NISTIR 7377

Summary of NIST Latent Fingerprint Testing Workshop

V. N. Dvornychenko
Michael D. Garris

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

NISTIR 7377

Summary of NIST Latent Fingerprint Testing Workshop

V. N. Dvornychenko
Michael D. Garris
Information Technology Laboratory
Information Access Division

November 2006

U.S. DEPARTMENT OF COMMERCE
Carlos M. Gutierrez, Secretary
TECHNOLOGY ADMINISTRATION
Michelle O'Neill, Acting Under Secretary of Commerce for Technology
NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
William A. Jeffrey, Director

Summary of NIST Latent Fingerprint Testing Workshop

V. N. Dvornychenko (vdvorny@nist.gov) and

Michael D. Garris (mgarris@nist.gov)

Abstract

The central focus of this paper is on the Latent Testing Workshop, held on the National Institute of Standards and Technology (NIST) Campus April 5 to 6, 2006 and the lessons that were learned from it. The primary goal of the workshop was to gather information for the creation of a "Latent Challenge," whose purpose is to stimulate Latent Automated Fingerprint Identification System (AFIS) vendors to submit their AFIS prototype systems for evaluation. To lay the foundations for such testing it is necessary that: 1) suitable test sets be identified and prepared; 2) the Application Programming Interface (API) is defined; and 3) the effective methods of performance scoring be defined. To provide background and context, past and present latent fingerprint activity at NIST is discussed. While the primary focus is on latent fingerprints, the paper also surveys relevant general biometrics activity.

Keywords: biometrics; Electronic Fingerprint; Extended Fingerprint Feature Set; EFFS; fingerprint image quality; latent fingerprints; Latent Workshop; level-3 features; lights-out systems; matcher architecture; performance testing

1. Introduction

A Latent Testing Workshop was hosted at the National Institute of Standards and Technology (NIST) on April 5 to 6, 2006. In all, 61 participants representing U.S. and international commercial, government, and academic interests came together to discuss the needs of the latent fingerprint community; assess the state of the art in latent matchers; and to obtain the views of latent matcher vendors on their ability and willingness to field "semi-lights-out" latent search systems for testing purposes. (The concept of "semi lights out matchers" is introduced in Section 2.6.) These topics were successfully explored, and in large measure resolved, in the workshop.

The purpose of this paper is to present the findings of the Latent Testing Workshop and to show how these findings relate to past and future fingerprint activity at NIST. Section 2 provides background to NIST's role in evaluating fingerprint biometrics; the concepts of "lights-out" and "semi-lights-out" latent processing are described; this material provides context as to why a latent workshop was necessary. Section 3 presents an overview of the Latent Testing Workshop including who presented on what topics, what lessons were learned, what are the perceived needs of the latent community, as well as additional topics. Section 4 concludes with directions for future work at NIST in response to the workshop.

2. Background

2.1 The Role of NIST

The traditional role of NIST has been the assessment of technology, the establishment of metrics, and the setting of standards [1][2]. In this connection NIST has a long-standing interest in biometrics. This interest has primarily focused on fingerprints, but selective work has been done on iris prints, facial recognition and Deoxyribonucleic acid (DNA) [3][4].

Recently (2006), the USA Patriot Act specifically mandated that NIST develop and certify a standard for verifying the identity of persons applying for a visa or seeking to enter the United States, and that NIST provide technical support to the Attorney General and Secretary of State in evaluation of biometric identification systems for Entry and Exit Data System for U.S. borders [5][6].

Since that time, other agencies, including the U. S. Department of Homeland Security (DHS), the Federal Bureau of Investigation (FBI), and the U. S. Dept. of Justice (DOJ), have sought technical guidance from NIST in several evaluations of biometric systems which may be considered for homeland security [7][8].

2.2 Fingerprints as a Biometric

Biometrics are measurements that assist in uniquely identifying ("individualizing") a person [9][10][11]. Fingerprints are currently the biometric of choice – and are likely to remain so for the foreseeable future. There are several reasons for this. Fingerprints are less intrusive than certain other biometrics, for example iris and DNA, though more intrusive than, say, facial photos or voice prints. They also have very good discrimination/identification power. Perhaps what is most important, they have a long and favorable track record with police departments and the courts. As a result, a tremendous amount of money and time has been invested into the fingerprint infrastructure.

Fingerprints may be roughly divided into three categories: 1) Rolled impressions, 2) Plain impressions or "flats," and 3) Latent fingerprints. "Palmprints" form a separate but closely related category. Partly because of their large size, lack of standard databases, and their complexity, they have been slower to develop than other types, especially regarding automation. Recently, however there has been an effort to bring palmprints to a level of development approximating the other types. Accordingly, the FBI's Criminal Justice Information Services (CJIS) Division is in the process of creating a repository of palmprints [12]. Commercial automated palmprint systems are becoming available. Examples of three types of fingerprints are provided in Figure 1. The soles of the feet also have friction ridge patterns similar to palms, but these are seldom encountered except for infant identification.

In a rolled impression the finger is carefully rolled (or rocked) from one side to the other so as to obtain the impression of a greater area. This has advantages and disadvantages. While a greater coverage results, the impression is slightly distorted by the rolling process. In a plain impression (or "flat") the finger is pressed down with a moderate pressure but not rolled. While the result covers a smaller area, there is less distortion. Rolled and plain impressions are typically taken from a cooperative subject, and are subject to a retake if lacking in quality (incomplete, smudged or otherwise unclear). As a result, they are typically of high quality and rich in information content.

Latent fingerprints are generally inadvertently left at a scene, are not subject to retake, and tend to be of considerably lower quality and information content. A second difference is the method of capture. Rolled and flats are captured electronically, or else as an inked impression (which may be subsequently scanned to create an electronic version). As the name suggests, latents are often "hidden" or at least not readily apparent. Consequently, they must be developed to render them suitable for searches and identification. “Development” may be as simple as photographing the print, but usually involves some type of processing, such as dusting with a powder, or chemical treatment.

Rolled and plain impressions may be used in one-to-many *identifications,* or in one-to-one *verifications*. In a fingerprint *identification,* or *search*, the given fingerprint is compared against many candidates comprising a *gallery* (also known as a *repository* or *background*), and zero, one, or more potential matches are reported back. For rolled or plain impressions, most identification systems will only report high-probability matches; for latents, many systems report back a fixed list of candidate matches regardless of similarity.

In verification the given fingerprint is compared to a single candidate and a match/no-match decision is made in “real time,” that is, in seconds. Fingerprint verification systems are used in physical access control and in logical access (e.g., banking). They employ “instant” livescan devices to capture a subject’s fingerprint, and compare this to a stored exemplar. Since the subject requiring verification provides his/her purported identity, the captured fingerprint need only be compared with a single stored exemplar. The process of entering the stored exemplar in the system’s database is referred to as *enrollment*.

Considerable amount of information on biometrics is now available online, for example refer to References [9][10][13][14].

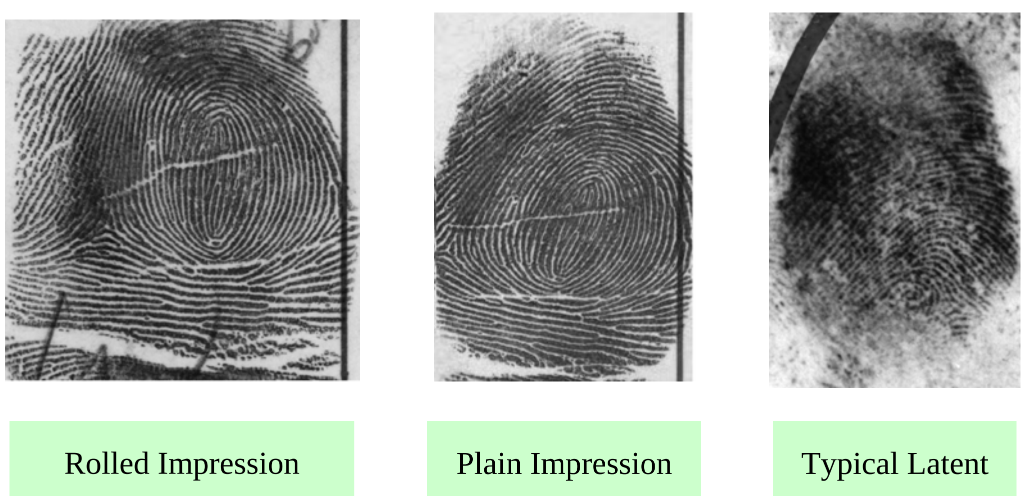


Figure 1. Example of Rolled Impression, Plain Impression, and Latent

2.3 Fingerprint Performance Testing

In line with its historical mission, NIST has performed a number of fingerprint tests whose purpose was to:

1. Verify and refine performance measurements of large operational fingerprint matchers, for example FBI/AFIS and US-VISIT IDENT [7][8];
2. Assess the state of the art (e.g., matching accuracy) of current vendor matching systems [15];

3. Determine the degree of interoperability between vendor systems when using common minutiae sets [16][17][18];
4. Assess the relative value of rolled vs. flat fingerprints in identification, and when matching against latent fingerprints [19]; and
5. Discover and consolidate "multiple enrollments" (the same subject appearing more than once in a database);

Much of this work is of considerable sophistication, and it is not possible to do it justice in a short summary. A great deal of information is found on the NIST website under

➔ http://fingerprint.nist.gov/

2.4 Need to Expand Scope of Latent Work

As mentioned, only a relatively small part of NIST testing has involved latent fingerprints. Recently there has been a large increase in interest in latent systems, driven by a need to: 1) perform real-time screening against latent databases at ports of entry; 2) improving interoperability between the various latent search systems; 3) improving response time for latents acquired at the crime scene by Crime Scene Investigators (CSI); 4) reducing human involvement in processing unsolved latent files and "dead files."

From NIST's standpoint these requirements lead to:

- A need to stimulate industry to produce latent search systems which are faster, more accurate, yet less labor intensive.
- NIST is interested in testing vendor systems to benchmark current performance levels and subsequent improvements.
- NIST would like these tests to be as automated as possible for the following reasons: a) simplify testing logistics; b) protect vendor proprietary data; and c) decouple the performance of the automated system from the performance of latent examiners.
- As will be shown, these requirements are best met by a "semi-lights-out" system.

Before introducing this concept we must first discuss what is meant by a "lights-out system."

2.5 The Concept of a "Lights-Out System"

A "lights-out" system is a computer system that normally requires no human intervention. In the present context, the system automatically performs all of the following:

- Accepts a set of fingerprints to be searched. Fingerprints may be rolled or plain, and may consist of several fingers to be searched jointly.
- Automatically extracts all the features required for searching
- Performs the search of the database (gallery)
- Determines if possible hits have sufficiently high scores so that no human verification is required (because of extremely high confidence in the result)
- Only these high-score/high-confidence candidates are output as "hits"

Of course even in a lights-out system there will be occasional "problem cases" requiring human review. Only recently has technology reached a level making semi-lights-out systems possible, but most ten-print matchers currently in use are not yet fully lights-out matchers. However, it is fully expected that successive generations will increasingly move in this direction.

2.6 The Concept of a "Semi-Lights-Out System"

Latent fingerprints contain considerably less information than other types (rolled or plain). Whereas a typical latent fingerprint might have fifteen usable minutiae, a typical rolled fingerprint will have sixty or more. In addition, rolled and plain fingerprints are generally searched as multi-finger searches of at least two fingers. Most latent searches consist of a single fingerprint – or more accurately, a single ***partial*** fingerprint. The reliability of a latent "hit" is therefore considerably less (on the average) than for ten-prints. Generally it is not possible to process them in a fully "lights-out" manner.

Current processing for latents is characteristically very human-intensive. The following steps must be performed by human experts:

1) Feature extraction is done almost entirely by human experts, with limited assistance from the machine
2) The human expert examines an extensive candidate list to see if it contains possible hits
3) The expert examines each possible hit to verify or reject it. (Of course by far the majority of examinations result in rejections.)

In a "semi-lights-out" system steps (1) and (2) are automated, and humans are only required for final verification. Limited human assistance may, however, be accepted in step (1). This can take the form of:

- Probable finger orientation
- Probable finger number
- Identifying ("lassoing") an area of interest (for example, to separate it from irrelevant background clutter)

2.7 Why are "Semi-Lights-Out Systems" of Special Interest?

There are time-sensitive scenarios in which the automation of feature extraction and screening of the output candidate list would be very useful. Examples include the following:

- In some applications there is a need to check an incoming fingerprint against an unsolved latent file, with a required turnaround time of seconds. In this particular case the feature extraction would probably be performed by humans at the time the latent is enrolled. However, "back end" processing would be required to ensure a candidate list of a length that can be processed within the time constraint. A human expert would of course make the final identification.
- Law enforcement agencies invariably have a large number of unprocessed latent images on file. They are unable to process these because they cannot spare examiners to mark up, search, and examine the returned candidates. Automated latent image searches and preliminary screening of candidates could identify high-probability hits with relatively little human effort.

These are just two examples in which semi-light-out latent searching has the potential to make hits that otherwise would not be made. Again, we emphasize that in all cases the final identification decision is made by a human latent examiner — no one is proposing automating the final decision.

2.8 Proposed Format of Latent Testing

NIST would like to be able to assess the performance of latent search systems developed by vendors. Prior to the workshop NIST had concluded that testing would be optimized if conducted under a "semi-lights-out" environment. The reasons are:

- Simplifies testing logistics;
- Protects vendor proprietary data; and
- Decouples the performance of the automated system from the performance of latent examiners to the greatest degree possible.

However, NIST was unsure if vendors could provide "semi-lights-out" systems for testing.

2.9 Unresolved Questions Leading to the Need for a Workshop

Since we are proposing that a "semi-lights-out" latent search system is the best environment for testing, it is important to determine if the latent fingerprint community and vendors are prepared to field such systems – if only for testing purposes. To answer this and similar questions, NIST decided to hold a Latent Testing Workshop. Specific topics to be answered during the workshop were:

- How are latent fingerprints being currently used by agencies?
- What are the specific needs of agencies using latent fingerprints?
- What improvements/innovations they would like to see?
- Do vendors believe a semi-lights-out system is within reach with present technology?
- If so, what might be the approximate performance level of such a system?
- Have there been large performance gains in the last couple years?
- Are vendors interested in providing a semi-lights-out latent system(s) for testing?
- Is NIST's proposed testing concept workable and fair to the vendors?
- What is a reasonable time frame for fielding a test?
- What changes in proposed NIST testing methodology does the latent community suggest?
- How difficult is it to acquire additional latent exemplars for testing purposes?
- What are promising sources of latent test data?
- What are promising new features ("Extended Fingerprint Feature Set (EFFS)") for latent matching?
- Are "special purpose" testing sets desirable?

3. The NIST Latent Testing Workshop

The two-day Latent Testing Workshop was held at NIST, Gaithersburg, Maryland on April 5 to 6, 2006. There were 61 participants in the workshop. A breakdown of participants by professional affiliation is given below:

Affiliation of Workshop Participants		
Organization		Number of Participants
United States		49
	Government	20
	State/Local Law Enforce.	8
	Commercial	20
	Academic	1
Foreign		12
	Government	6
	Commercial	5
	Academic	1
Total		61

Table 1. Breakdown of participants by professional affiliation

Twenty-seven presentations were given, followed by a wrap-up discussion. The two sections following provide a summary of what was learned. Additional information may be obtained from the NIST website ➔ http://www.itl.nist.gov/iad/894.03/latent/ and from ➔ fingerprint.nist.gov

The first website contains almost all presentations given, but in some cases in slightly redacted form. In one instance (presentation number 2) it was not possible to post the presentation at all.

3.1 Categorization of Workshop Presentations by Topics

This section categorizes each presentation in terms of a main topic, and one or more secondary topics. On the NIST website the presentations are in alphabetic order by presenter name. Table 2 is provided to facilitate the transition between the two (alphabetic order and presentation number). In its electronic form, this table contains "hot links" which can take one directly to the website presentation.

Cross Reference for Presentations			
Presentation Sequence No.	**Alphabetic No. (website)**	**Speaker's Name**	**Title of Presentation**
0	12	Marty Herman and Vladimir Dvornychenko	Opening Remarks
1	21	Stephen Meagher	Latent Processing at the FBI/LPU
2	28	Kasey Wertheim	Military Latent Needs
3	9	Danny Greathouse	Need for Rapid Turnaround at US-VISIT
4	23	Thomas Smith	Need for AFIS Vendors to Improve Search Capability
5	22	Francis P. Senese	Problems Experienced with Livescans
6	25	B. Scott Swann	Needs and Applications of Latents at FBI/CJIS
7	19	Deborah Leben	Applications ULW at DHS
8	20	Gordon Low	Needs and Applications of Latents at CA Dep't of Justice
9	6	Jeri Eaton and Wade Petroka	Needs and Applications of Latents at King County, WA, Sheriff Office
10	24	Ambika Suman	Evaluating Automated Finger and Palm Mark (latent) Searching
11	3	Mark Branchflower	Implementation of Remote Latent Search Capability at Interpol
12	7	Jean-Christophe Fondeur	Experience in Lights-out Processing
13	4	Wally Briefs	Maximizing Latent Identification Performance
14	2	Behnam Bavarian	Thresholds and Parameters for Automatic Decisions
15	11	Masanori Hara	Thoughts on Automatic Latent Processing and Matching Algorithms
16	16	Tom Hopper	ULW Approach for Sharing Latent Identification Services
17	11	Nigel Allinson and Ian Gledhill	Wireless Transmission of Fingerprints
18	13	Austin Hicklin	CDEFFS -- Extended Feature Sets
19	17	Anil Jain and Yi Chen	High Resolution Matching Using Level 3 Features
20	27	Phillip Wasserman	Level 3 Feature Detection Using Support Vector Machines
21	26	Elham Tabassi	Quality Measure Workshop Lessons Learned
22	28	Kasey Wertheim	Latent Quality Measures
23	14	Austin Hicklin	Quantifying Latent Quality
24	18	George Kiebuzinski	Lessons Learned during IAFIS Source Selection
25	10	Patrick Grother	Offline Biometric Testing at NIST
26a	29	Stephen Wood	Test Sets at NIST
26b	5	Vladimir Dvornychenko	Latent Test Sets "Wish List"
27	8	Michael Garris	Proposed Latent Testing Methodology
28	15	Austin Hicklin	CDEFFS Committee Meeting

Table 2. Workshop speakers and presentations

3.2 Lessons Learned from Workshop

We believe the workshop was very successful and answered most of the questions we hoped to resolve. These included:

1. Is the community, particularly the vendors, favorable toward automated testing of latent search systems?

   We found the responses to be generally favorable. Several vendors indicated that they had made significant progress in automated feature extraction, and the prioritization of candidates. This suggests that an automated search system could be ready for testing in the near future.

2. Are sufficient test data available?

   Prior to the workshop there was concern at NIST, as well as among other participants, that it might prove difficult to acquire a sufficient number of latent mated pairs for a really high quality test. By "mated pairs" is meant 1) a latent print along with 2) its mated rolled or plain-impression "exemplar." (Of course both a rolled and a plain impression would be ideal. Normally the latent is used as the "probe" or "search" and the rolled (or plain-impression) is seeded into the "gallery." However, sometimes a "reverse search" is conducted in which a rolled impression is searches against a gallery of latents.)

   Even assuming a statistically sufficient number of pairs are identified, it is highly desirable that these come from as diverse sources as possible so as to represent a good cross-section of the latent user community. Of particular concern is that the latents should not be largely selected from the population of AFIS hits. Such a selection would bias the results toward latents which exhibit characteristics "preferred by current-generation AFIS systems."

   In the course of the workshop a number of potential sources for latent test subjects were identified. Most of the sources were participants in the workshop, or else were identified by participants. The identified sources also fulfill the "diversity" requirements. It therefore appears that a fairly large quantity of latents, of diverse origin, may become soon available.

3. Do vendors think that such testing can be fairly conducted?

   On this issue there was some concern expressed. Some vendors raised the point that the series of one-to-one matches proposed by NIST, and based on the System Development Kit (SDK) model, would allow little or no data sharing between searches and that this would result in a suboptimal candidate list. Other vendors expressed the opinion that NIST was overly focused on the "ROC"[1] approach as a performance measure, and that for latents there existed better measures of performance, such as rank statistics.

4. The workshop also engendered a great deal of interest in image quality. It quickly became clear that two different types of image quality measures will be needed. The first is geared toward automated systems (computer oriented). The second is geared toward judiciary/evidentiary considerations (human oriented).

[1] "ROC" is an older term which stood for "Receiver Operating Characteristics." The more modern term is DET for "Detection Error Tradeoff"

Below we provide additional details. The topics generally follow the workshop program.

3.3 Needs of the Latent Community

3.3.1 Rapid Turnaround

A number of presentations stressed the need for rapid turnaround in processing latents. The typical scenario proposed was that of a latent print lifted at a crime scene by CSI, with a need to have this latent "screened" as rapidly as possible. The presentations stressed that the current typical turnaround time of two or three days is unacceptable.

What is patently needed is a method of rapid screening with a turnaround time of a few hours at most. This would give the criminal(s) less time to make good their escape. This initial screening may be at a reduced accuracy level if required to meet the time constraint. If the situation warrants it, this initial screening may be followed up with a more labor-intensive, but more accurate, search.

To provide this rapid response capability, three things are needed: 1) a good automated method of latent fingerprint image quality assessment for determining which latents are "more matchable," 2) a fast method of transmitting these latent to a central processing site, and 3) a semi-lights-out method of processing the image.

Semi-lights-out systems are not currently operational. However, several presentations (e.g., Allinson, P-17) presented information on rapid latent fingerprint transmissions. The presentation by FBI/CJIS (Swann, P-6) also stressed the shortage of latent experts, and the need for assistance via automation.

3.3.2 Workstations

Currently in the US there are several types of latent workstations in use. The three most common are: 1) Universal Latent Workstation (ULW) (developed by FBI/CJIS), 2) Remote Fingerprint Editing Software (RFES) (developed by Lockheed Martin[2] for the FBI) and 3) workstations developed by major AFIS vendors. Several presentations (#6, #7, #14) stressed the need to bring out new workstations combining the best features of ULW and RFES. At least one presentation (Bavarian, P-14) specifically recommended that ULW and RFES be retired and replaced by a workstation conforming to published standards for feature extraction.

3.3.3 Compression

"Data compression" is used for reducing data transmission times and data storage requirements. Latent fingerprints are rarely compressed because of a fear of losing information due to image degradation caused by compression/decompression. However in some cases the benefits of compression outweigh the losses, such as when it is desired to rapidly transmit the fingerprint to a central location using "handheld" wireless communication devices.

---

[2] This workshop was held for the US Department of Homeland Security in accordance with Section 303 of the border Security Act, codified as 8U.C.S.1732. Specific hardware and software products identified in this report were used to conduct work described in this document. In no case does identification of any commercial product, trade name, or vendor imply recommendation or endorsement by NIST, nor does it imply that the products and/or equipment identified are necessarily the best available for the purpose.

Only a few of the presentations dealt with compression per se. The presentation by Allinson (P-17) was the most detailed. The presentation discussed how the United Kingdom (UK) is using a handheld device to rapidly transmit latent fingerprints (called “marks” in the UK) to a central location for rapid screening. The compression algorithm used is JPEG2000. (JPEG stands for Joint Photographic Experts Group; 2000 is the year of the standard was introduced.) The researchers found that JPEG2000 slightly outperformed Wavelet Scalar Quantification (WSQ). The presentation by Branchflower (P-11) also mentions compression.

3.3.4 Other Needs

The presentation by Senese (P-5) discussed a problem encountered with a livescan device. The problem manifested itself as an incorrect or garbled ridge flow. (It was apparently caused by a “stitching” problem internal to the device when putting together several sub-images). While this occurred with an older model, similar problems still occur in newer models, and raise concerns in regard to standards, testing and certification.

3.4 Matcher Architecture

Although not singled out in the workshop program as an area of special interest, a number of presentations went into depth on the computer hardware/software architecture of current matchers. Some presenters proposed methods for matcher improvement. The presentations by Bavarian (P-14) and Hara (P-15) stand out in this regard. Bavarian characterizes advanced matchers as being “progressive matchers” and/or “fusion matchers.” Progressive matchers may be viewed as having a sequential architecture in which each stage acts as a filter to reduce the volume entering the next stage. This is a traditional design; the FBI/IAFIS basically falls into this category.

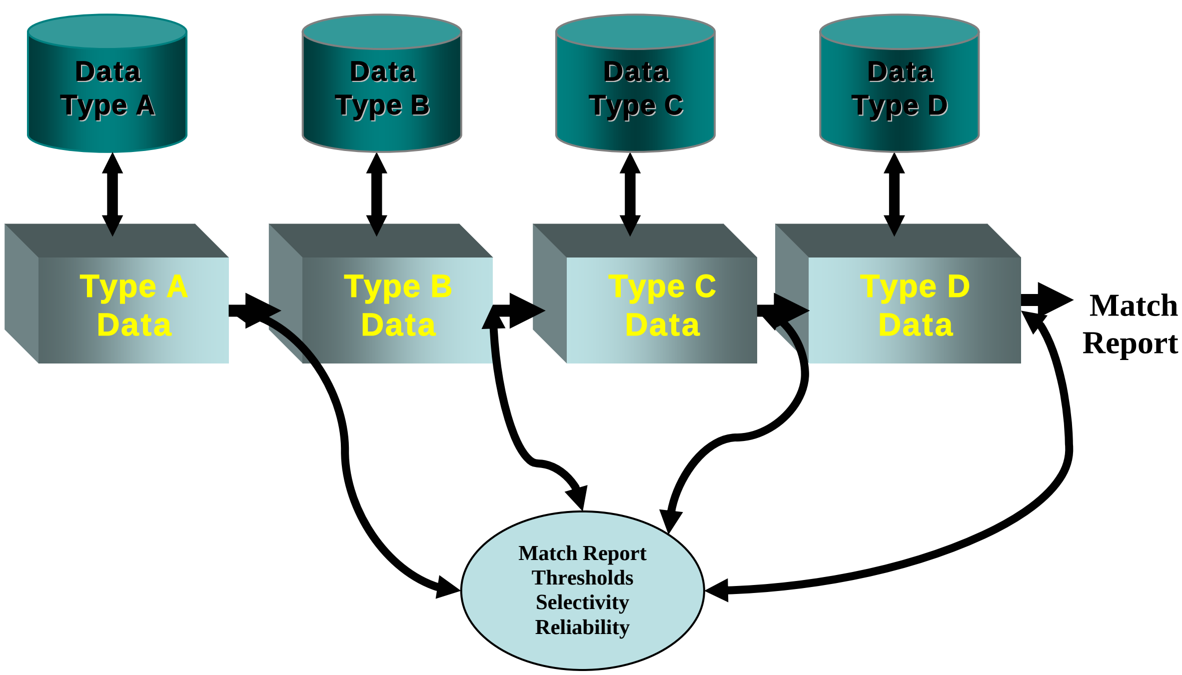


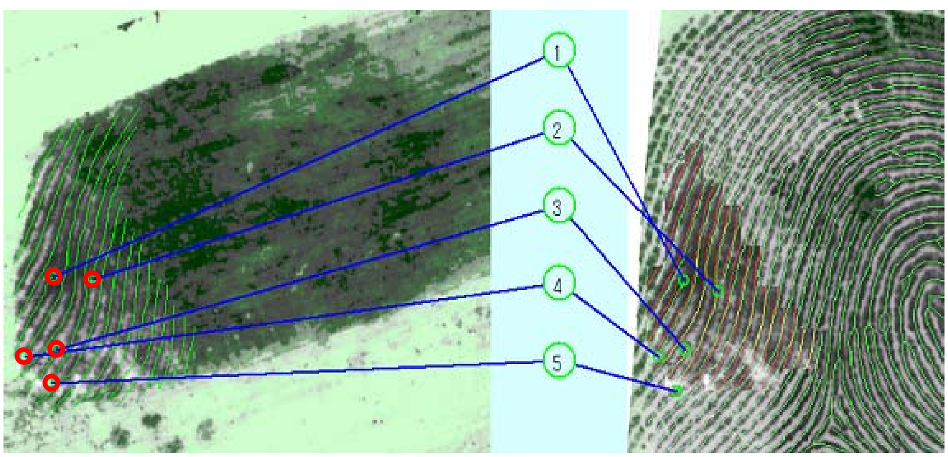
Figure 2. Multi-Stage Progressive Matcher (adapted from Bavarian)

Each "generic" matcher in the above diagram employs a different type of data. For example, "Type A Data" might refer to descriptive data such as *gender* and *age*; "Type B Data" might refer to fingerprint pattern classification; "Type C Data" might be minutiae; while "Type D Data" might refer to third level features. Regardless of the actual data type, for a candidate to pass through a given stage (and enter the next stage) its degree of similarity to the search exemplar must exceed some predefined threshold. Thus in the above diagram only candidates passing all four stages would emerge on the final "match report." Of course it is not necessary there be exactly four stages; some matchers might have more, while most would probably have fewer.

In this type of architecture it is common to refer to the initial stages as "filters" or "bulk filters." There is however a subtle distinction. A matcher might be employed as a "filter," but it is not necessarily true that a filter can always become a matcher. A necessary requirement for a matcher is that it can be "throttled down" until it passes only a few candidates (say less than five). With a filter this may not be possible. Thus, it may be possible to "throttle down" the filter to where it only passes 10 % of input, but attempts at further reduction might result in nothing passing the filter.

Regardless of how many stages the actual matcher has, it is virtually a certainty that with current matchers one of these will be a "minutiae matcher." This remains true for single-stage matchers. (The only exceptions appear to be matchers which use correlation techniques to directly match the fingerprint images. Prototypes of this type of matcher have been made using optical matching.)

Minutiae, it will be recalled, are defined as the endings or bifurcations of ridges. Figure 3 shows an example of extracted minutiae that are being matched between two fingerprint images.



Left: Latent fingerprint with five (5) minutiae identified ("extracted")

Right: The rolled impression "mate" of the latent on left. The corresponding minutiae are identified on the mate, and are connected via blue arrows.

Figure 3. Example of Minutiae Matching

A second type of matcher, fusion matcher, may be viewed as two or more matchers that work in parallel and independently of each other. Once all the component matchers have finished, a separate stage combines the information output by each individual matcher. This *combining logic* is generally at the "score level," in that the input into the combining logic consists of the output scores of the component matchers, possibly supplemented by ancillary information.

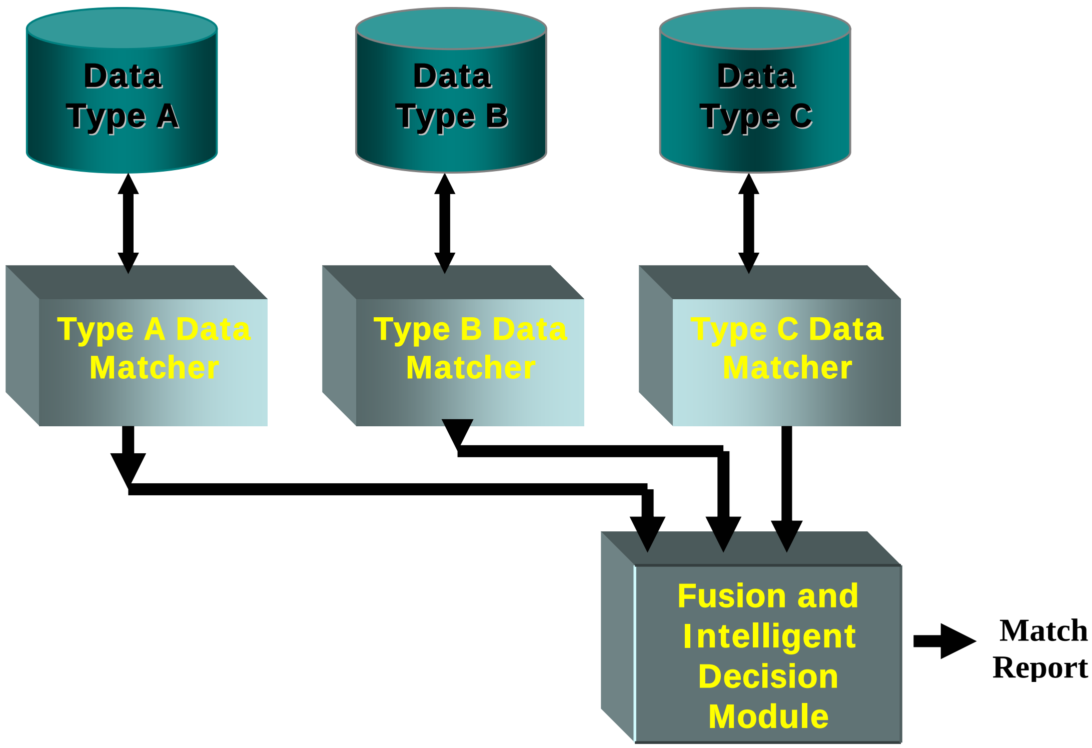


Figure 4. Multi-Stage Fusion Matcher (adapted from Bavarian)

The simplest way to combine the output scores is via straightforward addition, though this may not produce highest performance. To improve performance, a normalization step often precedes score fusion. The purpose of normalization is to make all the scores comparable in magnitude, and also to have scores reflect the intrinsic reliability of each component matcher. Some researchers have found that combining scores by score multiplication is more effective than addition. This may be because multiplication does not require normalization. Additional information on data fusion can be found in References [21][22].

Hara (P-15) also goes into architecture in detail, but the emphasis is somewhat different. The basic approach is that of "cost/benefit." Five different matchers (or matcher stages) with increasing complexity are considered. The relative processing cost (in terms of time and computer resources) as well as match accuracy is estimated for each matcher. The cost goes up progressively, so that the fifth matcher has a cost of about 200 times the first. Hara's presentation then goes on to outline a cost/benefit model for selecting the most cost-effective combination of matchers. In general this appears to be an excellent approach for identifying optimal combinations of matcher stages; but at a time of rapidly declining hardware costs, the "cost" of a stage might decrease drastically in less than a year.

Two presentations were directed toward architecture algorithms for implementing level-3 feature, Jain and Chen (P-19) and Wasserman (P-20). Chen presented some interesting data for implementing level-3 features, including some results using score fusion. The data presented shows about 10 % increase in performance. The paper also shows that going up in resolution from 500 pixels per inch (ppi) [19.7 pixels per mm (ppmm)] to 1000 ppi [39.4 ppmm] adds about 5 % to accuracy.

Wasserman's presentation (P-20) is concerned specifically with a Support Vector Machine (SVM) architecture. The presentation provides a good general introduction to SVMs. These are the modern descendants of neural nets, and are presently receiving considerable attention. It appears that the underlying theory of these machines is better developed than was the case for neural nets. The paper presents some work using sweat pores as features. The results obtained were impressive (100 % correct classification), but the test set was too small for strong conclusion. Also of interest is the fact that 500 ppi appears to provide adequate resolution for using pores under some conditions.

Two presentations, Garris (P-27) and Grother (P-25), presented top-level architecture diagrams showing how to structure latent matcher tests. Garris presents several diagrams showing the fusion of human and machine data. Figure 5, adapted from Garris, illustrates an architecture in which human data is merged with machine-generated data so as to boost performance.

Branchflower (P-11) provides some interesting diagrams for WAN-based matcher network systems.

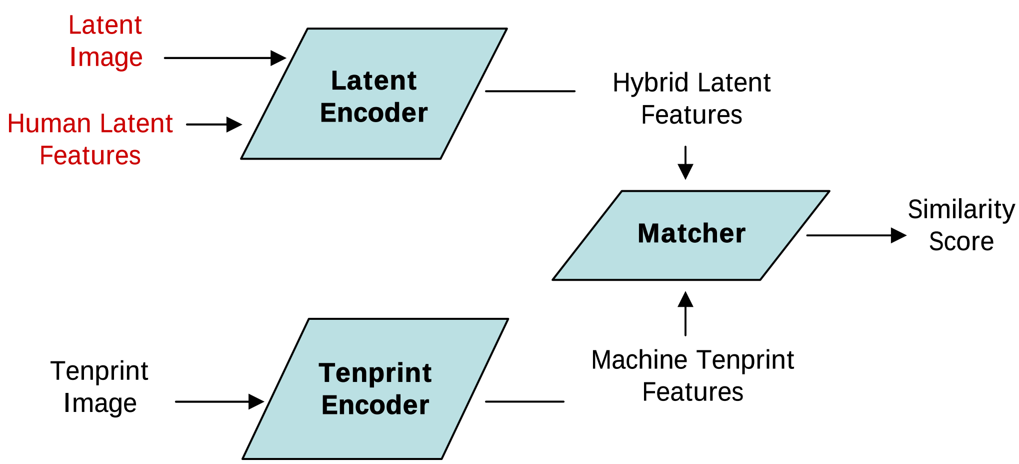


Figure 5. Hybrid Latent Encoding

3.5 Current Performance of Latent Matchers

A number of papers discussed the performance of present-generation latent matchers.

The presentation by Swann (P-6) provided insight into the performance of the FBI/IAFIS. The results confirm the tremendous importance of image quality. For example, when employing a mix of image qualities representative of actual case work (Test Set SD-27) the hit rate was 54 % when matched against rolled impressions, and 39 % when matched against plain impressions (flats). When using the higher quality New York DCJS latent images a hit rate of 94 % was obtained against rolled, and 63 % against flats. Even higher performance was obtained using Secret Service data: 97 % against rolled and 71 % against flats.

The presentation by Suman (P-10) also gives some performance data, but does not provide actual reliability figures (although providing references). The presentation includes the interesting statistic that out of about 100,000 monthly searches, 4000 are positively identified, for an "identification rate" of 4 %. (Although these numbers appear to be low, they are actually fairly good; most latent prints may be expected to belong to younger perpetrators who are not yet in the criminal master file.) The "identification rate" for latent searches by the FBI/LFPU is about 2 %.

A number of vendor presentations also provided some performance data. The vendor data seem to support the idea of 70-80 % reliability on "good" latents when matched against a combination of rolled and flats. Table 3 summarizes the results.

Performance of Current-Generation Latent Matcher			
Quality of Latents	Type of Mate		
	Rolled	Plain	Mixed rolled/plain
Good and Better	94 %	63 %	78 %
Average Case Work	54 %	39 %	47 %

Table 3. Representative latent matcher performance on a large background (40+ million subjects)

3.6 Projected Performance of Latent Matchers

A number of vendors, for example Fondeur (P-12) and Briefs (P-13), presented data on measured performance of semi-lights-out systems. The vendors are cautious about presenting specific numbers (understandably so), and tended to provide "delta numbers" or else employ graphs without numerical scales. Nevertheless a good deal can be inferred from their data. Two separate performance losses are identified, "front-end" and "back-end." Front-end losses result from automated feature extraction done by the machine. For example, when using "good" latents only, it appears that performance decreases by about 20 % when using machine-encoded features versus human-encoded features.

Back-end losses result from the need to restrict the candidate list to only those candidates the computer considers "probable hit." This is essential, otherwise the human is overwhelmed by the sheer number of candidates that need examination. Good back-end processing can potentially reduce the candidate list by a factor up to one hundred or so. According to the vendor's projections, the additional performance drop incurred by "backend" processing is about 15 %.

It is possible to make some very rough, order-of-magnitude, performance projection based upon the available data. We make the following assumptions:

- Latent images to be searched are all of "good" quality or better;
- Searches are performed against a very large background (about 50 million subjects);
- The mates (in the background) consist of a mix of rolled and plain impressions
- Feature extraction will be automated;
- The number of candidates presented to the human will be a small number (say 3 times the number of expected hits)

We perform our estimates in two independent ways. In the first method we start with the IAFIS performance, which assume a large background and a mix of plain and rolled mates. However, these IAFIS numbers apply to human encoding and humans verifying a large number of candidates (say ten for each search). We need to adjust for this. Thus, we begin with a 78 % reliability, subtract 20 % for auto-encoding, then another 15 % for backend candidate reduction. The result is a net value of 43 % for the semi-lights-out reliability. This is summarized in Table 4.

Semi-Lights-Out Performance (Estimated from IAFIS Data)	
Action	Resulting Value
Begin with “human” assist value	78 %
Subtract 20 % for auto-encoding	58 %
Subtract 15 % for backend processing	43 % (final estimate of accuracy)

Table 4. Estimated semi-lights-out performance extrapolated from IAFIS data

In the second method we begin with vendor-supplied reliability for auto-encoded search against a medium background (2.5M fingers) and no backend processing. We then make two corrections. The first is for backend processing (automated candidate reduction) and the second is for the small size of the background. In making the background adjustment we will assume a 1.5 % performance drop for each factor of two increase in the size of the background. The results are summarized in Table 5.

Semi-Lights-Out Performance (Estimated from Vendor Data)	
Action	Resulting value
Begin with auto-encoded value on medium background	71 %
Subtract 15 % for backend processing	56 %
Subtract 12 % for background increase	44 %

Table 5. Estimated semi-lights-out performance extrapolated from vendor data

Interestingly, both methods appear to lead to values in the low forties. This may appear somewhat low, but is in fact quite respectable for a preliminary screener. However, it must be kept in mind that these numbers are very preliminary.

3.7 Latent Quality Measures

Among the topics were given emphasis in the workshop program were “Latent Quality Measures.” The purpose of a quality measures is to assess the suitability of a fingerprint for some specific purpose, for example automated searching [23][24][25]. The quality measure algorithm might, for example, assign a value between zero and unity to an input fingerprint image, depending upon the algorithm’s assessment of how suitable the particular fingerprint is for automated searching. Assuming that 1.0 indicated the highest possible value, a computed value of 0.8 to 1.0 would indicate “very good;” a value between 0.6 to 0.8, “good;” 0.4 to 0.6, “fair;” 0.2 to 0.4, “poor;” and a value less than 0.2, “unusable.” The computed score can then be used to predict the success rate of a search. It can also be used to screen out fingerprints of very low value, as these would have a negligible success rate, and would waste system resources.

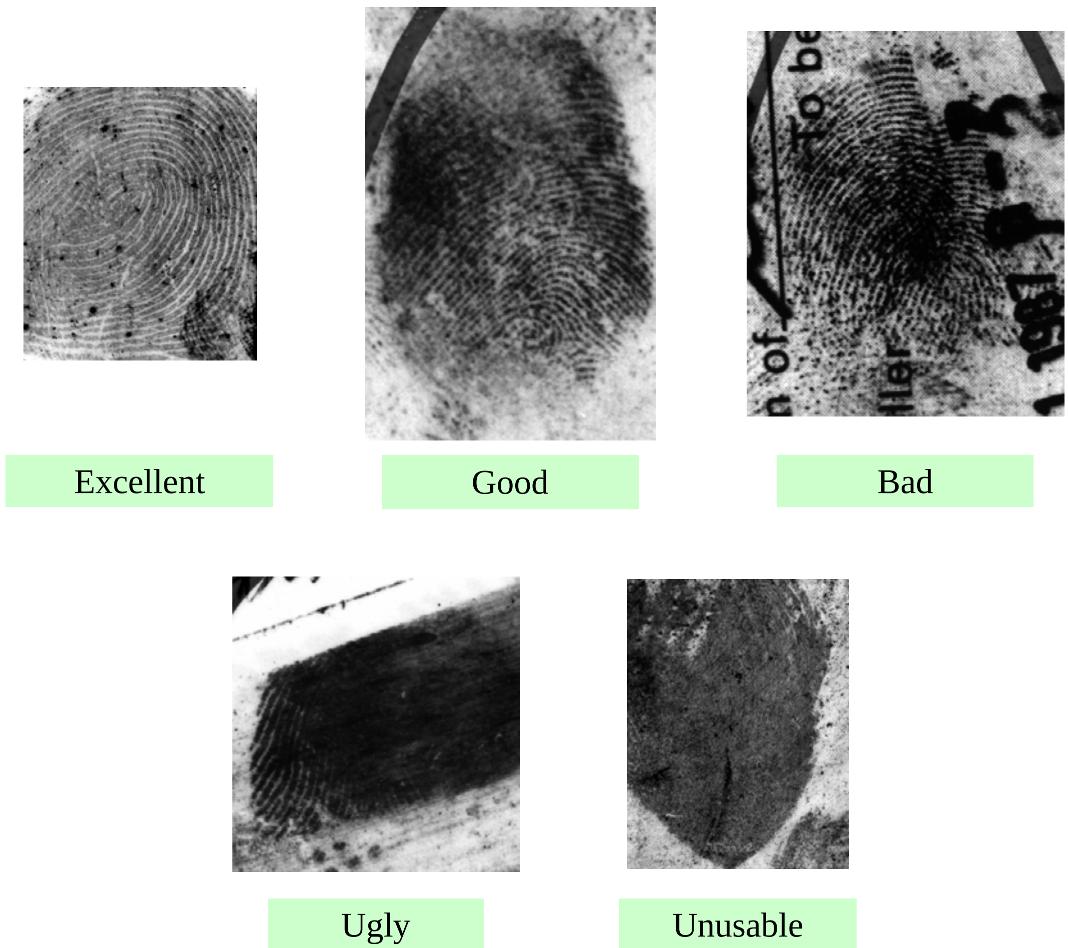


Figure 6. Representative latent fingerprints showing wide range of quality

Three presentations had as their primary topic “latent quality measures:” Tabassi (P-21), Wertheim (P-22), and Hicklin (P-23). The presentation by Tabassi stressed lessons learned in developing the NIST Fingerprint Image quality (NFIQ) algorithm. The NFIQ algorithm is geared toward flat fingerprints and uses a scale of 1 to 5, with 5 being **lowest** quality. Experiments have been performed to correlate the NFIQ with 1) the True Acceptance Rate (TAR) and 2) the False Acceptance Rate (FAR). (A high TAR is indicative of a good ability to recognize a true mate, and generally means good performance. A high FAR means many false candidates, and is generally indicative of poor performance.)

Experiments at NIST verified that a high correlation exists between the computed NFIQ value and the matcher False Alarm Rate (FAR). The higher the NFIQ value, the lower the quality, and the higher the FAR. (Recall that FAR means False Identification Rate or False Alarm Rate.) Similarly, a low NFIQ score indicates good quality, and generally produces high TAR.

It is believed that standard fingerprint quality measures such as the NFIQ algorithm cannot be simply transferred in toto to latent images. For one thing, since latents contain considerably less information the quality assessment needs to be more delicate. A "missing" portion of the fingerprint image is a relatively routine occurrence, and by itself does not indicate low quality. Also, even though most of the area of a fingerprint is blurred, the existence of a sufficiently large clear area might render the latent usable. Finally, since "semi-lights-out" matchers are immature, we do not have a clear understanding what their "preferences" might be. The presentations by Wertheim and Hicklin looked into the types of additional information that may be considered in quantifying latent quality. Wertheim defined the "Elements of Latent Difficulty" that he uses in training latent examiners [P-22, slide 2]:

- Quantity – percent or area of full print present
- Clarity – Focus, resolution, detail etc.
- Contrast – % of Grayscale range utilized, Average (shift), etc.
- Pressure – Vertical pressure, ridge to furrow thickness, etc.
- Slippage – Lateral pressure, ridge flow distortion, smearing, etc.
- Background – Substrate distortion, texture interference, etc.
- Focal Points – Core, Delta, Occasional features, Major flow convergence/divergence, etc.

Hicklin (P-23) underscored the issue that quality depends on the use of an image, and that characteristics that would constitute poor quality for an automated search may allow for great distinctiveness when performing human comparisons. The relation of quality to use is especially true for latents, because localized quality values can be used to improve the matching process. Hicklin defined quality metrics in terms of levels:

- Overall quality metrics (like NFIQ) provide an overall assessment of quality, tuned for a specific use
- Representation-based metrics (like the count of high-quality minutiae) are measurements of how well a given feature extraction task could be performed
- Feature-specific and localized quality metrics are assessments of the quality at a given point or small area within an image (such as the quality for a specific minutia)

Discussions during the workshop introduced an additional complexity. It became clear that one needs to consider two types of latent quality measures, one geared toward *machine searche*s, (machine oriented) and second toward *court evidence* (human oriented). These may require two separate paths of development. It is likely that NIST would initially concentrate on the first type.

3.8 Proposed New Features

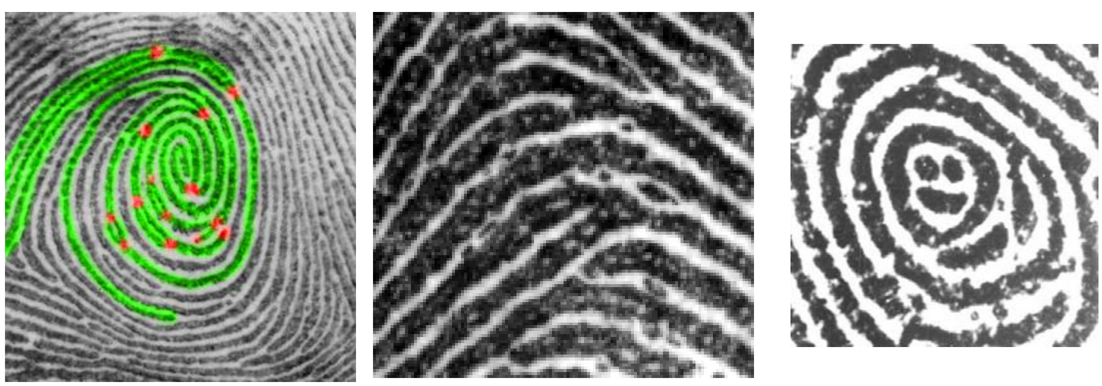
Another special topic covered in the workshop was that of new features. These are features not currently used, but proposed as future performance enhancers Four presentations had new features as their primary topic: Hicklin (P-18 and P-28), Jain and Chen (P-19), Wasserman (P-

20). At least two other presentations included new features as secondary topics. The ANSI/NIST Committee to Define Extended Fingerprint Feature Sets (CDEFFS) (P-18 and P-28) is in the process of defining a new ANSI/NIST standard for such features.

Within the current context, “features” basically mean information that will assist in either the search, or the verification of a candidate; of course the source of these features is the original fingerprint image. From an information theoretic standpoint, extracting features may be considered to be a type of data compression, the features themselves defining the compressed data. In this context, we should recall that data compression usually results in information loss; though in a data-rich environment some form of compression is essential.

One may define three categories of features: 1) features which assist in the search process – these are of primary interest; 2) features which do not assist in the search process, but are useful in ruling out a candidate (negative evidence) – these are of interest especially in connection with a semi-lights-out system; and finally 3) features which are (currently) only of interest to human experts and cannot be used by machines. Though important, these are of less interest in the present context.

Features are traditionally divided into three levels: 1) level-1 is the ridge-flow classification level; 2) level-2 is the minutiae level; and level-3) is the ridge detail level (dots, pores, protuberance, edge shapes). Recently is has been proposed that there should be a level-4, consisting of 3D features. The presentation by Hicklin (P-18) briefly covers 3D features. Examples of “new features” are found in Figure 7.



Ridge Path Analysis (Augmented Level-2 Feature)

Dots and Short Ridges (Examples of Proposed Level-3 Features)

“Happy Faces” are **not** Proposed New Features

Figure 7. Examples of newly proposed features

Many of the newly proposed features are level-3 features, for which Figure 7 (middle) provides some examples. The presentation by Jain and Chen (P-19) provides a detailed discussion of how these features might be used for search enhancement. Wasserman (P-20) focuses on using pores as the level-3 feature, and in addition employs a novel architecture. As features, “pores” while interesting and promising are not without problems. For one thing “pores” might require higher resolution (usually at least 1000 ppi) for best performance. This is a problem since most present day databases (files or galleries) have been scanned at 500 ppi. This may reduce the utility of pores in the near future, until higher resolution files become common. A second problem is that pores only manifest themselves in a minority of cases, even when scanned at high resolution. By some estimates less than 10 % of fingerprints exhibit good pore data. The reason appears to be that some degree of perspiration must occur for pores to manifest themselves strongly. It would appear that pores by themselves may only provide moderate performance gains.

Similar restrictions apply to many of the other level-3 features. That is, the features might be sporadic in appearance and may require greater resolution than currently used in practice. Thus, while many of the proposed features exhibit promise, a sorting-out process will be required to determine their relative effectiveness. This might require several years.

The "happy face" in the above figure was included for several reasons. The first of course is whimsy. But there are also more serious ones. A latent examiner seeing the "happy face" in one print and not in its purported mate might dismiss the "match." (Assuming that the central portions of both images are clear.) This illustrates the power of exclusionary features.

There are many strange characteristics/features in fingerprints which "jump out" at a human, but are unfortunately very difficult to program into a machine. This brings us to our third point: enhancements of computer performance are often better accomplished by following a path for which the computer is uniquely suited, rather than trying to mimic a human.

3.9 Test Sets

The presentation by Wood (P-26a) summarized the NIST in-house latent test sets. There are two: SD-27 and SS-1000. The first contains 258 exemplars with a wide range of quality. This data set is available with marked-up minutiae on the latent, as well as the matching minutiae identified on the ten-print. Although in most respects this is an excellent test set it has two shortcomings: 1) it is of rather modest size (258), and 2) since it is publicly available, vendors may be overly familiar with it (and may have over-trained their systems to this set).

SS-1000 is a much larger set containing about 1000 exemplars. However, this data has not yet been completely marked up (in the manner of SD-27), nor is it publicly available. In addition, since all the exemplars were obtained from AFIS hits (from latent submitted by the Secret Service), there is concern that the data are skewed toward "latents liked by FBI/IAFIS".

SD-27	
Database Size	258 latent prints from 233 subjects
Origin	Selected by FBI from FBI casework
Purpose	Originally selected to test IAFIS; later became an open-source test set for AFIS systems development.
History	Nominally 300 prints selected from FBI case files. Divided into three groups: 1) 100 "good", 2) 100 "bad" (representative of typical case work), and 3) 100 "ugly" (more difficult). Later reduced to 258 prints

	for various reasons.
Size of images	Type 13 records of varying size (TBD)
Mates	233 criminal ten-print cards obtained from CJIS, Clarksburg, WV. Type 14 records.
Compression	Latent uncompressed; ten-prints compressed via WSQ
Ancillary Data	Complete “mark ups are available in two forms: 1) all apparent minutiae on latent (“ideal minutiae”), and 2) all minutiae having matching minutiae on ten-print (“matched latent minutiae”). These are available as type 9 records, as are the rolled images.
Comments	Some of these are quite difficult, particularly in a “lights out mode.” Publicly available.

Table 6. Summary of SD-27 test set

SS - 1000	
Database Size	1000 latents corresponding to about 600 ten-prints
Origin	Secret Service
Purpose	Unknown (presumed for testing)
History	Selected by Secret Service form cases identified by IAFIS (operational data, 2001-2004)
Size of images	Type 13 records of varying size (TBD)
Mates	Criminal ten-prints obtained from CJIS
Compression	Latent uncompressed; ten-prints compressed via WSQ
Ancillary Data	Unknown (TBD)
Comments	Generally much easier than SD -27. May be more suitable for “lights out mode.” Not publicly available.

Table 7. Summary of SS-1000 test set

The presentation by Dvornychenko (P-26b) outlined how NIST proposes to remedy the identified shortcomings of our data. First, we need to acquire more exemplars. Second, we need to employ more varied data sources. This will prevent the skewing of data toward large agencies and/or large AFIS systems. It has been suggested that as few as 50 exemplars from (say) ten smaller law enforcement agencies would help balance the composition of the data. The total number of exemplars in the "final" test set might be on the order of 5,000. Based upon feedback obtained in the workshop, it appears that a test set of this size would not be difficult to put together. Several very promising sources were identified, and are being actively pursued.

The subject of "special test sets" was also raised in a number of presentations, including those by Meagher (P-1), Suman (P-10), Grother (P-25), and Dvornychenko (P-26b). The purpose of these "special test sets" is to identify problems which might otherwise escape notice. Problems might escape notice for any number of reasons, including: 1) the condition causing the problem is not identified, 2) there are too few exemplars with this condition, and 3) the condition occurs in conjunction with other factors which mask it.

The kinds of fingerprints that would go into the a special test set are: 1) latents developed or lifted by the various common processes in use (e.g., chemical, fluorescent, photo, etc.) and preferably from the same subject; 2) latents with special pattern classes (e.g., high curvature whorls, accidentals, etc.), 3) fingerprints exhibiting graded degrees of shear (elastic deformation) from low to very high. As mentioned, such a database could be used for discovering hidden problems and hence improving performance. It is probable that to acquire such a database would require the use of volunteers.

4. Conclusions

We now come to an important topic: How should NIST follow up the workshop initiative in creating a "latent challenge?" Our basic approach is four-pronged:

1) We will collect, electronically scan (digitize), categorize and archive latent fingerprint images along with their mates. We are currently actively engaged in this process and plan to expand the scope of the effort. In the process of scanning latents we also plan to selectively collect palmprint images.

2) We will continue to refine our proposed Application Program Interface (API) for latent fingerprint matching software in concert with other agencies. Our experience in fingerprint testing performed at NIST, together with lessons learned during the workshop, will guide us in putting together the proposed standards (API) for the software. We fully expect this to be an iterative process in which we would issue a preliminary specification for review and comments, followed by a revised version. The process might well require several iterations.

3) We will revisit methods of measuring the performance of latent matchers. We received a number of comments during the workshop along the lines that NIST is overly reliant on ROC type performance measurements, and that these are sub-optimal for latent searches. We will therefore conduct a review of theoretical alternative performance measurement concepts and issue the findings .

4) We will continue to assist groups that are actively engaged in defining new features for computer matching. We will make assessments as to which of these features might reasonably be included in the matching software submitted for testing, and what special provisions (if any) need to be made.

References

[1] NIST Special Publication 500-245, Information Technology: American National Standard for Information Systems –Data Format for the Interchange of Fingerprint, Facial, & Scar Mark & Tattoo (SMT) Information, September 2000. ➔ http://fingerprint.nist.gov/standard/index.html

[2] F.L. Podio, et al, "Roles for the National Institute of Standards and Technology (NIST) in accelerating the Development of critical Biometric consensus Standards for the US Homeland Security and the Prevention of ID Theft," 2003. ➔ www.itl.nist.gov/div893/biometrics/documents/White%20Paper%20Roles%20for%20NIST%20FP&MH%20March11%202003.PDF

[3] NIST Image Group's Mugshot / Facial Research (Website) ➔ http://www.itl.nist.gov/iad/894.03/face/face.html

[4] Iris Challenge Evaluation (Website) ➔ http://iris.nist.gov/ice/

[5] "Use of Technology Standards and Interoperable Databases With Machine-Readable, Tamper-Resistant Travel Documents," a joint report to Congress from Department of Justice, Department of State, and National Institute of Standards and Technology.

[6] NIST Image Group's Patriot Act Research (Website) ➔ http://www.itl.nist.gov/iad/894.03/pact/pact.html

[7] C.L. Wilson, M.D. Garris, C.I. Watson, NISTIR 7110, "Matching Performance for the US-VISIT IDENT System Using Flat Fingerprints," May 2004. ➔ ftp://sequoyah.nist.gov/pub/nist_internal_reports/ir_7110.pdf

[8] S.S. Wood and C.L. Wilson, NISTIR 7112, "Studies of Plain-to-Rolled Fingerprint Matching Using the NIST Algorithmic Test Bed (ATB)," April 2004. ➔ ftp://sequoyah.nist.gov/pub/nist_internal_reports/ir_7112.pdf

[9] Biometrics Technology: Identification and Verification (Website) ➔ http://www.itl.nist.gov/iad/programs.html#BIOMETRICS

[10] ITL Biometrics Resource Center Website: About Biometrics ➔ http://www.itl.nist.gov/div893/biometrics/about.html

[11] Biometric Authentication Technology: From the Movies to Your Desktop ➔ www.itl.nist.gov/div893/biometrics/Biometricsfromthemovies.pdf

[12] B.S. Swann (FBI/CJIS), "Needs and Applications of Latents at FBI/CJIS," presentation at Latent Testing workshop, NIST, April 2006. ➔ http://www.itl.nist.gov/iad/894.03/latent/workshop/proc/P6_Swann_LatentOverview_2.pdf

[13] Information Access Division (IAD) Webpage ➔ http://www.itl.nist.gov/iad/

[14] NIST Image Group's Fingerprint Research (Website) ➔ http://fingerprint.nist.gov/

[15] Fingerprint Vendor Technology Evaluation (FpVTE) (Website) ➔ http://fpvte.nist.gov/index.html

[16] SDK Tests Website ➔ http://fingerprint.nist.gov/SDK/index.html

[17] MINEX04 Website ➔ http://fingerprint.nist.gov/minex04/index.html

[18] Ongoing MINEX (Website) ➔ http://fingerprint.nist.gov/minex/index.html

[19] S.S. Wood, et al, NIST, "Latent Test Sets: Current Inventory," presentation at Latent Testing workshop, NIST, April 2006. ➔ http://www.itl.nist.gov/iad/894.03/latent/workshop/proc/P26a_S_Wood_ssw_latent_data_v2.pdf

[20] Latent Testing Workshop (Website) ➔ http://www.itl.nist.gov/iad/894.03/latent/

[21] E. Tabassi, G. Quinn, P. Grother, "When to Fuse Two biometrics," Computer Vision and Pattern Recognition, Biometrics Workshop, NY, 2006.

[22] B. Ulery, A. Hicklin, C. Watson, et al, NISTIR 7346, "Studies of Biometric Fusion," Sept. 2006. ➔ http://www.itl.nist.gov/iad/894.03/pact/ir_7346.pdf

[23] NIST Fingerprint Image Software (NFIS2) ➔ http://fingerprint.nist.gov/NFIS/index.html

[24] NIST Fingerprint Image Quality (Website) ➔ http://fingerprint.nist.gov/NFIS/index.html#nfiq

[25] Biometric Quality Workshop (Website) ➔ http://www.itl.nist.gov/iad/894.03/quality/workshop/index.html