NISTIR 7204

NIST Biometrics Evaluations and Developments

Michael D. Garris
Charles L. Wilson

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

NISTIR 7204

NIST Biometrics Evaluations and Developments

Michael D. Garris
Charles L. Wilson
Image Group
Information Technology Laboratory

February 2005

U.S. DEPARTMENT OF COMMERCE
Carlos M. Gutierrez, Secretary
TECHNOLOGY ADMINISTRATION
Phillip J. Bond, Under Secretary of Commerce for Technology
NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
Hratch G. Semerjian, Acting Director

NIST Biometric Evaluations and Developments

Michael D. Garris, Charles L. Wilson
National Institute of Standards and Technology, 100 Bureau Drive, Stop 8940, Gaithersburg, MD, USA 20899-8940

ABSTRACT

This paper presents an R&D framework used by the National Institute of Standards and Technology (NIST) for biometric technology testing and evaluation. The focus of this paper is on fingerprint-based verification and identification. Since 9-11 the NIST Image Group has been mandated by Congress to run a program for biometric technology assessment and biometric systems certification. Four essential areas of activity are discussed: 1.) developing test datasets, 2.) conducting performance assessment; 3.) technology development; and 4.) standards participation. A description of activities and accomplishments are provided for each of these areas. In the process, methods of performance testing are described and results from specific biometric technology evaluations are presented. This framework is anticipated to have broad applicability to other technology and application domains.

Keywords: biometric, consolidation, evaluation, fingerprints, identification, performance testing, standards, verification

1. INTRODUCTION

NIST has a long history of involvement in biometric research and biometric standards development. For over 30 years, NIST has collaborated with the Federal Bureau of Investigation (FBI) in the area of automated fingerprint recognition. Researchers at NIST (then the National Bureau of Standards (NBS)) began work on the first version of the FBI's Automated Fingerprint Identification System (AFIS) system back in the late 1960's. Over the years, NIST has conducted fingerprint research, developed fingerprint identification technology and data exchange standards, developed methods for measuring the quality and performance of fingerprint scanners and imaging systems, and produced databases containing fingerprint images for public distribution[1-20].

NIST has also conducted biometric evaluations in the area of face recognition. The Image Group has run a series of large scale face recognition system tests, the last of which was called FRVT2002[21]. More recently, it has conducted a large scale fingerprint vendor technology evaluation called FpVTE2003[22]. Conducting these biometric technology evaluations requires the collection and publication of large volumes of data as well as the development of scoring methods and technology for the computation of performance statistics. As a result, NIST has significant experience and expertise in managing and analyzing large repositories of biometric data, and it has conducted many large-scale biometric technology evaluations.

Based on this experience, it was not too surprising that Congress included NIST in its legislative response to the terrorist attacks on September 11, 2001 (9-11). Several pieces of congressional legislation were created that directly cited participation and contribution from NIST in the area of biometric standards development. These include the USA PATRIOT Act and the Enhanced Border Security and Visa Entry Reform Act. Both of these acts specify requirements for interoperable biometric systems that are being developed by the Department of Homeland Security (DHS) and the Department of State (DOS). Specifically, NIST was tasked to "develop and certify a technology standard ... that can be used to verify the identity of persons applying for a United States visa or such persons seeking to enter the United States." Within a couple of months, new initiatives were started that redirected work that had been focused on law enforcement to new work focused on border control.

To support this effort, the Image Group at NIST leveraged its experience in evaluating biometric technologies and began efforts to certify specific biometric components of the DHS US-VISIT program. This involved significant investment of time and energy in four key areas of activity: 1.) developing test datasets, 2.) conducting performance assessment; 3.) technology development; and 4.) standards participation. Together, these four activities form a general R&D framework for technology evaluation.

While the work presented in this paper has been applied to biometrically managing foreign travelers through our land, air and sea ports (a.k.a. border control), the framework deployed for biometric evaluation has much broader application. It is anticipated that similar technology evaluations will be needed to "develop and certify technology standards" for port and harbor security as well. To this end, this paper provides an overview of the work conducted at NIST. Sec. 2 discusses developing test datasets; Sec. 3 presents methods for performance assessment and presents some results; Sec. 4 describes the need to develop technology to support and bring credibility to the evaluations; Sec. 5 provides an overview of standards participation; with summary remarks provided in Sec. 6.

2. DEVELOPING TEST DATASETS

Perhaps the single most critical resource needed to successfully evaluate biometric systems is *data*. This is true for any pattern recognition application. A common challenge for researchers is that there is too little data available to adequately develop and test a new algorithm beyond a "toy" scenario. Fortunately, this has never been quite the situation within the Image Group at NIST. With years of FBI collaboration, NIST has acquired and distributes the largest publicly available collection of federal law enforcement fingerprint images[7,10,11,12,16,19,20]. With other federal agencies receiving new mandates and initiatives as of 9-11, NIST has considerably added to its fingerprint image repository, including operational data from federal agencies, state & county jurisdictions, and Department of Defense (DOD) applications. Nearly all this new data is considered "sensitive but unclassified," so it is not available to the general public.

These different datasets are represented in Tab. 1. As a result, NIST hosts and maintains one of the largest multi-jurisdictional fingerprint repositories in the world with approximately 16 million subjects and 95 million fingerprints.

Table 1: Multi-jurisdictional repository of fingerprint datasets at NIST

NAME	SCAN TYPE	PLAIN	ROLL	TESTS	SIZE	QUALITY
SD 14 (V2)	Ink w/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 (DFR-90)	Index		Plain:Plain	620×10^3 Subjects 3×10^6 Images	Operational
INS Benefits	96% Live; 4% Rescan	10	10	Roll:Roll, Plain:Plain, Plain:Roll	640×10^3 Subjects	Operational
DOS-BCC	Live (DFR-90)	Index		Plain:Plain	6×10^6 Subjects 240×10^3 Matched	Operational Office
INS CARD	Ink	10	10	Plain:Roll	100×10^3 Cards	Operational
TX	60 % Ink; 40 % Live	10	10	Plain:Roll	1×10^6 Cards	Operational
IAFIS	Ink w/Live		10	Roll:Roll, Plain:Roll	1.2×10^6 Cards	Operational
ESD	Live	10	10	Plain:Roll	3×10^3 Cards	Good
US-VISIT Jan. 04-Feb.04	Live	Index		Plain:Plain	1.7×10^6 Subjects 34×10^3 Matched	Good
US-VISIT Mar. 04-Jun. 04	Live	Index		Plain:Plain	3.7×10^6 Subjects	Good
LA County	90% Live; 10% Rescan	10	10	Roll:Roll, Plain:Plain, Plain:Roll	1.5×10^6 Subjects 100×10^3 Matched	Good
AZ	60% Ink; 40 % Live	10	10	Plain:Roll	179×10^3 Subjects 58×10^3 Matched	Operational
FBI 12K	Ink w/Live	10	10	Plain:Roll	12×10^3 Subjects	Operational
OHIO	Ink w/Live	10	10	Roll:Roll, Plain:Plain, Plain:Roll	925 Subjects	Very Good
IDENT IAFIS	Live	10&Index	10	Plain:Plain, Plain:Roll	3.8×10^3 Subjects	Operational

Beyond sheer quantity, it is important to have some understanding of how the type and quality of biometric data changes between data sources. There are several significant factors that apply to fingerprints. These include capture type: were the images of fingerprints generated by scanning paper cards of inked fingerprints, or were they generated using a live scan device? There is impression type: are the fingerprints rolled nail-to-nail, or are they a plain (flat) impression? Fig. 1 shows a rolled fingerprint captured from paper, while Fig. 2 shows a flat impression from the same finger. One final attribute that should be considered is fingerprint image quality. This has been the focus of significant R&D at NIST[23], and is discussed further in Sec. 4.2.1.

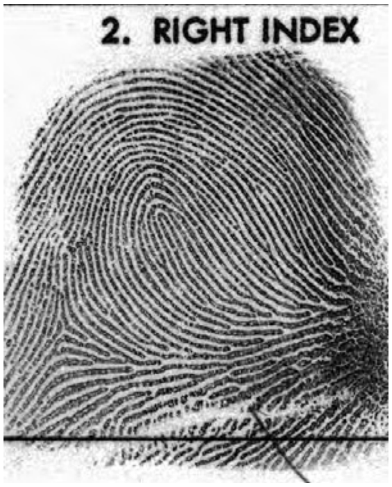

Figure 1: Rolled impression

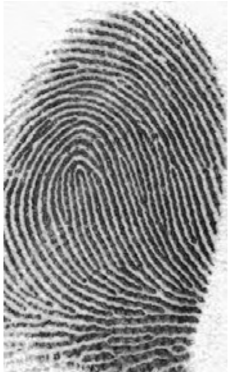

Figure 2: Plain (or flat) impression

3. CONDUCTING PERFORMANCE ASSESSMENT

The datasets described above are carefully sampled and utilized by NIST to test fingerprint matching algorithms and systems. These experiments are conducted and results are reported based on the elemental requirement that a biometric system reports a similarity score when two biometric templates are matched to each other. In general, the higher the score, the more likely the two templates come from the same person.

3.1. Terminology and definitions

A fingerprint matching test is typically comprised of two general sets of fingerprints. There is the set of fingerprints comprised from all those subjects enrolled in the biometric system, called the *gallery* set. The identities of all those in the gallery set are *known* at the time of search. The second set of fingerprints represents the users of the biometric system, called the *probe* set. The identities in the probe set are *unknown* at the time of search. As comparisons are computed between fingerprints in the probe and gallery sets, matcher scores are stored in a *similarity matrix* where the *ij*-th element in the matrix corresponds to the similarity between the *i*-th fingerprint of the gallery set compared to the *j*-th fingerprint of the probe set. Once a similarity matrix is populated with matcher scores, performance statistics are computed.

The scores in the similarity matrix fall into two general categories. A score computed between a probe and gallery belonging to the *same* person is referred to as a *match*, while a score computed between a probe and gallery belonging to *different* persons is referred to as a *non-match*. (Note that the terms, match and non-match, are being used here to characterize whether the probe and gallery fingerprints are from the same person, and not whether the matcher actually achieved a correct identification, which is also often referred to as a correct match or hit.) Significant insights into the performance of a fingerprint system may be gained by analyzing and comparing the distribution of match scores to the distribution of non-match scores such as those shown in Fig. 3.

The distribution of match scores in this figure is represented by the darker curve labeled with ‘+’ and nearly spans the entire horizontal range of the graph. A sample distribution on non-match scores is represented by the lighter curve labeled with ‘×’ and occupies the leftmost portion of the graph. One observation is that non-match scores are relatively and consistently low, whereas match scores may vary widely. The most interesting portion of the graph to focus on is where the right portion of the non-match score distribution overlaps with the left portion of the match score distribution. It is within this region of overlap that system errors occur. In this example, there are 24 cases where match scores are less than or equal to a score of 16 and overlap with non-match scores.

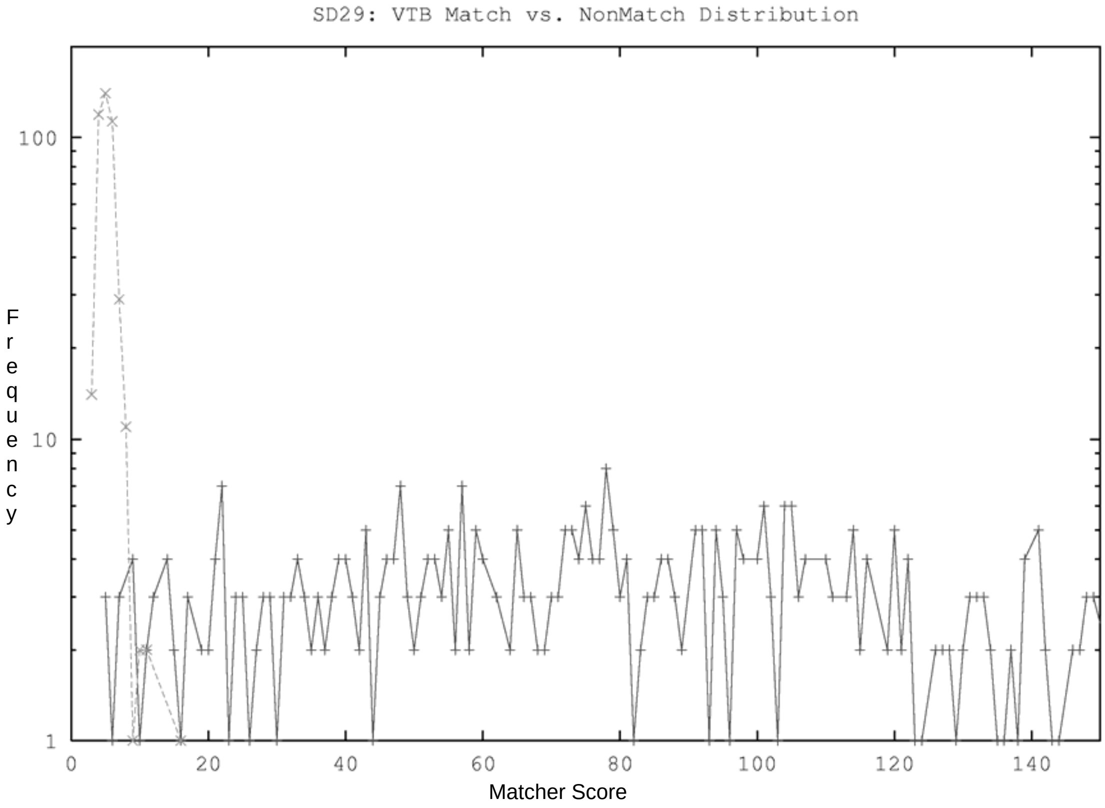

Figure 3: Match and non-match score distributions

3.2. Categories of biometric application

Biometric applications are typically categorized into two general types: verification and identification. 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. *Identification* is a term used to describe the process of matching a biometric record from a single unknown person against an entire repository of similar biometric records in order to determine the identity of the owner of the biometric record. It is a one-to-many comparison.

3.2.1. Verification (1-to-1) testing

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 or *true accept rate (TAR)*. This is the probability that a system will verify the identity of a legitimate claim. The performance statistic for rejecting false claims is referred to as the *false accept rate (FAR)*. This is the probability that that a false claim will be accepted as being true; i.e., someone fools the system and an unauthorized person is granted access.

A Receiver Operator Characteristic (ROC) analysis measures the trade off of TAR and FAR. A threshold is swept across the range of scores such as those in Fig. 3. At each step, the percentage of match scores above the threshold is recorded as TAR, and the percentage of non-match scores above the threshold is recorded as FAR. Plotting these (TAR, FAR) points creates an ROC curve like the ones shown in Fig. 4. This serves as a primary measurement of verification performance.

The score distributions in Fig. 3 contributed to the top curve in Fig. 4. In this graph, the matching performance is compared between: 1.) rolled probe prints matched to rolled gallery prints, 2.) plain probe prints matched to plain gallery prints, and 3.) plain probe prints matched to rolled gallery prints. Comparing the curves, it is observed that matching rolled to rolled prints achieves significantly higher performance than either of the other two combinations, and when matching with plain impressions, there isn't much discernable difference between matching to plain or to rolled gallery prints.

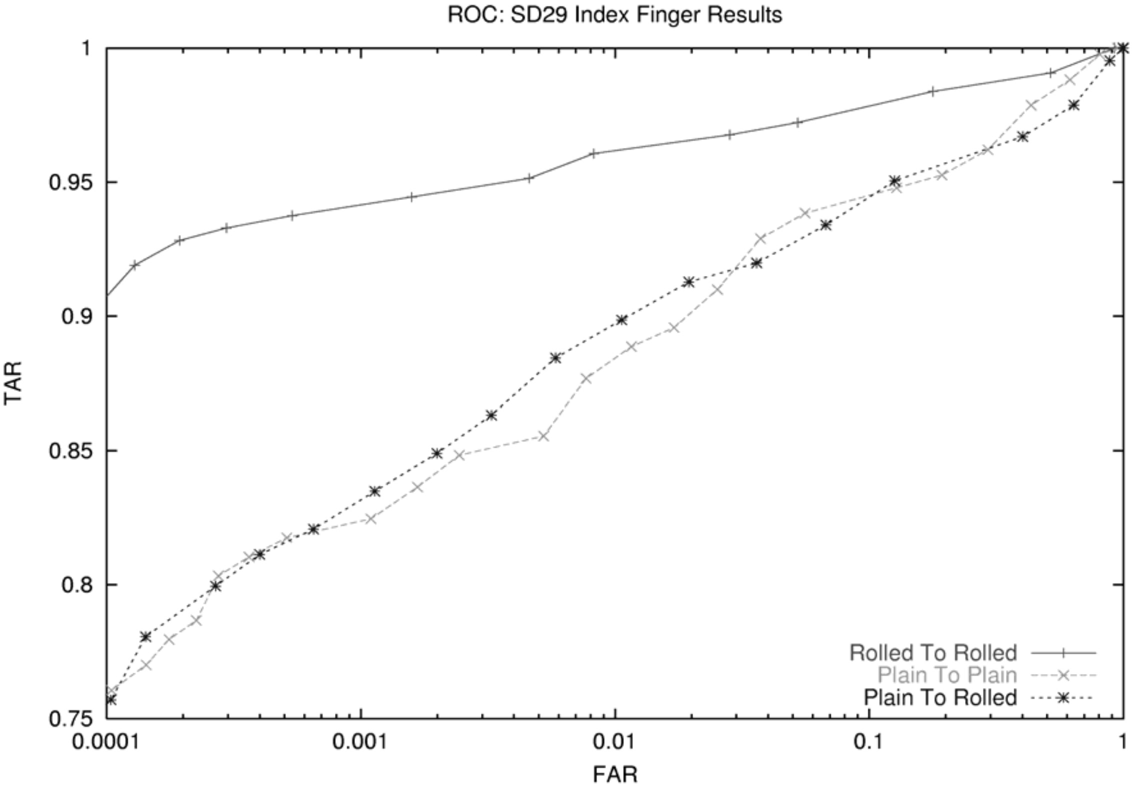

Figure 4: Comparison of verification ROC curves

If one computes performance statistics on an overly small sample of fingerprints, results will be quite unreliable. This instability is observed as significant variation in performance metrics when subsequent independent samples of the same size are computed and compared. As the size of the sample increases, the variation observed between independent trials becomes more stable. NIST has developed a method for determining when a verification study is of sufficient size to produce reliable results. Pairs of mated fingerprints are matched, but rather than create one large similarity matrix, the probe and gallery sets are partitioned into smaller blocks and a similarity matrix is computed for each block.

For example, one study conducted took a random set of 60×10^3 people and subdivided it into ten independent sets of 6×10^3 people, and the corresponding probe and gallery images were compared and resulting matcher scores were compiled into ten 6×10^3 by 6×10^3 similarity matrices. To look at the variation in performance, one could simply plot and visually compare the ten ROC curves, each corresponding to one of the ten similarity matrices. Fig. 5 plots a more sophisticated and useful analysis, called a *Multi-Trial* ROC. The solid curve in the graph plots the *mean* of the ten ROC curves. The small clusters of points along the curve contain synchronized values extracted from each of the ten ROC curves. The spread of the points within these clusters represent the variation in performance between each of the ten random trials. The ellipse overlaying each cluster represents a statistically standardized amount of variance across the trials. The radius of each ellipse is (2×Standard Error), measured from the points in the cluster along both the x-axis and the y-axis.

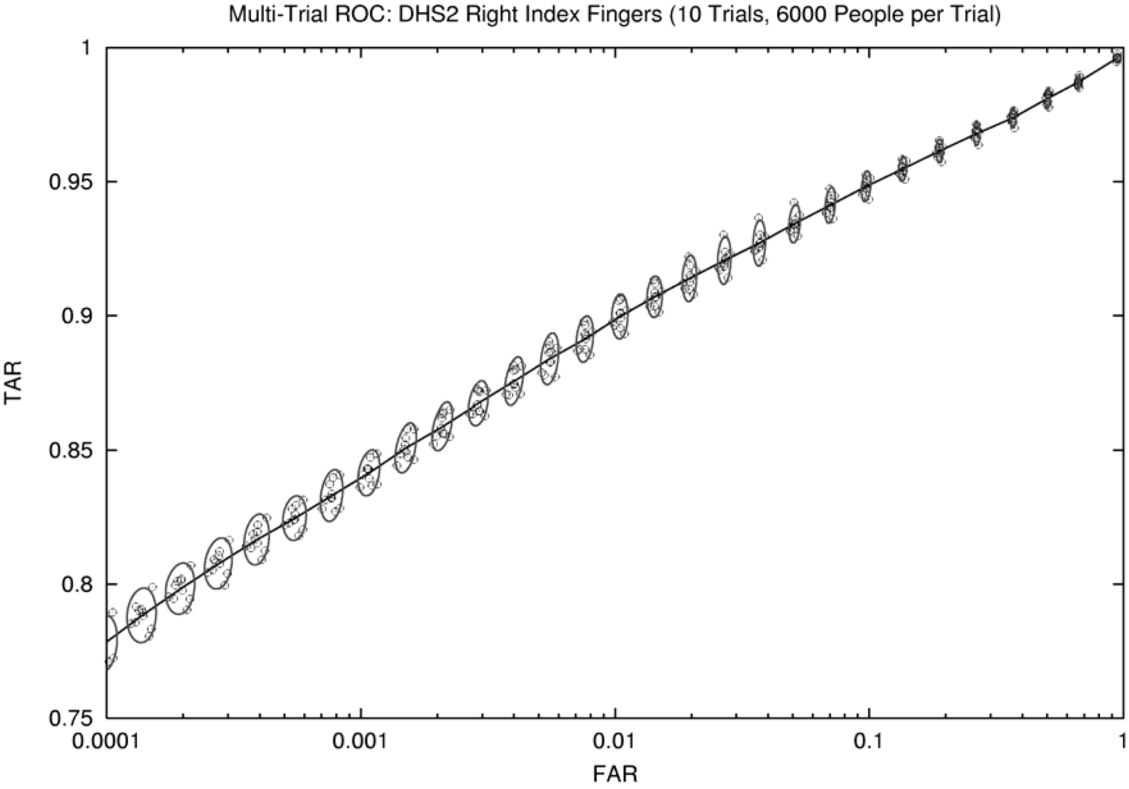

Figure 5: Multi-trial ROC curve

3.2.2. Identification (1-to-many) testing

Identification performance is measured by determining the ability of a biometric system to accurately identify an individual in a large database, given a single unknown biometric record. Again, two sets of data (a probe set and a gallery set) are used to test identification performance. The probes represent the unknown templates for which identity is sought. The gallery represents all subjects *enrolled* in the system. It is generally the case that for identification applications the size of the gallery far exceeds the size of the probe set. To compute identification performance, each probe must have a second mated template to be *seeded* into the gallery. The gallery is padded out with biometric templates (called the *background*) known to not be mates of the probes.

To know if a system correctly identifies a probe template, it is necessary to know beforehand that there is in fact one and only one mate in the gallery set. This way, if the biometric system reports the mate as a candidate match, then a true accept is tallied. A false accept rate is computed by excluding the mates of the probes from the gallery, matching the probes to the *unseeded* gallery, and then tallying how many times the system reports candidate matches. By sampling across a range of matcher score thresholds, an ROC plot can be generated charting the (TAR, FAR) response. Fig. 6 demonstrates how increased TAR is traded off for increased FAR by choosing different operational thresholds.

There is an expected correlation between FAR and the size of the gallery. As the size of the gallery used for identification increases, the probability increases that one of the templates in the gallery will produce a confusable match with one of the probes and therefore generate a false accept. To measure this relationship, a curve like the ones in Fig. 7 can be computed. It should be noted that with typical identification systems, TAR remains constant as the gallery size increases.

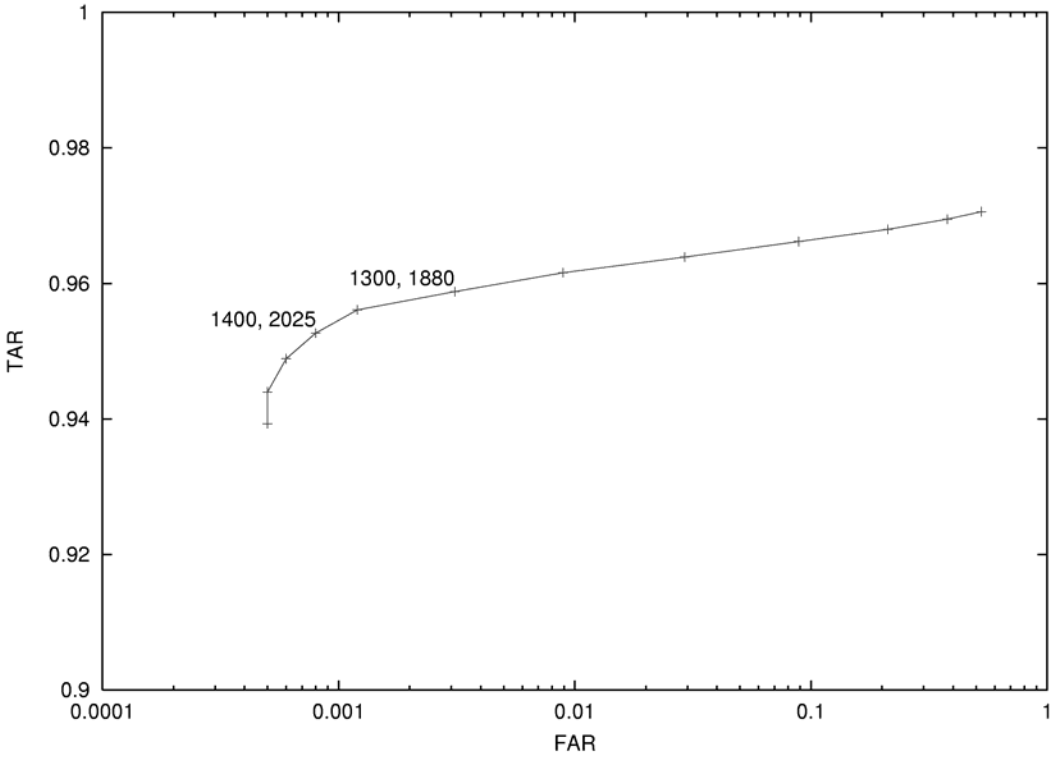

Figure 6: Identification ROC curve

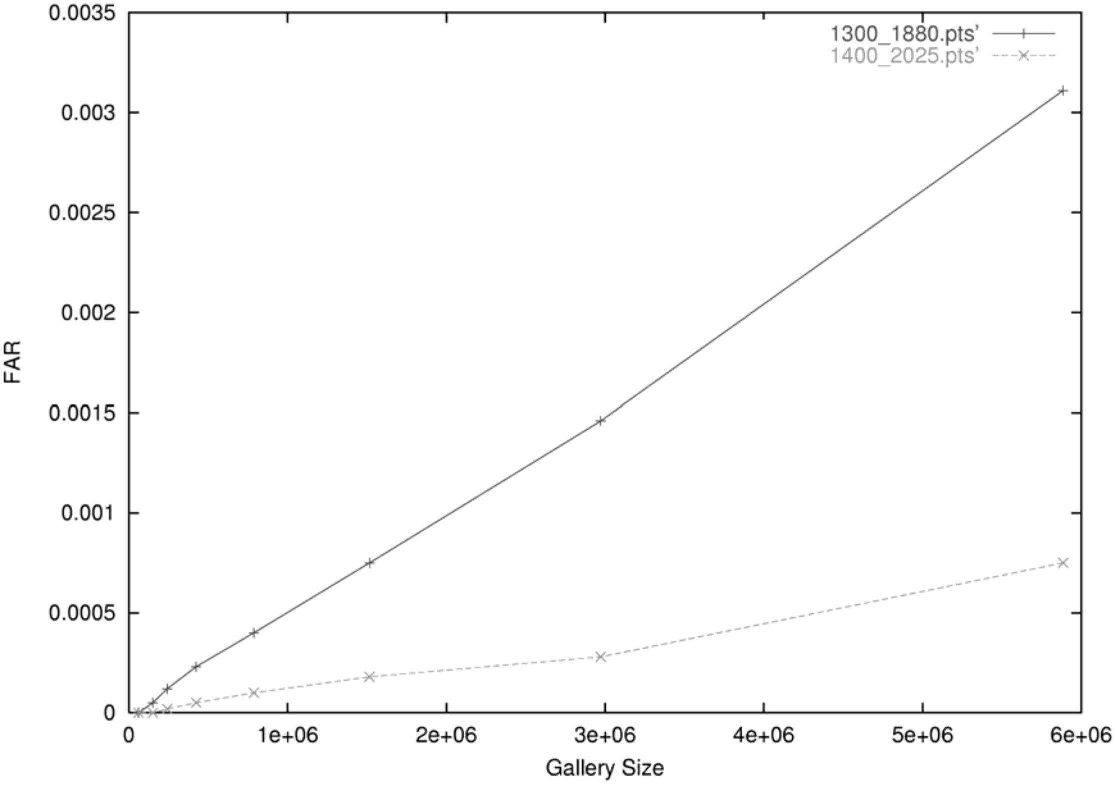

Figure 7: Identification FAR versus gallery size

There are two curves in Fig. 7. Each was created with a different set of operational thresholds applied to the identification system's matcher score. The analysis suggests that FAR can be significantly reduced, even out to a gallery of nearly 6 million, by using the thresholds associated with the bottom curve over the thresholds used to create the top curve.

3.3. Consolidation

All this, whether for verification or identification, relies upon accurately knowing the identities associated between the probes set and the gallery set. The datasets provided to NIST record a subject ID with each biometric data record. Unfortunately, operational errors are unavoidable, so situations arise where a subject has been assigned more than one ID over multiple biometric capture sessions, and cases arise where more than one subject have been assigned the same ID. There is also the added complication of sampling records across datasets from different sources to form probe and gallery sets. In the case of law enforcement fingerprints, a subject may have been arrested multiple times in different jurisdictions that collect independent archives of fingerprints, so that associating identity is not possible based on a single jurisdiction's subject ID. For example, it is quite possible for the same subject to have fingerprint records on file at both the state and federal levels, and it is quite possible for the same subject to be on file in neighboring states.

Because of errors in the data records and due to unknown identity associations across data sources, it is necessary to *clean* the data after initial probe and gallery set selections have been made for a particular test. This process involves using one or more biometric systems to match the probe set to the gallery set, and then analyzing the resulting matcher scores. The matcher scores of alleged mates are sorted, and the lowest scoring cases are identified. The matcher scores of alleged non-mates are sorted, and the highest scoring cases are identified. These cases (low scoring alleged mates and high scoring alleged non-mates) are sent to trained fingerprint examiners for human review and manual verification. The results of these reviews are then used to clean the probe and gallery sets. This process is known as consolidation, and is quite costly, but most necessary.

3.4. NIST fingerprint test

NIST has conducted a large number and a wide variety of fingerprint matching tests[*] using the data and methods described above. A baseline of performance for fingerprint matching technology was established using a NIST minutiae detector and matcher on a computer hardware system referred to as the Verification Test Bed (VTB)[24]. NIST has a fingerprint identification system replicating technology used in the FBI IAFIS system called the Algorithmic Test Bed (ATB). Extensive testing under various scenarios has been conducted on the ATB[25]. NIST also has a fingerprint identification system replicating technology used in the DHS US-VISIT system called the IDENT Test Bed (ITB), and a number of large scale tests[26] have been performed on this system. There are also numerous vendor tests of commercial technology that have been conducted. These include FpVTE[22], and many fingerprint verification tests evaluating different vendors' software development kits (SDKs)[27]. At the time of this paper, an evaluation of four-finger slap segmentation technology was under way called, SlapSeg04[28]. A summary of results and conclusions from a number of these tests follows.

3.4.1. FpVTE

This fingerprint vendor test was designed to: 1.) measure the accuracy of fingerprint matching, identification, and verification systems; 2.) identify the most accurate fingerprint matching systems; 3.) determine the viability of fingerprint systems for near-term deployment in large-scale identification systems; 4.) determine the effect of a wide variety of variables on matcher accuracy; and 5.) develop a well-vetted set of a variety of operational data for use in future research. The evaluations were not intended to: 1.) measure system throughput or speed; 2.) evaluate scanners or other acquisition devices; 3.) directly measure performance against very large databases; and 4.) take cost into consideration.

[*] These tests were performed for the Department of Homeland Security in accordance with section 303 of the Border Security Act, codified at 8 U.S.C. 1732. Specific hardware and software products identified in this report were used in order to adequately support the development of technology to conduct the performance evaluations described in this document. In no case does such identification 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.

Based on the results of participants, it is concluded that: 1.) there is a substantial difference in accuracy between the best fingerprint matching systems and the average or worst systems; 2.) the top three systems are more consistent in performance than the other systems; 3.) these top-tier systems perform consistently well over a variety of data, and are less affected by fingerprint quality and other variables; and 4.) the performance of these most accurate systems has been verified using the NIST SDK testing.

One of the most significant results demonstrated by FpVTE is that the accuracy of a fingerprint matching system consistently improves as more fingers are used in the system. In other words, a fingerprint identification system will have improved accuracy by capturing, storing and matching on 2 fingerprints per subject, rather than just one. This trend holds all the way up through 8 fingers. The conclusion is that the more fingers used, the more potentially accurate the fingerprint matching system.

3.4.2. SDK tests

The NIST SDK fingerprint matcher tests are a medium scale evaluation of 1-to-1 verification. Currently, 16 software matchers from 10 different vendors have been tested, each on samples from 20 different fingerprint datasets. Goals of this testing include: 1.) determine the feasibility of verification matching in US-VISIT and DOS application clients; 2.) evaluate vendor accuracy variability; 3.) evaluate vendor sensitivity to image quality; and 4.) these tests were used to scale evaluations in FpVTE.

Each fingerprint matcher SDK test involves a probe and gallery set of 6×10^3 by 6×10^3 for a total of 36×10^6 matches per test. To qualify for testing, the software matcher must compute a match in less than 10ms per fingerprint pair on a 3GHz Pentium platform. Results are reported in an ROC plot like those being compared in Fig. 8. As can be seen, there is a significant difference in performance between the best and worst performing matchers.

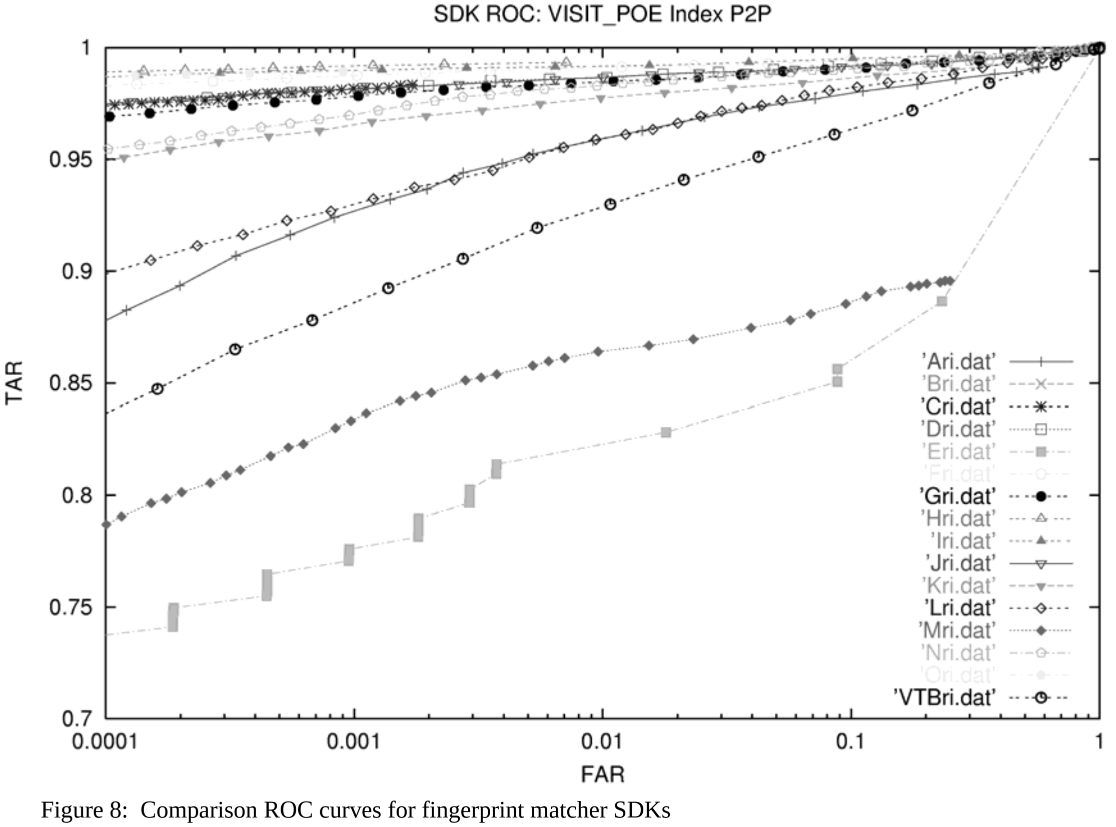

Figure 8: Comparison ROC curves for fingerprint matcher SDKs

Some conclusions from these SDK tests are: 1.) all vendors are sensitive to image quality; 2.) three algorithm vendors are clearly more effective; 3.) combining two fingers will provide very effective one-to-one verification for the US-VISIT program; and 4.) the NIST VTB algorithm is better than many commercial products.

3.4.3. US-VISIT certification

There are three main biometric functions provided by the DHS US-VISIT system: 1.) watch list checking at the time of enrollment; 2.) duplicate identification check for visa holders; and 3.) 1-to-1 verification for enrolled travelers. To date, NIST has conducted tests using the ITB and its fingerprint matchers to certify the last two biometric functions[26]. Evaluations to certify watch list functionality are currently under way.

The fingerprint matchers used in the US-VISIT system were provided to NIST as SDKs and evaluated. The results demonstrate that, for 1-to-1 verification, a TAR of 99.6% is achieved at a FAR of 0.1%. This performance has been measured to date across a collection of more than 800×10^3 operational fingerprints. Tests conducted on the ITB, demonstrate that, for a 1-to-many identification of two flat index fingerprints, a TAR of 96% is achieved at a FAR of 0.09%. This performance has been measured using a probe set of 60×10^3 subjects matched against a gallery set of 5.7×10^6 subjects.

4. TECHNOLOGY DEVELOPMENT

It should be of little surprise that technology must be developed in order to test technology. This involves R&D on two fronts. The first is the development of performance evaluation technology and the second is the development of application technology. Due to the relatively unique role that NIST plays, technology in both domains is actively being developed.

4.1. Performance evaluation technology

Given the methods outlined in Sec. 3, software is needed for compiling and manipulating similarity matrices of matcher scores, and then computing analyses from these scores such as the various ROC plots and results shown above.

4.2. Application technology

To be a credible technology evaluator requires understanding and expertise with the technology being tested. It is also important to have in-house algorithms and systems on which performance methods and scoring software can be developed. There is also the need to process and manipulate large collections of data and prepare them for evaluations. NIST has allocated significant resources over the years to developing application technology with respect to automated fingerprint matching. The cumulative result is a large collection of software tools referred to as the NIST Fingerprint Image Software 2 (NFIS2)[29]. This software distribution is currently in its second release and is available free of charge on CD-ROM subject to U.S. export control laws.

The provided utilities fall under seven general categories: 1.) NFSEG – a fingerprint segmentation algorithm, segments the four-finger plain impressions (slaps) found on the bottom of a fingerprint card into individual fingerprint images or it can be used to remove white space from around a single fingerprint image; 2.) PCASYS – a neural network based fingerprint pattern classification system, automatically categorizes a fingerprint image into the class of arch, left or right loop, scar, tented arch, or whorl; 3.) MINDTCT – a minutiae detector, automatically locates and records ridge endings and bifurcations in a fingerprint image; 4.) NFIQ – a fingerprint image quality algorithm, analyzes a fingerprint image and automatically assigns a quality value of 1 (highest quality) to 5 (lowest quality) to the image; 5.) BOZORTH3 – a fingerprint matching algorithm, is the minutiae-based fingerprint matching algorithm used in the VTB that has been studied extensively by NIST in both 1-to-1 verification and 1-to-many identification scenarios[24]; 6.) AN2K – a reference implementation of ANSI/NIST-ITL 1-2000[30], is a suite of utilities designed to read, write, edit, and manipulate files formatted according to this law enforcement data interchange standard; and 7.) IMGTOOLS – is a large collection of general-purpose image utilities that support the processing of fingerprint images such as baseline and lossless JPEG and the FBI's Wavelet Scalar Quantization (WSQ) encoders and decoders.

4.2.1. NIST fingerprint image quality (NFIQ)

Of all this technology, perhaps the most significant is the recent development of the NIST Fingerprint Image Quality (NFIQ) algorithm[23]. This tool is a predictor of fingerprint matcher performance, and has been designed to be matcher independent. The algorithm does not use traditional image processing attributes such as contrast and signal to noise; rather, image quality is defined in terms of characteristics and features of fingerprints that convey information for a matching algorithm. The algorithm assigns a quality value ranging between 1 and 5, where 1 is of highest quality and 5 is lowest (unusable) quality.

NFIQ has been extensively tested by taking the fingerprints comprising each of the 20 different datasets used in SDK testing and binning the fingerprints into the five levels of quality. The fingerprints in each dataset quality bin were then matched using 14 of the fingerprint matchers used in the SDK tests. In every case (280 in all), the ranking of the quality-based ROCs for a specific vendor's matcher on a specific dataset reflected accurately the assigned quality level. This clearly demonstrates NFIQ's ability to predict matcher performance. Results show that, as NFIQ goes from 1 to 5, matching accuracy falls from 99.6% to 26%.

Using this technology, it is now possible to take a fingerprint image and determine whether it is of sufficient quality to result in high matcher performance, or whether it is of sufficiently low quality to result in low matcher performance. The implications of this technology on traditional fingerprint capture and matching workflows is huge. Now, quality control over fingerprints can automatically take place at the time of capture. If the first fingerprint captured is of insufficient quality, it is possible to catch this in real time and request the subject's fingerprint be retaken on the spot. This technology also introduces the ability for fingerprint matching systems to devote different levels of computing resources according to the assessed quality of the fingerprint image. Those prints that are determined to be of low quality may be routed to slower more robust matching algorithms, while the higher volume of high quality prints may be routed to faster matching algorithms.

5. STANDARDS PARTICIPATION

Participation in the creation of biometric standards is another important activity of the Image Group at NIST. There are two major areas of focus: 1.) the establishment of standards for biometric data interchange and interoperability, and 2.) the development of standards methods for biometric testing and reporting performance results. NIST actively participates in both of these areas at both the national and international level.

At the national level, technology standards are developed within the American National Standard Institute (ANSI)[31]. NIST is one of a number of organizations that can sponsor standards within ANSI. In this role, NIST has sponsored the standard entitled, "Information Technology – American National Standard for Information Systems – Data Format for the Interchange of Fingerprint, Facial, & Scar Mark & Tattoo (SMT) Information," ANSI/NIST-ITL 1-2000[30]. This standard specifies a data interchange format containing fingerprints, mugshots, and other identifying images along with a subject's background information for use within the law enforcement community. Virtually all law enforcement fingerprint data interchange at the local, state, federal and international levels relies on this standard for interoperability.

The InterNational Committee for Information Technology Standards (INCITS)[31] is another organization that sponsors information and communication technology standards within ANSI. NIST participates in the M1 technical committee of INCITS that focuses on biometric standards. There have been significant standards developed and adopted in 2004 regarding fingerprint data formats. INCITS 381-2004, "Information Technology – Finger Image Based Interchange Format," specifies a header structure for exchanging fingerprint images. INCITS 378-2004, "Information Technology – Finger Minutiae Format for Data Interchange," specifies a structured format for exchanging fingerprint minutiae data. INCITS 377-2004, "Information Technology – Finger Pattern Based Interchange Format," specifies a structured format for representing pattern vectors derived from tiling the fingerprint image.

There is also significant standards work being conducted in the area of biometric testing and reporting. NIST is seeking to guide the development of these standards by leveraging the methods and experience gained in the fingerprint matching tests described in this paper, and also with other biometric tests, including face recognition. INCITS/M1 is currently working on a draft standard entitled, "Biometric Performance Testing and Reporting," INCITS 1602-D. There are four parts to this draft: Part 1 – Framework; Part 2 – Technology Testing and Reporting; Part 3 – Scenario Testing and Reporting; and Part 4 – Operational Testing and Reporting.

On the international level, standards are developed by the International Organization for Standards[33] and International Electrotechnical Commission (ISO/IEC). INCITS serves as ANSI's Technical Advisory Group to the Joint Technical Committee 1 (JTC 1) within ISO/IEC. JTC 1 is responsible for international standardization in the field of information technology. JTC 1 established Subcommittee 37 (SC 37) on Biometrics with the goal to ensure a high priority, focused, and comprehensive approach worldwide for the rapid development and approval of formal international biometric standards. Through this relationship, the biometric standards adopted nationally by ANSI are then passed on to seed international standards within SC 37.

For example, the INCITS/M1 work on 1602-D is currently reflected in the international draft standards from JTC 1/SC 37 entitled, "Information technology -- Biometrics performance testing and reporting -- Part 1: Test Principles," ISO/IEC CD 19795-1 and "Information technology -- Biometrics performance testing and reporting -- Part 2: Testing Methodologies," ISO/IEC WD 19795-2.

6. SUMMARY

NIST is an international leader in biometric systems testing and technology evaluation. Post 9-11, focus has shifted from law enforcement to border control, with the deployment of the DHS US-VISIT program. Today, NIST maintains one of the world's largest multi-jurisdictional fingerprint image repositories. To achieve new congressional mandates, NIST continues to develop methods and technology for evaluating large government biometric systems.

NIST offers the following Patriot Act recommendations as a result of extensive testing. For 1-to-1 matching, two index finger images are sufficient for accurate automated biometric verification, while a face image is recommended for manual human verification. For 1-to-many matching, ten flat fingerprint images are sufficient for accurate automated biometric identification. Large existing government databases require ten fingers because archives contain relatively poor quality images.

This paper has presented a biometric certification framework that focuses its activity in four key areas: 1.) developing test datasets, 2.) conducting performance assessment; 3.) technology development; and 4.) standards participation. Readers of this report are encouraged to consider how they might best leverage this approach for their particular technology domain.

REFERENCES

1. J.H. Wegstein, "A Semi-automated Single Fingerprint Identification System," NBS Technical Note 481, April 1969.
2. J.H. Wegstein, "Automated Fingerprint Identification," NBS Technical Note 538, August 1970.
3. R.T. Moore, "Results of Fingerprint Image Quality Experiments," NBS Technical Report NBSIR 81-2298, June 1981.
4. J.H. Wegstein, "An Automated Fingerprint Identification System," NBS Special Publication 500-89, February 1982.
5. R.M. McCabe, and R.T. Moore, "Data Format for Information Interchange," American National Standard ANSI/NBS-ICST 1-1986, August 1986.
6. R.T. Moore, "Automated Fingerprint Identification Systems - Benchmark Test of Relative Performance," American National Standard ANSI/IAI 1-1988, February 1988.
7. C. Watson, "NIST Special Database 4: 8-bit Gray Scale Images of Fingerprint Image Groups," CD-ROM & documentation, March 1992.
8. C.L. Wilson, G.T. Candela, P.J. Grother, C.I. Watson, and R.A. Wilkinson, "Massively Parallel Neural Network Fingerprint Classification System," Technical Report NISTIR 4880, July 1992.
9. R. McCabe, C. Wilson, and D. Grubb, "Research Considerations Regarding FBI-IAFIS Tasks & Requirements," NIST Technical Report NISTIR 4892, July 1992.

10. C. Watson, "NIST Special Database 9: 8-Bit Gray Scale Images of Mated Fingerprint Card Pairs," Vol. 1-5, CD-ROM & documentation, May 1993.
11. C. Watson, "NIST Special Database 10: Supplemental Fingerprint Card Data (SFCD) for NIST Special Database 9," CD-ROM & documentation, June 1993.
12. C. Watson, "NIST Special Database 14: Mated Fingerprint Card Pairs 2," CD-ROM & documentation, September 1993.
13. R.M. McCabe, "Data Format for the Interchange of Fingerprint Information," American National Standard ANSI/NIST-CSL 1-1993, November 1993.
14. J.L. Blue, G.T. Candela, P.J. Grother, R. Chellappa, C.L. Wilson, "Evaluation of Pattern Classifiers for Fingerprint and OCR Application," in Pattern Recognition, 27, pp. 485-501, 1994.
15. C.I. Watson, J. Candela, P. Grother, "Comparison of FFT Fingerprint Filtering Methods for Neural Network Classification," Technical Report NISTIR 5493 September 1994.
16. C. Watson, "NIST Special Database 18: Mugshot Identification Database of 8 bit gray scale images," CD-ROM & documentation, December 1994.
17. G.T. Candela, P.J. Grother, C.I. Watson, R.A. Wilkinson, C.L. Wilson, "PCASYS - A Pattern-level Classification Automation System for Fingerprints," Technical Report NISTIR 5647 & CD-ROM, April 1995.
18. R.M. McCabe, "Data Format for the Interchange of Fingerprint, Facial & SMT Information," American National Standard ANSI/NIST-ITL 1a-1997, April 1997.
19. C. Watson, "NIST Special Database 24: Digital Video of Live-Scan Fingerprint Data," CD-ROM & documentation, July 1998.
20. M.D. Garris and R.M. McCabe, "NIST Special Database 27: Fingerprint Minutiae From Latent and Matching Tenprint Images," CD-ROM & documentation, June 2000.
21. J. Phillips, P. Grother, R.J. Micheals, D.M. Blackburn, E. Tabassi, M. Bone, "Face Recognition Vendor Test 2002 – Overview and Summary," NIST Internal Report 6965, March 2003, http://www.frvt.org/
22. C. Wilson, R.A. Hicklin, H. Korves, B. Ulery, M. Zoepfl, M. Bone, P. Grother, R. Micheals, S. Otto, C. Watson, "Fingerprint Vendor Technology Evaluation 2003 – Summary of Results," NIST Internal Report 7123, June 2004, http://fpvte.nist.gov/
23. E. Tabassi, C. Wilson, C. Watson, "Fingerprint Image Quality," NIST Internal Report 7151, August 2004.
24. C. Wilson, C. Watson, M. Garris, A. Hicklin, "Studies of Fingerprint Matching Using the Verfication Test Bed (VTB)," NIST Internal Report 7020, July 2003.
25. S. Wood, C. Wilson, "Studies of Plain-to-Rolled Fingerprint Matching Using the NIST Algorithmic Test Bed (ATB)," NIST Internal Report 7112, April 2004.
26. C. Wilson, M. Garris, C. Watson, "Matching Performance for the US-VISIT IDENT System Using Flat Fingerprints," NIST Internal Report 7110, May 2004.
27. C. Watson, C. Wilson, K. Marshall, M. Indovina, R. Snelick, "Studies of One-to-One Fingerprint Matching with Vendor SDK Matchers," NIST Internal Report 7119, June 2004.
28. "Slap Segmentation Fingerprint Evaluation 2004," http://fingerprint.nist.gov/slapseg04/
29. C. Watson, M. Garris, "NIST Fingerprint Image Software 2 (NFIS2)," CD-ROM, October 2004, http://fingerprint.nist.gov/NFIS/
30. R.M. McCabe, "Data Format for the Interchange of Fingerprint, Facial, & Scar Mark & Tattoo (SMT) Information," American National Standard ANSI/NIST-ITL 1-2000, July 2000. Available from R.M. McCabe at NIST, 100 Bureau Drive, Stop 8940, Gaithersburg, MD 20899-8940.
31. American National Standards Institute (ANSI), http://www.ansi.org/
32. InterNational Committee for Information Technology Standards (INCITS), http://www.incits.org
33. International Organization for Standardization (ISO), http://www.iso.org