SUMMARY OF NIST STANDARDS FOR BIOMETRIC ACCURACY, TAMPER RESISTANCE, AND INTEROPERABILITY

November 13, 2002

EXECUTIVE SUMMARY

This appendix briefly outlines appropriate standards for biometric accuracy, tamper resistance, and interoperability based on current findings and test results. Due to the time constraints imposed by the Patriot and Enhanced Border Security acts, biometrics to be initially tested and certified by NIST as being highly accurate must conform to certain conditions. First, any biometric to be considered must be an available and established technology. Second, the captured biometric image outputs from the biometric devices must be available to NIST. Finally, large-scale databases of realistic samples must be available for testing.

Large Scale Testing

Biometric accuracy determination requires the use of large-scale databases for testing. Large realistic test samples of face and fingerprint images have been obtained from the State and Justice Departments. The scale of these tests is significantly larger than any tests previously published. At this time, only fingerprints and face recognition biometrics are included in the NIST accuracy certification studies; iris-based technology was not included, due to the lack of an iris database of sufficiently large sample size. All tests conducted by NIST used image-based biometrics, rather than proprietary templates from biometrics system vendors. In order to ensure interoperability among systems, all image data exchange has used the data format as specified in the ANSI/NIST-ITL 1- 2000 standard.

Known Accuracy of Face and Fingerprints

Results of tests performed have shown that fingerprints provide a higher accuracy rate than face imagery for identification but, under controlled conditions, has comparable accuracy for verification. Using realistic INS data, one index fingerprint can provide a 90% probability of verification with a 1% probability of false acceptance for verification on a sample of 6000 fingers. Tests using realistic face data, show that the best commercial facial recognition systems available can attain a 90% probability of verification with a 1% probability of false acceptance for verification on a sample of 3000 faces. Both these results are strongly dependent on the image quality of the biometrics. It is probable that face recognition on high quality images with good illumination control similar to that used in the State Department visa protocol will equal the accuracy possible with low quality fingerprints for verification (one-to-one matching). The initial and all subsequent biometric acquisitions must meet the same image quality standards. Under less constrained outdoor conditions face recognition accuracy falls to 47% for the best system.

Tests on large databases demonstrate that fingerprints have a greater accuracy for identification than faces. For a database size of 10,000 subjects, the rank one identification accuracy of a single finger is 90% while the rank one identification accuracy for a face is 77%. For a database size of 1,000 subjects, the rank one identification accuracy of a single finger is 93% while the rank one identification accuracy for a face is 83%. "Rank one identification accuracy" means that the image that has the highest score of any image in the sample is the correct match.

Previous work on fingerprint identity searches by Mitretek Corp. has shown that adequate identification can be obtained using the FBI's IAFIS (Integrated Automated Fingerprint Identification System). In this study, at least four fingers are required to perform identification on a database of 40 million individuals.

Tamper Resistance

For foreign nationals entering the country with travel documents, it will be necessary to provide evidence to uniquely authenticate the source of the document, to ensure that the electronic data has not been altered, and to protect the privacy of the data. To accomplish this, Public Key Infrastructure (PKI) technology can be used. PKI can support the key-enabled digital signature that is the analogue of a written signature. Existing Federal Information Processing Standard (FIPS) are approved standards and should be used for the Digital Signature Algorithm and other required components for the system.

Interoperability Standards

There are several approved standards that should be used for interoperability between systems for both identification and verification functions. The ANSI standard, *Data Format for the Interchange of Fingerprint, Facial, & Scar Mark & Tattoo (SMT) Information* (ANSI/NIST-ITL 1-2000), formats fingerprint data to perform background searches against the FBI or another Automated Fingerprint Identification System's (AFIS) criminal file. The Common Biometric Exchange File Format (CBEFF) accommodates any biometric technology and recognizes the ANSI/NIST fingerprint format. Finally, the *Government Smart Card Interoperability Specification Version 2.0* insures interoperability at the higher interface layers and defines a standard CBEFF-compliant container for biometric data on the card.

Projected Biometric Systems Requirements

Not all subjects can be easily fingerprinted with existing technology under the wide range of conditions expected in the entry/exit system. Tests by NIST using INS data show that for approximately 2% of the fingers in the INS database, the friction ridges are too damaged to be matched with existing technology. In addition, within the intelligence community, facial data is often the only biometric data that has been and is currently

being captured. Face data is one key source for "watch lists", and in many situations fingerprint data cannot even be captured to use in constructing a "watch list". NIST measurements indicate that a dual biometric system including two fingerprint images and a face image is needed to meet projected system requirements.

NIST STANDARDS FOR BIOMETRIC ACCURACY, TAMPER RESISTANCE, AND INTEROPERABILITY.

November 13, 2002

BACKGROUND

The USA Patriot Act of 2001 (PL 107-56) and the Enhanced Border Security and Visa Entry Reform Act of 2002 (PL 107-173) were created to provide appropriate tools required to intercept and obstruct terrorism in our country. The laws call for the Attorney General and the Secretary of State jointly, with the National Institute of Standards and Technology (NIST), and in consultation with the Secretary of the Treasury and other federal law enforcement and intelligence agencies to develop and certify technology standards to be used in visa control systems. Taken together, these laws specifically call for:

- Denying visas to those foreign nationals identified as having a criminal record or as being on a "watch list";
- Verifying that a person seeking admission to the US, who is in possession of a legitimate US visa, is the one who was issued the visa or other travel document;
- Establishing document authentication standards for tamper resistant entry and exit documents to the US; and
- Requiring NIST to develop and certify accuracy standards for biometric technologies in support of the identification and verification functions.

These biometric certification standards will be based on the use of fingerprint or other readers and scanners that NIST has determined to be highly accurate and shall be embedded in future immigration control systems. The document authentication standards to be incorporated will take advantage of currently available security standards for digital signatures.

BIOMETRIC TECHNOLOGIES

Due to the time constraints imposed by the Patriot and Enhanced Border Security acts, biometrics to be initially tested and certified by NIST as being highly accurate must conform to certain conditions. First, any biometric to be considered must be an available and established technology so that established testing and certification target dates can be met. New and promising leading-edge biometric technologies that are currently on the horizon can be tested during subsequent testing periods rather than delaying current certification procedures. Secondly, the captured biometric image outputs from the biometric devices must be available to NIST. Finally, biometric accuracy determination

requires large-scale realistic samples for testing. For each biometric tested, large-scale operational databases of at least 100,000 subject images must be available for testing purposes. Without this level of effort, NIST would not have the required confidence to establish biometric accuracy standards and error bounds on these certifications. This is a significant increase in the needed size for the testing database. The largest face database previously used to test face was 1200 faces and the largest fingerprint database had 2700 sets of fingerprints. The largest face test discussed here used 120,000 faces and the largest fingerprint test used 620,000 fingerprints.

Various biometric technologies currently available and suitable for identification or verification functions have been chosen for testing. The identification function is a term used to describe the process of matching a biometric record from a single subject probe against an entire database of similar biometric records in order to determine the identity of the owner of the biometric record. It is a one-to-many comparison. The term verification is used to describe the process of confirming that a person is who he/she claims to be by matching their biometric record against that of their claimed identity. It is a one-to-one comparison. At this time, the only biometrics that are included in the NIST accuracy certification studies are fingerprints and faces. Due to lack of adequate databases, the iris-based technology is not currently under consideration.

Combinations of four to ten rolled (the area of the finger spanning the distance between the two edges of the fingernail) or plain (the fleshy portion of the finger) fingerprint images can be used for enrolling subjects and certifying the identification accuracy rate. One or two plain impressions can be used for establishing subject verification and verification accuracy rates. Still images of faces from the State Department visa program have been used for both verification and identification testing. Using available fingerprint and facial databases, NIST has developed and completed tests used for setting accuracy standards and certifying proposed fingerprint and facial biometric technologies.

OVERVIEW

It is anticipated that fingerprint background checks for foreign nationals seeking a US visa may be performed using the FBI's IAFIS located in West Virginia. The most extensive test of the IAFIS system available was performed by Mitretek as part of the Image Quality Study (IQS). The results of these tests set baseline scores for the tenprint database matching process. Tests were designed and run on an AFIS prototype system to determine the minimum number of rolled or plain images required for a probable identification. For verification purposes, separate system performance criteria have been developed for fingerprints to determine whether a person seeking admission to the US and who is in possession of a legitimate US visa is the one who was issued the visa. These tests were performed on a verification test bed constructed at NIST.

Procedures have been developed to address and evaluate face-based recognition systems. Tests were performed and accuracy statistics gathered for both the identification and verification functions for face recognition systems. NIST, together with DARPA, NIJ, and a number of other Federal agencies, sponsored the Face Recognition Vendor Test 2002 (FRVT2002). This competition compared facial recognition systems

using large-scale facial image databases. Accuracy results from the FRVT 2002 tests can be used the set accuracy standards for identification and verification for travel documents. These results show that facial recognition is a viable technology for verification or “watch lists” but not for large-scale identification.

The first step for any biometric technology is to capture a specific biometric image that will be used. The image is then processed and a set of features or characteristics is extracted. The description of these features can then placed in a specific vendor-defined template used for matching. In the case of fingerprints, this template consists of a set of minutiae specified in terms of location, orientation, and proximity to other minutiae. The size of each template is usually a few hundred bytes as opposed to ten kilobytes used for an image, thus making the storage and transmission of a template much less a problem than that for an image. However, each manufacturer within a particular biometric technology generally has specific vendor-defined information in the template. As a result, most vendors state that their products only work at maximum efficiency if their proprietary definitions and templates are used. Therefore, all the tests that have been conducted by NIST have used image-based biometrics. Fingerprint or face templates were not used as input in any of the tests done by NIST. At the present time no standard template exists for face recognition. All face recognition templates in use are proprietary. Adhering to the exchange of biometric images (rather than templates) enables compliance with the interoperability requirements of the entry-exit system. For testing efforts and for future operational uses, the ANSI/NIST-ITL 1-2000 Data Format for the Interchange of Fingerprint and Facial Information should be used for the exchange of fingerprint and facial images.

FBI IAFIS ACCURACY

A proposed system for visa screening would require the use of the FBI’s IAFIS system for criminal background checking as part of the initial registration process. To do this, it is necessary to perform test on the IAFIS in order to determine its accuracy rate. Since the IAFIS is an operational production system, concurrent testing with production use is impractical. As an alternative, NIST intends to use the Algorithm Test Bed (ATB) to model the IAFIS identification performance. This system was used by Lockheed Martin to design and test the AFIS algorithm and throughput performance of the IAFIS. The ATB uses the same software and hardware matchers as the IAFIS but on a smaller scale. A copy of the ATB was delivered by Lockheed Martin for NIST use on September 16, 2002.

A part of the testing needed to certify the IAFIS for use in visa processing was performed by Mitretek as part of the Image Quality Study (IQS) under contract to the INS. This study concludes that adequate identification performance can be obtained using the FBI’s present IAFIS algorithms. When using the IAFIS ten-print algorithm suite, this accuracy level requires a minimum of four (preferably six or more) flat fingerprints, given the size of the present criminal database, 45 million individuals. Using less than four fingers will result in an unacceptable reduction in identification accuracy.

The IQS study determined that the factors that control flat to rolled fingerprint matching were:

- Number of Fingers
- Correspondence between Search and File images:
    - Overlapping areas
    - Lack of mutual distortion
- Quality of ***both*** Search and File images:
    - Quality of ridge detail
    - Number of features
    - Size of image

The quality of the fingerprints used was critical, particularly if either the search or file print was of poor quality. The quality of approximately 2-5% of the INS flat prints is so poor that it renders them virtually impossible to match using current AFIS technology. They are even difficult for human verification. Current IAFIS operations have a reject rate due to poor image quality of 0.5% for criminal search data and about 2.5% for civil search data. The number of searches that cannot be matched due to poor quality can be reduced by using more fingers, and/or by improving the quality of the capture process.

This study concluded that the IAFIS ten-print algorithm suite cannot meet the IDENT/IAFIS reliability and selectivity requirements for a two-finger search. Using four or more (preferably at least six) fingers with the IAFIS ten-print algorithm suite is likely to produce results at the desired performance level, but would require improvements in IAFIS capacity and workflow management. The use of more fingers not only increases system accuracy, but dramatically reduces the size and cost of the necessary hardware. Each additional pair of fingers used for searching reduces the cost by approximately 50%.

The IQS also showed that female fingerprints are more difficult to match than male fingerprints. On average, matching female fingerprints will require about 150% of the processing needed to match male fingerprints. A greater proportion of female fingerprints are poor or very poor quality. Performance and throughput may be engineering challenges for systems with large female populations.

IRIS RECOGNITION

Several small tests involving several hundred subjects have been performed on iris recognition. The results of these test do not agree on either the false positive rate or the failure to register rate and the results of these tests are not easily scalable to the sample size which is used in this report for face and fingerprints. After extensive discussions with the iris recognition vendor we have concluded that the non-match distribution for iris is very well characterized. The match distribution is an unknown function of image quality and failure to enroll has not been investigated adequately. Without knowledge of the match distribution, no estimate of detection rate verses false positive rate can be made.

NIST concludes that iris is a promising candidate for future testing. The test data available at this time suggests that iris may have accuracy comparable to fingerprints. The uniqueness of iris data for large samples has not been tested. This would require the collection of large test databases in an operational environment.

DATASETS

The tests developed and performed by NIST, including the FRVT 2002, require data, software/hardware algorithms, and equipment. Table 1, "Fingerprint Data Sets," summarizes the fingerprint databases currently available or those being acquired by NIST. Rolled and plain fingerprint images will be used to fulfill fingerprint identification and verification testing functions. These databases have originated from a variety of sources. The NIST Special Databases (SD) are publicly available from NIST. Other datasets from INS, Texas, and the FBI were provided for this effort on a restricted use basis and are therefore not available at this time for public distribution.

Table 1 - Fingerprint Data Sets

NAME	SCAN TYPE	PLAIN	ROLL	TESTS	SIZE	QUALITY
SD 14 (V2)	Ink/live		10	Roll:Roll	2,700 Card Pairs	Medium
SD 24	Live (DFR-90)	10		Plain:Plain	80 Fingers	Good
SD 29	Ink	10	10	Roll:Roll Plain:Plain Plain:Roll	216 Card Pairs	Medium
INS INDEX	Live	Index		Plain:Plain	620K Subjects 3M Images	Operational
STATE INDEX	Live	Index		Plain:Plain	2M Images	Operational
INS CARD	Live	10	10	Plain:Roll	100K Cards	Operational
TX	Ink/live	10	10	Plain:Roll	1M Cards	Operational
IAFIS	Ink/Live		10	Roll:Roll Plain:Roll	1.2M Cards	Operational
ESD	Live	10	10	Plain:Roll	3K Cards	Good

SD 14(V2), SD 24, and SD 29 are Special Databases produced by NIST and available to the public for testing and research purposes. SD 14, originally issued in 1993, used a non-certified version of the Wavelet Scalar Quantization (WSQ) algorithm to compress the rolled fingerprint images. Version 2 of SD 14 contains the same 2,700 card pair fingerprint images but are compressed using a NIST-developed and certifiable version of the WSQ algorithm.

SD 24 contains many plain images from 80 fingers captured with an Identicator DFR-90 live-scan reader. The images were all captured in a NIST laboratory under a controlled

environment. As the subjects were all very cooperative this database should contain some of the best images available.

SD 29 contains both the ten rolled and ten segmented plain images from each tenprint card of the 216 tenprint card pairs. SD 14 and SD 29 shall be used for determining baseline scores for rolled tenprint matches.

The INS INDEX (Immigration and Naturalization Service) recidivist fingerprint database contains over three million finger images. For each of 620,000 unique subjects, there are two or more samples from each index finger taken at different times. For one hundred of these recidivists there are 30 samples of each index fingerprint available.

The INS is also supplying NIST with a CARD dataset of 100,000 tenprint cards consisting of the ten rolled images and associated plain impressions. Among other applications, it will be used to study the scores resulting from comparing plain impressions to rolled.

The TX database comes from the Texas Department of Public Safety. It is a mix of scanned inked cards and live-scan images containing the rolled images, and the thumb and four-plain images from both hands of each subject. The plain impressions can be segmented into ten separate impressions and used for testing and obtaining results of plain versus rolled comparisons. At this time we have over 560,000 samples of this database in-house.

The FBI supplied the Escort Services Desk (ESD) database consisting of rolled and plain images captured by an Identix TP-600 live scan system. The plain images will be segmented into ten individual images. Like the TX, INS, and SD 29 databases, the ESD database may be used for testing plain against rolled images.

Table 2, “Face Data Sets,” summarizes the face databases currently available or those being acquired by NIST. These databases have been acquired from the INS, the State Department, DARPA, and the Army Research Lab. The face images can be used for facial image identification and verification testing

Table 2 - Facial Data Sets

NAME	IMAGE TYPE	VIEWS	SIZE	QUALITY
INS FACE	JPEG	2	620K Subjects 1.25M Images	Operational
STATE	JPEG	1 or 2	6.3M Images 388K Pairs	Operational
HUMANID	JPEG	20	859 Subjects	Controlled
FERET	TIFF	12	1204	Controlled

The INS FACE database contains the "mugshots" for the great majority of the 620,000-subject INS INDEX fingerprint database described above. There are two facial images for each of 620,000 subjects in the index database. A cross-reference between the fingerprints and facial images is also available

The State Department database consists of 6.3 million frontal images obtained from visa applications from approximately 6.1 million subjects. At least one facial image accompanied each application. For 388,000 subjects there were at least two images present. All of the images were obtained from the consular offices in Mexico. In the FRVT 2002 test, 120,000 of these image pairs were included.

The HUMANID facial database was created during the DARPA Identification at a Distance project. Different facial images from each of the subjects taken at different times are included in this database. It was used as part of the FRVT 2002 test. The test report will provide figures appropriate to assessing performance on tasks that include identification, verification, and watch-list operations.

The FERET face database was collected between 1993 and 1996 by the Army Research Lab (ARL) for DARPA. It consists of 14,051 images of 1,204 subjects gathered at ARL and George Mason University. It is the first large freely available face database to contain variations in lighting, pose, and subject aging. It was used in the FRVT 2000 test .

BASELINE SCORES FOR SINGLE FINGER MATCHES – VTB TESTS

To determine the reliability of systems used for verification purposes, and to be able to verify the identity of the person in possession of a visa, additional fingerprint tests have been performed at NIST that do not require the use of the ATB. Desktop and workstation systems together with RAID units for storing the large data sets are used to run single finger tests and develop performance metrics for verification and single finger identification. Previous research completed at NIST has resulted in the development of operational components for pattern classification, minutiae detection, and matching algorithms based on the mating of minutiae clusters. NIST has integrated computers, storage, and software into a standalone system referred to as the Verification Test Bed (VTB). The original purpose of the VTB was to provide the ability to do single finger processing and matching. However, results from these tests may also be combined or additional tests performed to produce results that can be used for identification purposes.

SD 24 TESTS

SD 24 is a NIST database of plain images acquired from an Identicator DFR-90 live scan reader. Images were captured at a resolution of 500 ppi and digitized to 8 bits of gray. Although the dataset represents only 80 unique fingers, there were four plain image samples taken for each finger from the 30 frames per second capture rate that was available.

Each of the 320 images were compared to every image in the database in order to form a similarity matrix. From this data, ROC curves for the plain-to-plain impressions were constructed. Verification performance for fingerprint recognition is reported on a receiver operator characteristic (ROC) plot. The purpose of a verification system is to simultaneously perform two tasks. The first is to correctly verify the identity of a person when the claim is legitimate. The second is to reject people who are not who they claim to be. Unfortunately, there is a trade-off between these two tasks, and one cannot simultaneously maximize the performance of both tasks. The performance statistic for verifying the identity is the probability of correct verification. This is the probability that a system will verify the identity of a legitimate claim. The performance statistic for rejecting false claims is the false alarm. This is the probability that a false claim will be accepted as being true; i.e., someone fools the system and an unauthorized person is granted access. A ROC measures that trade-off between probability of correct verification and the false alarm rate.

ROC plots of the probability of verification versus the probability of false accepts for each of the five finger positions are shown in Figure 1. The ring and little fingers showed the poorest performance for these verification tests. The thumbs showed the best performance giving a 99% probability of true verification with a 1% probability of a false accept. The index and middle fingers showed a 93% probability of verification with a 1% probability of false accept. Although these are very good performance rates, it should be noted that all of the samples for each finger were acquired within a few minutes. This is in contrast to other databases that acquired samples on the same finger over long periods of time. The sample size for SD24 is too small to generate confidence estimates.

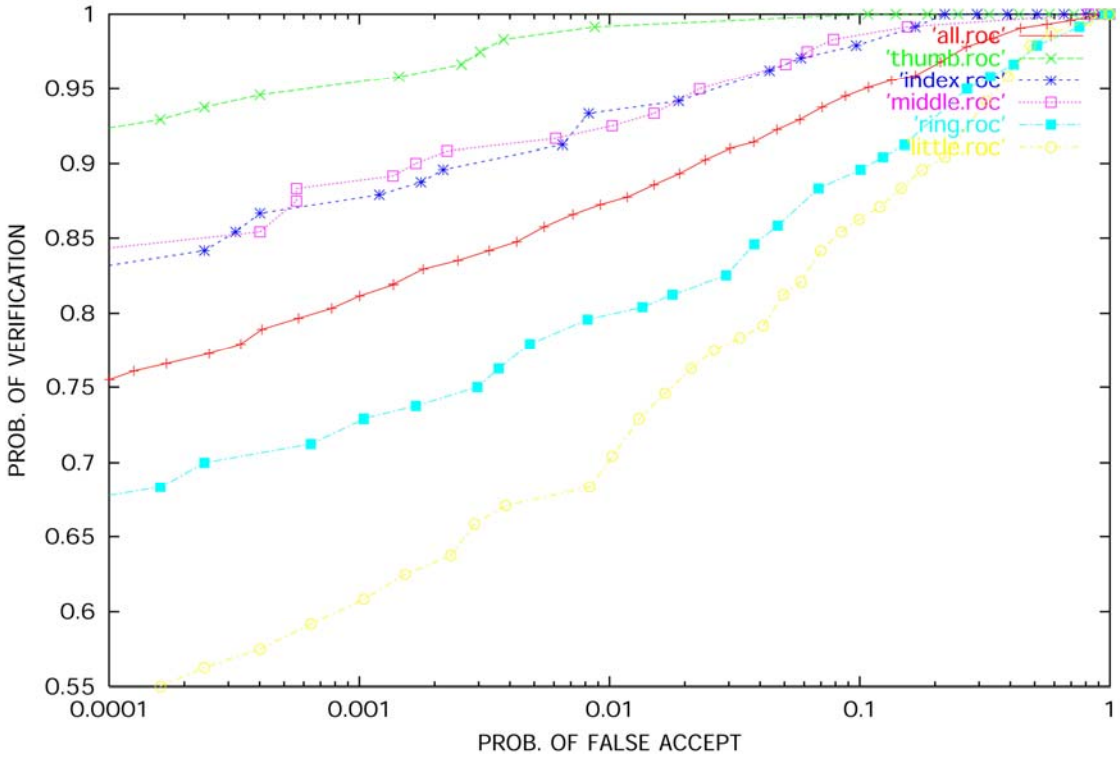

Figure 1. ROC Plots for SD 24 Plain to Plain Impressions

INS LIVE-SCAN INDEX DATASET

This INS dataset represents 620,000 subjects. For each subject, there are at least two samples of each index finger. For many of these subjects there are more than two samples of each index finger. Like SD 24, this dataset is being used to measure performances of plain-to-plain image comparisons. Due to the size of this dataset, it was processed in 6000 finger increments selected across the dataset. For each 6,000 block of images, a similarity matrix was constructed from which the ROC plots were generated. Ten of these blocks or 60,000 images were processed in this manner in the initial experiments using this data. The ROC plot for these samples is shown in figure 2. For a 1% probability of false accept there is a 90% probability of true verification. This verification rate compares favorably with the rate obtained for the controlled laboratory data in SD24. The two standard deviation confidence intervals shown by the red ellipses in figure 2 are indications that the errors associated with false accepts on the horizontal axis are correlated with the errors in verification on the vertical axis. This is shown by the fact that the ellipses in figure 2 are tilted.

Tests used to check for incorrectly labeled fingers have shown that data taken from the first year of the sample provided has 0.1% labeling errors. The labeling error rate then rises to greater than 1.3% in several quarters and in the year 2000 fell to approximately

0.6%. Although the finger labeling error is not a measure of image quality, this is consistent with significant sequential stratification of the data. The results for the 6000 sample blocks shown in figure 2 were taken using the part of the data that has large mislabeling rates.

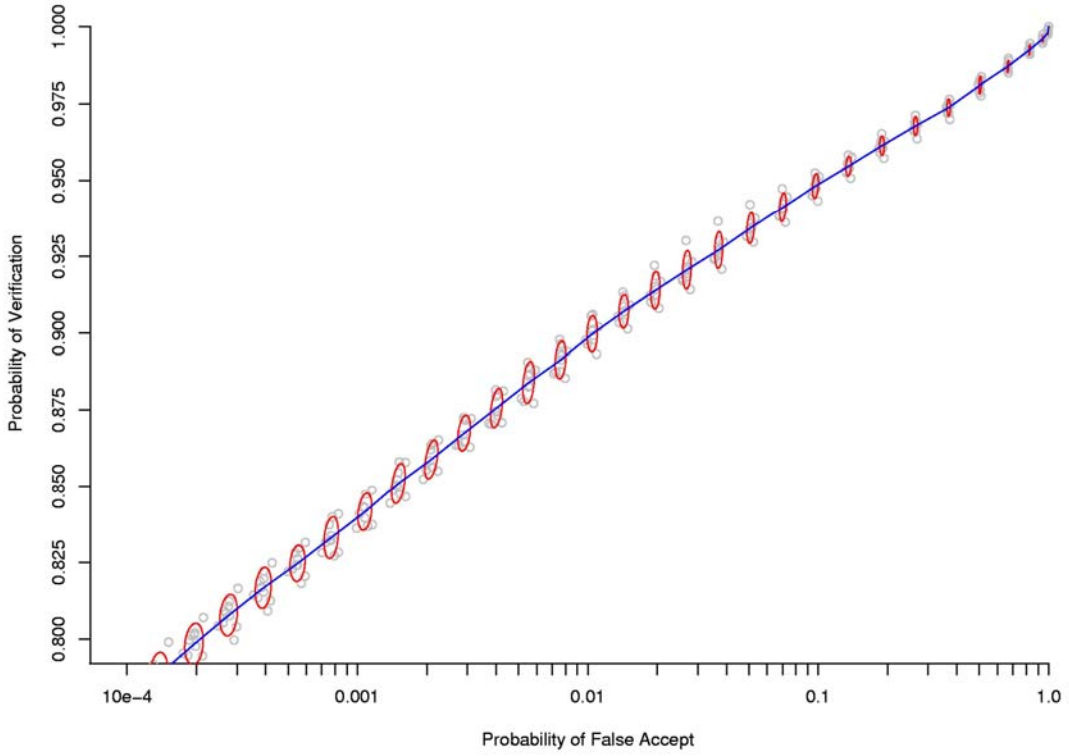

Figure 2. ROC Plots for INS Plain-to-Plain Index Fingers Using Average Quality Data

In addition to verification tests, the same algorithms were also used to evaluate identification performance achieved with the INS index dataset. A probe set of 450 subjects was matched against the entire 620,000 subjects in the INS database. The percent of the subjects correctly identified at rank one was then computed as a function of the gallery size. The results of this test are shown in figure 3. As expected for a gallery size of 500 the identification rate was 95%. The identification rate drops to 90% for a gallery size of 10,000 and to 86% for a gallery size of 100,000. Since the entire 620,000 INS database was used, the image quality was representative of that used for figure 2.

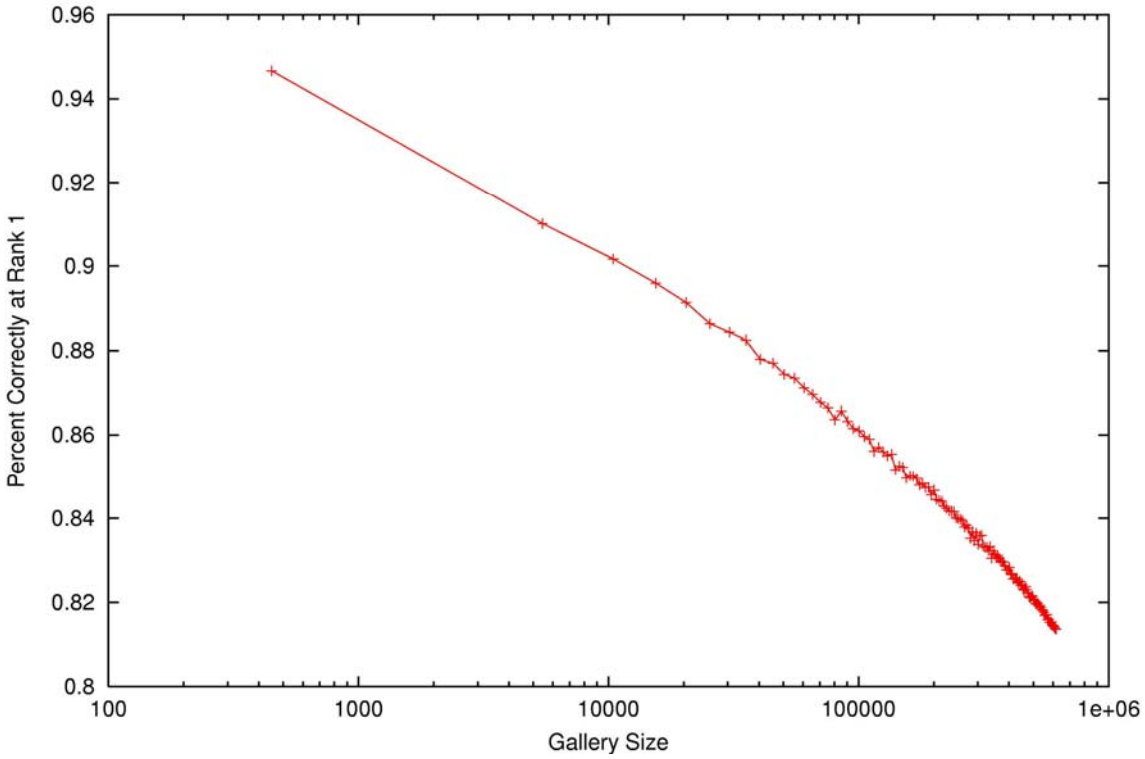

Figure 3. Probability of Detection at Rank 1 as a Function of 619,000 plain Index Fingers

IMAGE QUALITY TESTING

Since image quality has been shown to be critical in the fingerprint tests discussed above, NIST has developed a procedure that can correlate matched performance with image quality. In the INS data, there are exactly 30 samples from each of 100 individuals that are in the database. To evaluate the image repeatability of these fingerprints, each of the 30 fingerprints from all 100 individuals (3000 fingerprints) were matched against each of the 3000 fingerprints. The match scores were placed in a 3000x3000 table resulting in 9,000,000 entries. Of these there were 3000 (100x30) comparisons of the same exact image. There were 87,000 ((30x30x100) – 3000) possible match scores of similar images and 8,910,000 non-match scores.

The 870 matched scores for each individual were averaged and compared with the average matcher scores for each of the 9900 non-match blocks. Of the 100 subject match blocks, 98 had average scores that allowed them to be clearly distinguished from non-match blocks. Two of the match blocks had average match scores that were so low that they were only marginally distinguishable from non-match blocks.

When the 30 images of each of the marginal match blocks were examined, it was found that the friction ridges of these individuals were abraded to the point where no ridges were present. This did not appear to be an equipment problem. The images had sufficient contrast and were not blurred. The fingers did not have detectable or repeatable friction ridge patterns. From this experiment and similar results seen at NIST during the collection of SD24, we conclude that approximately 2% of individuals in the general population may not be easily fingerprinted. This number is in agreement with numbers provided by the FBI and the Mitretek IQS study.

FACE RECOGNITION ACCURACY

The most comprehensive test of commercial face recognition accuracy done to date was FRVT 2002 (Face Recognition Vendor Test 2002). The FRVT 2002 is a technology evaluation that measures the performance of core face recognition technology (For complete details on FRVT 2002 and FRVT 2000 see http://www.frvt.org). The FRVT 2002 consists of two parts: the High Computational Intensity Test (HCInt) and Medium Computational Intensity Test (MCInt). The MCInt reported results on small databases of images, and the High Computational Test reported results from a large database of images from the Department of State's Mexico Visa Application program. All results in this write up are from the HCInt.

The FRVT 2002 was an independent open evaluation organized and executed by the National Institute of Standards and Technology (NIST). FRVT 2002 is independent because it was organized by NIST and all the data in the evaluation was sequestered. The evaluation was announced on 25 April 2002 and was open to all providers of face recognition systems, with ten participants taking the HCInt. All systems were tested under strict Government supervision at the U.S. Naval facility at Dahlgren, VA. All systems were tested on exactly the same images, and therefore the performance among participants can be directly compared. The FRVT 2002 was administered in July and August 2002.

The images in the HCInt are from the Department of State's Mexico Visa Application program. The data set in the HCInt consists of 121,000 images of 37,000 people. There are either three or four images of each person in the data set.

Figure 4 shows the ROC for the top three participants in the FRVT 2002. The three solid blue lines are the ROCs for the participants generated from facial images of 33,000 individuals. It is well known in face recognition that performance varies when different people are in the gallery. For example, in access control, the verification rate for people in one building will be slightly different that for the people in a second. Measuring this variation is important to predicting performance of verification systems at different locations. To estimate this variation, the 33,000 people for the solid blue lines were divided in 11 non-overlapping sets; i.e., each of the 33,000 were only in one of the 11 sub groups. Verification performances were computed for each of the 11 subgroups, and are plotted as dots in Figure 4. The colored ellipses are the standard error for each of the 11 subgroups. For example, for the top participant, at a false alarm rate of 1%, the probability of correct verification for the 33,000 individuals is 89%. The standard error range at this performance point are 1%±1% for the false alarm rate and 89%±1% for

probability of verification. Under less constrained outdoor conditions face recognition accuracy falls for the best system fal1s to 47%.

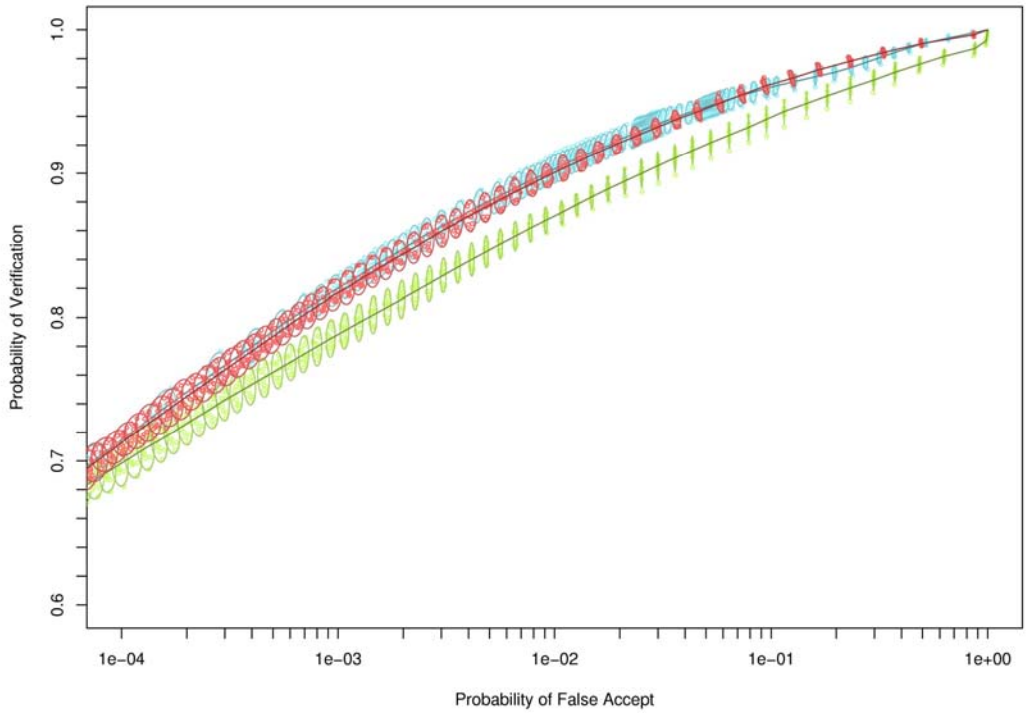

Figure 4. FRVT2002 ROC Curve for Face Verification for Three Most Accurate Vendors

One of the most interesting open questions in face recognition is how does performance change as the size of the database of known individuals increases. The FRVT 2002 provides the first substantial answer to this question. Figure 5 shows how performance changes as the database size increases from 100 to 37,000 individuals for the top three participants. The horizontal axis is the database size on a logarithmic scale, and the vertical access is the probability an unknown facial image of a person is correctly identified. Identification performance for the top three participants for a database of 37,000 is 73%, 69%, and 64%.

The FRVT 2002 shows substantial improvement since the FRVT 2000, where a comparable verification performance score was 80% verification at a 1% false alarm rate. There are no comparable identification performance figures because the largest gallery in FRVT 2000 was 1196 individuals. The most obvious question for identification

performance is what is performance on galleries of greater than a million individuals, and will identification performance continue to follow the log linear rule.

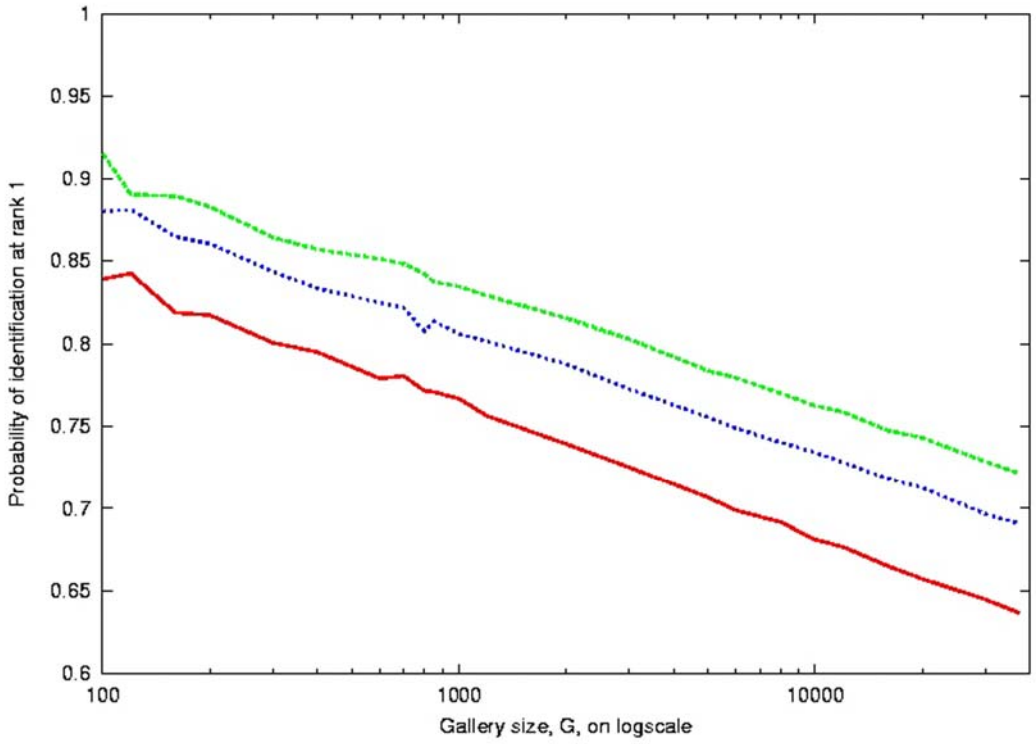

Figure 5. Probability of Identification at rank 1 as a Function of Gallery Size for the Best Three FRVT 2002 Vendors

STANDARDS FOR TAMPER RESISTANCE

In order for many foreign nationals to enter the country, they will be required to possess a U.S. government issued travel document that must be presented to an INS agent at the time of entry. These documents will contain identity information including one or more biometrics to be used for verifying that the person presenting the document was the one that was originally issued the document. In order to verify the presenter's identity, the subject will submit their appropriate biometrics for capture and comparison against those electronically recorded on the travel documents.

It is also necessary to provide evidence that the document was issued legitimately. Additional safeguards and procedures are required to uniquely identify the source of the travel document and to authenticate that it is a legitimate source. It is also necessary to ensure that none of the data recorded on the travel document has been deliberately or inadvertently altered since issuance. Mechanisms must also be in place to insure the

privacy and confidentiality of the data. Finally, it must not be possible for the source of the travel document to repudiate their responsibility for issuance of the document.

Public Key Infrastructure (PKI) can be used to support the key-enabled digital signature security service that will satisfy the above requirements. PKI is a combination of products, services, facilities, policies, and people that can provide and sustain secure travel documents. A digital signature is an electronic analogue of a written signature. The Digital Signature Algorithm (DSA) is used for generating a digital signature. A description of this algorithm can be found in the Federal Information Processing Standard FIPS PUB 186-2 "Digital Signature Standard (DDS)." The DSA is a pair of large numbers computed using a set of rules and a set of parameters to enable the verification of the signer and the integrity of the data. The digital signature is created and verified using two numbers referred to as the private and public key pair. In order to create the digital signature, a hash function is applied to the identity and biometric information to produce a new version of the data call a message digest. This message digest in combination with the user's private key is processed by the DSA to produce the digital signature that can be recorded on the travel document. The digital signature verification process (performed at the entry location to the U.S.) requires the public key of the pair to properly decrypt the digital signature. An issuing authority's private key is never shared unlike the public key that is known to the general public. By use of the public key the original information can be decrypted and verified.

Existing approved FIPS standards can be used for the creation and verification of travel documents. Currently there are three FIPS-approved algorithms for generating and verifying digital signatures: DSA, RSA, and ECDSA. All three are used in conjunction with the Secure Hash Algorithm SHA-1. Additional information and links to the required algorithms for the deployment of digital signatures can be found in the "Cryptographic Toolkit" maintained by NIST at: http://csrc.nist.gov/encryption/tkdigsigs.html.

INTEROPERABILITY STANDARDS

Currently, there are several approved standards that can be used for interoperability between systems for both the identification and verification functions. At the image level the WSQ and JPEG compression standards are applicable to fingerprints and facial images. There is a proven standard for formatting biometric data that had been used for years in the law enforcement community and an additional standard for supporting biometric technologies in a common manner. A specification for "smart card" technology is available to insure interoperability at higher interface layers.

A background check may be needed for some foreign nationals applying for a visa. This could require that those applying for a visa submit a set of their fingerprints for processing on either the current FBI IAFIS system or on a new AFIS system designed specifically for INS background clearance work. The ANSI standard, *Data Format for the Interchange of Fingerprint, Facial, & Scar Mark & Tattoo (SMT) Information* (ANSI/NIST-ITL 1-2000), is being used for this purpose. It was specifically developed for the FBI and other law enforcement and criminal justice agencies to provide a data format for the exchange fingerprint images, facial images, and other related identification data. The standard defines the content, format, and units of measurement for the exchange of

fingerprint and facial information that may be used in the identification process of a subject. Information compiled and formatted in accordance with this standard can be recorded on machine-readable media or transmitted to another criminal justice administration or organization that relies on an automated identification or verification system. Transmission of fingerprint image data between the FBI and each state relies on this standard for packaging and sending the data to the FBI in a form that can be recognizable and processed.

In addition to fingerprint and facial image data, the standard provides for the interchange of textual fields. When conducting a background search of a visa-applicant, the procedures to be followed must include a name search of the subject. This name search should be done in the subject's native language as well as English in order to improve the probability of finding a match if one exists. Although, text is normally expressed as 7-bit ASCII characters, the ANSI/NIST standard allows the use of Unicode to be used for searching databases with different variations and transliterations of the subject's native character set. This provides the ability to search with names normally expressed in a pictographic format. For more information and a downloadable copy of this standard refer to http://www.itl.nist.gov/iad/vip/fing/fing.html .

At the next higher standards level, the Common Biometric Exchange File Format (CBEFF) describes a set of data elements necessary to support biometric technologies in a common way. It promotes interoperability of biometric-based application programs and systems developed by different vendors by allowing biometric data interchange. It also provides forward compatibility for technology improvements, simplifies the software/hardware integration process and describes how new formats can be created. Data fields regarding the biometric data such as the type of biometric is available, the version number, vendor's name, etc. foster interoperability between different types of biometric systems, allows for the exchange of biometric related information between different systems, and enables systems with different requirements to translate between different formats.

CBEFF accommodates any biometric technology. It includes the definition of format and content for data elements such as a biometric data header, the biometric data itself, and any other required biometric data or data structures. The ANSI/NIST fingerprint format is recognized by and is fully compliant with CBEFF. More information or a copy of the document can be found at http://www.itl.nist.gov/div895/isis/bc/cbeff/ .

The BioAPI specification Version 1.1 was developed by the Biometric Consortium and is intended to provide a high-level generic biometric authentication model - one suited for any form of biometric technology. It covers the basic functions of Enrollment, Verification, and Identification, and includes a database interface to allow a biometric service provider (BSP) to manage the identification population for optimum performance. It also provides primitives that allow the application to manage the capture of samples on a client, and the Enrollment, verification and Identification on the server.

Once a background search is completed successfully, a travel document should be issued to the foreign national. This document may contain one or more biometric images belonging to the person. Upon entry to the United States, a similar set of biometric data can be captured from the visitor and compared to the image(s) contained

on the travel document. For this verification function a “smart card” with computational capability will not be required. But this application will require a chip or storage capability of approximately 32KB of storage. This size should adequately handle up to two fingerprint images and one facial image.

Although a “smart card” feature is not necessary, it may prove to be a better alternative providing it meets the data storage requirement. If this is the case, then the NISTIR 6887, "Government Smart Card Interoperability Specification Version 2.0" should be used to insure interoperability at the higher interface layers. Additional information can be found at http://smartcard.nist.gov/. This specification defines a standard CBEFF-compliant container for biometric data on the card, and a mechanism for making applications totally independent of the specifics of the underlying card technology.

SUMMARY

The Patriot Act and the Enhanced Border Security Act call for NIST to develop and certify testing procedures, accuracy standards for biometrics, interoperability standards, and standards for tamper resistance as part of future immigration control systems and travel documents. These standards will be applicable to background checks used for granting visas and for verification checks at the time of entry into the US and at the time of exit.

NIST has determined that face and fingerprints are the only biometrics available with large enough operational databases for testing at this time. Both technologies are mature. To properly certify any biometric, extensive tests must be performed using databases containing at least 100,000 subjects. Such databases have been acquired from NIST, FBI, INS, DOS, and Texas to perform the required testing.

Results from fingerprint testing based on the Mitretek study, and NIST testing using SD 24, and a sampling of INS data have been analyzed. To perform background identifications, ten plain image impressions should be used for enrollment and retention. As described in the “FBI IAFIS Accuracy” section of this report, Mitretek recommends a minimum of four plain finger impressions for background searches. With the live-scan fingerprint scanners currently available, the additional time required to capture the additional six fingers will be insignificant.

Results show fingerprint matching to be accurate. Verification can be performed on single fingers with 90% accuracy at a false accept level of 1%. Single finger identification can provide 95% accuracy for a gallery size of 500. The identification rate drops to 90% for a gallery size of 10,000 and to 86% for a gallery size of 100,000. This test illustrates the difficult nature of accurate database searches using a single fingerprint. High accuracy searching of 1 million subject or greater database will require more than one finger whether the FBI’s IAFIS is used or not.

Results indicate that fingerprints provide approximately the same verification accuracy as face . For facial recognition, the best packages available (based on FRVT 2002) provide a 90% probability of true verification with a 1% probability of false verification. This makes face recognition an excellent choice as an alternative to fingerprints for

verification and for situations where fingerprints are not available and where high quality face images with good illumination control similar to those taken using the State Department visa protocol are available. Under less constrained outdoor conditions face recognition accuracy for the best system falls to 47%. For identification the best available face recognition technology identification can provide 90% accuracy for a gallery size of 100. The identification rate drops to 83% for a gallery size of 1,000 and to 77% for a gallery size of 10,000. These numbers demonstrate that for identification, fingerprints are the preferred technology. However, not all subjects can be easily fingerprinted with existing technology resulting in a 2% failure to acquire rate. Furthermore, within the intelligence community, facial data is often the only biometric data that has been and is currently being captured. Based on these considerations, our measurements indicate that a dual biometric system including two fingerprint images and a face image may be needed to meet projected system requirements for verification. Each fingerprint and the facial image should require 10 kilobytes or less of storage apiece. Therefore, a card capable of storing two fingerprints and one face image will require a 32K byte chip to fulfill these requirements.