Studies of Biometric Fusion
Appendix A

Terminology, Experimental Design and Data Description

Brad Ulery,[1] Austin Hicklin,[1] Craig Watson[2]

[1] Mitretek Systems
[2] National Institute of Standards and Technology

29 June 2006

Abstract

This document describes the datasets and matchers used in these studies, and also their baseline performance prior to data fusion. The primary dataset is the multi-modal NBDF06 dataset, which contains fingerprint and face images from over 185,000 subjects; the FpVTE MST fingerprint dataset and the BSSR1 fingerprint and face dataset were also used. This document also summarizes the terminology used throughout these analyses.

Contents

1 Introduction

This document describes the datasets and matchers used in the various analyses in "Selected Topics in Biometric Fusion", and shows the baseline performance of the matchers on those datasets, before fusion techniques were applied.

This document also includes an overview of terminology as used in these studies.

2 Terminology

Biometric data derives from a ***population*** of subjects (persons). ***Biometric samples*** (generally images) are ***collected*** from each subject. Taken together, the population and sample collection process are a major determinant of overall performance. Therefore it is important to identify the ***source*** of sample data in a performance evaluation.

In ***multi-biometric systems***, more than one sample is collected per subject. The samples collected can come from multiple ***instances*** (such as different fingers), multiple ***modalities*** (such as face and fingerprint), or can be multiple samples of the same instance.

Samples are usually preprocessed to produce ***templates,*** which are representations of the information that facilitate efficient processing.

Samples from subjects previously encountered are included in the ***gallery*** (database). Subsequent encounters with subjects are known as ***probes*** (searches). Subjects that exist in both the probe and gallery are known as ***genuines*** (and the corresponding samples are ***mates***). Probes that do not exist in the gallery are known as ***impostors*** (the samples are ***non-mates***).

The ***accuracy*** of a biometric system is based on its ability to distinguish subjects previously known to the system (genuines) from subjects not known to the system (imposters). The ***comparison*** of a probe template and a gallery template produces a ***matcher score***, which is a measure of sample ***similarity***. The measurement scales and the distributions of the scores are unique to each system. In a perfect system, there would be no overlap between the genuine and imposter score distributions.

A ***decision criterion*** is applied to scores in order to ***classify*** pairs of subjects as genuines or imposters. Regardless of whether a decision is correct or not, a pair of subjects is said to ***match*** if the system classifies them as genuines; otherwise the pair is said to be a ***non-match***. Note that "match" is a determination, while "genuine" is ground truth, which may be unknowable. The decision criterion is often simply a ***score threshold***.

In multi-biometric systems, multiple scores per subject may be combined to better discriminate genuines from imposters. This process of combining scores is often modeled as having two stages called ***normalization*** and ***score-level fusion***. Normalization generally refers to a transformation of a univariate score distribution (such as rescaling), whereas score-level fusion is a function of multiple scores per subject. ***Decision-level fusion*** is a function of multiple decisions per subject.

Unfortunately, common usage of the terms normalization and score-level fusion makes simplifying assumptions about the underlying process that are not always valid. Often, normalization is based exclusively on score distributions as when scores are simply rescaled and recentered, or when transformed to probabilities or likelihood ratios. Sometimes, normalization, like fusion, refers to combining additional information, as when individual scores are adjusted based on information such as probe image quality, or gallery subject-specific score distributions. Depending upon how scores are combined in fusion, normalization is not always required.

Given multiple scores for each probe subject (i.e., prior to fusion), the decision criterion may be a higher-dimensional ***decision boundary***. For a given fusion method, the set of possible score thresholds corresponds to a family of decision boundaries.

A Receiver Operating Characteristic (***ROC***) curve plots the trade-off between classification errors that result from varying a score threshold. An ROC measures the ability of a system to discriminate genuines from imposters. The vertical axis of the ROC depicts the True Accept Rate (***TAR***) or equivalently the False Reject Rate (***FRR*** = 1-TAR). TAR is the fraction of genuines that are correctly classified. The horizontal axis of the ROC depicts the False Accept Rate (***FAR***), which is the fraction of impostors that are misclassified. Note that empirical measurements of an ROC are subject to statistical uncertainty and measurement errors.

An ideal ROC curve would include a point at TAR=1, FAR=0. Practically, a biometric system cannot achieve such perfect performance. When comparing two ROCs, one may be consistently superior (its TAR is higher at every FAR), or the two curves may cross. We say that one system is more ***accurate*** than another when its ROC is consistently superior. Many factors can be controlled to ***optimize an ROC***, such as the biometric collection process, the choice of matcher(s), and the method of fusion. System designers must additionally select an ***optimal threshold*** — a point on the ROC — for their design objectives and operational requirements. Optimization of the overall ROC is within the scope of this work, but selection of an optimal threshold is beyond scope. Determining appropriate thresholds for such decisions require knowledge of system scale, estimates of prior probabilities of genuine and imposter subjects, and risk/cost functions for false rejections and false accepts. These factors are system specific, and therefore such thresholds only serve as examples in these analyses.

3 Description of datasets

The datasets used in these studies came from three sources:

- ***NBDF06***: This operational dataset includes criminal arrest fingerprints and face images (mugshots) from more than 185,000 subjects. Three of the more accurate fingerprint matchers from the NIST SDK tests were used [SDK], as well as three recent commercially available facial recognition systems, which are required to remain anonymous. Most of our analyses use the NBDF06 dataset. The source of the NBDF06 dataset cannot be publicly released.
- ***FpVTE-MST***: The Fingerprint Vendor Technology Evaluation [FpVTE] Medium-Scale Test dataset contains 10,000 right index fingerprints from 5,600 subjects, from flat (single finger) or slap (four finger simultaneous) livescan devices; this study used one pair of mated fingerprints from each of 3,240 subjects. Fingerprint matching scores from fourteen of the participating systems were used in these studies.
- ***BSSR1***: The NIST Biometric Scores Set, Release 1 [BSSR1] contains face and fingerprint matching scores from two face and one fingerprint matcher. The BSSR1 dataset has the advantage of being publicly available, unlike the other datasets. BSSR1 is used only to a limited extent in these studies.

A detailed explanation of these datasets follows.

3.1 NBDF06 (NIST Biometric Data Fusion 2006) Data

The bulk of the empirical results reported in these analyses used matcher scores based on fingerprint and face data from this operational database.

3.1.1 Types of biometric samples

Each subject had one or more sets of face and fingerprint images. Each face/fingerprint set contained the following samples, collected at the time of the subject's arrest:

- ***Face:*** One frontal face image. The face images used are frontal, 24-bit color JPEG images compliant with ANSI/NIST-ITL 2000 image format specifications; the images are compliant with Best Practice Application Level 30 requirements [MugshotBP] except for size: image size is typically 384 x 480 pixels, smaller than the 480 x 600 minimum mandated by the Best Practices document. The images have controlled 3-point lighting, 18% gray backgrounds (with some exceptions), and uniform full frontal pose. The face occupies approximately 50% of the width of each image. The expressions are not controlled. A visual review of a sample of the face images shows that the images are fairly typical of recent mugshot photographs, and better than some databases such as BCC or US-VISIT (POE).
- ***Slap fingerprints:*** Slap livescan fingerprints from all ten fingers. Fingerprints were collected on FBI-certified livescan devices [FBI-Cert]. The four-finger slap images were segmented into individual fingerprint images using the NIST segmenter [NFIS]. Automated measures were used to identify probable segmentation failures, notably cases in which segmentation boxes touched or overlapped; these cases, comprising approximately 5% of the total, were excluded from the dataset. It would not be quite correct to regard these exclusions as failures to enroll (FTE) for several reasons:
  - These were probable segmentation failures, which are distinct from failures to enroll: failure to segment indicates a problem with the association of an individual fingerprint image with its finger position, and does not imply anything about the quality or content of the fingerprint features.
  - Today, slap segmentation is required at the time of collection in order to be compliant with FBI standards for identification slaps [EFTS-7.1]. The cases excluded in this evaluation presumably would have been flagged for recapture at the time of collection if the collection devices used had performed slap segmentation.
- The dataset as tested should be expected to contain some incorrectly segmented fingerprints, although small samples of manually inspected fingerprints (ones that resulted in low matcher scores) did not reveal segmentation failures; see [SlapSeg] for an evaluation of segmentation accuracy and a discussion of issues in slap segmentation.
- ***Rolled fingerprints:*** Rolled fingerprints were not used in these studies.

3.1.2 Number of subjects

Each subject had one or more face/fingerprint sets, as follows:

- NM = Non-Mated (imposter) subjects: 122,000 subjects had one face/fingerprint set each (a single encounter of each subject).
- M = Mated (genuine) subjects: 64,867 additional subjects had two face/fingerprint sets each (two encounters designated as M1 and M2). Of these, 4,015 subjects had three fingerprint sets each; these are the multi-sample subjects used in the analysis in Part X.

These sample sizes were limited by data availability and computing resources. Given these constraints, the samples (especially of mated subjects) were as large as possible to allow precise measurements at very high TAR, as anticipated from the use of high quality matchers and N-way fusion.

The association of mates was not biased through the use of matchers in selection: the mating information supplied with the dataset was determined through a combination of name-based (non-biometric) matching, and matching of the rolled fingerprints. The slap fingerprints and face images used in these studies were not used to identify the mates.

3.1.3 Matchers

Three fingerprint and three face matchers were used:

- The fingerprint matchers were three of the more accurate fingerprint matchers identified in the NIST SDK single finger and two-finger tests [SDK; SDK2]: matcher H, matcher I, and matcher Q.
- The face matchers were three recent (c. 2004-5) commercially available face recognition systems, which are required to remain anonymous.

3.1.4 Derivation of genuine and imposter matcher scores

As mentioned above (under "Number of subjects"), the sets of biometric inputs were:

- M1: 1 set of 10 slap fingerprints and 1 face image from each of the 64,867 mated subjects (M1=Mate set 1)
- M2: a second set of 10 slap fingerprints and 1 face image from each of the 64,867 mated subjects in M1 (M2=Mate set 2)
- NM: 1 set of 10 slap fingerprints and 1 face image from each of the 122,000 non-mated subjects (not in M1 or M2) (NM=Non-Mated set)

The similarity scores in the NBDF06 dataset follow the design developed for [SDK2]. Each set of scores consists of

- 64,867 genuine scores, corresponding to the diagonal elements of the M1 x M2 similarity matrix shown in Figure 1; and
- 122,000 imposter scores, generated by comparing each of the NM samples against one of the M1 samples, corresponding to the 2 "diagonals" of the NM x M2 similarity matrix shown in Figure 2.

Thirty-three sets of scores were produced by running face matchers A, B, C (one score set each) and finger matchers I, H, Q (one score set per finger). These are summarized in Table 1.

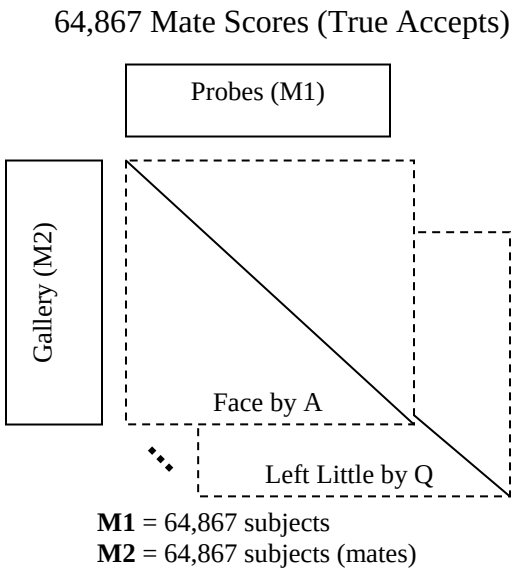

Figure 1: Derivation of genuine scores in the NBDF06 dataset. Diagonal elements of the M1 x M2 matrices provide the "true mates" (genuines) for construction of the ROC curves; off-diagonal comparisons were not performed.

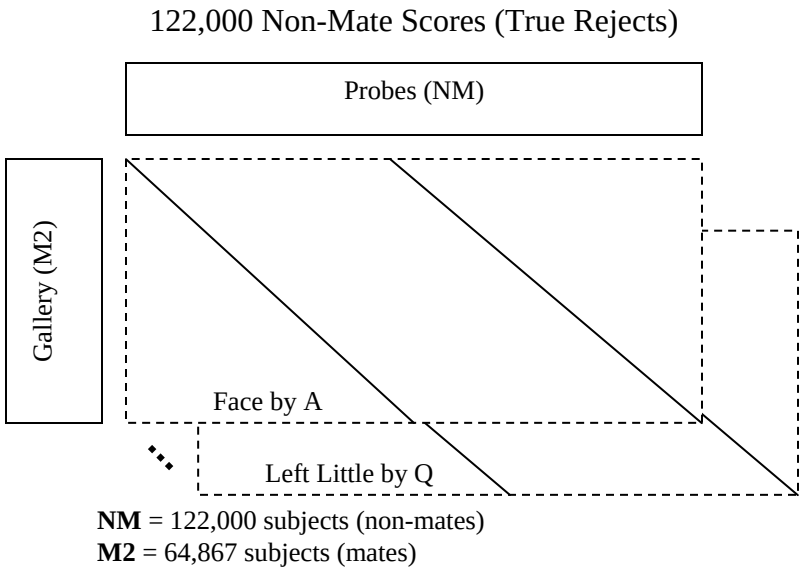

Figure 2: Derivation of imposter scores in the NBDF06 dataset. Diagonal elements of the NM x M2 matrices provide the "true non-mates" (imposters) for construction of the ROC curves. The diagonals represent the 122,000 imposter comparisons that were performed.

	A	B	C	H	I	Q
Face (frontal)	x	x	x			
Right Thumb				x	x	x
Right Index				x	x	x
Right Middle				x	x	x
Right Ring				x	x	x
Right Little				x	x	x
Left Thumb				x	x	x
Left Index				x	x	x
Left Middle				x	x	x
Left Ring				x	x	x
Left Little				x	x	x

Table 1: Thirty-three sets of 64,867 genuine and 122,000 imposter scores were generated from 186,867 distinct subjects, 251,734 distinct collection events (face/fingerprint sets), using 3 face matchers (A,B,C), and 3 fingerprint matchers (H,I,Q).

3.1.5 Fingerprint image quality

Figure 3 and Figure 4 show the distribution of fingerprint quality for the NBDF06 fingerprint data, using the NFIQ metric [NFIQ; NFIS]. 1.3% of all fingerprints were NFIQ 5; 4.8% were NFIQ 4 or 5. In 0.01% of the ten-finger sets, all fingerprints were NFIQ 5; in 0.22% of the ten-finger sets, all fingerprints were NFIQ 4 or 5.

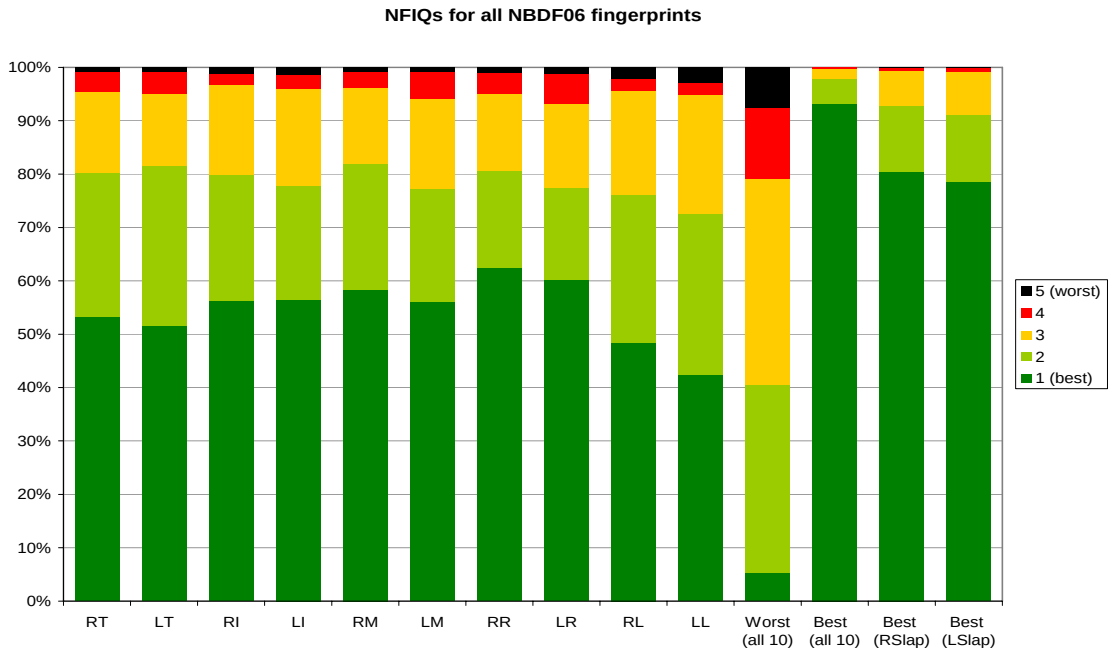

Figure 3: NFIQ results for all genuine and imposter fingerprints, including the best and worst fingerprint from each subject, and the best fingerprint from each slap from each subject.

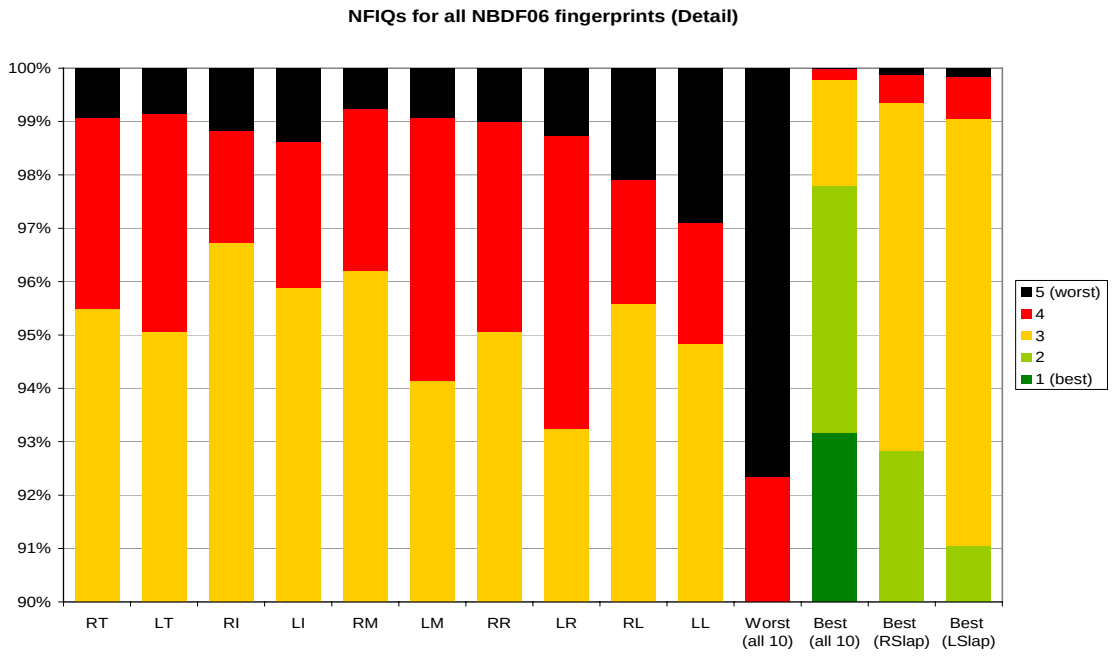

Figure 4: Detail of Figure 3 (NFIQ results for all genuine and imposter fingerprints)

Figure 5 and Figure 6 show the distribution of NFIQ fingerprint quality for the poorer quality fingerprint in all mated pairs (genuines).

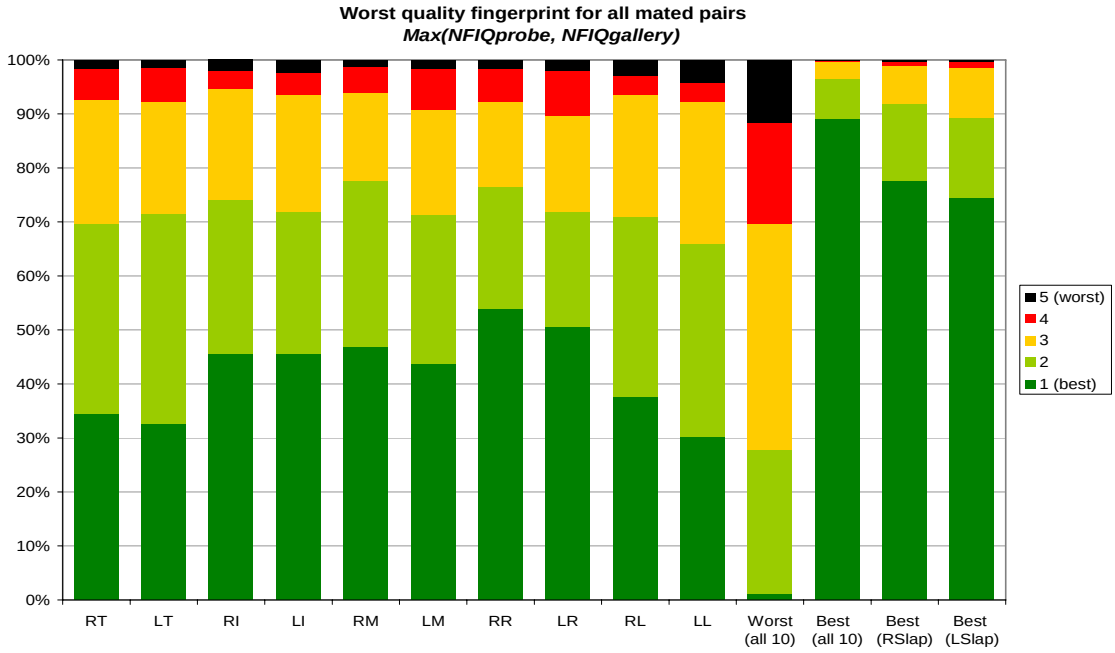

Figure 5: NFIQ results for the worst NFIQ value for every mated (genuine) comparison, including the best and worst finger position from each subject, and best finger position for each slap from each subject.

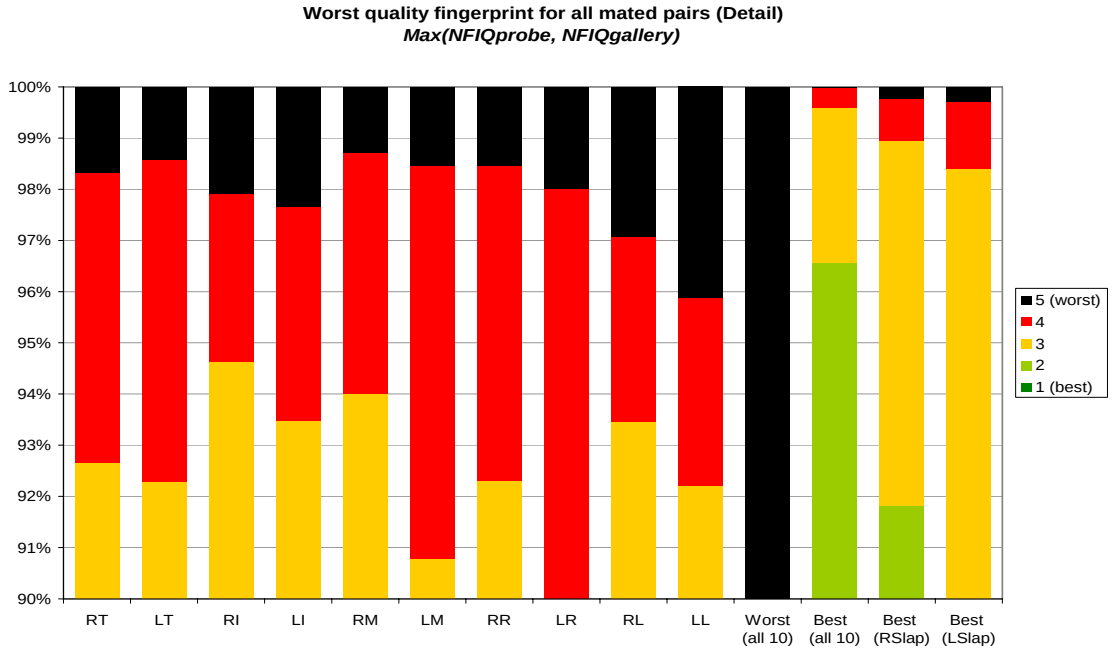

Figure 6: Detail of Figure 5 (NFIQ results for the worst NFIQ value for every mated comparison)

3.1.6 Data integrity issues in NBDF06 data: detection and results

Precise measurement of very small error rates on large datasets requires a detailed analysis of the data for potential data integrity issues; this process is often known as "groundtruthing". The specific issues that were considered in this study include:

- Unconsolidated records: A data integrity error (metadata error) in which the same person has records under different identifiers
- Misidentified records (misidentifications): A data integrity error in which records from different people are listed under the same identifier
- Swapped fingers: An error in which fingerprint images for one subject are incorrectly identified (as when the left and right thumb images are positionally reversed)
- Repeated fingers: An error in which one fingerprint is captured twice in a fingerprint set (as when the left thumb field of a record contains a second impression of the right thumb)

Such issues are often due to collection problems or administrative errors; see [DataQuality] for a full discussion. Note that fusion can be used to detect or overcome some data integrity issues.

The reason that detection of data integrity issues is so critical is that they set a limit on error rates. Given an accurate matcher:

- FRR cannot be lower than the rate of misidentifications
- FAR cannot be lower that the rate of unconsolidated records

3.1.6.1 Detecting data integrity problems (Groundtruthing)

Data integrity problems can be difficult to detect in large datasets. Redundant data (faces and multiple fingers) and multiple accurate matchers provide a basis for locating problems which can then be reviewed manually. To automate detection, metrics were developed based on two methods: score-level fusion and Mahalanobis analysis (which identifies outliers in the score data). Both methods were effective at identifying subject pairs that were misidentified as genuines. Mahalanobis analysis readily revealed partial errors, such as swapped thumbs of subjects whose fused scores were high because the face or other fingerprints matched.

Thorough human review was not possible. Cases were reviewed if metrics flagged all fingerprints, and a sample of cases was reviewed in which metrics flagged only some of the fingerprints. In practice, this means that we ***believe*** that we detected

- most or all cases in which subject pairs were incorrectly identified as genuines (misidentifications), i.e., did not belong in M1 and M2;
- most or all cases in which the thumbs were swapped;
- most or all cases in which the slaps were swapped;
- some of the cases in which only one thumb or slap was invalid (such as the right thumb repeated in both thumb images);
- some of the cases in which the faces were misidentified.

Some data integrity issues were almost certainly undetected, particularly if they only affected one source image (face, one thumb, or one slap). These cases are often difficult to separate from image quality problems.

3.1.6.2 Results

Among 64,867 subjects, we found 33 data integrity problems (0.051%):

- 24 had face and fingerprints misidentified (0.037%)

- 9 had some but not all fingerprints misidentified (0.014%)

This means that FRR smaller than 0.037% should not be possible at low FAR[1], and that FRR smaller than 0.051% is possible only when fusion tolerates these errors due to using additional fingerprints or face images that were not misidentified.

Among all 186,867 subjects, no unconsolidated records were detected in our analysis of this dataset. This is not surprising, as each non-mate (impostor) was compared to only one subject in the gallery, i.e., a full similarity matrix was not produced.

Among the very low-scoring genuines manually reviewed, an additional 46 subjects (0.071%) had some or all fingerprints noted as egregiously poor quality.

3.2 FpVTE MST Data

The Fingerprint Vendor Technology Evaluation (FpVTE) Medium-Scale Test (MST) dataset and matchers are fully described in [FpVTE]. The following is a summary. The FpVTE test was not rerun: the results in the "Multi-Algorithm Fusion" study are based on further processing of the FpVTE MST data from the fall of 2003.

3.2.1 Types of biometric samples

The MST dataset contained right index fingerprints from flat (single finger) or slap (four finger simultaneous) livescan devices. The fingerprints were collected from a range of governmental sources.

3.2.2 Number of subjects

There were 5,600 subjects, and a total of 10,000 fingerprints (all right index fingerprints). Among these, there were at least two fingerprints for 3240 subjects. The full similarity matrix for these subjects (3240x3240) was used in this study.

3.2.3 Matchers

The fusion analysis used fingerprint matching scores from fourteen of the fingerprint systems that participated in the FpVTE MST evaluation in 2003. Of the matchers from the original FpVTE MST, four were not included in the fusion analysis, due to redundancy:

- SAGEM and Ultrascan had multiple entries, of which only the more accurate was used in this study.
- Golden Finger and Raytheon returned equivalent similarity matrices on MST (differing in only 2 out of 100 million scores), so only Golden Finger was used in this analysis.

3.2.4 Derivation of genuine and imposter matcher scores

Each system in the MST evaluation generated a complete 10,000 x 10,000 similarity matrix, by comparing all pairwise combinations of fingerprint images. For these analyses, a subset of this matrix was used, with M=3,240 and NM=M (3,240 x 3,240), resulting in 3,240 genuine scores and about 10 million imposter scores.

[1] A coincidence of matching error and integrity error could result in a lower FRR.

3.3 NIST Biometric Scores Set, Release 1 (BSSR1)

Some of the experiments conducted used the NIST Biometric Scores Set, Release 1, which is described in [BSSR1]. The BSSR1 Dataset has the advantage of being freely available, and therefore is useful for replicating results. While the size of this dataset is sufficient for an exploratory study, the results from the NBDF06 dataset generally supersede BSSR1 results. BSSR1 contains scores based on both faces and fingerprints. The sources of these faces and fingerprints have not been made public.

3.3.1 Types and numbers of biometric samples

BSSR1 includes 3 datasets:

- Set 1 is based on face and fingerprint data from 517 subjects, with scores from two face matchers and one fingerprint matcher.
- Set 2 is based on fingerprint data from 6000 subjects.
- Set 3 (Face data from 3000 subjects) was not used in this study.

Elapsed time between a single individual's samples ranges from 1 day to 29 months with an average of 7 months. 80% of the elapsed times are less than 12 months.

3.3.2 Matchers

BSSR1 includes scores from two anonymous c. 2002 face matchers (C and G). No further information has been made public about these matchers.

BSSR1 includes scores from one fingerprint matcher: V, the NIST VTB Bozorth matcher, developed in 1993-95. A description of the VTB and analysis of its performance is included in [VTB], and in Appendix B of [FpVTE]. In the [FpVTE] and [SDK] evaluations, the performance of the VTB was shown to be about average in comparison to the performance of the commercially available matchers that participated in those evaluations. The source code for the Bozorth matcher is publicly available in [NFIS].

3.3.3 Derivation of genuine and imposter matcher scores

BSSR1 Set 1 is based on face and fingerprint data from 517 subjects, with scores from two anonymous c. 2002 face matchers (C and G), and one fingerprint matcher (V). Set 1 contains four 517x517 similarity matrices: right index fingerprints scored by matcher V, left index fingerprints scored by matcher V, frontal face images scored by matcher C, and frontal face images scored by matcher G. Each similarity matrix contains 517 genuine scores and 266,772 (517 * 516) imposter scores.

BSSR1 Set 2 is comprised of fingerprint scores from matcher V run on the fingerprints of 6000 subjects. There is one 6000 x 6000 similarity matrix for the left index fingerprints, and another for the right index fingerprints. Each matrix contains 6000 genuine scores and 35,994,000 (6000 * 5999) imposter scores. The similarity matrices are depicted in Figure 7.

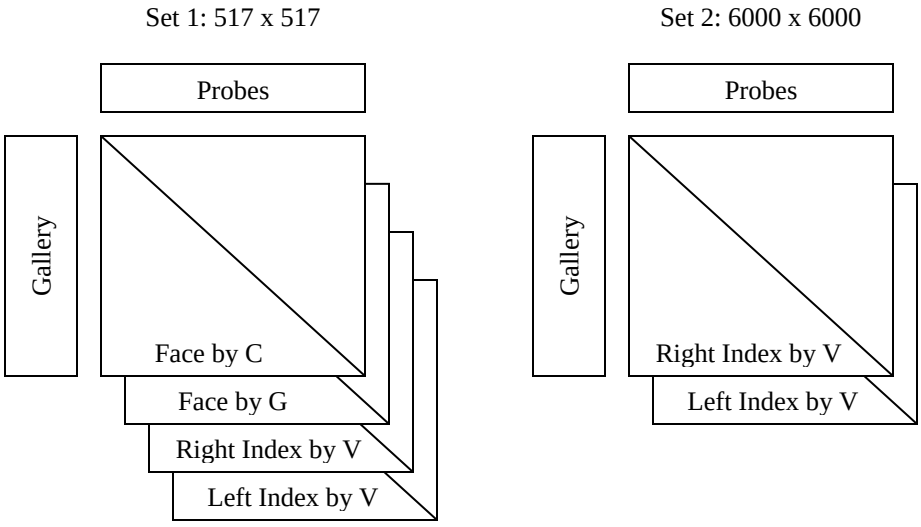

Figure 7: Structure of BSSR1 Set 1 and Set 2. Diagonal elements represent the "true mates" (genuines) and off-diagonal elements represent the "true non-mates" (imposters) for construction of the ROC curves.

4 Pre-fusion matcher performance

This section shows the accuracy of the matchers prior to fusion. It is provided as a baseline for comparison with fused results in the other studies in this collection.

4.1 Baseline performance for NBDF06 data

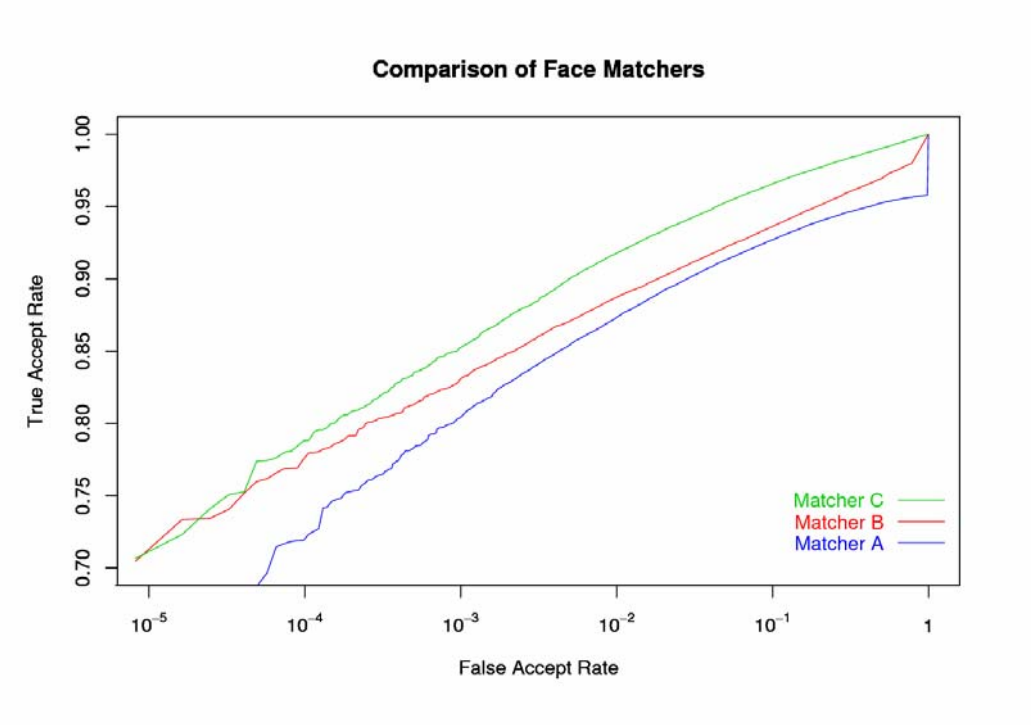

Figure 8: Comparison of face matchers A, B, and C on the NBDF06 dataset

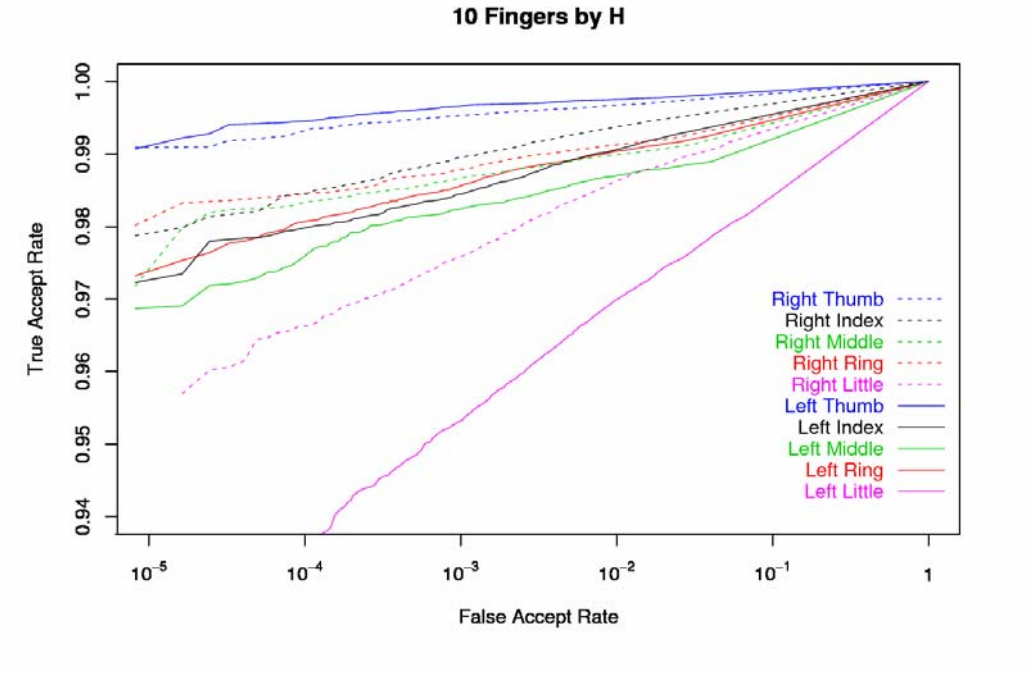

Figure 9: Fingerprint matcher H on the NBDF06 dataset, showing separate results for each finger

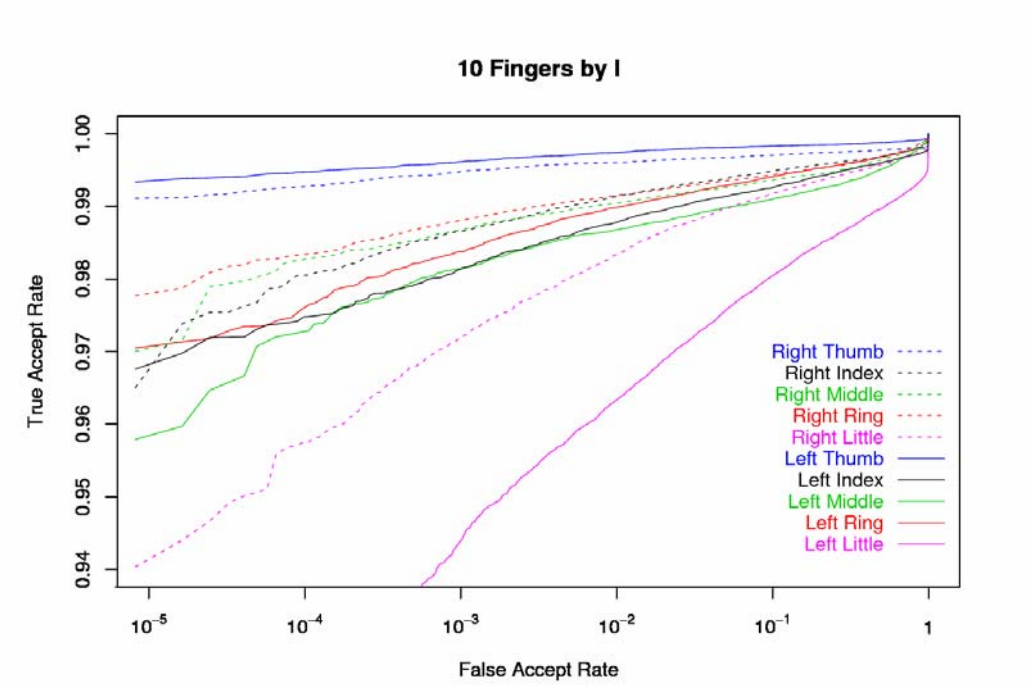

Figure 10: Fingerprint matcher I on the NBDF06 dataset, showing separate results for each finger

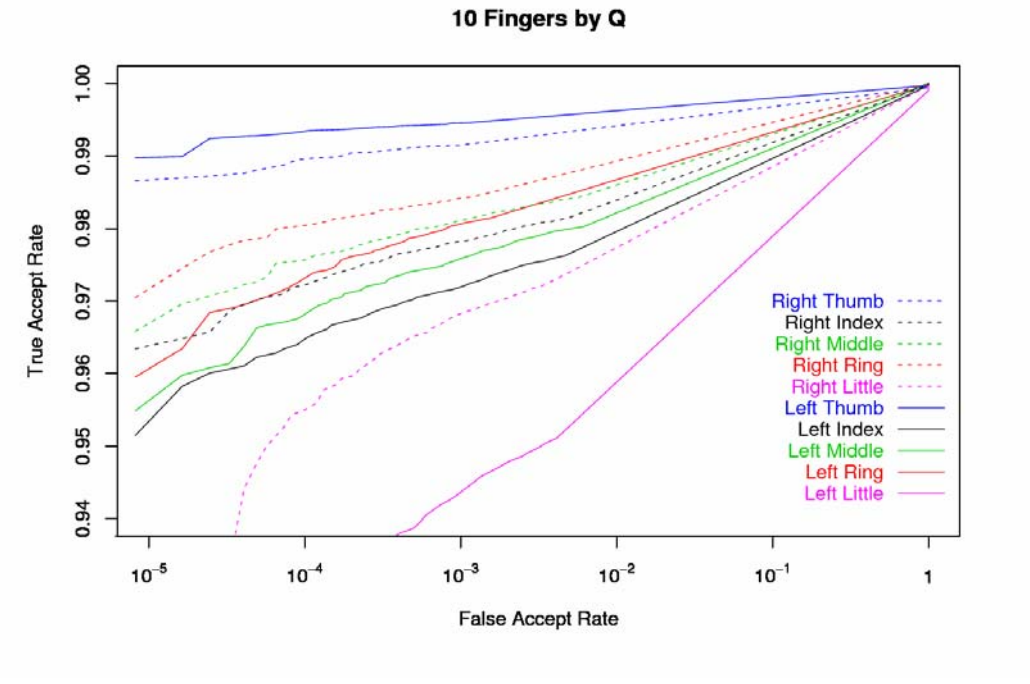

Figure 11: Fingerprint matcher Q on the NBDF06 dataset, showing separate results for each finger

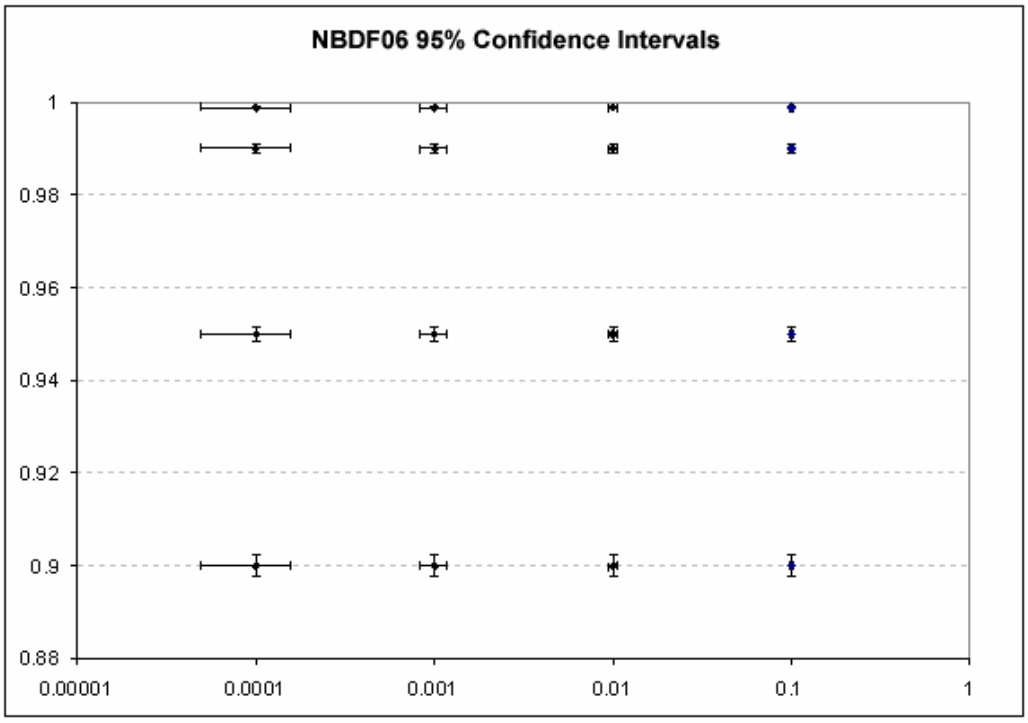

Figure 12: Confidence intervals at selected TAR/FRR and FAR points for the NBDF06 dataset

4.2 Baseline performance for BSSR1 data

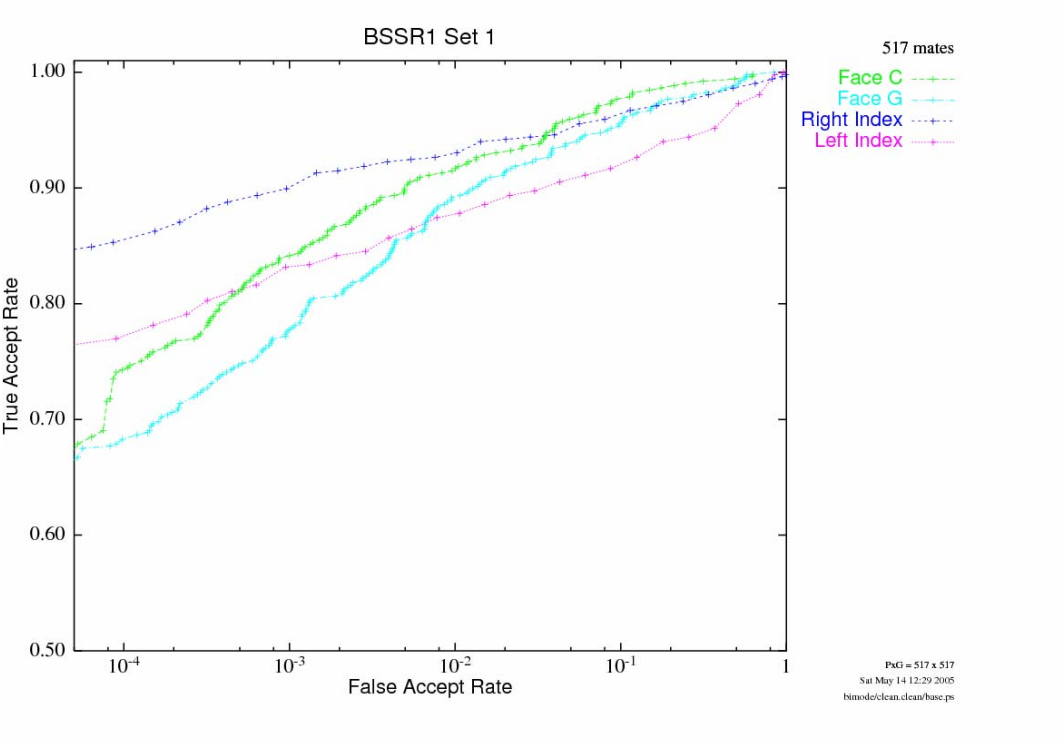

Figure 13: Face and fingerprint matchers on BSSR1 Set 1

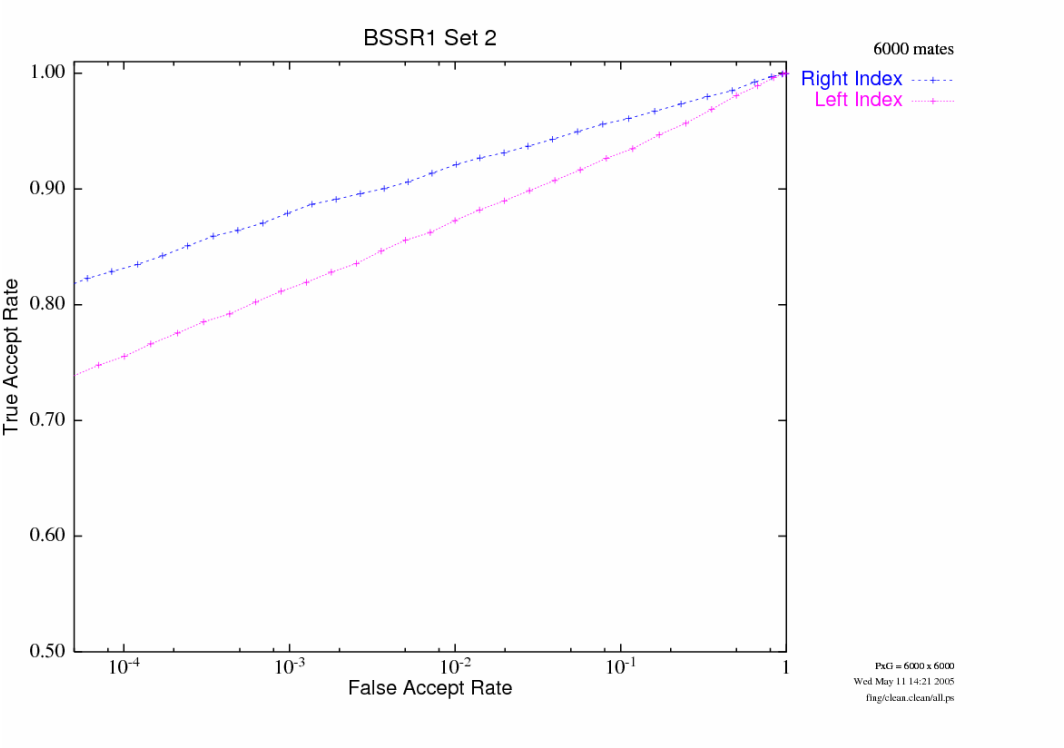

Figure 14: Fingerprint matcher on BSSR2 Set 2

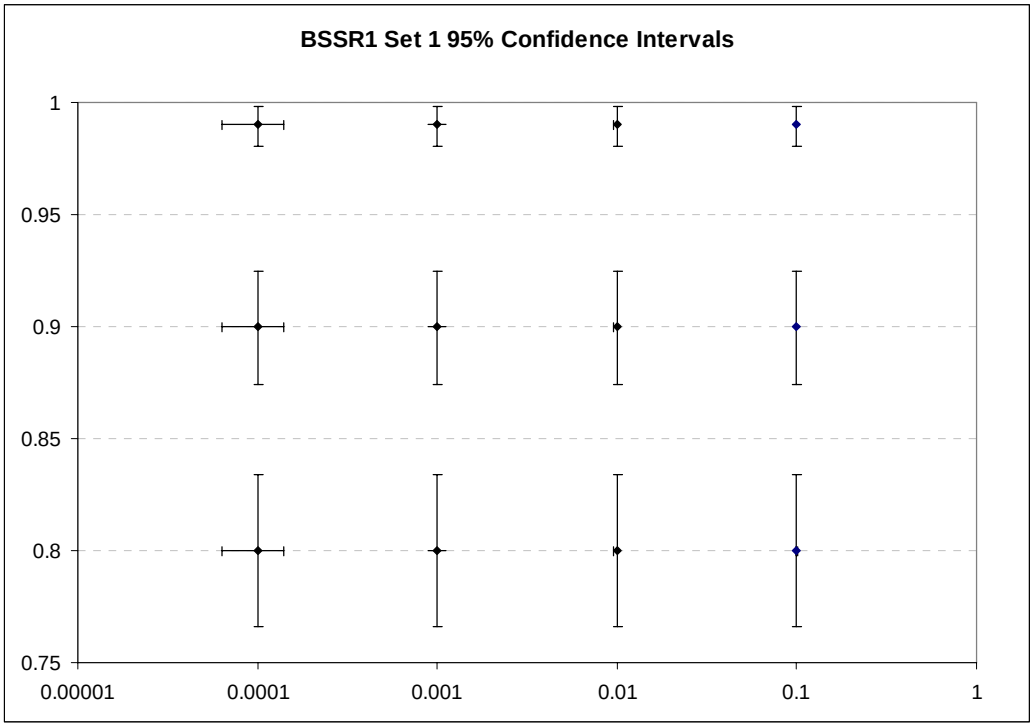

Figure 15: Confidence intervals at selected TAR/FRR and FAR points for the BSSR1 dataset

4.3 Baseline performance for FpVTE MST data

The baseline performance of the FpVTE MST matchers can be found in detail in Appendix C of [FpVTE].

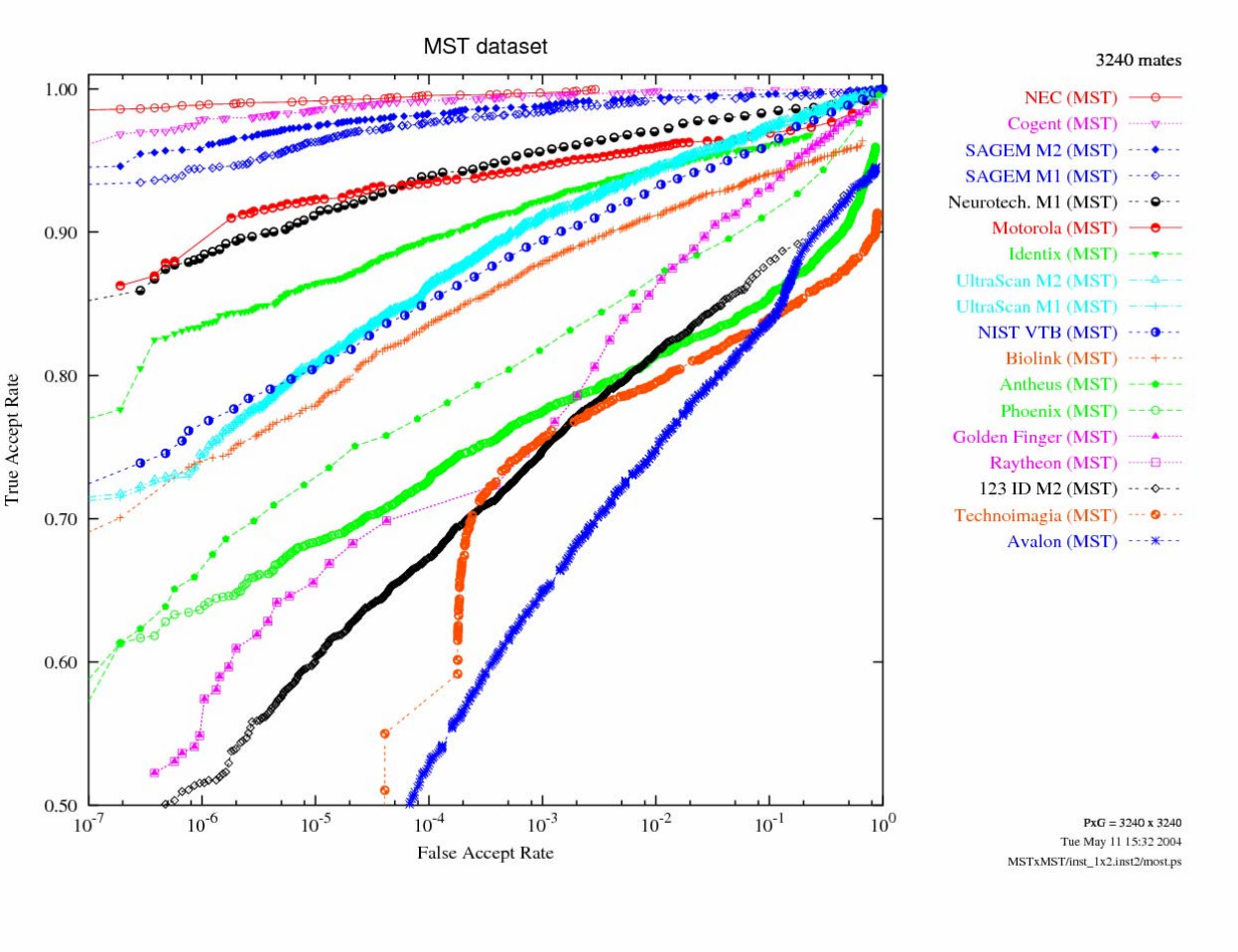

Figure 16: Baseline performance of 18 matchers on FpVTE MST dataset [FpVTE]

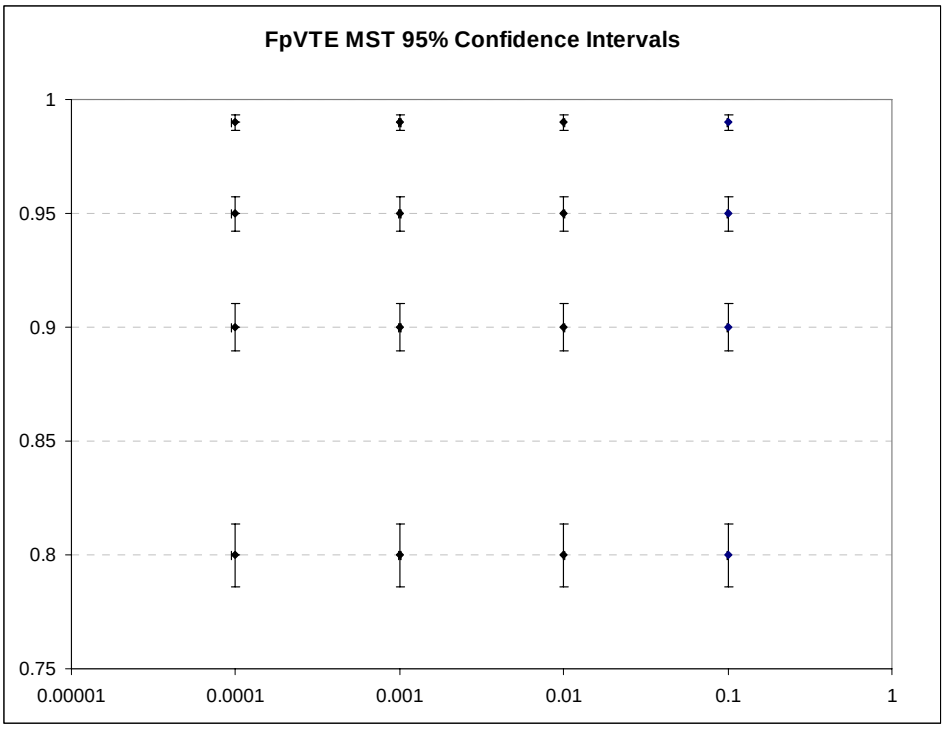

Figure 17: Confidence intervals at selected TAR/FRR and FAR points for the FpVTE MST dataset

5 References

[BSSR1] NIST; Biometric Scores Set –- Release 1 (BSSR1); September 2004. http://www.nist.gov/biometricscores/

[DataQuality] Hicklin, Khanna; "The Role of Data Quality in Biometric Systems"; February 2006; http://www.mitretek.org/Role_of_Data_Quality_Final.pdf

[EFTS-7.1] Federal Bureau of Investigation, Criminal Justice Information Services (CJIS); *Electronic Fingerprint Transmission Specification (EFTS)*; IAFIS-DOC-01078-7.1; May 2, 2005. http://www.fbi.gov/hq/cjisd/iafis/efts71/cover.htm

[FBICert] Products Certified For Compliance with the FBI's Integrated Automated Fingerprint Identification System Image Quality Specifications; http://www.fbi.gov/hq/cjisd/iafis/cert.htm

[FpVTE] C. Wilson, A. Hicklin, H. Korves, B. Ulery, M. Zoepfl, M. Bone, P. Grother, R. Micheals, S. Otto, C. Watson; "Fingerprint Vendor Technology Evaluation 2003"; NIST Interagency Report 7123. June 2004. http://fpvte.nist.gov/

[FRVT2002] Phillips, Grother, Micheals, Blackburn, Tabassi, Bone; Face Recognition Vendor Test 2002; March 2003. http://www.frvt.org

[MugshotBP] "Best Practice Recommendation for the Capture OF Mugshots"; Version 2.0; September 23, 1997. (http://www.itl.nist.gov/iad/894.03/face/bpr_mug3.html) Note: this is identical to "Best Practice Application Level 30" in the forthcoming ANSI/NIST ITL-1 2006 standard.

[NFIQ] Tabassi, E. et al., "Fingerprint Image Quality"; *NIST Interagency Report* 7151, August 2004.

[NFIS] C. Watson, et al; *NIST Fingerprint Image Software*. http://www.itl.nist.gov/iad/894.03/databases/defs/nist_nfis.html

[SDK] Craig Watson, Charles Wilson, Karen Marshall, Mike Indovina, & Rob Snelick; "Studies of One-to-One Fingerprint Matching with Vendor SDK Matchers"; NIST Interagency Report 7221; April 2005. http://fingerprint.nist.gov/SDK/ir_7221.pdf

[SDK2] C. Watson, C. Wilson, M. Indovina, B. Cochran; "Two Finger Matching With Vendor SDK Matchers"; NIST Interagency Report 7249; July 2005. http://fingerprint.nist.gov/SDK/ir_7249.pdf

[SlapSeg] Ulery, Hicklin, Watson, Kwong; Slap Segmentation Evaluation 2004, NIST Interagency Report 7209. March 2005. http://fingerprint.nist.gov/SlapSeg04/ir_7209.pdf

[Snelick-03] R. Snelick, M. Indovina, J. Yen, and A. Mink, "Multimodal Biometrics: Issues in Design and Testing", in Proceedings of Fifth International Conference on Multimodal Interfaces, (Vancouver, Canada), November 2003. http://w3.antd.nist.gov/pubs/ICMI_submit_4_23_03.pdf

[VTB] C. Wilson, C. Watson, M. Garris, A. Hicklin; "Studies of Fingerprint Matching Using the NIST Verification Test Bed (VTB)"; NIST Interagency Report 7020, July 2003. ftp://sequoyah.ncsl.nist.gov/pub/nist_internal_reports/ir_7020.pdf