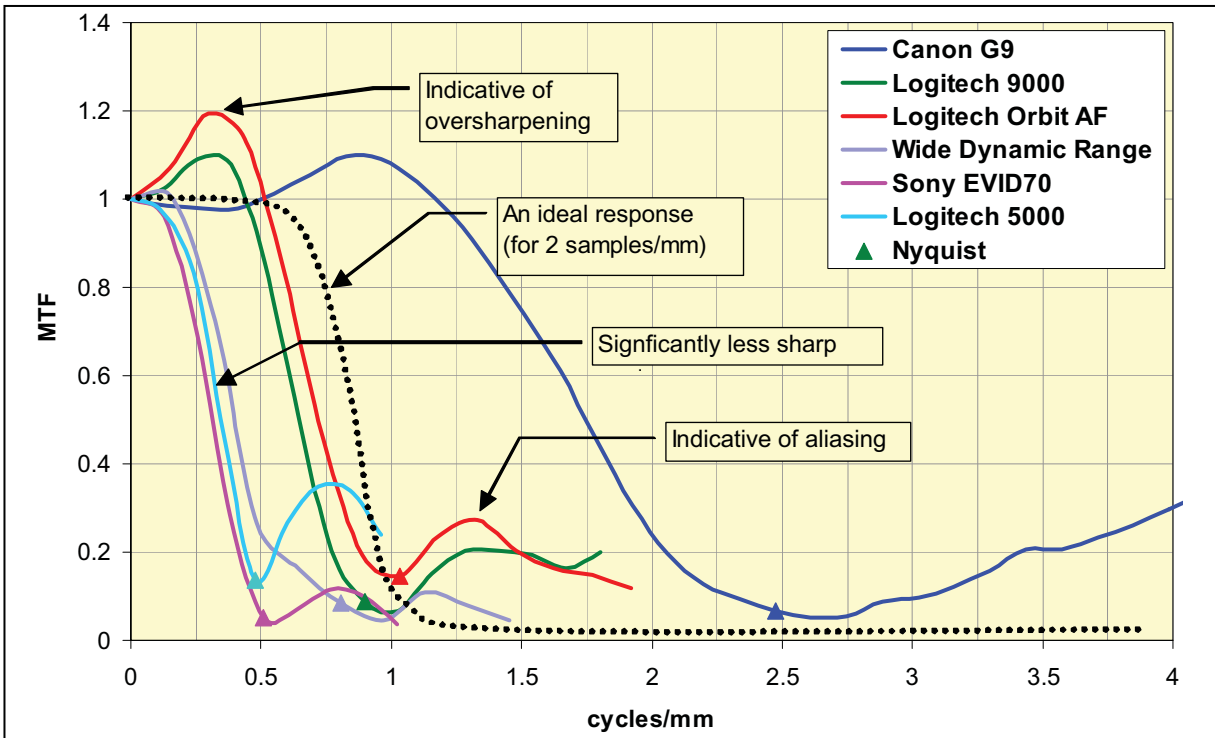


Figure 50: Combined SFR results (luminance channel)

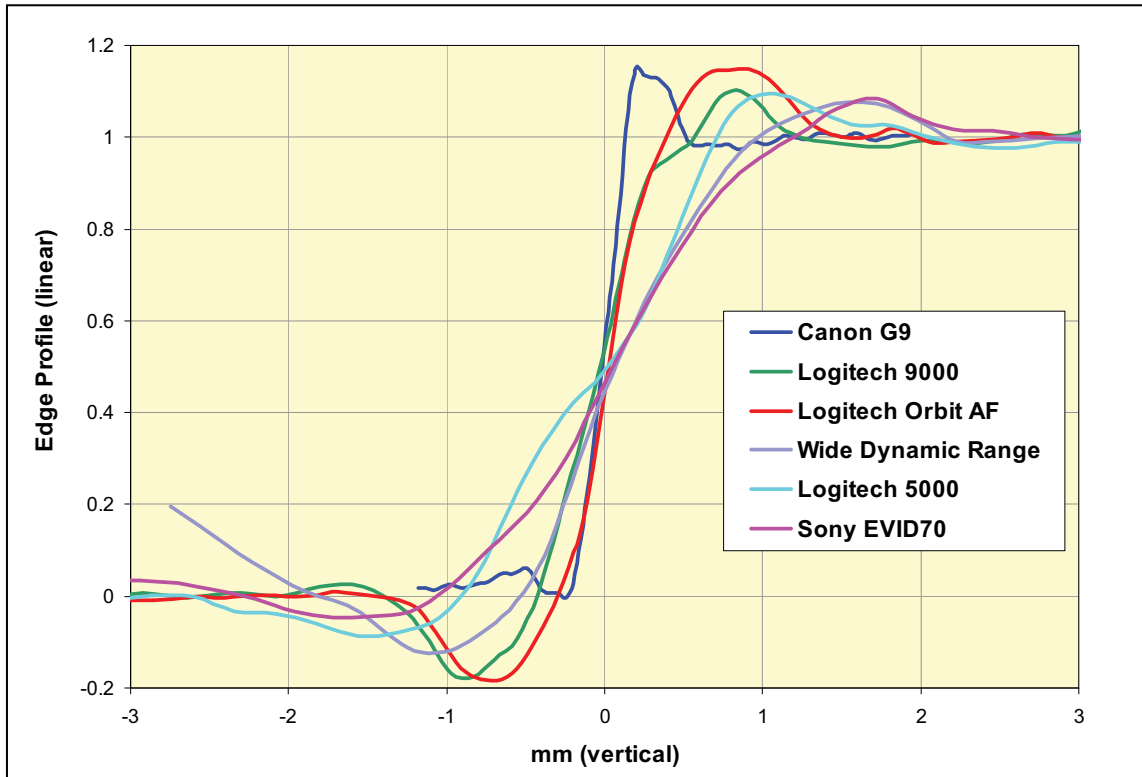


Figure 51: Combined horizontal edge profile results

All of the tested cameras exhibited overshooting in their edge profiles and aliasing beyond the Nyquist frequency. Table 8 summarizes the measurements from each camera for the sampling frequency, Nyquist frequency, MTF (50), and 10-90% rise, with the best value in green and the worst value in red in each column. The ideal response would have high MTF below the Nyquist frequency and low MTF at and above it. The MTF for the wide dynamic range camera dropped the fastest, relative to its Nyquist frequency, and it had the longest 10-90% rise (shorter is better). The Canon G9 had superior spatial resolution, as evidenced by its much higher Nyquist frequency and shorter 10-90% rise.

Table 8: Spatial Resolution Measurements

Camera	Sampling Frequency (mm/pixel)	Nyquist Frequency (cycles/mm)	MTF(50) (cycles/mm)	10-90% rise (mm)
Canon Powershot G9	0.2	2.47	1.75	0.27
Sony EVID70	0.98	0.51	0.31	1.53
Logitech QuickCam Pro 5000	1.04	0.48	0.35	1.31
Logitech QuickCam Pro 9000	0.56	0.9	0.65	0.61
Logitech QuickCam Orbit AF	0.52	1.03	0.72	0.47
Wide dynamic range camera	0.687	0.8	0.403	1.57

2.13 Light Sensitivity

The ability of a camera to create an image in a poorly lit environment is important for uncontrolled illumination at POEs. Light sensitivity is usually specified by the minimum lux required to produce a picture. A camera with good light sensitivity will have a lower minimum lux specification. The manufacturers' specifications for the light sensitivity for the Sony EVID70 and wide dynamic range camera were 1 lux and 0.8 lux, respectively. The minimum lux was not specified by the manufacturers of the other cameras in this study.

Still cameras do not use such a specification, since longer exposure times can generally be used to take pictures at very low luminance levels. The highest ISO sensitivity setting for the Canon G9 was 1600.

3 Conclusions

A desirable camera would be mountable, compact, and easy to use (e.g., auto-focus and auto-exposure), and would have high capture dimensions, fine spatial resolution, little noise, low compression, accurate color fidelity, and good light sensitivity and spatial uniformity.

3.1 Overall

- The Canon G9 image was visually superior to the images from other cameras in its color and fine detail, even at its medium capture dimensions.
- The newer Logitech models tested were superior to the webcam currently in use at the POEs (Logitech QuickCam Pro 5000) in resolution, focus, and color fidelity.
- The Logitech webcams were difficult to mount as procured. A custom mount would be necessary for use in the POE environment.
- The Logitech webcams have fixed compression levels and do not permit manual white balancing.
- Logitech QuickCam Pro 9000 and Orbit AF had similar output results.
- Logitech QuickCam Pro 5000, Sony EVID70, and the prototype wide dynamic range camera produced images with insufficient eye distances.
- When operated in portrait mode, the Logitech QuickCam Pro 9000 and Canon G9 had the largest field of view, followed by the Logitech Orbit AF in landscape mode.
- All of the cameras exhibited less than one percent of distortion, which is probably negligible for FR.

3.2 Size, Mounting, Connection, and Software

While the smallest cameras in the study were the Logitech webcams, all of the cameras were a reasonable size for the POE environment.

All of the cameras except for the Logitech webcams were mountable using a standard ¼-20 threaded screw. The Logitech webcams lacked a threaded hole, and therefore, use of them in the POE would require a custom mount. The Logitech QuickCam Orbit AF had a removable part, which may be vulnerable to loss or theft.

The Logitech webcams conveniently require only a USB connection from the computer to the camera. The rest of the cameras require additional cables. The Sony EVID70 and the prototype wide dynamic range camera have the most complex connections, with three cables and external capture card hardware.

Software drivers were provided by Logitech for the webcams. SDKs were provided for the Canon G9 and the prototype wide dynamic range camera. Sony only provided a user interface for setting camera parameters on the EVID70. A capture card and its accompanying software had to be purchased for saving output from the Sony EVID70 and the prototype wide dynamic range camera.

3.3 Visual Assessment

The Canon G9 image was visually superior to the images from other cameras in its color and fine detail, even at its medium capture dimensions. The Sony EVID70, Logitech QuickCam Pro 5000, and the prototype wide dynamic range camera exhibited much lower resolution. The Logitech QuickCam Pro 5000 and the prototype wide dynamic range camera images contained visible JPEG compression artifacts (blockiness). The images from the Logitech QuickCam Pro 9000 and Orbit AF cameras (which use the same sensor) had fairly good resolution and color fidelity. The image from the prototype wide dynamic range camera had very poor color fidelity. The newer Logitech models tested were superior to the webcam currently in use at the POEs (Logitech QuickCam Pro 5000) in resolution, focus, and color fidelity.

3.4 Capture Dimensions, Compression, and Field of View

The Canon G9 allowed for the highest capture dimensions (12 mega-pixels) and inter-eye distance, as well as the finest sampling frequency. The Logitech QuickCam Pro 9000 and Orbit AF had the next highest capture dimensions (2 mega-pixels) and a sufficiently high inter-eye distance. Images from all other test cameras had insufficient eye distances.

One shortcoming of the Logitech webcams was that their compression level is fixed. However, all of the cameras had an acceptable compression ratio of less than 20:1.

When operated in portrait mode, the Logitech 9000 and the Canon G9 had the largest vertical fields of view, with heights of approximately 3.9 and 3.5 head-lengths, respectively, followed by, in descending order, the Logitech Orbit AF (which had a height of three head-lengths), Logitech QuickCam Pro 5000, and Sony EVID70, with the prototype wide dynamic range camera having the smallest field of view, with a height of only 1.4 head-lengths. It should be noted that the lens, which affects the field of view, is interchangeable on the prototype wide dynamic range camera, but not on any of the other cameras evaluated herein.

3.5 Geometric Accuracy

The camera that produced the least amount of distortion was the Logitech QuickCam Orbit AF. All of the cameras exhibited less than one percent of distortion, which is probably negligible for FR.

3.6 Spatial Uniformity

For the most part, the results followed a trend - the larger the field of view of the camera, the less uniform was the intensity. This reduction in spatial uniformity is probably a fair trade-off for the significant benefit of having a large field of view.

3.7 Depth of Field

The only tested camera that had adequate depth of field is the Canon G9 (at capture dimensions of 2592 x 1944 pixels and higher).

3.8 Tonal response

The Canon G9 density response most closely resembled a step curve. Edge sharpening was evident in the Sony EVID70's response and in the prototype wide dynamic range camera's

response. The Logitech QuickCam Pro 5000 had the lowest maximum density, indicating that its image was too dark.

The Logitech QuickCam Orbit AF's gamma of 0.54 was closest to the optimal sRGB gamma of 0.45, while the Logitech QuickCam Pro 5000's gamma of 0.938 was the least desirable. The Logitech QuickCam Pro 5000 also detected the fewest number of zones in the density stepchart.

3.9 Noise

The cameras with the highest (best) SNR were the Logitech webcams, and the lowest (worst) SNR was from the Canon G9. The lower amount of noise in the webcam images may be due to the inherently smaller number of pixels per patch and by the JPEG compression, which reduces the variability in the patches.

3.10 Color Accuracy

A shortcoming of the Logitech webcams is that they do not permit manual white balancing, which is much more reliable than their automatic setting. While automatic white balance settings are usually acceptable, they can be incorrect when a strong color dominates the scene. The Canon G9, Logitech QuickCam Pro 9000, Logitech QuickCam Orbit AF, and Sony EVID70 had reasonable mean color errors, while the Logitech QuickCam Pro 5000 and the prototype wide dynamic range camera mean color errors were poor.

3.11 Spatial Resolution

The Canon G9, Logitech QuickCam Orbit AF, and Logitech QuickCam Pro 9000 exhibited over-sharpening. All of the tested cameras exhibited overshooting in their edge profiles and aliasing beyond the Nyquist frequency. The prototype wide dynamic range camera's MTF dropped the fastest, relative to its Nyquist frequency, and it had the longest 10-90% rise (shorter is better). The Canon G9 had superior spatial resolution, as evidenced by its much higher Nyquist frequency and shorter 10-90% rise.

3.12 Light Sensitivity

The ability of a camera to create an image in a poorly lit environment (i.e., a good light sensitivity) is important for uncontrolled illumination in POEs, although such images will possess more noise. The manufacturers' specifications for the light sensitivity for the Sony EVID70 and the prototype wide dynamic range camera were 1 lux and 0.8 lux, respectively. The minimum lux is not specified by the manufacturers of the other cameras in this study. Digital still cameras do not use such a specification, since longer exposure times can generally be used to take pictures at very low luminance levels. The Canon's highest ISO setting was 1600.

Multi-Biometric Fusion Research Plan Briefing

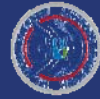
Human Factors/Behavioral Sciences Division
Science and Technology Directorate
U.S. Department of Homeland Security

Program Managers
Chris Miles, Arun Vemury, Patricia Wolfhope

07.13.09



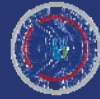
This document is confidential and is intended solely for the use and information of the client to whom it is addressed.





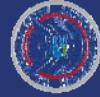
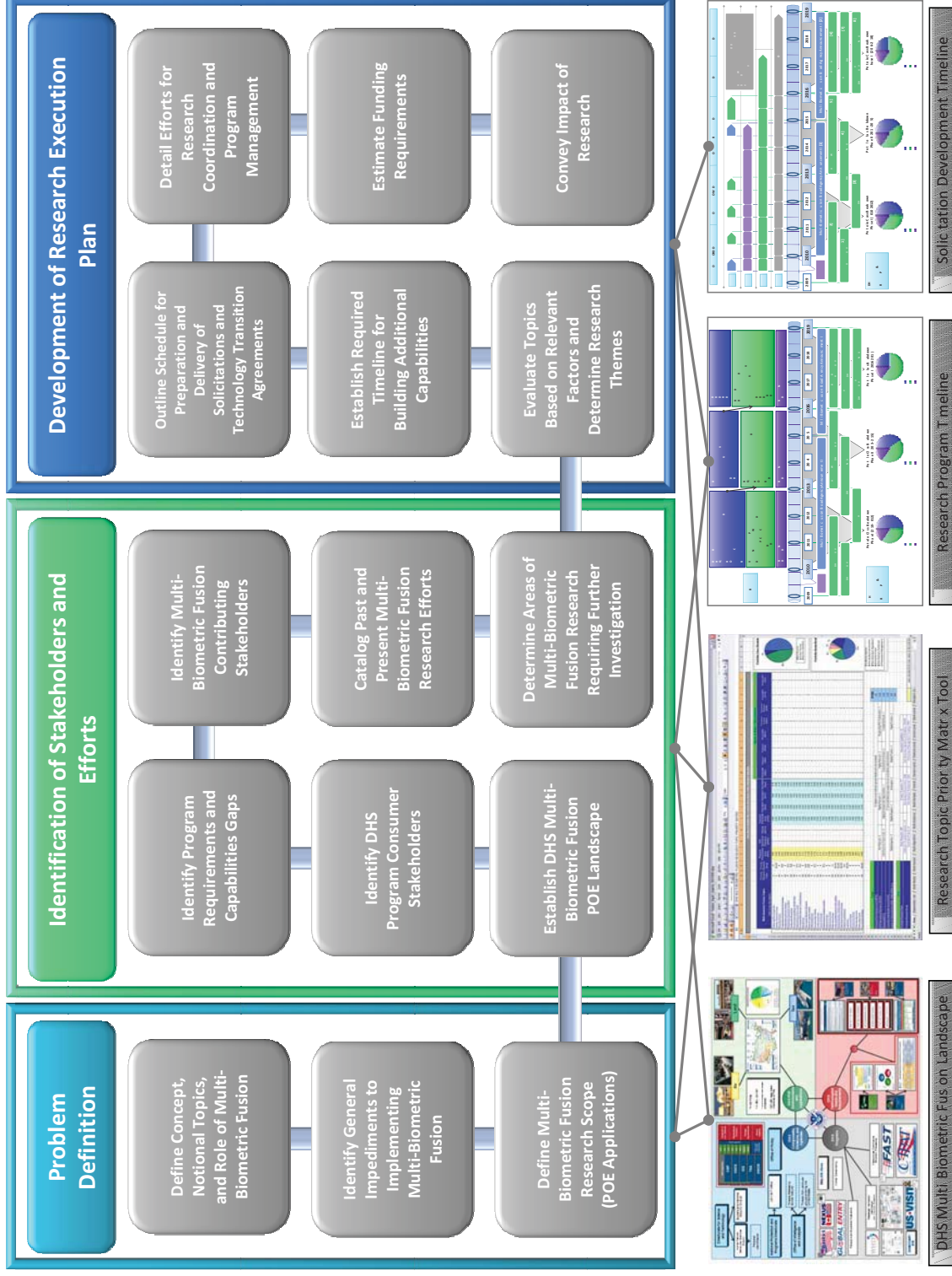
Objectives

1. **Introduce Task and Associated Deliverables**
2. **Summarize Problem Definition Content**
3. **Summarize Identification of Stakeholders and Efforts Content**
4. **Present Research Execution Plan**



Introduction

- ▶ The Department of Homeland Security (DHS) contracted Booz Allen Hamilton to address current issues related to multi-biometric fusion.
- ▶ The primary purpose of the contract was to deliver an execution plan for multi-biometric fusion research within the Human Factors and Behavioral Sciences Division of the Science and Technology Directorate of DHS.
- ▶ The development of the plan consisted of three steps:
 - Problem definition
 - Identification of stakeholders and efforts
 - Preparation of the final research execution plan



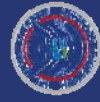
Deliverable #1 – Problem Definition

- ▶ Defining the role and importance of multi-biometric fusion in achieving an effective automated multi-biometric capability.
- ▶ Identifying impediments to the implementation of automated multi-biometric fusion for multi-biometric systems.
- ▶ Defining the role of an automated multi-biometric system for the DHS Operational Mission (DHS Mission Space).
- ▶ Synopsizing the role of multi-biometric fusion for automated multi-biometrics along with the impediments to its implementation in a comprehensive summary report (i.e., this report) and the role this report will have in the generation of the overall development of the research plan.



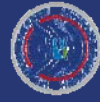
Deliverable #2 – Identification of Stakeholders and Efforts

- ▶ Identify the operational missions and operational components/offices within DHS that use or would benefit from the use of biometrics and have a need to evolve to multi-biometric systems. (face, iris, and fingerprints)
 - From these DHS operational components and offices, assess their operational need for automated multi-biometric fusion
- ▶ Identify government stakeholders outside of DHS that would benefit from the implementation of Multi-biometric systems.
- ▶ Identify how the following would impact multi-biometric decision fusion:
 - NIST’s involvement in biometrics quality research and biometrics standards.
 - Industry trends in biometrics.
 - Academia’s efforts in biometrics.
 - Standards Development Organization’s activities in biometrics.
- ▶ Synthesize identified stakeholders and efforts into a comprehensive summary report and the role this report will have in the generation of the overall development of the research plan.



Deliverable #3 – The Research Execution Plan

- ▶ Develop a comprehensive execution plan for a research program within the Department of Homeland Security, Science & Technology Directorate, to develop the integrated use of automated multi-biometric fusion systems for DHS applications.
- ▶ Creation of this plan shall take into account the following:
 - Due to the anticipated comprehensive/integrated nature of the program execution plan, DHS funding constraints may have a serious impact on the orchestration of future program activities. Program phases should be scalable and expandable, proportional to the level of funding, rather than one-time funded activities.
 - To the extent possible, the plan should seek to benefit as many common interests for multi-biometric fusion throughout the Federal Government but not at the expense of DHS-specific applications.
 - The plan should describe the relationships, resources (estimate as appropriate), responsibility, and timing of orchestration among the following entities:
 - ▶ Within DHS
 - ▶ Other Federal agencies
 - ▶ NIST
 - ▶ Academia
 - ▶ Standards Developing Organizations
 - ▶ Industry
- ▶ The research execution plan shall incorporate the synopsis plans from the Definition of the Problem and Identification of the Stakeholders and Efforts phases of this effort.
- ▶ The research execution plan should demonstrate how the proposed efforts will adequately address the impediments to implementing multi-biometric systems identified in the Definition of the Problem phase.

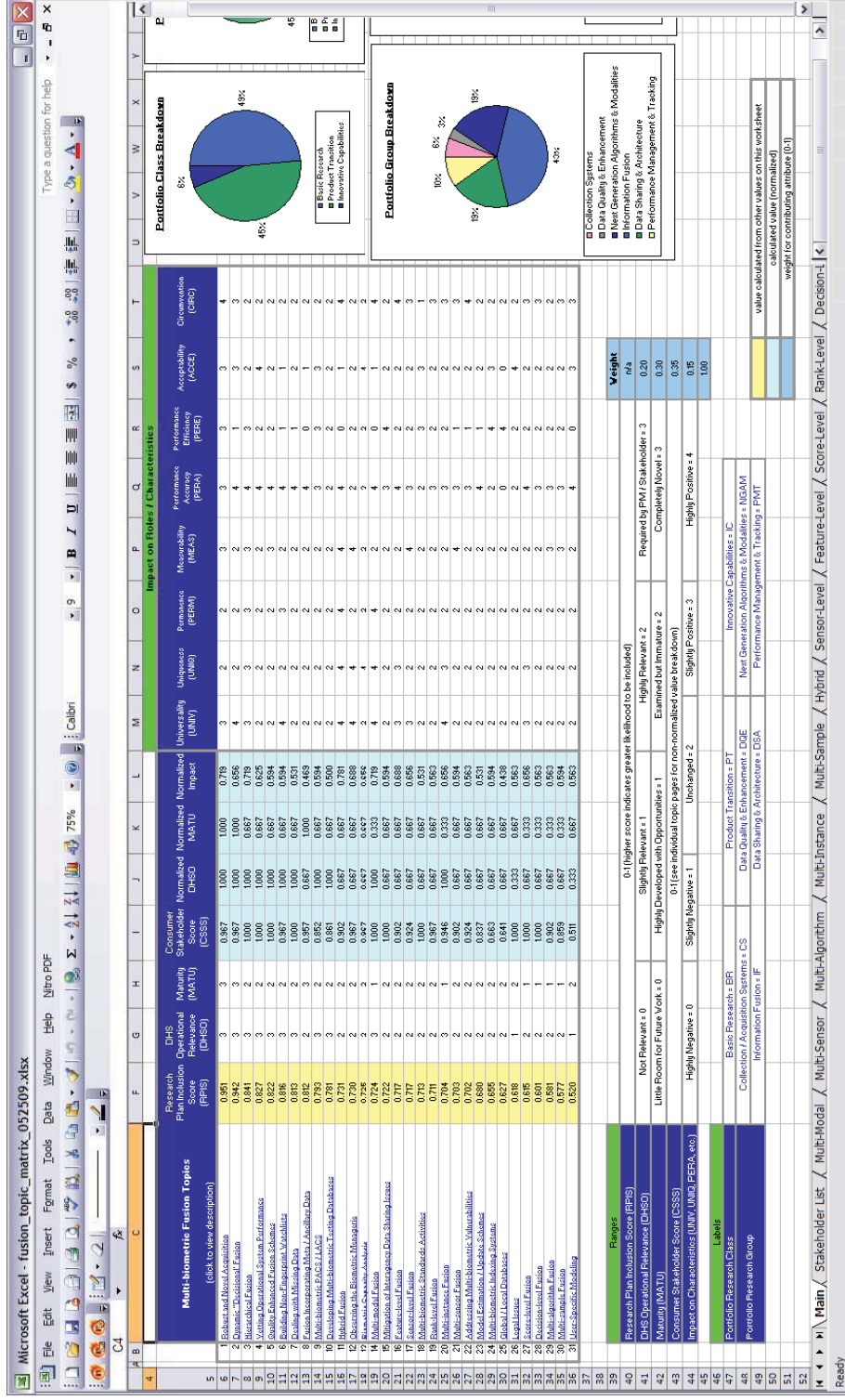


Multi-Biometric Fusion Research Topic Priority Matrix Tool

- **Multi-Biometric Fusion Research Topic Priority Matrix Tool** used to evaluate topics for consideration

- **DHS Operational Relevance**
- **Maturity of Research**
- **Consumer Stakeholder Score**
- **Impact on Characteristics of Biometric System**

- Also considers need for .lan flexibility,



Results of the Work

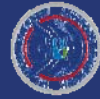
- ▶ **A 10 year research plan consisting of three phases**
 - Available on distributed CD
- ▶ **Plan outlines:**
 - Research topics
 - Timeline
 - Solicitation schedule
 - Estimated funding requirements
- ▶ **Plan describes additional capabilities and their impact on operational missions**
 -





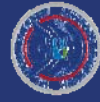
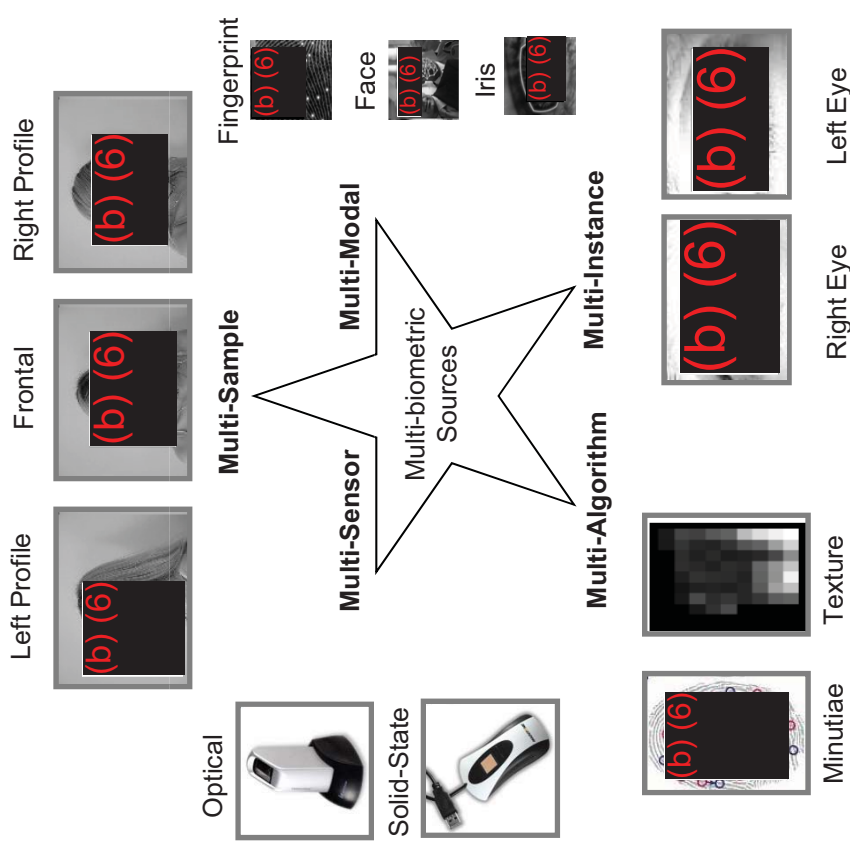
Objectives

1. Introduce Task and Associated Deliverables
2. Summarize Problem Definition Content
3. Summarize Identification of Stakeholders and Efforts Content
4. Present Research Execution Plan



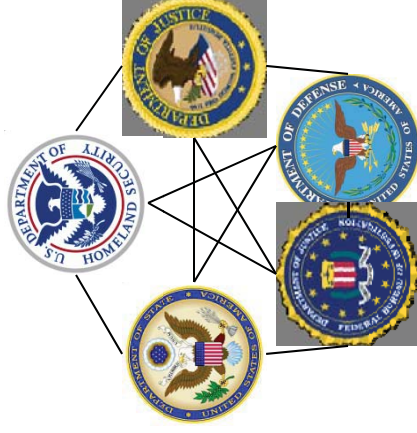
Defining the Role of and Importance of Multi-Biometric Fusion

- ▶ **Multi-Biometric Fusion** - *The process of consolidating evidence provided by multiple biometric sources* _RNJ08_.
- ▶ Generalized role of multi-biometric fusion is to improve the degree in which the 7 major characteristics of a biometric (system)
- ▶ Traditionally, this has focused on 4 characteristics
 - **Measurability** - Helps minimize noise by avoiding damaged biometrics and low quality signals
 - **Universality** - Ensuring the amount of individuals who cannot be enrolled is minimized.
 - **Performance** - Addressing upper bounds on single biometric matching accuracy.
 - **Circumvention** - Ability to address spoof attacks
 - Other characteristics include: uniqueness, permanence, & acceptability



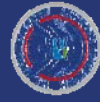
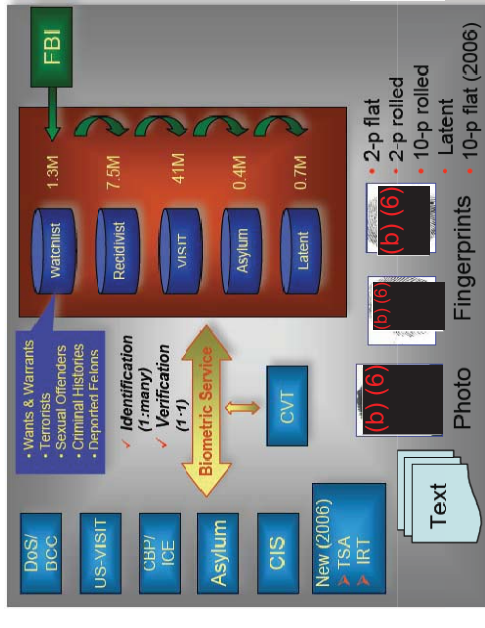
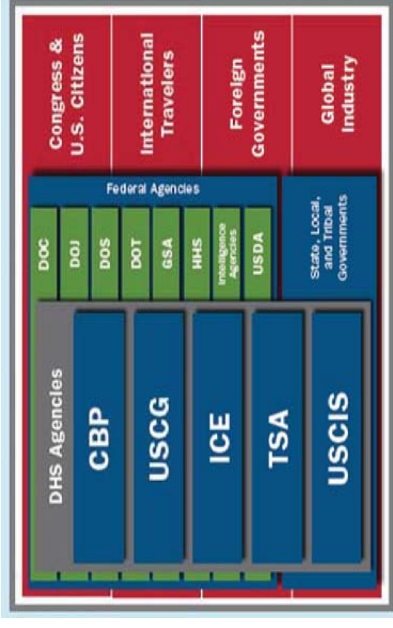
Impediments to Implementing Multi-Biometric Fusion

- ▶ A number of generalized impediments stand in the way of implementation:
 - **Acquiring Biometric Data**
Loosening constraints on data capture lead to new challenges as do increased throughput requirements.
 - **Creating New Databases**
New databases must be built before new biometric matching capabilities can be realized.
 - **Using Existing Databases**
The desired biometric data may already exist but across disparate organizations is often a difficult task.
 - **Meeting Efficiency / Throughput Requirements During Searching**
Database sizes continue to increase while the response time continues to decrease. This combination places high demands on enterprise level systems which must search against millions of enrolled identities.



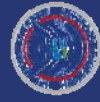
Role of Multi-biometric Systems for the DHS Operational Mission

- **Scope within the Department**
S&T, NPPD, OIA -> TSA, CBP, ICE, USCIS, USCG -> US-VISIT, FAST, SENTRI, etc.
- **Internal DHS Stakeholder Requirements**
PMs a_ need to limitin_ the o. erational environment to Port of Entry (POE) applications
- **Determine the Data DHS Collects**
10-print livescan fingerprints + 2D livescan face
Also collect non-biometric data -> name, DOB, sex, height, weight, eye color, etc.
- **Determine the Databases DHS Maintains and Uses**
DHS IDENT (ABIS) including multiple watchlists
- **Determine the Frameworks / Types of Searches Conducted by DHS Systems**
Verification, watchlist, and full identification tasks depending on client / scenario
- **Role of Multi-biometric Fusion Systems in Selected DHS Operational Locations**
Will improve characteristics such as universality, measurability, performance, etc. as well as serving to aid human inspection officers



Synopsis of the Problem

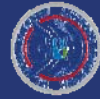
- ▶ **Generalized role of multi-biometric fusion is to improve the degree in which the 7 major characteristics of a biometric (system) are met**
- ▶ **Many considerations must be taken into account when selecting a multi-biometric fusion algorithm including**
 - Static vs. Dynamic Fusion Schemes
 - Training & Testing Considerations
 - Accommodating Missing Data
 - Model Update Schemes
 - Etc.
- ▶ **A number of generalized impediments stand in the way of implementation**
 - Acquiring Biometric Data
 - Creating New Databases & Using Existing Databases
 - Meeting Efficiency / Throughput Requirements During Searches
- ▶ **The scope of the problem focused on DHS Port of Entry (POE) environments**
 - Applications extend well beyond POE applications



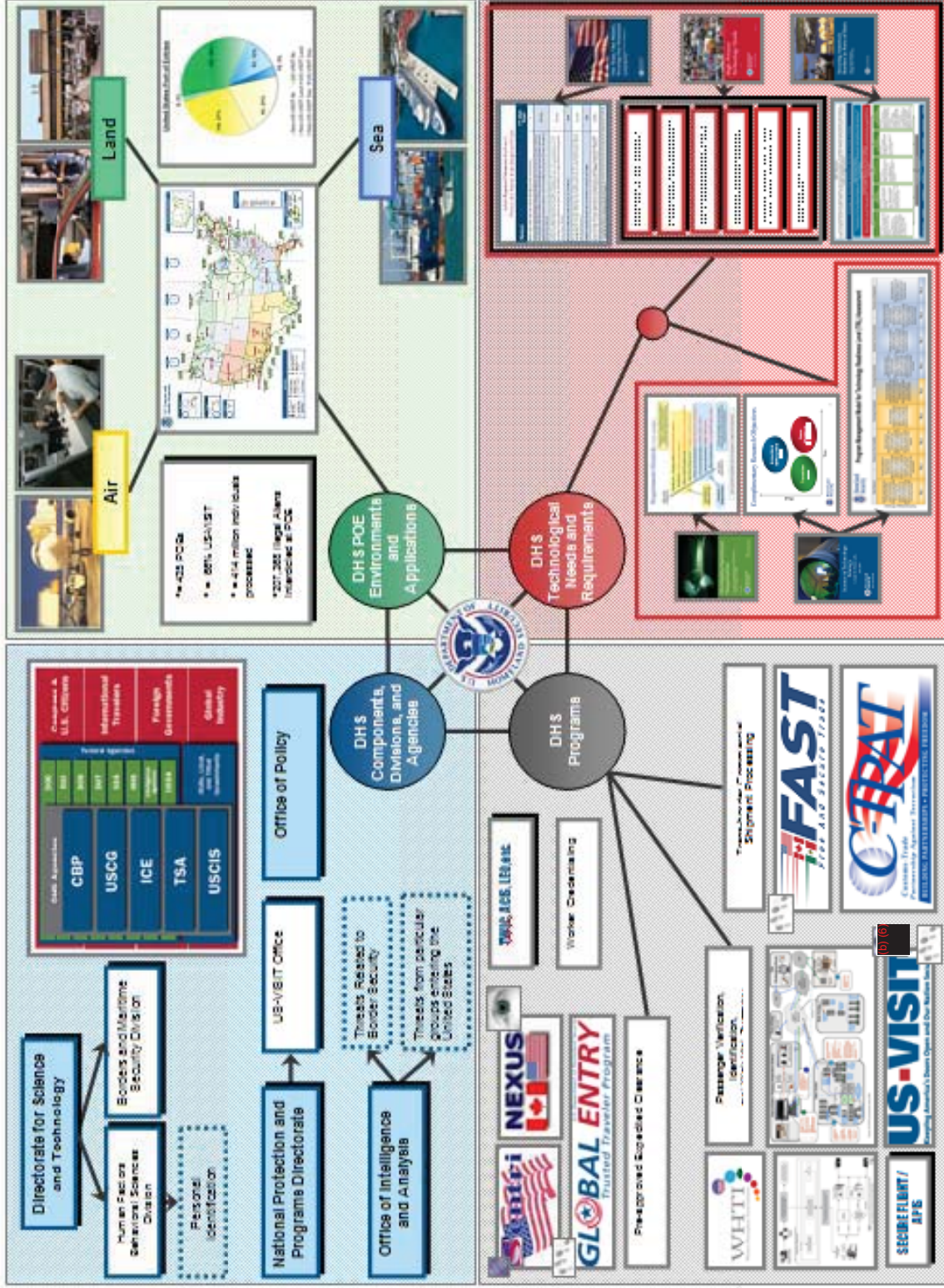


Objectives

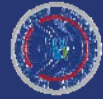
1. Introduce Task and Associated Deliverables
2. Summarize Problem Definition Content
3. Summarize Identification of Stakeholders and Efforts Content
4. Present Research Execution Plan



High Level View of DHS Multi-Biometric Fusion POE Application Landscape*



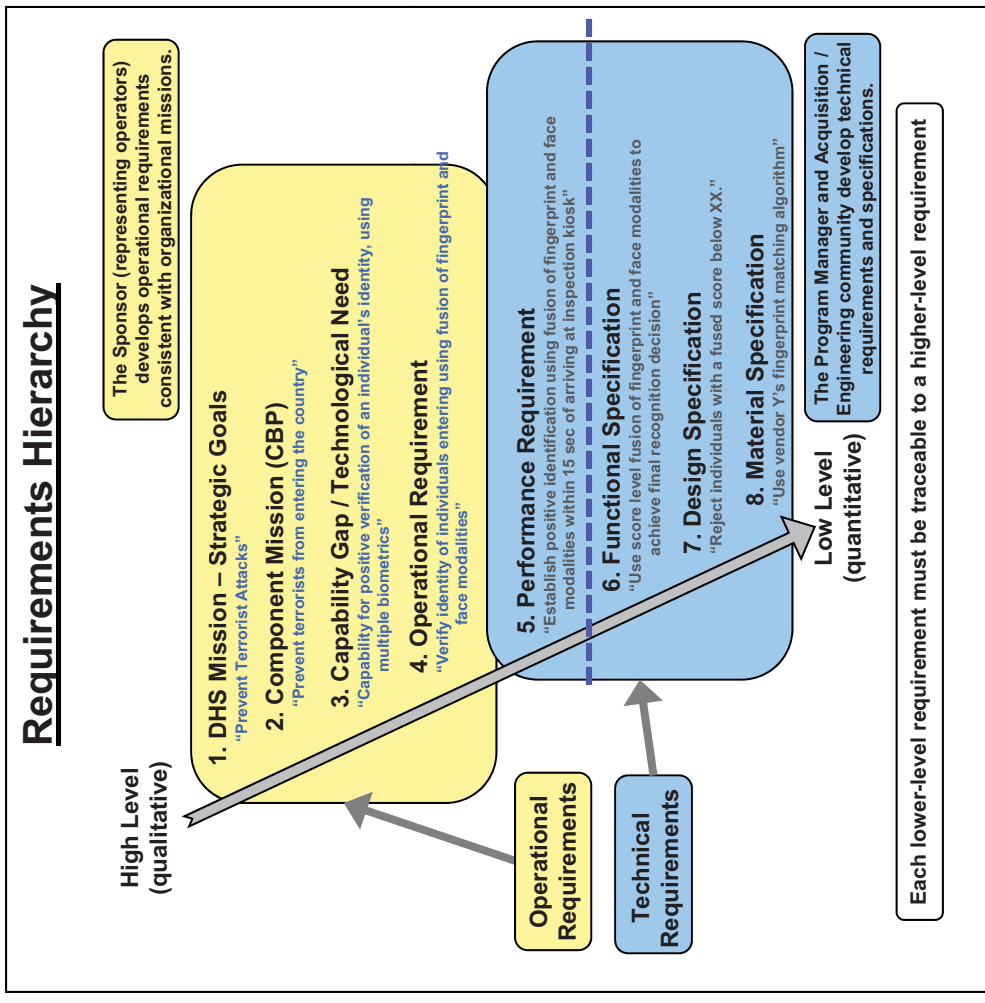
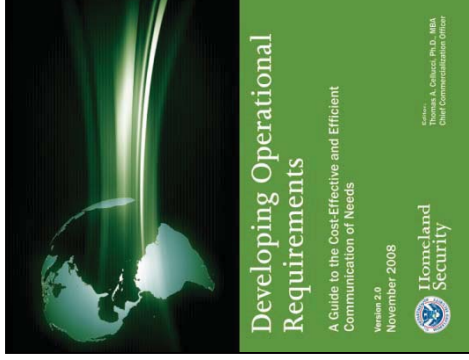
*Enlarged Version Available



Example DHS Operational Missions, Operational Requirements, and Performance Requirements

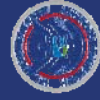
► Strategic Goals Level 1 Requirements

- **1. Protect Our Nation from Dangerous People**
 - Objective 1.1 - Achieve Effective Control of Our Borders
 - Objective 1.2 - Protect Our Interior and Enforce Immigration Laws
 - Objective 1.3 - Strengthen Screening of Travelers and Workers
- **5. Strengthen and Unify DHS Operations and Management**
 - Objective 5.2 - Advance Intelligence and Information Sharing
 - Objective 5.3 - Integrate DHS policy, planning, and operations coordination



Example DHS Operational Mission, Operational Requirements, and Performance Requirements

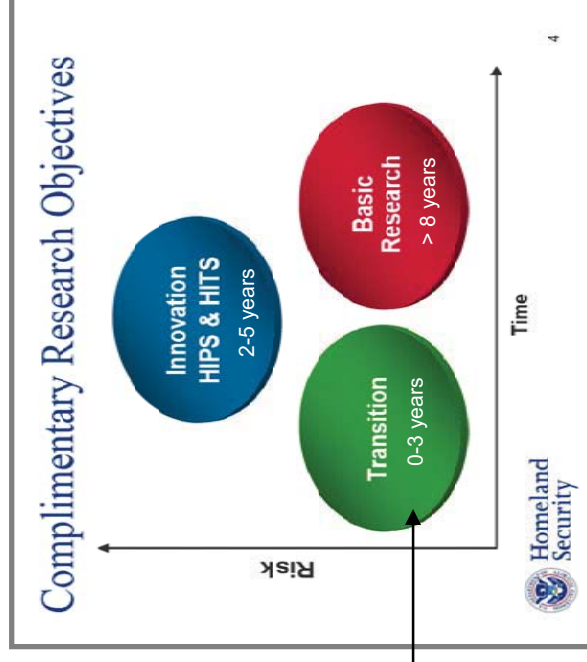
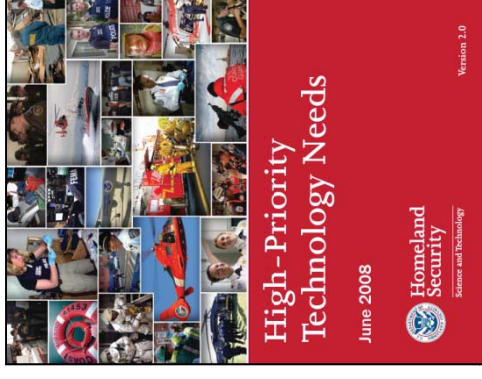
- ▶ **Component Missions (Level 2 Requirements)**
- ▶ **Example**
 - **Directorate for Science and Technology:** *“The S&T Directorate’s mission is to improve homeland security by providing to our customers, the operating components of DHS and state, local, tribal and territorial emergency responders and officials, state-of-the-art technology that helps them accomplish their missions [Coh07].”*
- ▶ **Relevant Mission Statements Identified for the following DHS Components**
 - **Directorate for Science and Technology**
 - **National Protection and Programs Directorate**
 - **Office of Intelligence and Analysis**
 - **Office of Policy**
 - **Transportation Security Administration (TSA)**
 - **Customs and Border Protection ,CBP ,**
 - **United States Citizenship and Immigration Services (USCIS)**
 - **United States Immigrations and Customs Enforcement (ICE)**
 - **United States Coast Guard USCG**



Example DHS Operational Mission, Operational Requirements, and Performance Requirements

► Technological Needs (Level 3 Requirements)

High-Priority Technology Need	Capstone IPT	Component Lead
Capability in real time for positive verification of an individual's identity, using multiple biometrics - In particular, face, fingerprint, and iris	People Screening	Human Factors / Behavioral Sciences Division
Mobile biometrics screening capabilities, including handheld, ten-fingerprint capture, environmentally-hardened, wireless, and secure devices	People Screening	Human Factors / Behavioral Sciences Division
Remote, standoff biometrics detection for identifying individuals at a distance	People Screening	Human Factors / Behavioral Sciences Division
Improved analysis and decision-making tools that will ensure the development and implementation of border security initiatives	Border Security	Borders & Maritime Division
Improved cross-agency reporting of suspicious activity - In particular, technologies that would improve real time awareness through alerting others to and sharing information about suspicious activities and persons	Information Sharing	Command, Control, & Interoperability Division
Management of user identities, rights, and authorities - In particular, technologies and standards to enable external identity adjudication	Information Sharing	Command, Control, & Interoperability Division
Information sharing within and across sectors on terrorist threats - In particular, analytic capabilities for structured, unstructured, and streaming data	Information Sharing	Command, Control, & Interoperability Division



Example DHS Operational Mission, Operational Requirements, and Performance Requirements

Component	Strategic Goal (Level 2)	Operational Requirements (Level 4)
CBP	<p>1. Advance Knowledge - increasing and improving the information and analysis CBP has about people, goods, and conveyances, before they arrive at the ports of entry</p> <p>2. Effective Inspections - screening all people, goods, and conveyances and examining them according to their assessed risk level.</p>	<p>1.1 Increase scope and accuracy of information gathered on people, goods, and conveyances ahead of arrival at the border.</p> <p>1.2 Implement a highly effective risk management process by performing advance analysis on collected information to identify potential threats prior to their arrival and to enable screening prioritization.</p> <p>2.1 Screen all people, goods, and conveyances crossing the border at the POE.</p> <p>2.2 Maintain flexible, agile, and streamlined inspection process.</p> <p>2.3 Improve recording and use of border crossing, inspection, and enforcement results.</p>
USCIS	<p>1. Strengthen the security and integrity of the immigration system</p>	<p>1.1. Enhance the security of the United States by ensuring that immigration benefits are granted only to eligible applicants and petitioners.</p> <p>1.2. Deter, detect, and pursue immigration-related fraud.</p> <p>1.3. Identify and share immigration-related information with partners.</p>
TSA, ICE, USCG		Component-wide strategic plans not available
US-VISIT	<p>1. Enhance the ability to determine which foreign nationals have legally entered and/or exited the United States.</p> <p>2. Establish and verify the identity of individuals in a timely, accurate, and reliable manner.</p>	<p>1.1. Ensure entry encounters are recorded and biometrically verified.</p> <p>1.2. Ensure exit encounters are recorded and biometrically verified.</p> <p>2.1. Associate an individual's encounters with homeland security organizations to that individual's unique biometric identity.</p> <p>2.2. Ensure US-VISIT systems can meet the growing demand for identification services.</p> <p>2.3. Advance the adoption of technical standards, processes, and specifications for the exchange of biometric data.</p>



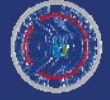
Example DHS Operational Mission, Operational Requirements, and Performance Requirements

Component	Goal / Objective	Performance Measure Level 5	FY '08 Result	FY '08 Target	FY '09 Target	FY '13 Target
USCG	1 1	Percent of undocumented migrants who attempt to enter the U.S. via maritime routes that are interdicted [As estimated, based upon data obtained from the U.S. Coast Guard and U.S. Customs and Border Protection]	62.7%	65%	69.9%	71.5%
CBP	1 1	Percent of apprehensions at Border Patrol checkpoints	2%	3-8%	>3%	N/A
ICE	1 2	Number of illegal aliens removed from the United States	N/A	N/A	342,251	N/A
		Percent of illegal aliens removed from the U.S. based on the number of illegal aliens processed for immigration law violations during the same period	N/A	N/A	68%	N/A
		Percent of closed investigations which have an enforcement consequence (arrest, indictment, conviction, seizure, fine or penalty)	46.3%	36.6%	47%	N/A
CBP	1 3	Air Passenger Apprehension Rate for Major Violations [Percent of the total number of individual passengers with major violations of customs and immigration laws and regulations that were apprehended based on statistical estimates of the total number of violations that came through our international airports]	25%	40%	25%	43.5%
		Land Border Apprehension Rate for Major Violations [Percent of the total number of vehicles travelers with major violations of customs and immigration laws and regulations that were apprehended based on statistical estimates of the total number of violations that came through the Points of Entry (POEs)]	28.9%	35%	28%	37.5%
		Percent of individuals screened against law enforcement databases for entry into the United States	73.5%	N/A	80%	N/A
US-VISIT(NPPD)	1 3	Average biometric watch list search time for queries from BioVisa	2.34 min	<5 min	< 5 min	N/A
		Average biometric watch list search times for queries from U.S. ports of entry	9.67 sec	<10 sec	< 10 sec	N/A
		Percent of biometrically screened individuals inaccurately identified as being on a US-VISIT watch list	0.0197%	<0.13%	<0.04%	N/A
TSA	1 3	Percent of in-country overstay leads deemed credible and forwarded to Immigration and Customs Enforcement for further investigation	25%	23%	25%	N/A
OIA	5 2	Passenger security screening assessment results	*	*	*	N/A
OIA	5 3	Percent of component-to-component information sharing relationships complying with Information Sharing and Access Agreement (ISAA) guidelines	70%	75%	80%	N/A
OIA	5 3	Percent of breaking homeland security situations disseminated to designated partners within targeted timeframes	N/A	N/A	80%	N/A



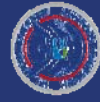
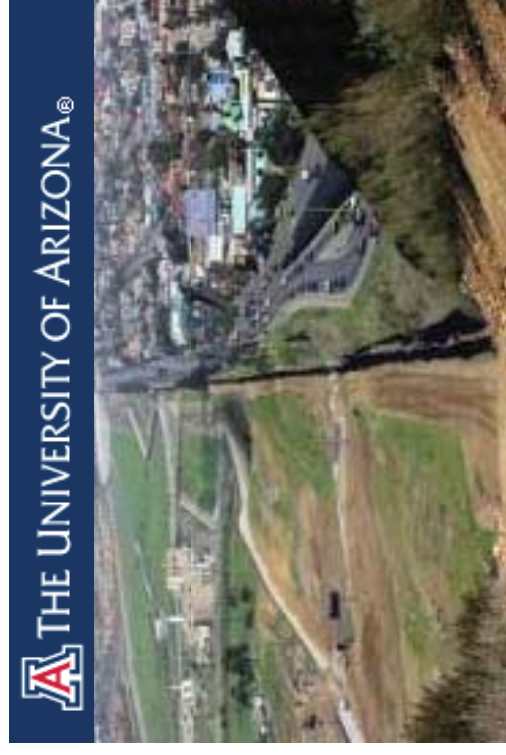
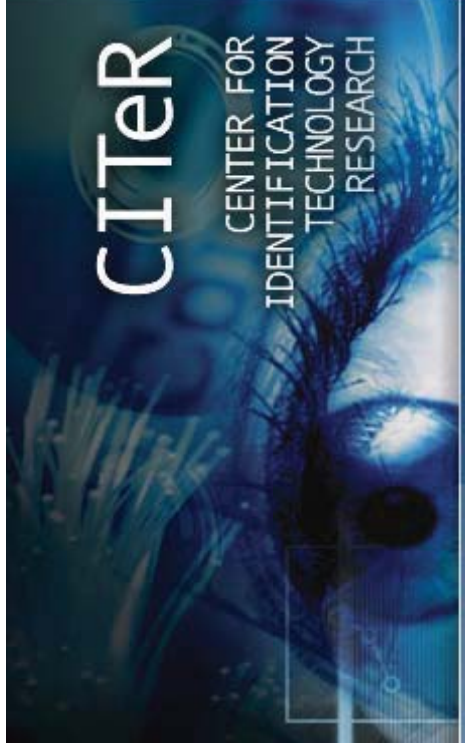
Example DHS Operational Mission, Operational Requirements, and Performance Requirements

Component	Goal / Objective	Performance Measure Level 5	FY '08 Result	FY '08 Target	FY '09 Target	FY '13 Target
US-VISIT(NPPD)	1 1	Number/percentage of total designated locations that can record and biometrically verify entry				
	1 2	Percentage of in-scope population recorded and biometrically verified at designated locations				
	2 1	Number/percentage of designated locations that can biometrically record exit				
		Number/percentage of in-country overstay leads deemed credible and forwarded to ICE				
		Biometric Watchlist FAR				
		Ratio of identity verifications to identity establishments				
	2 2	Number/percentage of SLA timeliness and reliability criteria met				
	2 3	Number/percentage of biometric standards proposed by the U.S. that are adopted by domestic and international standards organizations				
	3 1	Number/percentage of in-country overstay leads deemed credible and forwarded to ICE				N/A
	3 2	Out-of-country overstay lookout credibility rate				
		Biometric watchlist FAR				
	3 3	Number of U.S. and foreign government organizations as appropriate and authorized with which US-VISIT shares information				
		Number/percentage of open recommendations from GAO and OIG closed or acknowledged for substantial progress				
	4 2	Average cost performance of US-VISIT development and deployment activities				
	Average schedule performance of US-VISIT development and deployment activities					
	Number/percentage of management processes improved based on CMMI recommendations					
	Average processing time for privacy redress requests completed					



Stakeholders Who Might Contribute to DHS Multi-Biometric Fusion Research

- ▶ **Academic Centers of Excellence**
 - Center for Identification Technology Research (**CITeR**) WVU / Arizona
 - ▶ 13 academic institutions
 - National Center for Border Security and Immigration (**NCBSI**) Arizona
 - ▶ 16 academic institutions
 - Center for Applied Identity Management Research (**CAIMR**) Indiana University
 - ▶ 2 academic institutions
 - Center for Academic Studies in Identity Sciences (**CASIS**) Carnegie-Mellon University
 - ▶ 4 academic institutions



DHS S&T Funded Research Activities

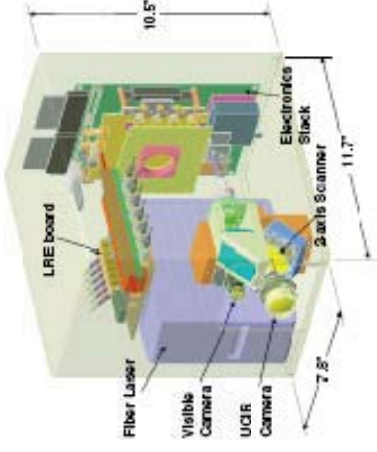
▶ CITER

- 14 Projects related to Multi-Modal Fusion over the last 7 Years



▶ Information Technology Research

- 6 Task Effort, 4 Universities
- Performance Evaluation (Dynamic Decisional Fusion)
- Reduction of Barriers (multi-biometrics / vulnerabilities)
- Social Impact Study
- Implications of Technology Maturity
- Biometric System Design Methodologies
- Realization of Broader Inquiries

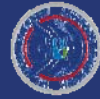
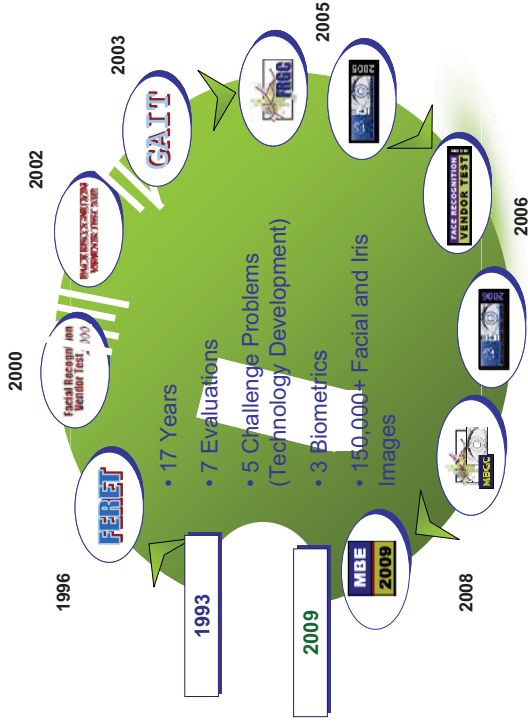


▶ SBIRS

- 7 Phase I Efforts with Relevance to Multi-biometric Fusion

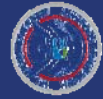
▶ NIST Challenge Events / Vendor Tests

- Multiple Biometric Grand Challenge (MBGC)
- Iris Challenge Evaluation (ICE) / Face Recognition Vendor Test (FRVT)
- Evaluation of Latent Fingerprint Technologies (ELFT-EFS)



Research That Can Be Performed If Funded

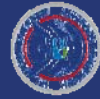
Portfolio Group	Definition
Collection / Acquisition Systems (CS)	Includes research topics specifically aimed at acquired biometric data / samples. In order to perform any sort of multi-biometric fusion, one must first have access to multiple sources of biometric input. Research in this area may involve development of new sensors, capture devices, communication strategies in multiple sensor environments, etc.
Data Quality & Enhancement	An important driver in multi-biometric fusion is data quality. Research in this grouping can consist of the development of quality metrics which can be used to automatically assess various factors impacting biometric feature extraction and matching processes. Additionally, it can include the ability to identify poor quality images or regions within biometric images as well as proposing approaches for rectifying or improving on poor quality samples and regions.
Next Generation Algorithms & Modalities	Next generation algorithms and modalities is a rather broad category in which many topics may fall. One aspect of next generation algorithms includes the development of algorithms which will have a drastic impact on the way systems operate. In other words, algorithms which may be capable of causing a paradigm shift in the application of biometric technology. Additionally, this category will include the research and development of new biometric modalities as well as extended feature levels of currently existing modalities such as level III face and fingerprint features or macro-level iris features. These different types of feature may have important uses as inputs to multi-biometric fusion systems.
Information Fusion	Information fusion is the core in which multi-biometric fusion research takes place. Whether investigating different types and levels of fusion, score normalization, or generalized information fusion approaches which consolidate biometric and non-biometric input, topics within this group should focus on efforts to meaningfully fuse multiple sources of information.
Data Sharing & Architecture	The data sharing & architecture group is another area which focuses on the availability of biometric input but also issues such as efficiency of performance, integration strategies, interoperability, etc. In many cases, biometric information may be available from a variety of disparate sources. In these cases there may be a number of technological as well as policy challenges which need to be overcome in order to implement a multi-biometric system. As outlined in the problem definition biometric programs are typically subject to ever increasing demands in terms of efficiency which brings to bear issues of integration and scalability. These types of issues fall within this grouping.
Performance Management & Tracking	The final research topic grouping involves the management and tracking of multi-biometric system performance. An important aspect of meeting quantitative goals is the ability to provide meaningful and accurate statistics to monitor progress toward meeting those goals. It is therefore necessary to research and develop systems which are capable of thoroughly tracking system performance on information which is available as well as estimating (to the greatest degree possible) performance information which is not readily available.





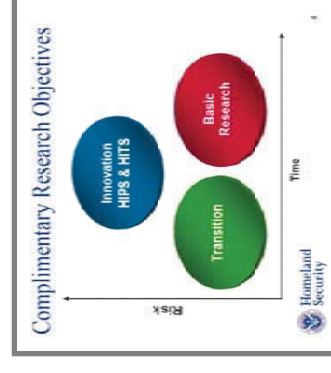
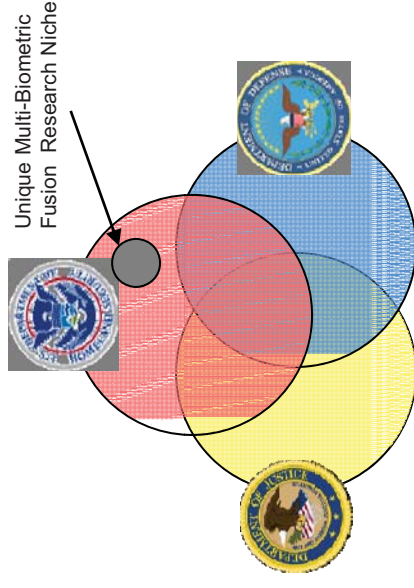
Outline

- 1. Introduce Task and Associated Deliverables**
- 2. Summarize Problem Definition Content**
- 3. Summarize Identification of Stakeholders and Efforts Content**
- 4. Present Research Execution Plan**



Rationale Behind the Development of the Plan

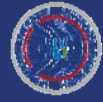
- ▶ Drivers in the Development of the Plan
 - Research resulting from the execution of the plan should support the stakeholders and programs that DHS S&T HFD serves
 - The plan should define the unique multi-biometric fusion research niche within the greater U.S. Government
 - Forward progress on research topics within the plan should positively impact the characteristics of a biometric system (universality, uniqueness, measurability, performance, etc.)
 - Development of the plan should abide by the previously defined framework of the DHS research system (basic research, product transition, innovative capabilities)



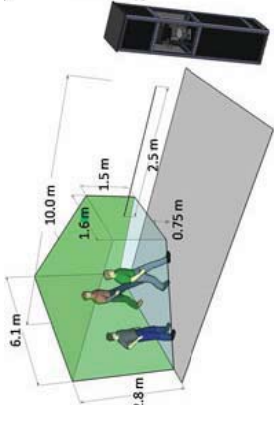
Research Scope, Timeline, and Plan Flexibility

- ▶ **Scope**
 - Consider totality of biometrics applications / DHS programs with focus on POE environments
- ▶ **Timeline**
 - 10 year timeline was required to accommodate **research life cycle** without extending beyond the scope of reasonable prediction
- ▶ **Plan Flexibility**
 - Adopted plan allows for flexibility and scalability of content to accommodate future changes

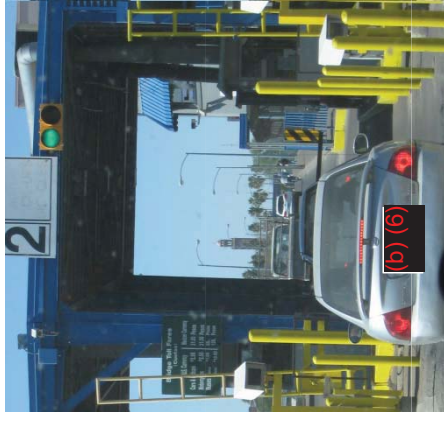
Research Class	Notional Description
Basic Research	<ul style="list-style-type: none"> • Enables future paradigm changes • University fundamental research • Government lab discovery and invention
Product Transition	<ul style="list-style-type: none"> • Focused on delivering near-term products / enhancements to acquisition • Customer IPT controlled • Cost schedule, capability metrics
Innovative Capabilities	<ul style="list-style-type: none"> • High Risk / High Payoff • “Game Changer / Leap Ahead” • Prototype, Test, and Deploy



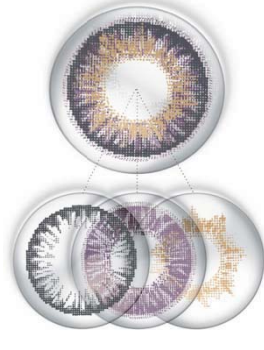
Capability Gap Analysis



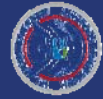
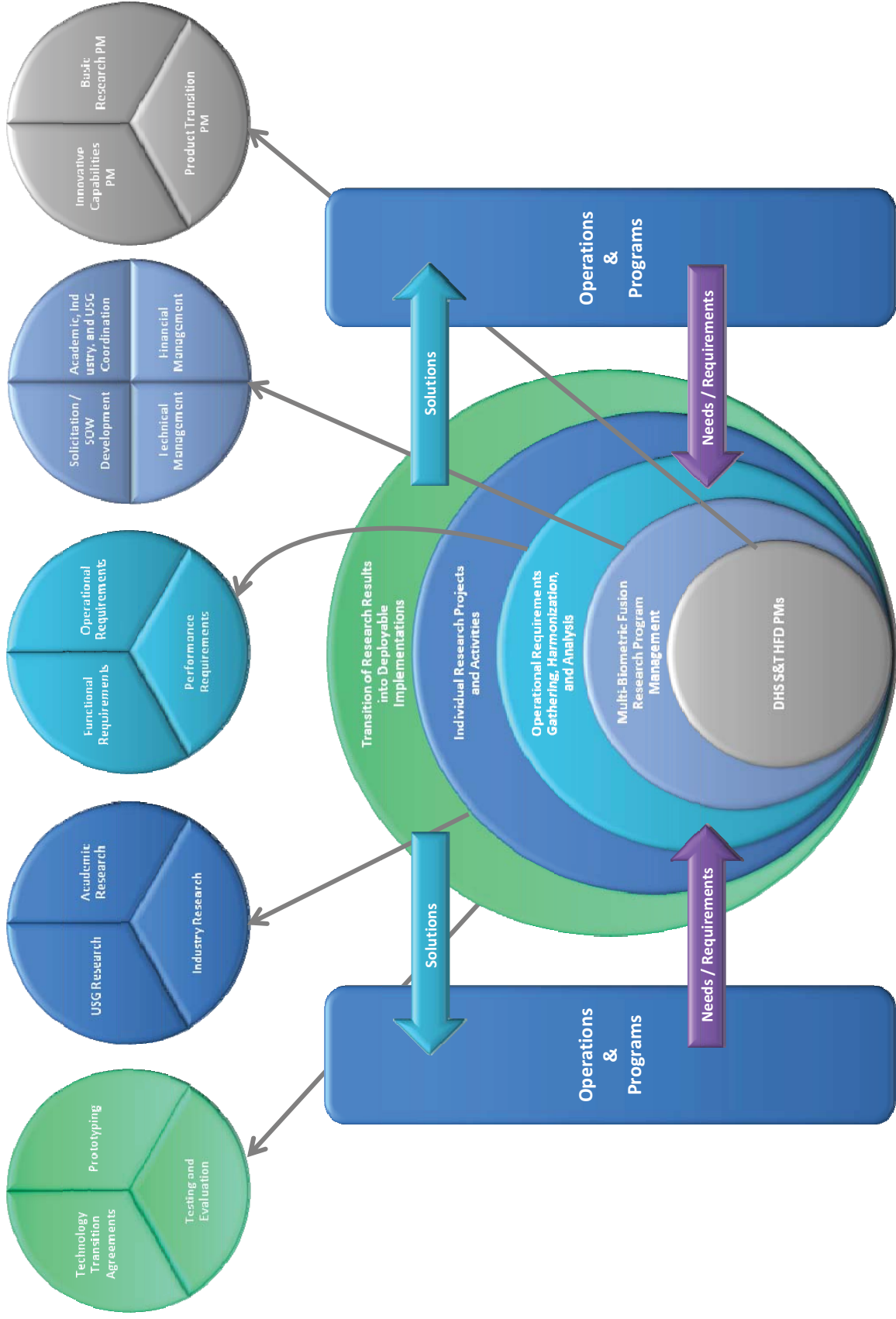
- ▶ **Required Capabilities Identified**
 - Multi-Modal Biometric People Screening
 - Mobile Biometric Screening Capabilities
 - POE Worker Multi-Biometric Authentication
 - Trusted Traveler Standoff Multi-Biometric Verification
 - Unmanned Walking Subject Multi-Biometric People Screening
 - Vehicular Multi-Biometric People Screening
 - Multi-Biometric Dynamic Decisional Fusion Systems
 - Multi-Biometric Vulnerability Countermeasures



- ▶ **Topics Requiring Further Investigation to Achieve Capabilities**
 - Over 30 topics developed
 - 22 topics selected based on previously described methodology

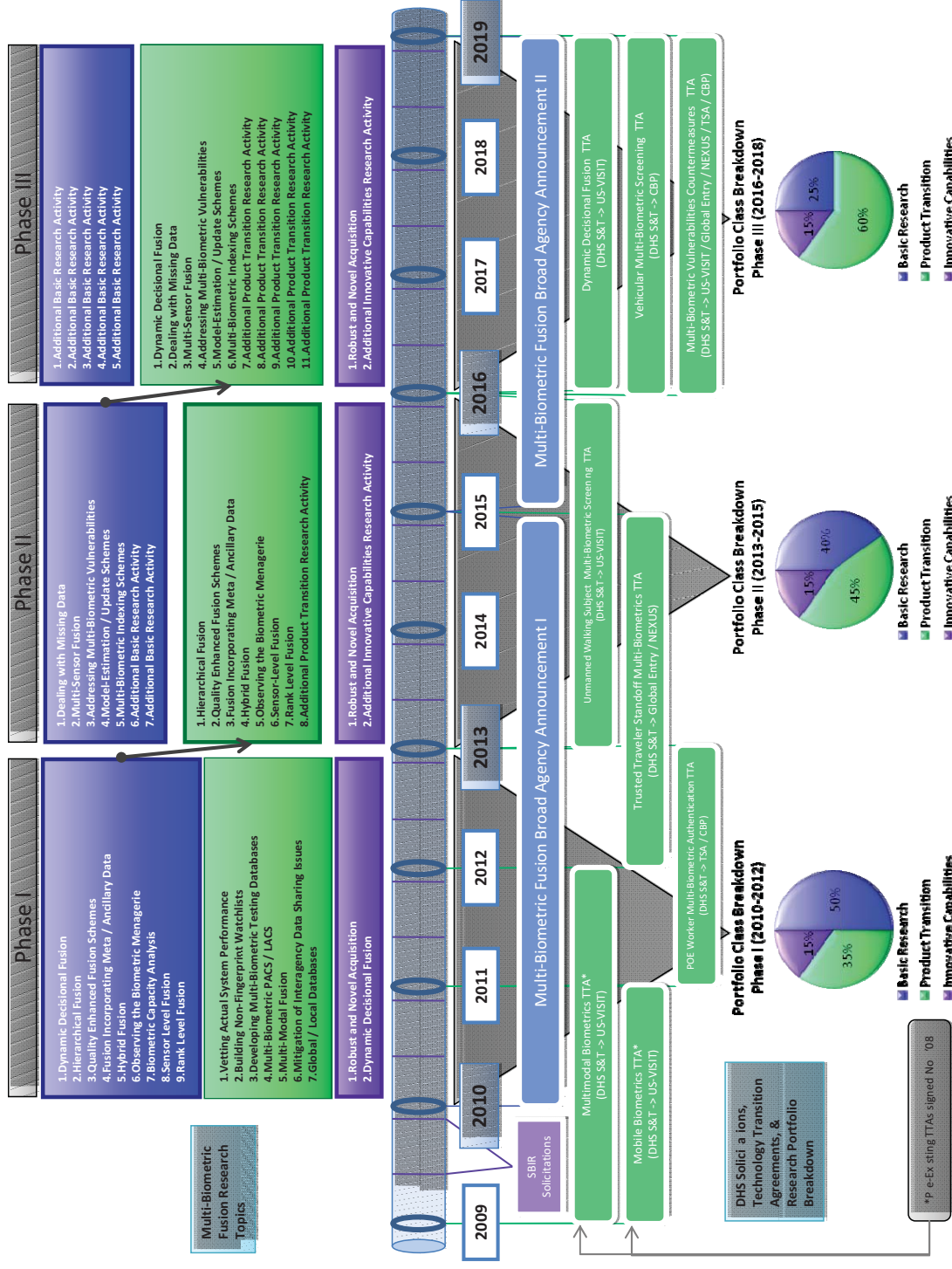


Notional Structure of Multi-Biometric Fusion Research Program



Multi-Biometric Fusion Research Plan Timeline

- ▶ 10 year timeline
- ▶ 3 Phases + an Integration / Transition Period
- ▶ Shifting balance of research topics across basic and product transition research
- ▶ Broad range of contract vehicles for research acquisition



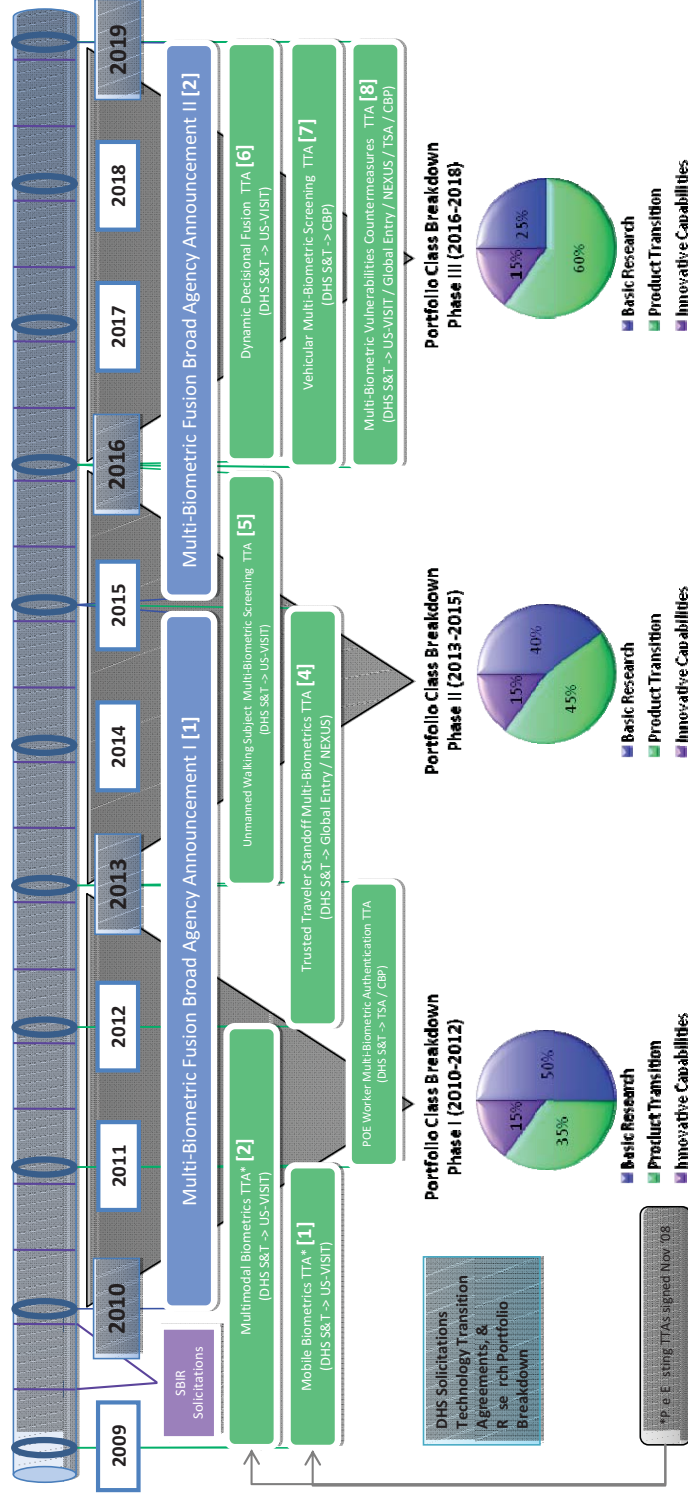
Multi-Biometric Fusion Research Solicitation Timeline

- ▶ Life cycles within each class of research activity



- ▶ Basic research efforts feed directly into product transition activities

- ▶ Innovative Capabilities efforts provide supplemental support to the research base



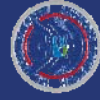
Expected Funding Requirements

Activity	Estimated Funding Requirement
Support for Development of Statement of Needs, Statement of Works, and Contract Vehicles	\$350,00
Program Management Support	\$75,000
Total Estimated Funding for Integration Period	\$425,000

Activity	Estimated Funding Requirement
Support for Development of Statement of Needs, Statement of Works, and Contract Vehicles	\$500,00
Program Management Support	\$450,000
Basic Research (7 projects 3 years each)	\$7,350,000
Product Transition Research (8 projects 3 years each)	\$36,000,000
Innovative Capabilities Research (2 projects 3 years each)	\$18,000,000
Total Estimated Funding Phase II	\$62,300,000

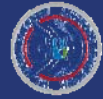
Activity	Estimated Funding Requirement
Support for Development of Statement of Needs, Statement of Works, and Contract Vehicles	\$400,00
Program Management Support	\$450,000
Basic Research (9 projects 3 years each)	\$9,450,000
Product Transition Research (7 projects 3 years each)	\$31,500,000
Innovative Capabilities Research (2 projects 3 years each)	\$18,000,000
Total Estimated Funding Phase I	\$59,800,000

Activity	Estimated Funding Requirement
Support for Development of Statement of Needs, Statement of Works, and Contract Vehicles	\$350,000
Program Management Support	\$450,000
Basic Research (5 projects 3 years each)	\$5,250,000
Product Transition Research (11 projects 3 years each)	\$49,500,000
Innovative Capabilities Research (2 projects 3 years each)	\$18,000,000
Total Estimated Funding Phase III	\$73,550,000



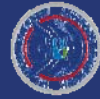
Expected DHS Impact

Capability	Expected Impact
<p>1. Multi-Modal Biometric People Screening</p>	<p>Universality: Significantly enhanced ability to include all members of the populace subject to screening activities at POEs.</p> <p>Uniqueness: Greatly expanded capacity of the US-VISIT system.</p> <p>Permanence: Incorporation of iris will likely increase the ability for US-VISIT to correctly identify individuals over extended periods of time.</p> <p>Measurability: By relying on multiple modalities, the biometric screening of US-VISIT will have an enhanced ability to tolerate noise from a single modality.</p> <p>Performance Accuracy: Increased theoretical and observed biometric performance accuracy of its people screening activities.</p> <p>Performance Efficiency: More biometric processing activities to take place in the same amount of time. Acceptability: Inclusion of non-contact biometrics can also contribute to the acceptability in some individuals and under some hygienic circumstances.</p> <p>Circumvention: With a spoofing / masquerade perspective, a malicious subject will have to simultaneously present multiple forged biometrics in verification activities.</p>
<p>2. Mobile Biometric Screening Capabilities</p>	<p>Universality: A greater portion of the incoming / exiting population of travelers would be subject to multi-biometric screening as a result of the addition of the capability.</p> <p>Acceptability: Expanded scope of biometric screening capability should result in a greater public approval.</p> <p>Uniqueness, Permanence, Measurability, Performance Accuracy, Performance Efficiency, and Circumvention: See 1.</p>



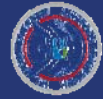
Expected DHS Impact

Capability	Expected Impact
<p>3. POE Worker Multi-Biometric Authentication</p>	<p>Acceptability: Improved public perception of POE security by imposing greater security controls on individuals working at POEs.</p> <p>Circumvention: Holistic approach to POE security combats malicious individuals and groups will exploit weak links in a system</p> <p>Universality, Uniqueness, Permanence, Measurability, Performance Accuracy, and Performance Efficiency: See 1.</p>
<p>4. Trusted Traveler Standoff Multi-Biometric Verification</p>	<p>Acceptability: Contact-less biometric systems for trusted traveler screening alleviates the hygiene related concerns of travelers providing biometric samples. Improved overall experience of presenting a biometrics</p> <p>Measurability: The ability to acquire and subsequently capture multiple biometrics at a standoff distance is currently, outside the scope of widely deployed biometric systems</p> <p>Universality, Uniqueness, Permanence, Performance Accuracy, Performance Efficiency, Acceptability, and Circumvention See 1</p>
<p>5. Unmanned Walking Subject Multi-Biometric People Screening</p>	<p>Performance Efficiency: POEs will be able to significantly increase traveler throughput.</p> <p>Measurability: Enable the ability to measure biometrics of walking travelers at distance.</p> <p>Acceptability: Increased efficiency in biometric screening will also result in increased levels of acceptance of travelers at POEs.</p> <p>Universality, Uniqueness, Permanence, Performance Accuracy, and Circumvention See 1.</p>



Expected DHS Impact

Capability	Expected Impact
<p>6. Vehicular Multi-Biometric People Screening</p>	<p>Measurability: Ability to tolerate increased speed of the travelers but also an increased tolerance for obstacles occluding biometrics such as windshields and other physical structures.</p> <p>Performance Efficiency: Greatly increased in vehicle throughput at POEs. Increased screening efficiency and facilitation of trade.</p> <p>Acceptability: Increased level of acceptance with the general public due to decreased wait times.</p> <p>Circumvention: More difficult to smuggle individuals across land POEs in the stated scope of vehicles.</p> <p>Universality, Uniqueness, Permanence, and Performance Accuracy: See 1.</p>
<p>7. Multi-Biometric Dynamic Decisional Fusion Systems</p>	<p>Measurability: Dynamic processing based on humidity, sunlight, etc. can vary with location, season, and other drivers.</p> <p>Performance Accuracy: Greater ability to correctly discriminate between matching and non-matching biometric samples based on improved tolerance of various factors.</p> <p>Performance Efficiency: Ability to modify biometric processing based on demand.</p> <p>Circumvention: A dynamic decisional fusion scheme would allow the biometric system to temporarily tighten the biometric screening process to account for this localized risk.</p> <p>Universality, Acceptability, and Uniqueness: See 1.</p>
<p>8. Multi-Biometric Vulnerability Countermeasures</p>	<p>Acceptability: By showing a biometric system's tolerance to known attacks, the public's acceptance of the technology will increase.</p> <p>Circumvention: Known methods of circumvention become too costly for malicious individuals to launch attacks on the biometric system in question.</p>



Expected DHS Impact

- ▶ **Research will provide publicly available information generally falling under the concepts of**
 - Information fusion
 - Collection and acquisition systems
 - Next Generation Algorithms and Modalities
 - Data quality and enhancement
 - Data sharing and architecture
 - Performance tracking
- ▶ **Research efforts will have uncovered solutions to overcoming the impediments associated with implementing multi-biometric solutions**
 - Acquiring the data
 - Creating new databases
 - Using existing databases
 - Meeting efficiency / throughput requirements



Wrap Up

- **The Following are Available on the Distributed CD's**
 - Extended version of this presentation
 - Problem Definition Deliverable
 - Identification of Stakeholders and Efforts Deliverable
 - Research Execution Plan Deliverable
- **The Entire Set of References Can be Found within the Deliverables**

Questions?



BioFuse: A Matlab™ Platform for Designing and Testing Biometric Fusion Algorithms

[Final Report]

Arun Ross¹ and Anil K. Jain²

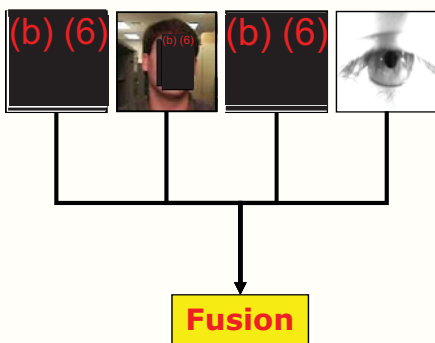
¹ West Virginia University

² Michigan State University

CITeR Meeting Spring 2010

Problem Statement

- Build a **software platform** that would provide its user with the ability
 - to **experiment** with a large number of fusion methods
 - to **evaluate** the relative performance of these methods on multiple datasets
- Generate a **collaborative environment** for updating software with newly developed fusion schemes





**BioFuse:
Biometrics Fusion
Platform**

Motivation

- **Recent research in multibiometrics has resulted in the development of several novel fusion algorithms at the data, feature, score, rank, and decision levels**
- **The development of a fusion toolkit incorporating these algorithms will benefit the following communities:**
 - **Researchers:** Comparing the performance of newly developed fusion methods against existing ones
 - **Practitioners:** Evaluating the performance of multiple fusion schemes and making appropriate recommendations for use in large-scale multibiometric systems
 - **Educators:** Access to an environment for training engineers and students

3

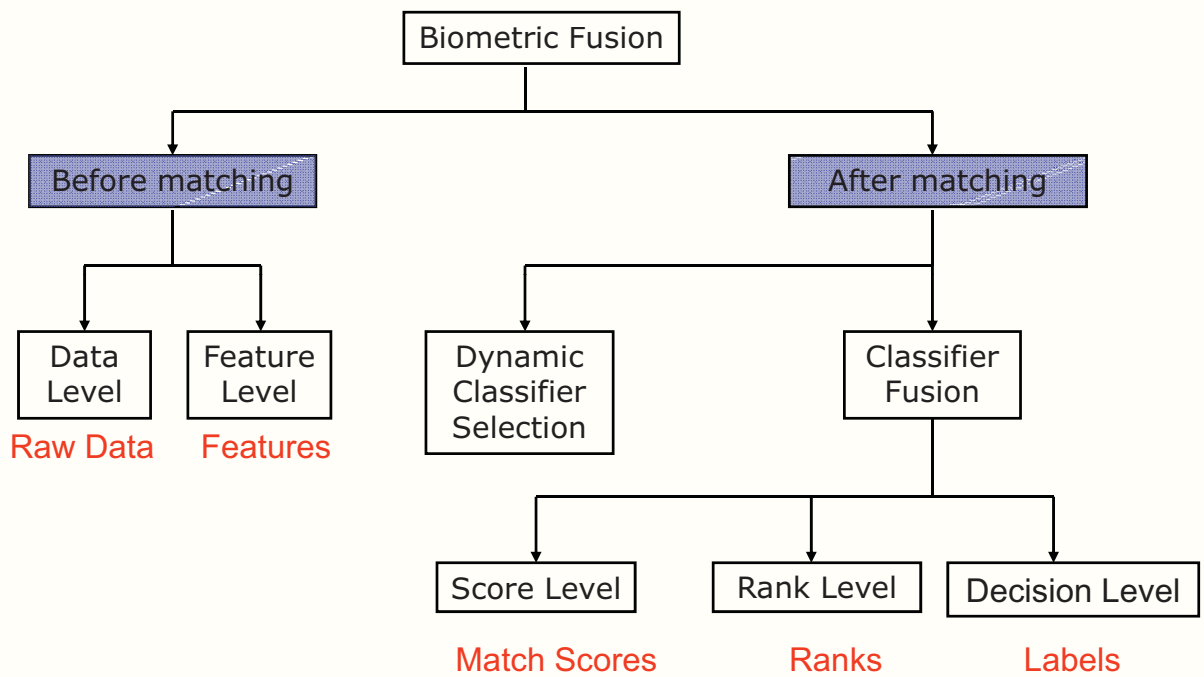
Milestones and Deliverables

Milestone	Description and Deliverable	Timeframe
1. Designing GUI	Design GUI and implement common fusion techniques at the feature, score, rank, and decision levels	0 - 5 mos
2. Modifying interface 	Based on feedback from CITeR affiliates, modify interface for ease of use	7 - 12 mos
3. Performance Evaluation	Performance evaluation on existing datasets	7 - 12 mos
4. Designing web interface 	Develop a wiki-type website for collaborative editing	7 - 12 mos

4

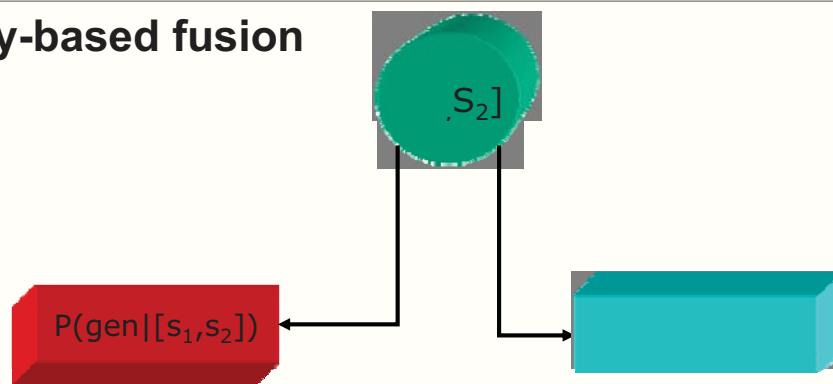
Levels of Fusion Considered

- Fusion after matching:



Score-level Fusion

- **Density-based fusion**



The likelihood ratio is used to define the fusion rule:
e.g.,

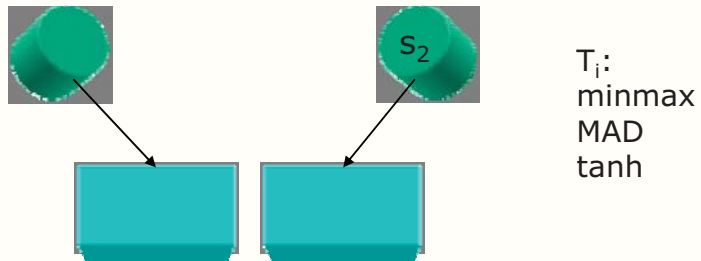
If $\frac{P(\text{gen} | [s_1, s_2])}{P(\text{imp} | [s_1, s_2])} > \lambda$, then genuine,
else impostor.

$$P(\text{gen} | s_i) = \frac{P(s_i | \text{gen})P(\text{gen})}{P(s_i)}$$

$$\frac{P(\text{gen} | s_1).P(\text{gen} | s_2)}{P(\text{imp} | s_1).P(\text{imp} | s_2)} > \lambda$$

Score-level Fusion

- Transformation-based fusion



- The transformed scores can be combined using several different rules
 - $\min[T_1(s_1), T_2(s_2)]$
 - $\max[T_1(s_1), T_2(s_2)]$
 - $\text{sum}[T_1(s_1), T_2(s_2)]$
 - $\text{prod}[T_1(s_1), T_2(s_2)]$

Implemented: Score-level Fusion

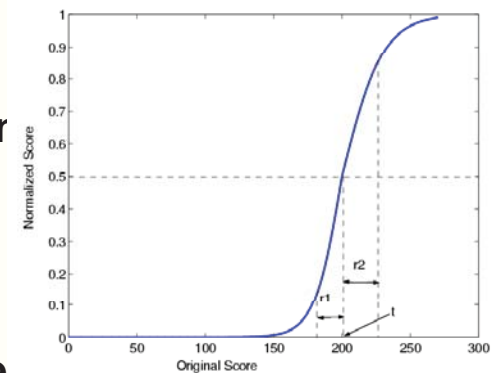
⑩ ✓ Density-based fusion

- Likelihood ratio
- Sum and Product rules
- Different density estimation schemes
 - Parametric
 - Non-parametric

⑩ ✓ Transformation-based fusion

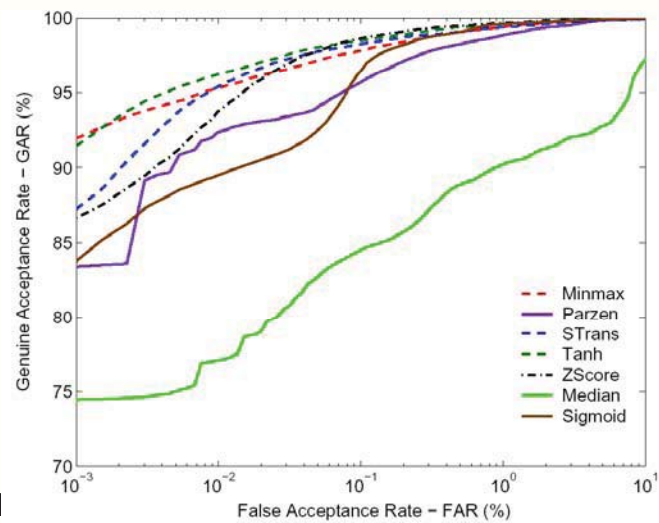
- Sum-rule, Max-rule, Min-rule
- Score normalization schemes

⑩ ✓ Score normalization schemes



Implemented: Score Normalization Schemes

- ⑩ ✓ Min-max
- ⑩ ✓ Decimal Scaling
- ⑩ ✓ Z-score
- ⑩ ✓ Median and MAD
- ⑩ ✓ Double sigmoid
- ⑩ ✓ Tanh estimators
- ⑩ ✓ Biweight estimator



Implemented: Decision and Rank Level

10 ✓ Decision-level:

- AND-rule
- OR-rule
- Majority-rule

$$R_j = \min_{i=1}^c r_{i,j} \quad R_j = \sum_{i=1}^c r_{i,j}$$

10 ✓ Rank-level:

- Borda Count
- Modified Borda Count
- Highest Rank
- Logistic Regression

$$R_j = \min_{i=1}^c r_{i,j} + e_j,$$

$$e_j = \frac{\sum_{i=1}^c r_{i,j}}{K}$$

$$\max_{i=0}^c r_{i,j} \leftarrow 0$$

10

Imputation of Missing Data

- Each row is a score vector

Fingerprint	Face	Iris
75	64	90
13	10	1
56	89	■
9	7	14
66	■	78
8	9	3
15	78	12
■	56	■
89	98	99

11

Implemented: Imputation of Missing Data

- ⑩ ✓ **Method 1: Maximum Likelihood Estimation (MLE) via EM Algorithm**

- ⑩ ✓ **Method 2: Multiple Imputation (MI) via Data Augmentation**

- ⑩ ✓ **Method 3: Imputation through Gaussian Mixture Models (GMM)**

12

Input and Output Data

⑩ ✓ Input Data

- Feature sets (only vectors)
- Match scores: Genuine and Impostor
- Quality

⑩ ✓ Output Data

- Statistics of scores (mean, variance, histogram)
- ROC curves

Datasets Available



- **WVU Multimodal Dataset**
- **MSU Match Score Dataset**
- **BioSecure Dataset**
- **DoD Multimodal Dataset**

Status

- **The GUI has not been finalized as yet**
 - Target date: 7/31/2010
- **Wiki site is still under development**
 - Target date: 7/31/2010
- **Contacted other researchers for developing modules**
 - Visiting scholars

Member Benefits

- **A software tool for testing a wide gamut of fusion rules on biometric data obtained from operational environments**
- **The tool can potentially be used to train engineers to apply fusion rules in a systematic way**
- **The capability to incorporate fusion modules developed by other researchers**





Improving Quality Enhanced Biometric Fusions Schemes

Progress Report



Bojan Cukic Afzel Noore
Nick Bartlow Mayank Vatsa
Nathan Kalka Richa Singh

West Virginia University
CITeR Fall 2009



The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research

www.citer.wvu.edu

Last Time - Recap

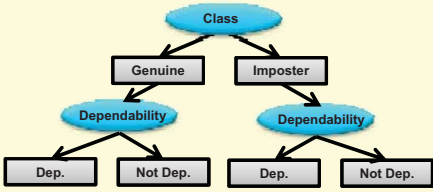
Computing Decision Dependability *Kryszczuk et. al

1. Estimate Error Conditional Distributions	$p(E CA) = p(E D \wedge GT), \quad p(E CR) = p(E \neg D \wedge \neg GT)$ $p(E FA) = p(E D \wedge \neg GT), \quad p(E FR) = p(E \neg D \wedge GT)$
2. Compute Credence Estimate For Correct Accept / Correct Reject Cases	$P(CA E_{TS}) = \frac{p(E_{TS} CA) \cdot P(CA)}{p(E_{TS} CA) \cdot P(CA) + p(E_{TS} FA) \cdot P(FA)}$ $P(CR E_{TS}) = \frac{p(E_{TS} CR) \cdot P(CR)}{p(E_{TS} CR) \cdot P(CR) + p(E_{TS} FR) \cdot P(FR)}$
3. Apply threshold to credence estimate to arrive at binary dependability decision	<i>If</i> $(D_{TS} = 1 \text{ AND } P(CA E_{TS}) \geq \theta) \Rightarrow \text{Dependable}$ <i>If</i> $(D_{TS} = 1 \text{ AND } P(CA E_{TS}) < \theta) \Rightarrow \text{Not Dependable}$ <i>If</i> $(D_{TS} = 0 \text{ AND } P(CR E_{TS}) \geq \theta) \Rightarrow \text{Dependable}$ <i>If</i> $(D_{TS} = 0 \text{ AND } P(CR E_{TS}) < \theta) \Rightarrow \text{Not Dependable}$

Labels


- E – Evidence
- CA – Correct Accept
- FR – False Reject
- CR – Correct Reject
- FA – False Accept
- GT – Ground Truth

Two Stage Dichotomization




```

graph TD
    Class([Class]) --> Genuine[Genuine]
    Class --> Imposter[Imposter]
    Genuine --> Dep1([Dependability])
    Imposter --> Dep2([Dependability])
    Dep1 --> Dep1a[Dep.]
    Dep1 --> NotDep1[Not Dep.]
    Dep2 --> Dep2a[Dep.]
    Dep2 --> NotDep2[Not Dep.]
                    
```




The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research

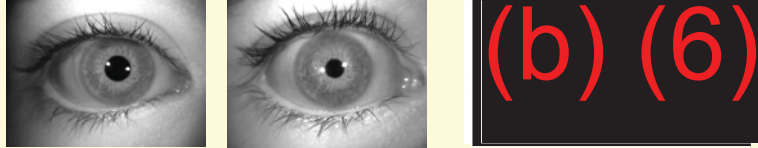
www.citer.wvu.edu



WVU Data Fingerprint + Iris

Examples of High Dependability Genuine Instances (GMM-LR)

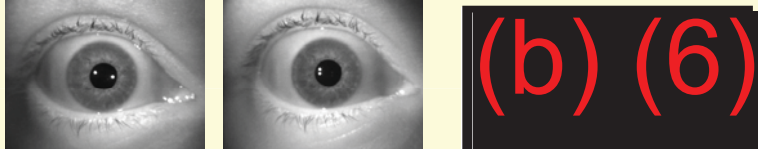




Correctly Predicted Match

Iris Quality (WVU)	Iris Quality (WVU)	Fing Quality (NFIQ)	Fing Quality (NFIQ)
0.82	0.90	1	1
Iris Score (WVU)		Fing Score (Identix BioEngine 5.0)	
0.28		358.45	


Row 72



Correctly Predicted Match


Iris Quality (WVU)	Iris Quality (WVU)	Fing Quality (NFIQ)	Fing Quality (NFIQ)
0.74	0.67	2	2
Iris Score (WVU)		Fing Score (Identix BioEngine 5.0)	
0.15		88.39	


Row 6




The Center for Identification Technology Research

An NSF I/UCR Center advancing integrative biometrics research






Last Time - Recap




Rectifying Decisions Likely to be Erroneous *Kryszczuk et. al

<p>1. Establish threshold for credence to arrive at dependability measures indicating how the chance of (in)accurate prediction</p>	$\begin{aligned} \text{If } (D_{TS} = 1 \text{ AND } P(CA E_{TS}) \geq \theta) &\Rightarrow \text{Dependable} \\ \text{If } (D_{TS} = 1 \text{ AND } P(CA E_{TS}) < \theta) &\Rightarrow \text{Not Dependable} \\ \text{If } (D_{TS} = 0 \text{ AND } P(CR E_{TS}) \geq \theta) &\Rightarrow \text{Dependable} \\ \text{If } (D_{TS} = 0 \text{ AND } P(CR E_{TS}) < \theta) &\Rightarrow \text{Not Dependable} \end{aligned}$
<p>2. Flip non-dependable decisions (falling below θ threshold likely to be inaccurate)</p>	$\forall D_{TS} : R(E_{TS}) < \theta \Rightarrow D'_{TS} \rightarrow \neg D_{TS}$
<p>3. Complement dependability of flipped decisions</p>	$\forall D_{TS} : R(E_{TS}) < \theta \Rightarrow R'(E_{TS}) \rightarrow 1 - R(E_{TS})$







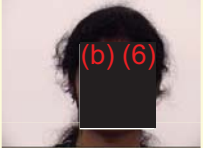

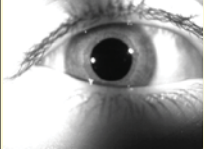
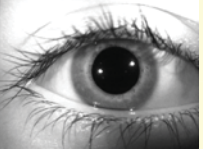
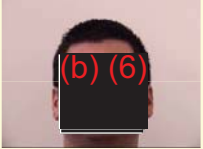
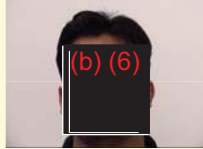
The Center for Identification Technology Research

An NSF I/UCR Center advancing integrative biometrics research



WVU Data Face + Iris
Examples of Decision Rectification (GMM-LR)

				Incorrect Before Gen - 0.10 Correct After Imp - 0.90
Iris Quality (WVU) 0.14	Iris Quality (WVU) 0.221	Face Quality (Facelt) 4.08	Face Quality (Facelt) 5.76	
Iris Score (WVU) 0.45		Face Score (Identix Facelt G6) 5.64		
				Row 20509 Row 28075 Incorrect Before Gen - 0.21 Correct After Imp - 0.79
Iris Quality (WVU) 0.43	Iris Quality (WVU) 0.62	Face Quality (Facelt) 4.78	Face Quality (Facelt) 4.98	
Iris Score (WVU) 0.38		Face Score (Identix Facelt G6) 3.01		

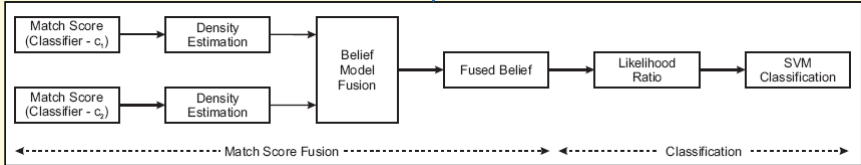
CITeR The Center for Identification Technology Research 5
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu

Problems


- Is it possible to improve the performance of state of the art multimodal score level classifiers?
- Can we improve classification performance when presented with conflicting evidence from independent modalities?
- Is it possible to make these improvements without significantly increasing complexity?

Proposed Solutions


- Explicitly incorporate feedback and potential for rectification in classification decisions
- Combine techniques from statistical, learning, and belief based fusion approaches to arrive at a hybrid match score fusion algorithm
 - Capable of dealing with conflicting evidence



CITeR The Center for Identification Technology Research 6
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu




Hybrid Fusion with Proportional Conflict Redistribution



Step 2: Utilize estimated density information as basic belief assignment for the application of PCR


- Estimate joint density for fingerprint match and quality scores. Similarly, the joint density for iris match and quality scores are estimated.
 - Densities estimated through Gaussian Mixture Models (GMMs)

$$m_{PCR5} = m_{12}(X) + \sum_{\substack{Y \in G \setminus \{X\} \\ c(X \cap Y) = \phi}} \left[\frac{m_1(X)^2 m_2(Y)}{m_1(X) + m_2(Y)} + \frac{m_2(X)^2 m_1(Y)}{m_2(X) + m_1(Y)} \right]$$




The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research

7




Hybrid Fusion with Proportional Conflict Redistribution



Step 3: Take the likelihood ratio after redistributing the conflicting mass


$$\Lambda = \frac{m_{PCR5}(A)}{m_{PCR5}(B)}$$

- PCR5 calculates the conflicting mass between all propositions and redistributes it proportionally to those propositions involved in the conflict.
- In the case of our data, mass will only be distributed to genuine and imposter propositions (other data may involve many more propositions)




The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research

8



Proportional Conflict Redistribution: PCR Rule 5




- PCR5
 - Apply the conjunctive rule

$$m_{\cap} = \sum_{\substack{X_1, \dots, X_s \in D^{\Theta} \\ X_1 \cap \dots \cap X_s = X}} \prod_{i=1}^s m_i(X_i)$$

- Calculate all partial conflicting mass


$$m_{12}(A \cap B) = m_1(A)m_2(B) + m_1(B)m_2(A)$$

- Redistribute all conflicting mass to A and B proportionally with respect to:
 - masses $m_1(A)$ and $m_2(B)$ respectively,
 - masses $m_2(A)$ and $m_1(B)$ respectively.




The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research

9



Proportional Conflict Redistribution: PCR Rule 5



- Add the redistributed conflicting masses to all sets involved in the conflict

Example

	A	B
m_1	0.7	0.3
m_2	0.5	0.5

$$\Theta = \{A(\text{genuine}), B(\text{imposter})\}$$


- Conjunctive rule yields: $m_{12}(A) = m_1(A) * m_2(A) = 0.35$ $m_{12}(B) = m_1(B) * m_2(B) = 0.15$
- Conflicting Mass: $m_{12}(A \cap B) = m_1(A) * m_2(B) + m_1(B) * m_2(A) = 0.35 + 0.15 = 0.5$
- Calculate partial conflicting mass to be proportionally redistributed to A and B :

$$x_1 = 0.7 * 0.292 = 0.204 \quad y_1 = 0.5 * 0.292 = 0.146$$

$$x_2 = 0.5 * 0.1875 = 0.09375 \quad y_2 = 0.3 * 0.1875 = 0.05625$$


$$m_{\text{PCR5}}(A) = 0.35 + 0.204 + 0.09375 = 0.64775$$

$$m_{\text{PCR5}}(B) = 0.15 + 0.146 + 0.05625 = 0.35225$$




The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research

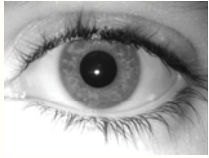
10



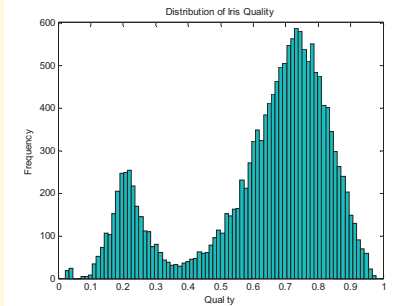
Description of Data: Stats and Quality

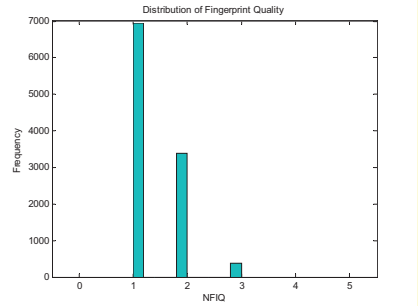



- DoD Operational Data
 - 7,818 genuine scores
 - 1000000 imposter scores
 - Left Iris
 - Left thumb



(b) (6)









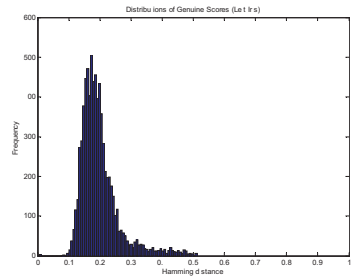
The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research

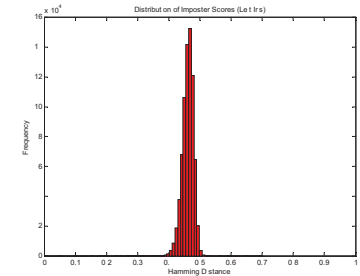
11
www.citer.wvu.edu

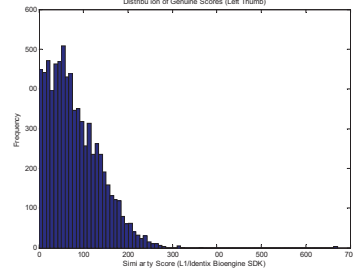


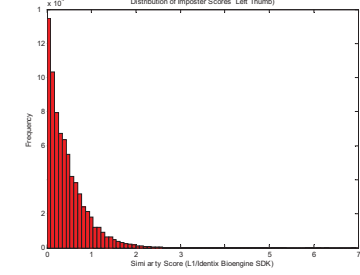
Description of Data: Match Score Histograms








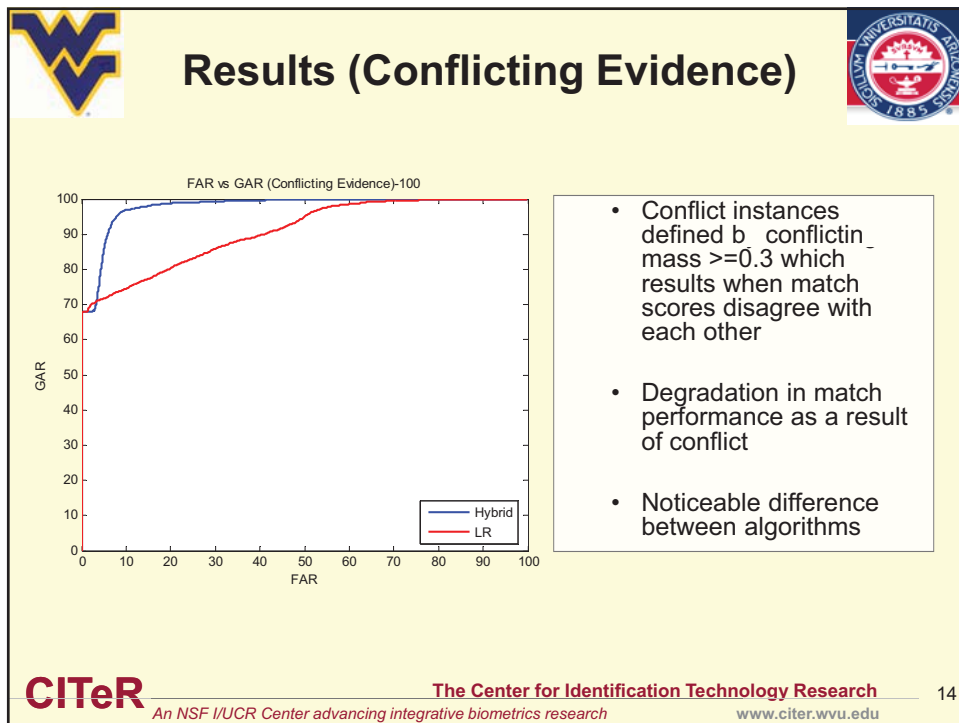
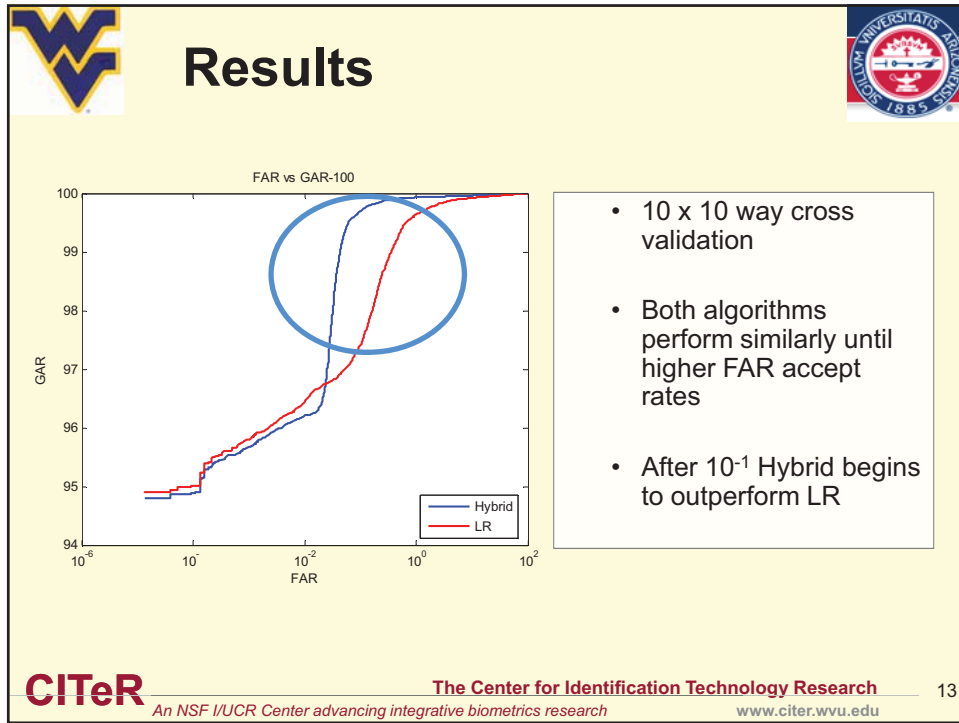







The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research

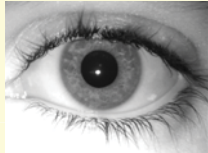

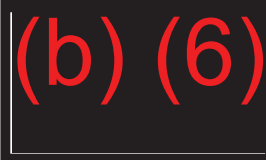

12
www.citer.wvu.edu



3.32, 13.86



Example Instances Considering Conflicting Mass

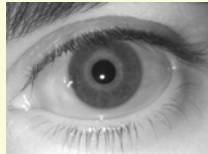

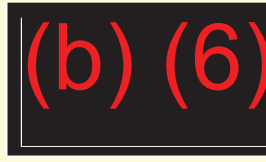





Iris Quality (WVU)	Iris Quality (WVU)	Fing Quality (NFIQ)	Fing Quality (NFIQ)
0.73	0.72	1	2
Iris Score (WVU)		Fing Score (Identix BioEngine 5.0)	
0.16		102.99	

Genuine Instance

Conflicting Mass
u.u1

LR	✓
260.81	✓
Hybrid	✓
268.44	






Iris Quality (WVU)	Iris Quality (WVU)	Fing Quality (NFIQ)	Fing Quality (NFIQ)
0.65	0.61	1	1
Iris Score (WVU)		Fing Score (Identix BioEngine 5.0)	
0.12		0.04	


Genuine Instance

Conflicting Mass
0.99

LR	✗
-0.94	✓
Hybrid	✓
4.60	




The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research



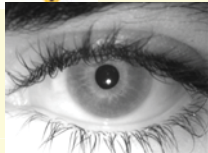
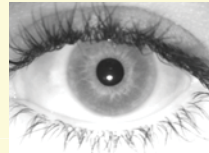
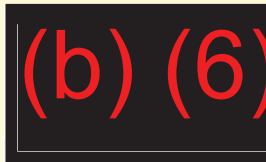

15

www.citer.wvu.edu

3.32, 13.86



Example Instances Considering Conflicting Mass


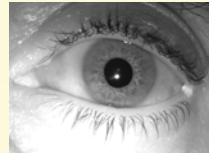
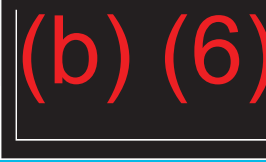





Iris Quality (WVU)	Iris Quality (WVU)	Fing Quality (NFIQ)	Fing Quality (NFIQ)
0.67	0.78	1	1
Iris Score (WVU)		Fing Score (Identix BioEngine 5.0)	
0.32		9.65	

Genuine Instance

Conflicting Mass
u.86

LR	✓
58.96	✓
Hybrid	✓
60.13	






Iris Quality (WVU)	Iris Quality (WVU)	Fing Quality (NFIQ)	Fing Quality (NFIQ)
0.80	0.89	1	2
Iris Score (WVU)		Fing Score (Identix BioEngine 5.0)	
0.15		0.37	


Genuine Instance

Conflicting Mass
0.99

LR	✗
-1.50	✓
Hybrid	✓
4.39	



The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research






16

www.citer.wvu.edu

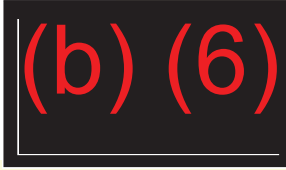
3.32, 13.86

Example Instances Considering Conflicting Mass



Iris Quality (WVU)	Iris Quality (WVU)	Fing Quality (NFIQ)	Fing Quality (NFIQ)
0.72	0.94	1	1
Iris Score (WVU)		Fing Score (Identix BioEngine 5.0)	
0.47		0.50	

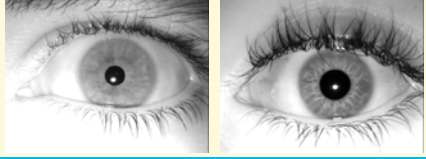


Imposter Instance

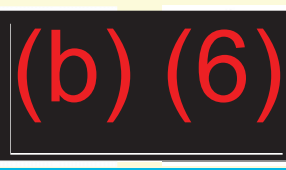
Conflicting Mass
0.03

LR ✔
-9.82

Hybrid ✔
-7.01



Iris Quality (WVU)	Iris Quality (WVU)	Fing Quality (NFIQ)	Fing Quality (NFIQ)
0.76	0.93	1	1
Iris Score (WVU)		Fing Score (Identix BioEngine 5.0)	
0.47		2.91	




Imposter Instance

Conflicting Mass
0.99

LR ✔
-1.22

Hybrid ✔
0.87





The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research


www.citer.wvu.edu

17

Summary / Observations

- Complement Quality based fusion algorithms with decision dependability.
 - Allows for decision discarding (secondary screening) and decision rectification.
 - Requires additional training for estimating error conditional densities but provides better results than using only class conditional densities.
 - Difficult in choosing a threshold for rectification in an operational setting where groundtruth isn't always available
 - Trade off between FAR/GAR
- Augmented Quality based Likelihood Ratio fusion with Proportional Conflict Redistribution (PCR) to create a "Hybrid Fusion" algorithm.
 - The application of PCR, specifically PCR rule 5 improves performance when dealing with conflicting evidence.
 - The addition of an SVM on the back-end of PCR did not provide improved performance.
 - Rectification handled implicitly.



The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research

www.citer.wvu.edu

18

Matching and Retrieving of Face Images Based on Facial Marks

Final Report

Anil K. Jain¹
Arun Ross²

¹ Michigan State University

² West Virginia University

CITeR Spring 2009©

Problem

- Current face recognition techniques do not explicitly extract any **distinguishing marks (wrinkles, scars, moles)**
- Some of the marks are **temporally invariant** and can be useful for face recognition and indexing
- **Goal: Automatically extract distinguishing marks and Use them in conjunction with commercial face recognition engines to enhance recognition accuracy**

2

Our Approach

- **Stage 1: Automatic technique to extract facial marks and encode them using a morphological scheme**
 - Generate statistics pertaining to these marks (e.g., frequency of occurrence, location, size, etc.)
- **Stage 2: Method to retrieve face images from a database based on a specified set of facial marks**
- **Stage 3: Fusion scheme to combine facial marks with a commercial face matcher to enhance recognition accuracy**

3

Milestones and Status

Milestone	Description and Deliverable	Timeframe
(1) Feature Extraction	Methods to extract facial scars, marks, moles and other irregularities, and to encode them using a morphological scheme; statistical analysis of the distribution of irregularities on a human face	6 months
(2) Retrieval System	Design of a face retrieval system that uses a specified set of facial marks to retrieve images from a digital face database	3 months
(3) Fusion Module	Design of a fusion system that combines facial marks with a commercial texture-based matcher to enhance recognition accuracy	3 months

4

Defined Mark Types

- **Freckle** - small spots from concentrated melanin
- **Mole** - growth on the skin (brown or black)
- **Scar** - marks left from cuts or wounds
- **Pockmark** - crater-shaped scar
- **Acne** - red lesions caused by pimples or zits
- **Whitening** - skin region appears white
- **Dark skin** - skin region appears dark
- **Abrasion** - wound (includes clots, temporary marks)
- **Wrinkle** - fold, ridge or crease in the skin
- **Other** - all other marks

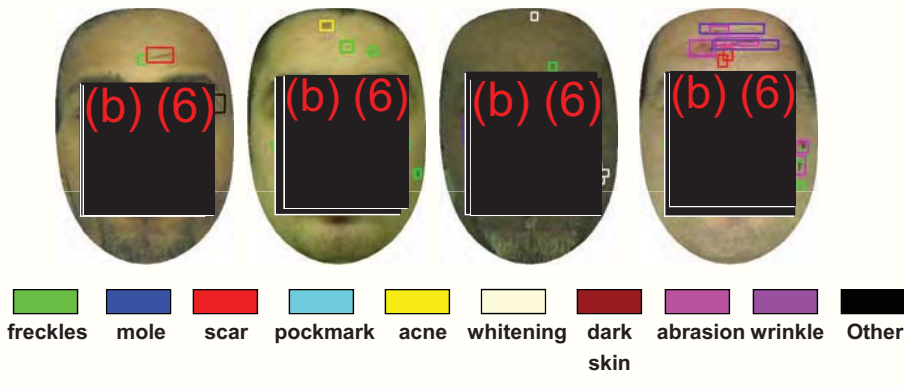
<http://en.wikipedia.org/wiki>

N. A. Spaun, Forensic Biometrics from Images and Video at the Federal Bureau of Investigation, BTAS, 2007

5

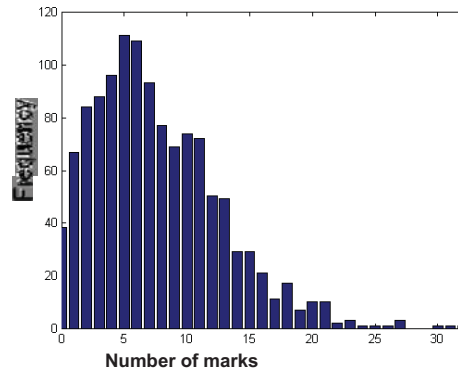
Ground Truth for Facial marks

- 1,225 images from 671 subjects in the Michigan Police mugshot database, 554 duplicate pairs
- Image resolution: 360 x 240 to 480 x 384 (low to medium quality)



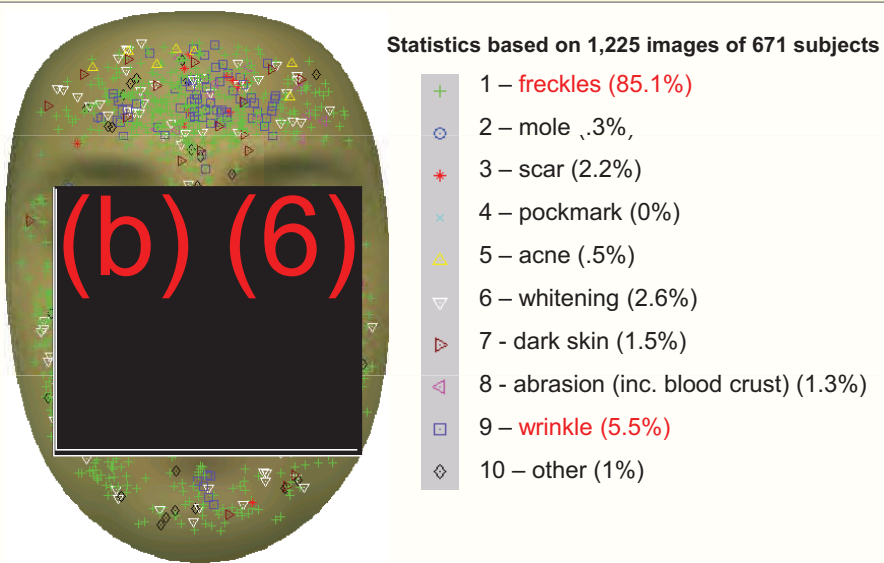
Statistics of Facial Marks

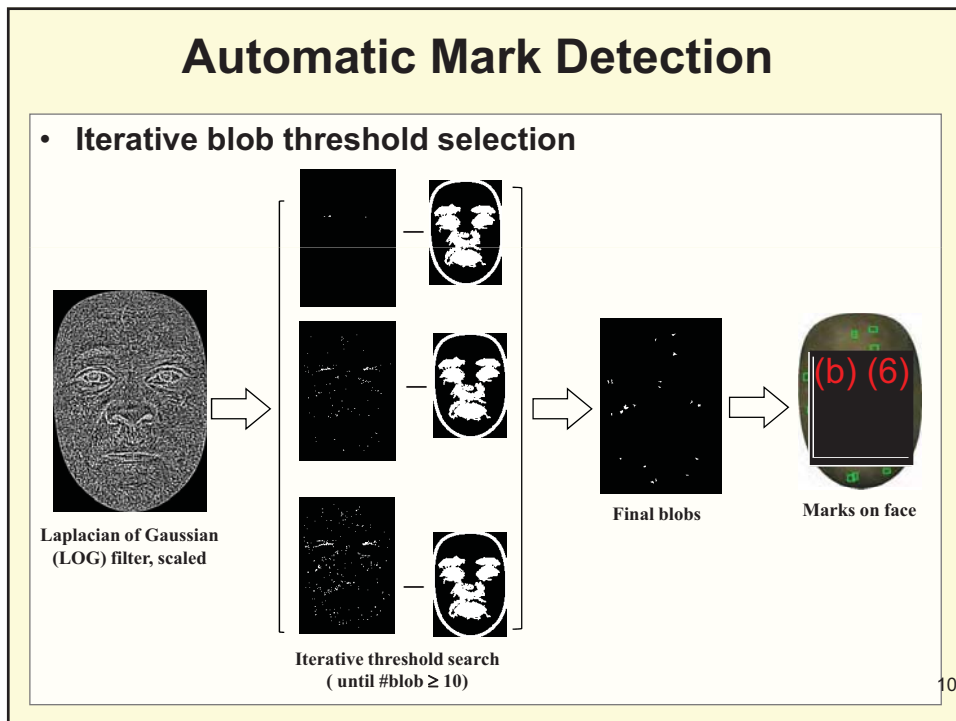
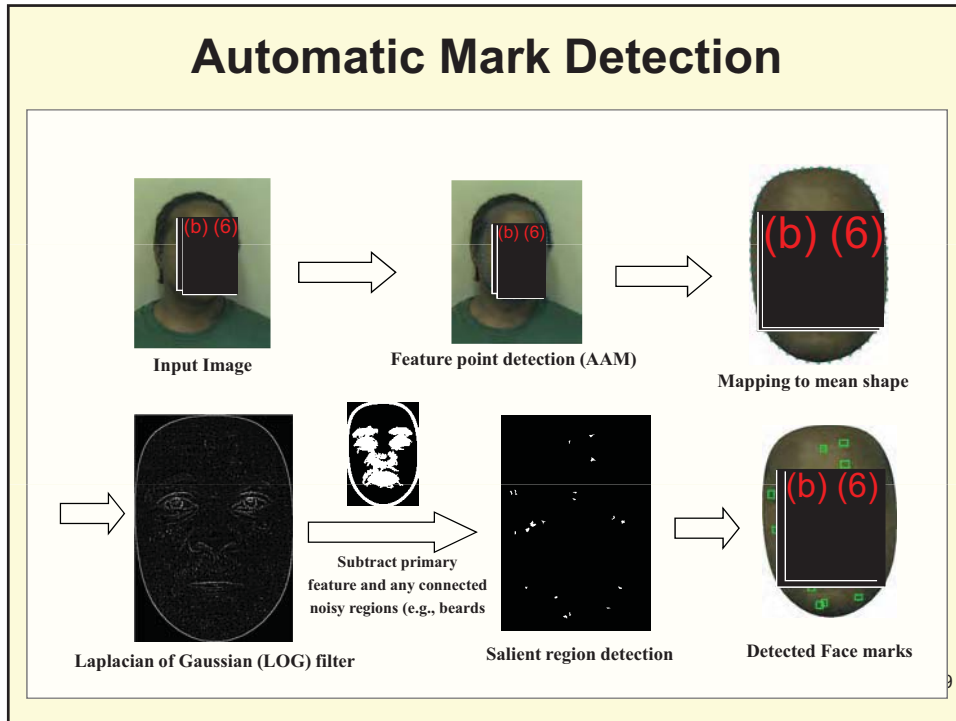
- **Average number of marks is ~7/subject**
 - All detected marks can be used for matching or retrieval
 - 97% of subjects in our database have at least one mark



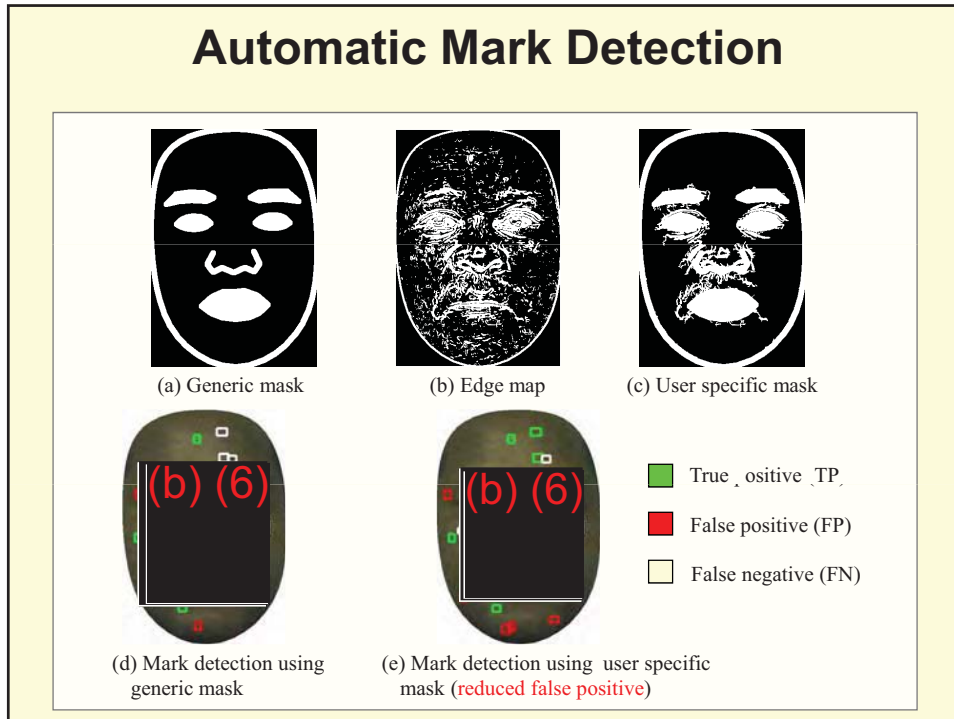
7

Statistics of Facial Marks





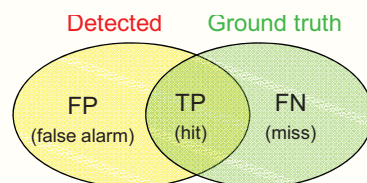
Automatic Mark Detection



Face Mark Detection & Matching

- **Mark detection accuracy**

- **Recall:** Percentage of true marks detected among all the ground truth marks
- **Precision:** Percentage of true marks out of all the detected marks



$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

- **Mark matching accuracy**

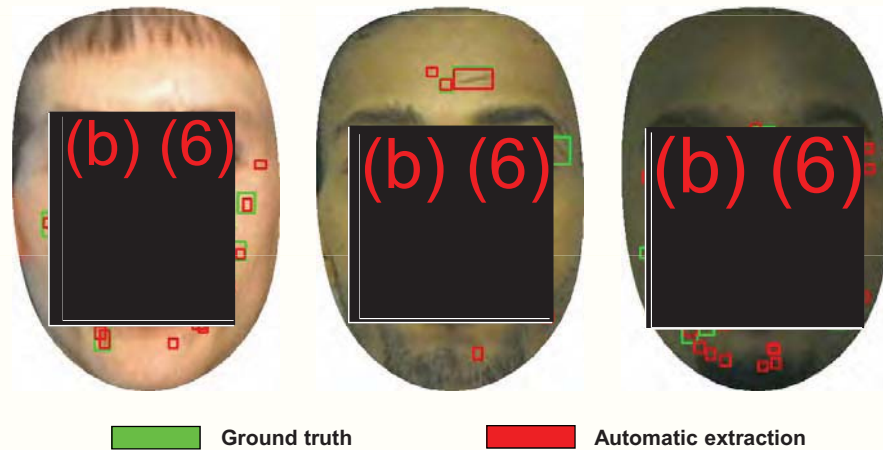
- Matching score between two face images is calculated based on the number of matching marks
- Two marks m_1 and m_2 are considered as match if

$$d(m_1, m_2) < \text{threshold}$$

$d(\dots)$ measures the distance in the mean shape space

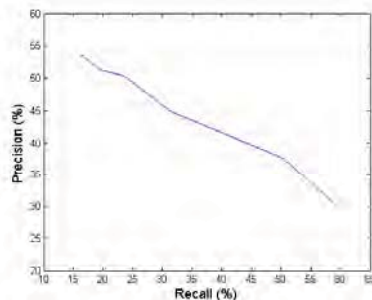
Limitations of Current Feature Extractor

- **Automatic extraction vs. ground truth**
 - Marks need to be larger than 2x2 pixels to be detected (~1x1 mm)



Fusion with Face Recognition Engine

- **Weighted score sum is used for the fusion**
 - 0.55 for the FaceVACS and 0.45 for the mark based matcher



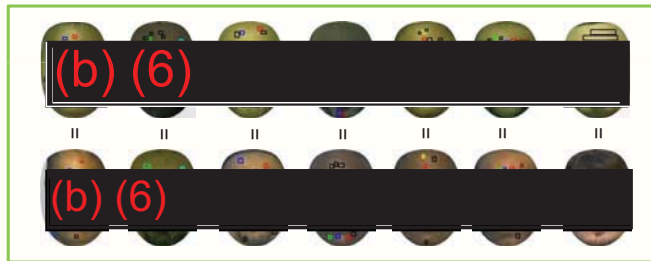
Matching with FaceVACS & Marks (%)					
FaceVACS only		Marks+ FaceVACS (Probe: manual; Gallery: automatic)		Marks + FaceVACS (Probe: manual; Gallery: manual)	
Rank 1	20	Rank 1	20	Rank 1	20
91.88	100	93.14	100	93.14	100%

14

Successful Matching Based on Marks



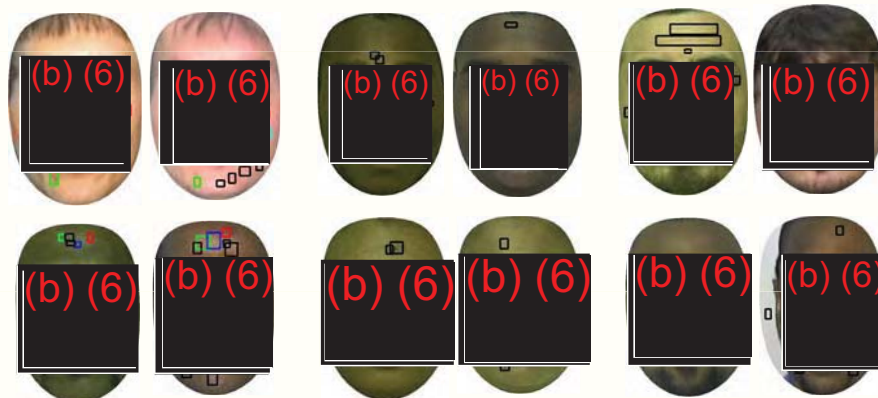
Failed at rank-1 by using
FaceVACS only



Succeeded at rank-1 by
FaceVACS + marks
(Probe: manual;
Gallery: automatic)

Unsuccessful Matchings Based on Marks

- Failed before and after fusion with mark based matcher



Insignificant contribution

No matching mark

No mark

Relevance to Members


- **Automatic facial mark extraction method is available for frontal face images**
- **The extracted marks can be used for image retrieval or to improve a commercial face matcher**
 - Performance improvement is shown using a leading commercial engine, FaceVACS
- **A. K. Jain and Unsang Park, *Facial Marks: Soft Biometric for Face Recognition*. ICIP, 2009 (submitted)**

17

Next Steps

- **Improve automatic mark extractor**
- **Extend it to non-frontal images and video frames**
- **Use marks to index and retrieve images from a large database**
- **We have submitted a Phase 2 proposal to continue this line of research**

18



 

Phase 0- Participation in MultiBiometric Grand Challenge

Final Report

Stephanie Schuckers, Natalia Schmid,
Besma Abidi, Uma Kandaswamy
Clarkson University, West Virginia University
and The University of Tennessee-Knoxville
CITeR Fall 2008©

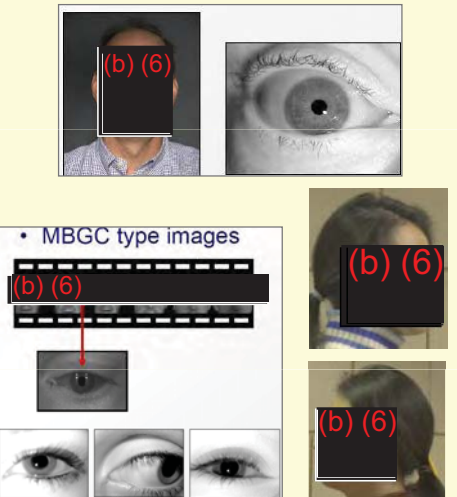
CITeR The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu

Problem



- Face and iris biometrics
 - Sensitive to the quality
 - Factors such as lighting, angle, pose, illumination, focus, resolution, and user's cooperation
- MultiBiometric Grand Challenge (MBGC) addresses some of the following questions:
 - Can face/iris video sequences improve the performance?
 - Can NIR face video provide sufficient information about iris for recognition based on iris?
 - What processing has to be done to use iris images from NIR face video for recognition of a large number of users?
 - What is the recognition performance of unconstrained face video with low to medium resolution?
 - Which way is better to fuse face and iris for biometric recognition, feature level? Match score level?
 - What's the effect of quality measure to the face and iris fusion?

• MBGC type images



Ref: Dr. P. Jonathon Phillips. "Multiple Biometric Grand Challenge Kick-Off Workshop" presentation, 18 April 2008

CITeR The Center for Identification Technology Research 2
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu



Our Work and Work Elsewhere

- Wang et al. combined face and iris biometrics for identity verification
- Liu et al. developed Gabor feature based classification using enhanced Fisher linear discriminant model for face recognition
- Schuckers et al. researched angle compensation in non-ideal iris recognition
- Zuo and Schmid researched robust segmentation of non-ideal iris.
- Kandaswamy et al. used error-encoded PDE-texton for face recognition
- Kong et al. researched multiscale fusion of visible and thermal IR images for illumination-invariant face recognition
- Yao et al. researched improving long range and high magnification face recognition



CITeR An NSF I/UCR Center advancing integrative biometrics research **The Center for Identification Technology Research** www.citer.wvu.edu 3



Our Approach

- Fast quality metrics used to select individual frames.
- Quality information to weigh the iris and face fusion.
- Robust iris segmentation developed for challenging iris sequences (NIR video).
- Advanced face recognition (texton, LBP) for encoding face information.
- Extended and refined Masek's algorithm for encoding iris images.
- Baseline: Masek's iris recognition and commercial Faceit for face recognition
- Fusion developed for multiple frames of face and iris at feature level or match score level.
- Advanced image/frame restoration techniques applied to HD video.



CITeR An NSF I/UCR Center advancing integrative biometrics research **The Center for Identification Technology Research** www.citer.wvu.edu 4



Research Plan

- Inventory of available algorithms for various components:
 - face localization, face tracking, face quality, face enhancement, face recognition, iris localization, iris segmentation, iris quality (including fast quality metric), iris recognition, and fusion strategies
- Determine which algorithms to be developed or modified to address challenges of MBGC data.
- Determine performance metrics for each and combined components and evaluate internal algorithms
- Create one or more super-algorithms. Evaluate their performance.
- Present results at 2008 and 2009 MBGC workshops



CITeR The Center for Identification Technology Research 5
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu




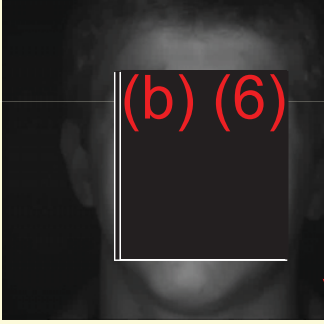

MultiBiometrics Grand Challenge

- **Portal Challenge (Face and Iris)** ← Our Focus
 - Still Face versus HD Video Face
 - Video Iris versus NIR Face Video
 - Still Iris versus NIR Face Video
 - Still Face / Video Iris versus HD Video Face / NIR Face Video
 - Still Face / Still Iris versus HD Video Face / NIR Face Video
- **Still Face Challenge**
- **Video Face Challenge**

CITeR The Center for Identification Technology Research 6
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu



 

Example Videos

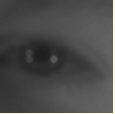
  

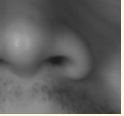
(b) (6) NIR-face
NIR-iris

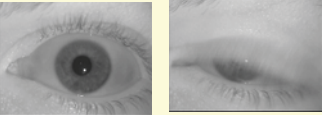
CITeR HD-face-video **The Center for Identification Technology Research** 7
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu

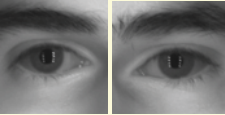
 

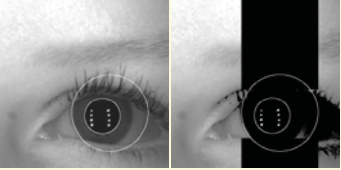
MBGC-IRIS: Challenges

 Low iris image quality caused by long ranges and poor illumination



 False alarms caused by misdetection or unintended iris presentation

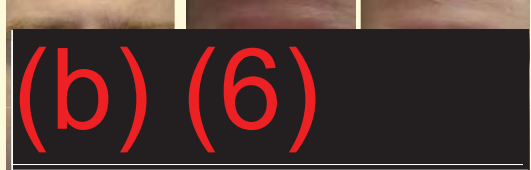
 Selection of best quality images (frames) from a video

 Fusion from multiple cropped images and multiple frames

 Standard MASEK algorithm:
Poor segmentation for NIR-face-video

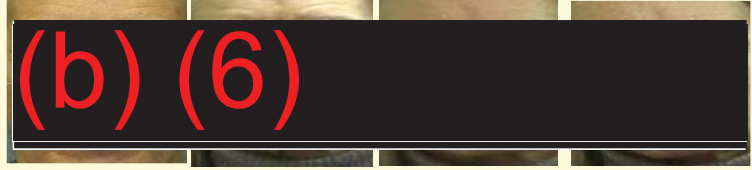
CITeR **The Center for Identification Technology Research** 8
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu

 **MBGC-FACE: Challenges** 



(b) (6)



Left: gallery high resolution still face image, middle: video frame with lower resolution and different illumination, right: video frame with lower resolution and out of focus



(b) (6)


From left to right: gallery high resolution still face image; video frame with lower resolution and non-visible eyes; video frame with lower resolution and motion blur; rotated face in video frame.

CITeR **The Center for Identification Technology Research** 9
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu

 **Inventory of algorithms** 

- **Face**
 - Localization: Facelt software
 - Frame selection: Quality algorithms (UT), integration into frame selection (Clarkson)
 - Encoding: Local binary pattern approach (Clarkson), Facelt software, PCA, LDA
- **Iris**
 - Localization: Specularity detector (WVU)
 - Frame selection: Defocus measure-based frame selection (WVU), application of quality maps
 - Segmentation: Robust segmentation for non-ideal iris (WVU), texture-based segmentation (Clarkson)
 - Automatic method for evaluating quality of iris segmentation (WVU)
 - Encoding: extended and refined Masek's implementation (WVU), Quality Biorthogonal Wavelets (Clarkson), Rosa.
- **Fusion**
 - Max, Sum, Multiple frames, Quality-based, Cohort-based.


CITeR **The Center for Identification Technology Research** 10
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu

 **Clarkson UNIVERSITY**
defy convention

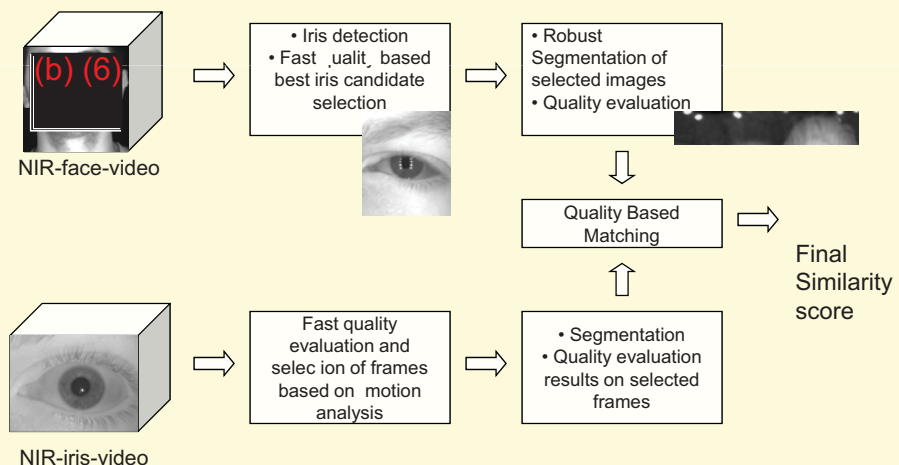
Metrics to Assess Algorithms

- Comparison with manual selection or manual check
- Improved matching performance with addition of component
- Speed/computational complexity
- Commercial software as a reference
- Automatic algorithm evaluating the precision of iris segmentation

CITeR The Center for Identification Technology Research 11
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu

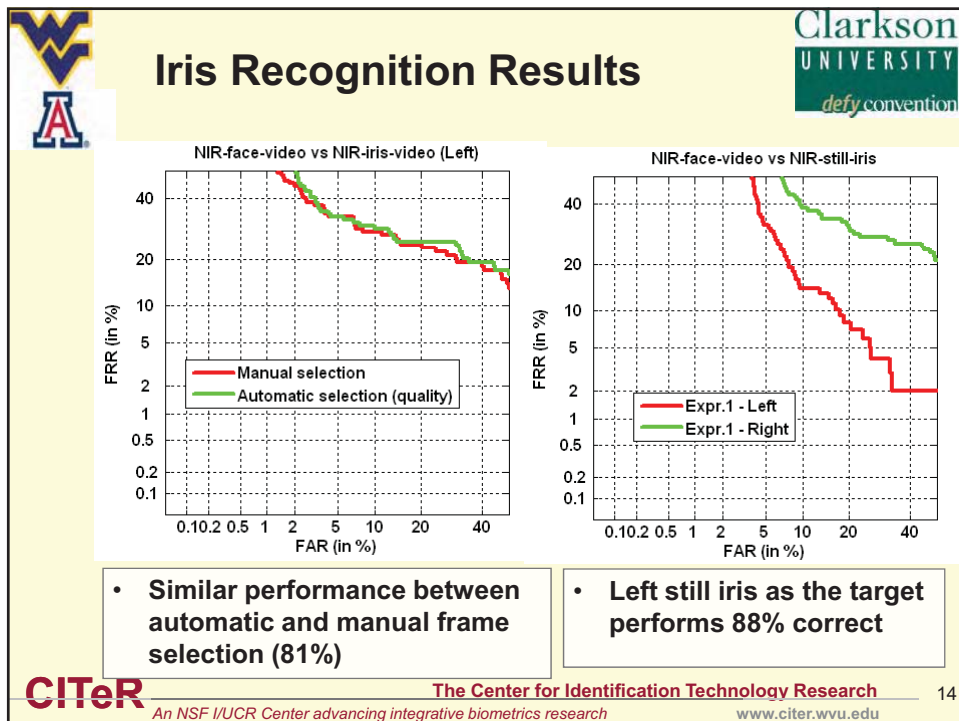
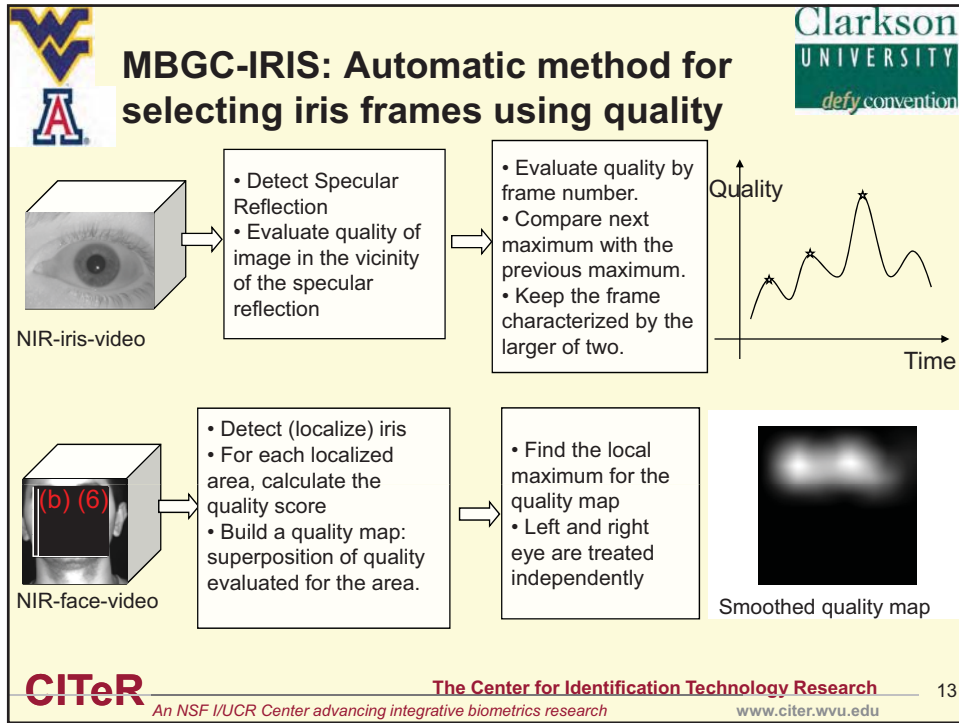
 **Clarkson UNIVERSITY**
defy convention

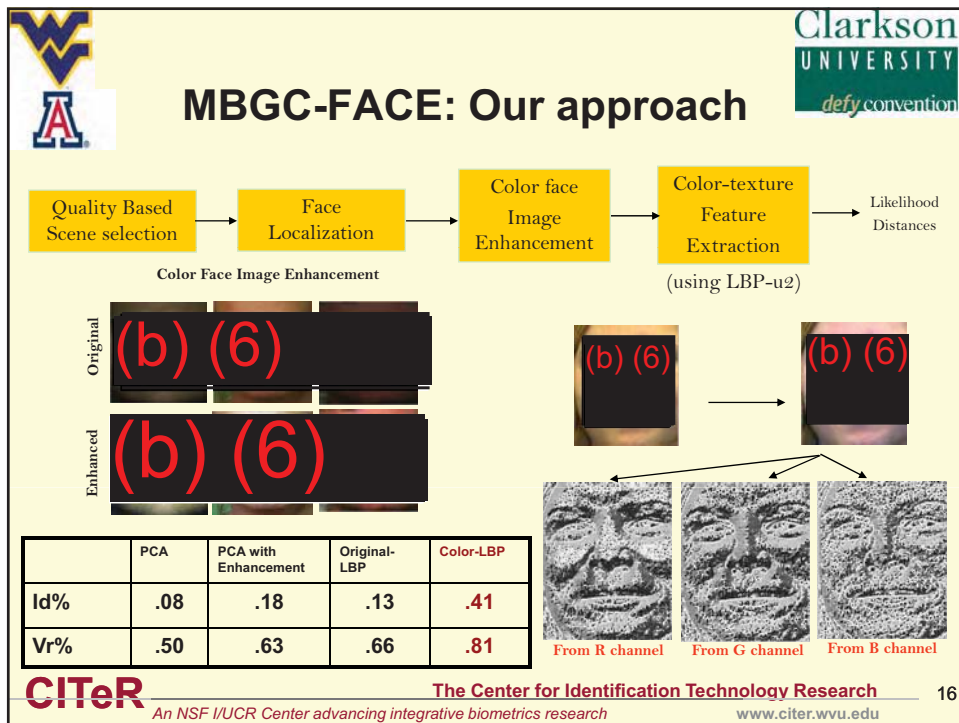
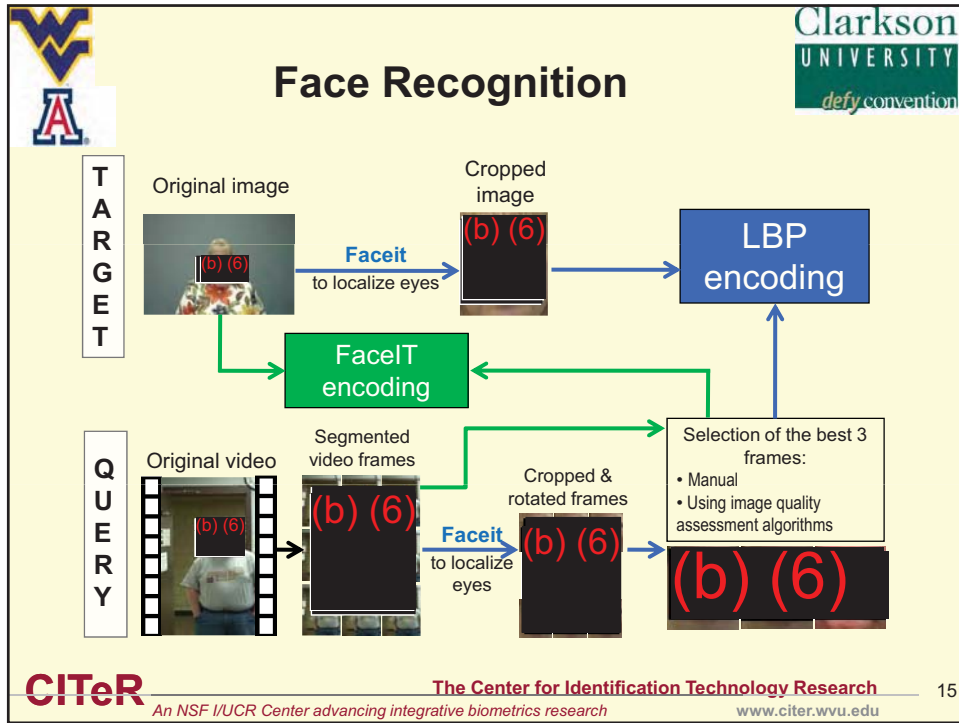
MBGC-IRIS: Our approach




```
graph LR; A["(b) (6)  
NIR-face-video"] --> B["• Iris detection  
• Fast, quality based  
best iris candidate  
selection"]; B --> C["• Robust  
Segmentation of  
selected images  
• Quality evaluation"]; C --> D["Quality Based  
Matching"]; D --> E["Final  
Similarity  
score"]; F["NIR-iris-video"] --> G["Fast quality  
evaluation and  
selection of frames  
based on motion  
analysis"]; G --> H["• Segmentation  
• Quality evaluation  
results on selected  
frames"]; H --> D;
```


CITeR The Center for Identification Technology Research 12
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu







Face Quality Metrics

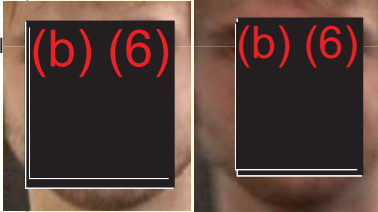


Seven different quality measures


- Laplacian
- Adaptive Laplacian - weighted according to the local activity in the image
- Tenengrad
- Adaptive Tenengrad –weighted according to the local activity in the image
- Entropy
- Histogram Difference – between the highest and lowest gray levels
- Gray Level Variance

Frames sorted by their best to worst quality values

Vote among all 7 quality measures to pick the highest quality frame




Sample best and worst frames from a video sequence based on their quality values




The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research

17
www.citer.wvu.edu



Automatic Selection of Frames Using Quality



Identification rates*

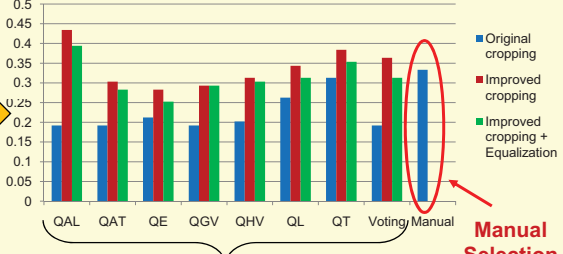
- Automatic frames selection comparable to manual
- Best quality shown by QAL

	LBP	PCA	FE-PCA
Id	0.13	0.08	0.18
Ver	0.66	0.50	0.60

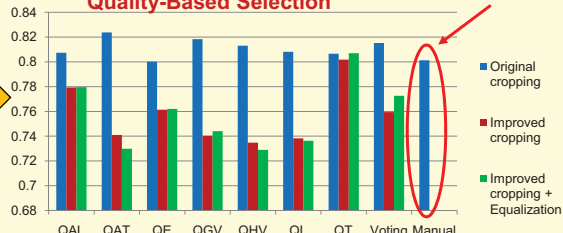
Verification rates* (normalized)


- Automatic frames selection comparable to manual
- Best quality shown by QT

Quality-Based Selection



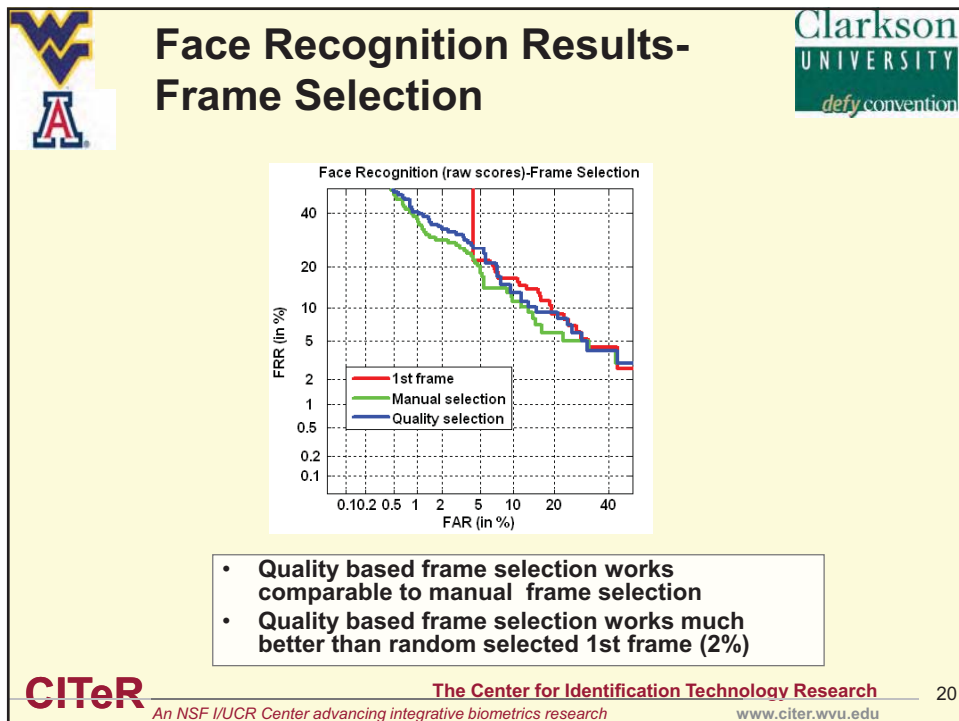
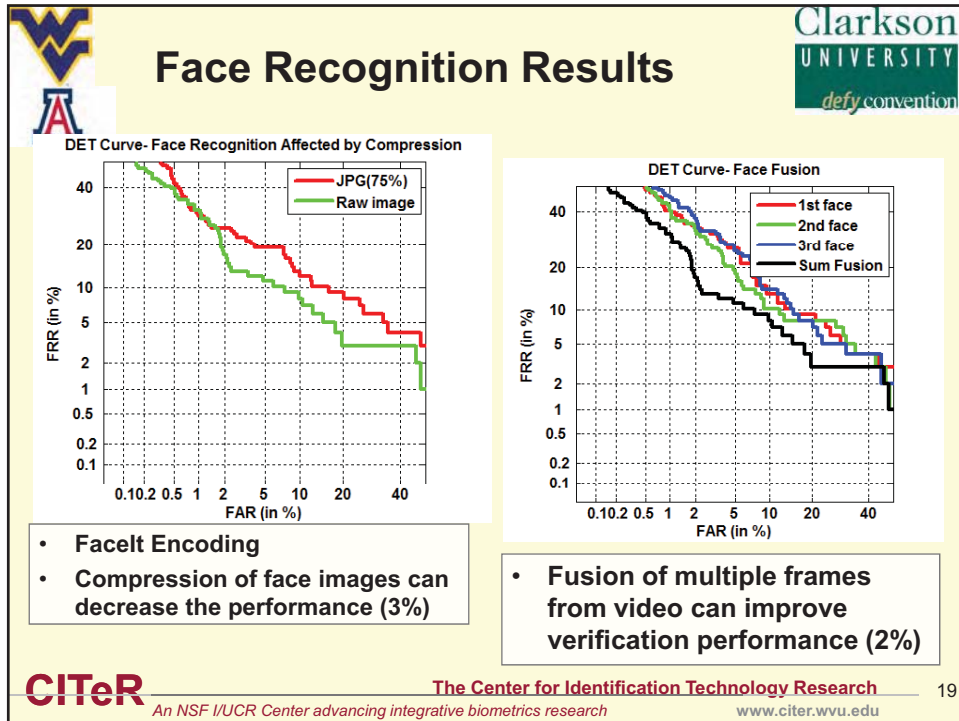
Manual Selection







The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research

18
www.citer.wvu.edu

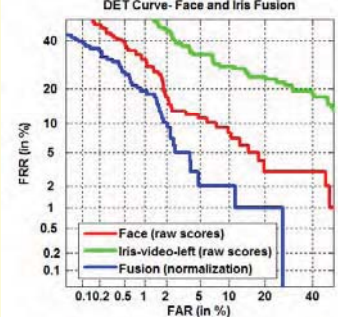


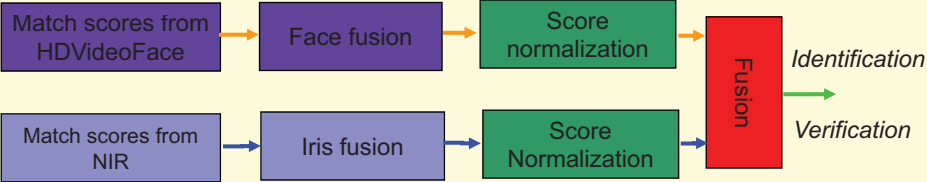



Multiple Biometrics Fusion



- Face:
 - 3 images from HDVideoFace automatically selected
 - Compared with StillFace
- Iris
 - 3 images from NIR manually selected
 - Compared with StillIris or VideoIris
- Fusion rules: Sum or Max
- Fusion after score normalization improves overall performance (4%)









The Center for Identification Technology Research

21


An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu



Importance





- Address face and iris problems found in operational data
 - Low to medium resolution face
 - Still and video iris
 - Near Infrared (NIR) & High Definition (HD) videos from portals
 - Unconstrained face recognition from still & video
- Assess the global and local quality measures
- Find the best way to do fusion: feature level? match score levels? fusion rules (Sum, Max, AND, OR, SVM, Decision Trees)?
- Assess the usage of quality measures at different levels



The Center for Identification Technology Research

22



An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu

Member Benefits

- Robust face and iris detection algorithms.
- Fast quality metrics for selection of “useful” frames.
- Robust algorithms for face and iris recognition using operational data
- New iris-face encoding for robust verification and identification
- Algorithm development for face and iris segmentation from challenging NIR and HD videos
- Optimal selection of fusion methods on different scenarios
- Systematic view of quality effects on biometric recognition and fusion
- Deliverable software prototypes



CITeR The Center for Identification Technology Research 23
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu

Milestones and Deliverables

Milestone	Description and Deliverable	Timeframe
(1) Participate...	Participation in the MBGC Oct 2008. Now changed to Dec 5 2008. Similarity scores submitted Nov. 3.	4 mos
(2) Propose...	Proposal to Fall meeting for Phase I	6 mos

CITeR The Center for Identification Technology Research 24
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu

Phase 1- Participation in MultiBiometric r . n . . h . l . l . n .



Final Report

Stephanie Schuckers, Clarkson University,
Natalia Schmid, West Virginia University
Besma Abidi, The University of Tennessee-Knoxville

Nadezhda Sazonova, Bozhao Tan, Uma Kandaswamy, Fang Hua
Jinyu Zuo, Francesco Nicolo, Yeon Rim

CITeR Fall 2009©

CITeR **The Center for Identification Technology Research**
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu



Problem

- Face and iris biometrics
 - Sensitive to the quality
 - Factors such as lighting, angle, pose, illumination, focus, resolution, and user's cooperation
- MultiBiometric Grand Challenge (MBGC) addresses some of the following questions:
 - Can face/iris video sequences improve the performance?
 - Can NIR face video provide sufficient information about iris for recognition based on iris?
 - What processing has to be done to use iris images from NIR face video for recognition of a large number of users?
 - What is the recognition performance of unconstrained face video with low to medium resolution?
 - Which way is better to fuse face and iris for biometric recognition, feature level? Match score level?
 - What's the effect of quality measure to the face and iris fusion?

• MBGC type images

Ref: Dr. P. Jonathon Phillips. "Multiple Biometric Grand Challenge Kick-Off Workshop" presentation, 18 April 2008



CITeR **The Center for Identification Technology Research** 2
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu



Research Plan

- Inventory of available algorithms for various components:
 - face localization, face tracking, face quality, face enhancement, face recognition, iris localization, iris segmentation, iris quality (including fast quality metric), iris recognition, and fusion strategies
- Determine which algorithms to be developed or modified to address challenges of MBGC data.
- Determine performance metrics for each and combined components and evaluate internal algorithms
- Create one or more super-algorithms. Evaluate their performance.
- Present results at 2008 and 2009 MBGC workshops

CITeR The Center for Identification Technology Research 3
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu

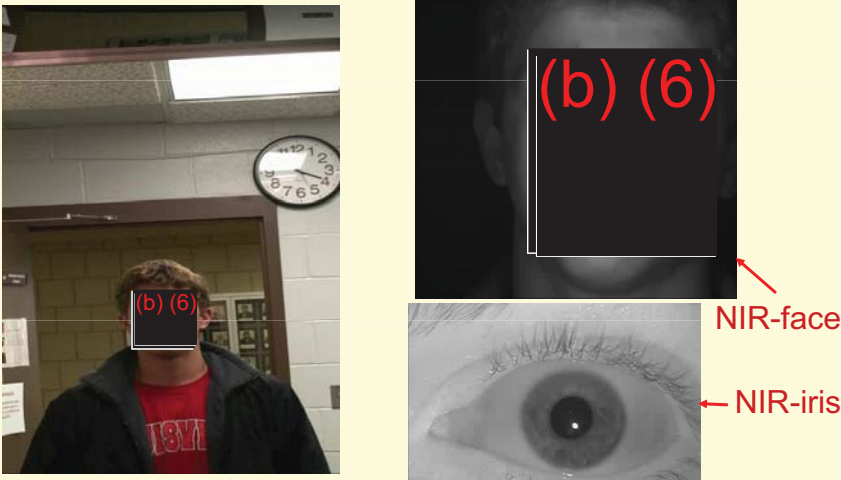


MultiBiometrics Grand Challenge

- **Portal Challenge (Face and Iris)** ← Our Focus
 - Still Face versus HD Video Face
 - Video Iris versus NIR Face Video
 - Still Iris versus NIR Face Video
 - Still Face / Video Iris versus HD Video Face / NIR Face Video
 - Still Face / Still Iris versus HD Video Face / NIR Face Video
- **Still Face Challenge**
- **Video Face Challenge**

CITeR The Center for Identification Technology Research 4
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu

Example Videos



Clarkson UNIVERSITY
defy convention

WV A

HD-face-video

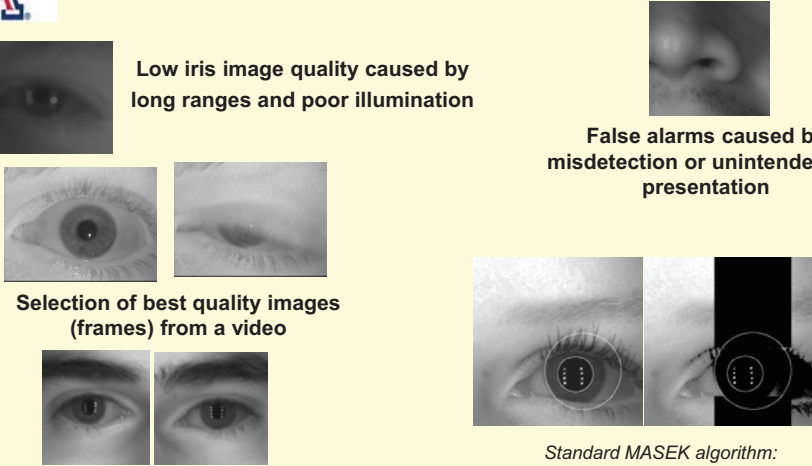
(b) (6)

NIR-face

NIR-iris

CITeR The Center for Identification Technology Research 5
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu

MBGC-IRIS: Challenges



Clarkson UNIVERSITY
defy convention

WV A

Low iris image quality caused by long ranges and poor illumination




False alarms caused by misdetection or unintended iris presentation

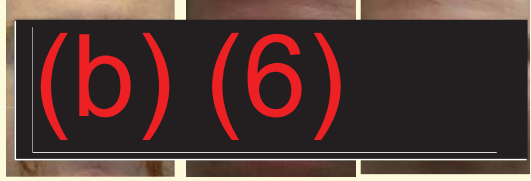
Selection of best quality images (frames) from a video

Fusion from multiple cropped images and multiple frames

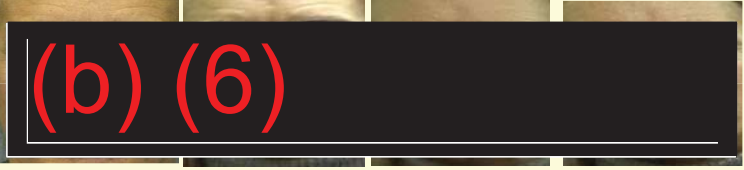
Standard MASEK algorithm:
Poor segmentation for NIR-face-video

CITeR The Center for Identification Technology Research 6
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu

 **MBGC-FACE: Challenges**  *defy convention* 






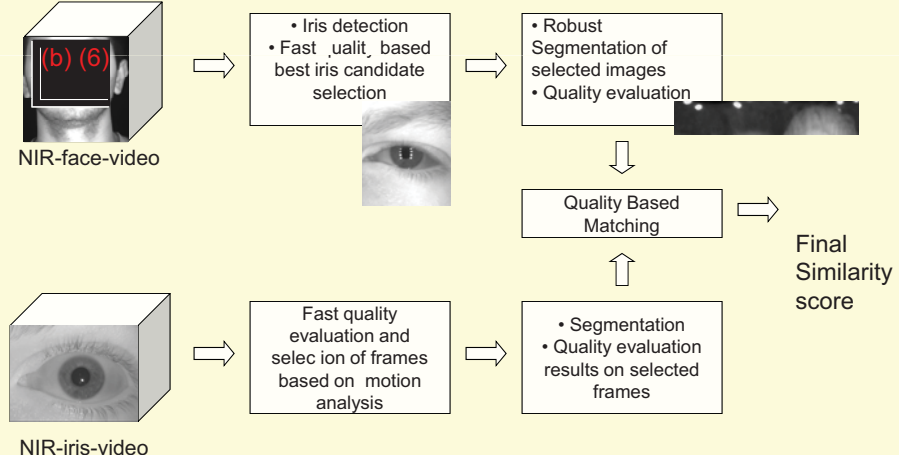
Left: gallery high resolution still face image, middle: video frame with lower resolution and different illumination, right: video frame with lower resolution and out of focus



From left to right: gallery high resolution still face image; video frame with lower resolution and non-visible eyes; video frame with lower resolution and motion blur; rotated face in video frame.

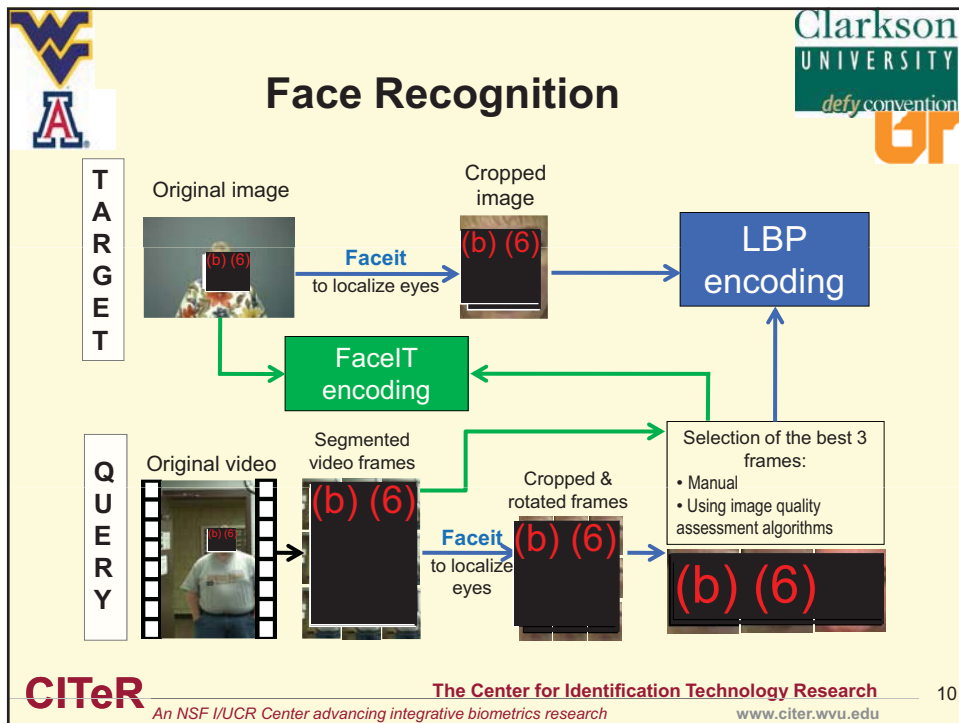
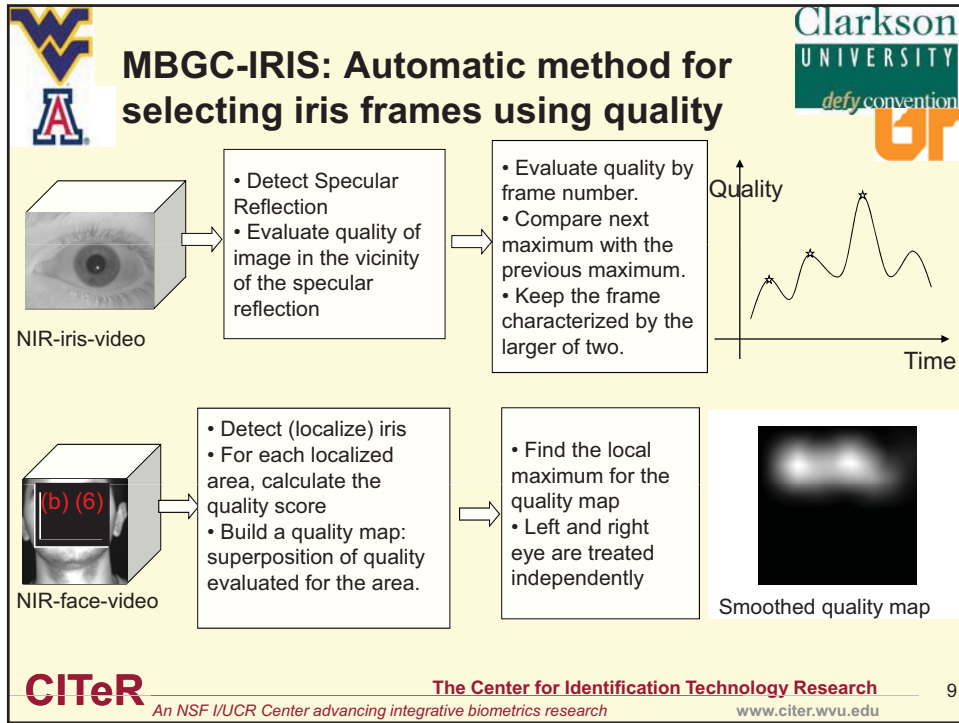
CITeR **The Center for Identification Technology Research** 7
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu


 **MBGC-IRIS: Our approach**  *defy convention* 




```
graph LR; A[NIR-face-video] --> B["• Iris detection  
• Fast quality based best iris candidate selection"]; B --> C["• Robust Segmentation of selected images  
• Quality evaluation"]; D[NIR-iris-video] --> E["Fast quality evaluation and selection of frames based on motion analysis"]; E --> F["• Segmentation  
• Quality evaluation results on selected frames"]; C --> G[Quality Based Matching]; F --> G; G --> H[Final Similarity score];
```

CITeR **The Center for Identification Technology Research** 8
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu

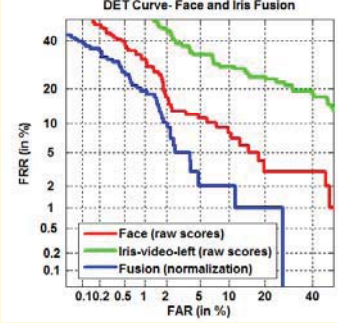




Multiple Biometrics Fusion

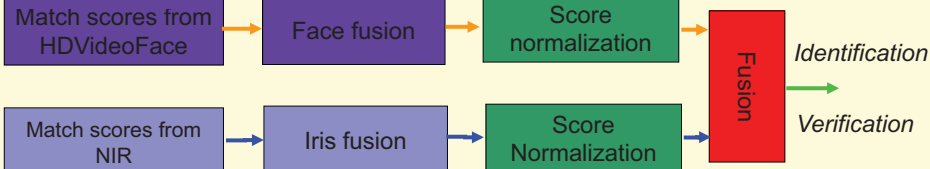


- Face:
 - 3 images from HDVideoFace automatically selected
 - Compared with StillFace
- Iris
 - 3 images from NIR manually selected
 - Compared with StillIris or VideoIris
- Fusion rules: Sum or Max
- Fusion after score normalization improves overall performance (4%)




DET Curve- Face and Iris Fusion


FAR (in %)	Face (raw scores) FRR (in %)	Iris-video-left (raw scores) FRR (in %)	Fusion (normalization) FRR (in %)
0.1	40	40	40
0.2	35	35	35
0.5	25	25	25
1	20	20	20
2	15	15	15
5	10	10	10
10	8	8	8
20	6	6	6
40	5	5	5




```
graph LR; A[Match scores from HDVideoFace] --> B[Face fusion]; B --> C[Score normalization]; C --> D[Fusion]; E[Match scores from NIR] --> F[Iris fusion]; F --> G[Score Normalization]; G --> D; D --> H[Identification]; D --> I[Verification];
```




The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research

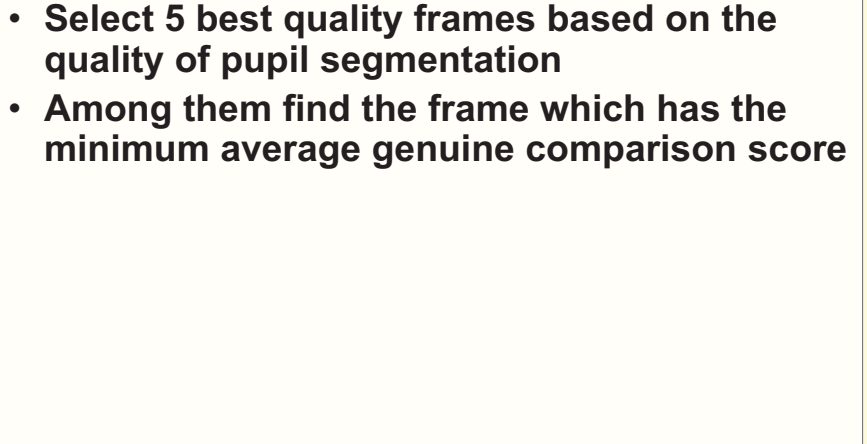





Frame selection within iris video

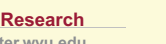


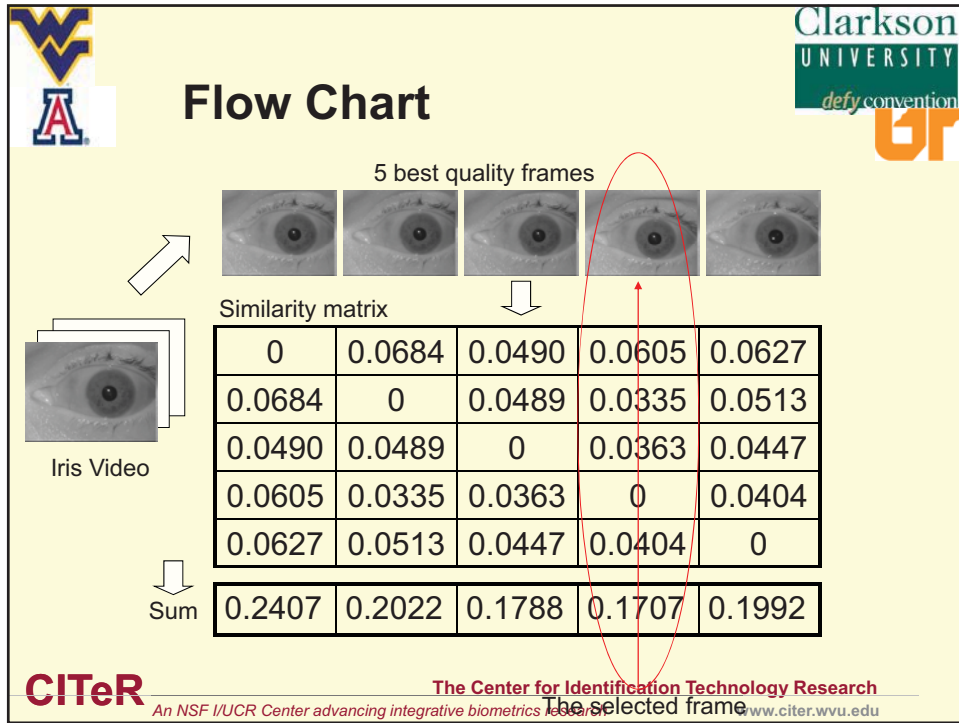
- **Select 5 best quality frames based on the quality of pupil segmentation**
- **Among them find the frame which has the minimum average genuine comparison score**





The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research

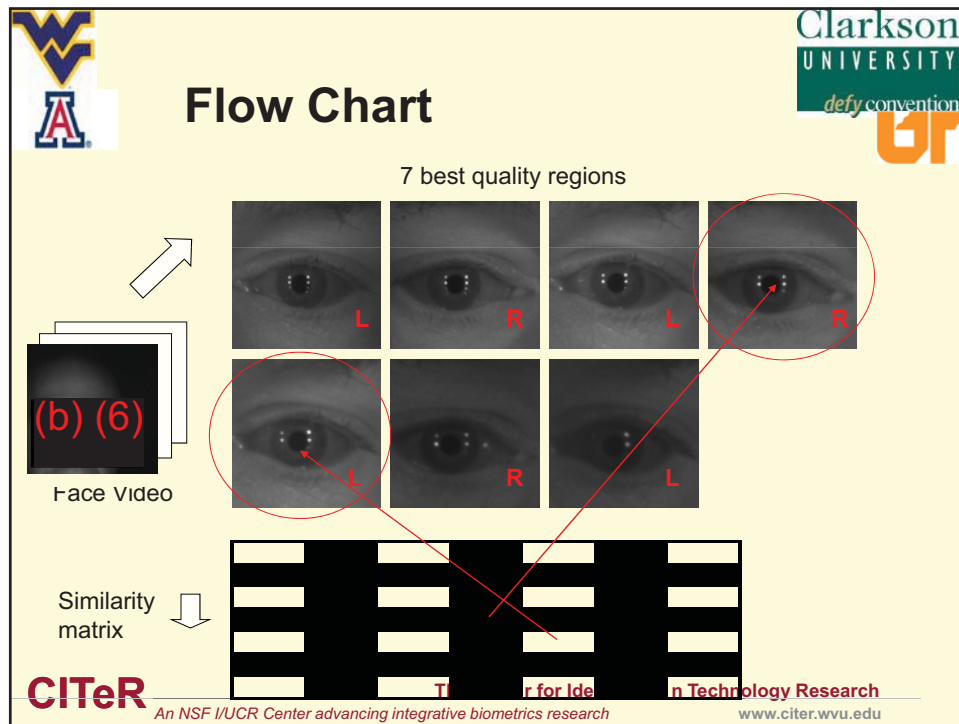




Area selection for NIR face video

- Detect the eye area (2 eyes)
- Evaluate sharpness
- Label each area with “left eye”, “right eye” or “unknown”
- Select about 10 best quality cropped regions based on the image quality (sharpness measure)
- Calculate the similarity scores between all regions
- Find the best cropped region which results in the minimum average genuine score for “left” or “right” eye
- Find the best cropped area which has the minimum average comparison scores for “unknown” region

CITeR The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu



Fusion

Clarkson UNIVERSITY
defy convention

- Log-Gabor, 2D-Gabor, Ordinal, Median Filter and SIFT based encoding techniques
- Not all of them are suitable for low quality NIR face video
- Fusion rules: weighted sum rule, AdaBoost-based fusion.

CITeR The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu

Face experiments

Clarkson UNIVERSITY
defy convention

(b) (6)

(b) (6)

(b) (6)

(b) (6)

(b) (6)

(b) (6)

C

idva

ica

Face: Encoding changes

Clarkson UNIVERSITY
defy convention

Updated
LBP encoding

NEW
Edge-based encoding

(b) (6)

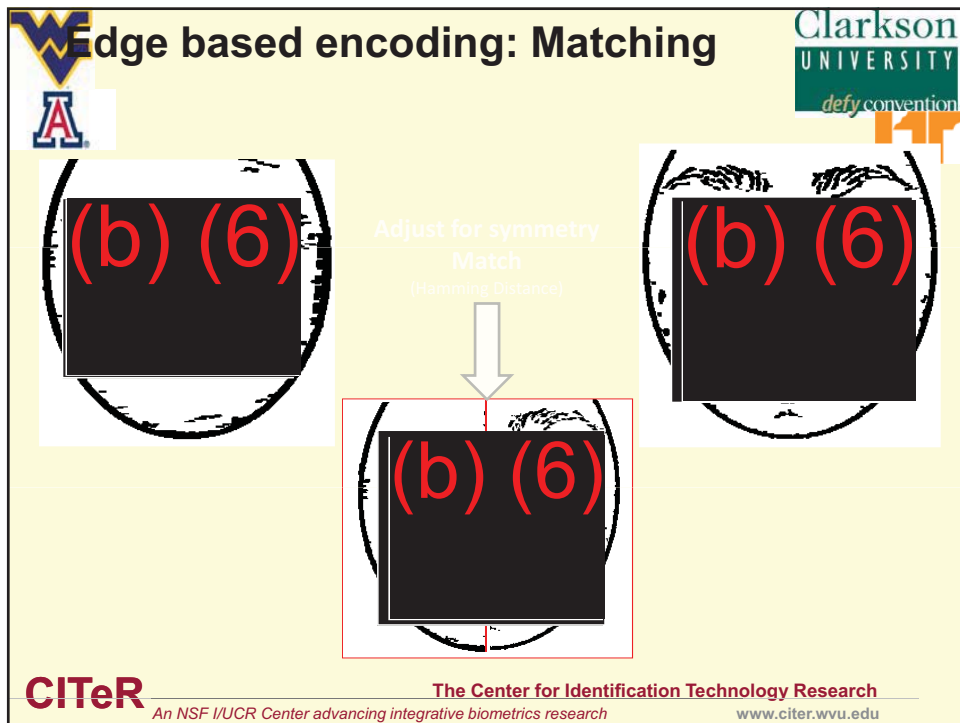
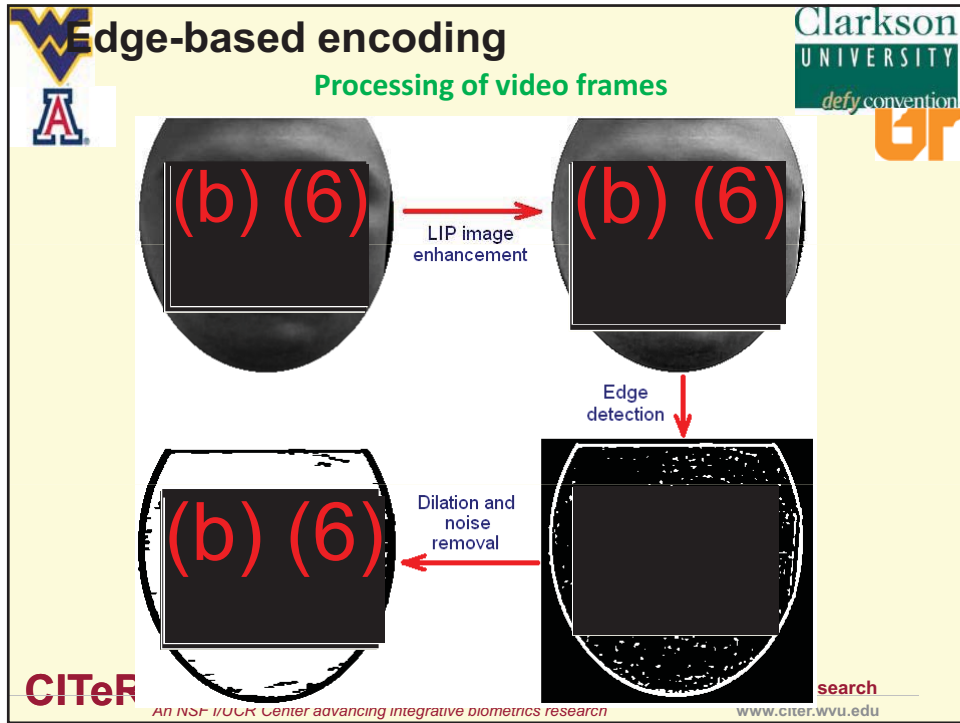
(b) (6)


Combined score

Bottom part of an image is weighted less during matching to account for the differences in facial expressions


Insensitive to illumination
Emphasizes distinctive facial features

CITeR The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu





Iris image enhancement



NIR face video

NIR iris video

(a) Original

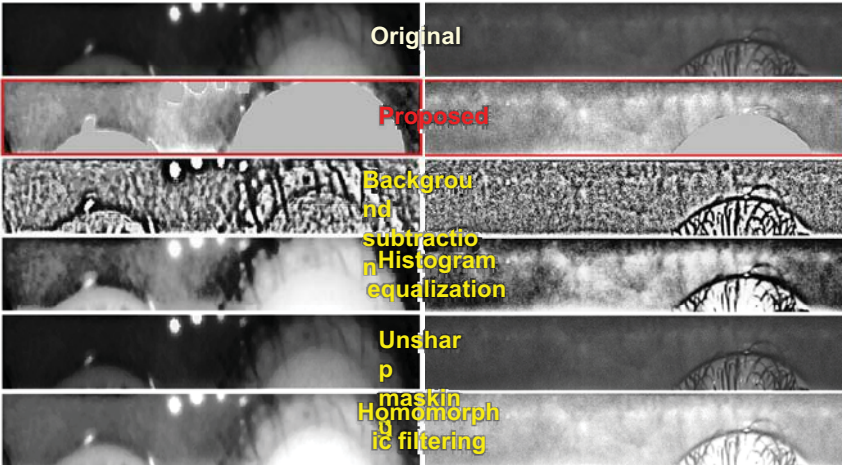
(b) Proposed


(c) Background subtraction

(d) Histogram equalization

(e) Unsharp mask


(f) Homomorphic filtering






CITeR
An NSF I/UCR Center advancing integrative biometrics research

The Center for Identification Technology Research
www.citer.wvu.edu

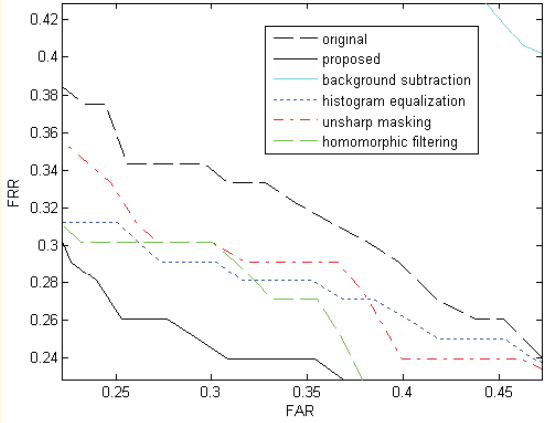


Iris image enhancement




Effect on iris recognition

(MBGC-1 data: video iris experiments)





FAR	original	proposed	background subtraction	histogram equalization	unsharp masking	homomorphic filtering
0.25	0.38	0.30	0.31	0.31	0.35	0.30
0.30	0.34	0.26	0.29	0.29	0.30	0.27
0.35	0.32	0.24	0.28	0.28	0.29	0.25
0.40	0.28	0.24	0.25	0.25	0.24	0.24
0.45	0.26	0.24	0.24	0.24	0.24	0.24




CITeR
An NSF I/UCR Center advancing integrative biometrics research

The Center for Identification Technology Research
www.citer.wvu.edu

Video De-Interlacing

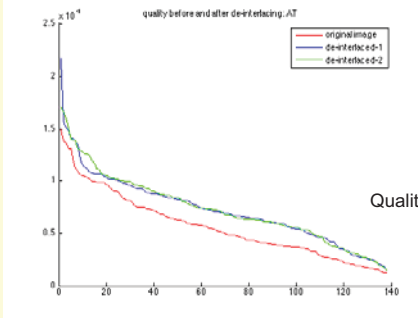



Interlaced De-interlaced



$$F_e(x, n) = \begin{cases} F_1(x, n) & \text{if } n \text{ is odd} \\ F_2(x, n) & \text{if } n \text{ is even} \end{cases}$$

$F_1(x, n)$ interpolated pixel



Quality values before and after de-interlacing



CITeR

An NSF I/UCR Center advancing integrative biometrics research

The Center for Identification Technology Research

www.citer.wvu.edu

Illumination Normalization/Compensation

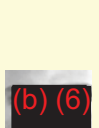



After processing

$I(x, y) = D' f_a(\theta) + S' f_s(\theta) \hat{n} \cdot \hat{i}$


$[k_1 \dots k_n \dots k_m] \cdot [H_1(\theta) \dots H_m(\theta)]$

(b) (6)



Original Input Image

(b) (6)

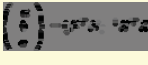


Specularity Removal

By means of rotation of coordinate system, generate specularity-independent image

Illumination Plane Subtraction

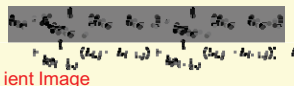
$I(x, y) = a \cdot x + b \cdot y + c$ P(x,y), the illumination plane



Illumination Plane Subtraction with Histogram Equalization

Shadow Illuminator

$I(x, y) = R(x, y)L(x, y)$ where $R(x, y)$ is the reflectance and $L(x, y)$ is the illuminance



Self-Quotient Image

$I(x, y) = \frac{I(x, y)}{WG \cdot I(x, y)}$ $W = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ $I \in R_2$ $I \in R_3$ $G =$ Gaussian kernel

Local Binary Pattern

5	3	4
7	5	2
9	6	1

Example

1	1	0
1	0	0
1	1	0


Thresholded

CITeR


An NSF I/UCR Center advancing integrative biometrics research

The Center for Identification Technology Research


www.citer.wvu.edu



Tested Quality Values on Various Enhancement




defy convention



Methods

Image No.	Method	AI (0-5)	TE (0-5)	AI (0-5)	TE (0-5)	MidW (0-5)	Comp (0-5)	Change
a.	Original	0.57	0.91	2.11	2.21	0.0000	0.00	0.79
b.	IP	0.51	0.90	2.16	2.21	0.0000	0.00	1.20
c.	IP+SI	0.16	0.22	0.04	0.00	0.0000	0.00	42.92
d.	LIP	0.09	0.04	0.01	0.00	0.0000	0.00	0.00
e.	SI	0.05	0.05	0.00	0.00	0.0000	0.00	0.00
f.	SUV	0.00	0.00	0.00	0.00	0.0000	0.00	0.00
g.	SUV+SI	0.02	0.05	0.00	0.00	0.0000	0.00	0.00
h.	SUV+SI+SI	0.07	0.07	0.00	0.00	0.0000	0.00	0.00
i.	SUV+SI+SI+SI	0.00	0.00	0.00	0.00	0.0000	0.00	0.00
j.	SUV+SI	0.02	0.00	0.00	0.00	0.0000	0.00	0.00
k.	SUV+SI	0.02	0.00	0.00	0.00	0.0000	0.00	0.00
l.	SUV+SI	0.02	0.00	0.00	0.00	0.0000	0.00	0.00
m.	SUV+SI	0.02	0.00	0.00	0.00	0.0000	0.00	0.00
n.	SUV+SI	0.02	0.00	0.00	0.00	0.0000	0.00	0.00
o.	SUV+SI	0.02	0.00	0.00	0.00	0.0000	0.00	0.00
p.	SUV+SI	0.02	0.00	0.00	0.00	0.0000	0.00	0.00
q.	SUV+SI	0.02	0.00	0.00	0.00	0.0000	0.00	0.00
r.	SUV+SI	0.02	0.00	0.00	0.00	0.0000	0.00	0.00
s.	SUV+SI	0.02	0.00	0.00	0.00	0.0000	0.00	0.00
t.	SUV+SI	0.02	0.00	0.00	0.00	0.0000	0.00	0.00
u.	SUV+SI	0.02	0.00	0.00	0.00	0.0000	0.00	0.00
v.	SUV+SI	0.02	0.00	0.00	0.00	0.0000	0.00	0.00
w.	SUV+SI	0.02	0.00	0.00	0.00	0.0000	0.00	0.00
x.	SUV+SI	0.02	0.00	0.00	0.00	0.0000	0.00	0.00
y.	SUV+SI	0.02	0.00	0.00	0.00	0.0000	0.00	0.00
z.	SUV+SI	0.02	0.00	0.00	0.00	0.0000	0.00	0.00


- SUV + SI is selected for best appearance, speed, and lowest quality value as measured by Entropy
- SUV removes highlights and specularities
- SI homogenizes illumination




An NSF I/UCR Center advancing integrative biometrics research

The Center for Identification Technology Research


www.citer.wvu.edu

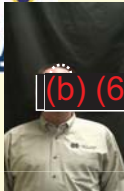


Sample Images and their Enhancements

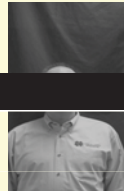


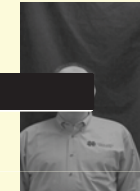
defy convention






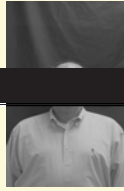
(b) (6)

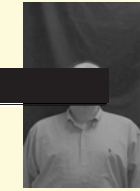


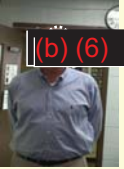




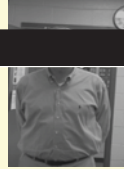
(b) (6)

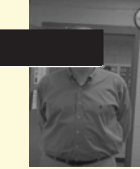







(b) (6)



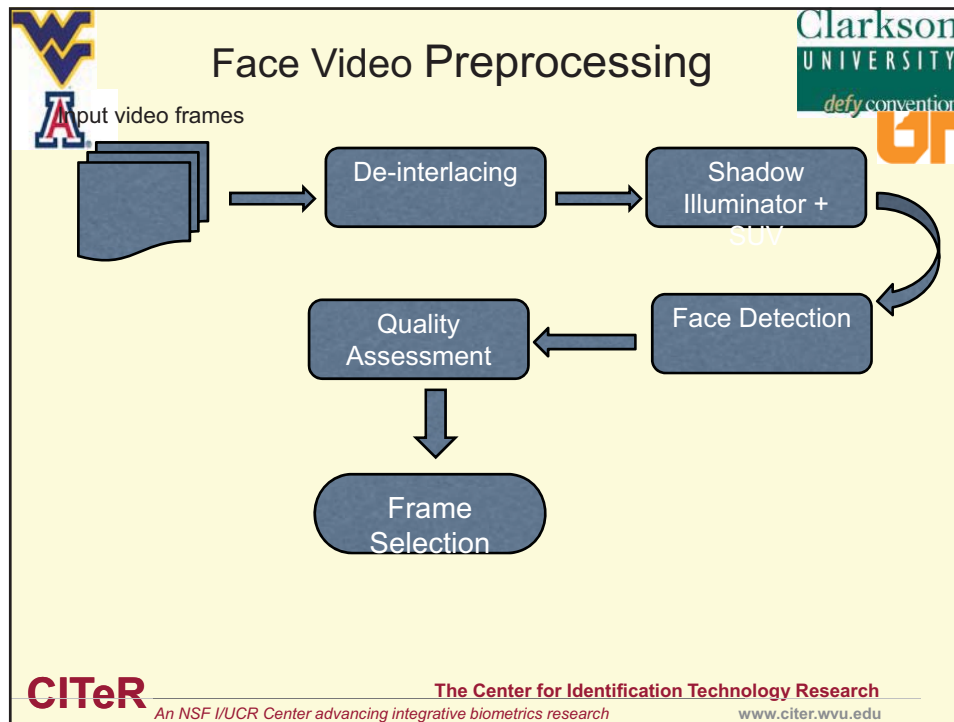




An NSF I/UCR Center advancing integrative biometrics research

The Center for Identification Technology Research


www.citer.wvu.edu




Work Remaining

- MBGC Challenge 2 due in early November
- MBGC Challenge 2 Workshop—Dec. 4, 2009
- Combine modules for final iris algorithm design which includes quality-based frame selection, segmentation, image enhancement, encoding, matching
- Combine modules for final face algorithm design which includes de-interlacing, enhancement, face detection, quality-based frame selection, encoding, matching


CITeR The Center for Identification Technology Research 28
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu



Milestones and Deliverables



Milestone	Description and Deliverable	Timeframe
(1) Study...	Study available face and iris recognition algorithms and fusion strategies with quality Determine performance metrics for subroutines, based on Challenge 1 data	1 month
(2) Develop...	Develop automatic iris segmentation on challenging NIR face video, NIR iris video or still iris image Automatic face locating, normalization,	3 months
(3) Validate...	Test the proposed algorithms on MBGC Challenge 1, Challenge 2 data	4 months & 9 months
(4) Publish..	Prepare description of algorithms for publication.	12 months



The Center for Identification Technology Research

An NSF I/UCR Center advancing integrative biometrics research

www.citer.wvu.edu

Unconstrained Face Recognition Under Non-Ideal Conditions

Final Report

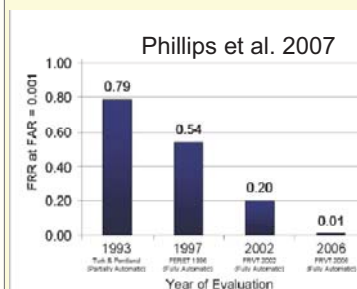
Thirimachos Bourlai¹, Anil K. Jain², Arun Ross¹

¹ West Virginia University

² Michigan State University

Introduction

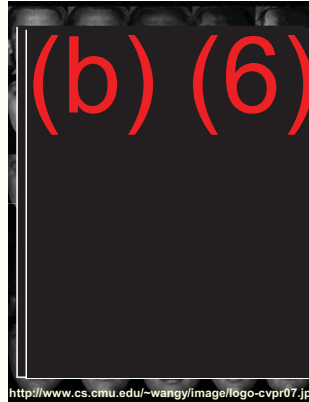
- The performance of face recognition (FR) algorithms has **significantly improved** over the past ~15 years
 - FRGC/FRVT evaluations
- However, **uncontrolled images** can severely impact the performance of FR algorithms
- At a FAR=0.1% (FRVT 2006)
 - FRR = ~2.5% (Controlled high-res)
 - FRR = ~12.5% (Uncontrolled high-res)



2

Introduction

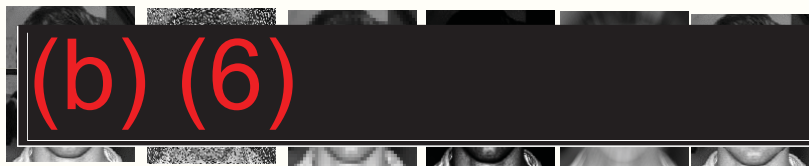
- The problem of matching face images acquired using **different cameras** and/or subjected to **severe photometric** (e.g., illumination) and **geometric** (e.g., compression) degradation continues to pose many challenges



3

Problem Statement

- **Task 1:** Design algorithms to match high-resolution face images against their degraded counterparts



- **Task 2:** Design algorithms to extract facial objects from these videos and images (esp. YouTube™)

4

Milestones and Deliverables

Milestone	Description and Deliverable	Timeframe
(1) Image Normalization	Design algorithms to perform photometric and geometric normalization of several de_rated images	3 months
(2) Image Matching	Design methods to match facial images using morphable templates and other sophisticated 2D shape/illumination models	6 months
(3) Performance Evaluation	Report ROC/CMC curves after applying the scheme on public high-resolution datasets	3 months

5

Image Degradations - 1

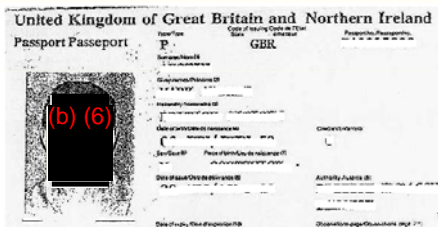
- **JPEG compressed images**

- Blocky artifacts



- **Fax images**

- Noise

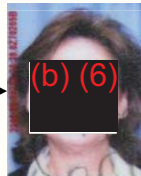


<http://media.canada.com/290cac5f-206a-453b-b3f1-ae7d176af64/ritter.jpg>

6

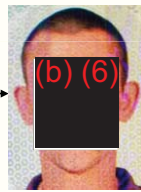
Image Degradations - 2

- **Hologram imposed images**
 - Image quality may be good
 - But additional features might be induced



Notice the lines/marks induced by the hologram

http://images.townnews.com/helenair.com/content/articles/2008/06/10/top/top/40st_080610_license.jpg



The facial skin texture of the subject is altered by the hologram.

http://www.lim-corlett.com/chris/CCorlett_PP.jpg

7

Mitigating the Effects of Degradation

- **Research issues**
 - Study the **impact of various types of degradations** on recognition performance
 - Develop **pre-processing methods** which enhance the quality of the degraded images so that they can be successfully matched

8

Project Direction

- Generated “**degraded**” datasets using the following two techniques:
 - Scanning passport photographs
 - Applying FAX compression on high-res face images
- Performed **verification** experiments involving high-resolution face images and degraded images
- Considered **multiple face recognition algorithms** for the study (Verilook, Identix G6, LDA+knn)

9

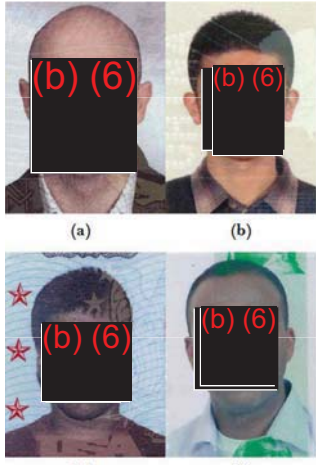
West Virginia University MICHIGAN STATE UNIVERSITY CITeR CITeR 2009

Passport Facial Matching

Passport Facial Matching

CHALLENGES

- ✚ **Person-related**
 - Hairstyle
 - Expression
 - Pose of the individual
 - Aging factor
- ✚ **Document-related**
 - Security watermarks embedded on passport photos
 - Variations in image quality
 - Tonality across the face
 - Color cast of the photographs
- ✚ **Scanning device-related**
 - Foibles of the scanner
 - Resolution
 - Artifacts due to lighting
 - Smaller size of the document photo
 - Image file format/compression used
 - Operator variability





The mug-shots taken from passports issued by different countries (a) Greece (issued 2006), (b) China (issued 2008), (c) US (issued 2008), and (d) Egypt (issued 2005).

West Virginia University MICHIGAN STATE UNIVERSITY CITeR CITeR 2009

Proposed Technique

Hardware and Settings

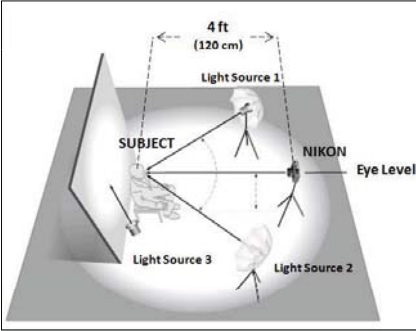
<p>NIKON Coolpix P-80</p> 	PURPOSE	SETTINGS
	<p>1) Capture face image of the live subject</p> <p>2) Capture the subject's passport face image</p>	<ul style="list-style-type: none"> • Portrait Mode • Spatial Resolution 3648x2736 • Auto-Focus • Optical Vibration Reduction • Image Stabilization • ISO AUTO • In-Camera Red-Eye Fix • JPEG format

<p>HP Officejet Pro L7780</p> 	PURPOSE	SETTINGS
	<p>Capture (scanner use) the subject's passport face image</p>	<ul style="list-style-type: none"> • Spatial Resolution 2900x2000 • JPEG format

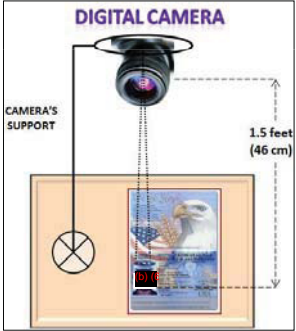
West Virginia University MICHIGAN STATE UNIVERSITY CITeR CITeR 2009

Proposed Technique

Image Capture Setup



(a) The **live subject-capture setup**. Taken from the *US State Department, Bureau of Consular Affairs*, and enhanced for visual purposes by the authors




(b) The **passport-capture setup** used for data collection

West Virginia University MICHIGAN STATE UNIVERSITY CITeR CTeR 2009

Passport DB

➤ **Passport Database (PassportDB)** is composed of three datasets:

- **NFaceD** - high-resolution face photographs from live subjects
- **NPassFaceD** - images of passport photos
- **HPassFaceD** - face images scanned from the photo page of passports



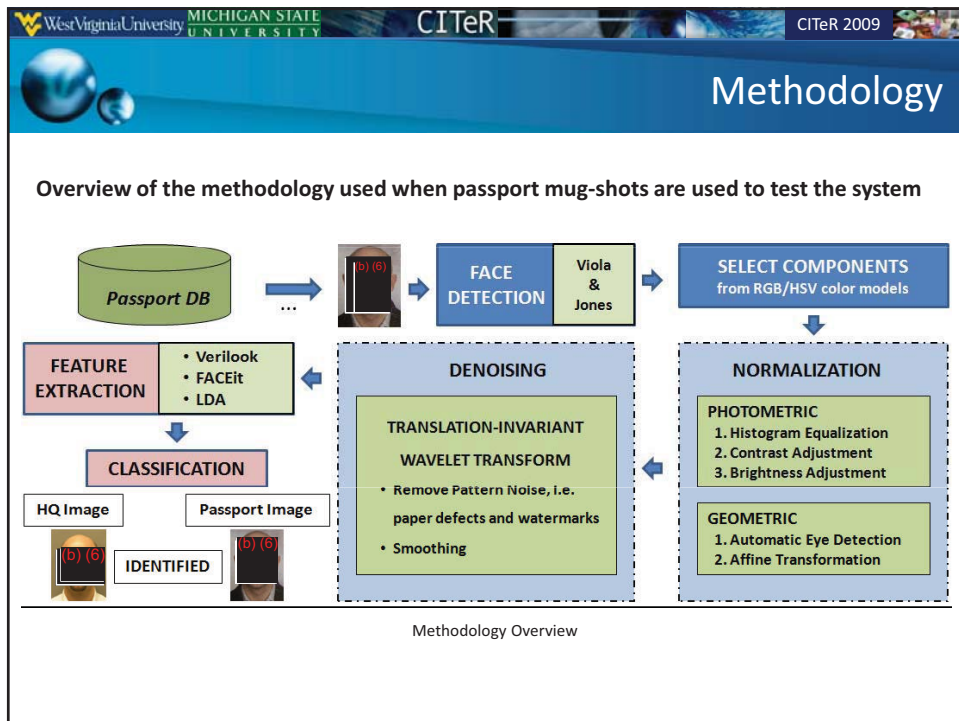
HQ Image (Nikon Camera) Passport Image (Nikon Camera) Passport Image (HP scanner)

Samples of a subject taken from the three datasets of the Passport Database.

➤ The **NPassFaceD** and **HPassFaceD** were assembled during the first session

➤ **NPassFaceD**: 3 samples of the photo page of the passport were acquired for each subject (to compensate for paper reflections & camera motion)

➤ **HPassFaceD**: one scan (per subject) was sufficient to capture a reasonable quality mug-shot from the passport



West Virginia University MICHIGAN STATE UNIVERSITY
CITeR
CITeR 2009

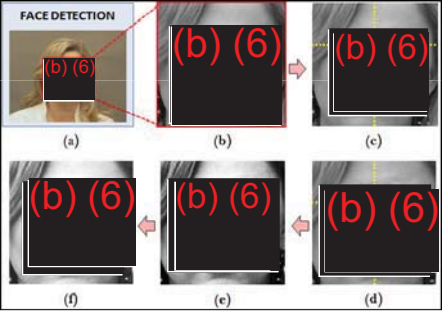
Empirical Evaluation

Overview of the **methodology** used to perform **face normalization**:

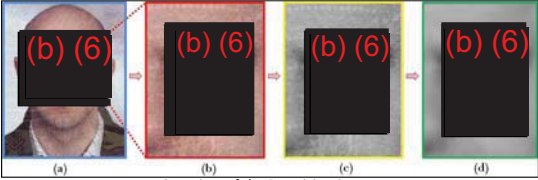
- (a) Face detection
- (b) Convert to gray scale
- (c) Eye detection
- (d) Geometric Normalization
- (e) Histogram Equalization
- (f) Contrast/Brightness Adjustment

Illustration of **denoising**:

- (a) A sample scanned passport face image taken from the HPassFaceD dataset
- (b) Selected ROI
- (c) ROI Grayscale version before denoising
- (d) ROI after denoising



Face Normalization Methodology Overview



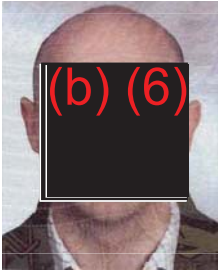
Overview of the Denoising Steps

West Virginia University MICHIGAN STATE UNIVERSITY
CITeR
CITeR 2009

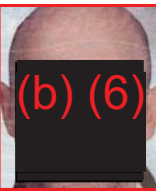
Visual Results

STEP 1: ORIGINAL
STEP 2: FACE DETECTION
STEP 3: CHANNEL SELECTION

ORIGINAL




FACE DETECTION




CHANNEL SELECTION

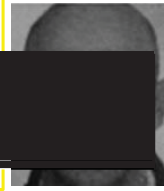
R




G




B




H



S



V



Face Recognition Systems

The techniques employed to perform the recognition experiments include:

- Commercial software provided by **Verilook** [3]
- Commercial software **FaceIT**® provided by L1 Systems [4]
- Linear Discriminant Analysis [5] in combination with the k-Nearest Neighbor (k-NN) algorithm [6]

Experiments

Experiment 1 : - **Train** with Session 1 of the NFaceD dataset, i.e. 175 images
 - **Test** with Session 2 of the NFaceD dataset, i.e. 175 images

Experiment 2 : - **Train** with the whole NFaceD dataset, i.e. 350 images
 - **Test** with the passport mug-shots (Nikon), i.e. 75 images - 25 subjects with 3 images per subject

Experiment 3 : - **Train** with the whole NFaceD dataset, i.e. 350 images
 - **Test** with the HP scanner, i.e. 25 subjects with 1 image per subject)

Experiment 4 : - **Tested the pre-processing effects on system performance**

[3] Neu otechnology. Ye look 3.2 x86 fo face ecogn t on. http://www.neu.otechnology.com/face-b-cmet_cs.html, 2009.
 [4] FaceIT, L1 Systems, <http://ep.c.s.p. vacy/su.ve/lanee/post/gh/1105/facefss.pdf>, 2009.
 [5] J. Hespanha et al. "E genfaces vs. f she faces" Recogn t on us ng class spec f c l nea p object on," n IEEE T ans. on PAMI, 19 4558, 1996.
 [6] T. M. Cove and P. E. Ha t. Nea est ne g hbo patte n class f cat on. IEEE T ans. info m. Theo y, 13(1) 2127, 1967.

West Virginia University
MICHIGAN STATE UNIVERSITY
CITeR
CITeR 2009

Pre-Processing Effects

Effect of Pre-Processing (ROC Results)

The main effects of face normalization and denoising on system performance were studied in the case of Exp 2, i.e. **Train NFaceD dataset – Test Passport mug-shots by Nikon** :

- Select Channels
- Geometric Normalization
- Photometric Normalization
- Denoising
- Data Refining

Pre-Processing case study. The experiments were conducted by using all available images of the NFaceD dataset for training and the images from the NPassFaceD for testing.

West Virginia University
MICHIGAN STATE UNIVERSITY
CITeR
CITeR 2009

Pre- vs. Post-Processing Effects

EXPERIMENTAL RESULTS

Table 4. PERFORMANCE RESULTS IN TERMS OF EER OF ALL EXPERIMENTS

Comparing HQ Digital Photos vs. Camera/Scanner Passport Photos : LDA & k-NN method

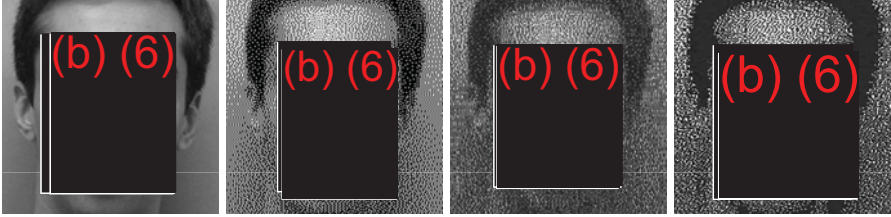
S1/S2=SESSION 1/2; PRE-PROC=PRE-PROCESSING

Train	Test	Pre-Proc	Method	EER (%)
NFaceD(S1)	NFaceD(S2)	-	Facelt	0.59
NFaceD(S1)	NFaceD(S2)	-	Verilook	1.11
NFaceD(S1)	NFaceD(S2)	-	LDA & k-NN	3.58
NFaceD(S1/S2)	NFacePassD	No	Facelt	70.83
NFaceD(S1/S2)	NFacePassD	No	Verilook	-
NFaceD(S1/S2)	NFacePassD	No	LDA & k-NN	24.05
NFaceD(S1/S2)	NFacePassD	Yes	Facelt	48.70
NFaceD(S1/S2)	NFacePassD	-	Verilook	-
NFaceD(S1/S2)	NFacePassD	Yes	LDA & k-NN	5.49
NFaceD(S1/S2)	HFacePassD	No	Facelt	47.02
NFaceD(S1/S2)	HFacePassD	No	Verilook	-
NFaceD(S1/S2)	HFacePassD	No	LDA & k-NN	26.82
NFaceD(S1/S2)	HFacePassD	Yes	Facelt	58.11
NFaceD(S1/S2)	HFacePassD	-	Verilook	-
NFaceD(S1/S2)	HFacePassD	Yes	LDA & k-NN	9.89

West Virginia University MICHIGAN STATE UNIVERSITY CITeR CITEr 2009

FAX Facial Matching

Probe images of a subject under different fax scenarios



ORIGINAL FAX Compression Printed FAX AFTER FAX

West Virginia University MICHIGAN STATE UNIVERSITY CITeR CITEr 2009

FAX Facial Matching

EXPERIMENTS

DATA : High Quality Images (the same that was used in the Passport DB)

SCENARIOS : Apply FAX compression on the data. Compare the original HQ images to the compressed FAX images before and after they are sent via FAX

FACE RECOGNITION SYSTEMS:

1. Verilook
2. FaceIT[®] provided by L1 Systems
3. LDA with k-NN

West Virginia University MICHIGAN STATE UNIVERSITY CITeR CTeR 2009

METHODOLOGY


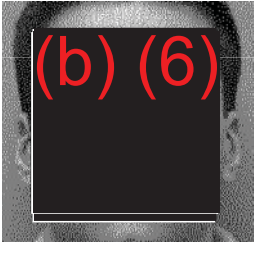
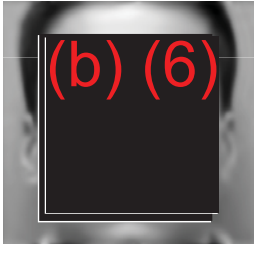
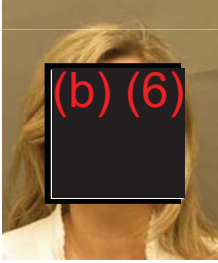
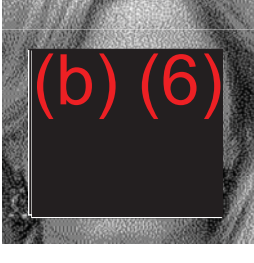
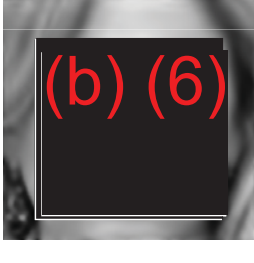
METHODOLOGICAL STEPS

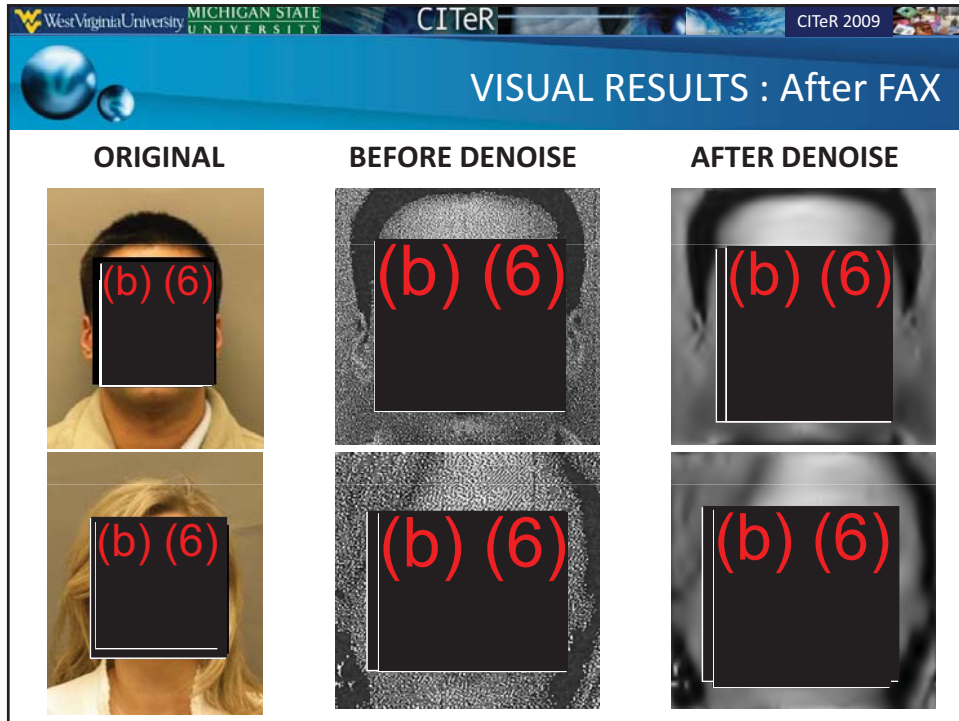
- Face Detection:** The Viola & Jones [1] face detection algorithm is used to localize the spatial extent of the face and determine its boundary
- Geometric Normalization:** In the next step, a geometric normalization scheme is applied to the original and degraded images after detection. It is composed of two main steps, viz., eye detection and affine transformation
- Denoising:** Translation-Invariant Wavelet Transform (TI-WT) [2]. It averages out the translation dependence and is used to remove the pattern noise from the fax face images
- Feature Extraction and Classification:** The face recognition techniques are then used to perform matching and generate match scores

[1] Viola, P. A. and Jones, M. J., "Robust real-time face detection," International Journal of Computer Vision 57(2), 137-154 (2004).
[2] Coifman, R. R. and Donoho, D. L., "Translation-invariant de-noising," In Wavelets and Statistics, Springer Lecture Notes in Statistics 103, 125-150 (1994).

West Virginia University MICHIGAN STATE UNIVERSITY CITeR CTeR 2009

VISUAL RESULTS : Before FAX

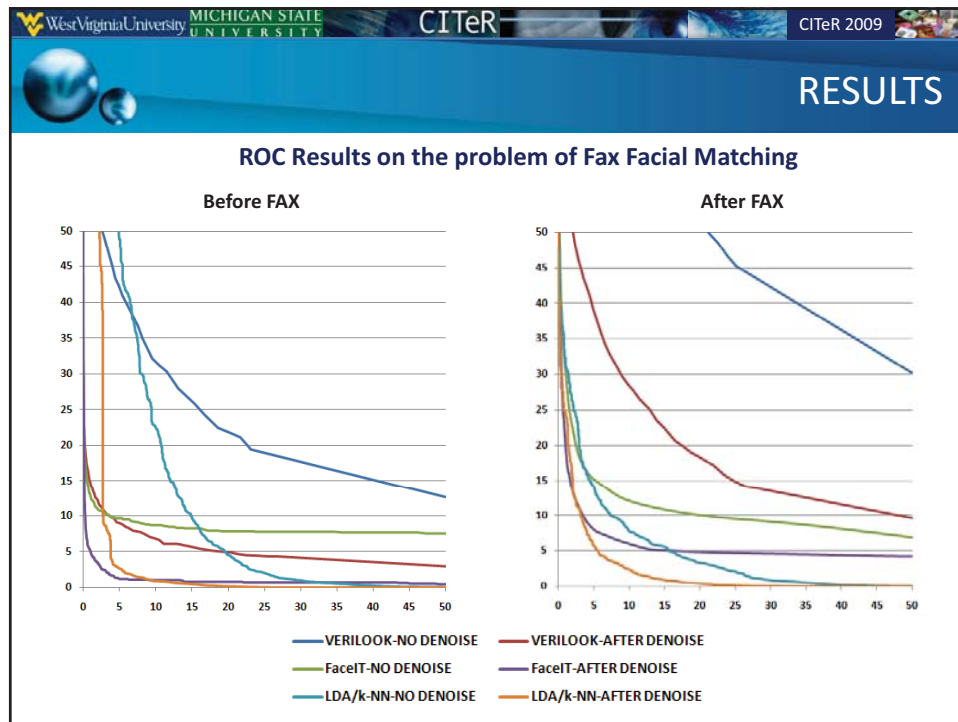
ORIGINAL	BEFORE DENOISE	AFTER DENOISE
		
		



RESULTS

TABLE : Results on the problem of Fax Facial Matching

Face Recognition Techniques	DENOISE	SCENARIOS (HQ vs.)	Performance (EER %)
LDA&k-NN	NO	HQ	3.26
		Before FAX	12.83
		After FAX	8.8
	YES	-	-
		Before FAX	3.72
		After FAX	5.34
VERILOOK	NO	HQ	1.25
		Before FAX	21.11
		After FAX	45.23
	YES	-	-
		Before FAX	7.1
		After FAX	19.23
FACEIT	NO	HQ	0.41
		Before FAX	8.65
		After FAX	12.95
	YES	-	-
		Before FAX	2.61
		After FAX	7.53



Member Benefits

- Method to pre-process **highly disparate facial images** obtained using different cameras, under varying illumination conditions, subjected to severe degradations
- Applications include:
 - Matching **scanned passport photos** to high-resolution digital images
 - Identifying face images in **YouTube™** videos
 - Identifying individuals in **surveillance videos (CCTVs)**

28

Publications and Future Work

- Currently, we are conducting **identification** experiments using the passport and fax photos
- The **number of identities** in the database has to be increased
- The impact of **severe geometric perturbations** has to be studied
- **Publications:**
 - T. Bourlai, A. Ross and A. K. Jain, "On Matching Digital Face Images Against Scanned Passport Photos," Proc. of First IEEE International Conference on Biometrics, Identity and Security (BIdS), (Tampa, USA), September 2009
 - Working on a journal paper that summarizes the results generated in this work

29

An Acquisition Platform for Non-Cooperative, Long Range Ocular Biometrics

Progress Report CITeR Conference, November 2008

Reza Derakhshani¹, Besma Abidi², and
Plamen Doynov¹

¹ University of Missouri, Kansas City

² University of Tennessee, Knoxville

CITeR

The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research

www.citer.wvu.edu

Problem Description

- Range, precision, and reliability of ocular biometric systems are significantly impacted by the quality of input data (e.g. long range surveillance)
- Goal: a COTS hardware platform for iris recognition from distances up to 10 meters, and possibly without cooperation from the subjects

CITeR

The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research

www.citer.wvu.edu

Our Approach

Build a COTS-based system where

•Cameras make eye contact with the subjects

- Customized XUUK™ system with for long-range gaze detection, in conjunction with a pan-tilt (PT) mounted area-scanning main camera

•Use advance imaging techniques, especially lucky imaging, for long range acquisition of the iris

- High-magnification resolution optics
- NIR-enhanced high-speed image sensor in burst mode
- Real-time software to evaluate image quality and to pick the best from each image sequence

Milestones and Deliverables

Milestone	Description and Deliverable	Timeframe
(1) Equipment acquisition and protocol approval	Research protocol development and IRB clearance, instrumentation acquisition (XUUK eyebox2, camera IR conversion, pan-tilt tracking mechanism, NIR illuminators)	3 months
(2) Hardware prototyping	Hardware prototype construction	3 months
(3) Image processing	Develop algorithms for image quality assessment and segmentation	5 months
(4) Test	Hardware-software field test and calibration	1 month

Progress to Date

Month 1-3: Equipment acquisition and protocol approval:

Research protocol development and IRB clearance, instrumentation acquisition (XUUK eyebox2, camera IR conversion, pan-tilt tracking mechanism, NIR illuminators) - **DONE**

Month 4-6: Hardware prototyping: Hardware prototype construction - **DONE** (pan-tilt in progress)

Month 7-11: Image processing: Develop algorithms for image quality assessment and segmentation - **DONE**

Month 12: Test: Hardware-software field test and calibration - **IN PROGRESS**

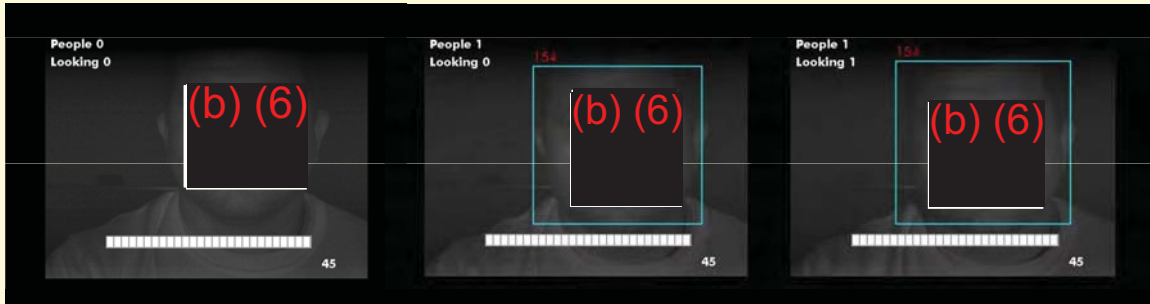
Project Directions

- Given the new budget line, one station (instead of two coordinated platforms) was built
- In the interest of faster progress towards stage 3, a phase 1 data collection was performed using stationary subjects and a stationary main camera (the XUUK-controlled, pant-tilt mounted camera adaptation is in progress. This dynamic configuration will be tested as a part of stage 4 and by the end of December)

Hardware Platform: Eye-finder Subsystem

XUUK™ eye via-eye effect, from up to 10 m

- Presence and location of any number of eyes
- Presence and location of any number of faces

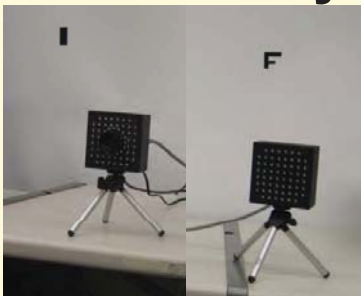


CITeR

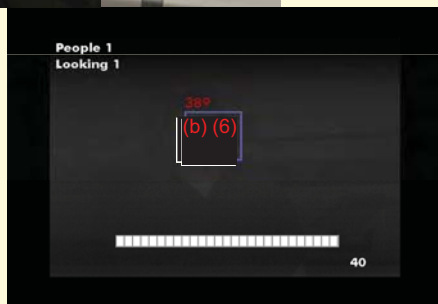
The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research

www.citer.wvu.edu

Hardware Platform: Eye-finder Subsystem



XUUK™ eye finder
Subsystem



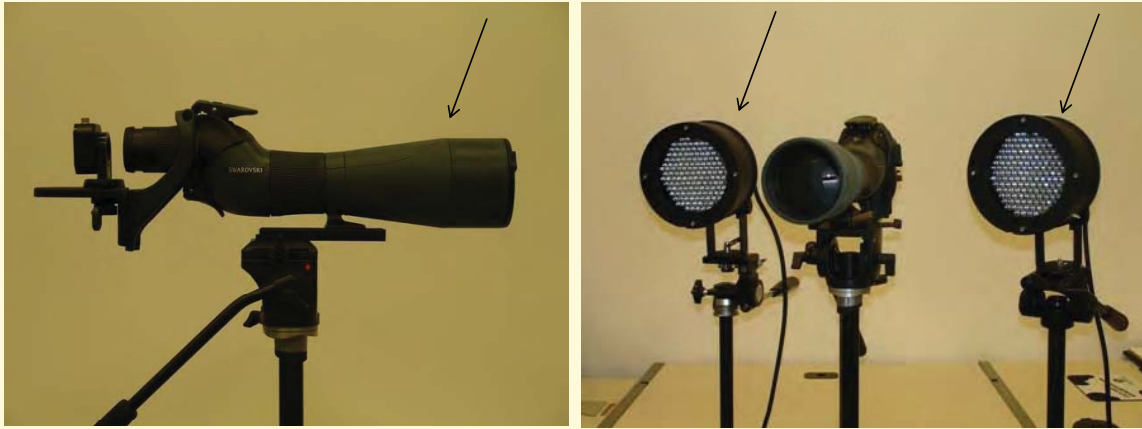
Above: XUUK locating the face and the eyes location from 7.5 m. Will be used to control the PT

CITeR

The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research

www.citer.wvu.edu

Hardware Platform: Main Optics and NIR Illuminators



Above: Main optical front end – lens, camera, and main NIR illuminators (850 nm). A Swarovski high-definition spotting scope was chosen after evaluation of different telephoto lenses and telescopes.

CITeR

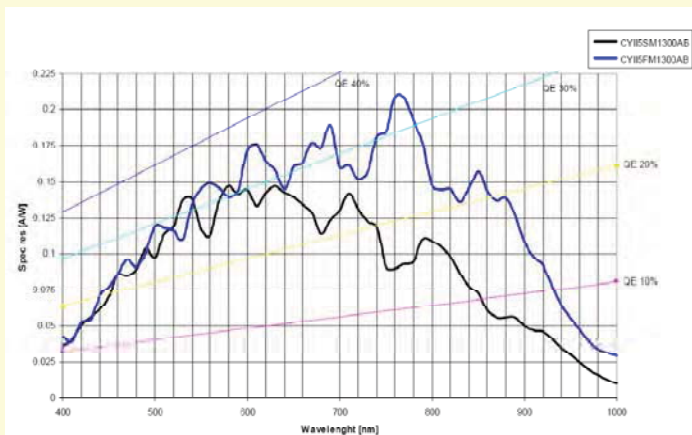
An NSF I/UCR Center advancing integrative biometrics research

The Center for Identification Technology Research

www.citer.wvu.edu

Hardware Platform: NIR CMOS Sensor

High speed, highly sensitivity, and
under full real-time software control



Spectral characteristics of SMX-150M Image Sensor

IBIS5-AE-1300 (blue)



Above: winner, SMX-150M CMOS Sensor

- Several cameras were tested
- NIR enhanced sensitivity
- Increased light collection quantum efficiency due to larger pixel size
- Monochromatic=all the pixels are used for a single wavelength
- High frame rate for real time NIR visualization

CITeR

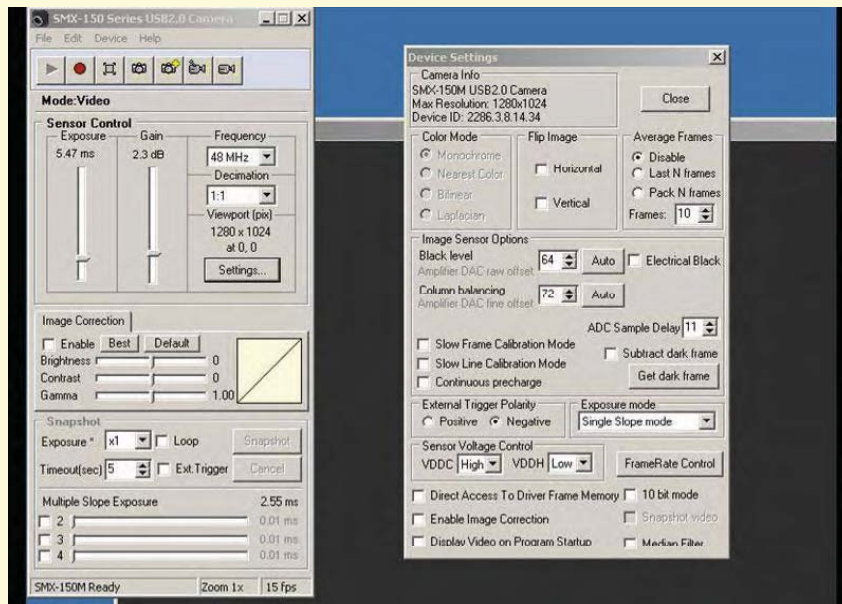
An NSF I/UCR Center advancing integrative biometrics research

The Center for Identification Technology Research

www.citer.wvu.edu

Hardware Platform: Camera Controllers

An example of the all electronic, real time control of the SMXC 150M NIR camera:



CITeR

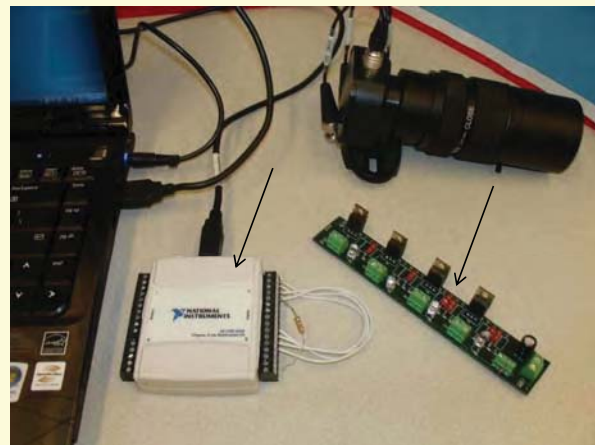
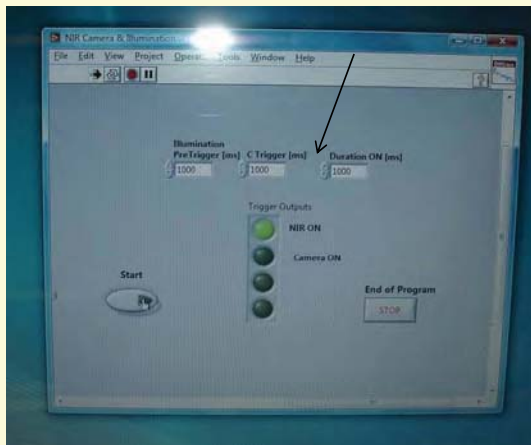
The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research

www.citer.wvu.edu

Hardware Platform: Illumination and Camera Controllers

In house master control
software (LabVIEW)

Hardware interfaces to control
camera and main NIR lights



Above: LabVIEW-based control of the NIR electronic switch and the main camera, via PC USB port.

CITeR

The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research

www.citer.wvu.edu

Data Collection

- **Trial 1: 14 Volunteers so far (i.e. early October)**
 - Challenging IRB process...
- **Multiple burst images from 0.75, , 6, 7, 8, and 9 meters**
- **Ages 21-51, of Asian and Caucasian descent, 11 males and 3 females**
- **Images of darker eyes were better in the NIR spectra (e.g. compared to light grey or blue)**
- **Background NIR for manual focus, synchronized electronic NIR “flash” for burst captures**

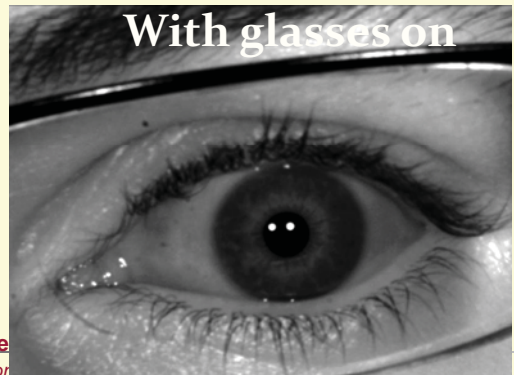
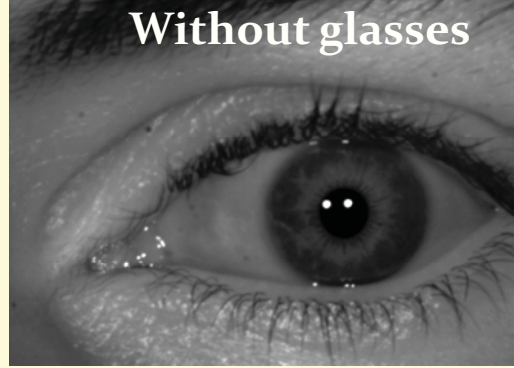
Sample Images

Burst mode capture for lucky imaging

Below: short range samples (enrollment)



Sample Images

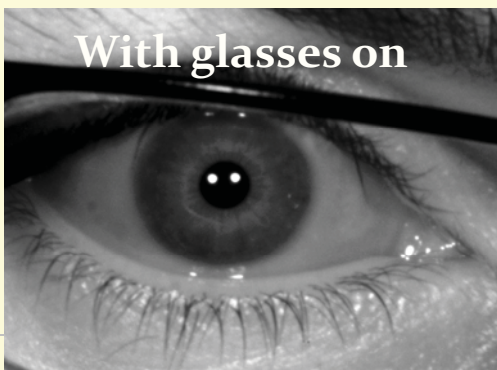


Images from 5 meters

CITeR

The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research

Sample Images

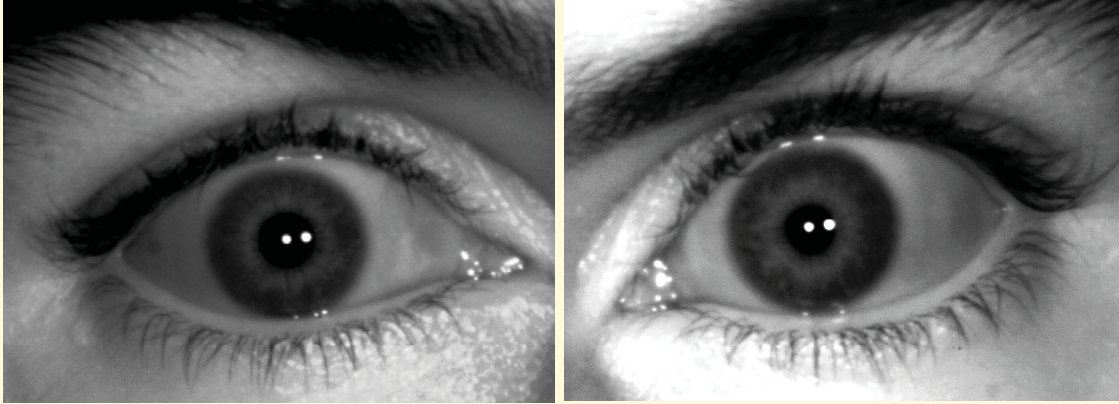


Images from 6 meters

The Center for Identification Technology Research
advancing biometrics research www.citer.wvu.edu

Sample Images

Images from 9 meters



Right eye

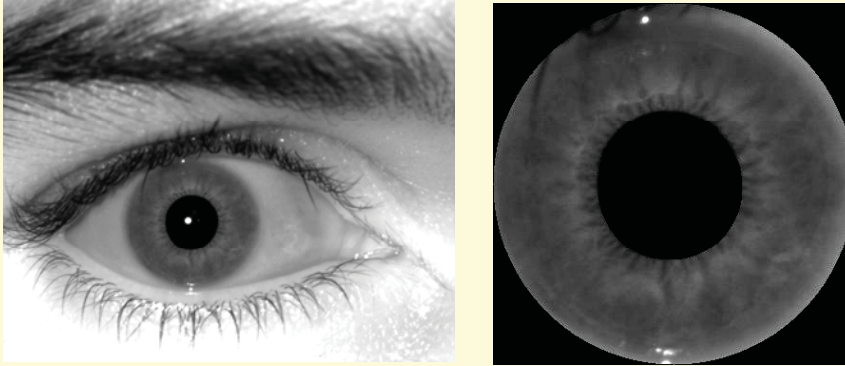
Left eye

Quality Assessment for Best Frame Selection

- A number of quality factors were implemented and used to select the best frames from a series of burst shots of each of 14 subjects, taken from distances of 1, 6, 7, 8, and 9 meters
- The frames were first segmented and the lower part of the iris used to compute the local quality measures for best frame selection
- Only correctly segmented frames were selected to compute the various quality factors
- One representative quality measure was selected to illustrate the use of quality measures for frame selection and show their variation with distance and with subjects at the same distance

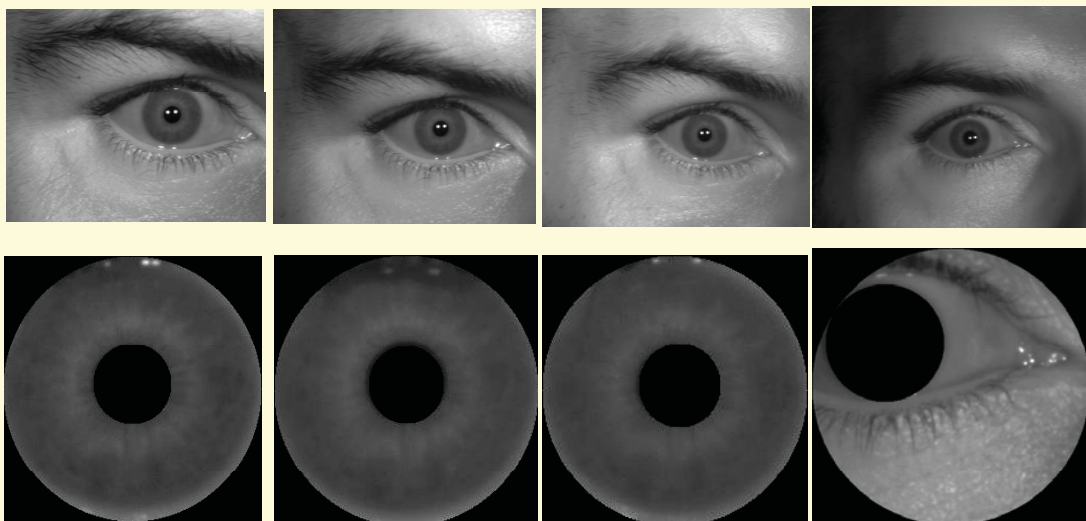
Segmentation Issues

- As the distance from camera to subject increases, the segmentation success rate decreases



Close up image (1m): segmentation performed very well.

Segmentation Issues



Sample images from 6, 7, 8, and 9m (left to right), and their respective segmentation results

Quality Factors Tested

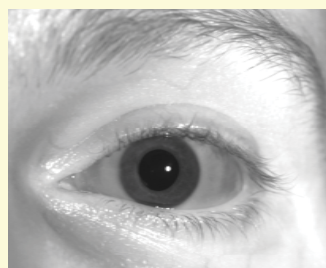
- Gradient-based measures (Tenengrad, adaptive, separable and non separable Tenengrad, Laplacian, Adaptive Laplacian)
- Correlation-based measures (autocorrelation function-single sample, area and height of central peak of the correlation function)
- Statistics-based measures (absolute central moment, grey level variance, Chebychev moments/ratios, entropy, histogram)
- Transform-based measures (Fourier transform: coefficients & magnitude, Cosine transform, multivariate kurtosis, Wavelets)
- Edge-based measures (step edge characteristics, transition width, local kurtosis)

Quality Assessment for Best Frame Selection

Frame	Quality measure
'SEQ_8.bmp'	345317.785504828
'SEQ_7.bmp'	325331.387383784
'SEQ_1.bmp'	324469.790982552
'SEQ_2.bmp'	323125.173731092
'SEQ_5.bmp'	322709.304476144
'SEQ_9.bmp'	319900.013367425
'SEQ_6.bmp'	318866.762008204
'SEQ_3.bmp'	309428.327550595
'SEQ_0.bmp'	300207.356147468
'SEQ_4.bmp'	294608.362633670
'SEQ_11.bmp'	251136.187098181
'SEQ_12.bmp'	248153.202589444
'SEQ_10.bmp'	245958.507221755
'SEQ_13.bmp'	245561.867981988
'SEQ_14.bmp'	241344.948234824

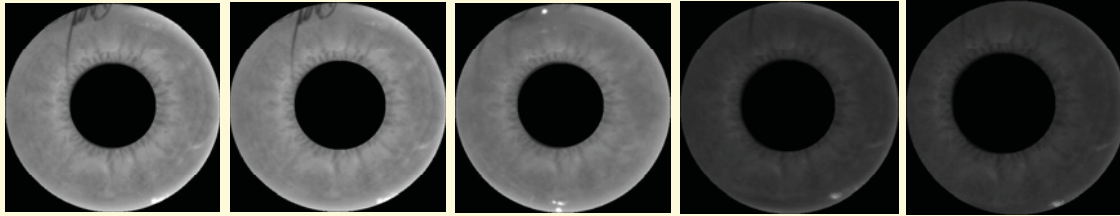


Frame 8



Frame 14

Best Frame Selection – at 1 and 6m



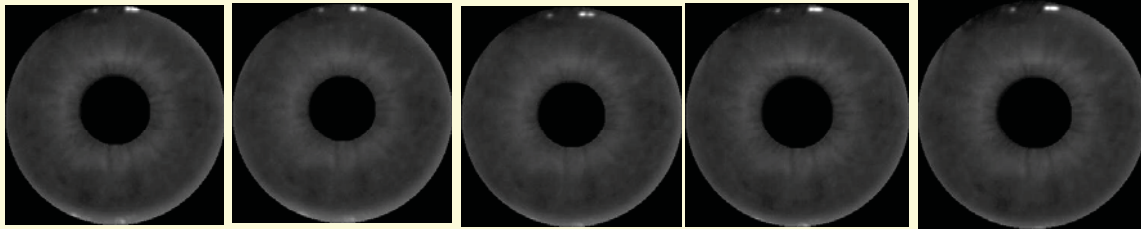
37e+04

35.9e+04

25e+04

5.9e+04

4.7e+04



11.3e+04

10e+04

9.3e+04

8.5e+04

8.2e+04

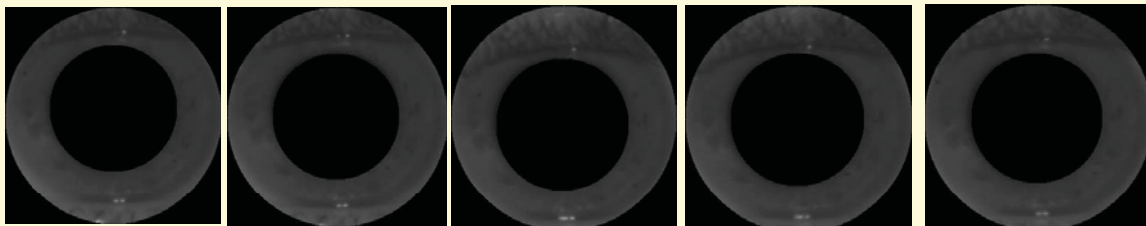
CITeR

The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research

The Center for Identification Technology Research 23

www.citer.wvu.edu

Best Frame Selection – at 7 and 8m



5.3e+04

5e+04

4.9e+04

4.8e+04

4.7e+04



5e+04

5e+04

4.99e+04

4.97e+04

4.94e+04

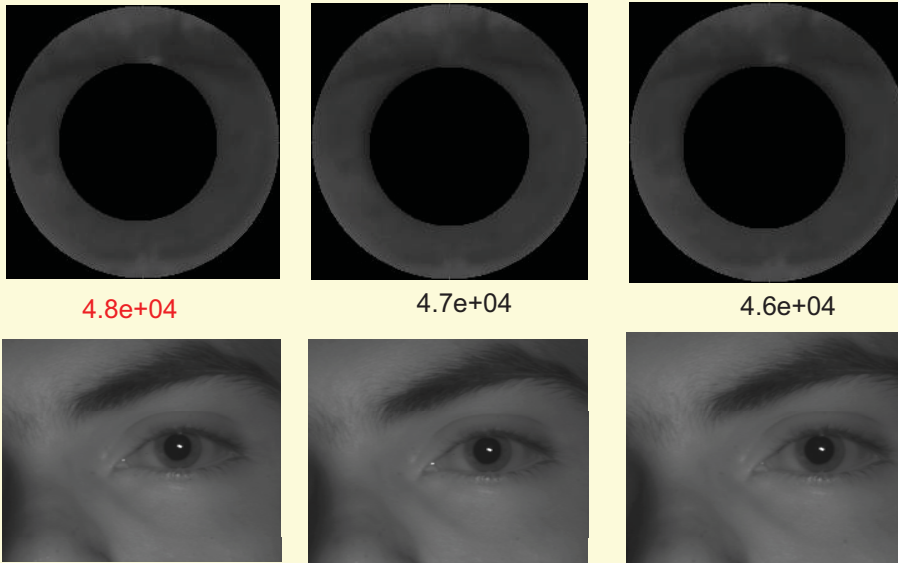
CITeR

The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research

The Center for Identification Technology Research 24

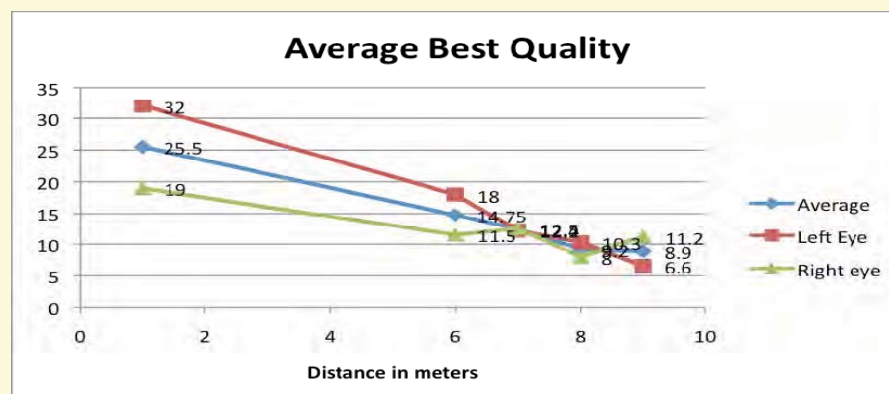
www.citer.wvu.edu

Best Frame Selection – at 9m



Average Quality Measure per Distance

Distance in m	Average quality	AQ-Left eye	AQ-Right eye
1	16.5	32	19
6	14.75	18	11.5
7	12.4	12.2	12.5
8	9.2	10.3	8
9	8.9	6.6	11.2



'SEQ_31.bmp'	68	'SEQ_0.bmp'	59	'SEQ_1.bmp'	52	'SEQ_13.bmp'	39
'SEQ_32.bmp'	68	'SEQ_5.bmp'	59	'SEQ_7.bmp'	52	'SEQ_18.bmp'	39
'SEQ_33.bmp'	64	'SEQ_10.bmp'	55	'SEQ_12.bmp'	48	'SEQ_19.bmp'	39
'SEQ_35.bmp'	64	'SEQ_12.bmp'	55	'SEQ_0.bmp'	42	'SEQ_14.bmp'	38
'SEQ_26.bmp'	63	'SEQ_17.bmp'	54	'SEQ_3.bmp'	42	'SEQ_15.bmp'	38
'SEQ_36.bmp'	63	'SEQ_18.bmp'	54	'SEQ_5.bmp'	42	'SEQ_1.bmp'	36
'SEQ_38.bmp'	63	'SEQ_3.bmp'	54	'SEQ_16.bmp'	41	'SEQ_11.bmp'	36
'SEQ_39.bmp'	63	'SEQ_2.bmp'	51	'SEQ_17.bmp'	41	'SEQ_12.bmp'	36
'SEQ_29.bmp'	62	'SEQ_14.bmp'	50	'SEQ_6.bmp'	41	'SEQ_3.bmp'	36
'SEQ_20.bmp'	61	'SEQ_10.bmp'	50	'SEQ_10.bmp'	40	'SEQ_4.bmp'	36
'SEQ_22.bmp'	61	'SEQ_4.bmp'	49	'SEQ_4.bmp'	40	'SEQ_17.bmp'	35
'SEQ_30.bmp'	61	'SEQ_6.bmp'	49	'SEQ_14.bmp'	39	'SEQ_10.bmp'	34
'SEQ_34.bmp'	61	'SEQ_8.bmp'	49	'SEQ_19.bmp'	39	'SEQ_5.bmp'	34
'SEQ_37.bmp'	61	'SEQ_11.bmp'	48	'SEQ_2.bmp'	38	'SEQ_7.bmp'	34
'SEQ_27.bmp'	59	'SEQ_19.bmp'	48	'SEQ_8.bmp'	38	'SEQ_0.bmp'	33
'SEQ_23.bmp'	58	'SEQ_7.bmp'	48	'SEQ_18.bmp'	37	'SEQ_16.bmp'	33
'SEQ_25.bmp'	58	'SEQ_9.bmp'	48	'SEQ_9.bmp'	37	'SEQ_6.bmp'	33
'SEQ_28.bmp'	56	'SEQ_11.bmp'	47	'SEQ_11.bmp'	35	'SEQ_8.bmp'	31
'SEQ_24.bmp'	55	'SEQ_16.bmp'	47	'SEQ_13.bmp'	35	'SEQ_9.bmp'	30
'SEQ_21.bmp'	54	'SEQ_13.bmp'	46	'SEQ_15.bmp'	35	'SEQ_2.bmp'	27
6m		7m		8m		9m	

Same subject at Various distances

CITeR

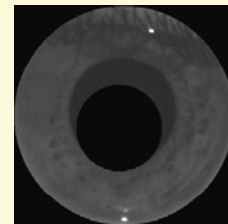
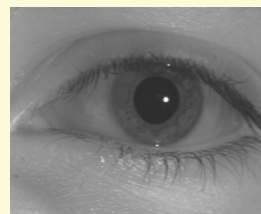
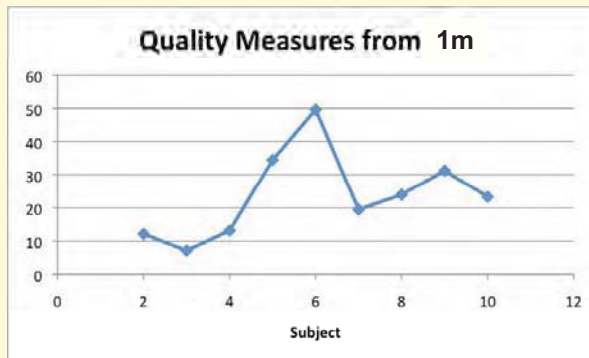
An NSF I/UCR Center advancing integrative biometrics research

The Center for Identification Technology Research

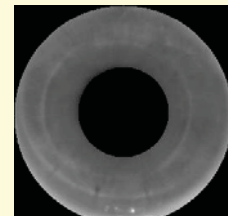
www.citer.wvu.edu

27

Best Quality Measure per Subject



Subject #3, QM = 7e+04



Subject #6, QM = 49e+04

CITeR

An NSF I/UCR Center advancing integrative biometrics research

The Center for Identification Technology Research

www.citer.wvu.edu

28

Conclusions-Software

- We implemented segmentation code and 8 different quality measures for iris images
- We applied the 8 different quality measures to all correctly segmented iris images
- We selected the best frame for each of 14 subjects at 5 different distances varying between 1 and 9m
- Each subject at each given distance has from 3 to 5 bursts or sequences of 5 images each
- Future work will involve implementing a robust segmentation algorithm that would work on images acquired from different distances, where the iris is at different scales, resolutions, locations in the images and at various degradation levels
- A voting system will also be implemented for the selection of the best frame and before and after matching scores compared

CITeR

The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research

www.citer.wvu.edu



Conclusions - Hardware

- XUUK system can be successfully used to locate distant face and eyes
- The embedded computing power is critical for the real time performance of the discovery and tracking algorithm
- Image sensor has to be monochromatic with high quantum collection efficiency at the wavelength of illumination (large pixel size), have high frame rate, and electronically controlled settings
- Optics have to have high magnification and quality (e.g. low geometric distortion, fast, highly transmissive at NIR, large aperture and depth of field)
- Optical front-end needs to be adapted for electronic focus (including a fast and accurate assessment of subject distance)
- The sensor was synchronized with the NIR illumination sources
- PT system has to have the ability to address reference coordinates with fast vector movement and motion stabilization
- Alternative and non-traditional PT (arc mounted) could be better suited for this application

CITeR

The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research

www.citer.wvu.edu



Collaborative acquisition of face images and real time face recognition using camera sensor networks



Vinod Kulathumani, Arun Ross, Bojan Cukic

Students: Srikanth Parupati, Raghavendra Jillela

Jan 1 2009 – Dec 31 2009

Project Report Oct 2009



CITeR **The Center for Identification Technology Research**
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu



Milestones and Deliverables

Milestone	Description and Deliverable	Timeframe
Assembling smart camera unit	Prototype deployment and test unit, comprising a set of 4 embedded smart cameras	4 months
Network Coverage	Positioning of cameras to maximize biometric content	2 months
Embedded and network system development	Software development on embedded cameras to take pictures of object of interest, extract relevant features and for using the partial snapshots from the cameras to construct full facial image for recognition	4 months
Performance analysis	Robustness of our algorithms in the presence of camera failures or camera views getting blocked Evaluating the improvement in accuracy / confidence of our face recognition by the use of a network of smart cameras	2 months



CITeR **The Center for Identification Technology Research**
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu



Project Goal

- **Face recognition using *network of embedded smart cameras***
 - Capable of local embedded processing
 - Uses multiple camera inputs
 - Intelligence within the network
 - Suppression of required bandwidth
 - Reduce workload at fusion center
- **Design and evaluation of a prototype**

CITeR An NSF I/UCR Center advancing integrative biometrics research **The Center for Identification Technology Research** www.citer.wvu.edu





Related work

We will compare with

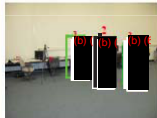
- Existing work on multi-camera networks for surveillance
- Existing work on video analytic systems

CITeR An NSF I/UCR Center advancing integrative biometrics research **The Center for Identification Technology Research** www.citer.wvu.edu

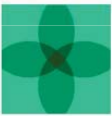


1. Centralized camera networks

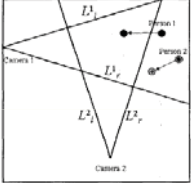
- Rely on transmitting all data to a central unit
- All processing at central unit
- Used extensively for tracking and surveillance
- Some specific focus areas have been:
 - (observe mode) Multi-camera, single and multi-person tracking
 - Camera handoffs, overlapping FOVs
 - Handling occlusions
 - Positioning for persistent surveillance




Chen et al. AVSS 08




Uniform and sufficient overlap
Persistent surveillance, Abidi et al. CVPR 08





Handoffs, Javed et al '00



The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research




www.citer.wvu.edu



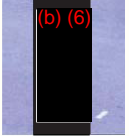
1. Centralized camera networks

- Rely on transmitting all data to a central unit
- All processing at central unit
- Used extensively for tracking and surveillance
- Some specific focus areas have been:
 - (control mode) Centrally controlled multiple PTZ and static cameras
 - Optimal close-ups for biometrics, surveillance
 - Master slave configuration
 - Optimal tasking [Krahnstoever et al, GE]




Master view


Slave view





Zhou et al, DHID 03



The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research





www.citer.wvu.edu

Centralized network architecture


- **When processing involves fusion from multiple cameras**
 - Scalability issues when number of cameras increase
 - **Processing bottleneck**
 - **Communication bottleneck**
 - Often requires tight calibration

CITeR **The Center for Identification Technology Research**
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu

2. Video analytics



- **Lot of existing work focus on video analytic systems**
- **E.g. cameras to deploy on vehicles, highways**
- **Processing**
 - Completely local
 - Code specific for a given task
 - Computationally intensive
- **Extensive systems**



Detecting illegally stopped cars
[Agent-vi technologies]

- **No networking and collaboration of data**

CITeR **The Center for Identification Technology Research**
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu





Distributed, embedded smart camera networks

- **Combines**
 - Local processing
 - Inputs from multiple units
 - Centralized fusion
- **Mode of operation**
 - Each unit extracts “features” relevant for particular application
 - Groups formed within network that fuse these features

And / Or

 - Features fused at central unit

CITeR An NSF I/UCR Center advancing integrative biometrics research **The Center for Identification Technology Research** www.citer.wvu.edu



Distributed, embedded smart camera networks

- **Advantages**
 - By locally extracting features, reduces communication bottleneck
 - Permits use of wireless and low power radios
 - Reduces processing bottleneck at base
 - Robust to individual camera failures, views, occlusions
 - For face recognition, we gain robustness wrt pose, illumination variances as well
 - Individual unit cheap, => , permitting large scale fine grained coverage
 - Portable and easy to deploy, can enable covert operations



CITeR An NSF I/UCR Center advancing integrative biometrics research **The Center for Identification Technology Research** www.citer.wvu.edu

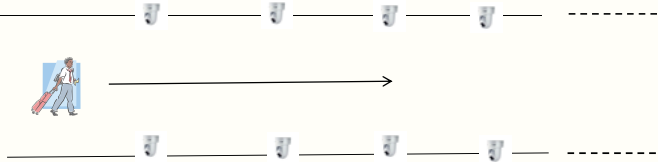
This project

Can we use this for real time face recognition ?



CITeR **The Center for Identification Technology Research**
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu

Specific problem statement


- **Given a long linear network of smart cameras**

- **Person walking through network**
- **Ensure barrier coverage**
 - Person accurately recognized using images acquired within network
 - Without overwhelming a central unit
 - Pose, illumination, resolution vary as person walks through different camera FOVs
- **Longer the barrier, greater the probability of recognition**
- **We consider a network of 7 cameras over 40 feet**

CITeR **The Center for Identification Technology Research**
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu





Outline

- **Assembly of smart camera platform**
- **Assembly of smart camera network**
- **Face recognition setup**
- **Results**
- **Extensions**




The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research

www.citer.wvu.edu



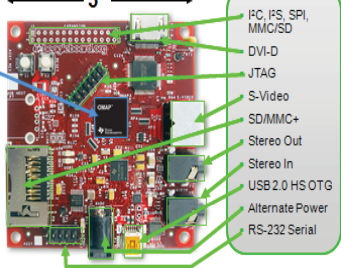
Assembly of smart camera unit



Laptop-like performance

- TI OMAP3530
- 600 MHz superscaler ARM® Cortex™-A8
- More than 1200 Dhrystone MIPS
- Up to 10 Million polygons per sec graphics
- HD video capable C64x+™ DSP core
- Memory**
- 128MB LPDDR RAM
- 256MB NAND flash


← 3" →



Flexible expansion



- PC, I2S, SPI, MMC/SD
- DVI-D
- JTAG
- S-Video
- SD/MMC+
- Stereo Out
- Stereo In
- USB 2.0 HS OTG
- Alternate Power
- RS-232 Serial

Beagleboard, camera, wireless card



The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research



www.citer.wvu.edu

Assembly of smart camera unit


- **Components**
 - TI OMAP chipset along with a c64+ DSP (\$149 + peripherals worth \$75)
 - 128 MB RAM, 600MHz processor
 - Attached to USB web camera [Logitech 9000]
 - Capable of image processing [faster than smartphone platforms]
 - Portable and low cost solution
- **Linux based development**
- **Ported OpenCV to this platform**

CITeR **The Center for Identification Technology Research**
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu

Can attach other cameras






- **Via USB or Ethernet**
- **Example: a SONY PTZ**
- **Can control parameters locally**



CITeR **The Center for Identification Technology Research**
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu

Other platforms considered

- **Intel imote2 with multimedia board**
 - Poor image resolution, frame-rate of capture
 - No extension to DSP
- **PDA phones**
 - Already have cameras embedded
 - Hp Ipaq 910
 - No access to camera API
 - Nokia N95
 - Code verification
 - Certification for installation
 - Google G1 with Android
 - Easy development, good camera
 - Currently no DSP extension,
 - Cannot optimize for specific application



CITeR The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu

Smart camera network assembly

- Each unit portable and easily configurable into a network
- Assembled a smart camera network testbed – *Hawk-eye*



A ceiling setup of cameras



Other components on table



6 feet setup

CITeR The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu

Hawk-eye test-bed features

- **Portable and easy to set up**
- **Indoor or outdoor _can run on 5V batter_ _**
- **Wireless reprogramming**
 - Can download and change programs on individual nodes, wirelessly
- **Single hop wireless data transfer to base unit**

CITeR The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu

Network deployment for face recognition

Schematics

- **Cameras placed 7 feet above ground**
- **Angled slightly downwards facing exit**
- **No tight calibration**

9 ft

ENTRANCE



EXIT



EXIT



40 ft

CPU



CITeR The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu

 **Actual setup** 

 Network view 



 Close-up of cameras 

CITeR **The Center for Identification Technology Research**
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu

 **Operation on each smart camera** 

- **Capture frames continuously [960 by 720]**
- **Perform background subtraction to detect any movement**
 - Suppresses transmission of non event frames
 - Reduces data input to face detector component
- **If event detected, run a face detector algorithm**
 - Haar cascade face detector
 - Minimum face size specified as 22 by 22
 - *By virtue of training data*, filters out faces if
 - Too small [we specify minimum sizes]
 - Poor pose [training data only for frontal faces]
 - Poor illumination [faces not detected by classifier if too dark]
- **If face detected**
 - extract only part of image containing face and send wirelessly to base



CITeR **The Center for Identification Technology Research**
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu



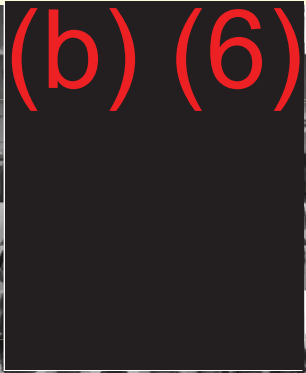
Experiment details

- **29 subjects walking through network**
- **Speed: 40 feet in 10 seconds: 2.75 mph**
- **Cameras extract filtered probe images, transmit to base**
- **Each subject has 5 gallery images in database**
 - Taken under good illumination, close-up view
 - *Probes will not be of same quality*
- **Database contains gallery of 100 individuals**
 - 29 subject gallery images mixed with 79 images from WVU's multibiometric database
 - 5 images per individual
- **Each probe scored against each gallery image**
 - Software used: VeriLook
- **Scores fused across probes to generate match**

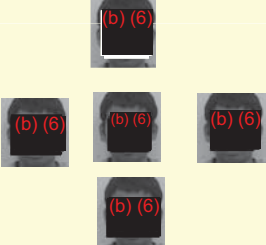
CITeR The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu



Gallery images



*Subject gallery images mixed with
WVU's Multibiometric database images*





Specific subject gallery images

CITeR The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu



Video

CITeR **The Center for Identification Technology Research**
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu



Recognition technique

- **Fusion algorithm**
 - Scores generated for each probes against each gallery image
 - Generate score matrix
 - For a given subject
 - Fuse match scores (max-rule) when all individual probes are compared against each gallery identity
 - The gallery identity resulting in the best score is deemed to be the identity of the probe

CITeR **The Center for Identification Technology Research**
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu

Recognition - illustration

	Subject 1	Subject 2
Gallery	(b) (6)	(b) (6)
(b) (6)	max	max
(b) (6)	MAX	MAX
(b) (6)	MAX	max

CITeR The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu

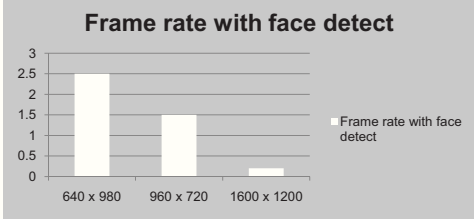
Network performance

- **Processing speed**
 - Camera initialization 100 ms
 - Frame capture time [960 x 720] 35 ms
 - Background subtraction 78 ms
 - Face detection
 - [image scaled to 480 x 360]
 - On entire image 1.9 s
 - On segmented image 470 ms
- **Average frame rate**
 - No back_ round chan_e 8 f_s
 - Running face detection 1.5 fps
- **Network latency – 80 ms for each extracted image**
- **Bandwidth saved on transmitted frames – 90%**

CITeR The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu

Impact of resolution

- **Frame rate versus image resolution**



Resolution	Frame rate with face detect
640 x 980	~2.5
960 x 720	~1.5
1600 x 1200	~0.2

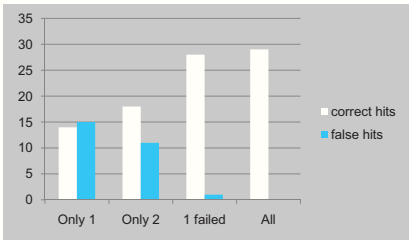
- Lower resolution =>
 - more frame rate & more probes but poorer quality
- We choose 960 by 720

CITeR The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu

Accuracy analysis

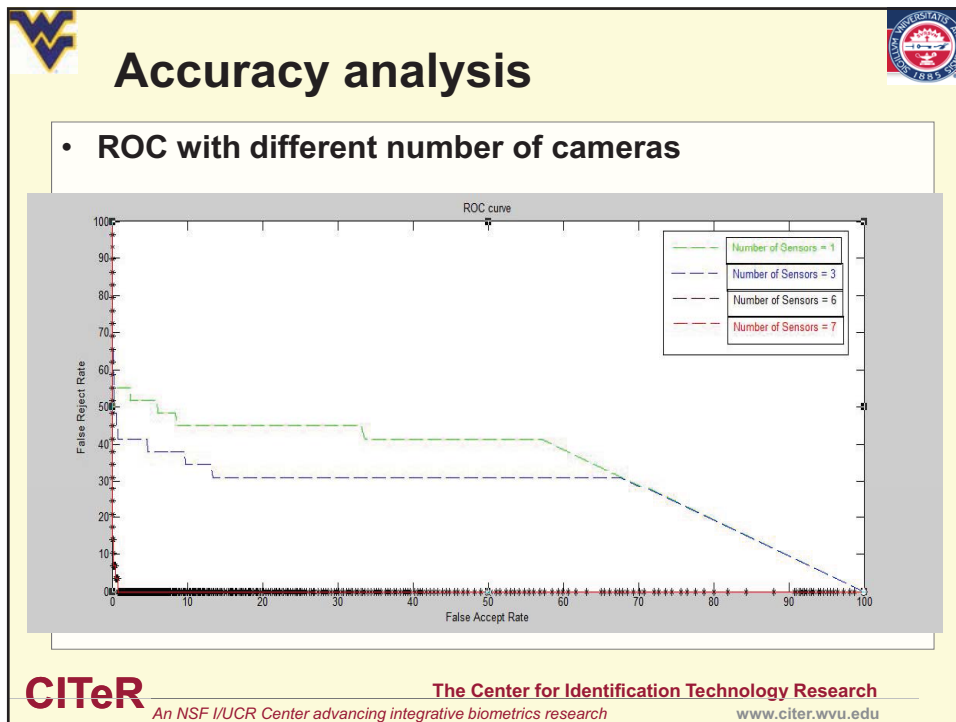
- **All probes with max based fusion**

- Probability of at least one camera able to get a “successful” probe: *very high*
- Rank 1 evaluation – *only top match is considered*



Number of sensors considered	Correct hits	False hits
Only 1	~15	~15
Only 2	~18	~10
1 failed	~28	~1
All	~29	~1



CITeR The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu



Results

- **Assembly of smart camera testbed**
 - Portable and reprogrammable
 - Permits local processing
 - Reusable for many applications and extensions
- **Prototype of distributed face recognition system**
 - Considered barrier coverage scenario
 - Locally selects potential good images
 - Max based fusion produces high accuracy
 - Reduces communication bandwidth by 90%
 - Reduces CPU bottleneck permitting real time recognition
 - Only 8-10 probes transmitted to base
 - Can be reduced further is nodes exchange “quality” data and reduce to 3-4 probes
 - Tolerates failures of individual cameras


CITeR The Center for Identification Technology Research
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu



Extensions

Multi-modal human identification
Combine with soft biometrics

Multi-layer sensor network
More energy efficiency




Integration with active control locally
Increase coverage with fewer cameras

Smart camera system

In-network face matching

Extend central fusion techniques
Use partial views, image repair, marks



Integrate with mobile robots
*Self-deploy,
Coordinated search*

Surveillance in general
*Spatiotemporal event detection
Detecting intrusion onto PRT tracks
Urban monitoring*

CITeR **The Center for Identification Technology Research**
An NSF I/UCR Center advancing integrative biometrics research www.citer.wvu.edu

3D Scanning for Biometric Identification and Verification

Anselmo Lastra, Henry Fuchs, Greg Welch
University of North Carolina at Chapel Hill

(b) (6)

Ali Farsaie
SIS, Incorporated

(b) (6)

Reliable and robust identification and verification of individuals is critical to homeland security applications such as surveillance, authorization for entry to secure areas, and passport identity verification. Traditional biometrics, such as mug shots, fingerprints, and voice recognition, have been used with some success. However, they exhibit serious disadvantages for some tasks. These three biometrics, for example, are problematic for surveillance (identification); even the traditional mug shot is difficult to use in automated surveillance applications because many factors, such as lighting and frontal visibility, cannot be controlled.

A relatively new biometric, 3D facial recognition, holds great promise. Even though the technology is nascent, recognition performance using 3D shape and texture matched that of the much more mature high-resolution image recognition (which featured controlled lighting) and iris recognition in a comprehensive 2006 study [1]. Additionally, 3D modeling promises to enhance recognition performance because it can be used to recognize people in profile as opposed to a typical forward-looking, mug-shot pose. Even when using 3D to match to a mug shot, an advantage is that a 3D model allows one to render a view of the person from any desired perspective—the pose, distance, and even lighting can be factored into the rendering to match any known/reference photos.

Scenarios in which 3D recognition could be profitably used include (a) verification of identity at an airport, for example the subject's face could be rapidly scanned while his or her smart-card ID is being examined; the system could then match the scan with data on the ID; (b) identification at a secure site or even at an airport while people are walking down the hallways or standing in line; (c) 3D pose extraction of a moving subject, thereby potentially enhancing recognition performance and enabling intent analysis.

In this brief, we present the technical background of the 3D scanning technologies, briefly survey related biometrics that may be combined with 3D recognition, provide an overview of the major technical issues and highlight research opportunities to overcome those issues.

Background

Probably the most studied technology for 3D modeling is baseline stereo vision. The idea is to image the scene with two (or perhaps more) cameras, and select corresponding points in the two images. If the cameras are calibrated (camera position and orientation, as well as lens and imager characteristics), the correspondences can be used via triangulation to determine the distance, and thus the geometry of the visible structures in the scene.

A major problem is to determine the correspondences. A scene with very uniform color, such as white walls, is clearly problematic. If the scene is highly textured, then correspondence points may be extracted automatically. The techniques fall into two categories, intensity matching and feature detection [2-3], with the latter having proven more reliable. Stereo reconstruction may also be performed from video sequences [4-6]. The problem of accurately finding correspondences, however, has proven to be difficult and not always robust, leading researchers to investigate active approaches. The major active approaches are laser scanning and structured light.

Laser scanning, when used for faces, human bodies or other objects at short distances, typically utilizes a triangulation method. A laser stripe scanned across the subject essentially provides correspondences for the camera(s). A well known laser scanner of this type, which has been extensively used in the movie industry, is made by Cyberware. A problem is that this scanning technique takes seconds to minutes; not a problem for scanning a seated and supported actor's face, but prohibitively long for identification purposes. Some laser techniques project complex patterns using interference of two beams. This is essentially structured light (see below).

Another laser-scanning technique uses time of flight (the time for illumination to travel to and from a surface, divided by the speed of light) to determine distance. This is also known as LIDAR. Typically this method is used for longer ranges. Some new devices, such as the Swiss Ranger [7] and Canesta [8] cameras work at ranges of a few meters, and at video rates. However, the low resolution, such as 160 x 120 (Canesta) or 176 x 144 (Swiss Ranger) pixels, makes them unsuitable for our purposes. The marketing focus for these devices seems to be in vehicle safety applications (backup alarms for cars, for example) and human-computer interaction (potentially for video games).

The second general approach, *structured light*, is very similar to laser triangulation except that a light projector is typically used to project a pattern onto the subject. This provides a rich field of correspondences all across the subject that can be used to extract a 3D model from the camera images. The use of time-multiplexed coded structured light patterns was first proposed by Posdamer and Altschuler [9], and has sparked a great deal of research. Typically a small number of patterns is projected in sequence and the result imaged. Monochrome cameras can be used to capture geometry, and a color camera to add texture. This is the technology used by the 3D

Snapshot system from SIS. Below we will focus on structured light, as it is the most suitable for human-subject scanning, and examine the challenges as well as possible research directions.

Issues

- The process should not disturb the subject. A major problem with conventional structured light approaches is that the rapidly flashing patterns are uncomfortable for the people being scanned. We can also foresee situations in which it would be important to scan a subject without his or her knowledge.
- Speed of capture is critical for any moving subject, especially for human biometrics. Many systems take less than a second (0.3 sec. for 3D Snapshot) to scan, but humans move significantly in that time. Ideally we'd be able to scan in $1/10^{\text{th}}$ to $1/30^{\text{th}}$ of a second.
- Speed of processing is also important. The result must be available within a second or two. Ideally, the processing could be done at real-time rates in order to generate 3D at video rates.
- Accuracy is a major issue, of course, especially under less than ideal lighting and environmental conditions.
- The scanner should have a reasonably wide field of view so the subject does not have to be in a very precise location. Analogously, the scanning device should have reasonable depth of field.
- Eyeglasses are a problem because of reflections from, and refraction through the lenses.
- Geometry of hair can be difficult to capture, and a beard can also be used to hide features.

Research Directions

In this section we propose research directions, in priority order. The ordering is based on the importance of the problem to be solved, as well as the amount of time we expect a technical solution to take.

Imperceptible Scanning. We see two fruitful technical directions to make the scanning process invisible to the subject. The first, *imperceptible structured light*, was invented at UNC Chapel Hill [10] to enable 3D modeling of persons for 3D video conferencing applications. The key idea of imperceptible structured light is that we can flash a pattern and its inverse rapidly enough so that it will appear to the subject as white light. A fast camera can be synchronized to the projector and will capture an image of the pattern. Most of the work in this area, ours and others, has been to calibrate projector systems shining on non-planar environments [11-13]. Although we have demonstrated the concept, many challenges remain with the hardware implementation.

The other potential approach is to use infrared illumination. Infrared may be imaged directly (essentially to detect skin temperature) [14-15], or we can project infrared patterns, much as we do visible light. There has been little work on infrared structured light. We only know of a bench prototype tested in Japan [16].

Speed. Two factors account for the time, acquisition and processing. Carefully synchronizing the camera with the projector, such as we've done with our prototypes [10-12], can make the image acquisition process faster. However, imaging in a shorter amount of time, or with less light, tends to make sensor noise more problematic, and we need to work to combat this using techniques such as those of Bennett and McMillan [17]. To make the processing faster, we can use the graphics processing unit, an approach, which we pioneered [18], that is now becoming popular. Speedups of 20 to 40 times are possible.

Improved Biometric Accuracy. It is possible to combine multiple biometrics, with the resulting biometric fusion potentially increasing accuracy. A promising approach may be to combine iris/retinal scanning with 3D scanning. The texture of the iris forms during the human's gestational period, and it exhibits a great deal of detail, including furrows, freckles, etc. [19]. The iris can be imaged unobtrusively, and the near infrared modality used brings out patterns even in person whose iris pigmentation is dark. Imaging of the iris requires cooperation from the subject, therefore may be less useful for identification from surveillance imagery [14]. A survey of techniques is presented in [20].

Field of View and Depth of Field. The ability to capture 3D models of people over a wide working area will provide a very powerful biometric tool. This is a very difficult problem, however. For the hardware part of the solution, we would propose overlapping, synchronized structured light projectors, and a set of cameras. Note that the prices for both of these devices is dropping very rapidly, so cost is not the primary barrier.

This net of projectors and cameras could be coupled with software algorithms for a progressive refinement of the biometric over time. For example, the scanning might occur as people are standing at the line waiting for the TSA screening. Even if there is no wait, just the walk through the cordoned area could serve.

A potentially powerful strategy is to combine structured light approaches with extraction of correspondences for a combined modeling approach. We can also take advantage of the fact that we are observing the people for a longer time to improve the models by predicting the subject's motion and tailoring the imperceptible structured-light patterns to improve the model.

Extraction of Subject Pose and Posture. The way in which a person walks is a very characteristic identifier. We can often recognize people in this way. Furthermore, pose and posture analysis could be used to analyze intent in certain situations. We have been working with the Navy to estimate the posture of Marines during training, and using the posture to analyze their performance. This work is done outdoors using multiple video cameras because a multitude of views is necessary. Structured light would be a very useful enhancement (not possible for the Marine-training scenario).

Project Contact: Anselmo Lastra

(b) (6)
(b) (6)
(b) (6)
(b) (6)

Bios

Anselmo Lastra is a Professor and Chairman of Computer Science at the University of North Carolina at Chapel Hill. He received a BS in Electrical Engineering from the Georgia Institute of Technology, and MS and PhD degrees in Computer Science from Duke University. He was co-chair of I3D 2005 and Graphics Hardware 2000 & 2004, and is an associate editor of IEEE Computer Graphics and Applications. His research interests are in the areas of image-based 3D modeling and rendering, and graphics hardware architectures.

Henry Fuchs is the Federico Gil Professor of Computer Science at UNC Chapel Hill. He has been active in computer graphics since the early 1970s, with rendering algorithms (BSP Trees), hardware (Pixel-Planes and PixelFlow), virtual environments, tele-immersion systems and medical applications. He is a member of the National Academy of Engineering, a fellow of the American Academy of Arts and Sciences, the recipient of the 1992 ACM-SIGGRAPH Achievement Award, the 1992 Academic Award of the National Computer Graphics Association, and the 1997 Satava Award of the Medicine Meets Virtual Reality Conference.

Greg Welch is a Research Associate Professor of Computer Science at the University of North Carolina at Chapel Hill. In 1986 he received a degree in Electrical Technology from Purdue University (with Highest Distinction), and in 1996 he received a Ph.D. in Computer Science from UNC-Chapel Hill. Previously he has worked at NASA's Jet Propulsion Laboratory and Northrop-Grumman's Defense Systems Division. His research interests include human tracking systems, 3D telepresence, projector-based graphics, computer vision and view synthesis, and medical applications of computers. He has co-authored over 50 peer-reviewed publications in these areas, is a co-inventor on multiple patents, maintains an internationally-recognized web site dedicated to the Kalman filter. He is a member of the IEEE Computer Society and ACM.

Ali Farsaie is President and CEO of Spatial Integrated Systems, Inc. (SIS). He provides strategic guidance and development for new program activities, within the Defense, Aerospace, Federal agencies, and commercial activities. Governor Perdue recently appointed him to North Carolina's Economic Development Board. Dr. Farsaie received M.S. and Ph.D. degrees in 1979 from North Carolina State University. Prior to SIS, Dr Farsaie was a chief engineer at the Naval Surface Warfare Center Dahlgren Division. He conducted, formulated and managed research and development of novel approaches in advanced information technology, sensors and robotics, training and system engineering to meet long term Navy mission requirements.

References

1. Phillips, P.J., W.T. Scruggs, A.J. O'Toole, P.J. Flynn, K.W. Bowyer, et al., *FRVT 2006 and ICE 2006 Large-Scale Results*. 2007, National Institute of Standards and Technology: Gaithersburg, MD. p. 56.
2. Faugeras, O., *Three-Dimensional Computer Vision -- A Geometric Viewpoint*. 1993, Cambridge: The MIT Press.
3. Forsyth, D.A. and J. Ponce, *Computer Vision: A Modern Approach*. 2002: Prentice Hall.
4. Pollefeys, M. and L.V. Gool, *From Images to 3D Models*. Communications of the ACM, 2002. **45**(7): p. 50-55.
5. Pollefeys, M., L.V. Gool, M. Vergauwen, F. Verbiest, K. Cornelis, et al., *Visual modeling with a hand-held camera*. International Journal of Computer Vision, 2004. **59**(3): p. 207-232.
6. Tomasi, C. and T. Kanade, *Shape and motion from image streams under orthography: a factorization method* International Journal of Computer Vision, 1992. **9**(2): p. 137-154.
7. Mesa Imaging, <http://www.mesa-imaging.ch>.
8. Canesta, <http://canesta.com>.
9. Posdamer, J.L. and M.D. Altschuler, *Surface measurement by space-encoded projected beam systems*. Computer Graphics and Image Processing, 1982.
10. Raskar, R., G. Welch, M. Cutts, A. Lake, L. Stesin, et al. *The Office of the Future: A Unified Approach to Image-Based Modeling and Spatially Immersive Displays*. in *SIGGRAPH98*. 1998. Orlando, FL.
11. Cotting, D., M. Naef, M. Gross, and H. Fuchs. *Embedding Imperceptible Patterns into Projected Images for Simultaneous Acquisition and Display*. in *ISMAR*. 2004.
12. Cotting, D., R. Ziegler, M. Gross, and H. Fuchs. *Adaptive Instant Displays: Continuously Calibrated Projections using Per-Pixel Light Control*. in *Eurographics*. 2005.
13. Zollmann, S. and O. Bimber. *Imperceptible Calibration for Radiometric Compensation*. in *Eurographics*. 2007.
14. Abayowa, B.O., *Thermal infrared exploitation for 3D face reconstruction*. Proceedings of SPIE - The International Society for Optical Engineering, 2009. **SPIE-7347**.
15. Colantonio, S. and M. Benvenuti, *Object tracking in a stereo and infrared vision system*. Infrared Physics & Technology, 2007. **49**(3): p. 266-271.
16. Akasak, K., R. Sagawa, and Y. Yagi. *A Sensor for Simultaneously Capturing Texture and Shape by Projecting Structured Infrared Light*. in *3DIM*. 2007.
17. Bennett, E.P. and L. McMillan, *Video enhancement using per-pixel virtual exposures*, in *ACM SIGGRAPH*. 2005. p. 845-852.
18. Harris, M.J., G. Coombe, T. Scheuermann, and A. Lastra. *Physically-Based Visual Simulation on Graphics Hardware*. in *SIGGRAPH / Eurographics Workshop on Graphics Hardware*. 2002.
19. Daugman, J., *How Iris Recognition Works*. IEEE Transactions on Circuits and Systems for Video Technology, 2004. **14**(1): p. 21-30.
20. Bowyer, K.W., K. Hollingsworth, and P.J. Flynn, *Image understanding for iris biometrics: A survey*. Computer Vision and Image Understanding, 2008. **110**(2): p. 281-307.

Registry of USG Recommended Biometric Standards

Version 2.0
August 10, 2009

**NSTC Subcommittee on
Biometrics and Identity Management**

1. Introduction

This *Registry of USG Recommended Biometric Standards* (Registry) supplements the [NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards](#), which was developed through a collaborative, interagency process within the Subcommittee on Biometrics and Identity Management and approved by the NSTC Committee on Technology. This Registry is based upon interagency consensus on biometric standards required to enable the interoperability of various Federal biometric applications, and to guide Federal agencies as they develop and implement related biometric programs.

The Subcommittee's standards and conformity assessment working group is tasked to develop and update the Registry as necessary. The Subcommittee will continuously review the content of this document, and release updated versions as required to assist agencies in the implementation and reinforcement process of biometric standards to meet agency-specific mission needs. The latest version of this document is available on the Federal government's web site for biometric activities at www.biometrics.gov/standards¹.

The maintenance of this Registry is supported by agencies providing appropriate personnel and resources to the Subcommittee's standards and conformity assessment working group. Federal agencies identifying issues with this Registry should notify their representatives to the Subcommittee's standards and conformity assessment working group.

Two other documents support this Registry and the *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards*:

- Supplemental Information in Support of the NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards ;
- Catalog of USG Biometric Product Testing Programs [DRAFT].

In support of specific cross agency biometric data interoperability requirements, this Registry is cited by NATIONAL SECURITY PRESIDENTIAL DIRECTIVE/NSPD -- 59/ HOMELAND SECURITY PRESIDENTIAL DIRECTIVE/HSPD -- 24, Biometrics for Identification and Screening to Enhance National Security.

For comments or to obtain additional information about this document, send e-mail to standards@biometrics.gov.

2. Scope

This Registry lists recommended biometric standards for USG-wide use. Only standards finalized and approved by a standards developing organization are eligible for analysis by the Subcommittee. Inclusion of a standard in this Registry requires consensus agreement of USG agencies through the Subcommittee's deliberative process. For dated references to standards, only the edition cited applies. For undated references to standards, the latest edition of the referenced standard (including any amendments) applies.

These recommendations take into account:

- the differences in how criminal identification and civil biometric authentication systems operate;
- the need to accommodate current implementations as well as new implementations;

¹ The latest version of this Registry is also available at www.standards.gov/biometrics.

- the movement to international versions of these national standards.

Therefore, along with recommended biometric standards, some high level guidance is often provided with respect to implementation, migration, and grandfathering of existing implementations. Further guidance may be found in the Supplemental document.

This Registry is divided into sub-registries of standards or profiles for:

- biometric data collection, storage, and exchange standards;
- biometric transmission profiles;
- biometric identity credentialing profiles;
- biometric technical interface standards;
- biometric conformance testing methodology standards;
- biometric performance testing methodology standards.

Additional biometric standards will be added to this Registry as other standards in the above categories (e.g., other modalities, such as DNA) or additional categories (e.g., biometric quality measurement standards) are approved by the standards developers and evaluated by the USG for USG-wide use.

This Registry may have supplements intended for use within specific communities of the USG. For information on the status of any such supplements, send email to standards@biometrics.gov.

3. Verbal forms for the expression of provisions

The following terms are used in this document to indicate mandatory, optional, or permissible requirements:

- the terms “shall” and “shall not” indicate requirements strictly to be followed in order to conform to this document and from which no deviation is permitted;
- the terms “should” and “should not” indicate that among several possibilities one is recommended as particularly suitable, without mentioning or excluding others, or that a certain course of action is preferred but not necessarily required, or that (in the negative form) a certain possibility or course of action is deprecated but not prohibited;
- the terms “may” and “need not” indicate a course of action permissible within the limits of this document.

4. Terms and definitions

For the purposes of this document, the following terms and definitions apply. The terms are grouped according by conceptual area, not alphabetic order.

- **standard** - document, established by consensus and approved by a recognized body, that provides, for common and repeated use, rules, guidelines or characteristics for activities or their results, aimed at the achievement of the optimum degree of order in a given context. [ISO/IEC Guide 2:2004]
- **base standard** - a fundamental standard with elements that contain options

NOTE Base standards can be used in diverse applications, for each of which it may be useful to fix the optional elements in a standardized profile with the aim of achieving interoperability between instances of the specific application. [ISO/IEC 24713-1]

- **biometric profile** - conforming subsets or combinations of base standards used to effect specific biometric functions
 - NOTE Biometric profiles define specific values or conditions from the range of options described in the relevant base standards, with the aim of supporting the interchange of data between applications and the interoperability of systems. [ISO/IEC 24713-1]
- **standards developing organization** - an organization that develops and approves consensus standards
 - NOTE Such organizations may be: accredited, such as ANSI accredited INCITS and ANSI accredited NIST ITL; or international treaty based, such as ICAO; or international private sector based, such as ISO/IEC; or a consortium, such as RTIC; or a government agency, such as the DoD, DHS, FBI, and NIST.
- **certification** - third-party attestation related to products, processes, systems or persons [ISO/IEC 17000:2004, Conformity assessment — Vocabulary and general principles]
 - NOTE 1 Certification of a management system is sometimes also called registration.
 - NOTE 2 Certification is applicable to all objects of conformity assessment except for conformity assessment bodies themselves, to which accreditation is applicable.
- **test** - technical operation that consists of the determination of one or more characteristics of a given product, process or service according to a specified procedure [ISO/IEC Guide 2:2004]
- **testing** - action of carrying out one or more tests [ISO/IEC Guide 2:2004]
- **conformance testing** - process of checking, via test assertions, whether an implementation faithfully implements the standard or profile
- **performance testing** - measures the performance characteristics of an implementation such as system error rates, throughput, or responsiveness, under various conditions
- **sample** - raw data representing a biometric characteristic, which is captured and processed by the biometric system or the digital representation of a biometric characteristic used internally by a biometric system
- **template** - encoded representation of features extracted from a sample suitable for direct comparison
- **sample quality** – properties of a biometric sample associated with its fidelity to its source and its expected performance in a verification or identification system
- **signal** - one dimensional time series data or spatial data
 - EXAMPLE 1 A speech recording
 - EXAMPLE 2 The coordinates and pressure of a pen in a handwriting recognition system, is an example of a multivariate signal (i.e. x and y and pressure).
- **image** - two or three dimensional spatial data
 - EXAMPLE 1 A fingerprint image
 - EXAMPLE 2 A three dimensional facial image (i.e. including shape information)
- **proprietary image** - image format defined in a privately controlled biometric data format specification
- **proprietary signal** - signal format defined in a privately controlled biometric data format specification
- **basic interoperability** - ability of a generator to create samples that can be processed by other suppliers' comparison subsystems, and the ability of a supplier's comparison subsystem to process input samples from other suppliers' generators [ISO/IEC 19795-4:2008 Interoperability Performance Testing]
- **interoperable performance** - performance associated with the use of generator and comparison subsystems from different suppliers

- **native performance** - performance associated with the use of generator and comparison subsystems from a single supplier
- **performance interoperability** - measure of the adequacy of interoperable performance
- **scenario test** - the online evaluation of end-to-end system performance in a prototype or simulated application in which samples collected from test subjects are processed in real time. [ISO/IEC 19795-2:2005 Testing Methodologies for Technology and Scenario Evaluation]

NOTE Scenario tests are intended for measurement of performance in modeled environments, inclusive of test subject-system interactions. Scenario Testing assesses biometric technologies in a manner representative of the operational application while maintaining control of performance variables.

- **technology test** - the offline evaluation of one or more algorithms for the same biometric modality using a pre-existing or specially-collected corpus of samples

5. Acronyms and Abbreviations

ABIS	Automated Biometric Identification System
ANSI	American National Standards Institute
APB	Advisory Policy Board
BDB	Biometric Data Block
BIAS	Biometric Identity Assurance Services
BioAPI	Biometric Application Programming Interface
BIR	Biometric Information Record
BSP	Biometric Service Provider
CBEFF	Common Biometric Exchange Formats Framework
CJIS	Criminal Justice Information Services
CTS	Conformance Test Suite
DHS	Department of Homeland Security
DoD	Department of Defense
EBTS	Electronic Biometric Transmission Specification
EFTS	Electronic Fingerprint Transmission Specification
FBI	Federal Bureau of Investigation
FDIS	Final Draft International Standard
FIPS	Federal Information Processing Standard
HSPD	Homeland Security Presidential Directive
IAFIS	Integrated Automatic Fingerprint Identification System
ICAO	International Civil Aviation Organization
IDENT	Automatic Biometric Identification System
IDMS	Identity management system
IEC	International Electrotechnical Commission
INCITS	InterNational Committee for Information Technology Standards
INT-I	Interpol Implementation of the ANSI/NIST ITL 1-2000 Standard
ISO	International Organization for Standardization
ITL	Information Technology Laboratory
IXM	IDENT Exchange Messages
JPEG	Joint Photographic Experts Group
LDS	Logical Data Structure

MINEX	Minutiae Interoperability Exchange Test
MRTD	Machine Readable Travel Document
NGI	Next Generation Identification
NIST	National Institute of Standards and Technology
NSTC	National Science and Technology Council
PIV	Personal Identity Verification
PNG	Portable Network Graphics
RT	Registered Traveler
RTIC	Registered Traveler Interoperability Consortium
SAP	Subject Acquisition Profile
SOAP	Simple Object Access Protocol
TWIC	Transportation Workers Identification Credential
TWPDES	Terrorist Watchlist Person Data Exchange Standard
USG	United States Government
US-VISIT	United States Visitor and Immigrant Status Indicator Technology
WSQ	Wavelet Scalar Quantization
XML	Extensible Markup Language

6. Registry concepts and standards nomenclature

The meanings for the headings of the columns in the following tables are as follows:

Validity Period: This column shall be updated periodically as new or improved standards are developed. This may result in the retirement or deprecation of a standard. In such cases, a migration strategy to facilitate backward compatibility will be needed because standardized data will likely exist in databases or on identity credentials. Agencies engaged in the design of biometrically enabled applications shall adhere to the standards called out below, and shall heed the "validity period" value.

Biometric Data²: This column is organized around the kind of data that is being stored. This derives from the particular biometric modalities chosen for an operation. In some cases, feature based data is stored, and thus the column identifies the captured or processed representation of the sample.

Domain of Applicability: The functions of a generic biometric application include an enrollment phase, and a subsequent identification or verification phase. The enrollment phase embeds capture of an initial sample. The capture may be from a cooperative, non-cooperative or uncooperative subject. Enrollment itself is usually an attended operation. These factors influence the selection of an appropriate data interchange standard because conformance to a standard might be unattainable (e.g., non-cooperative imaging will not always yield a frontal face, for example).

Conceptually a general biometric system³ might execute:

- data capture;
- transmission;
- image or signal processing;
- data storage;
- matching;

² This column appears only for the Biometric Data Collection, Storage, and Exchange Standards.

³ This description of biometric systems is expanded upon in ISO/IEC 24713-1:2008, Biometric Profiles for Interoperability and Data Interchange – Part 1: Overview of Biometric Systems and Biometric Profiles

- decision;
- administration;
- interface.

Recommended standards: This column enumerates those standards. The intent is that all biometric samples captured, or otherwise instantiated during the validity period, in a domain of applicability shall be encoded in formal conformity with the identified standards. In cases where two or more standards are specified, either or both may be used. In cases where the standards contain high level options or branches, values are mandated as needed.

Notes: This column provides implementation guidance and caveats on use and non-use of this and other standards. When the column includes guidance and refinements on the use of the standard (e.g., on compression) the use of the word *shall* is normative. That is, when users adopt one of the recommended standards, the guidance is required.

Standards nomenclature: The ISO standards identified in the following sections carry specific nomenclature. The example in the Table below explains the fields. The base standard, as originally developed in the international body, is shown in bold. The details of any subsequent US adoption which enclose this are shown in normal type.

INCITS/ISO/IEC 19794-6:2005[2007]					
INCITS	ISO/IEC	19794	-6	2005	2007
This is the name of the body in the U.S. that adopts the international standard	The parent standards development body	ISO/IEC 19794 is a multipart data interchange standard	The dash six denotes Part 6 which standardizes exchange of iris imagery	This is the year that the standard was published. Development was generally completed a few months prior.	This identifies the year the standard was adopted by the adopter.

For standards that have published amendments, the amendment itself is identified with the following syntax:
 INCITS/ISO/IEC 19784-1:2006/Amdt. 1 -2007[2008]

7. Biometric data collection, storage, and exchange standards

The biometric standards listed in Table 1 shall be used in all USG applications for which biometric data:

- are copied or moved between systems within an agency;
- are copied or moved to or by agencies;
- persist beyond the interaction of a subject with a sensor or system.

The biometric standards listed below cover:

- fingerprint images;
- latent fingerprint images;
- palm print images;
- fingerprint minutia records;
- facial images;
- iris images.

Standards for other modalities have been approved by the various standards developers. They are not listed here because the imperative for development of this Registry was ongoing or anticipated multi-agency or USG-wide applications. For parties seeking to collect, store and exchange data from modalities not covered by this Registry, they have the option of using standards approved by national or international standards developers⁴.

It is assumed that parent applications can properly embed or wrap biometric data formatted according to the standards enumerated below (e.g., EBTS transactions embedding Type 14 fingerprint records). Data records or sets of data records shall not be wrapped in a proprietary wrapper that requires a specific provider's software to decode or encode.

While Table 1 addresses collection, storage and exchange of biometric data, existing transmission profiles such as the FBI's EBTS (see Table 2) might further modify or restrict the recommended standards of Table 1.

⁴ The DoD tracks the development of biometric standards. For a copy of the DoD's "BTF Standards Development Status Update" contact standards@biometrics.gov.

Table 1 - Registry of Biometric Data Collection, Storage, and Exchange Standards

#	Validity period	Biometric data	Domain of applicability	Recommended standards	Notes	Implementation Guidance
Finger and Palm Recognition						
1.	October 2007 – current	Plain or rolled fingerprint images. For latent fingerprint images, see row 2.	Capture, storage and exchange of data (e.g., enrollment or registration)	ANSI/NIST-ITL 1-2007, Type 14	<p>PIV (FIPS 201-1, 2006) requires the use of INCITS 381:2004 for the retention of images.</p> <p>Other standards, or standardized records, including those enumerated below shall not be used as a substitute for the required standard; they may be used only in addition: ANSI/NIST-ITL 1-2007, Type 3, 5 or 6; INCITS 381:2004; INCITS/ISO/IEC 19794-4:2005[2007].</p>	<p>Capture and storage with resolution ≥ 19.69 pixels/mm.</p> <p>When images are captured at 19.69 pixels/mm and compressed with WSQ, the compression ratio shall not exceed 15:1.</p> <p>When images are captured at 39.37 pixels/mm and compressed using JPEG 2000, the compression ratio shall not exceed 15:1.</p> <p>NOTE: While ANSI/NIST-ITL 1-2007 Type 4 remains the predominant format for transmission of rolled fingerprint information, the Type 14 record is recommended because it is:</p> <ul style="list-style-type: none"> • used for plain impression transactions including segmentation coordinates; • supporting use of high resolution images; • a more flexible format for additional metadata. <p>However, users should check with receiving agencies that they are capable of accepting Type 14 data.</p>
1XML	December 2008 – current	Plain or rolled fingerprint images	Capture, storage and exchange of data (e.g., enrollment or registration) in XML format.	ANSI/NIST-ITL 2-2008, Type 14	<p>The following XML standards, or standardized records, shall not be used: ANSI/NIST-ITL 2-2008, Type 3, 5 or 6. This requirement deprecates the use of these legacy types in new applications even though ANSI/NIST-ITL 2-2008 included them by default.</p> <p>NOTE: Implementers migrating to ANSI/NIST-ITL 2-2008 should also migrate to Type 14 from Type 4.</p>	
2.	October 2007 -	Latent fingerprints or	Storage and exchange	ANSI/NIST-ITL 1-2007, Type 13	Other standards or standardized records, including those enumerated below shall not be	The latent image shall be acquired with a native resolution

#	Validity period	Biometric data	Domain of applicability	Recommended standards	Notes	Implementation Guidance
	current	latent palm print images	of data (e.g., enrollment or registration)		used as a substitute for the required standard; they may be used only in addition: ANSI/NIST-ITL 1-2007, Type 7; INCITS 381:2004; INCITS/ISO/IEC 19794-4:2005[2007]. Other standards, including those enumerated below shall not be used: ANSI/NIST-ITL 2-2008, Types 3, 4, 5, 6, 14, 15, 16 and 99. When latent minutia are extracted from a latent image and encoded in, for example, an ANSI/NIST-ITL 1-2007, Type 9, the parent image shall be retained.	of 394 pixels/cm or greater. Latent images should be uncompressed. If losslessly compressed, images shall be stored in conformance to the ISO/IEC 15948 format (PNG). Images shall not be compressed using a lossy compression algorithm If reduced resolution versions are prepared (e.g., for transmission) the parent high resolution image shall be retained.
2XML	December 2008 - current	Latent fingerprints or latent palm print images	Storage and exchange of data (e.g., enrollment or registration) in XML format	ANSI/NIST-ITL 2-2008, Type 13	Other standards, including those enumerated below shall not be used: ANSI/NIST-ITL 2-2008, Types 3, 4, 5, 6, 14, 15, 16 and 99. When latent minutia are extracted from a latent image and encoded in, for example, an ANSI/NIST-ITL 2-2008, Type 9, the parent image shall be retained.	Capture and storage with resolution ≥ 197 pixels/cm. When images are captured at 197 pixels / cm and compressed with WSQ, the compression ratio shall not exceed 15:1. This may be achieved by invoking the WSQ compressor with a target bit rate parameter greater than or equal to 8/15 bits per pixel.
3.	October 2007 – current	Palm prints (excluding latent palm prints)	Storage and exchange of data (e.g., enrollment or registration)	ANSI/NIST-ITL 1-2007, Type 15	Other standards or standardized records, including those enumerated below shall not be used as a substitute for the required standard; they may be used only in addition: INCITS 381:2004; INCITS/ISO/IEC 19794-4:2005[2007]. Other standards or standardized records, including those enumerated below shall not be used: ANSI/NIST-ITL 1-2007, Types 3, 4, 5, 6, 13, 14, 16 and 99.	When images are captured at 197 pixels/cm and compressed with WSQ, the compression ratio shall not exceed 15:1. This may be achieved by invoking the WSQ compressor with a target bit rate parameter greater than or equal to 8/15 bits per pixel.
3XML	December 2008 –	Palm prints (excluding	Storage and exchange	ANSI/NIST-ITL 2-2008, Type 15	Other standards or standardized records, including those enumerated below shall not be	When images are captured at 394 pixels/cm and compressed using JPEG 2000 the

#	Validity period	Biometric data	Domain of applicability	Recommended standards	Notes	Implementation Guidance
	current	latent palm prints)	of data (e.g., enrollment or registration) in XML format		used: ANSI/NIST-ITL 2-2008, Types 3, 4, 5, 6, 13, 14, 16 and, 99.	compression ratio shall not exceed 15:1. This may be achieved by invoking the JPEG 2000 compressor with a target bit rate greater than or equal to 8/10 bits per pixel. If images scanned at 1000 ppi and compressed using JPEG 2000 are to be converted to images at 500 ppi and compressed using WSQ, then the MITRE procedures [MITRE1000] shall be followed.
4.	October 2007 – current	Fingerprint minutiae, not latent minutiae For minutiae encoded in latent images, see row 7.	Storage and exchange outside and unrelated to personal identity credentials	INCITS 378:2004 or ANSI/NIST-ITL 1-2007 Type 9, Fields 1-4 and 13-23 or ANSI/NIST-ITL 1-2007 Type 9, Fields 1-4 and 126-150	Other standards or standardized records, including those enumerated below shall not be used as a substitute for the required standard; they may be used only in addition: INCITS/ISO/IEC 19794-2:2005[2008]. If ANSI/NIST-ITL 1-2007 Type 9 is used, vendor blocks (i.e. fields 31 - 125 and 151-175) shall not be used.	Verification applications (e.g., access control) shall not use the “vendor-defined extended data” fields of INCITS 378:2004 clause 6.6. Better accuracy will be obtained if, within the target application, it is possible to additionally exchange standardized image records, per row 1 of this Table. Identification applications shall use the INCITS 378:2004 standard. This may include proprietary template data in the “vendor-defined extended data” fields of INCITS 378:2004 clause 6.6. Proprietary template data is non-interoperable but some implementations have been shown to have improved accuracy over standardized data alone [MINEX04]. It is usually
4 XML	December 2008 – current	Fingerprint minutiae encoded in XML. For minutiae	Storage and exchange outside and unrelated to personal identity credentials	ANSI/NIST-ITL 2-2008 Annex G XML encoding of INCITS 378:2004 or	If ANSI/NIST-ITL 2-2008 Type 9 is used, vendor blocks (i.e. fields 31 - 125 and 151-175) shall not be used.	

#	Validity period	Biometric data	Domain of applicability	Recommended standards	Notes	Implementation Guidance
		encoded in latent images, see row 7.		ANSI/NIST-ITL 2-2008 Type 9, per Tables 216a and 216b		usable only if the data is prepared and matched by the products of a single supplier. Reliance on such proprietary data will promote vendor lock-in. In order to mitigate this risk, the parent images shall be retained. To eliminate this risk, standardized image records should be exchanged, per row 1 of this Table. To avoid abuse of this allowance of proprietary data, the standardized minutiae data required by clauses 6.1 through 6.5 of INCITS 378:2004 should be produced by MINEX compliant template generators.
5.	October 2007 – current	Fingerprint minutiae	Storage in, and transmission to, personal identity credentials for match-on-card	INCITS/ISO/IEC 19794-19794-2:2005[2008], clause 8 compact card format with clause 9 format types 0001, 0003, 0005	<p>INCITS/ISO/IEC 19794-2:2005[2008] (compact card format) shall be stored on the card for match-on-card.</p> <p>INCITS/ISO/IEC 19794-2:2005[2008] (compact card format) shall be sent to the card for verification against the reference template on the card.</p> <p>In both cases the minutiae may be prepared from parent INCITS 378:2004 records.</p> <p>For match-on-card, neither INCITS 378:2004 nor INCITS/ISO/IEC 19794-2:2005[2008] clause 7 (record format) shall be stored on the card.</p> <p>For match-on-card, neither INCITS 378:2004 nor INCITS/ISO/IEC 19794-2:2005[2008] clause 7 (record format) shall be sent to the card.</p> <p>Regarding INCITS/ISO/IEC 19794-2:2005[2008] card formats, the absence of a header and</p>	

#	Validity period	Biometric data	Domain of applicability	Recommended standards	Notes	Implementation Guidance
6.	October 2007 – current	Fingerprint minutiae	Storage in, and transmission from, personal identity credentials for match-off-card	INCITS 378:2004	<p>ambiguities inherent in the sort-ordering of minutiae mean that such records shall not be used for persistent storage off-card.</p> <p>ANSI/NIST-ITL 1-2007 and ANSI/NIST-ITL 2-2008 shall not be used.</p> <p>In match-off-card applications, none of the INCITS/ISO/IEC 19794-2:2005[2008] formats shall be used. This applies to both the reference and verification templates.</p> <p>ANSI/NIST-ITL 1-2007 and ANSI/NIST-ITL 2-2008 shall not be used.</p>	
7.	October 2007 – current	Latent fingerprint minutiae	Storage and exchange of data (e.g., enrollment or registration)	ANSI/NIST-ITL 1-2007, Type 9, Fields 1-4 and 13-23 or ANSI/NIST-ITL 1-2007 Type 9, Fields 1-4 and 126-140	<p>Other standards, including those enumerated below shall not be used as a substitute for the required standard; they may be used only in addition:</p> <p>INCITS 378:2004.</p>	<p>Standardized minutiae records afford only limited automated matching accuracy, and therefore parent latent images shall be retained with any extracted minutiae.</p> <p>Fields 13-23 are defined in Appendix J of the FBI's EBTS.</p>
7 XML	October 2007 – current	Latent fingerprint minutiae encoded using XML	Storage and exchange of data (e.g., enrollment or registration)	ANSI/NIST-ITL 2-2008, Type 9, Tables 216a and 216b		
Face Recognition						
8.	October 2007 – current	2D Face images	Storage of digital images in personal identity credentials	INCITS/ISO/IEC 19794-5:2005[2007], Full Frontal or Token	<p>The INCITS/ISO/IEC 19794-5:2005[2007] "basic" mode shall not be used.</p> <p>INCITS 385:2004 shall not be used.</p> <p>ANSI/NIST-ITL 1-2007 and ANSI/NIST-ITL 2-2008 shall not be used.</p>	

#	Validity period	Biometric data	Domain of applicability	Recommended standards	Notes	Implementation Guidance
9.	October 2007 - current	2D Face images	For capture and storage in MRTDs (e.g., e-Passport chip reading)	ICAO 9303	<p>The following informative material should be consulted.</p> <p>For general case: ISO/IEC 19794-5:2005, Amendment 1 adds an Annex to the base standard as guidance for producing or requiring either conventional printed photographs or digital images of faces that may be used in applications for passports, visas, or other identification documents and when those images are required to conform to the frontal image types of this standard (INCITS/ISO/IEC 19794-5:2005 [2007]).</p> <p>ICAO 9303 covers capture, storage and transmission.</p> <p>INCITS 385:2004 shall not be used.</p> <p>ANSI/NIST-ITL 1-2007 and ANSI/NIST-ITL 2-2008 shall not be used.</p>	
10.	October 2007 - current	2D Face images	Capture and storage (i.e., enrollment or registration processes) for which end-to-end subject capture times above 120 seconds are tolerable	ANSI/NIST-ITL 1-2007, Type 10 with subject acquisition profile (SAP) of level 10 or above or INCITS/ISO/IEC 19794-5:2005 [2007], Full Frontal or Token, with at least 90 pixels between the	<p>Other standards or standardized records, including those enumerated below shall not be used: INCITS 385:2004 ANSI/NIST-ITL 1-2007, Types 7, 16 and 99.</p>	<p>Failure to conform to the quality-related requirements of these standards will undermine facial recognition performance.</p> <p>ISO/IEC 19794-5:2005, Amendment 1 should be consulted. It adds an Annex to the base standard as guidance for producing either conventional printed photographs or digital images of faces that may be used in applications for passports, visas, or other identification documents.</p>

#	Validity period	Biometric data	Domain of applicability	Recommended standards	Notes	Implementation Guidance
10	December 2008 - current	2D Face images encoded in XML	Capture and storage (i.e., enrollment or registration processes) for which end-to-end subject capture times above 120 seconds are tolerable	eyes from all subjects ANSI/NIST-ITL 2-2008, Type 10 with subject acquisition profile (SAP) of level 10 or above	Other standards or standardized records, including those enumerated below shall not be used: ANSI/NIST-ITL 2-2008, Types 7, 16 and 99.	
11.	October 2007 – current	2D Face images	Non-cooperative or uncooperative capture and storage of images	ANSI/NIST-ITL 1-2007, Type 10 with subject acquisition profile (SAP) of level 1 or above or INCITS/ISO/IEC 19794-5:2005[2007] Basic type only	For images collected in applications in which subjects are imaged in a non-cooperative or uncooperative manner. The acquisition should be frontal when possible. Other standards or standardized records, including those enumerated below shall not be used: INCITS 385:2004 ANSI/NIST-ITL 1-2007, Types 7, 16 and 99.	
11	December 2008 – current	2D Face images	Non-cooperative or uncooperative capture and storage of images	ANSI/NIST-ITL 2-2008, Type 10 with subject acquisition profile (SAP) of level 1 or above	For images collected in applications in which subjects are imaged in a non-cooperative or uncooperative manner. The acquisition should be frontal when possible. Other standards or standardized records, including those enumerated below shall not be used: ANSI/NIST-ITL 2-2008, Types 7, 16 and 99.	
12.	October 2007 –	2D Face images	All other capture,	ANSI/NIST-ITL 1-2007, Type 10	Conformance to the ANSI/NIST-ITL 1-2007 SAP level 1 and the INCITS/ISO/IEC 19794-	

#	Validity period	Biometric data	Domain of applicability	Recommended standards	Notes	Implementation Guidance
	current		storage or exchange applications	with subject acquisition profile (SAP) of level 1 or above or INCITS/ISO/IEC 19794-5:2005[2007], Basic, Full Frontal or Token	5:2005[2007] "Basic" type allows storage of an arbitrarily poor photograph whose digital, scene, photometric and geometric properties are unlikely to yield acceptable face recognition accuracy. Other standards or standardized records, including those enumerated below shall not be used: INCITS 385:2004 ANSI/NIST-ITL 1-2007, Types 7, 16 and 99.	
12 XML	December 2008 - current	2D Face images	All other capture, storage or exchange applications	ANSI/NIST-ITL 2-2008, Type 10 with subject acquisition profile (SAP) of level 1 or above	Conformance to the ANSI/NIST-ITL 2-2008 SAP level 1 allows storage of an arbitrarily poor photograph whose digital, scene, photometric and geometric properties are unlikely to yield acceptable face recognition accuracy. Other standards or standardized records, including those enumerated below shall not be used: ANSI/NIST-ITL 2-2008, Types 7, 16 and 99.	
13.	October 2007 – current	Iris images	Capture, storage and exchange of data (e.g., enrollment or registration)	The rectilinear image format of INCITS/ISO/IEC 19794-6:2005[2007] or ANSI/NIST-ITL 1-2007, Type 17	Other standards or standardized records, including those enumerated below shall not be used: INCITS 379:2004 (standard withdrawn 2008) ANSI/NIST-ITL 1-2007, Types 7 and 16. The ANSI/NIST-ITL 1-2007, Type 17 format is a strict derivative of INCITS/ISO/IEC 19794-6:2005[2007], and may be used as an alternative. Other standards, including those enumerated below shall not be used as a substitute for the required standard; they may be used only in addition: All ISO/IEC 19794-6:2005 polar image formats.	If lossy compression is applied to iris images the compression ratio shall not exceed 6:1. For compression algorithms without a bit-rate parameter (e.g., JPEG), this may require iteration over the compression "quality" parameter.

#	Validity period	Biometric data	Domain of applicability	Recommended standards	Notes	Implementation Guidance
13 XML	December 2008 – current	Iris images encoded in XML	Capture, storage and exchange of data (e.g., enrollment or registration)	ANSI/NIST-ITL 2-2008, Type 17	<p>Iris stored in any of the polar image formats of INCITS/ISO/IEC 19794-6:2005[2007] may be retained only if their rectilinear image parents are also retained.</p> <p>The ANSI/NIST-ITL 2-2008, Type 17 format is a strict derivative of INCITS/ISO/IEC 19794-6:2005[2007], and may be used as an alternative.</p> <p>Other standards or standardized records, including those enumerated below shall not be used: ANSI/NIST-ITL 2-2008, Types 7 and 16.</p>	

8. Biometric transmission profiles

To facilitate interoperability, biometric base standards, such as the Biometric Data Collection, Storage, and Exchange Standards in Table 1, should normally be used in conjunction with a biometric profile. Such profiles specify application-specific criteria onto the base standard. This profiling could consist of establishing definitive values for performance related parameters in the base standard (e.g., resolution, maximum compression) or enumerating values for optional or conditional requirements (e.g., full-frontal face vs. token face in INCITS/ISO/IEC 19794-5:2005[2007]).

Biometric profiles developed for USG applications should address, on a clause-by-clause basis, all the normative requirements of the base standards, and where appropriate:

- call out values of parameters (e.g., number of finger);
- call out normative practice (e.g., encoding of core and delta positions in minutia records);
- promote informative material to become normative requirements (e.g., maximum face image compression ratios);
- demote normative requirements if compliance would be problematic. Such a step shall be undertaken only after an evidence-based justification can be established and documented. This practice should be undertaken with utmost caution because it breaks conformance to the standard, and may undermine interoperability.

Configurable elements of standards should be specified as part of requirements documents based on operational needs of the implementations.

Proprietary data

Some of the base standards enumerated in this document include fields for additional proprietary data. A biometric profile should disallow population of these fields because proprietary data is non-interoperable and is likely to be used in preference to standardized data thereby subverting interoperability via vendor lock-in.

USG applications shall not use proprietary image or signal formats when a national or international standard exists for images or signals related to that biometric.

Proprietary extensions

USG applications should generally prohibit inclusion of proprietary data in standardized records that contain standardized data. Applications may embed proprietary templates, and achieve interoperability at the image-level.

Biometric Profiles and Data Models for Large Scale Identification Applications

The biometric transmission profiles of Table 2 are specifications developed by federal and international organizations that permit electronic communication with the specified system. These documents are not base standards but are critical because they define current (“as is”) technical requirements that facilitate interoperability.

As of September 2007, the FBI EBTS Version 8.0 superseded the FBI EFTS Version 7.1. In April 2008, the FBI EBTS Version 8.002 superseded the FBI EBTS Version 8.001. The FBI's EBTS Version 8.002 clarified existing processing capabilities. In November 2008, FBI EBTS Version 8.1 was approved. It offers a superset of the functionality provided by prior “8.” versions. FBI EBTS Version 8.1 is backward compatible with prior “8.” versions. The FBI EBTS Version 8.1 is the current standard for interfacing with the FBI Integrated Automated Fingerprint Identification System (IAFIS). The FBI EBTS Version 8.1 contains a description of operational concepts, descriptors, and field edit specifications, image quality specifications, and other information related to IAFIS services. The scope of the FBI EBTS Version 8.1 has expanded over the FBI EFTS Version 7.1 to include additional biometric modalities (e.g., palmprint, facial, and iris) in recognition of the rapidly developing biometric identification industry.

ANSI/NIST-ITL 1-2000 is specified in the FBI EFTS Version 7.1. ANSI/NIST-ITL 1-2007 is specified in FBI EBTS Version 8.1. DoD has developed its own EBTS with the goal of being compatible with the FBI's EFTS and EBTS. ANSI/NIST-ITL 1-2000/FBI EFTS Version 7.1 and ANSI/NIST-ITL 1-2007/ FBI EBTS Version 8.1 will need to coexist for some time.

A standards-based service model for interacting with the US-VISIT Program's IDENT system has been in effect since September 2007. IDENT Exchange Messages (IXM) provides a common interface to IDENT for client applications. IXM is based on XML and provides a communication protocol embedded in the SOAP framework. The latest IXM standard provides an overview and detailed information on each message operation, the steps required to create an interface, and guidelines and examples intended to help external users interact with US-VISIT/IDENT applications via the IXM format.

The Terrorist Watchlist Person Data Exchange Standard (TWPDES) provides a comprehensive XML based standard for exchanging and sharing terrorist-related information across the entire intelligence and law enforcement communities, both in the United States and abroad with biometric and biographic support in a single package. It incorporates the ANSI/NIST-ITL 2-2008 standard for biometric identifiers.

TWPDES 1.2b is NIEM 2.0 compliant supporting all of the terrorist watchlisting requirements and encounter scenarios in the communities. Users may constrain the standard to support only the specific requirements in the users' domain. The specification also has built-in extension mechanisms that can be used for inter-agency terrorist-data exchange models. TWPDES 1.2b has been accepted by DHS, DoD, and DOJ as a recognized standard for exchanging data.

Table 2 - Registry of Biometric Transmission Profiles

#	Validity period	Domain of applicability	Recommended Transmission Profiles	Notes
1.	May 2009 – current	Applications sharing terrorist data with the U.S. Government's intelligence community and law enforcement.	TWPDES 1.2b	This version supersedes versions 1.0, 1.1, 1.2a and 2.0.
2.	Through October 2008	Applications exchanging data with the FBI IAFIS/NGI identification system	FBI EFTS Version 7.1	Superseded by FBI EBTS Version 8.1. FBI EFTS v7.1 exists within this registry for backwards compatibility with legacy systems.
3.	October 2007 – current	Applications exchanging data with the FBI IAFIS/NGI identification system	FBI EBTS Version 8.1	The Criminal Justice Information Services (CJIS) Advisory Policy Board (APB) has recently approved the FBI EBTS Version 8.1 for interfacing with the FBI Integrated Automated Fingerprint Identification System (IAFIS) and its successor Next Generation Identification (NGI). Version 8.1 offers a superset of the functionality provided by prior versions. Version 8.1 is backward compatible with prior versions. In any case version 8.1 should be adopted for new applications.
4.	October 2007 – current	Applications exchanging data with the DoD ABIS identification system	DoD EBTS v1.2	DoD EBTS v1.2 is a superset of the FBI EFTS v7.1 for DoD-specific needs. DoD EBTS v1.2 preceded the development of FBI EBTS v8.001.
5.	September 2007 – current	Applications exchanging data with the DHS IDENT identification system	IDENT eXchange Messaging (IXM)	The IXM specification provides detailed information on messaging operation, and steps required to create an interface for external users to interact

#	Validity period	Domain of applicability	Recommended Transmission Profiles	Notes
				with US-VISIT/IDENT applications.
6.	October 2005 - current	Applications exchanging data with the Interpol identification system	Interpol Implementation of ANSI/NIST-ITL 1-2000 (INT-I)	This standard is used to transmit information between nations for international law enforcement.

9. Biometric identity credentialing profiles

The FIPS 201 standard specifies the architecture and technical requirements for a common identification standard for all US Government employees and contractors. It contains two major sections. Part one describes the requirements for a personal identity verification system that meets the control and security objectives of Homeland Security Presidential Directive 12, including personal identity proofing, registration, and issuance. Part two provides detailed specifications that will support technical interoperability among PIV systems. It describes the card elements, system interfaces, and security controls required to securely store, process, and retrieve identity credentials from the card. The interfaces and data formats of biometric information are specified in NIST Special Publication 800-76, Biometric Data Specification for Personal Identity Verification.

The TWIC Reader Hardware and Card Application Specification leverages FIPS 201. For all transportation workers requiring unescorted physical and/or logical access to national facilities, the TWIC design defines the behavior at the card interface of the TWIC card application as well as the requirements for TWIC smart card readers to be used with the TWIC.

Similarly the Registered Traveler Technical Interoperability Specification leveraged the FIPS 201 standard to specify the identify management infrastructure requirements for a fully-interoperable, vendor-neutral RT program within the United States.

The biometric credentialing profiles of Table 3 should be considered for all USG applications.

Table 3 - Registry of Biometric Identity Credentialing Profiles

#	Validity period	Domain of applicability	Recommended standards	Notes
1.	October 2007 – current	Personal identity verification	FIPS 201-1, 2006 NIST SP 800-76-1, 2007	HSPD-12 is applicable to Federal employees and contractors. Applicability to other agency specific categories of individuals (e.g., short-term (i.e., less than 6 months) guest researchers; volunteers; or intermittent, temporary or seasonal employees) is an agency risk-based decision. The TWIC and RT specifications are based upon the PIV standards (FIPS 201, and supporting NIST Special Publications) with certain extensions and modifications for their unique application environment.
2.	October 2007 – current	Registered travelers	Registered Traveler Interoperability Consortium <i>Technical Interoperability Specification Version 1.7</i> April 15, 2008	Version 1.0 of this Registry recommended Registered Traveler Interoperability Consortium <i>Technical Interoperability Specification Version 1.5</i> December 21, 2007

10. Biometric technical interface standards

The biometric technical interface standards listed in Table 4 shall be used in all USG applications for biometric systems that include “plug and play” capability. This permits agencies to easily, rapidly and seamlessly integrate system components into functioning systems and swap components as needed without losing functionality, such as the ability to achieve data interchange and to protect the biometric data during transmission and storage.

The BioAPI standards support “plug and play” compatibility by specifying how applications communicate with biometric vendor software in a common way independently of the biometric modality. This supports the swapping of products and incorporation of new products with no application modification.

The CBEFF standards specify data structures that support multiple biometric technologies in a common way. CBEFF's data structures, termed BIRs, conform to a CBEFF Patron Format which allows exchange of biometric data and related metadata (e.g., time stamp, validity period, and creator) and support security of biometric data in an open systems environment.

The BIAS standard defines biometric services used for identity assurance that are invoked over a services-based framework. It is intended to provide a generic set of biometric and identity-related functions and associated data definitions to allow remote access to biometric services.

Table 4 - Registry of Biometric Technical Interface Standards

#	Validity period	Domain of applicability	Recommended standards	Notes
1.	October 2007 - current	<p>Client-side capture and verification (e.g., enrollment workstation, kiosk) or server-side verification for one-to-one and multi-biometric applications</p> <p>There is no requirement for embedded devices to conform to the current versions of the BioAPI standards.</p> <p>This does not apply to law enforcement applications and other large-scale identification applications that require conformance to biometric profiles such as FBI EBTS.</p>	<p>INCITS/ISO/IEC 19784-1:2006[2007]</p> <p>INCITS/ISO/IEC 19784-2:2007[2008]</p> <p>or</p> <p>INCITS 358:2002</p>	<p>NIST and DoD have publicly available Conformance Test Suites (CTSs)⁵ to test Biometric Service Providers that claim conformance to INCITS 358:2002.</p> <p>No publicly available CTSs are known to be available for ISO/IEC 19784-1.</p> <p>Since there is a publicly available reference implementation for INCITS 358:2002 this standard may be used as an alternative to the international version if the lack of availability of the publicly available reference implementation for the international version is a deterrent to adoption at the present time.</p> <p>A framework component for INCITS/ISO/IEC 19784-1:2006[2007] is commercially available (i.e., license fee), which can serve the same purpose as a publicly available reference implementation.</p> <p>A graphical user interface specification is available as INCITS/ISO/IEC 19784-1:2006/Amdt 1 -2007[2008]</p>
2.	October	Biometric Information Records	INCITS 398:2008	Although the user can specify a new Patron Format,

⁵ http://www.itl.nist.gov/div893/biometrics/BioAPI_CTS/index.htm and <http://www.biometrics.dod.mil/CurrentInitiatives/Standards/BioAPI/Default.aspx>

#	Validity period	Domain of applicability	Recommended standards	Notes																					
	2007 – current	<p>conforming to a CBEFF Patron Format for the exchange, protection, encapsulation, transmission and storage of biometric data</p> <p>Encrypt and sign biometric data contained in Biometric Data Blocks in CBEFF BIRs by relying on the BIR Security Block, unless other system security mechanisms are already provided by means external to the BIR</p> <p>Patron Formats for applications that require transmission or storage of BIRs that require cleartext biometric headers or making metadata available without processing the record (e.g., for the purpose of indexing BIRs)</p> <p>This does not apply to law enforcement applications and other large-scale identification applications that require conformance to biometric profiles such as FBI EBTS.</p>		<p>those specified in INCITS 398:2008 are preferred:</p> <p>In addition to citing the INCITS 398:2008 standard, parties to a biometric interchange shall agree on a Patron Format. The ones specified in the standard are tabulated below.</p> <table border="1"> <thead> <tr> <th>#</th> <th>Name</th> <th>Domain</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>Patron Format A</td> <td>General purpose - See NOTE 1.</td> </tr> <tr> <td>2</td> <td>BioAPI BIR</td> <td>BioAPI Interfaces</td> </tr> <tr> <td>3</td> <td>ICAO LDS</td> <td>e-Passports / MRTDs</td> </tr> <tr> <td>4</td> <td>PIV</td> <td>PIV</td> </tr> <tr> <td>5</td> <td>ANSI/NIST Type 99</td> <td>Other modalities</td> </tr> <tr> <td>6</td> <td>Patron Format B</td> <td>Complex structures</td> </tr> </tbody> </table> <p>NOTE 1 NIST has a publicly available conformance testing architecture and Conformance Test Suite (CTS)⁶ to test implementations of Patron Format A.</p>	#	Name	Domain	1	Patron Format A	General purpose - See NOTE 1.	2	BioAPI BIR	BioAPI Interfaces	3	ICAO LDS	e-Passports / MRTDs	4	PIV	PIV	5	ANSI/NIST Type 99	Other modalities	6	Patron Format B	Complex structures
#	Name	Domain																							
1	Patron Format A	General purpose - See NOTE 1.																							
2	BioAPI BIR	BioAPI Interfaces																							
3	ICAO LDS	e-Passports / MRTDs																							
4	PIV	PIV																							
5	ANSI/NIST Type 99	Other modalities																							
6	Patron Format B	Complex structures																							
3.	October 2007 – current	Biometric services for identity assurance that are invoked over a services-based framework	INCITS 442:2008																						

11. Biometric conformance testing methodology standards

Conformance testing methodology standards may specify physical test requirements, logical test requirements (e.g., test assertions, test cases), use of reference data, test reporting formats, and means of testing requirements. Such standards can serve as the basis for the development of test tools (e.g., executable test code, reference data) and reference implementations, which can be used by organizations operating conformance testing programs.

The biometric conformance testing methodology standards listed in Table 5 should be considered for all tests run, commissioned or otherwise sponsored by USG agencies.

⁶ http://www.itl.nist.gov/div893/biometrics/CBEFF_PFA_CTS/

Table 5 - Registry of Biometric Conformance Testing Methodology Standards

#	Validity period	Domain of applicability	Recommended standards	Notes
1.	September 2007 - current	FBI certification of fingerprint systems that scan and capture fingerprints in digital, softcopy form, including hardcopy scanners such as ten-print card scanners, and live scan devices, altogether called "fingerprint scanners"; and systems utilizing a printer to print digital fingerprint images to hardcopy called "fingerprint printers"	FBI EBTS Version 8.1, Appendix F	The procedures for conduct of an Appendix F test can be found at http://www.mitre.org/tech/mtf/
2.	October 2007 – current	Conformance testing of Biometric Service Provider (BSP) implementations claiming conformance to critical requirements specified in INCITS/ISO/IEC 19784-1:2006[2007] (BioAPI 2.0)	ISO/IEC 24709-1:2007 and ISO/IEC 24709-2:2007	BSP implementations that are tested according to the methodology specified in ISO/IEC 24709-1 and with the test assertions specified in this part of ISO/IEC 24709 can only claim conformance to those aspects of ISO/IEC 19784-1 that are covered by these test assertions.
3.	October 2007 - current	Conformance testing of application(s) or service(s) implementations claiming conformance to the ANSI INCITS 378:2004 standard	INCITS 423.1:2008 and INCITS 423.2:2008	

12. Biometric performance testing methodology standards

The biometric performance testing methodology standards listed in Table 6 should be considered for all tests run, commissioned or otherwise sponsored by USG agencies.

Use of the standards does not restrict testing laboratories from conducting additional activities or using different practices. The standards are therefore suitable for agencies sponsoring tests in experimental or developmental applications.

Table 6 - Registry of Biometric Performance Testing Methodology Standards

#	Validity period	Domain of applicability	Recommended standards	Notes
1.	October 2007 – current	Physical and logical access control tests	INCITS/ISO/IEC 19795-1:2005[2007] and ISO/IEC 19795-2:2006	ISO/IEC 19795-2:2006 defines "technology" and "scenario" tests. For access control tests, only the latter is required. The following technical report should be consulted for modality specific guidance: ISO/IEC 19795-3:2007 - Biometric Performance Testing and Reporting – Part 3: Modality-Specific Testing.
2.	October 2007 - current	Testing of performance and interoperability of cross-supplier implementations generating and	INCITS/ISO/IEC 19795-1:2005[2007]	The following technical report should be consulted for modality specific guidance: ISO/IEC TR 19795-3:2007 - Biometric

#	Validity period	Domain of applicability	Recommended standards	Notes
		matching instances of standardized biometric data interchange data	and ISO/IEC 19795-4:2008	Performance Testing and Reporting – Part 3: Modality-Specific Testing.

13. References

Identification of Standards

The ISO standards identified in this section carry specific nomenclature. The example in the Table below explains the fields. The base standard, as originally developed in the international body, is shown in bold. The details of any subsequent US adoption which enclose this are shown in normal type.

INCITS/ISO/IEC 19794-6:2005[2007]					
INCITS	ISO/IEC	19794	-6	2005	2007
This is the name of the body in the U.S. that adopts the international standard	The parent standards development body	ISO/IEC 19794 is a multipart data interchange standard	The dash six denotes Part 6 which standardizes exchange of iris imagery	This is the year that the standard was published. Development was generally completed a few months prior.	This identifies the year the standard was adopted by the adopter.

For standards that have published amendments, the amendment itself is identified with the following syntax:
INCITS/ISO/IEC 19784-1:2006/Amdt. 1 -2007[2008]

1.	ANSI/NIST-ITL 1-2007	Data Format for the Interchange of Fingerprint, Facial, & Other Biometric Information – Part 1. Published as NIST Special Publication 500-271, May 2007. http://biometrics.nist.gov/standard/
2.	ANSI/NIST-ITL 2-2008	Data Format for the Interchange of Fingerprint, Facial, & Other Biometric Information – Part 2: XML Version, Published as a NIST Special Publication 500-275, August 2008 http://biometrics.nist.gov/standard/
3.	EBTS Version 1.2	DoD Electronic Biometric Transmission Specification (EBTS) Version 1.2 http://www.biometrics.gov/Standards/Default.aspx
4.	EBTS Version 8.1	FBI Electronic Biometric Transmission Specification (EBTS) Version 8.1 http://www.fbibiospecs.org/fbibioimetric/biospecs.html
5.	EFTS Version 7.1	FBI Electronic Fingerprint Transmission Specification (EFTS) Version 7.1 http://www.fbi.gov/hq/cjis/iafis/efts71/efts71.pdf
6.	FIPS 201-1, 2006	Personal Identity Verification for Federal Employees and Contractors http://csrc.nist.gov/publications/fips/fips201-1/FIPS-201-1-chng1.pdf
7.	HSPD-12	Policy for a Common Identification Standard for Federal Employees and Contractors http://www.dhs.gov/xabout/laws/gc_1217616624097.shtm
8.	ICAO 9303	Part 1 - Machine Readable Passport - Volume 2 Specifications for Electronically Enabled Passports with Biometric Identification Capabilities http://mrt.d.icao.int/content/view/33/202/
9..	INCITS 358	INCITS 358:2002 - American National Standard for Information Technology – The BioAPI Specification http://webstore.ansi.org/
10.	INCITS 378	INCITS 378:2004 - American National Standard for Information Technology — Finger Minutiae Format for Data Interchange http://webstore.ansi.org/
11.	INCITS 381	INCITS 381:2004 - American National Standard for Information Technology — Finger Image-

		Based Data Interchange Format. http://webstore.ansi.org/
12.	INCITS 398	INCITS 398:2008 - Common Biometric Exchange Formats Framework (CBEFF) http://webstore.ansi.org/
13.	INCITS 423.1	INCITS 423.1:2008 - Conformance testing Methodology Standard for Biometric Data Interchange Format Standards – Part 1: Generalized Conformance Testing Methodology http://webstore.ansi.org/
14.	INCITS 423.2	INCITS 423.2:2008 - Conformance testing Methodology Standard for Biometric Data Interchange Format Standards - Part 2: Conformance Testing, Finger Minutia http://webstore.ansi.org/
15.	INCITS 442	INCITS 442:2008 - Biometric Identity Assurance Services (BIAS) http://webstore.ansi.org/
16.	INT-I	ANSI/NIST-ITL 1-2000 Date Format for the Interchange of Fingerprint, Facial & SMT Information INTERPOL Implementation, Version No. 4.22b - October 28, 2005 http://www.interpol.int/Public/Forensic/fingerprints/RefDoc/implementation6.pdf
17.	ISO/IEC 15948	ISO/IEC 15948:2004 Computer graphics and image processing – Portable Network Graphics (PNG): Functional specification. http://webstore.ansi.org/
18.	ISO/IEC 19784-1	INCITS/ISO/IEC 19784-1:2006[2007] BioAPI – Biometric Application Programming Interface – Part 1: BioAPI Specification http://webstore.ansi.org/
19.	ISO/IEC 19784-5/Amdt 1	INCITS/ISO/IEC 19784- 1:2006/AM1 -2007 [2008], Information technology - BioAPI - Biometric Application Programming Interface - Part 1: BioAPI Specification - Amendment 1: BioGUI specification
20.	ISO/IEC 19784-2	INCITS/ISO/IEC 19784-2:2007[2008] Biometric Application Programming Interface (BioAPI) – Part 2: Biometric Archive Function Provider Interface http://webstore.ansi.org/
21.	ISO/IEC 19794-2	INCITS/ISO/IEC 19794-2:2005[2008] — Information technology — Biometric data interchange formats — Part 2: Finger minutiae data. http://webstore.ansi.org/ ISO/IEC 19794-2:2005/Cor.1:2007 — Information technology — Biometric data interchange formats — Part 2: Finger minutiae data – Technical Corrigendum 1
22.	ISO/IEC 19794-4	INCITS/ISO/IEC 19794-4:2005[2007] — Information technology — Biometric data interchange formats — Part 4: Finger image data. http://webstore.ansi.org/
23.	ISO/IEC 19794-5	INCITS/ISO/IEC 19794-5:2005[2007] — Information technology — Biometric data interchange formats — Part 5: Face image data. http://webstore.ansi.org/
24.	ISO/IEC 19794-5/Amdt 1	ISO/IEC 19794-5:2005/Amdt 1:2007 — Information Technology — Biometric Data Interchange Formats — Part 5: Face Image Data - Amendment 1 - Conditions for Taking Photographs for Face Image Data. http://webstore.ansi.org/
25.	ISO/IEC 19794-6	INCITS/ISO/IEC 19794-6:2005[2007] - Information technology — Biometric data interchange formats — Part 6: Iris image data. http://webstore.ansi.org/
26.	ISO/IEC 19795-1	INCITS/ISO/IEC 19795:2005[2007] - Biometric Performance Testing and Reporting – Part 1: Principles and Framework http://webstore.ansi.org/
27.	ISO/IEC 19795-2	ISO/IEC 19795:2006 - Biometric Performance Testing and Reporting – Part 2: Testing Methodologies for Technology and Scenario evaluations

		http://webstore.ansi.org/
28.	ISO/IEC 19795-3	ISO/IEC TR 19795:2007 - Biometric Performance Testing and Reporting – Part 3: Modality-Specific Testing http://webstore.ansi.org/
29.	ISO/IEC 19795-4	ISO/IEC 19795:2008 - Biometric Performance Testing and Reporting – Part 4: Interoperability Performance Testing http://webstore.ansi.org/
30.	ISO/IEC 24709-1	ISO/IEC 24709-1:2007 - Information technology -- Conformance testing for the biometric application programming interface (BioAPI) -- Part 1: Methods and procedures http://webstore.ansi.org/
31.	ISO/IEC 24709-2	ISO/IEC 24709-2:2007 - Information technology -- Conformance testing for the biometric application programming interface (BioAPI) -- Part 2: Test assertions for biometric service providers http://webstore.ansi.org/
32.	ISO/IEC 24713-1	ISO/IEC 24713-1:2008 - Biometric Profiles for Interoperability and Data Interchange – Part 1: Overview of Biometric Systems and Biometric Profiles http://webstore.ansi.org/
33.	IXM	Automated Biometric Identification System (IDENT) Exchange Messages (IXM) Specification – v2.0, September 7, 2007, IDENT-TO007-MAN-IXMTSP-004-D. http://www.biometrics.gov/Standards/Default.aspx
34.	JPEG 2000	ISO/IEC 15444-1:2004 - Information technology - JPEG 2000 image coding system - Part 1: Core coding system http://webstore.ansi.org/
35.	MINEX04	P. Grother et al., <i>Performance and Interoperability of the INCITS 378 Template</i> , NISTIR 7296 http://fingerprint.nist.gov/minex04/minex_report.pdf
36.	MITRE1000	Margaret Lepley, <i>Profile for 1000ppi Fingerprint compression</i> , Version 1.1 April 2004. Mitre Technical Report 04B0000022. http://www.fbibiospecs.org/fbibometric/docs/J2K1000.pdf
37.	NIST SP 800-76-1	NIST Special Publication 800-76-1, Biometric Data Specification for Personal Identity Verification, Revision 1, January 24, 2007 http://csrc.nist.gov/publications/nistpubs/800-76-1/SP800-76-1_012407.pdf
38.	RTIC	Registered Traveler Interoperability Consortium (RTIC), Technical Interoperability Specification, Version 1.7, April 15, 2008. http://www.rtconsortium.org/docpost/RTICTIGSpec_v1.7.pdf
39.	TWIC	TWIC Reader Hardware and Card Application Specification, September 11, 2007. http://www.tsa.gov/assets/pdf/twic_reader_card_app_spec_091107.pdf
40.	TWPDES	Terrorist Watchlist Person Data Exchange Standard, Version 1.2b. http://www.niem.gov/TWPDES.php
41.	WSQv3	WSQ Gray-Scale Fingerprint Image Compression Specification, IAFIS-IC-0110(V3), December 19, 1997. http://www.fbibiospecs.org/fbibometric/docs/WSQ_Gray-scale_Specification_Version_3.pdf

Multiple-Biometric Evaluation (MBE)
2010

**Report on the Evaluation of 2D
Still-Image Face Recognition
Algorithms**

NIST Interagency Report 7709

Patrick J. Grother, George W. Quinn and P. Jonathon Phillips

Image Group
Information Access Division
Information Technology Laboratory
National Institute of Standards and Technology



June 22, 2010

EXECUTIVE SUMMARY

Background

- Facial recognition algorithms from seven commercial providers, and three universities, were tested on one laboratory dataset and two operational face recognition datasets, one comprised of visa images, the other law enforcement mugshots. The population represented in these sets approaches 4 million, such that this report documents the largest public evaluation of face recognition technology to date. The project attracted participation from a majority of the known providers of FR technology including the largest commercial suppliers.
- Accuracy was measured for three applications: One-to-one verification (e.g. of e-passport holders); one-to-one verification against a claimed identity in an enrolled database (e.g. for driver's license re-issuance); and one-to-many search (e.g. for criminal identification or driver's license duplicate detection).
- Face images have been collected in law enforcement for more than a century, but their value for automated identification remains secondary to fingerprints. In a criminal investigation setting, face recognition has been used both in an automated mode and for forensic investigation. However, the limits of the technology have not previously been quantified publicly, and, in any case, are subject to improvement over time, and to the properties of the images in use.
- Core algorithmic capability is the major contributor to application-level recognition outcomes. A second critical factor is the quality of the input images; this is influenced by design of, and adherence to, image capture protocols (as codified by face recognition standards) and also by the behavior of the person being photographed (e.g. whether they face the camera). Some data collection protocols can embed a human adjudication of quality (e.g. of a visa image by a consular official) while others cannot maintain such tight quality controls (e.g. because of non-cooperative subjects in police booking processes).
- This is the first time NIST has reported accuracy of face *identification* algorithms. Prior tests have assumed an equivalence of a 1:N search as N 1:1 comparisons. This new protocol formally supports use of fast search algorithms such as indexing, partitioning and binning. The benefits are more accurate predictions of scalability to national-size populations.
- The project used archival imagery to assess core algorithmic capability of algorithms. It did not do an instrumented collection of images as might be used in a scenario or operational test. It therefore did not measure human-camera transactional performance parameters such as duration of use and outcome. These would be of vital interest in, for example, e-Passport gate applications.

Core Accuracy

- As with other biometrics, recognition accuracy depends strongly on the provider of the core technology. Broadly, there is an order of magnitude between the best and worst identification error rates.
- Biometric identification algorithms return candidate lists. These enumerate hypothesized identities for a search sample. Face identification algorithms can be set up to be used in two distinct modes. The first, *investigational* mode, assumes the existence of a corps of human face examiners retained to examine perhaps dozens of images on candidate lists. In the second, *identification* mode, the algorithm is set up with a high threshold to give very short candidate lists and a small chance that a non-matching candidate is returned. The most accurate investigational algorithms are not the most accurate identification algorithms.
- Using the most accurate face recognition algorithm, the chance of identifying the unknown subject (at rank 1) in a database of 1.6 million criminal records is about 92%. For other population sizes, this accuracy rate decreases linearly with the logarithm of the population size. In all cases a secondary (human) adjudication process will be necessary to verify that the top-rank hit is indeed that hypothesized by the system.
- When the most accurate algorithm is used in an investigational mode to provide trained examiners with the top fifty ranked candidates 97% of searches will yield the correct identity in a fixed population of 1.6 million subjects. In cases where the top 200 candidates are searched, the correct match is present 97.5% of the time. The hit rate increases roughly linearly with the log of the number of candidates inspected.
- In criminal law enforcement applications, where recidivism rates are high and a pool of examiners is available to traverse lengthy candidate lists, facial recognition algorithms offer high success rates. The more accurate

MBE-STILL REPORT PARTICIPANT KEY	P = PITTPATT	R = SURREY U.	S = TSINGHUA U.	T = TOSHIBA	U = DALIAN TECH U.	PAGE 2 OF 58
	V = NEC	W = L1 IDENTITY	X = COGNITEC	Y = SAGEM	Z = NEUROTECHNOLOGY	

algorithms reduce the workload on the examiner by placing mates at low (i.e. good) rank. We define an overall performance metric as the expected number of candidates an examiner will need to compare before the mate is found. If the most accurate algorithm is used for identification in the population of 1.6M, an examiner willing to review 50 candidates will only need to look at 3 on average before the mate is found.

- A facial recognition algorithm can also be used to do “lights-out” face identification. This requires application of a high decision threshold that implements a selectivity policy. Here a threshold is adopted that gives, on the average, a particular number of false candidates per search. For the most accurate face recognition algorithm tested here, if one in two searches produces a false candidate on average, the hit rate will be 89%. If workload demands on human adjudicators require that only one in ten searches produce a false candidate, the hit rate reduces to 85% and a different algorithm is best in this regime. These numbers apply to a population of 1.6M. The threshold will need to be estimated over a calibration process. This threshold will need to be increased as the enrolled population increases.
- Facial recognition algorithms are more accurate on the visa images than the mug shot images. The visa images were collected c. 1996-2001. The imaging processes used for their collection have improved since that time. The mug shot images are contemporary, and operationally representative of current law-enforcement collection practices. On these images, the face recognition accuracy results reported here will be closely predictive of those that would be encountered in any near term deployment.
- The visa images are collected with careful cooperation of the subject, active compliance by the photographer to the image collection specification, and a yes/no review by an official. The visa images are subject to losses associated with JPEG compression. The mugshot images, while less compressed and of generally higher resolution, exhibit considerable pose, illumination and expression variation. The most accurate algorithm demonstrates better tolerance of non-frontal pose than others.
- On the one database used in 2002, 2006 and 2010, the best verification accuracy measurement has declined by an order of magnitude in each four year period. On the visa images, false non-match rates (at a fixed false match rate of 0.001) have reduced from 0.2 in 2002, to 0.026 in 2006, and to 0.003 now. This result is achieved on a dataset that has various deviations from formal standards and best practices.

Exploitation of all historical encounters

- The test was executed using an Application Programming Interface (API) that supported identification of an image against all prior images of a subject, not just the most recent. This allowed the face recognition algorithm developers to exploit the historical record. It also assigned responsibility for fusion to the algorithm developers, who could implement early-stage template-level fusion or the simpler late stage score-level fusion.
- All recognition algorithms derive accuracy improvements when all past images are enrolled as a single template. The benefits are uniform across algorithms. The template size for a person enrolled with K images is, for all algorithms tested, closely K times the template size produced from a single image. These two facts suggest that common and simple techniques are sufficient to realize the available gain.

Speed and template size

- For the first time, this NIST evaluation measures and reports the speed of face recognition algorithms. The main result is atypical in biometrics: The most accurate algorithms are among the fastest. This departs from observations in fingerprint and iris trials that showed an industry-wide tradeoff between accuracy and computational expense.
- For search algorithms from the two most accurate providers, the time required to execute a one-to-many search against an enrolled population of 1.6 million people is 0.4 and 1.2 seconds respectively. This is the duration of the core search computation as measured on contemporary high-end yet standard hardware consisting of 16 computational cores, 192GB of main memory, and a 64 bit address space. It assumes that a search template has been prepared and transmitted to the matching engine.
- In most cases the time required to execute a search does not scale linearly with the size of the enrolled population. While the most accurate algorithm does scale linearly, the second most accurate algorithm scales such that a ten-fold increase in database size produces only a 1.3-fold increase in search duration. This behavior has been confirmed on sizes up to 1.6 million.

MBE-STILL REPORT PARTICIPANT KEY	P = PITTPATT	R = SURREY U.	S = TSINGHUA U.	T = TOSHIBA	U = DALIAN TECH U.	PAGE 3 OF 58
	V = NEC	W = L1 IDENTITY	X = COGNITEC	Y = SAGEM	Z = NEUROTECHNOLOGY	

- Several participants elected to provide several implementations for evaluation. In so doing the provider demonstrated an ability to trade accuracy for speed, and to use large or small template sizes. This suggests a valuable ability to parameterize their algorithms to meet computational and accuracy requirements.
- The variance in search times is small. This arises because all search templates from a particular recognition algorithm have the same size. Across the algorithms tested here, template sizes range from about 5 to 75 kilobytes. By comparison, 90% of the law enforcement JPEG images used here are in the range 4 to 380 kilobytes, with median, 36 kilobytes.

Accuracy dependence on biographic data

- For the law enforcement images, it is empirically observed that men are more easily recognized than women, that heavier individuals are more easily recognized than lighter subjects, and that Asian subjects are more easily recognized than White. Younger persons are more difficult to recognize than their elders for some recognition algorithms, but the opposite is true for others.
- These results state marginal observations for the particular dataset. They do not explain the cause of the observation because there are confounding aspects to the data. So while men are more readily recognized than women, this may arise because women are generally shorter than men, and the height of a subject may induce non-optimal imaging angle if the camera height is not adjusted.

MBE-STILL REPORT PARTICIPANT KEY	P = PITTPATT	R = SURREY U.	S = TSINGHUA U.	T = TOSHIBA	U = DALIAN TECH U.	PAGE 4 OF 58
	V = NEC	W = L1 IDENTITY	X = COGNITEC	Y = SAGEM	Z = NEUROTECHNOLOGY	

ABSTRACT

The paper evaluates state-of-the-art face identification and verification algorithms, by applying them to corpora of face images the population of which extends into the millions. Performance is stated in terms of core accuracy and speed metrics, and the dependence of these on population size and image properties are reported. One-to-many search algorithms are evaluated in terms of their use in both investigational and identification modes. Investigational performance has implications for workload on an examiner reviewing the results of a search. Identification performance, using a high score threshold, can support fully automated operation and decision making if some quantified level of false match is acceptable. In addition, the paper establishes an initial approach toward calibration of false match accuracy.

ACKNOWLEDGEMENTS

- The authors wish to thank Federal Bureau of Investigation for their support of this work.
- In addition, we appreciate Michael Garris’ direction and tight coordination, and for his review of this document.
- In addition, NIST is indebted to Nick Orlans and the MITRE-led teams responsible for the intensive effort coordinating preparation of the public MEDS and private Photo-File image corpora.
- The authors thank Craig Watson, Brian Cochran and Wayne Salamon at NIST for their herculean and timely efforts to stand up the computers, power, air conditioning and software used to run the MBE-STILL trials.
- The authors thank Jay Scallan for review of the images.
- The authors are grateful to the experts who made comments on the drafts of the MBE-STILL Concept, Evaluation Plan and API document¹.
- Finally, the authors acknowledge the diligent work of the developers in implementing and supporting the MBE-protocol.

KEYWORDS

Face recognition; biometrics; verification; identification; recognition; identity management; watch-list; pattern recognition; reliability; scalability; calibration; mugshot.

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.

¹ See <http://face.nist.gov/mbe/>

MBE-STILL REPORT PARTICIPANT KEY	P = PITTPATT	R = SURREY U.	S = TSINGHUA U.	T = TOSHIBA	U = DALIAN TECH U.	PAGE 5 OF 58
	V = NEC	W = L1 IDENTITY	X = COGNITEC	Y = SAGEM	Z = NEUROTECHNOLOGY	

TIMELINE OF THE MBE-STILL EVALUATION

Date	Activity
June 8, 2010	Release of the first draft of the public MBE-STILL report.
May 14, 2010	Window for submission of FR implementations to NIST closes
February 28, 2010	First FR implementations arrive at NIST
February 1, 2010	Release of the final Still Face Image Track - Concept, Evaluation Plan and API Version 1.0.0
January 27, 2010	Window for submission of FR implementations to NIST opens
December 15, 2009	Release of sample data: http://face.nist.gov/mbe/NIST_SD32v01_MEDS_1_face.zip
December 09, 2009	Second draft evaluation plan (revised version of this document) for public comment.
November 16, 2009	Initial draft evaluation plan circulated for public comment.
July 29, 2009	Project Initiation: Briefing to the FBI, <i>Face Recognition Testing Tailored to the FBI Application</i> , Patrick Grother, NIST.

October 2009							November 2009							December 2009							January 2010							February 2010							
Su	Mo	Tu	We	Th	Fr	Sa	Su	Mo	Tu	We	Th	Fr	Sa	Su	Mo	Tu	We	Th	Fr	Sa	Su	Mo	Tu	We	Th	Fr	Sa	Su	Mo	Tu	We	Th	Fr	Sa	
				1	2	3	1	2	3	4	5	6	7			1	2	3	4	5						1	2			1	2	3	4	5	6
4	5	6	7	8	9	10	8	9	10	11	12	13	14	6	7	8	9	10	11	12	3	4	5	6	7	8	9	7	8	9	10	11	12	13	
11	12	13	14	15	16	17	15	16	17	18	19	20	21	13	14	15	16	17	18	19	10	11	12	13	14	15	16	14	15	16	17	18	19	20	
18	19	20	21	22	23	24	22	23	24	25	26	27	28	20	21	22	23	24	25	26	17	18	19	20	21	22	23	21	22	23	24	25	26	27	
25	26	27	28	29	30	31	29	30						27	28	29	30	31			24	25	26	27	28	29	30	28							
March 2010							April 2010							May 2010							June 2010							July 2010							
Su	Mo	Tu	We	Th	Fr	Sa	Su	Mo	Tu	We	Th	Fr	Sa	Su	Mo	Tu	We	Th	Fr	Sa	Su	Mo	Tu	We	Th	Fr	Sa	Su	Mo	Tu	We	Th	Fr	Sa	
7	8	9	10	11	12	13	4	5	6	7	8	9	10	2	3	4	5	6	7	8	6	7	8	9	10	11	12	4	5	6	7	8	9	10	
14	15	16	17	18	19	20	11	12	13	14	15	16	17	9	10	11	12	13	14	15	13	14	15	16	17	18	19	11	12	13	14	15	16	17	
21	22	23	24	25	26	27	18	19	20	21	22	23	24	16	17	18	19	20	21	22	20	21	22	23	24	25	26	18	19	20	21	22	23	24	
28	29	30	31				25	26	27	28	29	30		23	24	25	26	27	28	29	27	28	29	30				25	26	27	28	29	30	31	
														30	31																				

	Test and API Development	Test Execution	Analysis and Reporting	
--	--------------------------	----------------	------------------------	--

VERSION HISTORY

Date	Activity
June 22, 2010	1/ Improved reporting of class B vs. class A results in INVESTIGATION 8. 2/ Added results for Y04 to pose/sex/age in Figure 20 and onwards. 3/ Replaced boxplots for FNMR by sex with tabulated values. 4/ Added verification results for R00, W10, W11 to Figure 12. 5/ Added tabulated values to graphs showing effect of population size, and effect of rank.
June 18, 2010	1/ Fixed incorrect identification of Dalian University of Technology 2/ Updated Figure 16 – LEO Selectivity by number of prior encounters to include more class C algorithms and all 1000 bootstrap estimates of selectivity. 3/ Added result for Y04 to verification results in Figures 12 and 15.
June 16, 2010	First publication of this document, NISTIR 7709

TABLE OF CONTENTS

1. MBE-STILL Goals and Objectives 10

 1.1. MBE Context 10

 1.2. Market drivers 10

 1.3. Application scenarios 11

 1.4. Offline testing 12

2. Participation 12

3. Datasets 12

 3.1. Sizes of datasets 13

 3.2. Public sample images 14

4. Metrics 15

 4.1. Verification 15

 4.2. Identification 15

 4.3. Failure to acquire 17

5. Properties of the implementations 18

6. Results 20

 INVESTIGATION 1. Investigation-mode one-to-many search accuracy 20

 INVESTIGATION 2. Identification-mode one-to-many search accuracy 22

 INVESTIGATION 3. Dependence on population size 24

 INVESTIGATION 4. Dependence on rank 25

 INVESTIGATION 5. Impostor distribution stability 29

 INVESTIGATION 6. Search duration 31

 INVESTIGATION 7. Verification accuracy 33

 INVESTIGATION 8. Verification accuracy with and without an enrollment database 34

 INVESTIGATION 9. Exploiting all prior images 35

 INVESTIGATION 10. Exploiting all prior images: A false match hazard? 38

 INVESTIGATION 11. Evidentiary value 39

 INVESTIGATION 12. Dependence of accuracy on pose 42

 INVESTIGATION 13. Template size 45

 INVESTIGATION 14. Template creation time 47

 INVESTIGATION 15. Link between sex and accuracy 48

 INVESTIGATION 16. Link between subject age and accuracy 48

 INVESTIGATION 17. Face ageing 49

 INVESTIGATION 18. Is subject weight influential? 51

 INVESTIGATION 20. Value of biographic data 53

7. Progress in face recognition 55

8. References 56

 8.1. Publications and Reports 56

 8.2. Standards 58

LIST OF FIGURES

Figure 1 – Organization and documentation of the MBE 10

Figure 2 – Examples of law enforcement images (I) 14

Figure 3 – Examples of law enforcement images (II) 14

Figure 4 - Examples of law enforcement images (III) 15

Figure 5 – Identification accuracy vs. candidate rank 21

Figure 6 – Identification rate vs. selectivity 23

Figure 7 – LEO Identification accuracy dependence on population size 25

Figure 8 – LEO identification miss rate versus rank 27

Figure 9 – Workload implications of LEO cumulative match performance 29

Figure 10 - Reliability and selectivity at a fixed threshold 30

Figure 11 – Duration of LEO identification searches 32

MBE-STILL REPORT PARTICIPANT KEY	P = PITTPATT	R = SURREY U.	S = TSINGHUA U.	T = TOSHIBA	U = DALIAN TECH U.	PAGE 7 OF 58
	V = NEC	W = L1 IDENTITY	X = COGNITEC	Y = SAGEM	Z = NEUROTECHNOLOGY	

Figure 12 – Verification accuracies of class A algorithms.....33
 Figure 13 – LEO Verification accuracy with and without enrollment datasets.....35
 Figure 14 – LEO Identification accuracy by number of prior encounters.....37
 Figure 15 – LEO Verification accuracy with and without multiple enrollment samples 38
 Figure 16 – LEO Selectivity by number of prior encounters 39
 Figure 17 – LEO False Match Rate calibration curves..... 41
 Figure 18 – Face images from the FERET database demonstrating varying amounts of head yaw42
 Figure 19 – Histograms for yaw and roll angles for LEO images.....43
 Figure 20 – Dependence of accuracy on face yaw angle.....43
 Figure 21 – Dependence of LEO accuracy on yaw angle of enrollment and verification images 44
 Figure 22 – Dependence of LEO accuracy on reported roll angle..... 45
 Figure 23 – Duration of LEO template generation calls.....47
 Figure 25 – LEO Verification accuracy by age of subject at most recent capture 49
 Figure 26 – LEO Verification accuracy by time elapsed between photographs..... 50
 Figure 27 – LEO Verification accuracy by subject weight52
 Figure 28 – LEO Verification accuracy by ethnic category.....53
 Figure 29 – Progression of face recognition accuracy measurements 54

LIST OF TABLES

Table 1 – Abbreviations 9
 Table 2 – Biometric identification applications11
 Table 3 – Subtests supported under the MBE still-face activity11
 Table 4 – MBE-STILL Face Recognition Technology Providers..... 12
 Table 5 – Image dataset descriptions..... 13
 Table 6 – Image dataset sizes 13
 Table 7 -- Definition of False Non-match Rate 15
 Table 8 -- Definition of False Match Rate 15
 Table 9 – Verification Performance characteristics..... 15
 Table 10 – Definition of True Positive Identification Rate 16
 Table 11 – Definition of False Positive Identification Rate 16
 Table 12 – Definition of Reliability 16
 Table 13 – Definition of Selectivity 16
 Table 14 – Definitions of Type I error rates..... 16
 Table 15 – Definitions of Type II error rates..... 17
 Table 16 – Identification Performance characteristics 17
 Table 17 – Test design considerations 18
 Table 18 – Processing time limits in milliseconds 31
 Table 19 – Adjustment of search duration estimates by number of cores used..... 31
 Table 20 – Uses of K images of a MEDS dataset subject for testing..... 36
 Table 21 – Interpretation of impostor scores 40
 Table 22 – On-disk template sizes by SDK and template role 46
 Table 23 -- LEO Verification accuracy by sex..... 48

MBE-STILL REPORT PARTICIPANT KEY	P = PITTPATT	R = SURREY U.	S = TSINGHUA U.	T = TOSHIBA	U = DALIAN TECH U.	PAGE 8 OF 58
	V = NEC	W = L1 IDENTITY	X = COGNITEC	Y = SAGEM	Z = NEUROTECHNOLOGY	

TERMS AND DEFINITIONS

The abbreviations and acronyms of Table 1 are used in many parts of this document.

Table 1 – Abbreviations

FR	Face Recognition
MBE	NIST's Multiple Biometric Evaluation program
MBE-STILL	The track of the MBE concerned with recognition of 2D still images.
TPIR	True positive identification rate
FNIR	False negative identification rate
FPIR	False positive identification rate
FMR	False match rate
FNMR	False non-match rate
FTE	Failure to Enroll, also Failure to Enroll Rate.
Reliability	A Type I error rate expressing hit or miss rate.
Selectivity	A Type II error rate expressing false positive errors
DET	Detection error tradeoff characteristic: For verification this is a plot of FNMR vs. FMR (sometimes as normal deviates, sometimes on log-scales). For identification this is a plot of FNIR vs. FPIR.
ROC	Receiver Operating Characteristic
CMC	Cumulative Match Characteristics
SC 37	Subcommittee 37 of Joint Technical Committee 1 – developer of biometric standards
INCITS	InterNational Committee on Information Technology Standards
ISO/IEC 19794	ISO/IEC 19794-5: Information technology — Biometric data interchange formats — Part 5:Face image data. First edition: 2005-06-15. (See Bibliography entry).
I385	INCITS 385:2004 - U.S. precursor to the 19794-5 international standard
ANSI/NIST Type 10	The dominant container for facial images in the law enforcement world.
MEDS	Multiple Encounter Deceased Subjects
NIST	National Institute of Standards and Technology
PIV	Personal Identity Verification
SDK	The term Software Development Kit refers to any library software submitted to NIST. This is used synonymously with the terms "implementation" and "implementation under test".

1. MBE-STILL Goals and Objectives

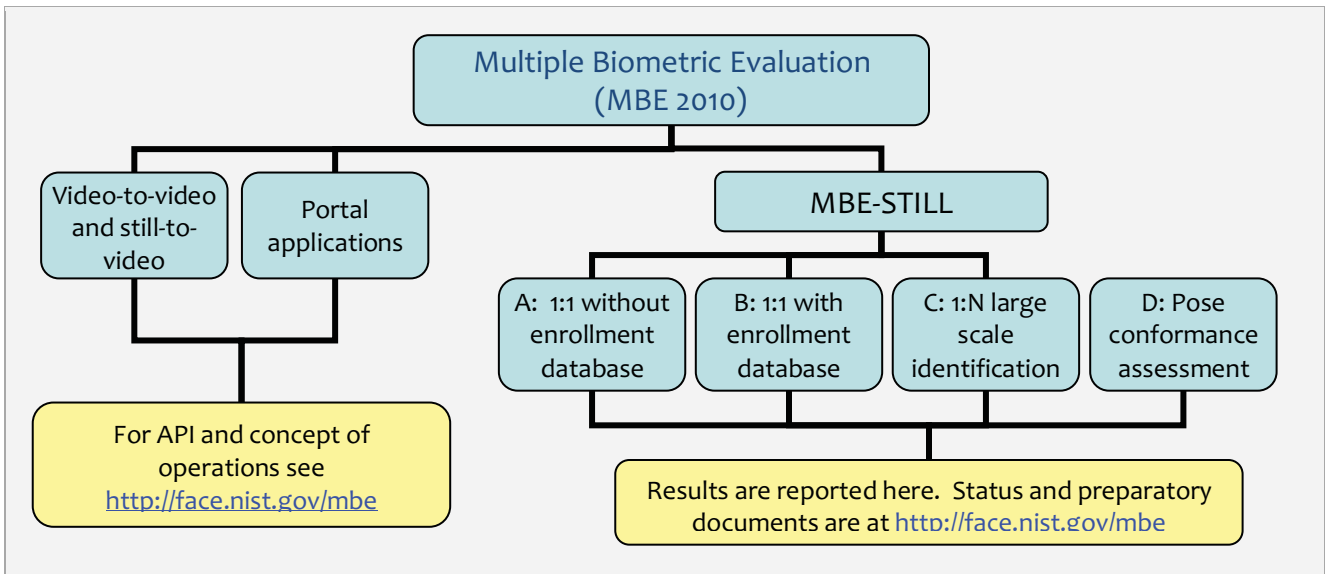
Initiated in summer 2009, the Multi-biometric 2D Still-Face Recognition evaluation was undertaken with the following objectives.

- To respond to governmental and commercial requests to assess contemporary facial recognition (FR) implementations.
- To leverage massive operational corpora. The availability of images from large populations (in the millions) ensures statistical significance of all studies, particularly across demographic groups. The use of operational images brings greater operational relevance to the test results.
- To evaluate face recognition technologies in a proper one-to-many identification mode. This departs from many prior evaluations in which 1:N search accuracy was simulated via computation of N 1:1 comparisons².
- To report parameters important to implementers and procurers. These include template size and processing times.

1.1. MBE Context

The still-face recognition track is a standalone part of the larger MBE parent program. As depicted in Figure 1, MBE also includes tracks for recognition-from-video and face-iris portals. See <http://face.nist.gov/mbe> for the status of all MBE activities.

Figure 1 – Organization and documentation of the MBE



1.2. Market drivers

This test is intended to support a plural marketplace of face recognition systems. While the dominant application, in terms of revenue, has been one-to-many search for driving licenses and visa issuance, the deployment of one-to-one face recognition has re-emerged with the advent of the e-Passport verification projects³. In addition, there remains

² NIST has previously only modeled identification scenarios. The simplest simulation mimics a 1:N search by conducting N 1:1 comparisons.

³ These match images acquired from a person crossing a border against the ISO/IEC 19794-5 facial image stored on the embedded ISO/IEC 7816 + ISO/IEC ISO 14443 chips. Such systems are fielded in Portugal (RAPID), Australia (SmartGate), Germany (EasyPASS) [NUPPENY], the United Kingdom (UKBA) and elsewhere. Such systems can viably establish a low false acceptance policy because traditional immigration processes are available to those rejected. Accuracy is dependent on both the quality of the standardized e-Passport image, and on the images produced by the automated access-control gate. The gate design typically includes supplemental lighting, and some mechanism to adjust for the height of the traveler.

considerable activity in the use of face recognition in a number of identification applications. The list given in Table 2 differentiates applications by population size and the kinds of images used.

Table 2 – Biometric identification applications

#	Application	Open set	Coop	Typical population size	Kind of reference or enrolled image	Search image
1	De-duplication e.g. for benefits fraud detection	Y	C	Starting at zero, increasing to millions, as database grows	Often collected for a credential, e.g. a visa or driver's license. Ideally ISO/IEC 19794-5 compliant.	Same as reference
2	Web-search. Social networking consolidation.	Y	N	Millions	Zero or more images existing on various web pages. Usually uncontrolled.	Also uncontrolled.
3	Forward criminal search	Y	CNU	Millions	Mugshot collected incident to an arrest (i.e. more or less compliant to ANSI/NIST Type 10 mugshot standards)	Usually the same properties as the reference
4	Reverse criminal search; unsolved photo file as defined in [EBTS].	Y	CNU	Tens of thousands	The photograph of a person associated with an adverse event.	Mugshot.
5	Watch-list, covert surveillance.	Y	N	Tens of thousands	Varied. Sometimes controlled but often adverse.	Varied, often dissimilar to the reference
6	Access-control without presentation of a credential or PIN	Y	C	Thousands	Attended enrollment in good conditions. Ideally ISO/IEC 19794-5 compliant.	Similar to reference, but with relaxed imaging constraints
7	Cruise-ship	N	C	Thousands	Controlled photo collected at time of boarding.	
8	Disaster post-mortem	N Y	N	Hundreds, thousands	Varied.	Varied.
9	Column 4 gives cooperation of the subjects: C = cooperative, N = non-cooperative, U = actively uncooperative or evasive.					

1.3. Application scenarios

The MBE-STILL activity includes one-to-one verification and one-to-many identification tests, as described in Table 3. Class A might be preferred by academic institutions because the API supports the elemental hypothesis testing verification function: "Are the images from the same person or not?"

Table 3 – Subtests supported under the MBE still-face activity

	Class A	Class B	Class C	Class D
Application area	1:1 verification without an enrollment database	1:1 verification with an enrollment database	1:N identification	Pose calibration
Description	Verification scenarios in which still images are compared.	Verification scenarios in which images are compared with entries in an enrolled database.	Close-to-operational use of face recognition technologies in identification applications in which the enrolled dataset could contain images from up to three million persons.	Assess whether the orientation of the head meets frontal imaging pose specifications.
Required	Yes, all participants must submit algorithms.	Optional	Optional	Optional
Participation	10	7	8	3

1.4. Offline testing

While this set of tests is intended as much as possible to mimic operational reality, this remains an offline test executed on databases of images. The intent is to assess the core algorithmic capability of face recognition algorithms. This test was conducted purely offline. That is, it did not include a live human-presents-to-camera component. Offline testing is attractive because it allows uniform, fair, repeatable, and efficient evaluation of the underlying technologies. Testing of implementations under a fixed API allows for a detailed set of performance related parameters to be measured.

2. Participation

The MBE program was open to participation worldwide. There were no requirements for entry, other than an ability to implement the interface protocol specifications. In the case of MBE-STILL, this requires conformance to a “C” language API which in turn requires software engineering skills associated with technology developers and researchers. As with all NIST technology evaluations, NIST did not charge a fee to participate.

This report documents SDK-based implementations submitted during a window of participation which ran from January 27, 2010 through May 14, 2010. The participants are tabulated in Table 4.

Table 4 – MBE-STILL Face Recognition Technology Providers

#	Organization	Code	Class A: One-to-one verification	Class B: One-to-one verification with enrollment db	Class C: One-to-many identification	Class D: Pose conformance
1.	Cognitec	X	Yes	Yes	Yes	
2.	Dalian University of Technology	U	Yes			
3.	L1 Identity Solutions	W	Yes	Yes	Yes	
4.	NEC	V	Yes	Yes	Yes	
5.	Neurotechnology	Z	Yes	Yes	Yes	
6.	Pittsburg Pattern Recognition	P	Yes	Yes	Yes	Yes
7.	Sagem	Y	Yes		Yes	Yes
8.	Surrey University	R	Yes			
9.	Toshiba	T	Yes	Yes	Yes	
10.	Tsinghua University	S	Yes	Yes	Yes	Yes

3. Datasets

This report documents the use of four datasets.

- **LEO:** The primary dataset consists of facial images collected by various law enforcement (LEO) agencies and transmitted to the FBI as part of various criminal records checks. This is known as the FBI Photo File.
- **DOS / Natural:** The secondary dataset consists of non-immigrant visa images. It is used here for one-to-many identification purposes.
- **DOS / HCINT:** This extract of DOS / Natural was used in the FRVT 2002 evaluation [FRVT2002], and subsequently the FRVT 2006 follow-on [FRVT 2006].
- **SANDIA:** A set of high resolution frontal-face images used as the high resolution dataset in the FRVT 2006 evaluation [FRVT2006]. The Sandia dataset was collected at the Sandia National Laboratory. Enrollment face images were collected with controlled illumination with cooperation from the subjects. Verification images were collected in two modes: First, with controlled illumination, and second without it. The Sandia images were taken with a 4 Megapixel Canon PowerShot G2.

The properties are summarized in Table 5.

MBE-STILL REPORT PARTICIPANT KEY	P = PITTPATT	R = SURREY U.	S = TSINGHUA U.	T = TOSHIBA	U = DALIAN TECH U.	PAGE 12 OF 58
	V = NEC	W = L1 IDENTITY	X = COGNITEC	Y = SAGEM	Z = NEUROTECHNOLOGY	

Table 5 – Image dataset descriptions

Property	LEO	DOS / Natural	DOS / HCINT	Sandia
Collection, environment	Law enforcement booking	Visa application process	Visa application process	Dedicated laboratory collection.
Collection era	~ 1960s-2008	~ 1996-2002	~ 1996-2002	
Live scan, Paper	Live, few paper	Mostly live	Mostly live	Live
Documentation	See NIST Special Database 32 Volume 1, available 12/09 ⁴ .		See NIST IR 6965 [FRVT2002]	See NISR IR 7408 [FRVT 2006]
Image size	Various, 480x640, 240x240, 768x960	Most 300 x 252	Most 300 x 252	
Compression	JPEG ~ 20:1	JPEG	JPEG mean size 9467 bytes. See [FRVT2002b]	JPEG, very little compression
Eye to eye distance	mean=156, sd=46	Median = 71 pixels	Median = 71 pixels	Controlled, mean = 350 Uncontrolled, mean = 110
Frontal	Moderate control. Known profile images excluded.	Yes, well controlled	Yes, well controlled	Controlled: yes Uncontrolled: yes
Full frontal geometry	Mostly not. Varying amounts of the torso are visible.	Yes, in most cases. Faces are more cropped (i.e. smaller background) than ISO FF requires.	Yes, in most cases. Faces are more cropped (i.e. smaller background) than ISO FF requires.	
Use in MBE-STILL	1:1 and 1:N	1:N	1:1	1:1
Parent	Operational data	Operational data	This is an extract of DOS / Natural persons 18+ years with 3 or more images. Introduces selection bias toward young men.	Self

3.1. Sizes of datasets

The databases are characterized by population sizes well in excess of all published biometric tests. The numbers of subjects and images are given in Table 6.

Table 6 – Image dataset sizes

#	Quantity	LEO	DOS / Natural	DOS / HCINT	Sandia
1.	Number of subjects	1802874	5738141	37440	263
2.	Num. subjects with 1 images	1428308	5294708	0	0
3.	Num. subjects with 2 images	253564	388975	0	0
4.	Num. subjects with 3 images	69527	44539	30701	2
5.	Num. subjects with 4 images	26509	7554	5069	12
6.	Num. subjects with 5 images	11659	1647	1111	14
7.	Num. subjects with 6 images	5825	477	376	1
8.	Num. subjects with 7 or more	7482 (Person max 26)	201 (Person max 13)	193 (Person max 13)	234
9.	Num. images in total	2407768	6249392	121589	13854
10.	Num. subjects used	1800000, selected randomly from line 1.	1850000	36000 in 12 partitions of 3000 subjects.	263
11.	Num. subjects used only as impostor / non-mates	200000, selected randomly from line 10.	50000	3000 from partition $k+1 \text{ mod } 12$	0

⁴ NIST Special Database 32, Volume 1, is available at: http://face.nist.gov/mbe/NIST_SD32v01_MEDS_I_face.zip. This link is temporary. The database will ultimately be linked from <http://face.nist.gov/mbe>.

12.	Num. images used only as impostor / non-mates	200000, last image of subjects selected randomly	50000	6000, two images used separately	0
13.	Num. subjects used in enrollment processes	1600000, selected randomly from line 10.	1800000	3000 per partition	263
14.	Num. images used in enrollment processes	1816170	1816743	3000 per partition	3404
15.	Num. images excluded	Profile, corrupt JPEG,	0	9, corrupt similarity files in FRVT 2002.	

3.2. Public sample images

NIST released the MEDS dataset in January 2010. MEDS stands for Multiple Encounter Deceased Subjects. The MEDS dataset is a representative and public sample of the non-public LEO set used in MBE-STILL. Specific examples are shown in Figure 2, Figure 3 and Figure 4 with respective commentaries.

Figure 2 – Examples of law enforcement images (I)




		
(a)	(b)	(c)
<p>Image (a) is about as conformant to facial recognition image standards as the law enforcement images get. The remaining images shown here are grossly non-conformant. Image (b) has a pitch angle that is likely fatal to automated facial recognition algorithms, lens distortion associated with the camera being too close to the subject, poor uniformity of illumination and low contrast. Image (b) and (c) have image dimensions 240x240 indicative of capture using a webcam. Such images can originate in non-traditional law-enforcement sites, such as at a border crossing.</p>		

Figure 3 – Examples of law enforcement images (II)


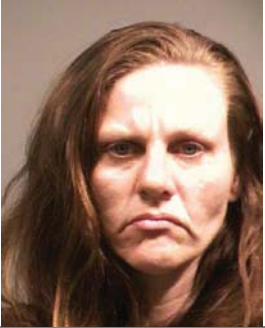

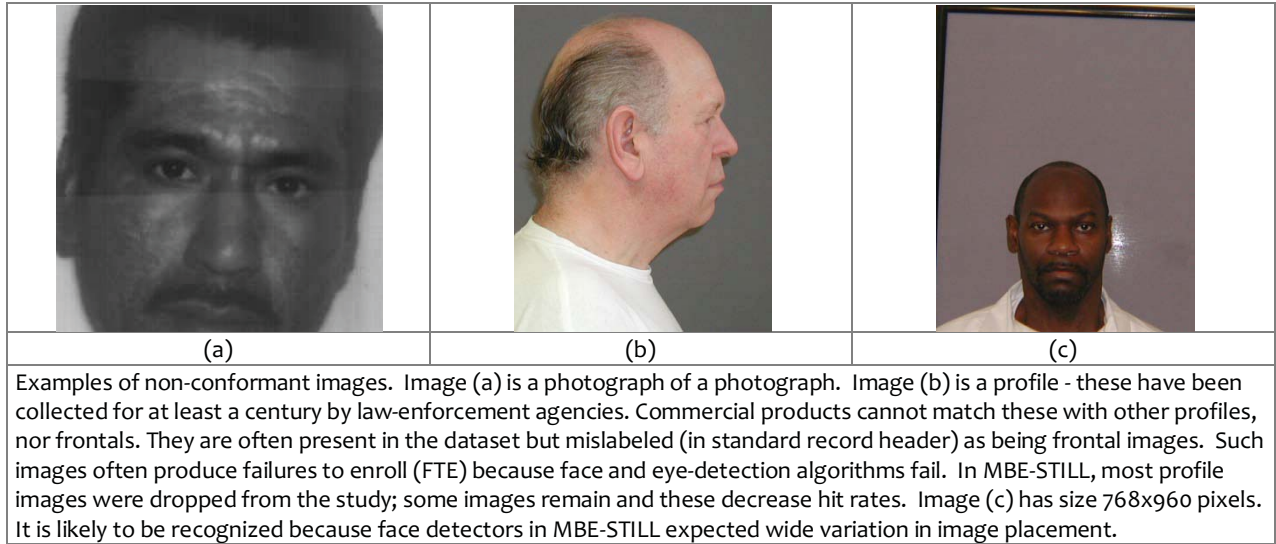
		
(a)	(b)	(c)
<p>Three images of one person from the MEDS dataset. All three images are of size 480x600 pixels. All depart from defined standards: In image (a) the subject's pitch angle is slightly too high; in image (c) the yaw angle is well beyond the 5 degrees limit established in standards; in image (b) there is considerable saturation and non-uniformity of the lighting</p>		

Figure 4 - Examples of law enforcement images (III)



4. Metrics

4.1. Verification

The fundamental matching error rates are defined in Table 7 and Table 8.

Table 7 -- Definition of False Non-match Rate

FNMR(T)	=	Number of genuine comparisons with similarity score less than threshold, T.	Equation 1
		Number of genuine comparisons attempted	

Table 8 -- Definition of False Match Rate

FMR(T)	=	Number of impostor comparisons that produce a score greater than or equal to threshold, T.	Equation 2
		Number of impostor comparisons attempted	

These are plotted as a DET characteristic.

Table 9 – Verification Performance characteristics

Metric	Measured over	Definition
DET	Searches with and without mates	The receiver operating characteristic is a plot of FNMR(T) vs. FMR(T), where T is any real-valued threshold.

4.2. Identification

MBE-STILL tested only open-set identification algorithms. This means that some searches have no enrolled mate. This is operationally typical: some subjects have not been issued a visa or drivers license before; some law enforcement searches are from first-time offenders. Searches for these people should return zero identities.

Open-set applications require estimation of two error rates: Type I errors are those in which a person's biometric data is incorrectly not associated with its enrolled mate; Type II errors are those in which a person's biometric data is

associated with other enrollees' data. Table 14 defines metrics for Type I identification errors used in this report, and notes various synonyms and complementary terms.

Table 15 defines metrics for Type II errors.

In a closed-set application, all searches have an enrolled mate. Operationally closed-universe applications are rare. One example is a cruise ship in which all passengers are enrolled and all searches should produce one, and only one, identity. Another example is forensic identification of dental records from an aircraft crash. Most practical applications of biometric identification are open-set problems. MBE-STILL did not address closed-set applications. This means that the SDK under test could make no assumption about whether or not it should return a high-scoring result.

Table 10 – Definition of True Positive Identification Rate

TPIR (R,T,L)	=	Num. searches with enrolled mate reported as candidate with score \geq threshold, T, and rank \leq R on a candidate list of length L	Equation 3
		Num. searches with enrolled mate	

Table 11 – Definition of False Positive Identification Rate

FPIR (T,L)	=	Num. searches without enrolled mate yielding one or more candidates with score \geq threshold, T when candidate list is of length L	Equation 4
		Num. searches without enrolled mate	

Table 12 – Definition of Reliability

REL (T,L)	=	TPIR(N,T, L) where N is the size of the enrolled population	Equation 5
-----------	---	---	------------

Table 13 – Definition of Selectivity

SEL (T,L)	=	Num. candidates with score \geq threshold, T produced in searches without enrolled mate when candidate list is of length L	Equation 6
		Num. searches without enrolled mate	

Table 14 – Definitions of Type I error rates

Metric	Measured over	Definition	Related terms
True Positive Identification Rate (TPIR)	Searches for which a mate is present in the enrolled dataset.	Table 10. Fraction of identification searches for which the enrolled mate is present on the candidate list with rank less than or equal to R, and score greater than or equal to threshold, T. Special cases: 1. The rank requirement can be set to be difficult, i.e. R = 1, or absent (i.e. R = N, where N is the number of enrolled identities) or any value in between. 2. The threshold requirement can be difficult (i.e. high value of T), or absent (i.e. T = 0), or any value in between.	Hit Rate and Reliability of synonyms FNIR and miss rate are synonyms for the complement 1 – FNIR
FNIR	See TPIR	FNIR = 1 – TPIR(R, T,L)	FNIR
Miss Rate	See TPIR	FNIR(R, T,L)	FNIR
Hit Rate	See TPIR	TPIR(R, T,L)	FNIR

Table 15 – Definitions of Type II error rates

Metric	Measured over	Definition	Related terms
False Positive Identification Rate (FPIR)	Searches for which a mate is not present in the enrolled dataset.	Table 11. Fraction of identification searches for which any (i.e. one or more) enrolled identities on a candidate list of length L are returned with score greater than or equal to threshold T.	Selectivity
Selectivity	See FPIR	Table 13. The mean, over a set of searches, of the number of candidates returned for which the score is greater than or equal to a threshold, T.	False positive identification rate

From these metrics the primary performance characteristics are defined in Table 16.

Table 16 – Identification Performance characteristics

Metric	Measured over	Definition
CMC	Searches with mates	The cumulative match characteristic is a plot of $1 - FNIR(R, o, L)$ vs. R, with $1 \leq R \leq L$
ROC	Searches with and without mates	The receiver operating characteristic is a plot of $REL(T,L)$ vs. $SEL(T,L)$

4.2.1. Best practice testing requires execution of searches with and without mates

MBE-STILL embedded 1:N searches of two kinds: Those for which there is an enrolled mate, and those for which there is not. However, it is common to conduct only mated searches. This is bad practice because if the information that a mate always exists is revealed to a test participant, or can be reasonably assumed, then unrealistic gaming of the test is possible.

The cumulative match characteristic is computed from candidate lists produced in mated searches. Even if the CMC is the only metric of interest, the actual trials executed in a test should nevertheless include searches for which no mate exists. MBE-STILL reserved disjoint populations of subjects for executing true non-mate searches.

4.2.2. Rank and threshold censoring

In a real operation, a search against an enrolled population of size N could produce a candidate list with N entries. This would occur if the operating threshold was set to zero. Practically, systems use an internal threshold T and they may only report a finite number of candidates, e.g. only the top 60.

4.2.3. Bootstrap uncertainty estimation

Bootstrapping is an empirical method of measuring the variability of a statistic, often employed when the variability cannot be determined analytically. In the context of this evaluation, bootstrapping is sometimes used to measure the distribution of error statistics (i.e. FNMR or FMR) at a fixed threshold. Each bootstrap iteration samples with replacement from the original set of comparisons. The statistic of interest is then computed over the sampled data. This process is repeated for a large number of bootstrap iterations to produce a distribution of the measured statistic. Bootstrapping relies on several assumptions, including the assumption that the sample data is iid (independent and identically distributed). However, when different comparisons involve the same individual, the comparisons are likely to be correlated due to the existence of Doddington’s zoo [DODDINGTON]. Thus, the independence assumption is violated. Determining the effect this has on the bootstrapped distributions is beyond the scope of this evaluation, but the likely result is an underestimation of the variability of FNMR and/or FMR in some cases.

4.3. Failure to acquire

Some biometric algorithms may fail to convert some input samples to templates. This can be the result of a software bug (e.g. buffer overrun), or an algorithmic limitation (e.g. failure to find eyes in small images), or an elective refusal to process the input (e.g. because the image is assessed to have insufficient quality). For these events, the result is a template of zero size. The NIST API specification required the verification function to nevertheless process such templates. For identification the result of a failed template generation was not passed to the search function.

NOTE: Some face recognition algorithms would fail to produce a template when the result of decoding a broken JPEG image was provided to the implementation under test. This is not really the fault of the implementation. MBE-STILL removed all broken images before the test. The term "broken" in this context means a malformed JPEG file, e.g. one that is syntactically incorrect, up to and including truncation of the JPEG stream. Integrators will generally need to detect broken data before passing to the SDK⁵.

5. Properties of the implementations

The objectives of the test were to

- Assess core capability of face recognition technology.
- Be fair tests
- Be repeatable
- Not constrain or favor particular algorithms.

Accordingly the test was administered with the following aspects.

Table 17 -- Test design considerations

Black Box Testing	NIST specified a "C" API and required MBE STILL participants to hide their face recognition algorithms behind it. The resulting technologies were submitted to NIST as windows or linux libraries. This constitutes black-box testing, i.e. the test laboratory has no exposure to, nor interest in, how the technology works.
Fair repeatable testing	The software implementations were given face images stored on hard drives. This defines offline testing (versus online testing where a person appears before a camera). Offline testing allows use of very large datasets (giving statistical significance), supports fair comparative testing, and allows experiments to be exactly repeated.
Modular components	The main operations supported are template generation (conversion of facial imagery to a proprietary, trade-secret, mathematical representation), template comparison for one-to-one verification, and template search for one-to-many identification.
Support for vendor-defined sample fusion	One or more face images of a person are bundled together and passed to template generation routines as a single object. This allows the implementation to fuse information derived from any or all of those images.
Separated verification and identification tests	The outputs of verification runs are real-valued similarity scores. These primarily support computation of the Receiver Operating Characteristic (ROC) as the basic statements of biometric accuracy. The outputs of identification runs are candidate lists which embed hypothesized identities for the search image. Each hypothesized identity is accompanied by a real valued similarity score. Candidate lists are sorted in decreasing order of that score. Candidate lists are of length 200, unless stated otherwise. Identification systems rarely return a score for all enrolled identities.
Support multistage matching identification algorithms	For reasons of efficiency, biometric identification systems can embed a multi-stage matching process that uses successively more accurate but more computationally expensive algorithms to reduce the population of candidate identities from an initial large value. MBE-STILL is a black box test in which templates embed vendor-defined mathematical representation of the face that are trade secrets. The API regards a template as a blob of binary data. Internally a provider may embed several different facial representations that can be used conditionally during a search. Thus a template of measurable size T could be composed of say two parts. A small template of size T1 for efficient search, and another for expensive end-stage matching. This might have size T2 such that $T = T_1 + T_2$ and $T_2 > T_1$.
Asymmetric Templates	Templates were generated for the specific purpose of enrollment, verification or identification. That is the templates have assigned roles. Templates are not used in other roles. This allows enrollment

⁵ For all images, NIST tested JPEG conformance by ignoring any file for which "djpeg -fast -outfile /dev/null image.jpg" gave any output to stdout or stderr (djpeg is the command-line JPEG decompressor supplied in the IJG JPEG implementation www.ijg.org).

templates to have a different size than an identification template, for example.

Experimental method: The face recognition implementations submitted to MBE-STILL were tested in a black box manner. The mathematical representation of facial input data was stored as a proprietary template. The content of a template is unknown and non-standard. It is always a trade secret. The SDK was used as follows.

- The SDK under test was initialized. For one-to-many enrollments, the SDK was informed of the size of the population.
- The NIST test harness bundled $k \geq 1$ images of a person were bundled into a MULTIFACE data structure.
- The MULTIFACE was passed to the template generation function of the SDK under test.
- The template was stored.
- In addition templates were concatenated to form an enrollment database (EDB).
- For one-to-one comparisons two isolated templates were passed to the SDK’s comparison engine.
- For one-to-many searches, the SDK was initialized. The SDK would typically read the EDB into main memory.
- For one-to-many searches, a template was passed to the search function and it returned a candidate list.
- For one-to-one searches with an enrolment database, the template would be passed to the search function with an explicit claim to a particular identity. The SDK returns a comparison score.

NIST stored comparison scores and candidate lists and later used these in computation of the metrics of section o.

MBE-STILL REPORT PARTICIPANT KEY	P = PITTPATT	R = SURREY U.	S = TSINGHUA U.	T = TOSHIBA	U = DALIAN TECH U.	PAGE 19 OF 58
	V = NEC	W = L1 IDENTITY	X = COGNITEC	Y = SAGEM	Z = NEUROTECHNOLOGY	

6. Results

The results are presented as series of “investigations”. Each addresses one quantitative aspect of performance.

INVESTIGATION 1. *Investigation-mode one-to-many search accuracy*

Are there accuracy differences between suppliers?

Demand driver: There are many applications of biometric identification algorithms. As summarized in Table 2, applications are differentiated by population sizes and the kinds of images being compared. However, in most applications, the core accuracy of a facial recognition algorithm is the most important performance variable. It quantifies the ability to answer the question are two samples from the same person, or not.

Experimental method: NIST used participants’ class C SDKs to enroll images from each of N subjects to form an enrollment database containing N templates. A template is an entirely vendor-defined and non-standard blob of data. It is not suitable for interoperable interchange between systems. The template contains at least one mathematical representation of the face of the person in the input image.

Two sets were used:

- LEO: The values of N were 10 thousand, 80 thousand, 320 thousand and 1.6 million. The enrollment sets are simple random samples from the parent set of 1.6M persons. The sets are otherwise not subsets of each other. Mated and non-mated searches were run against these enrollments. The number of mated searches was M = 9240, 40000, 40000 and 40000 respectively for the four enrolled population sizes.
In all cases, subjects were enrolled with $k \geq 1$ images. These were the oldest to the second-to-most recent image of a subject. Searches were made using $k = 1$ images per person. This was the most recent image of the subject.
- DOS / Natural: The values of enrolled population size N were 83981, 388800 and 1800000. In all cases, subjects were enrolled with $k \geq 1$ images. These were the oldest to the second-to-most recent image of a subject.

In each search, the SDK under test was asked to report the top 200 candidates.

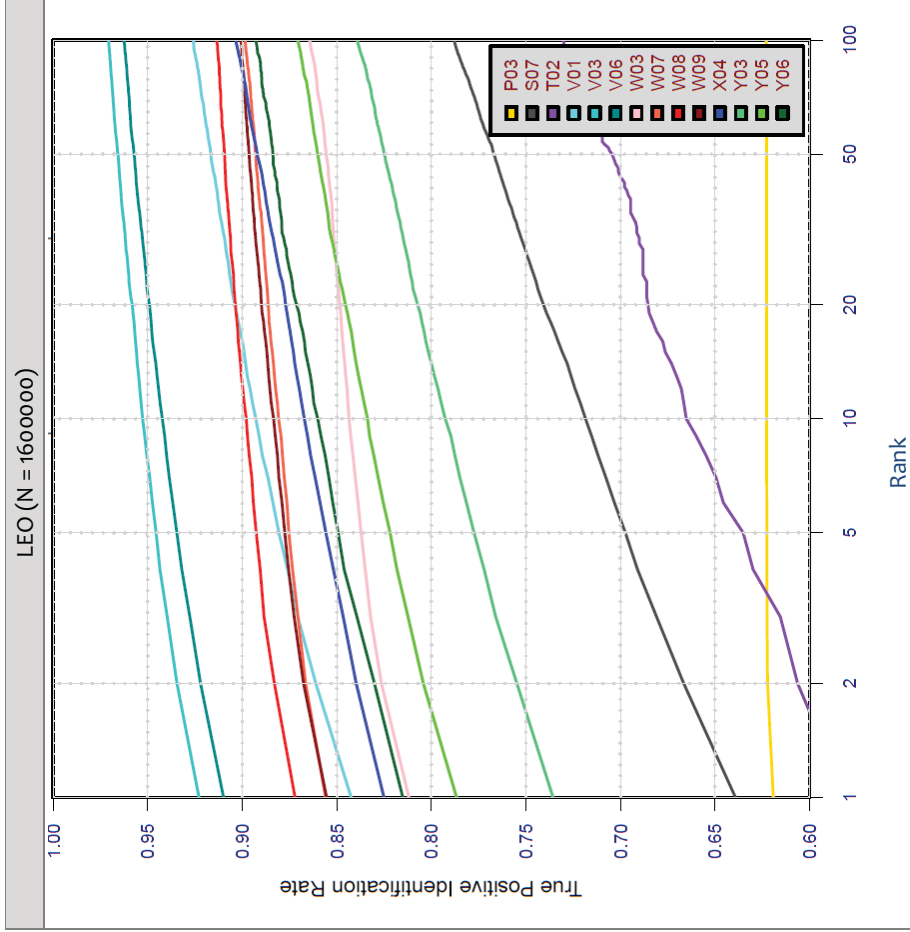
Results: The candidate lists from mated searches in the LEO dataset were used to compute the cumulative match characteristic of Figure 5. The population size was 1.6M. The table to the right of the figure shows the rank 1 hit rate. There is more than a factor of five range in the observed miss rates (i.e. 1-TPIR), from 0.08 to 0.42. In addition, some identification algorithms produce more rapidly rising CMCs. This is a valuable property with implications for the workload on an examiner tasked with finding a mate on a candidate list. This aspect is quantified later in this report.

Conclusions: As with other biometrics, accuracy of facial recognition implementations varies greatly across the industry. Absent other performance or economic parameters, users should prefer the most accurate algorithm. Note, however, that the results of this section are entirely rank-based – the CMC computations ignore score information. This befits use of face recognition in the investigational mode in which an examiner is willing to traverse candidate lists looking for mates. Subsequent investigations in this report consider threshold-based metrics appropriate for identification mode applications. There, the leading algorithms are different from those listed above.

Note that the absolute values of identification accuracy will always depend on the dataset used, specifically to the properties of the images in use. In particular this study includes some residual non-frontal images that eluded detection during the data preparation phase. These images include some 90-degree profiles, and, somewhat more frequently, images in which the face is at a 45 or 60 degree yaw angle to the camera. These images usually cause complete recognition failure and depress overall error rates

MBE-STILL REPORT PARTICIPANT KEY	P = PITTPATT	R = SURREY U.	S = TSINGHUA U.	T = TOSHIBA	U = DALIAN TECH U.	PAGE 20 OF 58
	V = NEC	W = L1 IDENTITY	X = COGNITEC	Y = SAGEM	Z = NEUROTECHNOLOGY	

Figure 5 – Identification accuracy vs. candidate rank



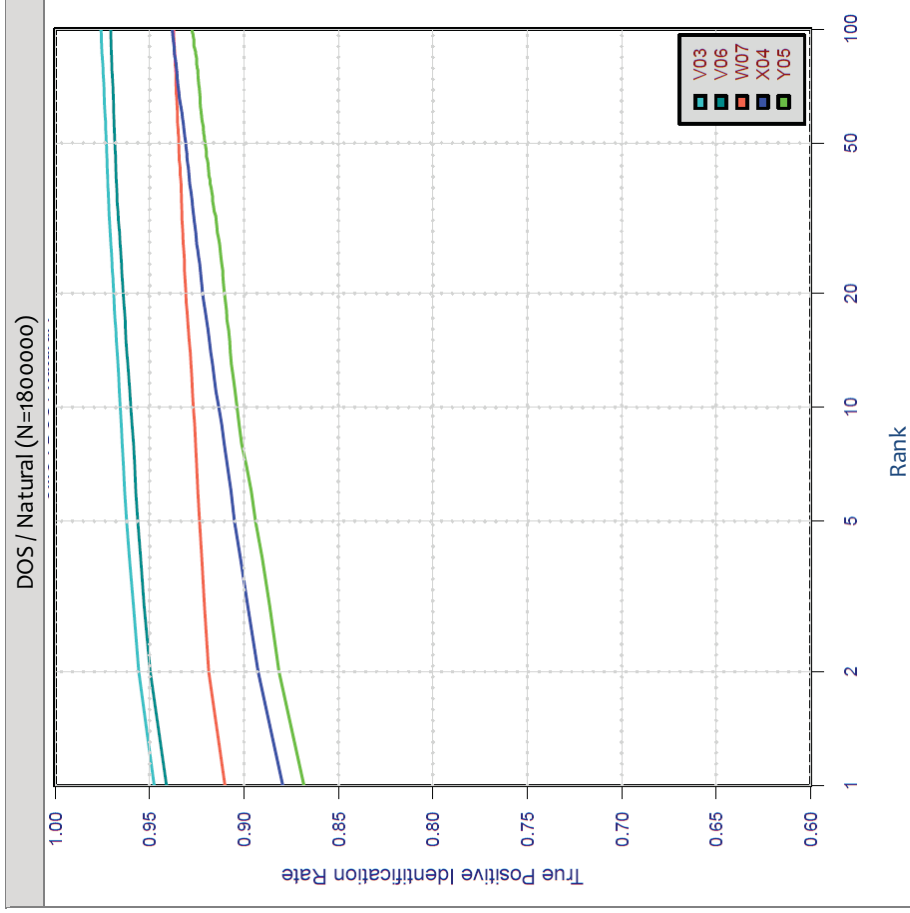
Number of searches with mate = 40000

Number of searches without mate = 40000

The vertical axis plots Equation 3, $CMC(R) = TPIR(R, T, L)$ with $T=0$ and $L = 200$.

The next rows give $CMC(1) = TPIR(1,0,200)$

SDK	V03	V06	W08	W07	W09	V01	X04	Y06	W03	Y05	Y03	S07	P03	T02
CMC(1)	0.92	0.91	0.87	0.86	0.86	0.84	0.83	0.82	0.81	0.79	0.74	0.64	0.62	0.58



Number of searches with mate = 40000

Number of searches without mate = 40000

The vertical axis plots Equation 3, $CMC(R) = TPIR(R, T, L)$ with $T=0$ and $L = 200$.

The next rows give $CMC(1) = TPIR(1,0,200)$

SDK	V03	V06	W07	X04	Y05
CMC(1)	0.95	0.94	0.91	0.88	0.87

INVESTIGATION 2. Identification-mode one-to-many search accuracy

Can a threshold be established such that the number of false matches produced can be targeted.

Demand driver: Fingerprint identification systems are most commonly used without exhaustive review of a candidate list⁶. Face-based systems can also be used in this identification mode. While higher error rates are typically higher, these may be useful and tolerable given the longstanding acceptance of face images in government issued credentials such as passports and driving licenses. In the one-to-many mode, duplicate enrollments are detected if and only if the comparison score is at or above some threshold.

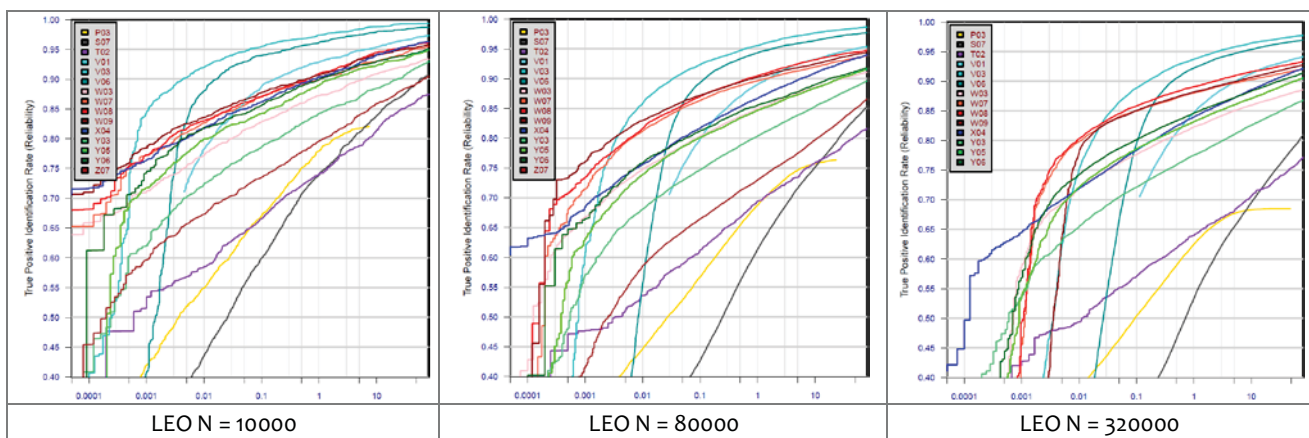
Prior work: A number of un-published tests have been conducted, usually as part of procurement processes. Prior NIST tests of facial recognition technology have computed identification accuracy from a complete set of N one-to-one comparison scores.

Experimental method: This investigation uses identically the same candidate lists produced in the prior investigations. Here the score values produced are not ignored; they are used to compute threshold-based estimates of accuracy. The result is plotted as a Receiver Operating Characteristic (ROC) which plots reliability against selectivity as a parametric function of threshold, i.e. REL(T) vs. SEL(T). Rank is completely ignored. While any set of real-valued thresholds can be used, we adopted the set of all observed genuine scores.

Results: Figure 6 shows the ROC for the SDKs operating on a population of 1.6M. The three subfigures show the ROCs for populations of 10K, 80K and 320K. The four ROCs are essentially horizontal translations of each other because the occurrence of high scoring non-matches is approximately linear in the population size.

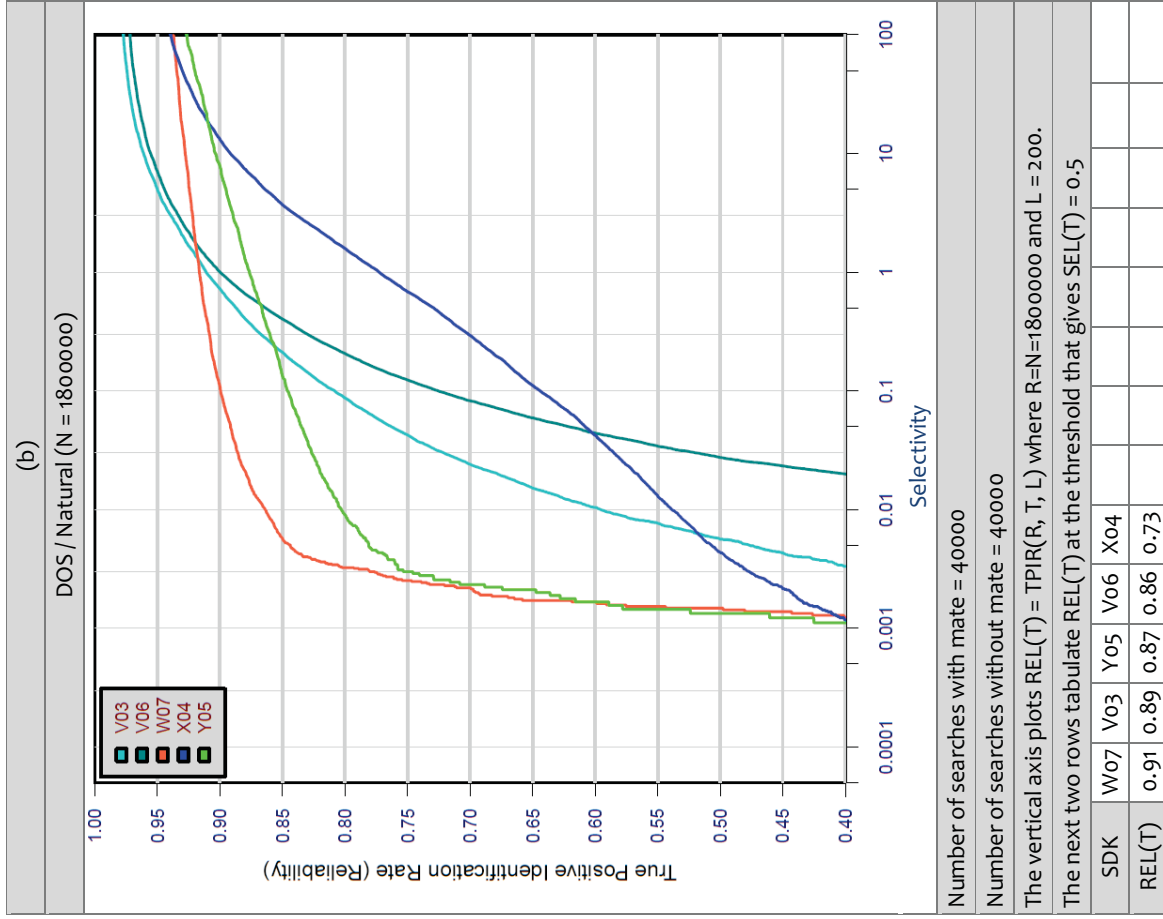
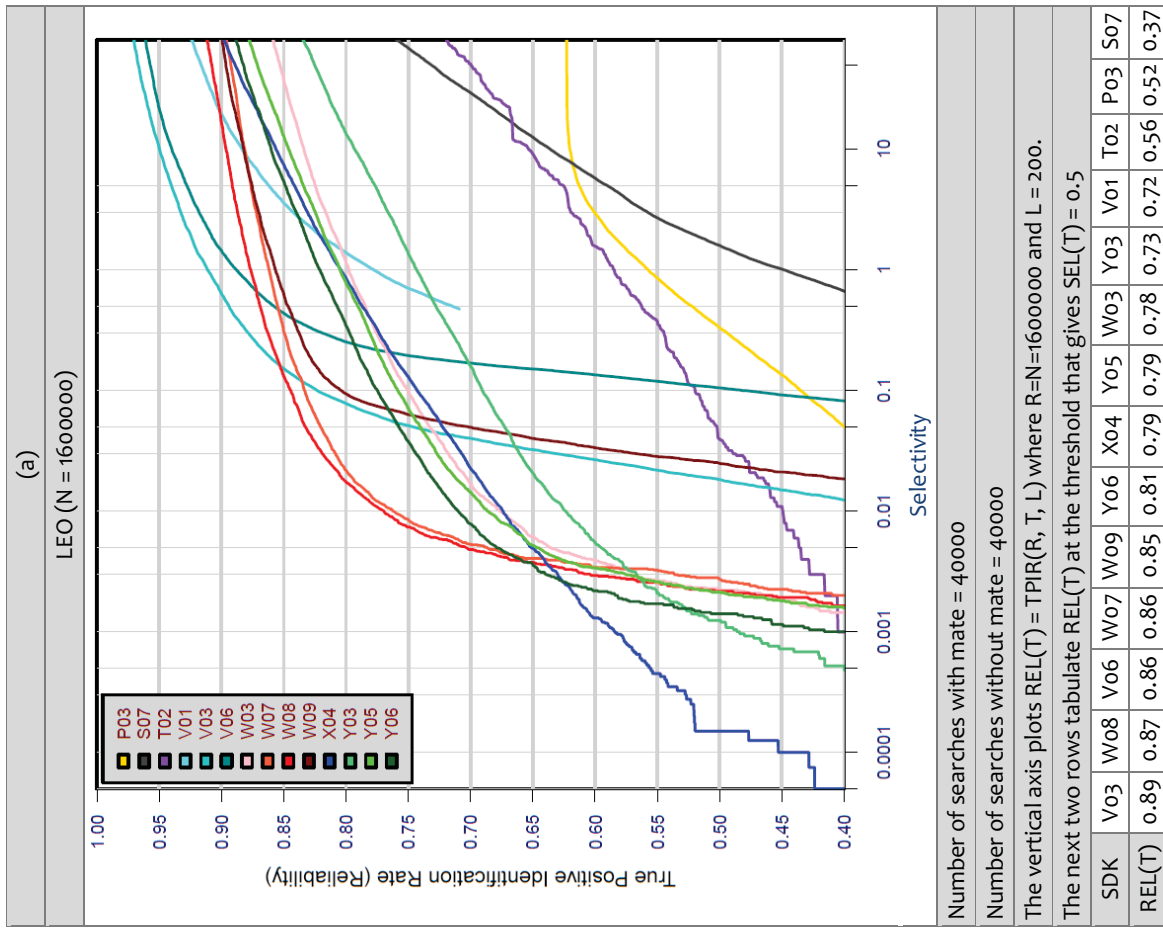
This ROC is one of the more interesting ROCs plots published in biometric test reports, because the ROCs cross. We can identify three selectivity regimes as follows. When selectivity is high (when $SEL > 1$, for example), the best performing *investigation* mode SDK, V03, gives the highest reliability. This is consistent with the cumulative match characteristics plotted earlier. However, in the middle-selectivity regime ($SEL < 0.1$), the reliability of V03 drops off rapidly such that the W08 SDK initially gives the best reliability. This lasts until the low selectivity region, $SEL < 0.005$, where ultimately X04 is most reliable. In this *identification* mode, the X04 SDK will produce false matches for only one in every 10000 searches, but reliability has dropped so that more than one in two mated searches will give a miss. Note that at such low selectivities, when the operating threshold T is high, the best *investigational* SDKs fail essentially completely (i.e. reliability approaches zero).

Conclusions: The leading algorithms at high selectivity rates are not the leading algorithms at low selectivity rates. However the declared intent of MBE-STILL was to support recognition in an *investigation* mode. Without soliciting SDKs for the specific purpose of identification mode operation, the results given in Figure 6 may not reveal the full potential for lights out identification with LEO images



⁶ This does not apply to latent fingerprint systems where examiners are involved in variety of ways [MEAGHER], including the forensic markup before a search, and in the adjudication of results from a search.

Figure 6 – Identification rate vs. selectivity



INVESTIGATION 3. *Dependence on population size*

How do the false negative and false positive error rates depend on the size of the enrolled population?

Demand driver: Face images are being collected in a number of civil and criminal applications [PINELLAS, JAPAN VISIT, US-VISIT, FBI]. The enrolled population increases with time. The number of subjects enrolled in the US-VISIT system has increased from 12 to 110 million in the seven years to 2010. The likelihood that a biometric sample collected during a prior encounter will be found in a one-to-many search is a function of the population size because the chance of one or more false matches increases with the population size.

Prior work: The FRVT 2002 study reported open-set identification accuracy as a function of the size of the enrolled population. In addition, a number of unpublished studies of 1:N facial recognition performance have been conducted as parts of procurement processes [WAGGETT].

The degree to which results can be extrapolated to large populations is a subject of debate. The academic studies have modeled empirical data [HUBE], or made assumptions of binomial independence [BOLLE, GROTHOR, SHERRAH] in empirical data. One study ambitiously used experimental results from a population in the hundreds to predict performance in the billions [WEIN]. Leading commercial providers have also aware of the need to quantitatively model scaling [FONDEUR, JAROSZ, MARTIN].

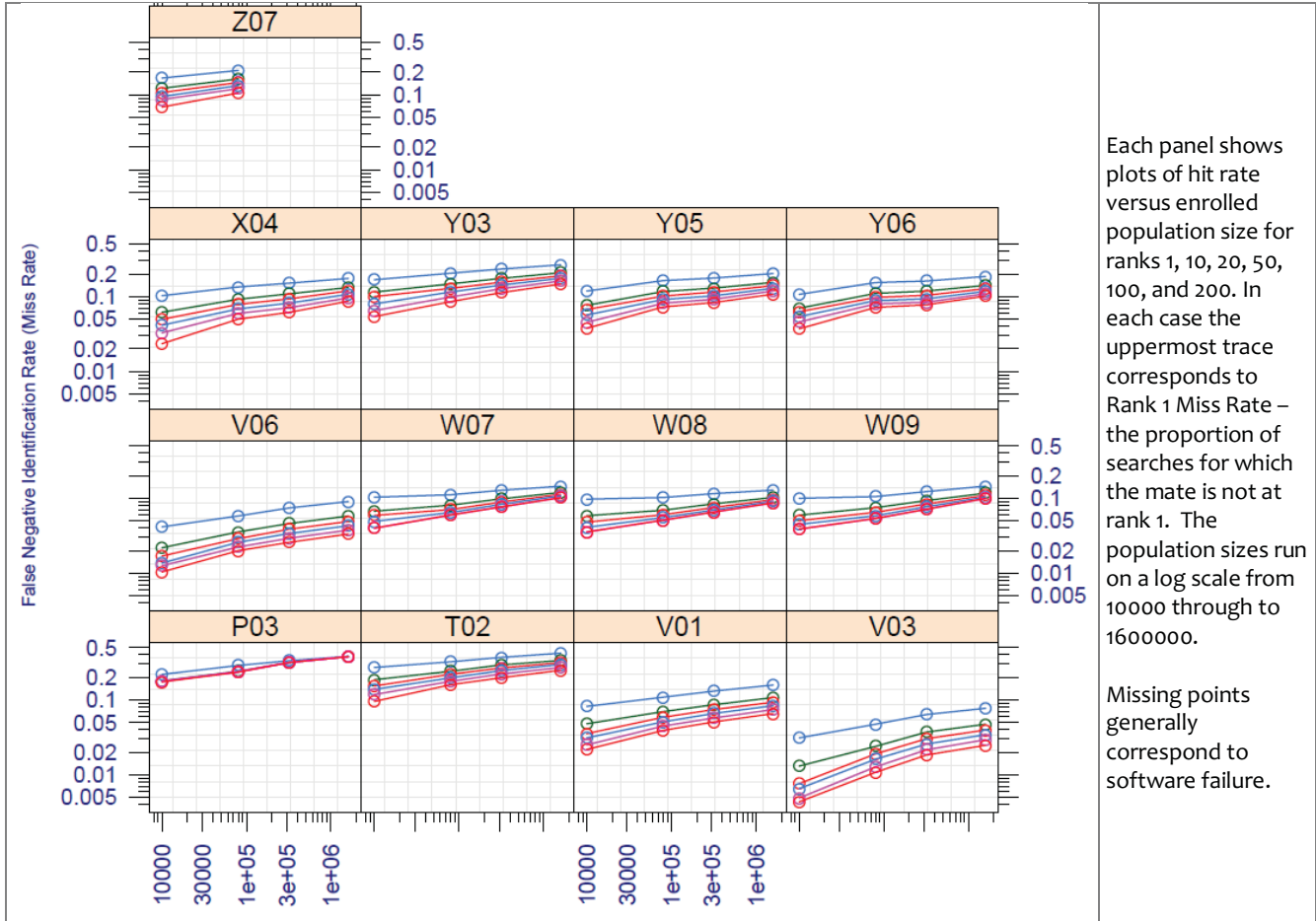
Experimental method: Identical to INVESTIGATION 1.

Results: The plots of Figure 7 show the increase in false negative identification rates (i.e. miss rates) for each class C SDK as the size of the enrolled LEO population increases. The text in the side panel explains the format in more detail. The overall result is that FNIR increases approximately linear with the logarithm of population size N . The topmost curve for the V03 panel shows that rank 1 recognition error rate in a population of 10000 is 0.031 rising to 0.077 for 1.6 million. For the W08 SDK the curves are notably flatter: The miss rate at 10000 is 0.1 and rises to 0.13 at 1.6 million.

Conclusions: There is an approximate dependence of accuracy on log of the population size. This is not an exact model.

The observed results have applicability for the LEO dataset at the population sizes used. For larger populations, either an empirical trial will be conducted, or careful extrapolation will be needed to estimate performance.

Figure 7 – LEO Identification accuracy dependence on population size



Each panel shows plots of hit rate versus enrolled population size for ranks 1, 10, 20, 50, 100, and 200. In each case the uppermost trace corresponds to Rank 1 Miss Rate – the proportion of searches for which the mate is not at rank 1. The population sizes run on a log scale from 10000 through to 160000.

Missing points generally correspond to software failure.

SDK	Enrolled Population			
	TPIR(1, N, 200)			
	N=10000	N=80000	N=320000	N=1600000
P03	0.782	0.712	0.667	0.619
T02	0.730	0.678	0.631	0.581
V01	0.918	0.892	0.869	0.842
V03	0.969	0.953	0.936	0.923
V06	0.959	0.942	0.926	0.910
W07	0.897	0.889	0.873	0.856
W08	0.903	0.898	0.885	0.872
W09	0.900	0.895	0.878	0.855
X04	0.898	0.866	0.849	0.826
Y03	0.833	0.796	0.767	0.736
Y05	0.882	0.837	0.824	0.798
Y06	0.893	0.847	0.839	0.815
Z07	0.833	0.789		

The Table quantifies True Positive Identification Rate, at rank 1 for four enrolled population sizes.

This is 1 minus the quantity graphed above:
FNIR = 1 –TPIR.

INVESTIGATION 4. Dependence on rank

What are the chances of finding a mate far down the candidate list?

Driver: In an investigational mode, a face recognition algorithm is used to provide a list of candidates to an examiner. The search image may have an enrolled mate, in which case the examiner may confirm the identity at some rank, or there may be no prior encounter of the individual, in which case the examiner would traverse and stop after finding no mates.

Experimental method: Identical to INVESTIGATION 1.

Results: Figure 8 re-plots the data of Figure 7 to show the dependence on rank. Each panel shows the dependence of hit rate on rank for a particular class C, one-to-many, SDK. The rank axis runs from 1 to 200 on a logarithmic scale. For any given population size, false negative identification rates exhibit an approximately linear dependence on the logarithm of the rank. While 200 candidates were requested, some systems (e.g. W07) returned candidate lists of length 200 for which the last 100 entries were zero. This produces a flat cumulative match characteristic. This may have occurred because the search algorithm is more efficient when fewer candidates are returned.

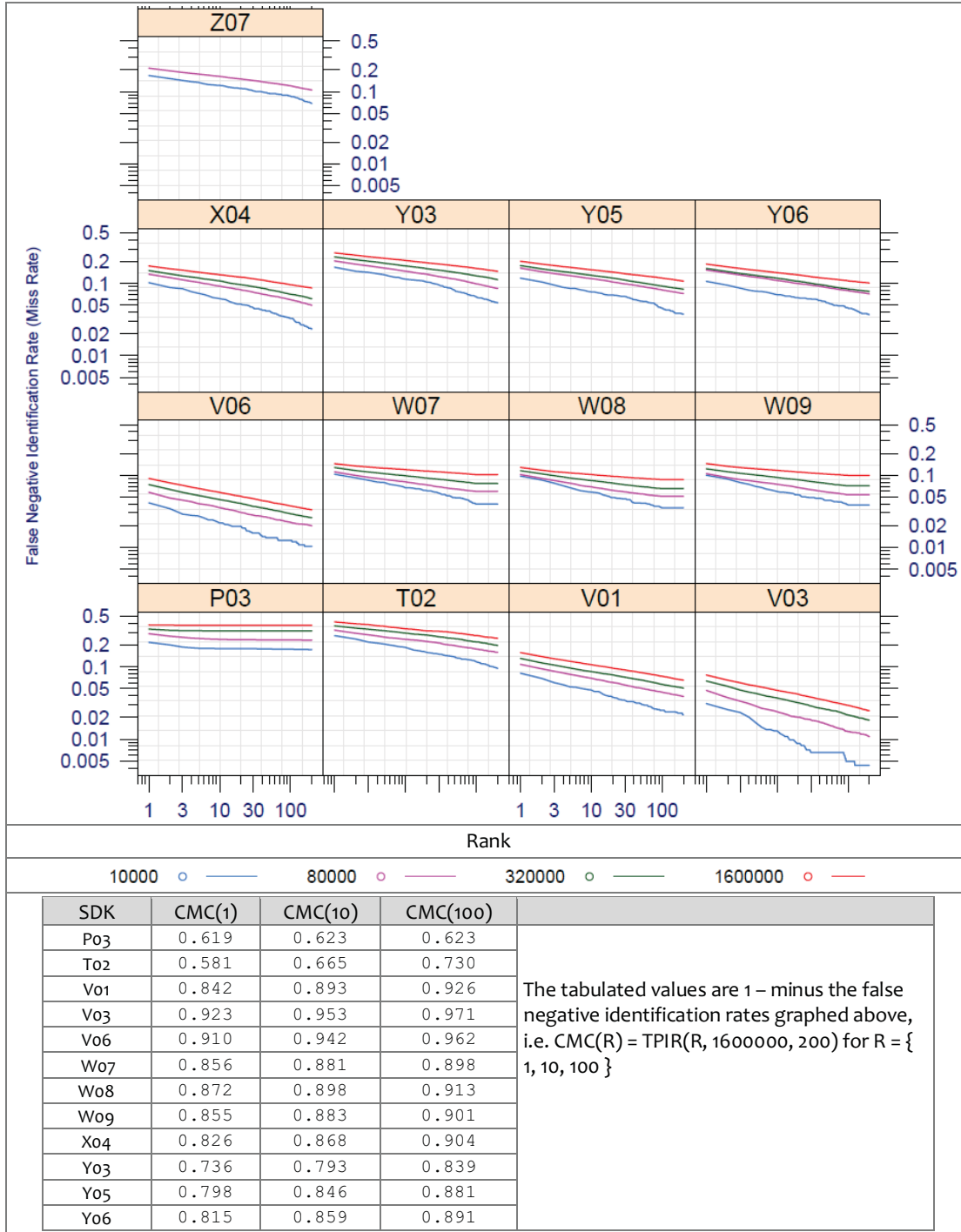
For some SDKs, FNIR decreases more rapidly with rank. This is especially true in smaller populations. This dependency was observed in the first CMC of Figure 5. Face identification algorithms which “front-load” a candidate list with mates offer workload benefits as discussed below.

Conclusions: As with other biometrics, accuracy of facial recognition implementations varies greatly across the industry. Absent other performance or economic parameters, users should prefer the most accurate algorithm. Note, however, that the results of this section are entirely rank-based – the CMC computations ignore score information. This befits use of face recognition in the investigational mode in which an examiner is willing to traverse candidate lists looking for mates. Subsequent investigations in this report consider threshold-based metrics appropriate for identification mode applications. There, the leading algorithms are different from those listed above.

Note that the absolute values of identification accuracy will always depend on the dataset used, specifically to the properties of the images in use. In particular this study includes some residual non-frontal images that eluded detection during the data preparation phase. These images include some 90-degree profiles, and, somewhat more frequently, images in which the face is at a 45 or 60 degree yaw angle to the camera. These images usually cause complete recognition failure and depress overall error rates.

MBE-STILL REPORT PARTICIPANT KEY	P = PITTPATT	R = SURREY U.	S = TSINGHUA U.	T = TOSHIBA	U = DALIAN U.	PAGE 26 OF 58
	V = NEC	W = L1 IDENTITY	X = COGNITEC	Y = SAGEM	Z = NEUROTECHNOLOGY	

Figure 8 – LEO identification miss rate versus rank.



Workload implications: In a law enforcement scenario, for example, a human examiner might review the candidates returned in an identification search. Typically the examiner would compare each candidate with the search image starting with the highest scoring candidate and proceeding in descending order of similarity score. The examiner would stop early if he is able to positively confirm a mate. In this case, an expression for the total workload associated with resolving candidate lists of length K is derived as follows.

- The examiner will always inspect the first ranked image. Num reviewed = 1

- The examiner will inspect the fraction of candidates not found at rank 1. Num reviewed = 1-CMC(1)
- The examiner will inspect the fraction of candidates not found at rank 1 or 2. Num reviewed = 1-CMC(2)

Etc. Thus if the examiner will stop at after a maximum of K candidates the expected number of candidate reviews is

M(K)	=	$1 + (1-CMC(1)) + (1-CMC(2)) + \dots + (1-CMC(K-1))$	Equation 7
	=	$K - \sum CMC(r)$ where the sum runs from $1 \leq r \leq K-1$	

A recognition algorithm that front-loads the cumulative match characteristic will offer reduced workload for the examiner. This workload is defined only over the searches for which a mate exists. In the cases where there truly is no mate, the examiner would review all K candidates. Thus, if the proportion of searches for which a mate does exist is β , which in the law enforcement context would be the recidivism rate, the full expression for workload becomes:

M(K)	=	$\beta (K - \sum CMC(r)) + (1 - \beta) K$	Equation 8
	=	$K - \beta \sum CMC(r)$ where the sum runs from $1 \leq r \leq K-1$	

Figure 9 shows the dependence of M(K) as a function of K, for each class C identification SDK operating on LEO images. The text in the side panel explains each plot in more detail. Importantly we restrict the analysis to the case where there is always a mate, i.e. $\beta = 1$. This is done because the goal is to compare algorithms. However, note that if $\beta < 1$, examiners will have to review more candidates than are plotted here.

The plots show that if an examiner is willing to review, say, K = 60 candidates, then the expected number of candidates actually needing review will often be fewer than 10. For the V03 SDK the number is about 3.

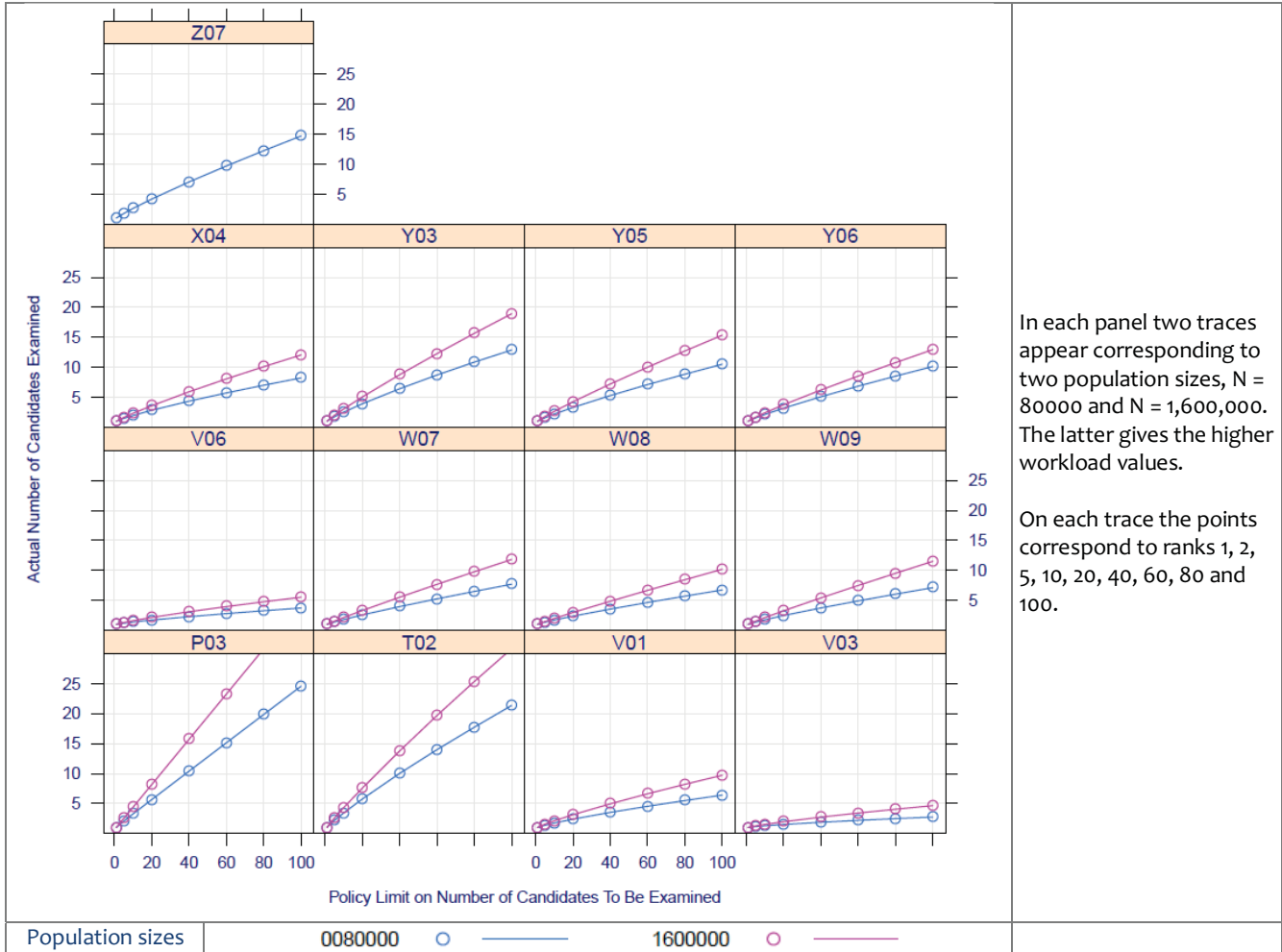
Cost implications: The above expressions for examiner workload could be multiplied by suitable time and salary factors to estimate cost. Such a cost formulation should be extended to capture the cost of missing a mate altogether - This is a societal cost of failing to find a mate in the first K candidates.

Conclusions: The use of a face recognition system can dramatically decrease workload on a human examiner. The expected number of candidates before a mate is found is a useful performance metric for identification systems.

Assumptions: This workload model assumes the following:

- The candidates are reviewed serially, not all at once in a large screen GUI, for example.
- The candidates are searched in decreasing order of similarity score.
- The time taken to confirm or exclude a candidate is independent of the rank of the candidate.
- The time taken to confirm or exclude a candidate is independent of population size. This is potentially incorrect since identification in a very large population will produce candidates more similar to the search sample, than in smaller populations.
- Examiners will stop after confirming a mate.
- The database is correctly consolidated such that the number of mates is zero or one.
- Examiners always find a mate if it is present, and examiners do not assign a wrong mate. Particularly that examiner success is independent of β . Effects of fatigue and boredom have been reported when β is very low [GREATHOUSE].
- Scores are ignored and score thresholds are not applied.

Figure 9 – Workload implications of LEO cumulative match performance



INVESTIGATION 5. Impostor distribution stability

Selectivity is defined above as the number of false matches produced in a search against an operational database. If future photographs with different image properties are searched against this database, does selectivity change?

Demand driver: The operating threshold of a biometric system is set to meet some accuracy criterion such as a selectivity requirement.

Prior work: Most of the academic literature addresses improvement of Type 1 error rates such as better hit rates. The primary performance metrics are 1:1 FNMR at fixed FMR, and closet-set CMC. The importance of a stable impostor distribution is little discussed.

Experimental method: For four different providers' algorithms, V, W, X and Y, the identical candidate lists used INVESTIGATIONS 1 and 2, were analyzed as follows. ROCs were plotted for DOS/Natural and LEO on the same graph. For a small set of selectivities between 0.01 and 10, the thresholds that give those selectivities on the LEO dataset were computed. This computation uses only searches without mates. If for algorithm, i, the thresholds are T_i , then the point $(SEL(T_i), REL(T_i))_{LEO}$ is joined to $(SEL(T_i), REL(T_i))_{DOS}$ with a grey line.

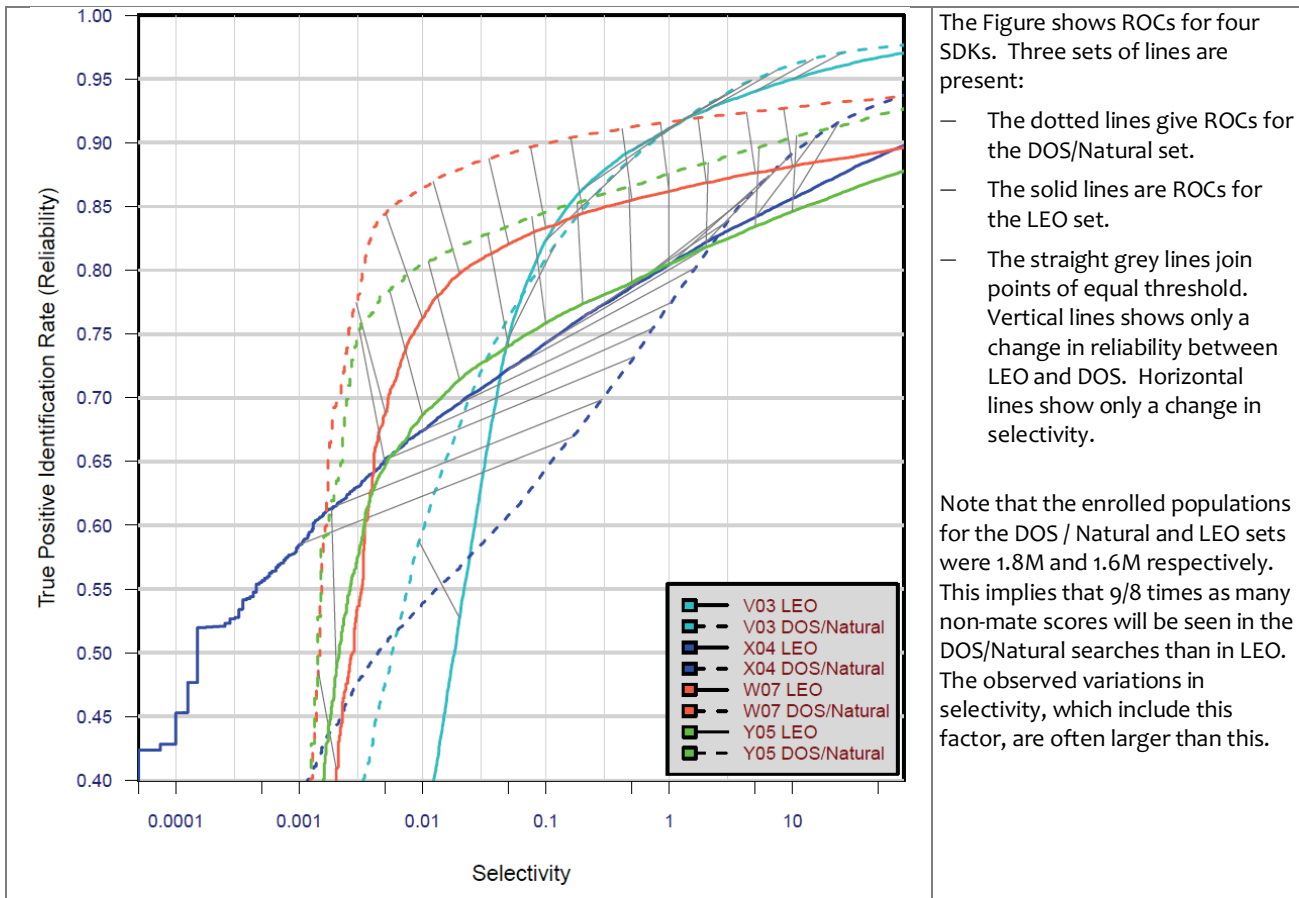
NOTE: The experiment did not enroll LEO and DOS images together into a single database. This would have supported the use of score normalization.

Results: ROCs are shown in Figure 10. With DOS data, all algorithms give generally higher reliability than with LEO data. Three of the four algorithms give better selectivity also. The one exception is the X04 algorithm which produces many more candidates above a fixed threshold value with the DOS / Natural imagery than with LEO.

A desirable property for a biometric system is to have a stationary impostor distribution. This allows a threshold to be set so that false match statistics are known and controlled. This is difficult if the properties of the images are unknown. The alternative is to set the threshold for the specific database. The approach to threshold setting is out-of-scope of this test.

The algorithms were informed of the kind of the images being used. Thus each DOS / Natural image was tagged with a label “visa”, and each LEO image with the label “mugshot”. The SDK could, in principle, invoke completely different template generation and matching algorithms for the two image variants. This might necessitate setting of database-specific thresholds.

Figure 10 - Reliability and selectivity at a fixed threshold



Conclusions: Algorithms exhibit variation in selectivity (the number of non-mates returned in a search) when a fixed threshold is used on two different enrollment databases. Thus, depending on the application requirements and the algorithm, the threshold may not be portable across datasets, and may need to be calibrated using a set of non-mated searches.

INVESTIGATION 6. Search duration

And does the time to identify scale linearly with the size of the enrolled population?

Demand driver: In most deployments, the enrolled population increases over time. This may be a continuous process or the result of merging separate datasets. If the database doubles in size, does the search time? This has major implications for planning, and system cost.

Prior work: There are no publically reported tests in operational populations. There is a large and mature literature on fast search algorithms, although most of this is outside of the biometric arena. The term fast refers to algorithms for which average search time increases better-than-linearly with population size N, for example as log N.

Experimental method: Calls to the one-to-many search function were made on a dedicated computer. The computer was not running any other processes except those back-grounded as part of the operating system. The durations are measured by wrapping the elemental identify_template function [MBE-API, Table 27] in a wall time counter⁷. The measurements do not include disk access unless the SDK under test elected to access enrollment or configuration data during a search – this is not necessary because the API supported initialization prior to searching.

The MBE-STILL test plan formally stated the durations of Table 18 as limits on the core elemental functions of the SDKs. The times were stated as 90-th percentiles.

Table 18 – Processing time limits in milliseconds

1	2	3	4	5
Function	1:1 verification without enrollment database	1:1 verification with enrollment database	1:N identification	Pose conformance estimation
Feature extraction enrollment	1000 (1 core)	1000 (1 core)	1000 (1 core)	500 (1 core)
Feature extraction for verification or identification	1000 (1 core)	1000 (1 core)	1000 (1 core)	
Verification	5 (1 core)	10 (1 core)	NA	
Identification of one search image against 1,000,000 single-image MULTIFACE records.	NA	NA	10000 (16 cores) or 160000 (1 core)	

In identification trials, the SDK was permitted to use the available hardware as it saw fit. For, the implementation could elect to start any number of threads [1,16] and this could be varied dynamically and as a function of N.

Table 19 – Adjustment of search duration estimates by number of cores used

SDK Identifier	Number of cores used by SDK in a search	For a search of duration, T, the time reported by NIST	Remarks
Vxx Txx	1	T / 16	The divisor is applied because 16 searches can be executed independently on the standard hardware used in this test. Operationally this is unrealistic unless 16 separate search transactions are actually outstanding. Otherwise the hardware is wasted.
All others	16	T	These implementations use threads to fully utilize the available cores hardware for a single search.

Duration measurements were made by executing searches involving mates and non-mates in a random order. The search population was N = 10000, 80000, 320000, and 1600000. Only LEO images were used.

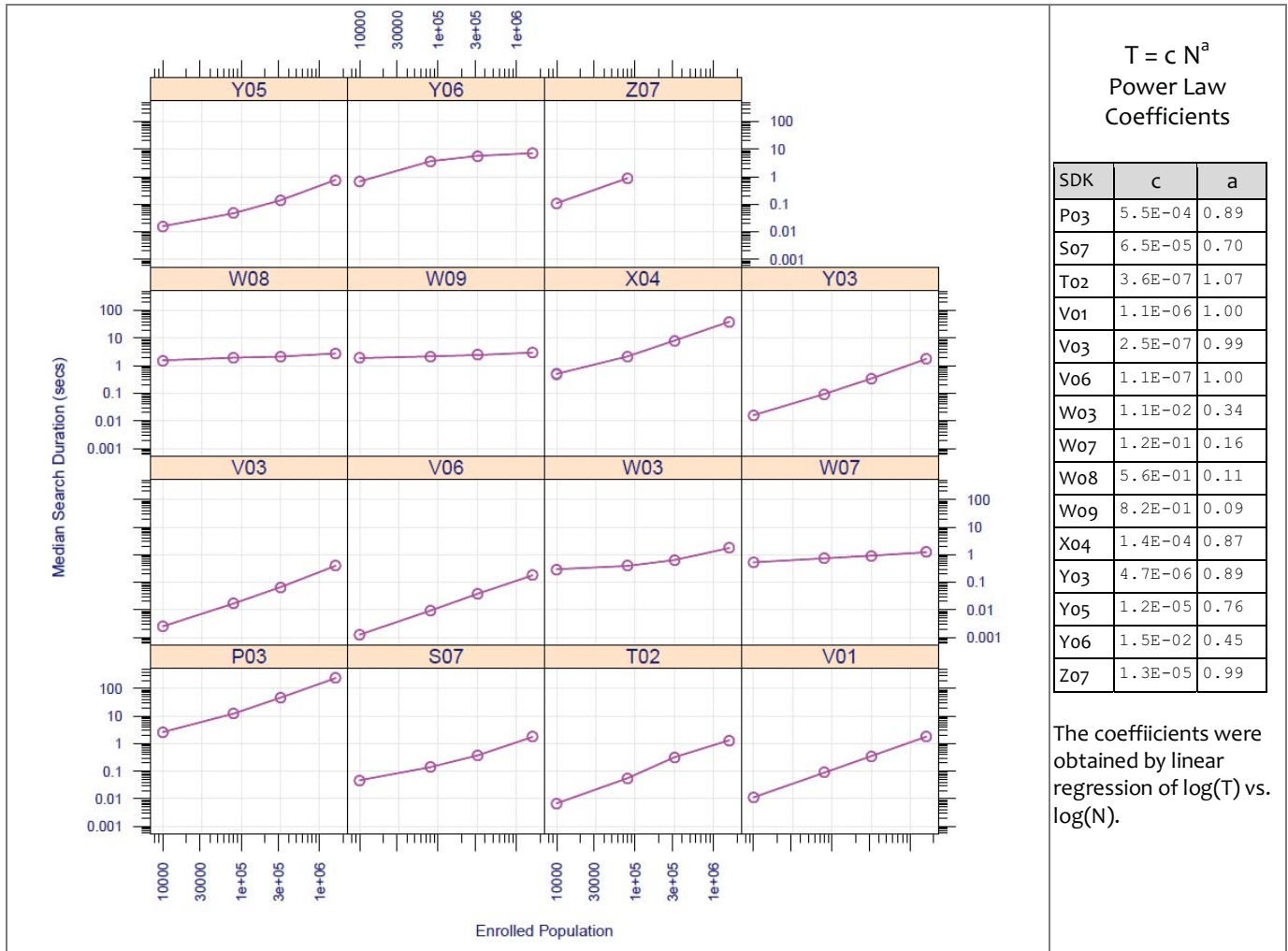
The estimates reported below are median values estimated over 1000 searches for which a mate exists, and 1000 searches for which a mate does not exist.

⁷ The standard "C" call get_time_of_day() function has resolution of 16ms on windows hosts but microseconds under linux.

Results: Figure 11 shows the duration, T, of a search as a function of enrolled population size, N, for the LEO images. In each panel there are two traces, one each for searches with and without mates. The median times in these two cases are so close together that the traces lie directly on top of each other, and only one trace is visible⁸. The functional form in most cases is an approximately straight line on a log-log plot. This observation, $\log T = a \log N + b$, corresponds to a power-law form $T = cN^a$ where the constant $b = \log c$ determines the intercept on the observed plot, and the constant a is the slope.

For the V-series SDKs the dependence is linear ($a \sim 1$). For the W-series SDKs the scaling is better (the coefficient “a” is as low as 0.1). However, while the V-series SDKs are absolutely faster, this only applies for populations in the millions. In the tens of millions the W-series SDKs would be faster unless algorithmic changes were made.

Figure 11 – Duration of LEO identification searches



N	P03	S07	T02	V01	V03	V06	W03	W07	W08	W09	X04	Y03	Y05	Y06	Z07
10000	2.64	0.047	0.007	0.011	0.003	0.001	0.297	0.531	1.562	1.891	0.510	0.016	0.016	0.672	0.109
80000	12.31	0.141	0.056	0.089	0.017	0.009	0.391	0.734	1.953	2.188	2.170	0.094	0.047	3.656	0.875
320000	47.07	0.375	0.322	0.357	0.067	0.038	0.641	0.921	2.141	2.437	8.181	0.344	0.141	5.718	
1600000	238.9	1.750	1.334	1.793	0.394	0.177	1.750	1.234	2.781	3.001	39.70	1.765	0.766	7.141	

LEO search times in seconds by SDK and population size, N.

Conclusions: Search durations scale approximately as a power of the database size. The coefficients are dependent on the algorithm. There are approximately two orders of magnitude difference in the search durations measured for the two most accurate algorithms.

⁸ This result differs from some applications [MINEXII] where matching times depend on whether a genuine or impostor comparison is being conducted.

INVESTIGATION 7. Verification accuracy

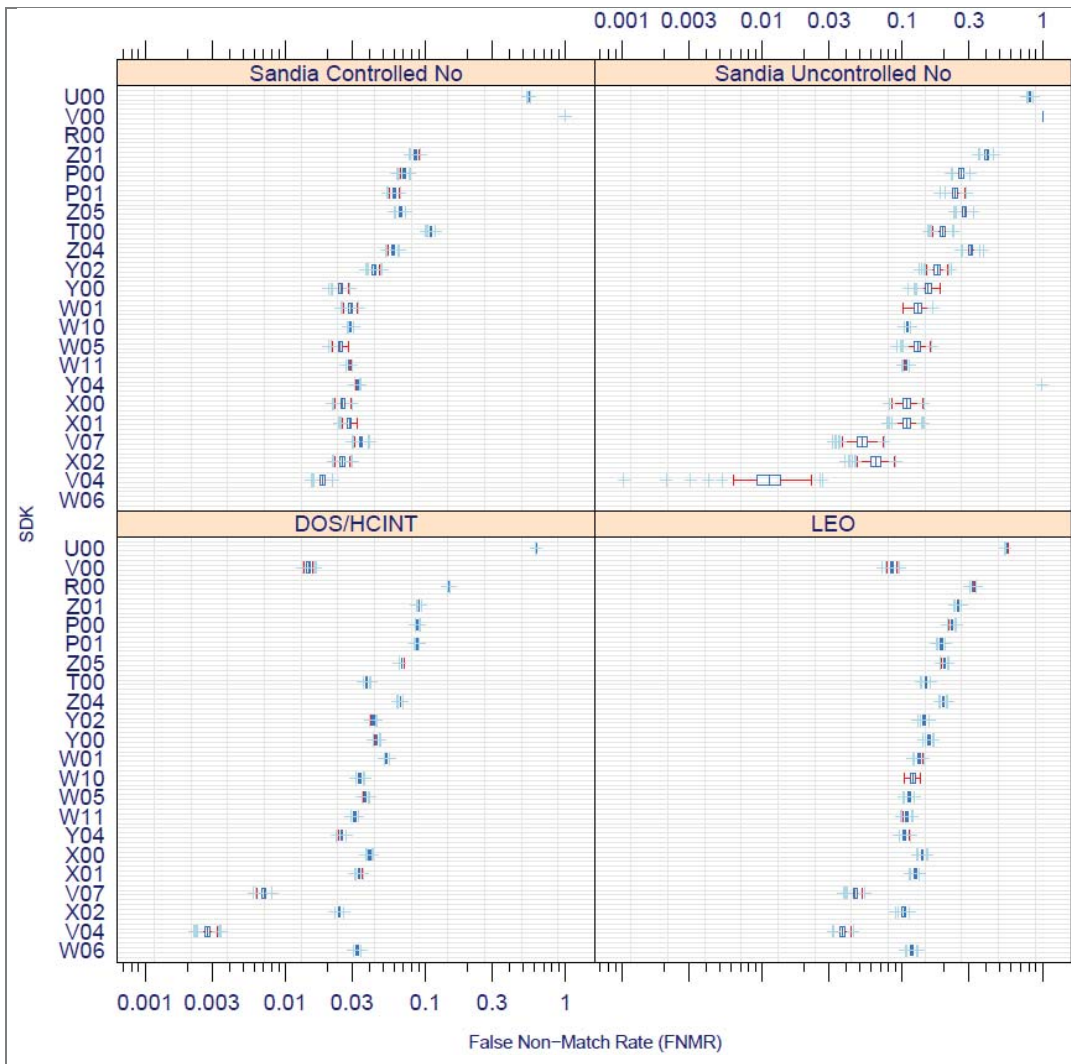
Face recognition is increasingly being used in access control systems. What is the accuracy?

Demand driver: The accuracy of the core verification algorithm is an important part of a face-based identity verification algorithm. However, unlike the case for identification, a recognition transaction may consist of several captures and comparisons, with the possibility to provide feedback to the user and to re-acquire a photograph. These aspects will produce better transactional accuracy. Identification systems, often incorporating backend matching systems, do not benefit from transactional cooperation of the user. Nevertheless accuracy of the core algorithm is influential on outcome.

Prior work: One-to-one verification has been measured in innumerable academic studies, and also in larger scale independent testing efforts [FRVT2002, FRVT2006].

Experimental method: Two datasets were employed. Forty thousand subjects randomly drawn from the LEO set formed the enrollment set against which 9240 individuals were verified. These comparisons produced the genuine scores. Single images from a further, disjoint, population of 10000 individuals were used to execute impostor comparisons. The second dataset, DOS / HCINT, was used in exactly the same manner as in prior NIST tests [FRVT2002, FRVT2006]. Twelve sets of 3000 persons were compared with 2 images of those persons to produce 12 times 6000 genuine scores. Those same persons were then compared with individuals from the next, disjoint, 3000-person set. This produced 18 million true impostor scores. All the persons had K=1 enrollment samples.

Figure 12 – Verification accuracies of class A algorithms.



Results: Figure 12 shows boxplots of FNMR for one-to-one verification accuracy for the DOS/HCINT and LEO datasets. FNMR is stated at FMR = 0.001. The threshold was set to give this FMR for each particular SDK and dataset. All the SDKs are class A, running without an enrollment database.

There is an order of magnitude variation in FNMR between verification algorithms running on the LEO images. For the DOS / HCINT images this increases to two orders of magnitude.

Conclusions: Error rates on DOS/HCINT have reduced dramatically in the last 8 years. In 2002, the best FNMR values at FMR = 0.001 was 0.2. This reduced to 0.026 in FRVT 2006, and to below 0.003 for the leading SDK in this test. The DOS images are overly compressed and exhibit some too-close-to-camera distortion effects.

INVESTIGATION 8. Verification accuracy with and without an enrollment database

Face verification can proceed by comparing a live capture with a face stored on an identity credential, or by comparing the live capture with an entry in an enrolled database. Is accuracy the same?

Demand driver: Face verification applications are deployed with and without a database of enrolled identities:

- The e-Passport gate task compares an image from height-adjustable cameras with the ISO/IEC 19794-5:2005 image read from the DG2 structure of the ICAO 9303 passport. There is purely comparison of images from the passport holder: There is no possibility to compare the image with other images of passport holders⁹.
- In a physical access control system (e.g. time-and-attendance, or a gymnasium access), all users could be enrolled and their templates maintained as an enrollment database. One might be selected via PIN or card presentation.

The use and maintenance of a face database is sometimes contra-indicated by communication constraints and by privacy policy considerations.

Prior work: To support the use of normalization across the enrollment dataset, the FRVT 2002 test allowed application of a post-processing function to the N scores produced by comparing a verification image with images in the enrollment set. The FRVT 2006 test explicitly solicited SDKs with and without normalization. In both cases, accuracy benefits were documented.

Experimental method: Using a fixed set of 40000 enrollment subjects, both genuine and impostor verification trials were conducted using class B SDKs. The class A implementations execute purely independent comparisons of template pairs. The class B implementations execute a comparison of a verification template with a specific enrolled identity and may internally compare just those two templates, or may undertake to utilize the remaining entries of the enrollment database in some effective way.

Results: Figure 13a shows FNMR at FMR = 0.001 for some providers who shipped both class A vs class B SDKs. Class A performance is represented by solid lines and class B performance by dotted lines. Curves are color coded by provider. False non-match rates at select false match rates are labeled. The Class B algorithms for V05/V08 produce identical scores to their V04/V07 class A counterparts. Thus, in the figure the dotted and solid line performance curves follow the same paths, with the result that only the solid lines are visible. This implies that the Vxx implementations do not perform any normalization across the enrollment dataset. For the P and X algorithms, the class B algorithms produce lower error rates than their corresponding class A algorithms.

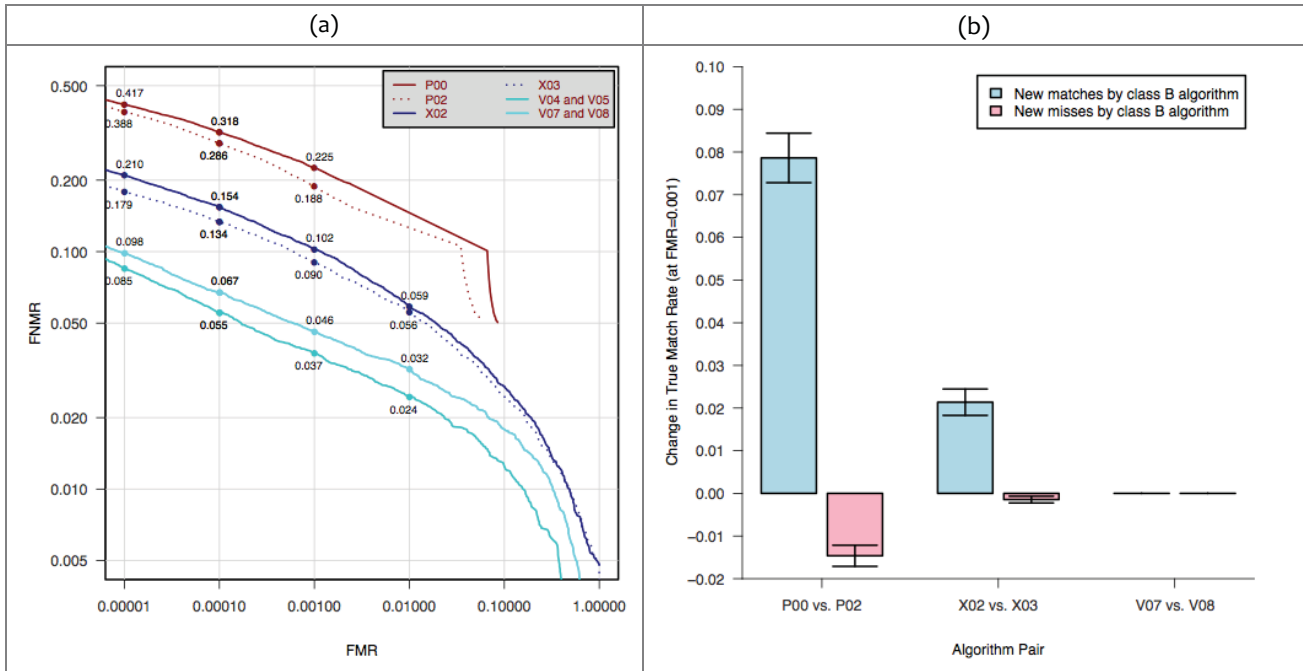
The barplot in Figure 13b shows how class B algorithms compare to their corresponding class A algorithms. The class B algorithms tend to perform better, but there are cases where people correctly matched by the class A algorithm are missed by the corresponding class B algorithm (indicated by the light-red bars in the figure).

Conclusions: The Class B algorithms from provider V give identical scores to their class A counterparts. The X03 algorithm performed better than X02. However, the best performing algorithms (V02 and V05) do not perform any normalization across the enrollment dataset.

⁹ A face recognition installation might include an internal normalization dataset. Such a set is unlikely to optimally represent the population of images that the e-Passport holder is presenting.

MBE-STILL REPORT PARTICIPANT KEY	P = PITTPATT	R = SURREY U.	S = TSINGHUA U.	T = TOSHIBA	U = DALIAN U.	PAGE 34 OF 58
	V = NEC	W = L1 IDENTITY	X = COGNITEC	Y = SAGEM	Z = NEUROTECHNOLOGY	

Figure 13 – LEO Verification accuracy with and without enrollment datasets



INVESTIGATION 9. Exploiting all prior images

If a facial recognition implementation is provided with the complete set of historical images of a person, does accuracy improve?

Demand driver: Many operational applications include collection and enrollment of biometric data from subjects on more than one occasion. This may be done on a regular basis, as might occur in passport issuance for example, or irregularly, as might happen in a criminal recidivist situation. In any case, the question arises whether accuracy can be improved if the face recognition implementation is allowed to exploit all prior images. This contrasts with typical practice in which the image from the most recent encounter replaces prior enrollments.

The number of images per person will depend on the application area:

- In civil identity credentialing (e.g. passports, driving licenses) the images will be acquired approximately uniformly over time (e.g. five years for a Canadian passport). While the distribution of dates for such images of a person might be assumed uniform, a number of factors might undermine this assumption¹⁰.
- In criminal applications the number of images would depend on the number of arrests¹¹. The distribution of dates for arrest records for a person (i.e. the recidivism distribution) has been modeled using the exponential distribution, but is recognized to be more complicated.

Fundamental concept: This document defines a template to be the result of applying feature extraction to a set of $K \geq 1$ images and merging the results. That is, a template contains the features extracted from one or more images, not generally just one. This is depicted in Table 20. The template is a single proprietary block of data. There are no facial template standards.




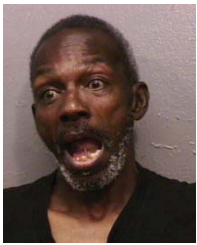

¹⁰ For example, a person might skip applying for a passport for one cycle (letting it expire). In addition, a person might submit identical images (from the same photography session) to consecutive passport applications at five year intervals.

¹¹ A number of distributions have been considered to model recidivism, see for example [BLUMENSTEIN]. "Random parameter stochastic process models of criminal careers." In Blumstein, Cohen, Roth & Visher (Eds.), Criminal Careers and Career Criminals, Washington, D.C.: National Academy of Sciences Press, 1986.

All verification comparisons and identification searches operate on such combined templates. This delegates the responsibility for fusion to the technology provider. This implies that end-users and system integrators should procure multi-image fusion capability; they should not implement this themselves.

Prior work: Use of multiple images per person has been shown to elevate accuracy over a single image [FRVT2002b]. While there are many academic publications in this area [MIN], many refer to the recognition-from-video problem which benefits from an ability to track the subject through time, and the use of single sensor. This covers early to late-stage integration strategies [SHAKNAROVICH]. The former is typically template-level fusion in which an algorithm might internally fuse K feature sets into a single representation. In the case of score-level fusion, the algorithm might match against the K feature sets separately and, for example, take the sum-score [KITTLER] or maximum score.

Table 20 – Uses of K images of a MEDS dataset subject for testing

	Enrolled images					Search (aka probe)
Image				...		
Encounter	1	2	3	...	K-1	K
Capture time	T ₁	T ₂	T ₃	...	T _{K-1}	T _K
Role RECENT	Not used	Not used	Not used	...	1 image enrolled	Probe
Role LIFETIME	N-1 images provided to SDK together and enrolled into a single template					Probe

Experimental method: Some of the proposed datasets includes K > 2 images per person for some subjects. This affords the possibility to model a recognition scenario in which a new image of a person is compared against all prior images¹². We ran two tests. The first shows the effect of using multiple images in a verification scenario. The second breaks out identification performance by the number of images enrolled. These are described as follows.

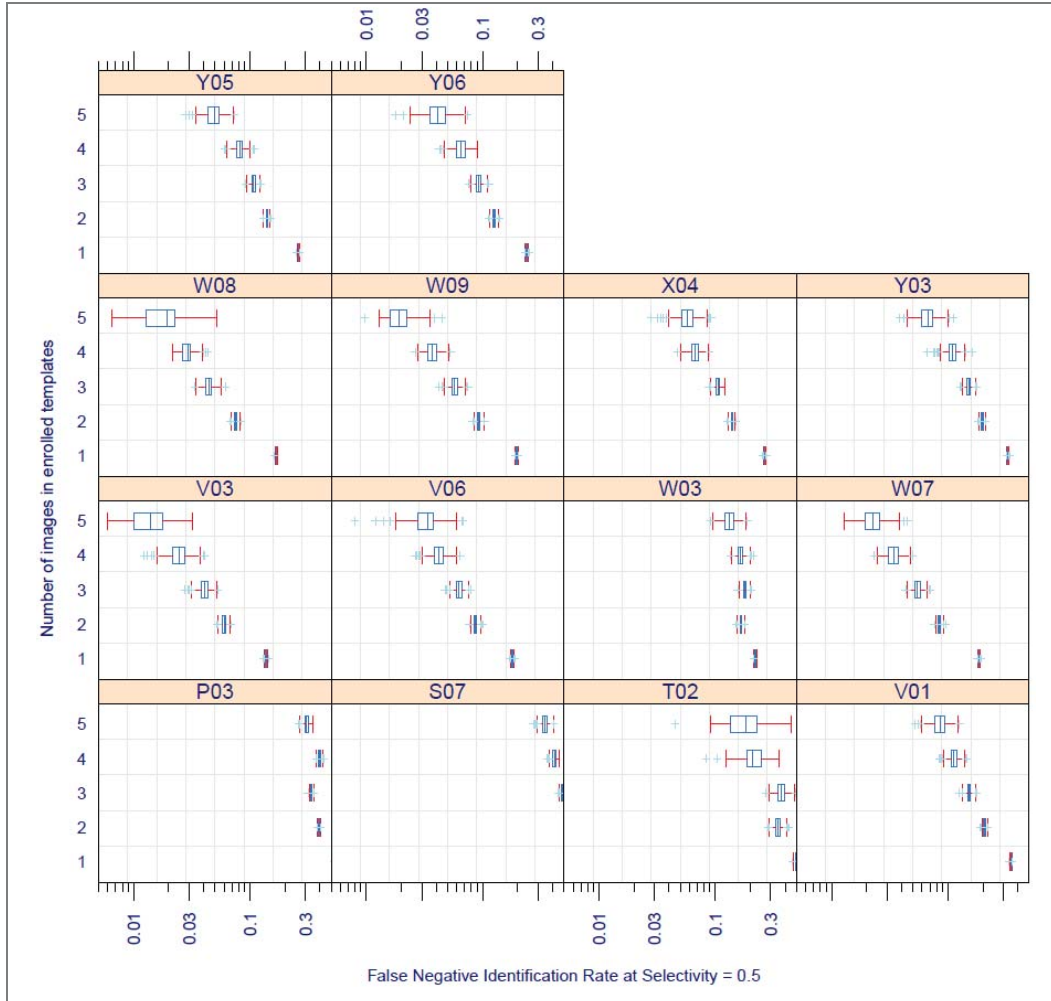
– **Identification:** The data and identification trials are identical to that used in investigations 1-4. The analysis is threshold based.

Verification: The verification test uses a population of 40000 persons from the LEO set. A fixed disjoint set of 10000 persons, one image per person was used to generate 400M impostor scores. A fixed set of 9240 persons was used to generate genuine scores. These were the most recent images of the enrolled subjects¹³. In the case where multiple images per person were enrolled, the total number of enrolled images was 45395. The method by which the face recognition implementation exploits multiple images is not regulated: The test seeks to evaluate vendor provided technology for multi-instance fusion. This departs from some prior NIST tests in which NIST executed fusion algorithms ([e.g. [FRVT2002b], and sum score fusion, for example, [MINEX]).

¹² For example, if a banned driver applies for a driving license under a new name, and the local driving license authority maintains a driving license system in which all previous driving license photographs are enrolled, then the fraudulent application might be detected if the new image matched any of the prior images. This example implies one (elemental) method of using the image history.

¹³ To mimic operational reality, NIST intends to maintain a causal relationship between probe and enrolled images. This means that the enrolled images of a person will be acquired before all the images that comprise a probe.

Figure 14 – LEO Identification accuracy by number of prior encounters

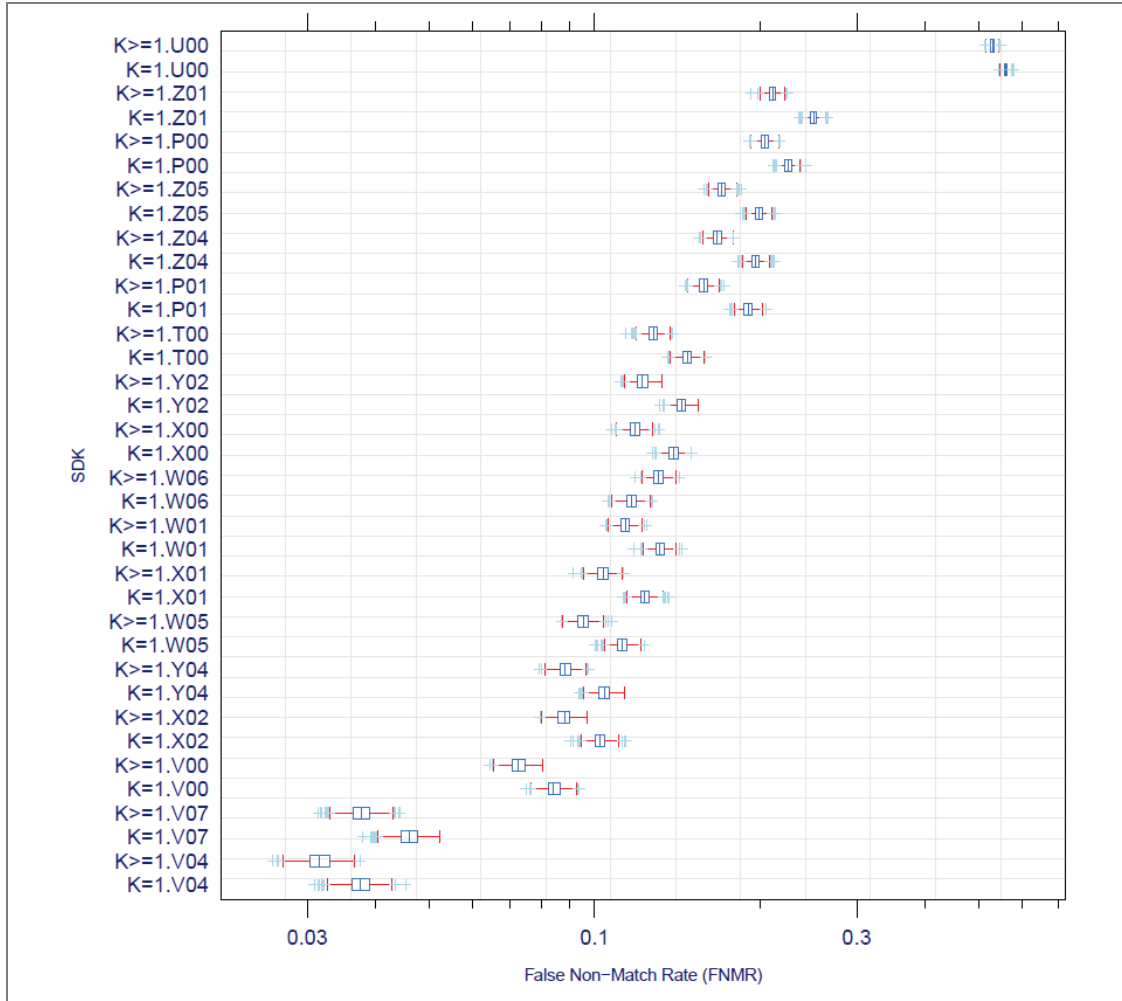


Results: Figure 14 shows identification accuracy broken out by the number of enrolled images per person. Specifically the plots show estimates of the FNIR(T) for persons enrolled with 1, 2, 3, 4 or 5 images, when the threshold T is set to produce a selectivity, $SEL(T) = 0.5$. Each boxplot summarizes 1000 bootstrap estimates of FNIR(T).

Some SDKs (P03, W03, for example) do not show accuracy gains from the use of multiple images. These SDKs may have elected to ignore all but the most recent, or the best enrollment image. Most SDKs do realize accuracy gains, and these are substantial: For SDKs W09 and V03, the FNIR values are between 5 and 10 times lower for persons enrolled with five images than with one.

The operational relevance of this result is dependent on the natural occurrence of multiple encounter data, the distribution of which is captured in the image counts of Table 6. To assess the overall effect we use verification results to quantify overall gains. Figure 15 gives boxplots of verification accuracies for 1:1 class A SDKs. Each boxplot summarizes 1000 bootstrap estimates of false non-match rate (FNMR) at a fixed false-match rate (FMR) of 0.001. The center of the box gives the median. There are two boxplots for each SDK. The first gives FNMR for the population when multiple enrolled images are used, i.e. $K \geq 1$. The second gives FNMR when identically one image per person is enrolled, i.e. $K = 1$.

Figure 15 – LEO Verification accuracy with and without multiple enrollment samples



Results from verification trials using LEO images show that FNMR decreases for all SDKs. The improvements are significant. Importantly, the overall benefit observed here depends on the fraction of the enrolled population with multiple encounters in the enrollment database. In the LEO population, the fraction of subjects who have multiple enrollment images is about 14%. The use of multiple-encounter data is effective because any given image may exhibit defects that cause recognition failure.

Conclusions: When all prior images or a person are enrolled under one identity, accuracy improvements in both verification and identification trials are realized. The value of multiple images increases with the number of images. Some algorithms exploit the availability more than others. The overall operational impact is related to the distribution of the number of images per person. Subjects from the LEO dataset with multiple images are more likely to be identified in a subsequent search.

INVESTIGATION 10. Exploiting all prior images: A false match hazard?

So by enrolling multiple images of a person, there is an increase in hit rate, but is there also a greater chance that un-enrolled subjects will false match against such enrollments?

Demand driver: The prior section showed that most algorithms realize reductions in the Type I error rate by exploiting the lifetime image history. However, there is a possibility that subjects enrolled with multiple images may attract more false matches. An enrollee producing elevated false-match rates is known as a lamb in the biometric zoo

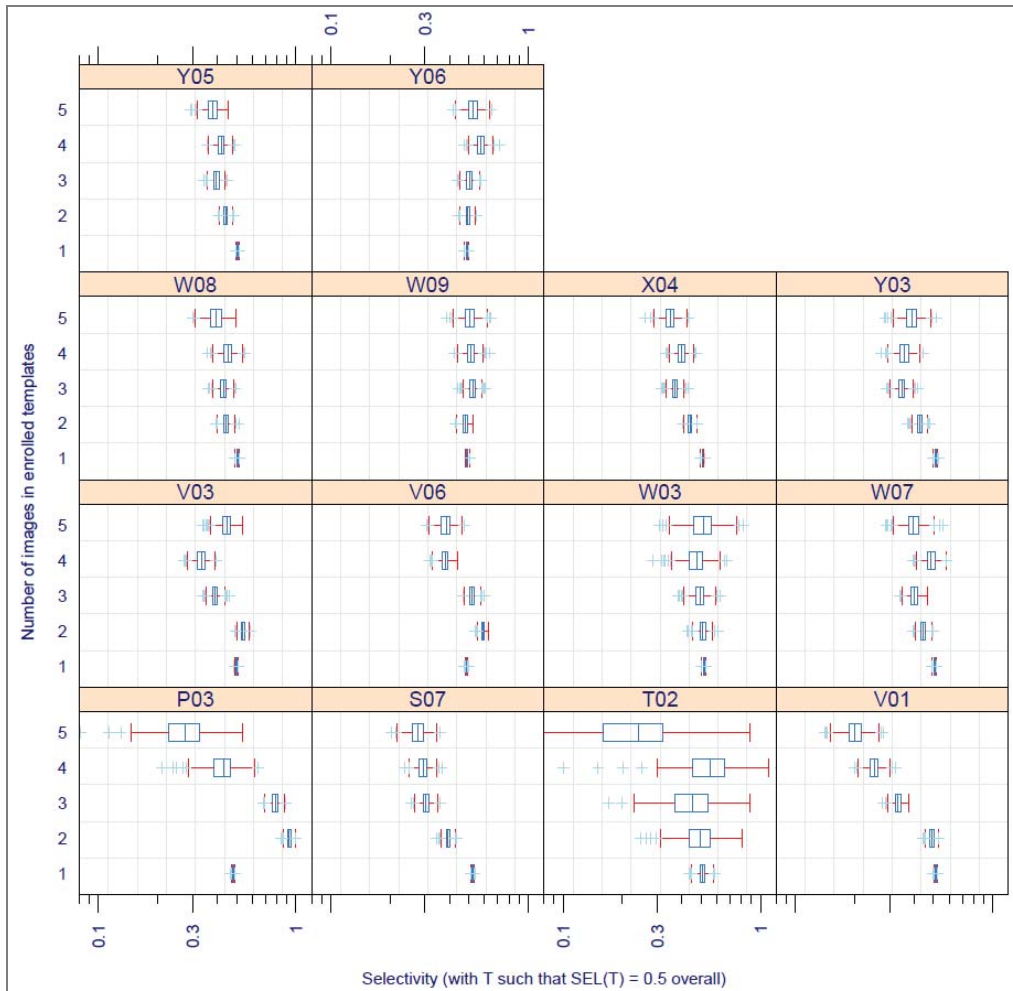
[DODDINGTON]. The face recognition algorithm, particularly its approach to fusion, is responsible for mitigating this risk.

Experimental method: As in prior investigation.

Results: Figure 16 shows, for each SDK, the selectivity associated with enrollees who have K=1, 2, 3, 4 and 5 LEO enrollment images in their template. The threshold is fixed to give an overall selectivity of 0.5. The desirable behavior is that selectivity is not a function of K. Also acceptable is that selectivity reduces with increasing K. The undesirable result an increase in selectivity because this would, depending on the application, give elevated workload to examiners.

Conclusions: The algorithms demonstrate benign false positive behavior when enrolling a person with $k > 1$ image.

Figure 16 – LEO Selectivity by number of prior encounters



INVESTIGATION 11. Evidentiary value

Can face recognition algorithms be used to support a statement of the form, "the chance that these two faces come from the different subjects is less than one in ten thousand"?

Demand driver: All biometric access control systems require a calibration of the false match rate so that a threshold, T, can be set. A score exceeding the threshold might result in an undetected false acceptance. Likewise, in an identification application, a high score might trigger an investigation, typically involving human adjudication of the result. The threshold T is used to implement a decision policy. It will usually be set according to stated cost considerations, estimates of impostor likelihoods, and an empirical calibration of the recognition algorithm response.

The intention then is to set a threshold so that the chance of two different persons matching is actually less than an FMR requirement. In practice, while the threshold might be set on the basis of theoretical considerations, or on an ad hoc basis, it is in practically set empirically. Thus, a calibration exercise is undertaken: A particular facial recognition algorithm is used to compare many images from a large population of persons, and the empirical cumulative distribution function of the observed impostor scores, $N(s)$, gives an estimate of the chance that images of different persons will reach a high value, s , via

$$P(s | \text{impostor}) = \text{FMR}(s) = 1 - N(s)$$

The use case is as follows. A laboratory is given two images and is asked do the images come from the same person. The lab passes the images through a one-to-one comparison engine to obtain a comparison score. If the score, s , is far above the mean of the observed impostor distribution, then the $\text{FMR}(s)$ calibration serves to assert that the chance we'd see this outcome from different persons is less than 1 part in a million, say.

The argument here is that a high score would imply the persons are same. The opposite is not true: a low score does not necessarily mean the images are of different persons. This arises because defective images produce low scores even in same-person comparisons. The term defective might mean low contrast, blurred, non-frontal pose, and exaggerated expression.

Table 21 – Interpretation of impostor scores

Score	Image properties	Supported conclusion
High	Good quality	Same person – but see the critical caveats below
High	Poor quality	This outcome should not occur
Low	Good quality and no sign of manipulation or evasive behavior	Different persons, modulo any
Low	Poor quality	Indeterminate

The actual reliability of the threshold will depend on the impostor distribution stability over the lifetime of the operation. However a number of factors may undermine the calibration. These include:

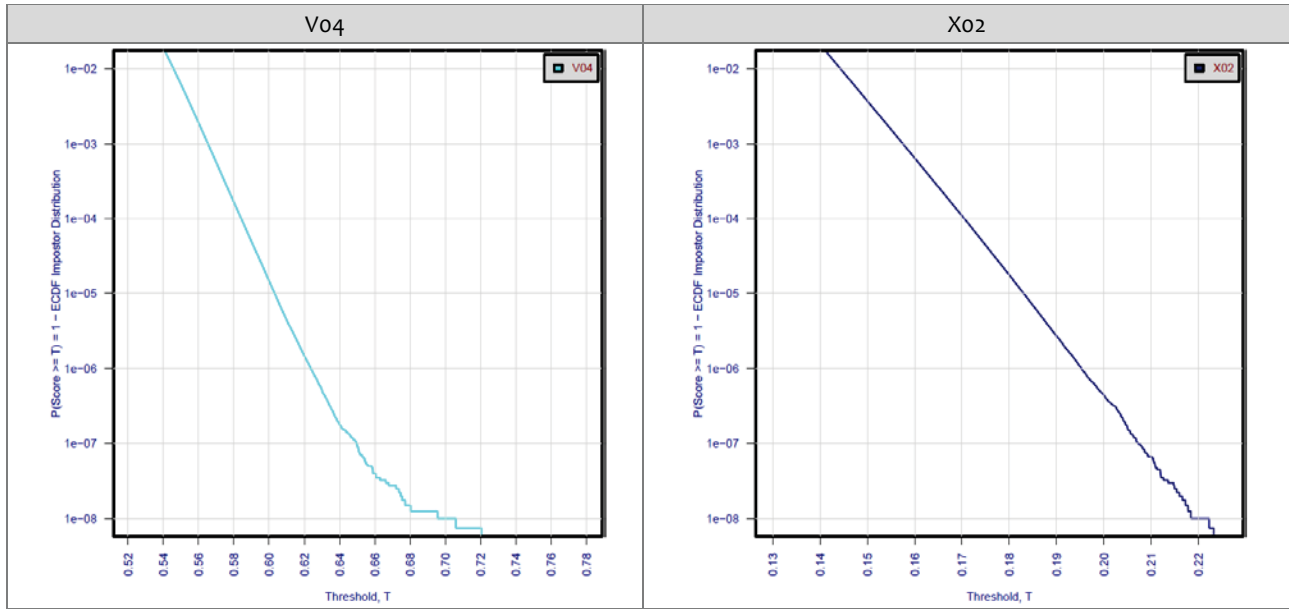
- Changes in the photometric and geometric properties of the images.
- Changes in the compression applied to the images.
- Changes in the demographics of the population.
- Changes in the ethnic mix of the population.

A very important caveat on this analysis is that the calibration method here is based on image corpora that do not include all photometric and geometric variants. For example, it does not include highly compressed images, taken with a flash, that exhibit fish-eye distortion. It is possible that such images, from truly different people, might give impostor scores much higher than those used to establish the calibrations here.

Experimental method: NIST used the class A SDKs to execute 400 million one-to-one comparisons of LEO images, and to use the resulting scores to compute empirical cumulative distribution functions.

Result: Figure 17 shows plots the right tail of one minus the empirical cumulative distribution function (ECDF) of the impostor distribution, i.e. they plot FMR against threshold, i.e. $\text{FMR}(T)$. The vertical axis is logarithmic. The curves are uninteresting in the sense that they show the expected monotonic decrease of FMR with threshold. Their utility, however, is as calibration curves - they allow a user to set a threshold to target a particular FMR. The calibration is worthwhile only to the extent that the impostor distribution is stable for the lifetime of the operation.

Figure 17 – LEO False Match Rate calibration curves



Conclusion: The calibration curve supports a statement of the form: “According to the automated face recognition calibration, the likelihood that this image pair come from different people is 1 part in 300,000”. The method is applicable in identification scenarios if any given candidate from an identification search is compared with the hypothesized enrollment sample using the 1:1 SDK.

Caveats: The above results are subject to the following caveats and assumptions, and should be used with caution.

- **Algorithm effects.** The calibration may not apply if the face recognition implementation changes. This would include any change in the entire algorithmic chain: front-end image analysis routines (e.g. face and eye detection); image processing routines (e.g. morphable models to correct for pose), and the feature extraction and back-end matching algorithms.
- **Database effects.** The calibration only applies to images with the properties present in the set.
- **Finite population.** While the curves were estimated using from 400 million comparisons, the population size is still only of size 40000 enrollees, and 10000 impostors.
- **Familial similarities:** The physical structure of the face is genetically linked such that the closely similar facial appearance of identical twins¹⁴ changes only over decades of environmental and developmental influences. Twins have been exploited for criminal purposes [BASIA]. Some face recognition systems appear able to differentiate identical twins by extracting information from the skin texture. Going further, parental and sibling appearance similarities are obvious hazards - the extent to which these could produce anomalously high similarity scores has not been studied here.
- **Ethnic origin:** The calibration applies to the specific population used in the test. This is a mixture of the U. S. persons many of whom have global ethnic ancestry. It is well known that size and appearance are dependent on national and even regional origin. In addition, globalization introduces a time dependency to those categories.
- **Deliberate image manipulation:** It is known [GALBALLY, ADLER] that an image can be automatically adjusted to produce a false match. This has been conceived of to defeat an access control system. While the images produced in an automated hill-climbing scheme can deviate from usual human-form, the extent to which an image can be manipulated to produce a high comparison score while preserving human form is clearly large when the manipulator is a skilled human [MORPH]. In any case, any such activity would be deliberate and fraudulent –

¹⁴ And triplets, quadruplets, etc. The natural incidence of n-tuplets is small, and tends to include fraternal rather than identical siblings.

the risk of this would be mitigated by the usual evidentiary controls, and could be detected by forensic data analysis [FARID].

INVESTIGATION 12. Dependence of accuracy on pose
In 2004, the ISO/IEC 19794-5 standard established limits for deviations from frontal pose (5deg, 5deg, 8deg). These were instituted because pose was known to known to adversely affect facial recognition. Is this still true?

Demand drivers: Previous evaluations have demonstrated that deviations from a fully frontal pose adversely affect recognition accuracy, making individuals more difficult to recognize. The ISO and ANSI/NIST standards limit deviations to avoid this problem, but previous studies have demonstrated that a significant portion of the images in IAFIS have poses deviating more than this requirement.

Prior work: There is an enormous literature on both pose estimation and on improving face recognition under three-axis rotation of the head.

Experimental method: The LEO images were accompanied by subject-specific and image-specific metadata values. In particular each image was accompanied by an estimate of the head pose. Head pose is quantified using the Tait-Bryan angles roll, pitch, and yaw. The estimates were produced by an SDK supplied by an MBE-STILL participant. The SDK was supplied to, and used by, an organization involved in the preparation of the data. It was not used by NIST. The SDK reported an estimate of yaw and roll. All estimates of pitch were zero degrees. The probable reason is the well known lack of a datum for zero pitch in a frontal image.

Yaw measures the degree to which the head is facing left or right (see Figure 18), while roll measures the amount of in-plane rotation of the head (or the camera) about the roll axis.

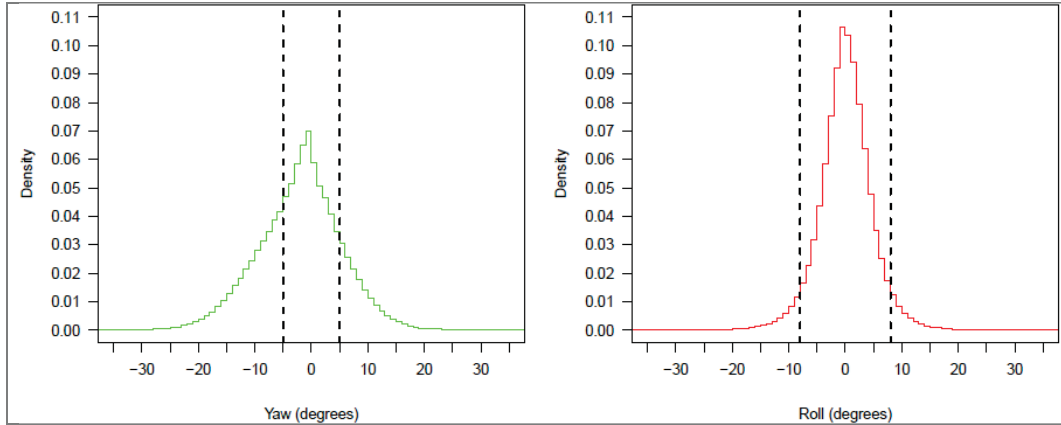
Figure 18 – Face images from the FERET database demonstrating varying amounts of head yaw.



Figure 19 shows the measured distribution of yaw and roll for 590,105 LEO face images. Of these images, 33.6 percent had yaw measurements between -5 and 5 degrees, 86.2 percent had roll measurements between -8 and 8 degrees, and 30.2 percent had yaw and roll measurements within both ranges. Thus, according to these measurements, less than a third of the images fall within the ISO/IEC 19794-5 limitations for deviations from a frontal pose. The black vertical lines highlight the ISO/IEC 19794-5 best practice recommendations for the minimum amount of pose deviation.

Note however that because pitch estimates were unavailable, many images for which yaw assumes a good value will include a significant pitch deviation from what a human observer would consider zero degrees.

Figure 19 – Histograms for yaw and roll angles for LEO images.



Experimental method: For each SDK we generate templates for all LEO images with one or more mates. We then compute the 1:1 comparison score for each mated pair. This produces 590105 scores. For each pair, we lookup the two yaw-angle estimates.

Results: Figure 20 shows the effect of yaw on the FNMR. The yaw for each comparison was taken as the maximum of the two face images. For most algorithms, error rates are increased when the yaw angle is between 6 and 16 degrees from frontal. Catastrophic failure tends to occur when the yaw angle is greater than 20 degrees. V07 appears the most robust to small-to-moderate deviations in yaw, having the widest U-shaped curve.

Figure 20 – Dependence of accuracy on face yaw angle

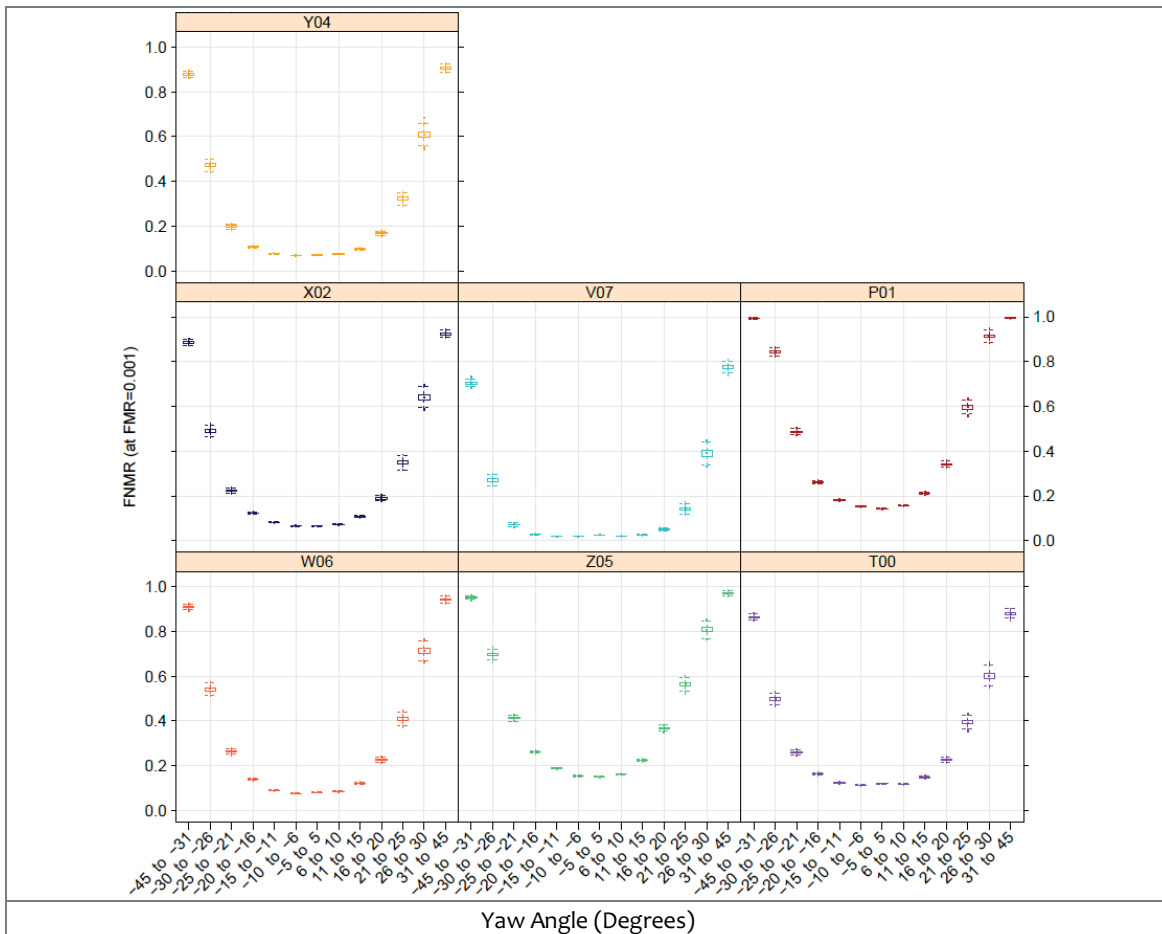


Figure 21 uses *heatmaps* to show the effect of head yaw on the FNMR. Darker colored cells signify higher error rates. This visualization technique has been used before [GROSS] to show increased resistance to pose variation. White colored cells indicate no data was available to compute an FNMR for that cell. The lighter colored cells tend to lie along the main diagonal, indicating error rates are lowest when both face images have similar amounts of yaw. However, unless the yaw angle is similar in both images, catastrophic failure tends to occur if the yaw is greater than 20 degrees in either image.

Figure 21 – Dependence of LEO accuracy on yaw angle of enrollment and verification images

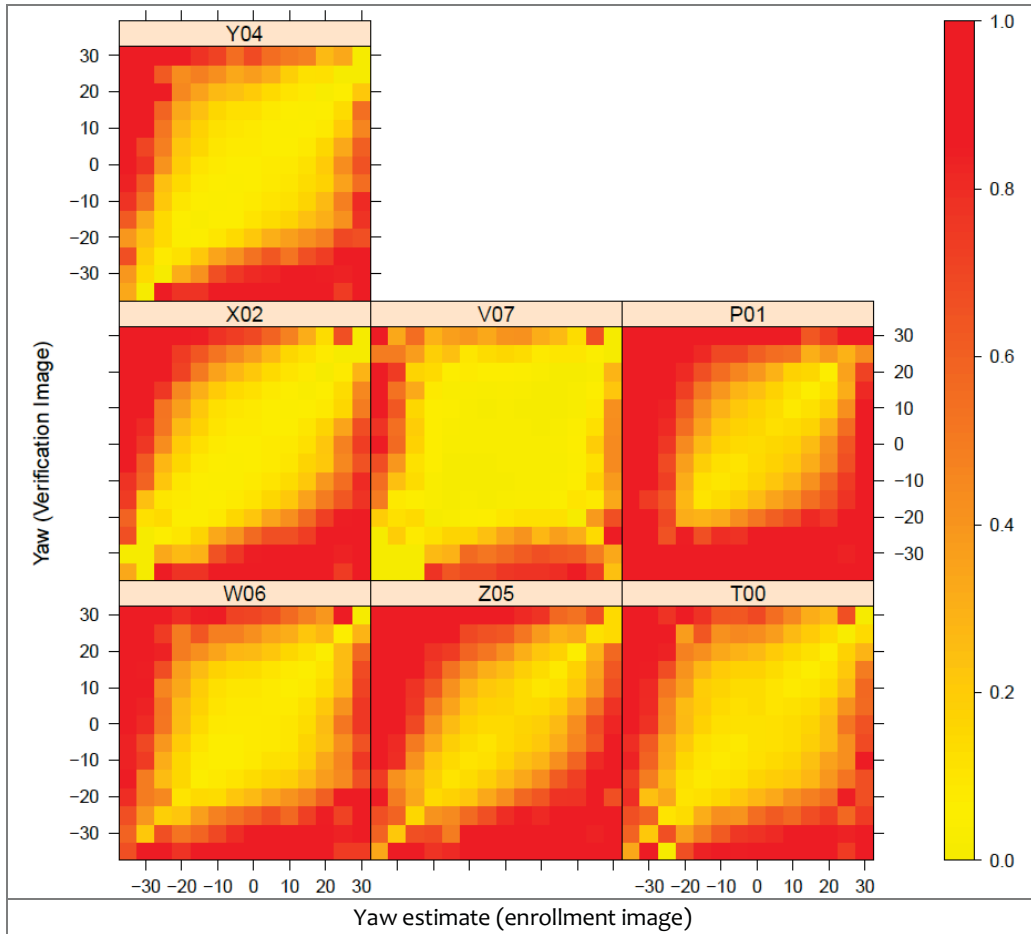
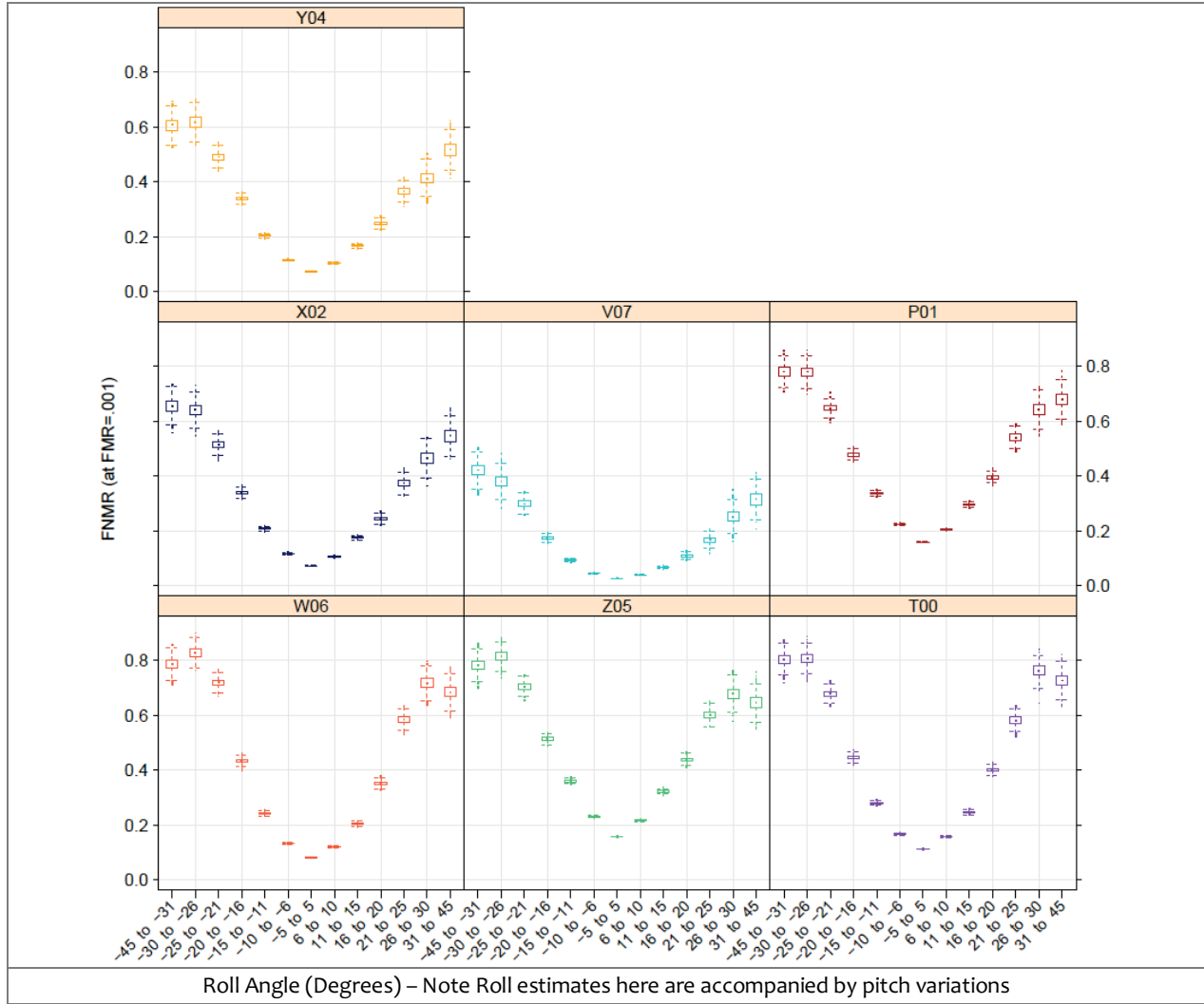


Figure 22 shows the relationship between reported head roll and FNMR. The large increase in error rates as a function of measured roll are mostly due to the roll measurement being accompanied with unmeasured pitch variations (i.e. compound rotations of the head): high amounts of roll also tend face downward. In many cases, high amounts of head roll were reported for images where the eyes were incorrectly located, usually due to considerable amounts of glare from eyeglasses. In these instances, the images suffer from poor sample quality due to a reason other than what was reported. Nevertheless, when a high amount of head roll is reported, it is often indicative of other problems with the image.

Figure 22 – Dependence of LEO accuracy on reported roll angle



Roll Angle (Degrees) – Note Roll estimates here are accompanied by pitch variations

Conclusions: While the pose problem has received considerable attention in the academic literature, most algorithms tested here will give increased error rates when non-frontal images are acquired and passed on to recognition engines. However, some algorithms are less sensitive to pose angles than others.

INVESTIGATION 13. Template size

How big are facial recognition templates? How big are facial images? How do these sizes compare with those of other biometric modalities?

Demand driver: Templates contain the mathematical representation of one or more images of a person. Biometric templates are proprietary, non-standard¹⁵, and their content is protected as a trade-secret.

Template size is clearly influential on storage requirements, both on-disk and in-memory, and on transmission bandwidth requirements. In addition, a large template may be associated with computational complexity and computational expense of the matching algorithm.

Experimental method: The MBE-STILL CONOPS Evaluation Plan and API¹⁶ explicitly supported measurement and reporting of facial recognition template size. When NIST passed $K \geq 1$ images to the implementation under test, we

¹⁵ Fingerprint minutia templates are the exception in that they are standardized [1378]. While standardized templates can be interoperable (across providers), they offer accuracy below that of proprietary templates [MINEX].

pre-allocated KT bytes, where maximum template size, T, was returned by a function call provided by the implementation under test. The function returned two values, one for maximum enrollment template size, and one for maximum verification or identification template size. For any given input, the actual template size was returned and used to save the template to disk.

Table 22 – On-disk template sizes by SDK and template role

SDK	Class	Enrollment	Verification	SDK	Class	Enrollment	Identification	Notes
P00	A	31349	31349	P03	C	31344	31344	During identification searches the SDK was allowed access to the enrolled templates on hard disk. That is, the API in no way required that all N templates be kept in memory, or to keep whole templates in memory.
R00	A	27500	27500					
R01	A	44200	44200					
S00	A	5520	5520	S06	C		8276	
				S07	C	5520	5520	
T00	A	21760	21760	T02	C	21760	21760	
U00	A	2200	2200					
V00	A	5025	5025	V01	C	5025	5025	
V04	A	5069	5069	V03	C	5069	5069	
				V06	C	2553	2553	
W01	A	5712	18196	W03	C	5698	18196	The SDK was free to initiate disk access, and to do partial reads of the enrolled data (via, for example, fseek, fread, mmap). This may have been done conditionally, for example reading in proprietary data blocks only during end-stage matching of high scoring candidates.
W05	A	5712	18196	W07	C	7775	20273	
W06	A	5698	61830	W08	C	8556	21068	
W10	A	7775		W09	C	8556	21068	
W11	A	6143						
X00	A	4304	4304					
X01	A	7240	7240	X04	C	7376	7376	
X02	A	7376	7376					
Y00	A	5320	5320	Y03	C		5320	
Y02	A	5752	5752	Y05	C	74056	74056	
				Y06	C	74056	74056	The test did not make measurements of peak or mean memory usage during a search.
Z01	A	20488	20488	Z03	C	20488	20488	
Z04	A	24396	24396	Z07	C	33484	33484	
Z05	A	33484	33484					

Results: Table 22 shows template sizes in bytes. These sizes reflect the size of the template in permanent storage (hard disk). More than one provider noted that the matching system does not need to load the entire template into memory for a search. We make the following observations.

- In all cases the size of an enrollment template is independent of the size of the enrolled population. This is not necessarily so because the API supported variable size templates by informing the SDK during initialization of the number of subjects about to be enrolled.
- In all cases except two, the size of the enrollment template size grows linearly with the number of images that went into its creation. The two exceptions are S00 and W03.
- The API supported asymmetric or role-specific templates. This allows a template to be used only for enrollment, or only for recognition but not vice-versa. Many template sizes are independent of role. Notably the enrollment templates for Vendor W are smaller than verification and identification templates. Operationally, a verification template is not stored permanently – it exists only for the duration of a recognition transaction.
- Template sizes are generally larger than those reported for iris recognition [IREX – 10 providers, 19 algorithms].
- Some modest reductions in size are possible via lossless compression (e.g. bzip2).

¹⁶ See http://face.nist.gov/mbe/MBE_STILL_Eval_Plan_v1.pdf

As noted previously these template sizes may actually be compound sizes of fast-search and end-stage matcher templates.

Conclusions: Template sizes vary between 2KB and 75KB with strong vendor dependence. Template sizes for verification and identification images sometimes differ from those of enrollment images. Note, however, that several vendors demonstrate an ability to tailor template size quite considerably.

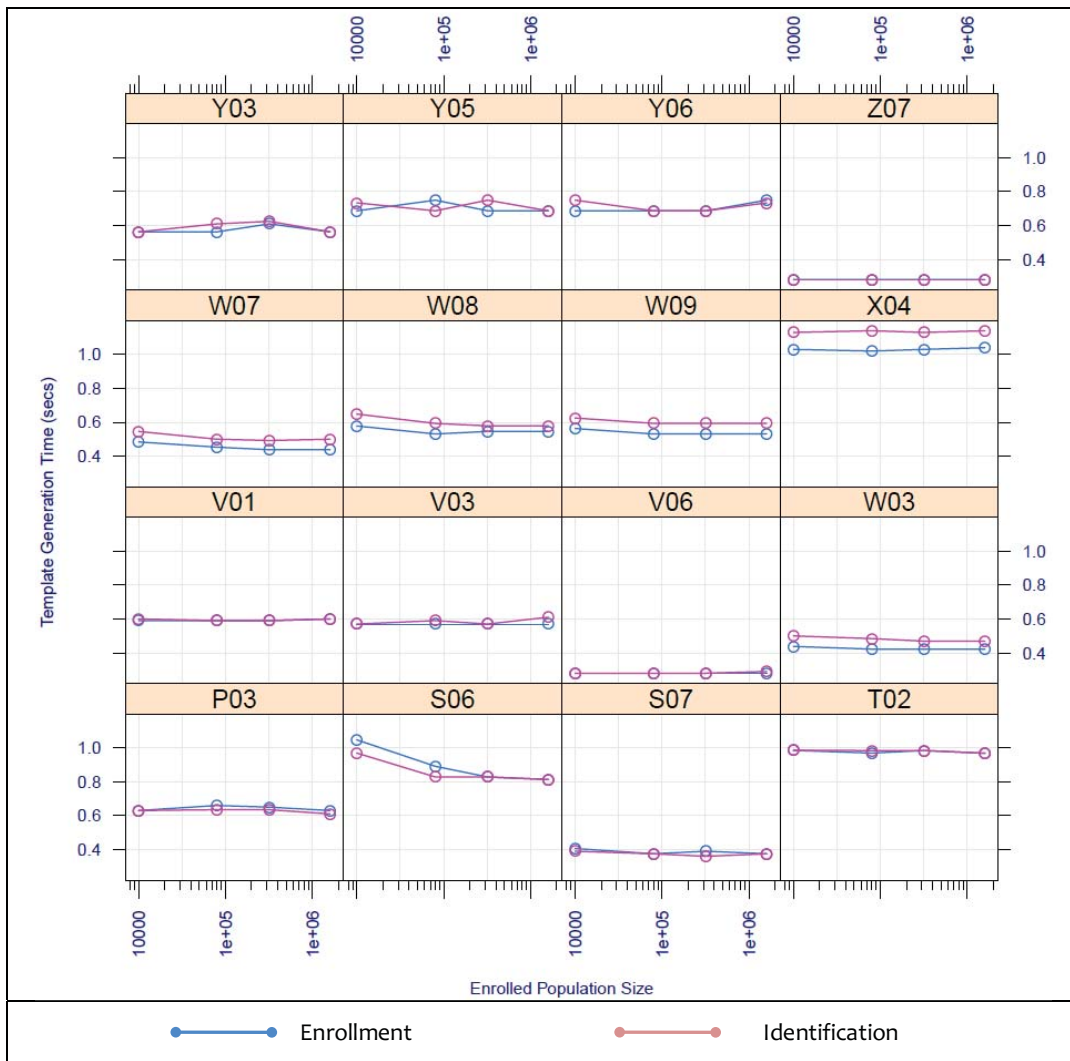
For comparison, templates sizes associated with standardized minutia templates are 0.1 - 0.8 kilobytes [MINEX], with iris recognition 0.5 - 10 kilobytes [IREX], and 1KB to 100KB for proprietary fingerprint templates [PFT].

INVESTIGATION 14. Template creation time

How long does it take to extract features from an image and make a template? Does this depend on the width and height of the input image? Does it depend on whether the template is used for enrollment, verification or identification?

Demand driver: Template generation time is often a large component of a 1:1 authentication attempt. For 1:N searches, the fraction will depend on N. Template generation time will be important if an existing image corpus is going to be re-enrolled by a new provider. For example, re-enrollment of a 18M person driving license database takes $1 \times 18 \times 10^6 / 64 / 3600 = 156$ hours if a one second template generation were sustained on a 32 core blade installation.

Figure 23 – Duration of LEO template generation calls



Experimental method: Each template creation function call was wrapped in a timer.

Results: The median duration of the template generation function is reported in Figure 23. The units are seconds. Each panel includes two traces, one for enrollment templates, one for verification templates. One SDK took longer than the 1 second template extraction time limit established in the MBE-STILL API. Note, that algorithm developers did not have access to the target machine, nor to detailed statistics on image dimensions.

Conclusions: Template creation times are independent of the target population size, suggesting that developers did not tailor their algorithmic representation to the size of the identification search.

INVESTIGATION 15. Link between sex and accuracy
Are photographs of one sex more readily recognized than those of the other?

Demand driver: Face recognition algorithms should not be too biased in how they treat individuals having certain demographic traits. Preferably, males should not produce score distributions substantially different than females.

Prior work: Previous evaluations have demonstrated that males are easier to recognize than females (FRVT 2002).

Experimental method: MBE-STILL separated 590,105 genuine comparisons from the FBI set into male and female sets. For each algorithm, a distribution of FNMRs (at FMR=0.001) was computed for each sex using 2000 bootstrap iterations. The resulting boxplots show how false non-match error rates differ for the two sexes. The false match rate used in the plots was computed using results from 1:1 comparisons of LEO images.

Certain genuine comparisons were excluded from consideration based on the following criteria:

- If the recorded sex of the individual was not consistent across all captures for that individual.
- If the sex was specified as “Unspecified” or “Unknown”.

Results: Table 23 shows FNMR for class A verification SDKs broken out by sex.

Table 23 – LEO Verification accuracy by sex

Class A SDK	FNMR at FMR = 0.001		Notes
	Male	Female	
Wo6	0.111	0.113	The standard errors for these measurements are about 0.001 for females and 0.0004 for males.
Zo5	0.193	0.200	
To0	0.145	0.137	
Xo2	0.095	0.109	
Vo7	0.039	0.042	
Po1	0.187	0.214	
Yo4	0.094	0.111	

Conclusions: Males generate fewer false non-matches than females for five of the six algorithms, although the disparity is small in every case. Since the FRVT 2002 Evaluation, the link between sex and genuine scores appears to have diminished. However, this may arise because different datasets with different demographic properties were used (LEO vs. DOS/HCINT). The relationship between sex and impostor scores was not investigated here.

INVESTIGATION 16. Link between subject age and accuracy
Are older subjects easier or more difficult to recognize?

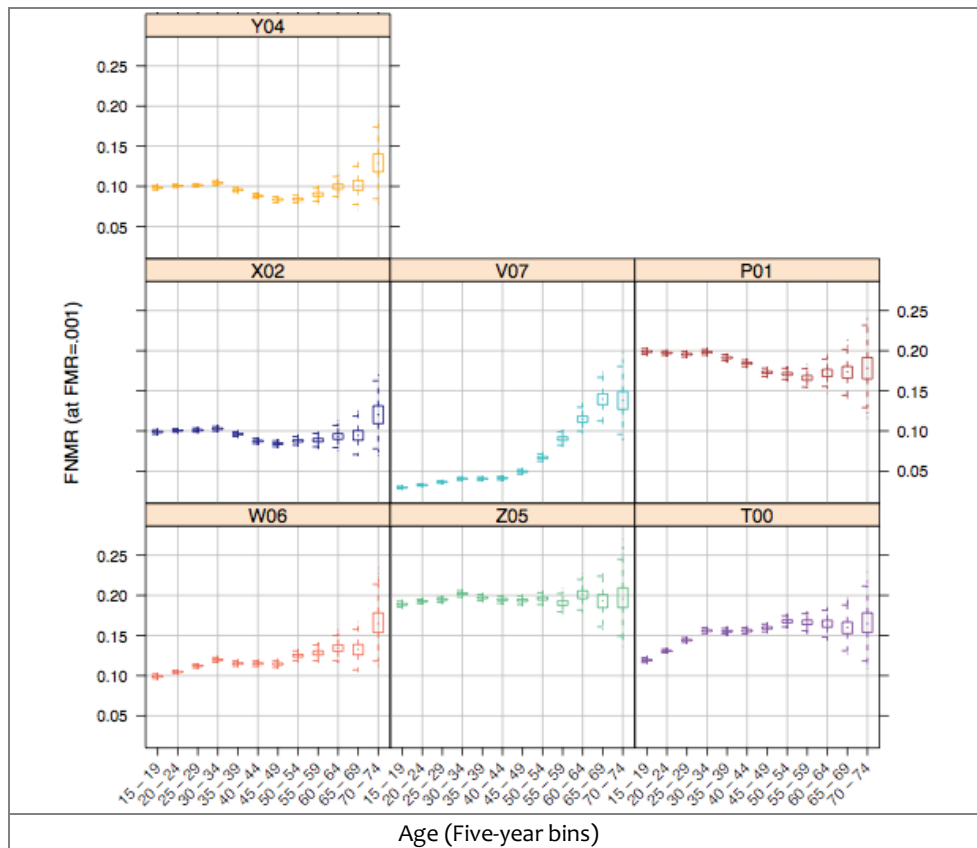
Demand driver: Face recognition algorithms should not be too biased in how they treat individuals having certain demographic traits. In addition, investigating a possible age effect can potentially identify aspects of automated face recognition that require improvement.

Prior work: Previous evaluations have demonstrated that older individuals are easier to recognize than younger ones [FRVT 2002].

Experimental method: MBE-STILL binned 590,105 genuine comparisons from the FBI set into 5-year age groups. For each algorithm, and within each age group, a distribution of FNMRs (at FMR=0.001) was computed using 2000 bootstrap iterations. The resulting plots show how false non-match error rates differ for the different age groups. Age was assigned to genuine comparisons based on the time elapsed between the individual’s birth date, and the date at which the first (i.e. oldest) image was captured. The false match rate used in the plots was computed using results from 1:1 comparisons of LEO images.

Results: While an age effect is clearly displayed for most algorithms, the precise behavior differs for each algorithm. The most defined trend is with V07, where an older individual (≥ 60 years old) is several times more likely to be missed than a younger person (< 30 years old). Most of the other trends are not as severe. Nor are they monotonic, since a jump in the FNMR is often present around the 30-34 age group. While an obvious concern is that the age effect may be confounded with the time elapsed between photographs, this doesn’t seem likely for the V07 SDK which shows resistance to elapsed time (see Figure 26).

Figure 24 – LEO Verification accuracy by age of subject at most recent capture



Conclusions: The effect of subject age depends on which algorithm is used. In most cases the effect is small, and smaller than that reported previously [FRVT2002]. The one exception, for one of the more accurate implementations, is an increase in FNMR of more than a factor of five. While, this result may prompt consideration by the developer, the effect is again subject to confounding factors in the data.

INVESTIGATION 17. Face ageing

Faces change over time. While no large and long-term face image collection exists, has the resistance of face recognition algorithms to age-related changes improved since it was reported in FRVT 2002?

Demand driver: False rejection errors will increase if the facial appearance changes significantly over time. The causes are not limited to just ageing. Other drivers include weight change, sun exposure, drug use, facial hair growth or removal, and posture changes related to skeletal changes (scoliosis). While research has been conducted to model

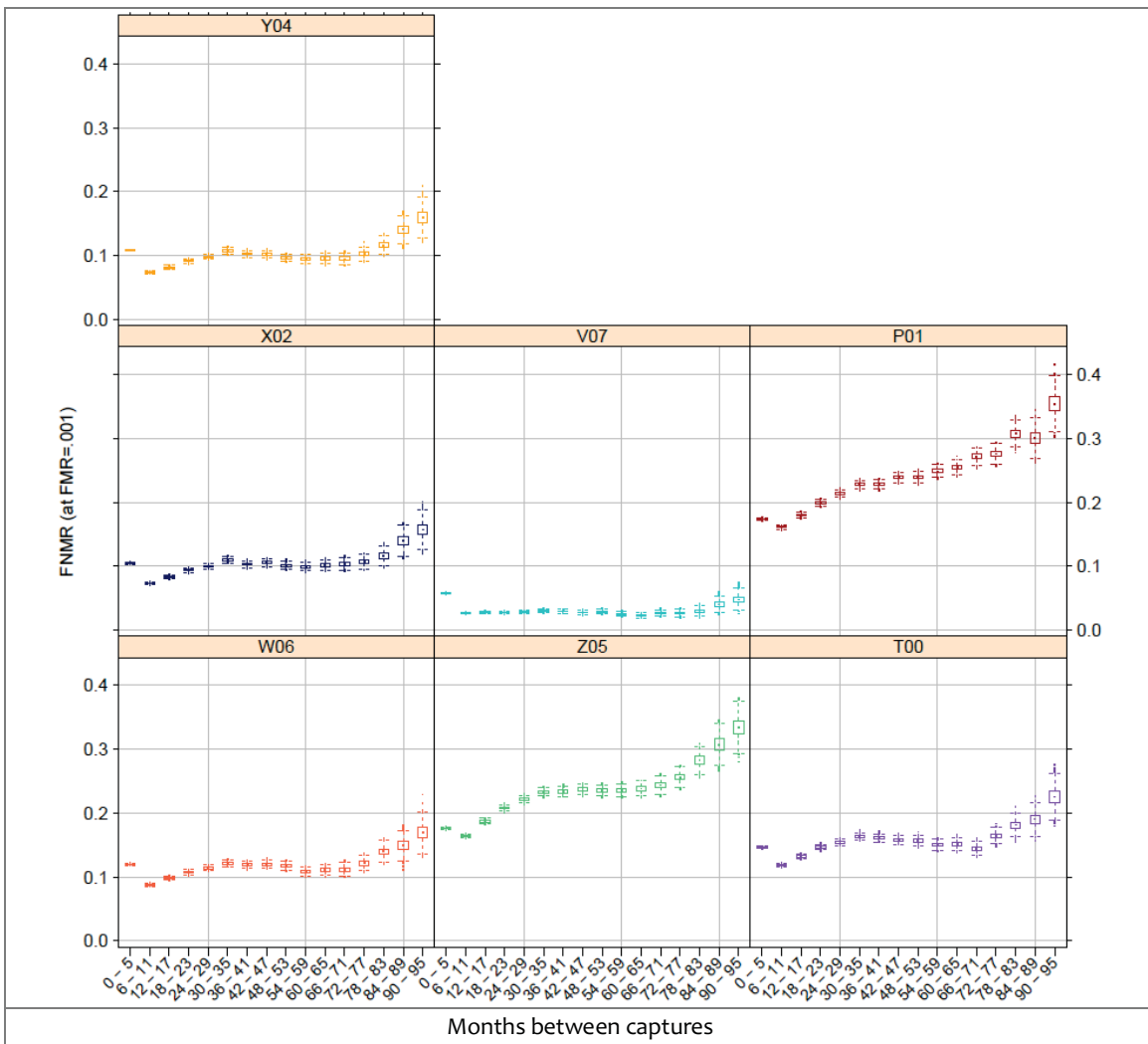
ageing [RAMANATHAN] and to build age-independent representations [PARK], the primary means to mitigate time-related changes is to re-enroll cooperative users, when it is practical and cost-effective.

Prior work: Several projects have specifically set out to collect facial images for the purposes of supporting research and development in this area (See particularly the large datasets collected under the MORPH project (Craniofacial Longitudinal Morphological Face Database), www.faceaginggroup.com, and the FG-NET work www.fgnet.rsunit.com. NIST's primary approach has been to leverage operational data for which date-of-capture metadata is available.

Experimental method: A key assumption of the analysis is that while any given pair of face images might yield a false rejection due to non-age related reasons (typically pose, illumination or expression effects), the average over a large number of facial comparisons will quantify age-related effects. This also assumes that there is no systematic change in the imaging collection practice and design over the interval.

Results: Figure 26 shows an increase in FNMR occurs as the time between captures increases, although some algorithms appear more robust to aging than others. The horizontal axis stretches from 0 months to 95 months (~8 years). Based upon a visual inspection, the increased FNMR for the 0-5 month bin is the result of same-day profile images being mislabeled as frontal. For most of the algorithms, the FNMR increases as the time between captures extends from 6-11 months to 30-35 months. Beyond that, many of the algorithms display a counter-intuitive decrease in the FNMR as the time between captures extends from 30-35 months to about 66-71 months. This may be the result of hidden factors (i.e. there may be some property of the face images within this range that makes them easier to recognize that is not directly related to aging of the face).

Figure 25 – LEO Verification accuracy by time elapsed between photographs



Conclusions: For most verification algorithms, false non-match rates increase by roughly a factor of two over the eight year interval represented in the LEO dataset. Any proposal to extend re-enrollment intervals for face-based verification systems is not supported by the results here. However, one algorithm does exhibit greater resistance to elapsed time. This is unlikely to be a random effect, but given the presence of confounding factors, such as subject age, a more detailed statistical analysis is warranted.

INVESTIGATION 18. *Is subject weight influential?*
Are photographs of lighter subjects more readily recognized than those of heavy subjects?

Demand driver: A person’s weight is reflected in his physical appearance. Knowing how weight affects recognition accuracy may provide information that could be exploited to improve the accuracy of recognition systems.

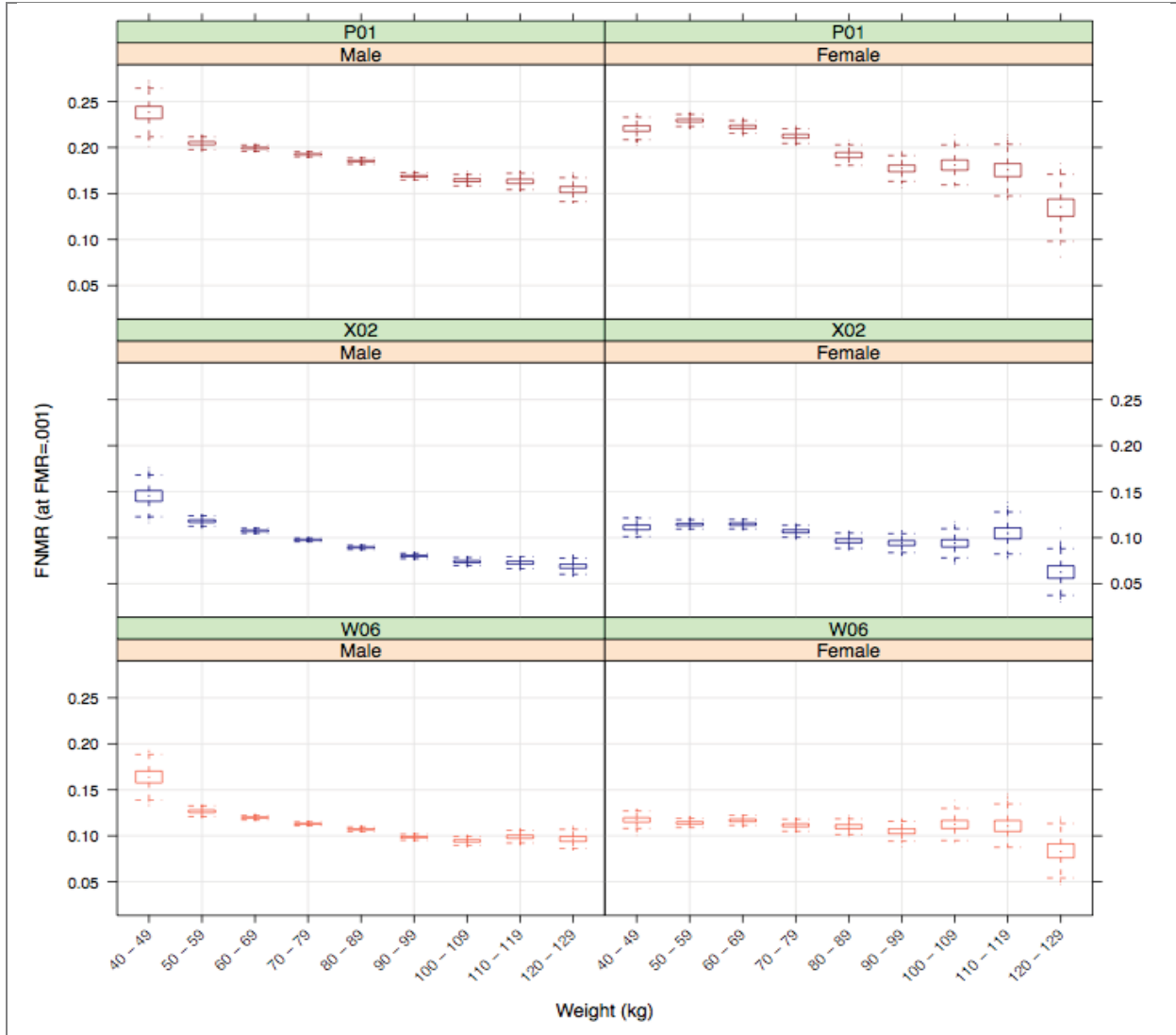
Prior work: The author is not aware of any.

Experimental method: MBE-STILL binned 590,105 genuine comparisons from the FBI set into 10 kg weight increments. For each algorithm, and within each weight increment, a distribution of FNMRs (at FMR=.001) was computed using 2000 bootstrap iterations. Comparisons were further separated into male and female groups since females tend to weigh less, which could introduce a bias if recognition accuracy differs for the different sexes. The resulting plot shows how the false non-match rate changes across weight increments. Weight was assigned to genuine comparisons based on the average of the weights reported for the two captures. The false match rate was computed using results from study 3.

Results: Figure 27 shows, for class A verification SDKs, the dependence of FNMR on subject weight, broken out by subject sex. The threshold is set to produce FMR = 0.001. In all cases, heavier set individuals appear easier to recognize, as most of the figures display a downward trend from left to right. It is possible that a higher amount of body fat introduces additional distinctive features in the face (e.g. folds under the chin). The trend may also be the result of one or more hidden factors, although separating comparisons into male and female sets precludes sex as such a factor. A follow-up investigation (not shown) revealed that the change in weight between captures of an individual had only a very small effect on recognition accuracy, much less than the individual’s average weight.

MBE-STILL REPORT PARTICIPANT KEY	P = PITTPATT	R = SURREY U.	S = TSINGHUA U.	T = TOSHIBA	U = DALIAN U.	PAGE 51 OF 58
	V = NEC	W = L1 IDENTITY	X = COGNITEC	Y = SAGEM	Z = NEUROTECHNOLOGY	

Figure 26 – LEO Verification accuracy by subject weight



Conclusion: While an accuracy trend is clearly evident for both males and females, the result is largely unimportant operationally because weight is not usually a controllable factor, and because the highest error rates are associated with relatively rare individuals with weight below 50 kilograms. The current analysis does not reveal whether this result is confounded with the presence of minors in the dataset. Further statistical analysis is warranted.

INVESTIGATION 19. Link Between race and accuracy

How does race affect the ease of recognition?

Demand driver: Face recognition algorithms should not be too biased in how they treat individuals having certain demographic traits. In addition, investigating a possible race effect can potentially identify aspects of automated face recognition that could be improved.

Prior work: The link between race and automated face recognition has been analyzed in several prior studies [QUINN, FRVT 2002].

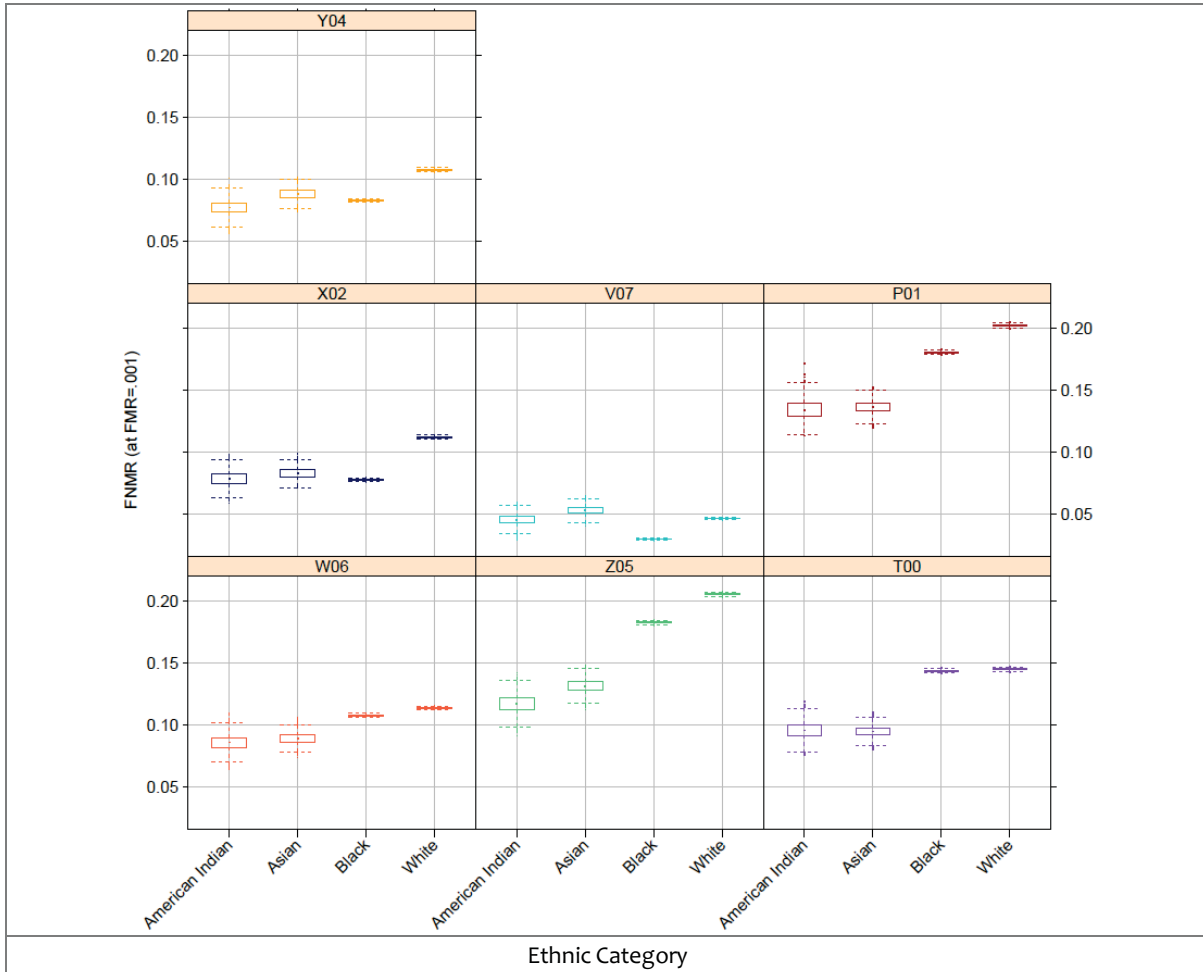
Experimental method: 590,105 genuine comparisons from the FBI set were separated by race. For each algorithm, and for each race, a distribution of FNMRs (at FMR=0.001) was computed using 2000 bootstrap iterations. The resulting plot shows how the false non-match rate differs for different races. A given genuine comparison was only

retained if the recorded race was consistent across all image captures for the given individual. The false match rate was computed using results from study 3.

Results: Figure 28 shows, for class A verification SDKs, the dependence of FNMR on subject weight, broken out by subject ethnicity code. The threshold is set to produce FMR = 0.001.

Conclusions: A race effect clearly exists for each of the algorithms. Blacks are easier to recognize than whites for 5 of the 6 algorithms. American Indians and Asians were clearly easier to recognize for 3 of the algorithms (P01, Z05, and T00), while for V07 American Indians and Asians appeared more difficult to recognize. Disparities in the performance of face recognition algorithms across races has been documented previously [PHILLIPS], and in many cases may simply be due to differing training procedures that are aimed at optimizing performance for an expected demographic.

Figure 27 – LEO Verification accuracy by ethnic category



INVESTIGATION 20. Value of biographic data

If a facial recognition implementation is provided with subject-specific biographic metadata, can accuracy be improved?

Demand driver: In many large-scale identity management applications, biometric data is collected with accompanying metadata such as sex, weight, height or ethnicity. Some of these pieces of information meet certain of the qualifications for being biometric data in their own right but are, by themselves, obviously of limited value for identification.

Such data is often entered by a human operator and is subject to error. This can arise due to clerical and typographic errors, and systemic effects (e.g. non-compliance to the ISO 8601 standard for dates). Unreliable data can undermine

identity management. Indeed biometrics is often advanced as an answer to clerical errors. An additional operational concern is that in some applications such data can be clearly incorrect or spoofed. That said, identity management applications such as PIV routinely protect the integrity of biographical information by computing the digital signature over biometric records (i.e. data + header, for example CBEFF [PIV]). In any case, the MBE-STILL was initiated to include a study of whether biometric recognition process could be augmented by the use of metadata such as sex and age.

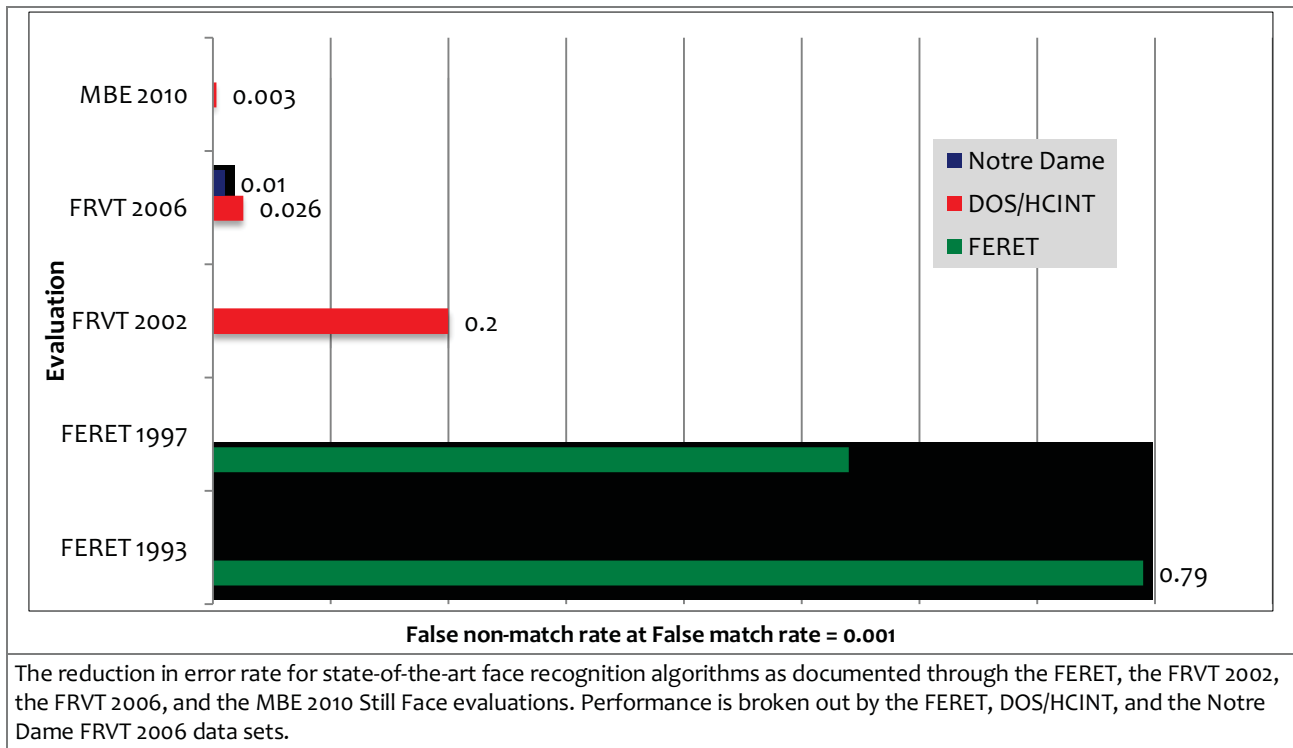
Prior work: This issue has been studied in the academic literature [FIERREZ, ROSS].

Experimental method: The MBE-STILL API supported the provision of the biographical metadata for an image to the SDK¹⁷. The list of variables is: sex, age, height, weight, ethnicity, date of birth, and date of photograph. These are supplied to the face image data-structure that is input to the template generator. The experiment proceeds by running an identical face recognition trial with and without metadata, and comparing the measured accuracy.

NIST only provided the developers with metadata information in late April 2010 just weeks before the closure of the submission window and the scheduled issuance of the first report. The data provided was for the MEDS dataset. This gave algorithm providers a very short period for its analysis and exploitation. In late May 2010, the authors asked whether vendors attempted to use metadata. None made any claim that this data was valuable.

Conclusions: We found no evidence that algorithms exploit date information. This is likely a consequence that algorithm developers have a) never been contracted or challenged to affirmatively incorporate date information, and b) have insufficient data with which to calibrate their reliance on the metadata. This negative result may be valuable to others' considering this issue.

Figure 28 – Progression of face recognition accuracy measurements



¹⁷ These were sex, ethnicity, date of birth, date of capture, height and weight. See Table 9 of the NIST Concept, API and Evaluation plan for the data structures and units.

7. Progress in face recognition

The face recognition community has benefited from a series of U.S. Government funded technology development efforts and evaluation cycles, beginning with the FERET program in September 1993. The evaluations have documented roughly three orders-of-magnitude improvement in performance from the start of the FERET program through the MBE 2010 Still Face.

Figure 29 quantifies this improvement at five key milestones. For each milestone, verification performance is reported. Performance report is the false non-match rate (FNMR) at a false match rate (FMR) of 0.001 (1 in 1000) and is given for a representative state-of-the-art algorithm. The 1993 milestone is a retrospective implementation of Turk and Pentland's eigenface algorithm [TURK], which was partially automatic (it required that eye coordinates be provided). Performance is reported on the eigenface implementation of Moon and Phillips [MOON] with the FERET Sept96 protocol [FERET], in which images of a subject were taken on different days (dup 1 probe set). The 1997 milestone is for the Sept97 FERET evaluation, which was conducted at the conclusion of the FERET program. Performance is quoted on the U. of Southern California's fully automatic submission to the final FERET evaluation [WISKOTT, OKADA]. The 1993 and 1997 results are on the same test dataset and show improvement in algorithm technology under the FERET program. Technology improved from partially automatic to fully automatic algorithms, while error rate declined by approximately a third.

The 2002 benchmark is from the FRVT 2002 [FRVT2002]. In the FRVT 2002 verification performance was reported for the Cognitec, Eyematic, and Identix submissions on the DOS/HCINT dataset. Because both the FERET and DOS/HCINT datasets are low-resolution and have similar performance on the baseline algorithm (see Table V in Phillips et. al [FRVT2006]), one can make the case that they are comparable and a significant portion of the decrease error rate was due to algorithm improvement.

The 2006 benchmark is from the FRVT 2006 [FRVT2006]. In Figure 29, performance is reported for both the Notre Dame high-resolution controlled-illumination still images and the DOS/HCINT dataset. The Notre Dame data set was collected under laboratory conditions. On the Notre Dame data set, the submission from Neven Vision achieved a FNMR of 0.008 at a FMR of 0.001. On the DOS/HCINT data set, Toshiba achieved a FNMR of 0.026 at a FMR of 0.001.

The 2010 benchmark is from the DOS/HCINT data set in Investigation 7. In Investigation 7, a FNMR of 0.003 at a FMR of 0.001 was achieved for the NEC submission. This performance shows a decrease in the FNMR at a FMR = 0.001 from 0.79 in 1993 to 0.003 in 2010. The 1993 benchmark is for a partial automatic algorithm on the FERET data set, which was a laboratory-collected data set. The 2010 benchmark was on an operational data set. The decrease the error rate is roughly three orders-of-magnitude while moving from performance of a partially automatic algorithm on laboratory data set to a fully automatic commercial system on operational data.

MBE-STILL REPORT PARTICIPANT KEY	P = PITTPATT	R = SURREY U.	S = TSINGHUA U.	T = TOSHIBA	U = DALIAN U.	PAGE 55 OF 58
	V = NEC	W = L1 IDENTITY	X = COGNITEC	Y = SAGEM	Z = NEUROTECHNOLOGY	

8. References

8.1. Publications and Reports

ADLER	A. Adler, <i>Images can be regenerated from quantized biometric match score data</i> , in Canadian Conference on Electrical and Computer Engineering, pages 469–472, May 2004. A. Adler, <i>Vulnerabilities in biometric encryption systems</i> , in International Conference on Audio and Video based Biometric Person Authentication, pages 1100–1109, July 2005.
BLUMENSTEIN	<i>Random parameter stochastic process models of criminal careers</i> . In Blumstein, Cohen, Roth & Visher (Eds.), <i>Criminal Careers and Career Criminals</i> , Washington, D.C.: National Academy of Sciences Press, 1986.
BOLLE	Ruud Bolle, Jonathon Connell, Sharanthchandra Pankanti, Nalini Ratha, Andrew Senior, <i>Guide to Biometrics</i> Springer, November, 2003.
CHUTORIAN	E. Murphy-Chutorian and M. Trivedi, <i>Head Pose Estimation in Computer Vision: A Survey</i> in IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), April 2009, vol. 31, no. 4.
DODDINGTON	G. Doddington, W. Liggett, A. Martin, M. Przybocki, and D. Reynolds, <i>Sheep, Goats, Lambs and Woves: An Analysis of Individual Differences in Speaker Recognition Performance</i> , in the International Conference on Spoken Language Processing (ICSLP), Sydney, 1998.
FARID	H. Farid, <i>Exposing Digital Forgeries from JPEG Ghosts</i> , in IEEE Transactions on Information Forensics and Security, Vol. 1, No 4. Pp 154-160, 2009. See also H. Farid, <i>A Survey of Image Forgery Detection</i> , IEEE Signal Processing Magazine, No. 26, Vol. 2. 2009.
FERET	P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, <i>The FERET evaluation methodology for face-recognition algorithms</i> , IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 22, 1090-1104, 2000.
FIERREZ	J. Fierrez-Aguilar et al. <i>Exploiting general knowledge in user-dependent fusion strategies for multimodal biometric verification</i> . In Proc IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2004.
FONDEUR	Jean-Christophe Fondeur, <i>Biometric Testing and Performance Extrapolation</i> , in Proc. International Biometric Performance Conference (IBPC), March 4, 2010. Linked from: http://biometrics.nist.gov/ibpc2010/presentations.html
FRVT 2002	P. Jonathon Phillips, Patrick Grother, Ross J. Micheals, Duane M. Blackburn, Elham Tabassi, Mike Bone, <i>Face Recognition Vendor Test 2002: Evaluation Report</i> , NIST Interagency Report 6965.
FRVT 2004	Patrick Grother and George W. Quinn, <i>Unpublished study for DHS</i> . May 2004.
FRVT 2002b	Patrick Grother, <i>Face Recognition Vendor Test 2002: Supplemental Report</i> , NIST Interagency Report 7083,
FRVT 2006	P. J. Phillips, W. T. Scruggs, A. J. O’Toole, P. J. Flynn, K. W. Bowyer, C. L. Schott, M. Sharpe, “FRVT 2006 and ICE 2006 Large Scale Results,” IEEE Trans. Pattern Analysis and Machine Intelligence, Vol 32, pp 831—846, 2010. P. Jonathon Phillips, W. Todd Scruggs, Alice J. O’Toole, Patrick J. Flynn, Kevin W. Bowyer, Cathy L. Schott, and Matthew Sharpe. <i>FRVT 2006 and ICE 2006 Large-Scale Results</i> . NISTIR 7408, March 2007.
GALBALLY	J. Galbally, C. McCool, J. Fierrez, S. Marcel, and J. Ortega-Garcia, <i>On the vulnerability of face verification systems to hill-climbing attacks</i> , Pattern Recognition, Volume 43, Issue 3, March 2010, pp. 1027-1038.
GROSS	Ralph Gross, Simon Baker, Iain Matthews, and Takeo Kanade, <i>Face Recognition Across Pose and Illumination</i> , Chap 9 in Handbook of Face Recognition, Li et al eds. Springer, 2005.
GROTHER	Patrick Grother, P. Jonathon Phillips, <i>Models of Large Population Performance</i> , Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 68-75, Vol 2. June 2004.
GREATHOUSE	DHS Latent Print Activities, D. Greathouse, in Proceedings of the Latent Fingerprint Testing Workshop, NIST, March 19-20, 2009. http://fingerprint.nist.gov/latent/workshop09/presentations.htm
HUBE	Jens Peter Hube, <i>Using Biometric Verification to Estimate Identification Performance</i> , In Proc. Biometrics Symposium 2006 (BSYM), http://www.biometrics.org/bc2006/program.htm

IREX	Patrick Grother, Elham Tabassi, George W. Quinn and Wayne Salamon, <i>Performance and Interoperability of Iris Images Performance of Iris Recognition Algorithms on Standard Images</i> . NIST Interagency Report 7629, October 30, 2009. Linked from http://iris.nist.gov/irex
JAROSZ	Herve Jarosz and Jean-Christophe Fondeur, <i>Large Scale Identification System Design</i> , Chap 9 in <i>Biometric Systems</i> Wayman et al. eds. Springer 2005.
MARTIN	Brian Martin, <i>Biometric Identification: Metrics & Models</i> , in Proc. International Biometric Performance Conference (IBPC), March 4, 2010. Linked from: http://biometrics.nist.gov/ibpc2010/presentations.html
MEAGHER	Stephen B. Meagher, <i>Defining AFIS Latent Print “Lights-Out”</i> , in Proc. Eval. of Latent Fingerprint Technologies Workshop, NIST, March 19-20, 2009. http://fingerprint.nist.gov/latent/workshop09/presentations.htm
MIN	J Min, K W Bowyer, P Flynn, <i>Using multiple gallery and probe images per person to improve performance of face recognition</i> , Notre Dame Computer Science and Engineering Technical Report (2003).
MINEX	P. Grother et al., <i>Performance and Interoperability of the INCITS 378 Template</i> , NIST IR 7296 http://fingerprint.nist.gov/minexo4/minex_report.pdf
NUPPENY	EasyPASS – Evaluation of face recognition performance in an operational automated border control system, in Proc. International Biometric Performance Conference (IBPC), March 2, 2010. Linked from http://biometrics.nist.gov/ibpc2010/presentations.html
MOC	P. Grother and W. Salamon, <i>MINEX II - An Assessment of ISO/IEC 7816 Card-Based Match-on-Card Capabilities</i> http://fingerprint.nist.gov/minex/minexII/NIST_MOC_ISO_CC_interop_test_plan_1102.pdf
NANDAKUMAR	Karthik Nandakumar, Arun Ross and Anil K. Jain, <i>Biometric Fusion: Does Modeling Correlation Really Matter?</i> In Proc. Third IEEE International Conference on Biometrics: Theory, Applications and Systems – BTAS 2009.
OKADA	K. Okada, J. Steffens, T. Maurer, H. Hong, E. Elagin, H. Neven, and C. von der Malsburg, <i>The Bochum/USC face recognition system</i> , in <i>Face Recognition: From Theory to Applications</i> , H. Wechsler, P. J. Phillips, V. Bruce, F. Fogelman Soulie, and T. S. Huang, Eds. Berlin: Springer-Verlag, 1998, pp. 186-205.
PARK	Unsang Park, Yiyong Tong, Anil K. Jain, <i>Age-Invariant Face Recognition</i> , IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 947-954, May, 2010
PHILLIPS	J. Phillips, F. Jiang, A. Narvekar, J. Ayyad, and A. O’Toole, <i>An Other Race Effect for Face Recognition Algorithms</i> . NIST IR 7666. http://face.nist.gov/NISTIR-7666_Algorithm_Other_Race_Effect.pdf .
PINELLAS	Jim Main and Scott McCallum, <i>Face Recognition; The Pinellas County Sherriff’s Office Experience</i> , Proc. Biometrics Consortium, 2003. http://biometrics.org/bc2003/program.htm Case Study, Pinellas Country Sherriff’s Office (PCSO) Improves Law Enforcement, L1 ID. http://www.l1id.com/files/488-SCD_Pinellas_Case_study_FINAL.pdf
QUINN	G. W. Quinn, P. Grother, <i>False Matches and Non-independence of Face Recognition Scores</i> , in the 2 nd international Conference on Biometrics: Theory, Applications and Systems (BTAS) 2008.
RAMANATHAN	Narayanan Ramanathan and Rama Chellappa, <i>Modeling Age Progression in Young Faces</i> , Proc. IEEE Conference on Computer Vision and Patten Recognition (CVPR), pp. 387-394, Vol 1. June 2006.
ROSS	Arun Ross and Norman Poh, <i>Multibiometric Systems: Overview, Case Studies and Open Issues</i> , in <i>Handbook of Remote Biometricsfor Surveillance and Security</i> , Tistarelli, Li, Chellappa (Eds.), Springer, 2009.
SHAKNAROVICH	Gregory Shakhnarovich, John W. Fisher, and Trevor Darrell, <i>Face recognition from long-term observations</i> , Proceedings of the 7th European Conference on Computer Vision (ECCV), Part III, pp. 851 - 868, 2002. AI Laboratory, MIT.
SHERRAH	Jamie Sherrah, <i>False Alarm Rate: a Critical Performance Measure for Face Recognition</i> in Proc. IEEE Conf on Automatic Face and Gesture Recognition, (FG’04), 2004. Seoul, Korea
TURK	M. Turk and A. Pentland, <i>Eigenfaces for recognition</i> , J. Cognitive Neuroscience, Vol. 3, No. 1, pp. 71-86, 1991
WAGGETT	Peter Waggett, Henry Bloomfield, Bill Perry John Marc Gibbon Jeremy Monroe, Jean-Christophe Fondeur, <i>Reducing Risk Through Large Scale Testing</i> . In Proc. International Biometric Performance Conference (IBPC), March 2, 2010. Linked from http://biometrics.nist.gov/ibpc2010/presentations.html

WEIN	Wein, L. and Baveja, M., "Using fingerprint image quality to improve the identification performance of the U.S. Visitor and Immigrant Status Indicator Technology Program," Proc. National Acad. Sci., v102, pp. 7772-7775, 2005.
WISKOTT	L. Wiskott, J.-M. Fellous, N. Kruger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.-17, pp. 775-779, 1997.

8.2. Standards

AN27	NIST Special Publication 500-271: American National Standard for Information Systems — <i>Data Format for the Interchange of Fingerprint, Facial, & Other Biometric Information – Part 1.</i> (ANSI/NIST ITL 1-2007). Approved April 20, 2007.
ISO STD05	<p>ISO/IEC 19794-5:2005 — <i>Information technology — Biometric data interchange formats — Part 5: Face image data.</i> The standard was published in 2005, and can be purchased from ANSI at http://webstore.ansi.org/</p> <p>Multipart standard of "Biometric data interchange formats". This standard was published in 2005. It was amended twice to include guidance to photographers, and then to include 3D information. Two corrigenda were published. All these changes and new material is currently being incorporated in revision of the standard. Publication is likely in early 2011. The documentary history is as follows.</p> <ol style="list-style-type: none"> 1. ISO/IEC 19794-5: Information technology — Biometric data interchange formats — Part 5: Face image data. First edition: 2005-06-15. 2. International Standard ISO/IEC 19794-5:2005 Technical Corrigendum 1: Published 2008-07-01 3. International Standard ISO/IEC 19794-5:2005 Technical Corrigendum 2: Published 2008-07-01 4. Information technology — Biometric data interchange formats — Part 5: Face image data AMENDMENT 1: Conditions for taking photographs for face image data. Published 2007-12-15 5. Information technology — Biometric data interchange formats — Part 5: Face image data AMENDMENT 2: Three dimensional image data. 6. JTC 1/SC37/N3303. FCD text of the second edition. Contact pgrother AT nist DOT gov for more information.
PERFSTD INTEROP	ISO/IEC 19795-4 — <i>Biometric Performance Testing and Reporting — Part 4: Interoperability Performance Testing.</i> Posted as document 37N2370 . The standard was published in 2007. It can be purchased from ANSI at http://webstore.ansi.org/ .

**Supplemental Information in Support of
the NSTC Policy for Enabling the
Development, Adoption and Use of
Biometric Standards**

August 10, 2009

**NSTC Subcommittee on Biometrics and
Identity Management**

Table of Contents

1	INTRODUCTION.....	5
1.1	OVERVIEW	5
1.2	ABOUT THIS REPORT	6
2	SUPPLEMENTAL INFORMATION	6
2.1	CONFORMITY ASSESSMENT.....	6
2.2	USG MODEL CRITERIA FOR THE ADOPTION/MAINTENANCE OF BIOMETRIC STANDARDS.....	7
2.3	USG PARTICIPATION IN BIOMETRIC STANDARDS DEVELOPMENT.....	10
2.4	APPLICATION OF BIOMETRIC STANDARDS IN PROCUREMENT ACTIONS	11
2.5	EXCHANGE OF PROPRIETARY DATA FORMATS.....	12
2.6	ACCESS TO COPYRIGHTED BIOMETRIC STANDARDS FOR USG-WIDE USE.....	14
2.7	BACKWARDS COMPATIBILITY OF STANDARDS	14
2.8	LIFECYCLE HANDLING OF BIOMETRIC SAMPLES.....	14
2.9	COLLECTION AND USE OF METADATA TO ACCOMPANY BIOMETRIC DATA	15
2.10	FUTURE USG-WIDE REQUIREMENTS FOR BIOMETRIC TECHNOLOGIES.....	17
	BIBLIOGRAPHY	19
	ANNEX A – HISTORY.....	23
	<i>A.1 Fingerprint and Palm Image Standard.....</i>	<i>23</i>
	<i>Analysis of Issue.....</i>	<i>23</i>
	<i>Potential Solutions</i>	<i>24</i>
	<i>A.2 Fingerprint Minutiae Standard.....</i>	<i>25</i>
	<i>Issue</i>	<i>25</i>
	<i>Analysis of Issue.....</i>	<i>25</i>
	<i>Potential Solutions</i>	<i>25</i>
	<i>A.3 Latent Fingerprint Standard.....</i>	<i>26</i>
	<i>Issue</i>	<i>26</i>
	<i>Analysis of Issue.....</i>	<i>27</i>
	<i>Potential Solutions</i>	<i>27</i>
	<i>A.4 Face Image Standard (2D).....</i>	<i>27</i>
	<i>Issue</i>	<i>27</i>
	<i>Analysis of Issue.....</i>	<i>28</i>
	<i>Potential Solutions</i>	<i>28</i>
	<i>A.5 Iris Image Standard.....</i>	<i>29</i>
	<i>Issue</i>	<i>29</i>
	<i>Analysis of Issue.....</i>	<i>30</i>
	<i>Potential Solutions</i>	<i>30</i>
	<i>A.6 Voice Standard</i>	<i>31</i>
	<i>Issue</i>	<i>31</i>
	<i>Analysis of Issue.....</i>	<i>31</i>
	<i>Potential Solutions</i>	<i>32</i>
	<i>A.7 DNA Data Standard</i>	<i>33</i>
	<i>Issue</i>	<i>33</i>
	<i>Analysis of Issue.....</i>	<i>33</i>
	<i>Potential Solutions</i>	<i>34</i>
	<i>A.8 Multi-biometric Fusion.....</i>	<i>34</i>
	<i>Issue</i>	<i>34</i>
	<i>Analysis of Issue.....</i>	<i>34</i>
	<i>Potential Solutions</i>	<i>35</i>
	<i>A.9 Application Profiles.....</i>	<i>35</i>
	<i>Issue</i>	<i>35</i>
	<i>Analysis of Issue.....</i>	<i>36</i>

<i>Potential Solutions</i>	36
<i>A.10 Large Scale Identification Applications</i>	37
<i>Issue</i>	37
<i>Analysis of Issue</i>	37
<i>Potential Solutions</i>	38
<i>A.11 Smart Cards Applications</i>	39
<i>Issue</i>	39
<i>Analysis of Issue</i>	39
<i>Potential Solutions</i>	39
<i>A.12 Mobile and Portable Biometric Devices</i>	40
<i>Issue</i>	40
<i>Analysis of Issue</i>	40
<i>Potential Solutions</i>	40
<i>A.13 Conformance Testing</i>	41
<i>Issue</i>	41
<i>Analysis of Issue</i>	41
<i>Potential Solutions</i>	42
<i>A.14 Performance Testing</i>	43
<i>Issue</i>	43
<i>Analysis of Issue</i>	43
<i>Potential Solutions</i>	43
<i>A.15 Interoperability Testing</i>	44
<i>Issue</i>	44
<i>Analysis of Issue</i>	44
<i>Potential Solutions</i>	45
<i>A.16 Security Testing</i>	46
<i>Issue</i>	46
<i>Analysis of Issue</i>	46
<i>Potential Solutions</i>	46
<i>A.17 Establishment of USG QPL Based on Conformance, Performance, and Interoperability Testing</i>	47
<i>Issue</i>	47
<i>Analysis of Issue</i>	47
<i>Potential Solutions</i>	47
<i>A.18 Reference Implementations and Data Sets</i>	48
<i>Issue</i>	48
<i>Analysis of Issue</i>	48
<i>Potential Solutions</i>	48
<i>A.19 Technical Interface</i>	49
<i>Issue</i>	49
<i>Analysis of Issue</i>	49
<i>Potential Solutions</i>	49
<i>A.20 Standardized Measurements for Biometric Sample Quality</i>	50
<i>Issue</i>	50
<i>Analysis of Issue</i>	51
<i>Potential Solutions</i>	51
<i>A.21 Human Factors (Usability and Accessibility)</i>	52
<i>Issue</i>	52
<i>Analysis of Issue</i>	53
<i>Potential Solutions</i>	54
<i>A.22 Privacy</i>	55
<i>Issue</i>	55
<i>Analysis of Issue</i>	55
<i>Potential Solutions</i>	55
ANNEX B - ACRONYMS	56

1 Introduction

1.1 Overview

In 2005, the NSTC Subcommittee on Biometrics & Identity Management established a standards & conformity assessment working group (SCA WG) to facilitate coordination of USG entities that participated in national and international biometric standards bodies. By 2007, the SCA WG members of the NSTC began working at a more systemic level on topics such as conformity assessment and government-wide adoption of appropriate, approved and published standards.

The collaborative efforts of the SCA WG members resulted in the development of a draft comprehensive policy analysis report, which served as a basis to develop the USG policy document on biometric standards entitled “*NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards*”. This policy was drafted by the NSTC Subcommittee on Biometrics and Identity Management and was approved by the NSTC Committee on Technology in September 2007. It identifies policy issues that impact improving USG mission effectiveness, by delivering standards-based biometric technology.

The NSTC Subcommittee on Biometrics & Identity Management has tasked its standards and conformity assessment working group to maintain the *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards*.

This policy builds on the previous work of the NSTC Subcommittee on Biometrics & Identity Management (e.g., the *National Biometrics Challenge*, dated August 2006) to support biometric data exchange and interoperability across USG agencies, as well as the broader NSTC goal of harmonizing policy and guidance for biometric applications throughout the USG. The policy states that the USG should be guided by the following principles:

- ***Continued development of voluntary consensus standards for biometrics is vital to the security of our Nation and the stability of the US-based biometrics community.*** Agencies should support national and international voluntary biometric standards development activities.
- ***Rigorous testing is required to ensure vendor and system compliance with biometric standards.*** Agencies should support the development of harmonized conformance, interoperability, performance, security, human factors, and operational scenario testing programs in support of procurement actions for biometric products, programs and services.
- ***Standards and conformity assessment processes must be identified and adopted across all agencies to ensure full interoperability.*** Agencies should participate in an interagency process led by the Subcommittee to review available standards and develop consensus recommendations regarding which standards should be adopted across the USG.

- ***The biometric standards and conformity assessment processes recommended by the Subcommittee should be promulgated.*** The Subcommittee shall develop a registry of adopted biometric standards at www.standards.gov/biometrics¹.
- ***The biometric standards and conformity assessment processes recommended by the Subcommittee should be integrated into agency plans whenever feasible.*** Agencies should strive to build and operate biometric systems that are based on the Subcommittee's recommended standards.
- ***Timely adoption and use of appropriate standards is critical to achieving biometrics goals.*** Following selection of recommended standards, the Subcommittee should work to advance adoption of standards for use in Federal biometrics programs and services.

1.2 About this Report

The initial draft comprehensive policy analysis report developed by SCA WG members by June 2007 served as a basis for:

- *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards;*
- *Registry of USG Recommended Biometric Standards;*
- *Supplemental Information in Support of the NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards* (this document);
- *Catalog of USG Biometric Product Testing Programs* [DRAFT].

These documents are developed and maintained by the NSTC Subcommittee on Biometrics and Identity Management and the Subcommittee's Standards Conformity Assessment Working Group. The latest approved versions of these documents are available on the Federal government's web site for biometric activities at: www.biometrics.gov/standards/.

2 Supplemental Information

To assist Federal agencies support biometric system interoperability, this section provides supplemental standards and testing related information in support of the *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standard* and the *Registry of USG Recommended Biometric Standards*.

2.1 Conformity Assessment

Conformity assessment² of products or equipment to a given set of standards and/or operational requirements enhances the user's confidence that the product will perform in

¹ This information is also available on the Federal government's web site for biometric activities at www.biometrics.gov/standards.

² Conformity assessment is defined in ISO/IEC 17025:2004 as: "demonstration that specified requirements (3.1) relating to a product (3.3), process, system, person or body are fulfilled."

accordance to the given set of standards and operational requirements. The specification of operational and performance requirements should express the users' expectations of the equipment and products' performance when used in realistic applications. These requirements must include technical operational characteristics that can be effectively tested and evaluated. Conformity assessment can be performed by testing laboratories that may or may not be accredited. Accreditation of laboratories that perform the tests and evaluations of products and equipment increases confidence that test results are developed with competence and integrity.

Currently there are several USG chartered programs for biometric product testing and certification. These programs are as follows:

- GSA's FIPS 201 Evaluation Program for credential and identity management
- FBI's fingerprint scanner certification
- TSA airport access control performance certification
- TSA TWIC product certification (under development)
- DOD biometrics certification program
- NIST NVLAP program (under development)

For further information on the above programs, refer to the *Catalog of Biometric Product Testing Programs*.

2.2 USG Model Criteria for the Adoption/Maintenance of Biometric Standards

The principle driving force for most USG systems is to improve mission effectiveness by delivering the technology required to support specific applications. Over the course of the last two decades and in accordance with US law and policy, many USG agencies (e.g., DHS, DoD, DoJ, NIST) have promulgated policies and procedures for the adoption of Information Technology (IT) standards, for intra-agency or inter-agency use, in order to facilitate interoperability across applications and systems. In support of standards-based USG biometric systems, the following model criteria for the adoption and maintenance of biometric standards for USG use have been developed.

In 2006, the NSTC SC on Biometrics, Working Group on Standards and Conformity Assessment developed an Interagency Coordination Plan, which included model criteria for the adoption of biometric standards. These criteria were based upon two main factors: the maturity of the standards as evaluated by the USG and the USG business need driving adoption. In terms of maturity, it was recommended that the USG categorize biometric standards and develop three categories: Emerging, Stable, and Mature. Building upon that work, the following criteria for categorizing each biometric standard and guidelines for adoption of a biometric standard are:

Criteria for Emerging Standards (E - Emerging)

- **Availability** – The standard is published and publicly available
-

- **Authoritative** – The standard was developed and is maintained by a recognized Standards-developing organization (SDO), such as INCITS M1, JTC 1 SC37, or NIST, through a process open to participation by the USG.

Criteria for Stable Standards (S - Stable)

- Includes criteria for Emerging standards in addition to the following:
- **Technical Maturity** – The standard is stable and its technical content is mature. No major revisions or amendments are in progress that will affect backward compatibility with the approved standard. If a revision or amendment is in progress that will have a great impact on compatibility with the approved standard, then the standard should be categorized as an emerging standard.
- **Commercial availability** – Several products from different vendors exist on the market to implement this standard.

Criteria for Mature Standards (M - Mature)

Includes criteria for stable standards in addition to the following:

- **Implementability** – Several commercial or government organizations have developed implementations of this standard.
- **Conformance Testing Tool & Certification** – Conformance testing captures the technical description of a specification and measures whether an implementation faithfully implements the specification. A conformance testing methodology and a tool implementing this methodology, and/or conformance testing program that allows preparation of a certified or otherwise approved validated/qualified product list is available.
- **Interoperability Testing** – Interoperability testing tests one implementation (e.g., device, subsystem, system) with another to establish that they can work together properly. A testing methodology and reference implementations or interoperability testing programs are available.
- **Performance Testing** – Performance testing measures one or more characteristics of an implementation under test (e.g., device, subsystem, system) such as its accuracy, human factors, quality, responsiveness, robustness, speed, throughput, etc., under various conditions. Technology, scenario, and operational performance test results based on recognized testing methodologies are available that provide confidence in sufficient performance to meet the requirements of a recognition application.

The NSTC Subcommittee on Biometrics and Identity Management should establish definitions for emerging, mature, and stable biometric standards and, based upon those definitions, establish model criteria for the adoption and maintenance of biometric standards for USG use. The model criteria for USG agencies to mandate and adopt biometric standards should include the following:

The Registry should adopt standards that may be categorized as either stable or mature;

The Registry should not adopt emerging standards the content of which is not stable or for which there is no product that implements it;

The Registry should include migration strategy concerning the adoption and use of new standards. This strategy should provide guidance for agencies to replace existing standards to mitigate the risk of lack of interoperability. This strategy should provide guidance for the adoption and use of new standards that may replace existing standards to mitigate the risk of possible loss of backward compatibility and/or interoperability. The following questions are examples of the questions that should be addressed in the analysis:

- Is the national standard a subset of the international standard?
- Is compatibility required by implementations of the standard?
- Can implementations conform to both the national and international standards?
- Is there an installed/implemented base using the national or international standard?
- Is the national standard already supporting interagency requirements for interoperability?
- Is the international standard sufficient for international (e.g., NATO Interpol) requirements?
- Are there approved national or international biometric profiles (implementation agreements) available?
- Are there sound conformance test methodologies and tools for the national or international standard?
- Are there conformity assessment programs with validated product lists for the national or international standard?

For new applications implementing biometric standards

- Case 1: Stable or Mature ANSI and Emerging ISO standards exist.
- If there is need to migrate to the ISO version in the future, and then perform comparative analysis and future migration plan.
- Based on the complexity of the future migration plan, decide whether to implement the ANSI standards now or work with industry and SDOs to accelerate the maturity of the international standard and implement the international standards.
- Case 2: Stable or Mature ANSI and ISO standards exist
- Absent technical issues, preference is given for implementation and adoption of the ISO standard.

For existing applications implementing biometric standards

- Case 1: Stable or mature ANSI standard exists, and there are no equivalent international standards.
- Continue implementation of ANSI standards.
- Consider sponsoring the development of an international standard while maintaining backward compatibility with the ANSI standard to protect previous investment.
- Case 2: An international standard becomes Stable or Mature, while an already implemented ANSI or government standard exists
- Determine the business need for migration to the international standard.
- If necessary, develop a future migration plan.
- Develop implementation guidelines for each of the approved standards that will assist the USG in its adoption and implementation of the biometric standards for various applications.
- Perform analysis of the relationship between standards and select the appropriate ones for specific applications based on business models or business cases. Select business cases. Then develop appropriate use scenarios for some of the choices available and discuss some emerging items that should be considered for future applications.
- Develop or identify a mechanism to communicate the USG evaluation criteria and adoption guidelines to the vendor community and SDOs to provide clarification concerning USG standards requirements for adoption by biometric systems.

2.3 USG Participation in Biometric Standards Development

In accordance with US law and policy, USG experts are participating in various national and international standards development organizations to ensure the timely development of technically sound biometric standards. The motive for this participation is to improve mission effectiveness by delivering standards-based biometric technology in support of specific agency applications.

Ongoing USG participation will be required in the future so that:

- Timely, technically sound biometric standards continue to be developed and maintained;
- USG has sufficient technical knowledge about these standards to make savvy adoption decisions; and
- USG can develop a testing infrastructure that supports successful procurements and deployments of standards-based biometric systems.

USG leadership in biometric standardization includes:

- FBI Electronic Fingerprint Transmission Specification (EFTS)/Electronic Biometric Transmission Specification (EBTS) standardization activity;

- DoD EBTS standardization activity;
- National Institute of Standards and Technology Information Technology Laboratory (NIST/ITL) development of standards under its American National Standards Institute (ANSI) accreditation, provide:
 - The Chair of InterNational Committee for Information Technology Standards -Technical Committee INCITS M1
 - The Chair and the Secretariat for ISO/IEC Joint Technical Committee 1-Subcommittee 37 (JTC 1 SC 37)
 - Technical editors for many important biometric standards development projects
- The Departments of State and Homeland Security provide USG representation to the UN International Civil Aviation Organization (ICAO) Technical Advisory Group (TAG) New Technologies Working Group (NTWG) dealing with travel identification and use of biometrics;
- Additionally, USG experts are providing substantive technical contributions for many biometric standards development projects, which are of high priority to the USG. USG coordination of agency positions and contributions to biometric standards development projects is successfully occurring through groups such as the FBI's Criminal Justice Information Systems (CJIS) Advisory Policy Board (APB), the DHS Biometrics Coordination Group (BCG), the DoD Biometric Standards Working Group (BSWG), and the NSTC Subcommittee on Biometrics & Identity Management's Standards & Conformity Assessment Working Group.

The USG should continue to provide administrative and technical leadership for national and international biometric standards development, and should coordinate USG positions and contributions to these standards developers.

2.4 Application of Biometric Standards in Procurement Actions

An important aspect of the adoption of biometric standards is the incorporation of applicable standards into procurement actions. To support the data interchange and interoperability goals for USG use of biometrics, agencies should follow USG guidelines and standards for procurement of biometric devices, hardware and software systems.

In procurement actions standards provide several advantages. The major advantages are:

- In equipment purchases, standards can set specifications that give confidence that products will function as intended;
- Data formats and system interfaces developed to standards support data interchange and USG system interoperability goals; and
- Standards widen the vendor base which leads to increased competition which, in turn, can result in reduced costs.

Unfortunately, it is not always obvious in that a standard is available or applicable to a procurement action. Therefore, many USG agencies have developed processes to identify, vet and adopt standards pertinent to their national security and homeland security needs. Those standards that are adopted will be compiled into a central database that program managers, systems developers, procurement officers and all others performing procurement actions will be able to access. The goal of this effort is to create a one-stop-shopping-center for standards related to national security and homeland security requirements.

Within DoD, the DoD Information Technology Standards Registry serves as a central repository for DoD-approved information technology standards, including biometric standards. The standards selection criteria focus on mandating only those items critical to net-centricity and interoperability. Standards must successfully satisfy the following seven criteria for submission and acceptance into DoD Information Technology Standards Registry (DISR): net-centricity, interoperability, maturity, implementability, public availability, and consistency with authoritative sources. Use of the DISR is mandated for the development and acquisition of new or modified fielded IT and National Security Systems throughout the DoD.

In another example, the DHS has developed a two stage adoption process. The first stage is a technology vetting step. When a document is submitted for consideration as a DHS adopted standard a determination is made by a standards coordinator in the DHS S&T Office of Standards as to the need for a review by technical experts in the pertinent field to determine on a technical level if a document has a sufficiently wide or critical application in the homeland security domain that warrants its adoption.

The second stage involves vetting the document at a policy level. The DHS S&T Office of Standards has formed a DHS Standards Council that works jointly with the DHS Biometrics Coordination Group's Standards Working Group. This is a group of DHS component employees who manage standards issues within their component. As such these representatives either are in a position to make policy decisions on standards matters or have access to those within their component who have that authority. Therefore, they are in a position to have standards vetted within their component.

At the policy level, documents are considered for application to the component's responsibilities, including procurement requirements, as well as whether or not they will encumber the activities of the component. Documents that are deemed acceptable at the policy level are then registered into the central database and publicized by the DHS S&T Office of Standards.

Agencies should develop internal procedures to ensure citation of relevant standards from the *Registry of USG Recommended Biometric Standards* (Registry) in all biometric procurement actions.

2.5 Exchange of Proprietary Data Formats

The issue is whether USG applications should allow standardized records to also include additional proprietary data. The hazard is that within one organization or deployment, a single supplier may use entirely proprietary data for matching, and have partial support

for data that may sometime arrive. For example, an employee of one government department visits another and presents an identity credential containing standardized minutia records to a system that is incapable of processing it.

The vast majority of biometric systems currently in use embed proprietary template data. They are either not interoperable at all, or achieve interoperability only at the input image or signal level. For example most current biometric laptop logon systems are purely proprietary. Alternatively while the FBI's IAFIS system uses a proprietary fingerprint template (minutiae plus other commercially-protected feature data) for matching, it achieves interoperability with the outside world (i.e. state and local law enforcement) only via standardized image formats, primarily ANSI-NIST image records.

However, while image based interoperability is common, there are some standardized biometric templates in existence. Some of these include fields for proprietary data. The format of such data is usually unpublished, is known only to the company that provided it, and could even be strongly encrypted. By definition then, such content is not interoperable i.e. it cannot be used by a system unless that system includes the (proprietary) components to handle it.

Some standards exist that address the issue of exchange of data in proprietary formats. They are stable, but some have revisions underway to correct minor errors:

- ISO/IEC 19794-2:2005
- INCITS 378-2004
- ANSI/NIST ITL 1-2007

With INCITS 378-2004 and ISO/IEC 19794-2:2005 the standard fingerprint minutiae data may be accompanied by either standardized ridge count, core and delta information or by fully proprietary data.

An ANSI/NIST ITL 1-2007 minutiae record can contain standardized minutiae data, very similar to INCITS 378-2004 minutiae data, or full proprietary minutiae data from one of six large commercial fingerprint concerns. The presence of standardized data is not required by an ANSI/NIST ITL 1-2007 record itself.

The technical differences between these standards for core minutia data the differences are syntactic. An ANSI/NIST ITL 1-2007 record can encapsulate purely proprietary data. The other standards can serve to add proprietary data to standardized data.

All USG biometric systems should employ standards to achieve interoperability and avoid proprietary formats to the maximum extent possible.

Agencies should use the proprietary data fields in standardized data formats from the registry of USG recommended biometric standards for the exchange of proprietary data.

Agencies with closed systems that do not require system or interagency interoperability should only use proprietary data formats if standardized data formats can be documented to be inadequate.

2.6 Access to Copyrighted Biometric Standards for USG-wide Use

USG planning/procurement/use of standards-based biometric applications would be greatly facilitated if USG persons involved in such activities had ready access to electronic copies of all biometric standards, which are being specified for USG biometric data exchange and interoperability. Biometric standards that are not copyrighted, such as USG developed standards, are most often, freely available for downloading from the Web. Also, some standards developing organizations copyright their standards and make them available at no cost. However, other standards developing organizations rely on the sale of their copyrighted standards to support their operation.

USG employees and contractors require access to biometric standards to design, procure, and implement systems that use biometric technologies. Providing access to these standards will allow a larger community within USG to be aware of standards, their applicability, and recommended best practices.

2.7 Backwards Compatibility of Standards

In the context of biometric systems, backwards compatibility can only be achieved by ensuring interoperability of new systems with legacy data, or new data with legacy systems. Adequate control and documentation of both the biometric data and biometric interfaces are necessary conditions for this, and while proprietary data and interfaces do not necessarily preclude migration to newer systems, these will most often be from the same supplier. Thus formal biometric standards offer benefits in two areas. First the ability to migrate to another vendor supports a competitive marketplace of improving products. Second this supports continuity of operations should the supplier have difficulties.

Biometric systems often achieve interoperability at the unprocessed image or signal level, but the actual identification or verification comparisons involve proprietary template data. In most cases, particularly for identification systems, this is a necessary condition because accuracy available from standardized templates (when they exist), lags that of the proprietary solutions. If an application is to successfully migrate from one supplier to another, there will be a need to re-enroll the raw image or signal data. In very large scale operations this will entail a transitional arrangement.

Not all applications migrate to new standards at the same rate. Historical data may be necessary to be used, therefore, systems should be able to use older data and formats, perhaps recognizing that utility may be reduced for legacy data; or current data captured according to previous benchmarks or standards.

While participating in SDO activities, the USG should promote the concept that voluntary consensus standards be backward compatible to the maximum extent possible to ensure interoperability of new systems with legacy data, or new data with legacy systems.

2.8 Lifecycle Handling of Biometric Samples

When a biometric sample is entered into a data set, its usefulness depends upon how it has been handled since the time of capture. The data sample may pass quality check algorithms and have the proper data storage format and data attributes, but not be reflective anymore of the biometric sample collected from the subject. This can be caused by a variety of factors, to include, but not limited to, multiple compressions/restorations of a data record, or scanning of an original image at an unsuitable resolution.

While the circumstances of data collection (particularly for watch list information) may not be controllable, once the data is captured, care should be taken so as not to unnecessarily degrade the data in handling of it. By following procedures recommended for selecting parameters at all stages of data handling and not employing certain means of data handling or transmission, the watch list data will be more suitable to actually identifying persons of interest.

Known or suspected terrorist (KST) and other watchlist data should be of the best possible quality. Mishandling of the data could produce false matches that would not be recognizable as such (for instance by introduction of artifacts into a fingerprint image with JPEG used to compress the image). Systems should be reviewed to ensure that data handling meets the best practices defined as a result of this issue.

2.9 Collection and Use of Metadata to Accompany Biometric Data

USG agencies often have requirements for metadata to facilitate the use and management of biometric data, and the storage and transmission of biometric records containing biometric data. The required metadata may include descriptive elements affecting the processing of biometric data as well as some operational system capabilities. The metadata may include information on the types of pre-processing done on the sample data, data that supports verification of the authenticity of the biometric data itself, source of the data, time stamping as well as data that support protection of the biometric data and the integrity of the biometric record. USG agencies often have requirements to associate the biometric data with user-defined challenge data and/or published or unpublished payload data. USG agencies often have requirements to efficiently determine whether a particular biometric data record is of interest by using attributes of their biometric-specific data without exposing the biometric data itself to applications. The best way to meet these types of requirements is for USG agencies to use appropriate standard biometric data structures defined in INCITS M1/JTC 1 SC 37 biometric interface standards, in instantiations of the ANSI/NIST ITL 1-2007 standard or in data structures that use a combination of the standards above.

Metadata can be categorized as “processing,” “operational,” or “demographic.” These categories are somewhat arbitrary, especially the first two. As discussed below, a metadata element may fall within one category or the other depending on the processor, the system and the application. Processing metadata is defined as the minimum information related to the biometric data that is required in order to process the captured biometric data. Length, width and resolution of an image are considered processing metadata. Operational metadata could be seen as information that is not required for the processing of a specific biometric record but that could be crucial for the effective system

operation. Information related to the origin of the biometric data, the product identifier, or the validity period of the biometric sample may fall within this category. In some instances, the distinction whether specific metadata is “processing” or “operational” is blurred. A data structure that contains biometric data could include metadata indicating the product (and version) of the software that generated the biometric data. Whether these are “operational” or “processing” metadata may depend on the system design and matcher functionality. The matcher may require these metadata to process the biometric data, or the metadata may be used only to pre-select a subset of records in a database. Finally, demographic metadata includes biographic and descriptive metadata pertaining to a subject but is not required to process the biometric data.

Metadata specified in the biometric data interchange standards developed by INCITS M1 and JTC 1/SC 37 contain processing metadata and also some operational metadata such as the product identifier and the equipment ID. Whether these metadata are sufficient to achieve the requirements depends of the applications, system design and expected functionality. Usually, a system requires more operational metadata elements than generally specified in biometric data interchange standards in order to achieve full data interchange and interoperability. The interface standards developed by INCITS M1 and JTC 1/SC 37 contain additional operational metadata. Therefore, in an open systems environment, both biometric data interchange format standards and these biometric interface standards are necessary to achieve full data interchange and interoperability for biometric recognition. In many cases, application profiles for the data interchange format standards and/or the technical interface standards are also necessary (e.g., Electronic Biometric Transmission Specifications).

Many applications may also need to incorporate in the system design, means of selecting biometric data based only on metadata external to these data. An example is instances where the biometric data is encrypted and a pre-selection of the records that contain these data needs to be made based on privacy-irrelevant information at the pre-decryption processing stage.

INCITS M1 and JTC 1/SC 37 have developed technical interface standards (e.g., Common Biometric Exchange Formats Framework (CBEFF) and Biometrics Application Programming Interface (BioAPI)) that specify self-describing Biometric Information Records (BIRs) that reveal the format and other attributes of their biometric-specific data without exposing the biometric data itself to applications. The metadata contained in these BIRs provide a means for applications to efficiently determine whether a particular biometric data record is of interest, and if so, which biometric services to call to process the biometric-specific data. The ANSI/NIST ITL 1-2007 standard specifies records that define biometric data of several modalities.

Agencies should develop agency-specific guidelines for the collection, maintenance, and use of metadata for USG biometric applications. This is a factor in OMB program review.

This policy does not apply to law enforcement applications and other large-scale identification applications that only require conformance to standards such as DoJ/FBI/CJIS, EBTS V8.0, DoD EBTS V1.2 or ANSI/NIST ITL 1-2007 nor does it apply to applications that can achieve full system requirements with metadata contained

within the biometric data records specified in INCITS M1/JTC 1 SC 37 biometric data interchange format standards, self-describing data such as JPG 2000 images and instantiations of ANSI/NIST ITL 1-2007 data structures that contain the required metadata.

All new USG biometric applications that require plug and play capability without losing functionality and required descriptive “processing”, “operational” or “demographic” data that is not contained in standards or biometric data records described in the note above should:

- Require Biometric Information Records (BIRs) conforming to a CBEFF Patron Format (PF) for the processing, exchange, protection, encapsulation, transmission and storage of biometric data. Use existing Patron Formats that permit incorporating in the data structure the required level of additional “processing” and “operational” metadata and data elements that support payload, security/integrity options and creation date/validity period. (Note: Patron Formats specified in INCITS 398-2008, or instantiations of BioAPI BIRs are preferred.). Part 3 of the international version of CBEFF offers other alternatives.
- Require conformance to the CBEFF Patron Formats detailed above for applications that require transmission or storage of BIRs that require clear text biometric headers or making metadata available without processing the record or exposing the biometric data itself to applications (e.g., for the purpose of indexing BIRs).
- Encrypt biometric data within the BIRs and sign BIRs by relying on information in the CBEFF BIR Security Block, unless other system security mechanism are already provided by means external to these biometric data structures.

USG agencies may define data structures that use a combination of the standards above (e.g., CBEFF BIRs containing ANSI/NIST ITL 1-2007 data structures).

USG applications should adhere to the standards detailed in this issue to the maximum extent possible but with the recognition that strict adherence may require agencies to defined their own CBEFF Patron Formats to meet the requirements for metadata not defined in existing Patron Formats. These Patron Formats may be published or unpublished. The goal is to assure interoperability and data interchange using still standardized data structures. A requirement is that the “owner” of the Patron Format be registered with the International Biometric Industry Association who acts as the Registration Authority for CBEFF. The IBIA organization identifier: Hex “FEFE” has been reserved for private use, not uniquely assigned by IBIA. A Patron Format can also be registered.

Note: Embedding these biometric data structures in other encapsulators not defined in the above standards may be needed to meet some system requirements. Their use is application-dependent. These can be published or unpublished data structures.

2.10 Future USG-wide Requirements for Biometric Technologies

The USG consists of many agencies with many different operational environments and business needs. In addition, new requirements may arise over time that will affect the potential use of biometrics by these agencies.

The Registry is focused on biometric technologies that are considered to be high priority for USG-wide use in the near term (i.e., fingerprint, 2D-face, and iris) or may be high priority by 2013 (i.e., voice and DNA modalities). Other biometric technologies (i.e., 3D-face, vascular, hand geometry, signature, etc.) may be included in subsequent revisions of this report.

Voice recognition is an excellent example of an emerging biometric technology that may have potential use in the USG. For example, voice recognition could be used in cases such as a driver of a vehicle on an airport tarmac approaching an airplane. Voice recognition software may be able to determine whether that particular driver has authority to enter a specific restricted zone.

DNA is not traditionally considered a real-time biometric due to the requirements for DNA processing and analysis. However, there is now more acceptance of DNA as a practical biometric tool as the processes for taking DNA samples and the actual 'laboratory' process becomes simplified and less time consuming.

Bibliography

Reports

- Report of the Defense Science Board on Defense Biometrics*, March 2007
<http://www.acq.osd.mil/dsb/reports/2007-03-Biometrics.pdf>
- NSTC Subcommittee on Biometrics, *The National Biometrics Challenge*, August 2006
<http://www.biometrics.gov/NSTC/pubs/biochallengedoc.pdf>
- NSTC Subcommittee on Biometrics, *Privacy and Biometrics*, September 2006
<http://www.biometrics.gov/docs/privacy.pdf>
- NSTC Subcommittee on Biometrics, Introduction to Biometrics Web Page
<http://www.biometrics.gov/ReferenceRoom/Introduction.aspx>
- NSTC Subcommittee on Biometrics, *Biometrics Glossary*, September 2006
<http://www.biometrics.gov/Documents/Glossary.pdf>
- NSTC Subcommittee on Biometrics, *Biometrics Standards*, August 2006
<http://www.biometrics.gov/Documents/BioStandards.pdf>
- NSTC Subcommittee on Biometrics, *Biometrics Testing and Statistics*, August 2006
<http://www.biometrics.gov/docs/biotestingandstats.pdf>
- NIST, *Guidance on Federal Conformity Assessment Activities*, August 2000
<http://ts.nist.gov/Standards/Global/caguidance.cfm>
- NISTIR 6025, *Metrology for Information Technology (IT)* , May 1997
<http://www.itl.nist.gov/lab/nistirs/NISTIR%206025.pdf>
- INCITS M1, *Report to M1 on Issues for Harmonizing Conformity Assessment to Biometric Standards*, March 2005
http://www.incits.org/tc_home/m1htm/docs/m1050067.pdf
- INCITS M1, *Study Report on Biometrics and E-Authentication*, March 2007
http://www.incits.org/tc_home/m1htm/2007docs/m1070185.pdf
- U.S. Army Biometrics Task Force, *U.S. Army BTF Technical Contribution to M1.3 - National and International Iris Image Interchange Format Standards: Comparative Analysis Report*, November 2006
http://www.incits.org/tc_home/m1htm/2006docs/m1060977.pdf
- U.S. Army Biometrics Task Force, *U.S. Army BTF Technical Contribution to M1.3 - National and International Face Recognition Format for Data Interchange Standards: Comparative Analysis Report*, November 2006

http://www.incits.org/tc_home/m1htm/2006docs/m1060976.pdf

Making the Confidence Connection -- Conformity Assessment System Design, Gordon Gillerman, Standards Engineering, the Journal of the Standards Engineering Society, Vol. 56, No. 6, November/December 2004

http://www.astm.org/SNEWS/DECEMBER_2004/gillerman_dec04.html

USG Laws and Policy

NIST National Technology Transfer and Advancement Act (NTTAA) Web Page

<http://ts.nist.gov/Standards/Conformity/nttaa.cfm>

OMB Circular A-119; *Federal Participation in the Development and Use of Voluntary Consensus Standards and in Conformity Assessment Activities*, February 1998

<http://ts.nist.gov/Standards/Conformity/upload/fr-ombal19.pdf>

Freely Available Standards and Guidelines

NIST, *ANSI/NIST-ITL 1-2000 Data Format for the Interchange of Fingerprint, Facial, & Scar Mark & Tattoo (SMT) Information*

ftp://sequoyah.nist.gov/pub/nist_internal_reports/sp500-245-a16.pdf

NIST, American National Standard, *ANSI/NIST-ITL 1-2007 Data Format for the Interchange of Fingerprint, Facial, & Other Biometric Information - Part 1*

<http://fingerprint.nist.gov/standard/>

NIST, American National Standard, *ANSI/NIST-ITL 2-2008 Data Format for the Interchange of Fingerprint, Facial, & Other Biometric Information - Part 2 -XML*

<http://fingerprint.nist.gov/standard/index.html>

NIST FIPS 201-1, *Personal Identity Verification (PIV) of Federal Employees and Contractors*, June 2006

<http://csrc.nist.gov/publications/fips/fips201-1/FIPS-201-1-chng1.pdf>

NIST SP 800-76-1, *Biometric Data Specification for Personal Identity Verification*, January 2007

http://csrc.nist.gov/publications/nistpubs/800-76-1/SP800-76-1_012407.pdf

NISTIR 6529-A, *Common Biometric Exchange Formats Framework*, April 2004

<http://www.itl.nist.gov/div893/biometrics/documents/NISTIR6529A.pdf>

DoJ/FBI/CJIS, *Electronic Fingerprint Transmission Specification (EFTS), Version 7.1*, May 2005

<http://www.fbi.gov/hq/cjisd/iafis/efts71/efts71.pdf>

DoJ/FBI/CJIS, *Electronic Biometric Transmission Specification (EBTS) Version 8.1*,
http://www.fbibiospeccs.org/fbibioimetric/documents/EBTS_v8.1_11-24-08.pdf

DoJ/FBI/CJIS, *IAFIS Wavelet Scalar Quantization (WSQ) Grayscale Fingerprint Image Compression Specification*, December 1997
http://www.fbibiospeccs.org/fbibioimetric/docs/WSQ_Gray-scale_Specification_Version_3.pdf

Image Quality Specifications for ten-print fingerprint capture systems, Appendix F, EBTS
www.fbibiospeccs.org

Personal Identity Verification (PIV) Image Quality Specifications for Single Finger Capture Devices
<http://www.fbi.gov/hq/cjisd/iafis/piv/pivspec.pdf>

JPEG 2000 Profile for 1000ppi Fingerprint Compression
www.mitre.org/work/tech_papers/tech_papers_04/lepley_fingerprint/

JPEG 2000 and WSQ Image Compression Interoperability
www.mitre.org/work/tech_papers/tech_papers_01/lepley_jpeg2000/

Testing Documents

WSQ Fingerprint Image Compression Encoder/Decoder Certification Guidelines
www.itl.nist.gov/iad/894.03/fing/cert_gui.html

Test Procedures for Verifying IAFIS Image Quality Requirements for Fingerprint Scanners and Printers
<http://www.mitre.org/tech/mtf>

Test Procedures for Verifying Image Quality Requirements for Personal Identity Verification (PIV) Single Finger Capture Devices
<http://www.mitre.org/tech/mtf>

Available Test Tools/Reference Data

DoD Conformance Test Suite (CTS) for ANSI INCITS 358-2002 BioAPI Specification (BioAPI 1.1)
<http://www.biometrics.dod.mil/CurrentInitiatives/Standards/TestingToolsets.aspx>

NIST Conformance Test Suite (CTS) for ANSI INCITS 358-2002 BioAPI Specification (BioAPI 1.1)
http://www.itl.nist.gov/div893/biometrics/BioAPI_CTS/index.htm

NIST Minutiae Exchange Interoperability Test for INCITS 378-2004
<http://fingerprint.nist.gov/minex/>

NIST Conformance Testing Architecture and Test Tool for CBEFF Patron Format A
(specified in INCITS 398-2008)

http://www.itl.nist.gov/div893/biometrics/CBEFF_PFA_CTS/index.htm

Certified Product Lists

GSA FIPS 201-1 certification: <http://fips201ep.cio.gov/>

TSA Airport Access Control certification:

http://www.tsa.gov/join/business/biometric_qualification.shtm

FBI certification: <http://www.fbi.gov/hq/cjisd/iafis/cert.htm>

TSA TWIC certification: http://www.tsa.gov/assets/pdf/twic_ice_list.pdf

The information provided in this Annex is a summation of the analyses performed by the SCA WG in the **first part of 2007**. Therefore some of the information below is now out-of-date. These analyses are being provided to give the background on the issues that shaped the subsequent first editions of *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards* and the *Registry of USG Recommended Biometric Standards (Registry)*.”

Annex A – History

2007 Analyses by SCA WG

The information provided in this Annex is a summation of the analyses performed by the SCA WG **in the first part of 2007 and therefore some of the references below are now out-of-date**. These analyses served as a basis for the subsequent first editions of *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards* and the *Registry of USG Recommended Biometric Standards (Registry)*.

A.1 Fingerprint and Palm Image Standard

Issue

USG agencies have ongoing requirements to capture, use, store, and exchange fingerprint and palmprint image biometric data. The best way to meet these requirements is for USG agencies to use the same biometric data interchange format standard for fingerprint images and to use the one voluntary consensus data interchange format standard available for palmprint images. While the standards support data exchange, conformance to them alone is not sufficient to satisfy the USG's high level objective to have the best quality finger images available for watchlists and other applications.

Analysis of Issue

There are three voluntary consensus data interchange format standards for fingerprint images presently available:

- ISO/IEC 19794-4:2005 Fingerprint Image Interchange Format
- ANSI INCITS 381:2004 Fingerprint Image Interchange Format
- ANSI/NIST-ITL 1-2007 Data Format for the Interchange of Fingerprint, Facial, & Other Biometric Information - Type-14 fingerprint image record

All three fingerprint image standards are stable and compatible with one another. The Number of Fingers, the capture resolution, compression ratio and the compliance of sensor is specified by each application. At the time of this writing, revision projects are underway for ISO/IEC 19794-4:2005 and INCITS 381-2004, which should result in improved standards.

There is only one voluntary consensus data interchange format standards for palmprint images presently available:

- ANSI/NIST-ITL 1-2007 Data Format for the Interchange of Fingerprint, Facial, & Other Biometric Information - Type-15 palmprint image record

There are many options within the standards and these should be rigorously addressed in a dedicated profile of the standards for specific application.

The information provided in this Annex is a summation of the analyses performed by the SCA WG in the **first part of 2007**. Therefore some of the information below is now out-of-date. These analyses are being provided to give the background on the issues that shaped the subsequent first editions of *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards* and the *Registry of USG Recommended Biometric Standards (Registry)*.”

Existing or planned USG/other procurements that should result in the deployment of standards-based fingerprint products include the Federal Government Personal Identity Verification (PIV) Program, FBI Next Generation Identification, DHS US-VISIT IDENT, and the DoD Automated Biometric Identification System. The FBI NGI specifies the ANSI/NIST ITL 1-2007 Type-14 fingerprint record and PIV will result in INCITS 381-2004 and ANSI/NIST ITL 1-2007 Type 4 and 14.

Potential Solutions

In all new USG biometric applications in which fingers are imaged for enrollment or registration, the images collected should conform to ANSI/NIST ITL 1-2007 Type-14 fingerprint image record requirements. The resolution should be at least 197 pixels per centimeter.

The Type-14 record, which permits information exchange beyond that of the Type-4 record (e.g., variable resolution images, greater than .8 bits of grayscale), is used for new USG fingerprint applications. The use of the ANSI/NIST ITL Type-4 record is deprecated

In all new USG biometric applications in which the palms of cooperative subjects are imaged for enrollment or registration, the images collected should conform to ANSI/NIST ITL 1-2007 Type-15 palmprint image record requirements.

For all new USG biometric fingerprint and palmprint applications, the image standards should be formally profiled. This should enumerate which of the options are permitted and instantiate minimum and maximum values for variables that the generic base standards do not prescribe. Particularly the profile should establish minimum criteria for the sensor resolution and the sensor area. It should enumerate the allowed compression algorithms and should specify maximum compression ratios.

USG should develop default or candidate profiles for fingerprint image retention and transmission.

USG should develop technical means, including open-source tools, for transcoding fingerprint images between instances of the standards (e.g., fingerprint images conforming to ISO/IEC 19794-4:2005 transcoded to the ANSI/NIST ITL 1-2007 Type-14 fingerprint image record).

USG applications should adhere to the ISO and ANSI standards to the maximum extent possible but with the recognition that strict adherence may not be feasible, advisable, or cost efficient. Therefore, a specific application profile should be developed that deals with the issue of which parts of the standards are not to be adhered to in any particular application. The goal is to assure machine interoperability and accuracy.

USG agencies should reflect this policy in any agency specific adoption processes (e.g., DoD DISR, DHS TRM).

Agency standards registries should harmonize with this policy.

The information provided in this Annex is a summation of the analyses performed by the SCA WG in the **first part of 2007**. Therefore some of the information below is now out-of-date. These analyses are being provided to give the background on the issues that shaped the subsequent first editions of *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards* and the *Registry of USG Recommended Biometric Standards (Registry)*.³

A.2 Fingerprint Minutiae Standard

Issue

USG agencies have ongoing requirements to exchange fingerprint minutiae biometric data. The best way to meet these requirements is for USG agencies to use the same biometric data interchange format standard for fingerprint minutia. Minutiae-based exchange has been demonstrated to be less accurate, but faster, than image-based fingerprint interoperability.³

Analysis of Issue

There are three data interchange format standards for fingerprint minutiae:

- ISO/IEC 19794-2:2005 Finger minutiae data
- INCITS 378-2004 Finger Minutiae Format for Data Interchange
- ANSI/NIST ITL 1-2007 Data Format for the Interchange of Fingerprint, Facial, & Other Biometric Information - Type-9 minutia data record

All three standards are stable. INCITS 378-2004 is being revised to correct minor flaws.

ISO/IEC 19792-4:2005 allows specification of either the record or (smart) card format and requires specification of the format type code to describe the minutia placement specification.

Existing applications allow the use of standardized fingerprint templates. The FBI CJIS, Electronic Biometric Transmission Specification (EBTS) Version 8.0 - requires conformance to the ANSI/NIST ITL 1-2007 Type-9 fingerprint record. NIST Special Publication 800-76-1 requires storage of INCITS 378-2004 fingerprint templates on the PIV credential.

Potential Solutions

All new USG identification applications should only use standardized minutiae records, even if parent images or associated proprietary template data are also available for matching.

All new USG verification applications which specify storage or use of standardized minutia records should use the ISO/IEC 19794-2:2005 formats of type 0001, 0003 or 0005. Such applications should allow inclusion of proprietary data in associated extended data fields.

The use of any of the standardized minutiae records for encoding latent fingerprint minutiae is insufficient, and should only be used as a supplement to the parent latent image.

³ NISTIR 7296 http://fingerprint.nist.gov/minex04/minex_report.pdf

The information provided in this Annex is a summation of the analyses performed by the SCA WG in the **first part of 2007**. Therefore some of the information below is now out-of-date. These analyses are being provided to give the background on the issues that shaped the subsequent first editions of *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards* and the *Registry of USG Recommended Biometric Standards (Registry)*.”

The use of any of the standardized minutiae records for encoding enrollment records to be used in the background or search in civil or criminal searches is insufficient, and the standards may only be used as supplemental material to a fingerprint image.

For all new USG fingerprint minutiae-based applications, the standards should be formally profiled. This will enumerate which of the options are permitted and instantiate minimum and maximum values for variable that the generic base standards do not prescribe.

NIST should coordinate USG positions on the revision of minutiae standards.

USG should develop default or candidate profiles for verification of fingerprint minutiae applications.

NIST should conduct further Minutiae Exchange (MINEX) research, development, test and evaluation rounds to improve minutiae-based accuracy and interoperability. Such work should include extant standardized records and emerging Extended Fingerprint Feature Sets.

USG should develop technical means, including open-source tools, for transcoding minutiae between instances of the standards, e.g. a minutiae record conforming to ISO/IEC 19794-2:2005 transcoded to the ANSI/NIST ITL 1-2007 Type-9 minutia data record.

NIST should conduct further Evaluation of Latent Fingerprint Technologies (ELFT) research, development, test and evaluation rounds to improve accuracy, and to evaluate performance of standardized latent fingerprint feature encodings.

NIST should conduct or otherwise coordinate evaluation of standardized encoding of fingerprint information.

USG applications should adhere to the ISO and ANSI standards to the maximum extent possible but with the recognition that strict adherence may not be feasible, advisable, or cost efficient. Therefore a specific application profile should be developed that deals with the issue of which parts of the standards are not to be adhered to in the particular application. The goal is to assure machine interoperability and accuracy.

USG agencies should reflect this policy in any agency specific adoption processes (e.g., DoD DISR, DHS TRM).

Agency standards registries should harmonize with this policy.

A.3 Latent Fingerprint Standard

Issue

The ability to transmit and process latent fingerprint images is of critical importance in the criminal law enforcement and homeland and national security domains.

The information provided in this Annex is a summation of the analyses performed by the SCA WG in the **first part of 2007**. Therefore some of the information below is now out-of-date. These analyses are being provided to give the background on the issues that shaped the subsequent first editions of *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards* and the *Registry of USG Recommended Biometric Standards (Registry)*.”

Analysis of Issue

Performance of latent examiners and automated biometrics system is strongly dependent on the acquisition and transmission of the latent images. The ability of modern latent matching technologies to conduct accurate one-to-many searches remains problematic and a high-end research topic. Two search paradigms are: Search of latent images against massive repositories of ten-print records (the typical criminal case); and comparison of a single ten-print record against a watchlist of latent images (the KST case).

The relevant acquisition and transmission standards may be incomplete in supporting lights-out evaluation of automated latent matching technologies (for example, in connoting mirror-imaging).

Potential Solutions

In all future applications, latent fingerprint and palm images should be stored in Type 13 records of the ANSI/NIST ITL 1-2007 standard. That standard's Type 7, 9 and 14 records should not be used. The INCITS 381 and ISO/IEC 19794-4 standards should not be used.

NIST should continue its ELFT series of performance-based evaluations of latent fingerprint technologies. These evaluations should be extended to include evaluations of standardized feature sets. NIST should propagate successfully evaluate feature data through the international standards community.

NIST should initiate and support formal standardization of one-to-many latent evaluations in SC 37 Working Group 5.

NIST should coordinate an interagency and international collaboration to collect and construct reference latent fingerprint and palm image databases. Such collections should include acquisition of mated ten-print records. These should be made available for research and development. NIST should sequester test data for its ELFT evaluations. NIST should support research and development by allowing testing via its Rapid Evaluation infrastructure.

USG agencies should reflect this policy in any agency specific adoption processes (e.g., DoD DISR, DHS TRM).

Agency standards registries should harmonize with this policy.

A.4 Face Image Standard (2D)

Issue

USG agencies have ongoing requirements to capture, store and exchange face biometric data. The best way to meet these requirements is for USG agencies to use the same biometric data interchange format standard for face images. While the standards support

The information provided in this Annex is a summation of the analyses performed by the SCA WG in the **first part of 2007**. Therefore some of the information below is now out-of-date. These analyses are being provided to give the background on the issues that shaped the subsequent first editions of *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards* and the *Registry of USG Recommended Biometric Standards (Registry)*.⁴

data exchange, they also contain requirements for the capture of the image in such a manner as to optimize the performance of facial biometric matching systems.

Analysis of Issue

There are three data interchange standards for face images. They establish formats for the data, but they also include quality-related requirements for the photographic capture process.

- ISO/IEC 19794-5:2005 Biometric Data Interchange Format - Part 5: Face Image Data
- INCITS 385-2004 Face Recognition Format
- ANSI/NIST ITL 1-2007 Data Format for the Interchange of Fingerprint, Facial, & Other Biometric Information - Type-10 Facial and Scars, Marks and Tattoos image record

ISO/IEC 19794-5:2005 and INCITS 385-2004 both support three face image types: basic, full-frontal, and token image. The basic image can be any image of a face. The full-frontal image is a well-posed passport-style frontal image. The token image is a geometrically constrained frontal image that requires an eye-finding algorithm to drive correction of rotation, scale and position.

The INCITS 385-2004 and ISO/IEC 19794-5:2005 standards are primarily intended to support formal enrollment processes. The ANSI/NIST ITL 1-2007 standard supports a greater diversity of applications.

The 2D content of all three standards is stable. The 3D content of INCITS 385-2004 is recently final but is likely to differ from that of ISO/IEC 19794-5:2005, which remains under development. Revisions also include information concerning the acquisition of face images.

A detailed comparison of the differences between ANSI INCITS 385-2004 and ISO/IEC 19794-5:2005 has been published.⁴ The differences between the base standards are minor. The ISO standard has been formally amended to include an informative annex on how best to acquire images from cooperative subjects.

There are many options within these standards and, each application must specifically determine which parts are to be used. The compilation of these specifications should be included in the Application Profile.

Potential Solutions

In all new USG biometric face applications in which cooperative subjects are photographed for enrollment or registration, the images collected should conform to the ISO/IEC 19794-5:2005 Face Image Data standard, for the capture, storage, and data exchange of face image data. This should include the Amendment 1 constraints on image

⁴ http://www.incits.org/tc_home/m1htm/2006docs/m1060976.pdf

The information provided in this Annex is a summation of the analyses performed by the SCA WG in the **first part of 2007**. Therefore some of the information below is now out-of-date. These analyses are being provided to give the background on the issues that shaped the subsequent first editions of *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards* and the *Registry of USG Recommended Biometric Standards (Registry)*.”

capture. The acquisition should be designed to be frontal and the result should be a conformant Full Frontal or Token instance. Applications should be designed to capture at least 90 pixels between the eyes from all subjects.

The images collected in all new USG biometric face applications in which subjects are imaged in a non-cooperative or covert manner should conform to the ANSI/NIST ITL 1-2007 Type-10 face record with subject acquisition profile (SAP) of level 1 or above. The acquisition should be frontal when possible.

For Machine Readable Travel Documents (MRTDs) (e.g., e-Passports and Visas), USG should follow the ISO/IEC 19794-5:2005, which is specified by ICAO 9303.

For all new USG biometric face applications, the standards should be formally profiled. These profiles should enumerate which of the options are permitted and instantiate minimum and maximum values for variables that the generic base standards do not prescribe. This should include specification of the maximum compression ratios and compression algorithms.

USG should develop default or example or candidate profiles for face image enrollment.

USG should develop technical means, including open-source tools, for transcoding images between instances of the standards, e.g. face images conforming to ISO/IEC 19794-5:2005 transcoded to the ANSI/NIST ITL 1-2007 Type-10 face record.

USG applications should adhere to the standards to the maximum extent possible but with the recognition that strict adherence may not be feasible, advisable, or cost efficient, therefore a specific application profile must be developed that deals with the issue of which parts of the standards are not to be adhered to in the particular application. The goal is to assure machine interoperability and accuracy.

As an example, at Ports of Entry (POE) the background is not controllable as required in the standards. This is a deviation, and while it may lead to some drop in face detection performance, it is unlikely to affect machine readability. For this reason a specific application profile may include limited, specified, deviations from the standard. Such deviations should be reported to NIST in each agency’s annual reporting in accordance with the NTTAA.

USG agencies should reflect this policy in any agency specific adoption processes (e.g., DoD DISR, DHS TRM).

Agency standards registries should harmonize with this policy.

A.5 Iris Image Standard

Issue

USG agencies have ongoing requirements for the capture, storage, use, and exchange of iris biometric data. The best way to meet these requirements is for USG agencies to use the same biometric data interchange format standard for iris images.

The information provided in this Annex is a summation of the analyses performed by the SCA WG in the **first part of 2007**. Therefore some of the information below is now out-of-date. These analyses are being provided to give the background on the issues that shaped the subsequent first editions of *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards* and the *Registry of USG Recommended Biometric Standards* (Registry).”

Analysis of Issue

There are three voluntary consensus data interchange format standards for iris images presently available:

- ISO/IEC 19794-6:2005 Biometric Data Interchange Format - Part 6: Iris Image Data
- ANSI INCITS 379-2004 Iris Image Interchange Format
- ANSI/NIST-ITL 1-2007 Data Format for the Interchange of Fingerprint, Facial, & Other Biometric Information - Type-17 Iris image record

As of this writing, INCITS M1 issued 30-day letter ballots to approve the withdrawal of ANSI INCITS 379-2004 as an American National Standard and to approve the withdrawal of Project 1576-D – Revision of INCITS 379-2004.

Should these letter ballots not pass, it is important to note that these two standards have options that result in a potential implementation issue for the USG. ISO/IEC 19794-6:2005 and ANSI INCITS 379-2004 both specify two alternative image interchange formats for biometric authentication systems that utilize iris recognition. The first format is based on a rectilinear image storage format that may be a raw, uncompressed array of intensity values or a compressed format such as that specified by ISO/IEC 15444. The second format is based on a polar image specification. A detailed comparison of the differences between ANSI INCITS 379-2004 and ISO/IEC 19794-6:2005 has been published.⁵ A major difference between these two standards is the polar coordinate conversion.

The ANSI/NIST-ITL 1-2007 Type-17 Iris image record only specifies a rectilinear image storage format, which is compatible with the rectilinear image storage format in ISO/IEC 19794-6:2005.

Potential Solutions

All new USG biometric iris applications should conform to the rectilinear image format requirements of ISO/IEC 19794-6:2005, *Biometric Data Interchange Format - Part 6: Iris Image Data*, for the capture, storage, and data exchange of iris image data. These requirements are compatible with the ANSI/NIST ITL 1-2007 Type-17 Iris image record. (Note: The ANSI/NIST ITL 1-2007 Type-99 CBEFF biometric data record is explicitly disallowed for use to exchange the rectilinear image storage format in ISO/IEC 19794-6:2005.)

Iris images conforming to the polar image format requirements of ISO/IEC 19794-6:2005 may be retained only if their rectilinear parents are also retained. If the USG receives a polar image only, the data may be retained but should be transcoded to a rectilinear

⁵ http://www.incits.org/tc_home/m1htm/2006docs/m1060977.pdf

The information provided in this Annex is a summation of the analyses performed by the SCA WG in the **first part of 2007**. Therefore some of the information below is now out-of-date. These analyses are being provided to give the background on the issues that shaped the subsequent first editions of *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards* and the *Registry of USG Recommended Biometric Standards (Registry)*.”

format. If USG receives an unformatted raw raster image, it should be encoded as a rectilinear image.

For all new USG biometric iris applications, the standards should be formally profiled. These profiles should enumerate which of the options are permitted and instantiate minimum and maximum values for variables that the generic base standards do not prescribe. Each profile should include specification of the maximum compression ratios and compression algorithms.

USG should develop default or example or candidate profiles for iris image enrollment in rectilinear format.

USG should develop technical means, including open-source tools, for transcoding images between instances of the standards, e.g. an iris image that conforms to the rectilinear image format requirements of ISO/IEC 19794-6:2005 transcoded to the ANSI/NIST ITL 1-2007 Type-17 Iris image record.

USG applications should adhere to the standards to the maximum extent possible but with the recognition that strict adherence may not be feasible, advisable, or cost efficient. In cases where a deviation from this policy is necessary, the specific application profile for that project must be developed that deals with the issue of which parts of the standards are not to be followed. This deviation must be listed in the agency’s annual report to NIST on compliance with the terms of the National Technology Transfer and Acquisition Act (NTTAA).

USG agencies should reflect this policy in any agency specific adoption processes (e.g., DISR, DHS TRM).

Agency standards registries should harmonize with this policy.

A.6 Voice Standard

Issue

USG agencies have ongoing requirements to capture, storage, use, and exchange voice biometric data for personal recognition. The best way to meet these requirements is for USG agencies to use the same biometric data interchange format standard for voice data.

Analysis of Issue

There are two published standards related to the identification of speakers using voice information:

- VoiceXML2.0
- Speaker Verification API

Additionally, two data interchange formats are under development at the national and international levels that allow the exchange of speaker data.

The information provided in this Annex is a summation of the analyses performed by the SCA WG in the **first part of 2007**. Therefore some of the information below is now out-of-date. These analyses are being provided to give the background on the issues that shaped the subsequent first editions of *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards* and the *Registry of USG Recommended Biometric Standards (Registry)*.”

- INCITS Project 1821-D: Speaker Biometrics Format for Data Interchange, which is the product of collaboration between the Speaker Biometrics Committee of the VoiceXML Forum (SBC), a liaison member of M1, and INCITS M1.
- ISO/IEC 19794 - Biometric data interchange format – Part 13: Speech data interchange format for speaker recognition, which is a recently approved project within JTC 1 SC 37.

The INCITS M1 project intends to define a method for characterizing the speech produced by an end user for enrollment, verification, or identification. It supports transmission of raw speech data with an optional extension for proprietary data. It defines the attributes that are needed to generate a voice model from the dialog and turns and includes the XML representation of those attributes. The USG has the option (recommended at this point) to require only the raw data and deprecate use of the optional extended data. Although as stated above, it currently specifies an optional extension for proprietary data (this could include vendor-dependent feature data or other types of data).

The JTC 1 SC 37 project intends to specify speech data interchange format(s) for speaker recognition. One data interchange format will support raw speech; other formats could include formats for interchange at the feature vector level.

At this time there are no known major implementations that include biometric standards for speaker identification or verification.

Potential Solutions

The standards need to become more stable before policy can be determined. A preliminary assertion is that all future USG biometric voice applications might require conformance to the national voice standard (once published). Although allowing extended optional data might be application dependent, the policy might require deprecating use of this extended optional data perhaps through profiling the base standard or affecting its content before the standard is completed.

USG should invest in the standards development and progress of R&D to support agency needs and implementations for voice applications.

USG should participate in the development of the national and international standards including the INCITS M1, ISO/JTC1/SC37.

Agencies should participate in interagency biometric standards working groups to communicate and define agency-specific requirements on operational use for voice applications.

The information provided in this Annex is a summation of the analyses performed by the SCA WG in the **first part of 2007**. Therefore some of the information below is now out-of-date. These analyses are being provided to give the background on the issues that shaped the subsequent first editions of *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards* and the *Registry of USG Recommended Biometric Standards (Registry)*.⁶

A.7 DNA Data Standard

Issue

USG agencies have ongoing requirements to capture, storage, use and exchange DNA biometric data for personal recognition. The best way to meet these requirements is for USG agencies to use the same biometric data interchange format standard for DNA.⁶

Analysis of Issue

A data interchange format is under development at the international level to allow the exchange of DNA data.

- ISO/IEC 19794 - Biometric data interchange format – Part 14: DNA Data, which is a recently approved project within JTC 1 SC 37.

INCITS M1, the U.S. TAG for JTC 1/SC 37 on Biometrics is concerned that the scope of the working draft for 19794-14:

- Exceeds international DNA data exchange intent;
- Requires core loci that are primarily European-centric; and
- Contains searching, matching, and reporting requirements.

INCITS M1 further recommends that 19794-14 should concentrate on the following issues:

- Standardize DNA profile nomenclature;
- Standardize data exchange format;
- Remove core loci requirement and allow each country or each application domain to define which core loci they require through their respective application profiles;
- Eliminate searching, matching, and reporting requirements or move them to an informative annex; and
- Establish liaisons with multi-national advisory committees, such as European Network of Forensic Science Institutes (ENFSI), the Scientific Working Group on DNA Analysis Methods (SWGDM), and the European DNA Profiling Group (EDNAP).

This project is intended to support the future emergence of DNA profiling systems that can produce electronic results (without manual intervention) within a few hours (automatic identification). Such automatic identification systems are not yet a reality; laboratory equipment, expert human supervision, and a lengthy identification period is

⁶ Note: This issue does not address how to collect, store, transfer, or protect DNA samples. It is solely concerned with consistent data formatting of information used by DNA matching processes.

The information provided in this Annex is a summation of the analyses performed by the SCA WG in the **first part of 2007**. Therefore some of the information below is now out-of-date. These analyses are being provided to give the background on the issues that shaped the subsequent first editions of *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards* and the *Registry of USG Recommended Biometric Standards (Registry)*.”

the current state of the art, but this is expected to improve on a time-scale similar to the production of an International Standard. This International Standard anticipates a reduction of human involvement and a reduction of the identification (enrollment and comparison) time, so that the identification becomes fully automatic.

Potential Solutions

The USG should develop a consolidated, consistent approach to DNA data reporting and participate in the standards development organization bodies proposing formats for DNA data storage and transmission.

USG should participate in the development of DNA standards through INCITS M1 and coordinate activities across disciplines, including biometric and medical standards bodies.

A.8 Multi-biometric Fusion

Issue

Multi-biometric fusion refers to any mechanism for combining information from:

- Multiple modalities, e.g. iris and fingerprint;
- Multiple samples, e.g. images of the right index and right middle fingers;
- Repeated samples, e.g. three passport images of a person over time;
- Samples gathered in sequential or otherwise staged process biometrics;
- Multiple algorithms, e.g. matching implementations from providers A, B and C.

Analysis of Issue

These offer substantial improvements in biometric accuracy, with the benefit decreasing in the order listed above, and they work because more information is available. Thus multimodal fusion is effective because two (or more) modes are more uniquely identifying. Multiple-sample fusion is a potent mechanism for utilizing more information in the recognition process. Repeated-sample fusion is particularly effective in cases where a first sample is weak. Multi-algorithmic fusion is effective in situations where different products fail on different samples.

The most readily implemented form of fusion is score-level fusion, in which matcher output scores are fused. It is easy to implement, and is highly effective. A further benefit is that it may be interoperable, i.e. the match scores from products X and Y feed a fusion module provided by supplier Z. Score level fusion is supported by standardized markup for statistical information from each matching implementation.

One draft standard exists. It is presently at stage of public review:

- INCITS 43X, Project Number 1790D, Fusion Information Format.

The standard supports multimodal or otherwise multi-algorithmic fusion processes. It is not needed for multi-sample and repeated sample fusion.

The information provided in this Annex is a summation of the analyses performed by the SCA WG in the **first part of 2007**. Therefore some of the information below is now out-of-date. These analyses are being provided to give the background on the issues that shaped the subsequent first editions of *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards* and the *Registry of USG Recommended Biometric Standards (Registry)*.”

Potential Solutions

All future USG applications should produce a documented assessment of the costs and benefits of using single-supplier multimodal applications vs. interoperable different-supplier applications.

USG should institute mechanisms to identify operational needs and prioritize support of operationally relevant multi-biometric research. The USG should prioritize research into fusion for identification applications. The USG should de-prioritize research that addresses just the matching error rates in verification systems.

USG should support standards development for supporting accurate fusion processes, and for supporting the transmission and storage of fused biometric data, and for storage, interchange and use of multiple or repeated biometric samples.

USG should develop best practices for implementation of multi-biometric fusion.

Agencies should identify agency-specific requirements on use of multi-biometric standards.

USG should support near term development of mechanisms for accurate fusion processes, which should include the transmission and storage of fused biometric data, as well as storage, interchange and use of multiple or repeated biometric samples for large identification systems.

USG should support near term research use of multiple modalities for rectification of extant Type I and Type II consolidation errors in large biometric systems and databases.

USG should support research, development, testing and evaluation of the following items, in each case specifically targeting reduced matching error rates and improved efficiency.

Fusion of biometric modalities:

- Repeated-sample fusion paradigms
- Inclusion of quality values into fusion processes
- Use of un-segmented four-finger fingerprint images as a single biometric
- Use of un-segmented index and middle fingerprint images as a single biometric

A.9 Application Profiles

Issue

In any given application, it will often be insufficient to simply cite a biometric standard and require conformance to it. This arises because the standards have often been drafted to be application independent, and the standards developers had all along intended that the standard should be layered beneath an application profile or a requirements document, or both.

The information provided in this Annex is a summation of the analyses performed by the SCA WG in the **first part of 2007**. Therefore some of the information below is now out-of-date. These analyses are being provided to give the background on the issues that shaped the subsequent first editions of *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards* and the *Registry of USG Recommended Biometric Standards (Registry)*.”

Analysis of Issue

The list below enumerates the kinds of options or open-ended issues that the standards leave to the user:

- In the ANSI/NIST-ITL 1-2007 standard, the Type 14 variable resolution fingerprint record has a set of fields that are optional, and the standard assumes that a consumer of the data (e.g. the FBI) would require the presence and legal population of those fields. For example, while the standard allows the original scanning resolution of a fingerprint to be recorded, it does not require it. Along the same lines a fingerprint quality field is provided but not required.
- Again, in reference to Type 14, the record allows variable resolution data. It does not, for example, mandate acquisition at 500 pixels per inch. Instead an application profile or requirements specification should call out 500 ppi, or perhaps 500 and 1000.
- In ISO/IEC 19794-6:2005, a standard for iris image data interchange, there is the possibility to save a captured iris image in one of two formats, rectilinear or polar. The latter requires use of image processing algorithms substantially more complicated than those needed to store the former conventional line scan image.
- In the ISO/IEC 19794-5:2005, a standard for face image data interchange, there is the possibility to allow acquisition of basic, full frontal, or token images. While the former may be non-frontal, the latter two require the subject's face to be frontal (to within specified limits) and this will drive design.

Potential Solutions

Any USG biometric application profile should select the proper parts of relevant standards for the different biometric modalities or for multi-modal biometric data capture and conform to appropriate selected standards. For unforeseen combination of factors, the biometric profile should provide a methodology to determine the proper combinations of parts from the relevant standards and document results. The resulting application profiles will be published as USG best practices.

USG applications should embed strong line-by-line profiling of the standards. As an example the following table shows an extract of the NIST Special Publication 800-76 profile of the INCITS 378-2004 minutiae record. It specifically calls out 500 ppi acquisition (line 22+23) of two index fingers (line 27) that must not be rolled fingerprints (line 29).

Extract of INCITS 378-2004 profile showing refined specification of requirements

Line	Clause of the base standard	Application-specific requirements	Application – Rationale for Requirement
22	X (horizontal) resolution (6.4.9)	500	Parent images must be 500 ppi. This ensures interoperability with legacy data.
23	Y (vertical) resolution (6.4.10)	500	
24	Number of Finger Views (6.4.11)	2	Application requires two finger templates

The information provided in this Annex is a summation of the analyses performed by the SCA WG in the **first part of 2007**. Therefore some of the information below is now out-of-date. These analyses are being provided to give the background on the issues that shaped the subsequent first editions of *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards* and the *Registry of USG Recommended Biometric Standards (Registry)*.”

27	Finger Position (6.5.1.1)	1-10	These should be index fingers, but any allowed
29	Impression Type (6.5.1.3)	0 or 2	Flat impressions (not rolled) live or from paper
30	Finger Quality (6.5.1.4)	20,40,60,80,100	A quality values
31	Number of Minutiae (6.5.1.5)	$0 \leq M \leq 128$	Cap the maximum size of the record
36	Extended Data Block Length (6.6.1.1)	0	No proprietary extensions allowed

Each new USG biometric system (or grouping of systems) or application should develop an application profile. The profile should address on a line-by-line basis all the normative clauses of the target standard. Where appropriate:

- Values of parameters should be called-out,
- Normative practice should be called out,
- Informative material should be elevated to normative requirements,
- Normative requirements should be dropped if compliance would be problematic (such a step should be undertaken only with a well document rationale based on empirical evidence). This practice should be undertaken with utmost caution because conformance to the standard may no longer be claimable.

Configurable elements of standards should be specified as part of requirements documents based on operational needs of the implementations.

A.10 Large Scale Identification Applications

Issue

The Electronic Fingerprint Transmission Specification (EFTS) Version 8.0 is the current specification for interfacing with the FBI Integrated Automated Fingerprint Identification System (IAFIS). The EFTS contains a description of operational concepts, descriptors, and field edit specifications, image quality specifications, and other information related to IAFIS services. ANSI/NIST-ITL 1-2000 is specified in EFTS Version 8.0. This is a revision to EFTS Version 7.1. DoD has developed its own EBTS with the goal of being compatible with the FBI’s EFTS and EBTS. ANSI/NIST-ITL 1-2000/EFTS Version 7.1 and ANSI/NIST-ITL 1-2007/ EBTS Version 8.0 will need to coexist for some time. A migration strategy for the USG is needed.

Analysis of Issue

The Department of Homeland Security’s principal biometric system (IDENT) has moved to the Automated Biometric Identification System (IDENT) Exchange Messages (IXM) Specification.

There is a large movement to move more into the XML based transmission standards but these standards have not been completely flushed out as of yet.

The following standards exist:

- DoJ/FBI/CJIS, Electronic Fingerprint Transmission Specification (EFTS), Version 7.1, May 2005

The information provided in this Annex is a summation of the analyses performed by the SCA WG in the **first part of 2007**. Therefore some of the information below is now out-of-date. These analyses are being provided to give the background on the issues that shaped the subsequent first editions of *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards* and the *Registry of USG Recommended Biometric Standards (Registry)*.”

- DoJ/FBI/CJIS, Electronic Biometric Transmission Specification (EBTS) Version 8.0, June 2007
- ANSI/NIST-ITL 1-2000 *Data Format for the Interchange of Fingerprint, Facial, & Scar Mark & Tattoo (SMT) Information*
- ANSI/NIST-ITL 1-2007 Data Format for the Interchange of Fingerprint, Facial, & Other Biometric Information
- U.S. Army Biometrics Task Force, Department of Defense Electronic Biometric Transmission Specification, November 2006

The scope of the FBI EBTS has expanded over previous versions to include additional biometric modalities (e.g., palmprint, facial, and iris) in recognition of the rapidly developing biometric identification industry. The most recent update of the ANSI/NIST-ITL 1-2000 (ANSI/NIST-ITL 1-2007) standard includes new record types to facilitate data sharing for new biometric modalities. The FBI EBTS integrates biometric data in accordance with the ANSI/NIST-ITL 1-2007 standard. A logical record Type-99 was added to the ANSI/NIST-ITL 1-2007 standard to contain and exchange biometric data that is not supported by other ANSI/NIST-ITL logical record types (e.g., voice records), thus providing a basic level of interoperability and harmonization with the ANSI INCITS biometric image interchange formats. This is accomplished by using a basic record structure that is conformant with INCITS 398-2005, the Common Biometric Exchange Formats Framework (CBEFF) and a biometric data block specification registered with the International Biometrics Industry Association (IBIA). The Type-99 logical record type was created for “exotic” biometric data types and should not be used for existing ANSI/NIST data types. IAFIS will provide identification services for many of these evolving biometric modalities at some time in the future.

Potential Solutions

USG should support interoperability and harmonization between IDENT and IAFIS and affected systems utilizing XML based transmission standards. The Executive Steering Committee (DHS, DoJ, DoS) for the Interim Data Sharing Model (IDSM) should continue operation. This should address:

- Real-time connection of biometric systems operational
- DoJ/FBI/IAFIS ‘wanted’ data, known and suspected terrorist (KST) data to DHS/US-VISIT/IDENT
- DHS deportation, expedited removal data to FBI
- DoS Category 1 visa refusal information to FBI
- Funding to extend capabilities
- Expanding access to additional Governmental entities

The information provided in this Annex is a summation of the analyses performed by the SCA WG in the **first part of 2007**. Therefore some of the information below is now out-of-date. These analyses are being provided to give the background on the issues that shaped the subsequent first editions of *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards* and the *Registry of USG Recommended Biometric Standards (Registry)*.”

DoD/DoJ data linkages should be maintained and promoted, utilizing the DoD data center at FBI/CJIS West Virginia site.

DHS/DoS data linkages should be maintained and promoted, including:

- DoS screening of visa applicants using IDENT
- DHS real-time inspection access to DoS database of visa applicants

Affected agencies should appoint representation to the NSTC Subcommittee on Biometrics and Identity Management as well as to the specific committees and working groups necessary to implement and maintain the capabilities referenced above.

A.11 Smart Cards Applications

Issue

Identity management applications based on user-carried credentials typically store biometric data on un-powered token devices. The archetype here is the ISO/IEC 7816 smart card credential (the US Government PIV card) which is a cryptographically enabled token embedding the cardholder's biometric data. These devices are additionally attractive because a number of FIPS 140-2 certified products exist today.

Analysis of Issue

Smart cards typically offer limited storage. In addition the computational resources needed to implement certain biometric operations and cryptographic encryption and digital signature computations is high and is often dependent on the size of the data in question. For these reasons, it is imperative that the stored biometric data is compact and standardized encodings need to support such constraints.

Potential Solutions

All biometric sample data stored on ISO/IEC 7816 smart cards, whether raw or processed, in standardized or proprietary format, should be stored in conformance with ISO/IEC 7816-11. In such data should be accompanied by digital signatures specified in NIST Special Publication 800-78 as revised.

USG should provide funds to extend MINEX series of evaluations. These should be directed at the identification of fingerprint templates that offer improved interoperability.

NIST should base these evaluations on the elemental minutia representations of the INCITS 378:2004 and ISO/IEC 19794-2:2005 standards.

NIST should identify performance-based improvements to these formats and propagate them through the formal standards development process.

NIST should initiate and support formal standardization of match-on-card evaluations SC 37 Working Group 5.

The information provided in this Annex is a summation of the analyses performed by the SCA WG in the **first part of 2007**. Therefore some of the information below is now out-of-date. These analyses are being provided to give the background on the issues that shaped the subsequent first editions of *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards* and the *Registry of USG Recommended Biometric Standards (Registry)*.”

A.12 Mobile and Portable Biometric Devices

Issue

Lightweight, portable fingerprint collection devices are increasingly in operational US Government applications. The US Government maintains large repositories of operational ten-print fingerprint data that is almost universally collected on optical sensors, often EFTS/F certified. The US Government therefore has a compelling objective that data collected on low power devices in the field is interoperable with other systems.

Application of mobile and portable biometric devices for screening and counterterrorism applications and for the agencies within the counterterrorism community, biometrics of known or suspected terrorists (KSTs) can be used to enhance and expand existing watchlist and screening functions.

Standardization will reduce the likelihood of deployment of mobile biometric systems that do not perform in the manner desired or afford interoperability with other systems.

Analysis of Issue

The collection of biometric data is often performed in an uncontrolled environment, particularly when dealing with KSTs. The biometric data itself may be for enrollment and include associated biographic and situation descriptive material or it may be used to check against existing databases to determine if there has been previous contact with this individual (possibly under a different assumed identity). Thus, it is extremely important that the biometric data itself be of the highest possible quality and that the biometric sample be collected in a manner so as to minimize potential harm to the data collector or the subject. The time for collection must be reasonable for the given circumstances. In addition, the biometric sample(s) must be usable in the other systems which might rely upon KST watch lists. All of these issues are important and the adoption and application of relevant standards can significantly improve the likelihood of easier and more accurate/usable data collection efforts.

Potential Solutions

USG should continue to develop and support mobile, rugged, and portable biometric collection devices to work in austere environments. Mobile biometric solutions must demonstrate long operational life as well as rapid and high-quality data capture and collection at stand-off ranges sufficient to ensure operator safety.

USG should develop application profiles describing which parts of existing biometric/ergonomic/safety/security and other relevant standards are applicable for mobile biometric data collection activities. This should address both the ‘store and forward’ type of operation as well as those with direct/real-time links to databases. It will also need to address local checking against a limited database.

The development of an “application profile” that is required for all procurement of mobile biometric capture devices will ensure that data is collected consistently and in a

The information provided in this Annex is a summation of the analyses performed by the SCA WG in the **first part of 2007**. Therefore some of the information below is now out-of-date. These analyses are being provided to give the background on the issues that shaped the subsequent first editions of *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards* and the *Registry of USG Recommended Biometric Standards* (Registry).”

usable form. It will also ensure that when mobile devices are used in a verification/identification capability that the probe image is of sufficient quality to be likely to yield a correct match/non-match response from the matchers. This applies to all biometric modalities, including fingerprint, face, iris, voice and DNA.

US Government agencies should field certified devices. The certification procedure should embed the Federal Bureau of Investigation's single finger assessment of the imaging properties of the device and a performance interoperability test.

USG should develop multiple profiles to support various operational requirements for handheld biometric devices. This should be the responsibility for each Agency proposing a system using handheld biometric devices.

The proposed “application profiles” should select the proper parts of relevant standards for the different biometric modalities (and for multi-modal biometric data capture) and map them to generalized mobile scenarios. For unforeseen combinations of factors, the document will provide a methodology to determine the proper combinations of parts from the relevant standards. The resulting “application profile” will be published as a USG standard.

USG should establish a performance-based evaluation program. A submitted capture device should be used in a scenario-test collection conformant to the provisions of ISO/IEC 19795-2. The resulting samples should be assessed for interoperability with optical data conformant in a test conformant to ISO/IEC 19795-4.

NIST should test and publish reports that include empirical data about limited size, resolution, and other factors on performance in order to allow application profile developers to examine trade-offs in the designing of systems for their specific requirements.

A.13 Conformance Testing

Issue

To establish a high level of confidence that standards-based biometric equipment, software and data perform as expected in USG biometric applications, standards based conformity assessment is critical. Standards alone are insufficient to ensure interoperability and proper performance of USG systems, components, and applications.

Analysis of Issue

Conformity testing is the process of testing a technology implementation that claims to support a standard to determine if the implementation adheres to the referenced standard. *Conformance assessment* standards specify the manner in which a conformity assessment should be performed and recorded. Conformance testing captures the technical description of a standard and measures whether an implementation faithfully implements the standard. This is the most obvious type of testing. For instance, when a photograph is taken of an individual, does it meet the requirements for use in a face recognition system? Are there a sufficient number of pixels between the eyes? Is the pose full frontal? Do

The information provided in this Annex is a summation of the analyses performed by the SCA WG in the **first part of 2007**. Therefore some of the information below is now out-of-date. These analyses are being provided to give the background on the issues that shaped the subsequent first editions of *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards* and the *Registry of USG Recommended Biometric Standards (Registry)*.”

biometric data records, data structures and applications conform to required open system standards? Standards-based conformance testing tools help both developers and users by validating conformance claims, leading to greatly increased levels of confidence in products. Conformance testing can also help ensure interoperability between standards-based products and systems.

Standards bodies are developing and have published several conformance testing standards for technical interfaces and data interchange formats applicable for many biometric modalities. However, the USG is unable to verify vendor claims of conformance without established second or third party conformity assessment programs. Although other industries have established conformity assessment programs, this area, while critical, remains undeveloped in the biometrics industry.

There are no ongoing or planned USG second or third party conformity assessment programs. As an initial step, the DoD developed in May 2004 a technical report titled “Biometrics Conformity Assessment Program for DoD”. The report details the necessary steps, policies and activities necessary to establish a Biometric Conformity Assessment program within DoD. An article on DoD Biometric Conformity Assessment Initiative has been published in the Defense Standardization Program Journal in April/June 2005 issue.

Potential Solutions

USG should establish Biometric Conformance Assessment (BCA) programs for validating to standards and performance of biometric devices and systems for certain USG biometric applications.

USG should establish a Second- or third-party testing program(s) to achieve a high level of assurance of standards conformance by systems and components required for government standards implementations. USG should ensure that the BCA programs do not rely on vendor claims of conformance since first-party (vendor) testing is not sufficient.

USG should designate a USG entity (or entities) as a Certification Authority within the BCA responsible for evaluation (certification) of test results and for issuance and maintaining of the validated product lists/qualified product lists or certificates of conformance.

Agencies should establish agency requirements and needs for a USG Biometric Conformity Assessment Program.

Agencies should develop a unified Conformity Assessment guidelines document for circulation within the USG.

The information provided in this Annex is a summation of the analyses performed by the SCA WG in the **first part of 2007**. Therefore some of the information below is now out-of-date. These analyses are being provided to give the background on the issues that shaped the subsequent first editions of *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards* and the *Registry of USG Recommended Biometric Standards (Registry)*.”

A.14 Performance Testing

Issue

In many large scale USG, cross-border or otherwise federated applications, biometric data captured and processed using one system may later be involved in verification or identification transactions with data collected and processed using another system. For example, fingerprints collected on a mobile sensor might be submitted to an identification system containing sets of fingerprint images captured during consular interviews. This presents interoperability issues:

- Are the sensors interoperable?
- Are the data interchange format standards compatible with one another?

Analysis of Issue

Biometric performance testing is concerned with measurement of the verification and identification error rates, and throughput performance, of biometric algorithms, components, devices and systems.

There are three published standards applicable to performance testing of biometric systems:

- ISO/IEC 19795-1:2006 Biometric Performance Testing and Reporting - Part 1: Principles and Framework
- ISO/IEC 19795-2:2007 Biometric Performance Testing and Reporting - Part 2: Testing Methodologies for Technology and Scenario evaluations
- ISO/IEC 19795-4:2007 Biometric Performance Testing and Reporting - Part 4: Interoperability Performance Testing

The standards are intended to do different things. ISO/IEC 19795-1 is a framework that should be required for all tests. ISO/IEC 19795-2 is appropriate to scenario or technology tests. There are options within ISO/IEC 19795-2 that should be profiled to govern the conduct of just a scenario test or just a technology test.

Potential Solutions

All new USG sponsored or mandated laboratory tests of commercial verification systems should conform to ISO/IEC 19795-1 and the scenario testing provision of 19795-2. When a test does not conform to specific sub-clauses, explanatory statements, excerpting the standard, should be included in the test reports.

All new USG sponsored laboratory tests of matching algorithms should conform to the technology testing provisions of ISO/IEC 19795-2. When a test does not conform to

The information provided in this Annex is a summation of the analyses performed by the SCA WG in the **first part of 2007**. Therefore some of the information below is now out-of-date. These analyses are being provided to give the background on the issues that shaped the subsequent first editions of *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards* and the *Registry of USG Recommended Biometric Standards (Registry)*.”

specific sub-clauses, explanatory statements, excerpting the standard, should be included in the test reports.

All new USG sponsored tests of access control devices should conform to scenario testing provisions of ISO/IEC 19795-2

For all new USG applications, a policy and approach toward operational testing of the fielded system should be formulated. This should address, at least, the procurement of zero or more instances of the system that would be specifically instrumented to support capture of operational samples, and offline analysis thereof.

USG should revisit the above-stated policy on the conduct of the access control tests once the new ISO/IEC 19795-5 Scenario Evaluation of Biometric Access Control Systems has been completed.

USG should develop a strategy approach toward operational testing of potential fielded biometric systems and institute consistent testing procedures to support procurement actions.

Agencies should participate in standards development organizations and should advocate for, and support, tests of the effectiveness of biometric standards both during and after their development.

A.15 Interoperability Testing

Issue

In many large scale USG, cross-border or otherwise federated applications, biometric data captured and processed using one system may later be involved in verification or identification transactions with data collected and processed using another system. For example, fingerprints collected on a mobile sensor might be submitted to an identification system containing sets of fingerprint images captured during consular interviews. This presents interoperability issues:

- Are the sensors interoperable?
- Are the data interchange format standards compatible with one another?
- Are sensors and matching systems by different vendors interoperable?

Analysis of Issue

Biometric interoperability testing is concerned with the ultimate ability of cross-vendor, cross-implementation and cross-format biometric samples to be accurately verified or identified. This might involve assessing sensor performance, the viability of a data interchange format standard, the ability to upgrade a system from one provider to another. Interoperability testing is particularly important when different suppliers and manufacturers may provide software and/or hardware to various parts of the system that is viewed as a whole. The importance of testing is highlighted by this real-life example: At an ICAO NTWG meeting in October 2003, manufacturer representatives claimed to

The information provided in this Annex is a summation of the analyses performed by the SCA WG in the **first part of 2007**. Therefore some of the information below is now out-of-date. These analyses are being provided to give the background on the issues that shaped the subsequent first editions of *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards* and the *Registry of USG Recommended Biometric Standards (Registry)*.”

ICAO that their products would be interoperable since they would conform to ICAO-established standards for the chip, data transmission, data format and content as well as for the chip readers. In February 2004 DHS co-sponsored a test session, where manufacturers of chips and readers were invited to demonstrate interoperability. The result was that no chip product was interoperable with the set of chip readers. Subsequent venues allowed manufacturers to develop and test products so that e-passports would be interoperable.

Some tests, such as Minutiae Exchange Interoperability Test (MINEX-conducted by NIST in an on-going basis for fingerprints) can assist in determining the relative levels of performance and interoperability based upon the capture device, minutia extraction and matcher. This concept of allowing vendors to self-test when ready should be expanded to the full range of biometric modalities.

Specific uniform procedures and standards must be established for interoperability testing for a wide variety of biometrics products.

Interoperability testing has been standardized in ISO/IEC 19795-4 FCD Biometric Performance Testing and Reporting - Part 4: Performance Interoperability Testing.

One mechanism to ensure sensor interoperability is to set acceptable minima for the relevant physical properties of the sensor. This has been done by the FBI for fingerprint sensors:

- For ten-print capture, see EFTS Appendix F IAFIS Image Quality Specifications
- For single finger capture see Personal Identity Verification (PIV): Image Quality Specifications for Single Finger Capture Devices.

As another example, NIST Special Publication 800-76-1 cites ISO/IEC 19795-4 to regulate fingerprint minutia interoperability testing.

Potential Solutions

USG should continue to support interoperability and performance testing for large scale biometric and identity management applications to ensure cross-vendor, cross implementation, and cross format biometric samples are accurately verified or identified.

All USG large scale applications, cross-border or otherwise federated applications, involving interoperable data formats, should reference, sponsor or conduct tests conforming to ISO/IEC 19795-4.

USG, and USG agencies participating in standards development organizations, should advocate for, and support, tests of the effectiveness of biometric standards both during and after their development.

Each Agency should institute consistent testing procedures as part of any new biometric application.

The information provided in this Annex is a summation of the analyses performed by the SCA WG in the **first part of 2007**. Therefore some of the information below is now out-of-date. These analyses are being provided to give the background on the issues that shaped the subsequent first editions of *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards* and the *Registry of USG Recommended Biometric Standards* (Registry).”

USG should develop a strategy approach toward operational testing of potential fielded biometric systems and institute consistent testing procedures to support procurement actions.

Agencies should participate in standards development organizations and should advocate for, and support, tests of the effectiveness of biometric standards both during and after their development.

A.16 Security Testing

Issue

Biometric systems may be actively attacked, in an attempt to illicitly gain access (in access control), or to insert/modify/delete identities or to evade detection in a one-to-many search. A number of kinds of attack are known, and these may be modeled in testing. Such testing is distinct from biometric performance testing which usually addresses system or component accuracy and capability.

Analysis of Issue

The principal question is: Does the standard include device attacks or attacks to circumvent portions of the system? This has been addressed by ISO/IEC 19792 Security Evaluation of Biometrics, which is under development in SC 27. It considers active attacks and differentiates between biometric components, systems, and applications. It quantifies security in terms of error rates, including the error rate encountered given specific active impostor attempts. It includes requirements on testing of vulnerability and on protection.

Secure biometric systems begin at conception. Red teaming and security involvement should occur throughout major system development to include system design. Similar efforts should be continually employed against the various underlying biometric algorithms, components, and devices. Red Teaming should also be focused on the underlying IT and telecommunications infrastructure upon which the biometric system rely. (Red teaming is the use of a person or group of people who attempt to defeat a system, reporting back their findings to the system owners/operators).

Certain security systems depend on a biometric comparison to serve as a supplemental authentication factor. The security module may need information from the biometric device concerning the context in which it was tested or certified status. For example, if the device has only been tested to a false acceptance rate of 0.02, this may be insufficient for the high security application.

Potential Solutions

USG should support development and adoption of biometric security testing standards.

USG should conduct security tests of biometric algorithms, components and devices.

USG should formulate a position on the use of such standards as they become available.

The information provided in this Annex is a summation of the analyses performed by the SCA WG in the **first part of 2007**. Therefore some of the information below is now out-of-date. These analyses are being provided to give the background on the issues that shaped the subsequent first editions of *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards* and the *Registry of USG Recommended Biometric Standards (Registry)*.”

USG should formulate policy on classification of methods vs. results of such studies.

USG should sponsor specific research into security related properties of algorithms.

The USG should maintain an active role in ISO/IEC JTC 1/SC27.

The USG should review ISO/IEC 19792 Security Evaluation of Biometrics (when published) for applicability in Federal environments and develop a best practices document.

A.17 Establishment of USG QPL Based on Conformance, Performance, and Interoperability Testing

Issue

Presently, there is no commonality in approach across the various USG testing programs and across agencies in developing QPLs or QPL-like lists. This can cause multiple testing of the same product for conformance/performance to the same (or similar) requirements.

Analysis of Issue

In order to ensure that equipment and software is procured that will properly function and meet specifications, pre-qualification of items (based upon specified procedures) may be done. This could result in Qualified Product Lists, Validated Products Lists, Basic Ordering Agreements (IDIQ type of acquisition), or certificates of compliance. For instance, DHS has established a testing program for biometric devices that may be purchased by airport authorities for use in airport access control. The actual testing of devices has been contracted to specific testing laboratories. DHS defines the tests and the test procedures. Another case is GSA and NIST developing test specifications for PIV applications. Yet another example is the testing of slap-print readers according to specifications developed on an interagency basis by DoD, DHS, DoJ/FBI and DoS.

Potential Solutions

USG should examine the principal qualification criteria for product/unit/system qualification, starting with a particular agency. Based upon that agency’s findings and any additional information available from other agencies, the Subcommittee should adopt a USG-wide approach to the testing and certification of biometrics-related products/units/systems.

USG should develop a model to establish a consistent testing approach in developing Qualified Product Lists (QPLs) or QPL-like lists that can be used by various programs for selecting biometric products for new applications. USG should examine the principal qualification criteria for product/unit/system qualification, starting with a particular agency. Based upon that agency’s findings and any additional information available from other agencies, the Subcommittee should adopt a USG-wide approach to the testing and certification of biometrics-related products/units/systems.

The information provided in this Annex is a summation of the analyses performed by the SCA WG in the **first part of 2007**. Therefore some of the information below is now out-of-date. These analyses are being provided to give the background on the issues that shaped the subsequent first editions of *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards* and the *Registry of USG Recommended Biometric Standards* (Registry).”

Agencies should build upon existing applications such as the PIV (DHS/GSA) and Airport Access Control (DHS/TSA) to use these findings and any additional information available from other agencies, to develop a model and a commonality in approach across the various USG testing programs and across agencies in developing QPLs or QPL-like lists.

A.18 Reference Implementations and Data Sets

Issue

In order for the USG to have a robust testing infrastructure in support of deployment of standards-based biometric applications, there is a need for the availability of high quality reference implementations and standard reference data sets.

Analysis of Issue

An important aspect of developing and improving biometric systems and applications is that of reference implementations. This can take several forms, such as:

- NFIQ (for fingerprint image quality) that allows vendors and/or users to examine the quality of fingerprint samples in a common framework.
- Laboratory mock-ups of typical operational environments (such as a mock port-of-entry inspection station).
- Software and hardware ‘duplication’ of operating systems (used to test possible enhancements without disrupting the operational system).

By having a standard reference set of data and specified operating conditions, vendors can evaluate their products and product improvements. Reference data sets should be releasable to the biometrics community, but care must be taken to ‘anonymize’ the data as part of privacy protection guidelines.

Sequestered testing datasets are available, but large-scale test data suites are not (particularly for multi-modal work). This is due to several factors, including the cost of gathering the data; privacy rights of the individuals from whom the samples were taken; administrative requirements; and access rights on data sharing.

Potential Solutions

USG should support development and dissemination of reference data sets for reading, writing and validating conformant instances of the standards.

USG should support development and dissemination of reference data sets for reading, writing and validating conformant instances of the standards.

USG should promote that reference data sets be releasable to the biometrics community, but ensure data sets are utilized in a manner that meets the privacy protection guidelines.

USG should develop public domain software platforms for reference implementation and demonstration prototyping.

The information provided in this Annex is a summation of the analyses performed by the SCA WG in the **first part of 2007**. Therefore some of the information below is now out-of-date. These analyses are being provided to give the background on the issues that shaped the subsequent first editions of *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards* and the *Registry of USG Recommended Biometric Standards (Registry)*.⁷

A.19 Technical Interface

Issue

USG agencies often have requirements for biometric systems that include plug and play capability. This permits agencies to easily/rapidly/seamlessly integrate system components into functioning systems and swap components as needed without losing functionality, such as the ability to achieve data interchange and to protect the biometric data during transmission and storage. Also, USG agencies often have requirements for deploying and invoking biometrics-based identity assurance capabilities that can be readily accessed using web services. The best way to meet these types of requirements is for USG agencies to use appropriate biometric technical interface standards.

Analysis of Issue

Product specific Application Programming Interfaces (APIs) provided with vendor software development kits (SDKs) require application developers and system integrators to develop custom interfaces for each biometric product they use. A biometric API standard known as the Speaker Verification API (SVAPI) was first developed in 1996.

The current BioAPI series of standards support plug and play compatibility by specifying how applications communicate with biometric vendor software in a common way independently of the biometric modality. This supports the swapping of products and incorporation of new products with no application modification. The Common Biometric Exchange Formats Framework (CBEFF) series of standards specify data structures that support multiple biometric technologies in a common way. CBEFF data structures allow exchanging of biometric data and metadata and support security of biometric data in an open systems environment. The Biometric Identity Assurance Services (BIAS) standards define a framework for deploying and invoking biometrics-based identity assurance capabilities that can be readily accessed using web services. The X9.84 and ISO 19092 standards define requirements for the use and management of biometric data and the processes that accompany that use.

Potential Solutions

USG should promote biometric industry product standardization and use of common interface standards such as BioAPI.^{7 8}

⁷ Note: This policy does not apply to law enforcement applications and other large-scale identification applications that require conformance to standards such as FBI EBTS V8.0, DoD EBTS V1.2 or ANSI/NIST-ITL 1-2007.

⁸ Note: There is no requirement for embedded devices to conform to the current versions of the BioAPI standards. This is deprecated because there would be no favorable cost-benefit.

The information provided in this Annex is a summation of the analyses performed by the SCA WG in the **first part of 2007**. Therefore some of the information below is now out-of-date. These analyses are being provided to give the background on the issues that shaped the subsequent first editions of *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards* and the *Registry of USG Recommended Biometric Standards (Registry)*.”

All new USG biometric applications that require plug and play capability without losing functionality (such as may be the case for access control systems) should:

- Require Biometric Information Records (BIRs) conforming to a CBEFF Patron Format (PF) for the processing, exchange, protection, encapsulation, transmission and storage of biometric data. Use existing Patron Formats that permit incorporating in the data structure the required level of additional “processing” and “operational” metadata and data elements that support payload, security/integrity options and creation date/validity period. (Note: Patron Formats specified in INCITS 398, or instantiations of BioAPI BIRs are preferred.). Part 3 of the international version of CBEFF offers other alternatives.
- Encrypt biometric data within the BIRs and sign BIRs by relying on information in the CBEFF BIR Security Block, unless other system security mechanism are already provided by means external to these biometric data structures.
- Require conformance to INCITS 398-2005, Revision 1 Patron Formats for applications that require transmission or storage of BIRs that require clear text biometric headers or making metadata available without processing the record (e.g., for the purpose of indexing BIRs).
- Require conformance to BioAPI standards V1.1 or V2.0 for client-side verification (e.g., enrollment workstation, kiosk) or server-side verification for one-to-one and multi-biometric applications. (Note: The international standard is preferred.)
- Require conformance to SVAPI for applications based only on voice verification.
- Require conformance to the BIAS standard (including the BioAPI requirement) when the application requires the use of biometric technologies in a Service-Oriented Architecture (SOA).

USG agencies should reflect this policy in any agency specific standards adoption process (e.g., DoD DISR, DHS TRM).

Agency standards registries should harmonize with this policy.

A.20 Standardized Measurements for Biometric Sample Quality

Issue

Biometric systems can fail or yield questionable results when sample quality is poor. Biometric sample quality can in large part be ensured by adequate system design. However any inability to regulate the design or the environment, or any adverse behavioral or interactive effects, may cause samples to be ill suited for biometric use despite attempts to follow established procedures.

The information provided in this Annex is a summation of the analyses performed by the SCA WG in the **first part of 2007**. Therefore some of the information below is now out-of-date. These analyses are being provided to give the background on the issues that shaped the subsequent first editions of *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards* and the *Registry of USG Recommended Biometric Standards* (Registry).”

Analysis of Issue

Biometric quality can be quantified by values that are indicative of subsequent matching accuracy. Such quality values, if computed at the time of acquisition, can be used to initiate reacquisition of a sample should the quality be poor. The quality values may also be used to augment subsequent matching processes, or to trigger use of a second biometric modality.

No universally meaningful scale for biometric quality values exists, and a mechanism for tagging samples in the data record with a source designation is only now being standardized. Existing data format standards are under revision to include such attributes. This supports surveying of operational quality and promises increased effectiveness of USG capture, use, and exchange of biometric data.

Biometric sample quality assessment algorithms exist for a number of biometric modalities, both open-source and commercial. The issue of how to conduct a performance test of such algorithms has only recently been investigated and published (NIST, IEEE PAMI, April 2007). A comparative test of such algorithms has never been run, and a standard is warranted to regulate procedures and establish metrics.

Within industry, there are numerous biometric sample quality measurement algorithms. However, the effectiveness of these algorithms in predicting future matching performance has not been evaluated. With the exception of the NIST Fingerprint Image Quality (NFIQ), DoD Fingerprint Image Quality Measurement tool, and DoD prototype Face Image Quality measurement tool, almost all quality tools are proprietary ‘black box’ implementations with no publicly available performance statistics. As such, it is extremely difficult for the USG to make informed decisions with regard to the deployment of specific quality measurement measures and tools without extensive testing.

Potential Solutions

All new USG applications should compute quality scores of all collected biometric modality samples using a consistent methodology suited for the specific modality. When practical, USG entities must avoid the collection and use of insufficient quality biometric samples, as identified by deployed quality measurement algorithms. Quality measurement algorithm identifiers and quality summary statistic within the range [0-100] should accompany each biometric sample.

USG should continue progress towards Quality Score Normalization Dataset (QSND) standardization methods to ensure a consistent and interoperable interpretation of the quality scores.

USG should develop technical means for detecting defective biometric samples and assessing biometric sample quality. Such capabilities should support accuracy-based interoperable standardization of quality values.

The information provided in this Annex is a summation of the analyses performed by the SCA WG in the **first part of 2007**. Therefore some of the information below is now out-of-date. These analyses are being provided to give the background on the issues that shaped the subsequent first editions of *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards* and the *Registry of USG Recommended Biometric Standards (Registry)*.”

USG should establish procedures for evaluating quality assessment implementations in terms of their relation to matching accuracy.

In all new USG biometric applications, the enrollment process should include a quality assessment for each biometric sample and a have a standardized criterion for initiating reacquisition if quality is poor.

All new USG applications should follow the Recommendations on Biometric Quality Summarization across the Application Domain published as [NIST Interagency Report 7422](#).

USG should foster research, development, test and evaluation, and deployment of methods for the rapid detection of defective samples and the quantitative assessment of biometric quality during the acquisition process. This should be done for, at least, fingerprint, facial, iris and speech modalities.

USG should foster research, development, test and evaluation of methods for quantifying quality suitable for human examiner review of samples. USG should support development of methods for appropriate delivery of feedback to users and operators during biometric sample acquisition.

A.21 Human Factors (Usability and Accessibility)

Issue

A system and its components may meet all of the tests mentioned above but still cause system failure. If operators, users and subjects cannot effectively use the system, it is worthless. Usability can include factors such as human factors, accessibility, interpretability of results and instructions, ease of integration, size of unit, required facility modifications for installation, interfaces to existing parts of the system and other factors. The usability factors must be determined for each application; however, some standards can be developed for general types of applications. Human factors and accessibility are particularly good candidates for development of standards. DHS has begun work with NIST in this area.

Some areas of focus for all biometric systems include (but are not limited to):

- Operator interface
- Attended / Unattended / Covert
- Subject Interface
- Acclimated / Non-acclimated
- Cooperative / Non-cooperative / Uncooperative
- Assisted / Non-assisted

The information provided in this Annex is a summation of the analyses performed by the SCA WG in the **first part of 2007**. Therefore some of the information below is now out-of-date. These analyses are being provided to give the background on the issues that shaped the subsequent first editions of *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards* and the *Registry of USG Recommended Biometric Standards (Registry)*.”

- Physical requirements [touching a unit / looking at a camera / ...]
- Output
- Presentation of possible matches above a specified threshold
- Ability to interpret results [red/green condition or specified detailed results depending upon the circumstance]
- Etc.

USG agencies have ongoing requirements for biometric systems that are effective and efficient for users and user performance. To address these requirements USG agencies require guidelines for biometric user interfaces and standards for testing the usability of biometric systems in operational environments that measure user performance including timing, quality, and satisfaction.

Analysis of Issue

ISO 9241-11(1998): *Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs) - Part 11: Guidance on Usability* defines usability as “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use”. The standard identifies three areas of measurement: effectiveness, efficiency, and user satisfaction, where

- 1 Efficiency is a measure of the resources expended in relation to the accuracy and completeness with which users achieve goals. Efficiency is related to productivity and is generally measured as task time
- 2 Effectiveness is a measure of the accuracy and completeness with which users achieve specified goals. Common metrics include completion rate and number of errors.
- 3 User satisfaction is the degree to which the product meets the users’ expectations—a subjective response in terms of ease of use, satisfaction, and usefulness

This standard definition requires identification of the:

- Context of use: The users, tasks, equipment (hardware, software and materials), and the physical and social environments in which a product is used. Examples include: environmental factors such as temperature, humidity, indoors versus outdoors, stationary or mobile system, height of unit, assisted versus unassisted.
- User: The person who interacts with the product. Examples include: users with disabilities, non-English speaking users, cooperative versus un-cooperative users, acclimated versus non-acclimated users.
- Goal: An intended outcome of user interaction with a product. Specific goals relating to user interaction may be referred to as 'task goals'. Examples include: time constraints or the time required to collect biometric samples and the quality threshold for the samples.

The information provided in this Annex is a summation of the analyses performed by the SCA WG in the **first part of 2007**. Therefore some of the information below is now out-of-date. These analyses are being provided to give the background on the issues that shaped the subsequent first editions of *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards* and the *Registry of USG Recommended Biometric Standards (Registry)*.”

- Task: The activities required to achieve a goal. Examples include: the specific process for acquiring the sample, the instructional set, or the order of slaps for fingerprints.

ISO 25062:2006: *Common Industry Format for Usability Testing* provides a standard format for reporting the results of a usability test.

Potential Solutions

USG should include human factors as a significant factor in the design and implementation of any biometric system. Specific design requirements to include at a minimum are:

- Accessibility
- Usability
- Environmental Factors
- Size
- Weight
- Health Effects
- User Interface
- Etc.

USG should have a coordinated interagency strategy for human factors and usability testing for biometric systems that require:

- Identification of the significant characteristics or requirements from the context of use, users, goals and tasks;
- Usability tests to understand the performance implications of these characteristics in terms of efficiency (timing), effectiveness (quality) and user satisfaction;
- Development of standards and/or guidelines that can be instituted in operational environments that compensate for or mitigate the influence of these factors in biometric systems;
- Acceptance test criteria for biometric systems to determine that systems have been tested and meet these standards and requirements before deployment.

USG should support analysis of human factors interfaces to biometric systems and development guidelines for future adoption.

Agencies should work with NIST to coordinate the USG interagency strategy for human factors and usability testing for biometric systems.

The information provided in this Annex is a summation of the analyses performed by the SCA WG in the **first part of 2007**. Therefore some of the information below is now out-of-date. These analyses are being provided to give the background on the issues that shaped the subsequent first editions of *NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards* and the *Registry of USG Recommended Biometric Standards* (Registry).”

A.22 Privacy

Issue

Privacy as a term can signify many different concepts, the extraordinary advances and popularity of information technology bring one conceptualization of privacy – information privacy to the forefront of the privacy protection discussion. Biometric information is, by definition, personally identifiable information. Biometric systems use information generated from observing individuals to recognize a particular individual. Since personal information is any information that *could* be used in *any way* to identify an individual, biometric information is personal information even in those situations where the identity of the individual associated with the biometric information is unknown⁹.

Analysis of Issue

A privacy assessment of a biometric system should be conducted when there is a direct use of personal information to analyze the impact that the use of this data may have on individual privacy interests and to ensure that personal information is used appropriately.

A privacy assessment should examine the stated purpose of the system and compare the purpose to the underlying authority of the organization and the specific authority for the program office that manages the system. The purpose for the system should align with the program office’s specific authority, and the organization’s general authority.

Potential Solutions

USG should request agencies to conduct privacy impact assessment to protect personal data for the implementation of any new biometric systems.

Agencies should recognize that biometric data is personally identifiable information and ensure that all applicable privacy compliance requirements are met prior to loading or using biometric data.

Agencies should conduct privacy impact assessment of biometric systems when there is a direct use of personal information to ensure that personally identifiable information is used appropriately.

⁹ “Privacy & Biometrics: Building a Conceptual Foundation”, September 2006. www.biometrics.gov

Annex B - Acronyms

Acronym / Abbreviation	Definition
AAMVA	American Association for Motor Vehicle Administrators
AHGBEA	Ad-Hoc Group on Biometrics and E-Authentication
ANSI	American National Standards Institute
APB	Advisory Policy Board
BFC	Biometrics Fusion Center (U.S. Army Biometrics Task Force)
BIAS	Biometric Identity Assurance Services
BIP	Biometric Inter-working Protocol
BSP	Biometric Service Provider
BSWG	Biometric Standards Working Group (DoD)
BTF	U.S. Army Biometrics Task Force
CBEFF	Common Biometric Exchange Format Framework
CIO	Chief Information Officer
CJIS	Criminal Justice Information Services Division (FBI)
COTS	Cost Off-the-Shelf
CTS	Conformance Test Suite
DHS	Department of Homeland Security
DISR	DoD Information Technology Standards Registry
DoD	Department of Defense
DoJ	Department of Justice
DoS	Department of State
DoT	Department of Transportation
EBTS	Electronic Biometric Transmission Specification
EFTS	Electronic Fingerprint Transmission Specification
FBI	Federal Bureau of Investigation
FDIS	Final Draft International Standard
FICC	Federal Identity Credentialing Committee
FIPS	Federal Information Processing Standard
FISMA	Federal Information Security Management Act
FY	Fiscal Year
GOTS	Government Off-the-Shelf
GSA	General Services Administration
HSPD	Homeland Security Presidential Directive
IAFIS	Integrated Automated Fingerprint Identification System
ICAO	International Civil Aviation Organization
ICT	Information and Communications Technologies

Acronym / Abbreviation	Definition
IEC	International Electrotechnical Commission
IETF	Internet Engineering Task Force
INCITS	International Committee for Information Technology Standards
IOE	INCITS Organizational Entity
IPMSCG	Identity Protection and Management Senior Coordinating Group
ISO	International Organization for Standardization
IT	Information Technology
ITL	Information Technology Laboratory
ITU-T	International Telecommunication Union Telecommunication Standardization Sector
JTC	Joint Technical Committee
NBSP	National Biometric Security Project
NIJ	National Institute of Justice
NIST	National Institute of Standards and Technology
NISTIR	NIST Interagency Report
NSTC	National Science and Technology Council
NTTAA	National Technology Transfer and Advancement Act
NTWG	New Technologies Working Group
OASIS	Organization for the Advancement of Structured Information Standards
PKI	Public Key Infrastructure
QUAHOG	Ad-Hoc Group on Data Quality
SC	Subcommittee
SDO	Standards-developing organization
TAG	Technical Advisory Group
TBD	To be determined
TBF	The Biometric Foundation
TC	Technical Committee
TF	Task Force
TSA	Transportation Security Administration
USG	U.S. government
US-VISIT	U.S. Visitor and Immigrant Status Indicator Technology
W3C	World Wide Web Consortium
WD	Working draft
WG	Working group
XCBF	XML Common Biometric Format
XML	eXtensible Markup Language

Annex C - Glossary

Acceptance Testing: The process of determining whether an implementation satisfies acceptance criteria and enables the user to determine whether or not to accept the implementation. This includes the planning and execution of several kinds of tests (e.g., functionality, quality, and speed performance testing) that demonstrate that the implementation satisfies the user requirements. *[ISO/IEC 15444-4]*

Accreditation: Procedure by which an authoritative body gives formal recognition that a body or person is competent to carry out specific tasks. *[ISO/IEC - Guide 2]*

Assertion:

a) The specification (description) for testing a conformance requirement. These are specific class of conditions that can be tested. *[NIST]*

b) The specification for testing a conformance requirement in an Implementation Under Test (IUT) in the form defined in [this] standard. *[ISO/IEC 9646-1]*

Certification: Procedure by which a third party gives written assurance that a product, process, or service conforms to specified requirements. *[ISO/IEC - Guide 2]*

Conformance Testing (or conformity testing):

a) Captures the technical description of a specification and measures whether an implementation faithfully implements the specification. *[NIST]*

b) Conformity evaluation by means of testing. *[ISO/IEC - Guide 2]*

Conformity: Fulfilment by a product, process or service of specified requirements. *[ISO/IEC - Guide 2]*

Conformity Evaluation: Systematic examination of the extent to which a product, process or service fulfills specified requirements. *[ISO/IEC - Guide 2]*

Interoperability Testing: The testing of one implementation (product, system) with another to establish that they can work together properly. *[NISTIR 6025]*

Means of Testing: Hardware and/or software, and the procedures for its use, including the executable test suite itself, used to carry out the testing required. *[ISO/IEC 9646-1]*

Performance Testing: Measures the performance characteristics of an Implementation Under Test (IUT) such as its throughput, responsiveness, etc., under various conditions. *[ISO/IEC 15444-4]*

Reference Data: In information technology, reference data is any data used as a standard of evaluation for various attributes of performance. *[NISTIR 6025]*

Reference Implementation: Implementation whose attributes and behavior are sufficiently defined by standard(s), tested by certifiable test method(s), and traceable to standard(s) that the implementation may be used for the assessment of a measurement method or the assignment of test method values. *[NISTIR 6025]*

Robustness Testing: The process of determining how well an implementation processes data which contains errors. *[ISO/IEC 15444-4]*

Test: Technical operation that consists of the determination of one or more characteristics of a given product, process or service according to a specified procedure. *[ISO/IEC - Guide 2]*

Test Assertion: A specification for testing a conformance requirement in an IUT in the form of a software or procedural methods that generate the test results (also named test outcomes or test verdicts) used for assessment of the conformance requirement. *[this MI Ad Hoc Group]*

Test Case:

a) A description of the actions (e.g., condition of the test, expected results) required to achieve a specific test purpose or combination of test purposes. *[NIST]*

b) A specification of the actions required to achieve a specific test purpose or combination of test purposes. *[ISO/IEC 9646-1]*

Test Method: Specified technical procedure for performing a test. *[ISO/IEC Guide 2]*

Test Procedure: *[definition to be determined in the future]*

Test Purpose: A prose description of a narrowly defined objective of testing, focusing on a single conformance requirement. *[ISO/IEC 9646-1]*

Test Scenario: *[definition to be determined in the future]*

Testing: Action of carrying out one or more tests. *[ISO/IEC - Guide 2]*