Successful Design of Biometric Tests in a Constrained Environment

V. N. Dvornychenko
NIST, Gaithersburg, MD
vdvorny@nist.gov

Abstract

The National Institute of Standards and Technology (NIST), with participation of the biometrics community, conducts evaluations of biometrics-based verification and identification systems. Of these, one of the more challenging is that of automated matching of latent fingerprints. There are many special challenges involved. First, since participation in these tests is voluntary and at the expense of the participant, NIST needs to exercise moderation in what, and how much, software is requested. As a result, it may not be possible to design tests which cover and resolve all possible outcomes. Conclusions may have to be inferred from studies that have limited results.

1. Introduction

The National Institute of Standards and Technology (NIST), with participation of the biometrics community, conduct evaluations of biometrics-based verification and identification systems. Several biometrics and modes are being looked at by NIST, including several types of friction ridge impressions such as rolled, plain and latent impressions (including palm); facial images; vein patterns; and iris prints. These biometrics may be examined singly, or may also be combined in synergistic combinations. However, in this paper we will restrict ourselves to one specific mode, namely latent fingerprints. The purpose of this paper is to show how limitations in control over the testing procedure may require special planning of the test, as well as interpretation of the results. References [1] and [2] provide some introductory material, which is also of historical interest.

Evaluation of Latent Fingerprint Technology, or ELFT (ref. [3]), conducted by the Information Access Division (IAD), Image Group, is dedicated toward advancing the state-of-the-art in latent fingerprint searches via: a) decreased dependence on human experts through greater automation; b) standardization of feature sets to facilitate data interchange; and c) standardized scores and performance measures. Milestones along the path to achieving these goals are periodic assessment of the state-of-the-art. These help gauge progress, and identify weaknesses. The diagram of Figure 1 provides a pictorial overview of the testing process.

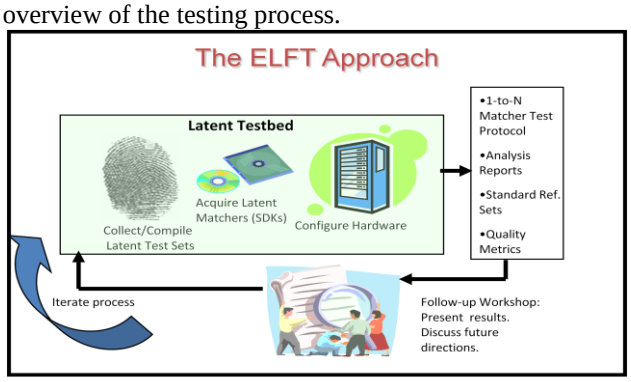

Figure 1

Two things are to be noted: a) the tests rely strongly on outside participation; and b) it is an ongoing process in that each test partially answers the questions proposed, but also may raise new questions requiring follow-on resting. Also, as mentioned, this a developing technology, so periodic testing is required to establish the state of the art.

There are special challenges involved in this kind of testing. First, since participation in these tests is voluntary and at the expense of the participant, NIST needs to exercise moderation as to what, and how much, is requested. This extends to requesting too many variants of the software, or to overly specifying the contents or the operation of any version. Second, NIST is obligated not to divulge any information which could be interpreted as proprietary to the contributor. Thus, should in the course of testing certain technical details reveal themselves, NIST must exercise care what is publically reported. Finally, each test may require great computer resources and execution time. Therefore, the total number of sub-tests must be limited.

These restrictions pose significant challenges regarding the design, planning, execution, and interpretation of the test as well as the test results. Since an important desired outcome is to provide feedback to the community, it is imperative to design the tests so that useful conclusions can be drawn without fully understanding of the details of the software. Considerable

effort has been devoted to the optimal planning and execution of such tests, and these ideas have been tested in several series of tests. However, this is still a developing discipline.

2. Overview of Latent Fingerprint Matching

The following figure illustrated a latent fingerprint on the right, and its prepared mate on the left. (The fingerprint on the left is called a rolled-impression, because the finger is gently rocked, or "rolled," during acquisition. These "prepared" or "controlled" mates are taken under carefully monitored conditions.)

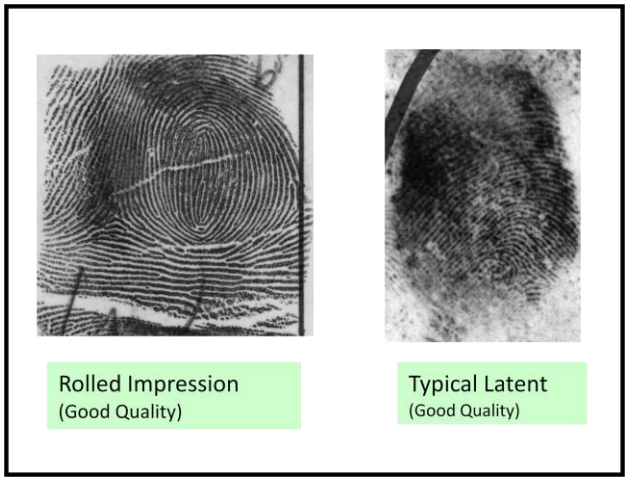

Figure 2

The latent fingerprint itself is typically referred to as the "search," "unknown," or the "probe." Each latent will be searched against a database (often called a "gallery") of potential mates. The fingerprints comprising the database are typically referred to as "mates," "knowns," and also sometimes as "exemplars." Typically, these fingerprints were acquired in a controlled environment from cooperative subjects, and whose identity is known (whence "knowns"). (In Figure 1, both types – "unknowns" and "knowns" -- are encompassed in the call-out "Latent Test Sets.")

The typical task performed by a matcher is to search a database containing many thousands -- potentially millions -- of fingerprints of the type shown on the left. Reduced to its basics, the search process consists of comparing each fingerprint in the database with the latent "search" fingerprint. Each comparison results in a matching score. The numerical scale of the matching scores is arbitrary, but a higher score indicates a better match. A final candidate list containing the highest "n" scores encountered is then output, together with the identification index of the mate which produced this score. . (A typical value of "n" (candidate list length) might be 10 or 20; but for analysis purposes a much longer candidate list, say 100, might be employed.)

In actual practice, e.g., criminal searches, it often happens that there is no mate in the database. Thus while there might be a candidate list, all candidates are false mates. For performance testing, a mate is almost always present, except in a few cases, where it is intentionally left out for control purposes.

For a search to be considered successful, the correct mate must appear on the candidate list, preferably in the top (first) position. Even with relatively good data this might not always be the case, and there are several reasons for this. The first, already mentioned, is that an actual mate might not be contained in the database. A second reason is that the search fingerprint might be of such low quality (smudged, fragmentary) that it could not be sufficiently differentiated from the many other fingerprints in the database (all of which are non-mates). This results in "impostors" outscoring the correct mate. In severe cases, high-scoring impostors can be so numerous as to crowd the correct mate completely off the candidate list. Finally, there might be algorithmic shortcomings in the matching algorithm (software), resulting in an abnormally low score being computed when matching against the true mate. (The above types of problems are quite common, and even a good algorithm will occasionally miss what should have been an easy "hit.")

3. Feature Extraction

Contrary to what might be expected, the matching (comparison) process does not typically involve comparing the images directly. Instead, it relies on comparing certain "features" which have been extracted from the images. These features are certain detailed characteristic of the fingerprint. These must be discernible in the fingerprint image, and ideally they must be unique to that fingerprint (at least in their aggregate). The features are extracted in a separate and distinct process prior to the actual search/matching.

To successfully search through a large database containing many thousands of fingerprints it is necessary that sufficient information be extracted from both the search and the database "known" so as to be able to differentiate the correct mate from every other fingerprint in the database.

There are many types of features that can potentially be extracted from the image. The traditional (mainstay) features are the so-called minutiae. These are illustrated in the top portion of the following diagram, Figure 3. (Going back to Figure 2, careful scrutiny of the rolled impression (left part of Figure 2) will reveal a large number of minutiae of both types.) Minutiae by themselves (possibly

augmented by core, delta, and pattern class) are enough for a successful search in the case of ***high quality*** fingerprint images. Methods for gauging performance are covered in [4] and [5].

To ensure best performance, a fingerprint should be both clear, and covers a large part of the finger. However, for latent fingerprints high quality is the exception rather than the rule. There are, therefore, many cases where minutiae by themselves provide insufficient information.

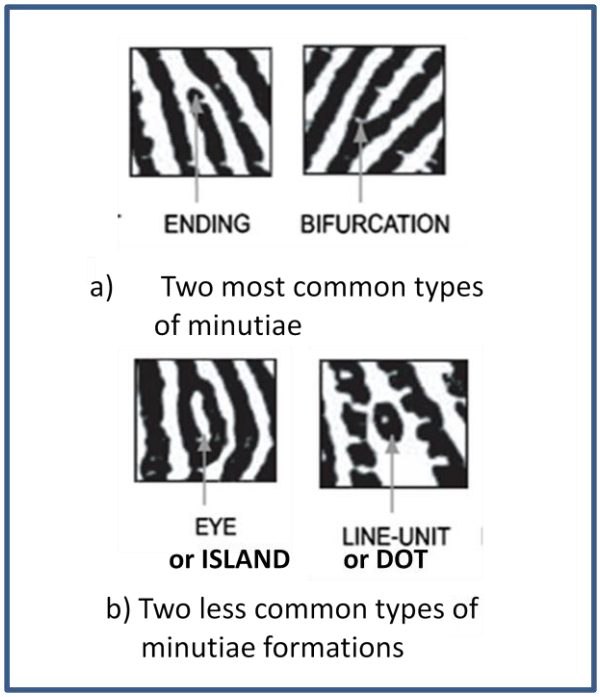

Figure 3

4. Extended Features

For cases where insufficient minutiae (say less than about eight) are found in the latent/search fingerprint it might be possible to augment these with additional and different features. The increase in the net amount of discriminatory information might then be sufficient to identify the true mate – and turn what would have been a "miss" into a "hit." There are many choices for such additional or "extended" features; two are shown at the bottom of Figure 3. A much more comprehensive list is found in the CDEFFS report, ref. [6]. We refer to these features as "Extended Features," "Extended Feature Sets", or simply as EFS.

Recall that the purpose of using extended features is to improve search performance by augmenting the available information. Whether a feature is ideally suited for this depends upon several factors. First of all, it cannot be too commonplace, or it provides little discriminatory information: to say that a ridge passes through a certain point by itself provides very little information. On the other hand, many features are rather uncommon, and possibly only a very small minority of fingerprints exhibits these features. Being so rare, it might be unproductive to include software to extract and match on this feature. (The "island" of Figure 3b may be such a case.) Finally, there are features such as sweat pores (the opening of sweat glands), which while quite common – in fact ubiquitous – might not actually show up consistently and strongly in every print. Finally, a feature might be difficult to define and capture due to its complex nature (e.g., bumps on ridges). Because of these complications, extended features are not routinely included among the traditional/mainstay features.

NIST desires to test the efficacy of the proposed EFS. First, we would like to know how much adding the entire set of extended features can potentially increase performance. Secondly, we would like to estimate the computational load incurred by adding EFS. Lastly, we wish to know if there exists a small subset of the features which provides most of the benefit.

To answer these is not as straightforward as it might first seem. As suggested previously, searching using EFS alone (and not including traditional features) may not be productive, because too low a percentage of true mates might be place on the candidate list. To circumvent this, a procedure was defined in which: a) the test-set is first searched using traditional features only, and then b) the same test-set is searches when EFS is included, with the traditional features retained. For this scheme to be successful, there needs to exist a careful balance between the amount of information provided by each class of features.

The following series of conceptual schematics is intended to lay out the principal alternatives regarding, a) the amount of information required, and b) the amount of information available. Consider Figure 4a. The larger region at the bottom is intended to convey the amount of information provided by the traditional features alone. Above this region are suggested the extended features (EFS). Together these two regions illustrate the total amount of available information for discriminating between fingerprints. The superimposed circular region, labeled "minimal information for matching," depicts how much information is required to successfully identify the true mate (and thus differentiate it from every other fingerprint in the database.) For the case of Figure 4a we note two things: a) first, the traditional features provide more than enough information for a successful search; so that b) including EFS would make little or no impact on performance.

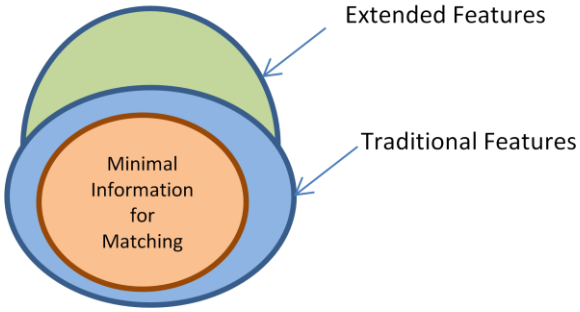

Figure 4a

The next figure differs in that now traditional features alone may no longer have sufficient information (depicted by the fact that the circular region no longer fits within the "Traditional Features" region). It is now necessary to add the extended features to obtain sufficient information.

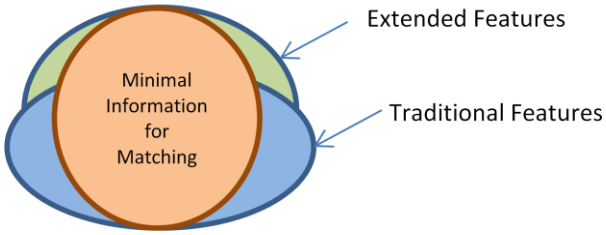

Figure 4b

The final figure of the series illustrates the case where the combined amount of information from both types of features are insufficient. In this case performance will be poor regardless whether extended features are included or not.

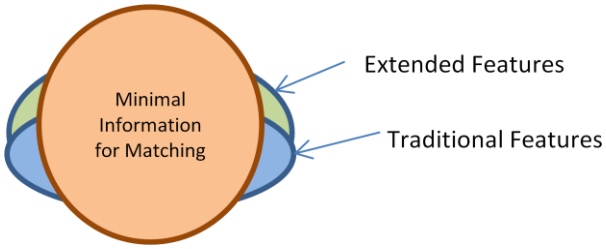

Figure 4c

We can summarize the three situations illustrated as follows: In the first case (Figure 4a) performance will be good; but this will be the case whether extended features are included or not, so that the benefits of extended features will not clearly emerge. In the second case (Figure 4b) performance will be poor if extended features are excluded, but good if they are included; so in this case the benefits of extended features will clearly emerge. In the third case (Figure 4c) performance will be poor in either event. For cases (a) and (c) the change in performance is difficult to measure, and in fact, may provide no reliable information. It follows that (4b) is best for measuring the performance increase.

5. Testing Extended Features

In testing latent fingerprint matching NIST relies heavily on the outside fingerprint community. NIST solicits software (from the outside community) in the form of Software Development Kits, or SDKs. The general form and function of these, and the input/output interfaces are specified by NIST through an Application Programming Interface, or API. (Despite the name, SDKs are not general tool-building kits to be experimented with. Rather, NIST must link the component parts following exact instructions provided by the submitter.)

As previously explained, gauging the improvements which accrue from including EFS entails the following steps: a) first conducting a test using only traditional features; b) rerunning this test when EFS is included; and c) computing the difference in performance, and assessing whether the differences are statistically significant.

As mentioned, NIST does not wish to burden the participants by overly specifying the contents and mode of operation of their SDKs. The compromise approach taken by NIST is that participants submit a form of their software which represents "best effort using image only." That is, the software uses only the fingerprint images as input, but the submitter is free to extract any features they desire from these images. It is fully expected that traditional feature are included among the features extracted, but there is nothing to prevent the submitter from including additional features, including extended features, or variants thereof. This "image only" software is taken for the baseline during testing. The submitter is requested to also submit a version which accepts, in addition to the images, a list of extended features which have been extracted by human experts in accordance with published standards.

As part of its policy of not overly specifying the software, NIST does not specify exactly ***how*** the additional information (features) should be used. For example: a) should these features ***replace*** any similar features already extracted from the image; b) should the features be used only if similar information have ***not*** already been extracted; c) should they be used only if it deemed superior to similar information extracted; and lastly, d) the submitter has the option of ignoring some or all the additional features.

Each test may require great computer resources and execution time. Therefore, the total number of all such tests run must be limited. This limits the number of "control" tests which can be run, for example intentionally including bad features to gauge the effect

The following figure shows the results of this type of testing, test, using five different SDKs.

6. Specimen Results

The following graph provides results of the type discussed in the previous section, although it does not focus on extended features per se. The results of five different SDKs/matchers are shown. For each case the leftmost bar represents the percentage of true mates placed in first (top) position when employing only the fingerprint image for input. For the second bar only certain preselected features may be used, and not the image. These features were extracted by human experts with the help of latent workstations. The third bar shows the result of using both the image and features as input. Finally, the fourth bar provides the expected performance based on simplified fusion theory.

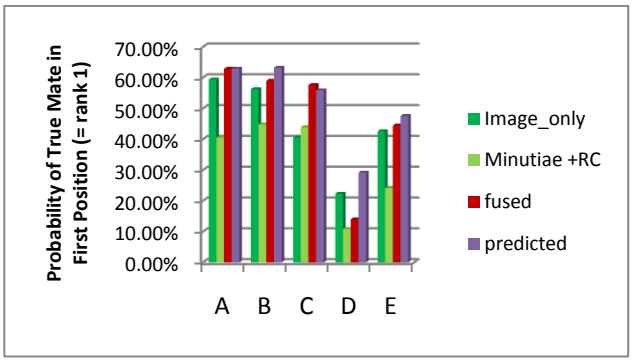

Figure 5

It will be seen that in four out of five cases using image only provides superior results to feature only. (The exception is case C, and here the feature-only result is only marginally better.) In all but one case, the combined result is superior to both the image and the feature results separately. (The exceptional case, D, may have problems processing the features input.) The last bar shows a theoretical estimate. In all but one case this is equal to, or slightly higher, than the actual results. Additional test results are covered in refs. [6-10].

7. Conclusions

The testing restrictions pose challenges regarding the design, planning, execution, and interpretation of the test results. Since an important desired outcome is to provide feedback to the community, it is important to design the tests so that useful conclusions can be drawn, even without a full understanding of the details of the software. Considerable effort has been devoted to how to optimally plan such tests. These ideas have been tested in several series of tests, but this is still a developing discipline.

We would like to thank the Department of Homeland Security's Science and Technology Directorate and the Federal Bureau of Investigation's Criminal Justice Information Services Division for sponsoring this work.

10. References

[1] Joseph H. Wegstein, NBS special Publication 500-89, "An automated fingerprint Identification system," February 1982

[2] Moore, R. T., "Automatic Fingerprint Identification Systems," chapter 6 of Advances in Fingerprint Technology, H. C. Lee and R. E. Gaensslen, Eds., Elsevier Science, 1991

[3] V.N. Dvornychenko, P. Grother, M. Indovina, "Concept of Operations (CONOPS) for Evaluation of Latent Fingerprint Technologies" NIST Publication (NIST website)

[4] Patrick Grother and P. Jonathon Phillips, "Models of Large Population Recognition Performance,"

[5] A. J. Mansfield and J. L. Wayman, "Best practices in testing and reporting performance of biometric devices," Tech. Rep., August 2002

[6] Data Format for the Interchange of Extended Friction Ridge Features
http://biometrics.nist.gov/cs_links/standard/CDEFFS_DraftStd_v05c_2010-05-10.pdf

[7] Final Report on Phase II Testing
http://fingerprint.nist.gov/latent/NISTIR_7577_ELFT_PhaseII.pdf

[8] NIST Latent Fingerprint Testing Workshop 2009, March 19 & 20, 2009
http://fingerprint.nist.gov/latent/workshop09/index.html

[9] NIST Evaluation of Latent Fingerprint Technologies: Extended Feature Sets, Evaluation #1
http://fingerprint.nist.gov/standard/cdeffs/Docs/ELFT-EFS_PreliminaryReport_DRAFT_2010-01-26b.pdf

[10] ELFT-EFS
NIST Evaluation of Latent Fingerprint Technologies: Extended Feature Sets -- Public Challenge Results
http://fingerprint.nist.gov/standard/cdeffs/Docs/ELFT-EFS_AppB_PublicChallenge.pdf