MINEX

Performance and Interoperability of the INCITS 378 Fingerprint Template

NISTIR 7296

Patrick Grother
Michael McCabe
Craig Watson
Mike Indovina
Wayne Salamon
Patricia Flanagan
Elham Tabassi
Elaine Newton
Charles Wilson

National Institute of Standards and Technology

March 21, 2006

Contents

List of Figures

Disclaimers

These tests were performed for the U.S. Department of Justice and U.S. Department of Homeland Security in accordance with section 303 of the Border Security Act, codified as 8 U.S.C. 1732.

Specific hardware and software products identified in this report were used in order to perform the evaluations described in this document. In no case does identification of any commercial product, trade name, or vendor, imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the products and equipment identified are necessarily the best available for the purpose.

Acknowledgements

The authors would like to thank the MINEX sponsors: The US-VISIT program of the U. S. Department of Homeland Security, the Justice Management Division's IDENT/IAFIS Project within U.S. Department of Justice, and the National Institute of Standards and Technology (NIST). The authors are also grateful to the Department of Homeland Security's US-VISIT program and to the U.S. Department of State's Consular Systems Division for test data.

The authors thank Tom Hopper of the FBI, Michael Garris of NIST, and George W. Quinn of NIST for their reviews. Finally, the authors are grateful to John Campbell of Bion Biometrics for advice on conformance and interoperability issues.

Errata

March 19, 2006 Table 15 was produced without including vendor B in the computation of the interoperable matchers. This exclusion has been fixed.

March 19, 2006 The tables of the supplemental documents erroneously excluded vendor B in some cases. These errors have been fixed.

Release Notes

- Thoughout this report the names of the vendors are associated with a single letter. This association was instantiated to support automated administration of the test and to effect a containment of the vendor identities within NIST. The letter codes were assigned in approximate order of receipt of the implementation and its passing of subsequent shakedown and conformance trials. The ordering is separated by operating system used by the implementation. The use of these letters is maintained in this report to conserve space in its many tables. For reference, the letters are associated with the vendors' names in a permanent footnote.
- A glossary of terms and definitions is given in section 2
- The files listed and hyperlinked below accompany this document. They contain tables too numerous to include in this report.
    1. MINEX Supplement A - Native, Non-Interoperable Performance of the MIN:A , MIN:B and Proprietary Templates
    2. MINEX Supplement B - Typical Interoperabile Authentication
    3. MINEX Supplement C - Matching Arbitrary Template Pairs
    4. MINEX Supplement D - Template Generator Substitution
    5. MINEX Supplement E - Matching Same-source Templates
    6. MINEX Supplement F - Matching Same-image Templates
- Much of the tabulated content in this report was produced automatically. This involved the use of scripting tools to generate directly typesettable LaTeX content. This reduces transcription errors and improves flexibility. The authors considered this important given the desire to disclose as much of the cross-vendor interoperability data as possible.
- This PDF file is likely to be better viewed in print than on-screen.
- Many of the tables in this report give summary biometric error rate accuracy statistics. These are hyperlinked to supplementary graphs and tables maintained on the MINEX website. This site will be made available shortly after publication of this report.
- Readers are asked to direct any correspondence regarding this report to the MINEX organizers.

1 Executive Summary

The approval of the INCITS 378 fingerprint template standard creates the possibility of a fully interoperable multivendor marketplace for applications involving fast, economic, and accurate interchange of compact biometric templates. This document addresses the outstanding questions surrounding the new standard: Does the template give accuracy comparable with proprietary (image-based) implementations? Can template data be generated and matched by different vendors without attendant increase in error rates? The MINEX evaluation was designed to answer these questions. This report summarizes the MINEX comparison of proprietary templates and two variants of the INCITS 378 format - MIN:A which codes minutiae $(x, y, \theta, type, quality)$ and MIN:B which supplements it with ridge count, core and delta information. Fourteen vendors participated. All of them implemented the MIN:A template, six elected to implement the MIN:B enhancement and all of them were baselined against their proprietary technology. By using very large scale trials and four archived operational datasets this report presents the headline assessments of proprietary vs. standard accuracy and cross-vendor interoperability. In due course, this document may be supplemented by more detailed analyses on causes and effects.

The headline results of the test are as follows:

1.	Proprietary templates are superior to MIN:A templates. With a single index finger, the three most accurate systems produce half as many false non-matches at a fixed false match rate of 0.01. Alternatively they produce an order of magnitude fewer false matches at a false non-match rate of 0.01.	Sec. 5.6
2.	The reduced accuracy obtained using standard templates compared to proprietary templates can be adequately compensated for by using two fingers for all authentication attempts. Two finger performance using MIN:A templates is an order of magnitude superior to single finger proprietary performance but again inferior to the proprietary two-finger operation.	Sec. 5.4
3.	The enhanced MIN:B template performed only marginally better than the basic MIN:A template.	Sec. 5.6
4.	Some template generators produce standard templates that are matched more accurately than others. Some matchers compare templates more accurately than others. The leading vendors in generation are not always the leaders in matching and *vice-versa*.	Sec. 5.7.1
5.	Authentication accuracy of some matchers can be improved by replacing the vendor's template generator with that from another vendor.	Sec. 5.7.4
6.	When a matcher's operating threshold is set to achieve some level of accuracy on a vendors own standard templates, there will be an increase in false non-match rates and decrease in false match rates when templates from other vendors are input during interoperable usage.	Sec. 5.7.2
7.	Thirteen of fourteen vendors avoided the use of minutiae type *other* in all cases. The templates from the one company who used *other* contained only this type and were matched poorly. An INCITS 378 application profile may reasonably disallow use of this type.	Sec. 5.7.1
8.	Certification of an interoperable group of products requires some prior specification of the required accuracy. Large numbers of products will interoperate when the accuracy requirement is low. Fewer vendors are interoperable in high performance interoperability scenarios.	Sec. 5.7.7
9.	Larger groups of products can be certified if the group's mean error rate is required to be below a threshold, than if their worst interoperable pair is used for certification. The choice has operational consequences.	Sec. 5.7.7
10.	As with most recent NIST tests [6] and tests conducted by other organizations [7], the error rates between matching algorithms vary by an order of magnitude.	Sec. 5.6

11.	Performance is sensitive to the quality of the dataset. This applies to both proprietary and interoperable templates. Two higher quality datasets, POEBVA and POE , provide reasonable interoperability. Two lower quality datasets, DOS and DHS2 , do not.	Sec. 5.10

2 Terms and Definitions

Table 1 gives MINEX-specific definitions to various words and acronyms found in this report.

No.	Term	Definition
0	ANSI	American National Standards Institute
1	ISO	International Organization for Standardization
2	IEC	International Electrotechnical Commission
3	INCITS	International Committee for Information Technology Standards
4	INCITS 381	U.S. standard governing fingerprint images (see [5])
5	INCITS 378	U.S. standard governing the MIN:A and MIN:B templates (see [4])
6	ISO/IEC 19795-2	International variant of the INCITS 378 format s (see [4])
7	MIN:A	The standard $(x, y, \theta, type, quality)$-based minutiae template
8	MIN:B	The MIN:A template plus ridge count, core and delta information
9	standard template	MIN:A or MIN:B template
10	proprietary template	Template regarded in MINEX to be comparable only with a template from the same vendor
11	enrollment template	Template generated from the first sample of a subject
12	authentication template	Template generated from a second sample of a subject, or from an impostor's sample
13	matcher	Software function that compares two templates to produce a similarity score
14	generator	Software function that accepts an image and produces a template
15	native matching	Comparison, by vendor X, of standard MIN:A or MIN:B templates generated by vendor X
16	BDB	Biometric Data Block (See SC37's *Harmonized Vocabulary*[1])
17	FNMR	False non-match rate
18	FMR	False match rate
19	DET	Detection Error Tradeoff characteristic
20	ROC	Receiver Operating characteristic
21	SDK	Software Development Kit
22	API	Application Programming Interface
23	transaction	The comparison of two templates
24	genuine transaction	Comparison of templates from the same person
25	impostor transaction	Comparison of templates from different individuals
26	verification	One-to-one comparison
27	authentication	Synonym for verification
28	DHS	U. S. Department of Homeland Security
29	DOJ	U.S. Department of Justice
30	DOS	U. S. Department of State
31	FBI	Federal Bureau of Investigation
32	JMD	Justice Management Division
33	NIST	National Institute of Standards and Technology
34	POE	Port of Entry
35	BVA	Biometric Visa
36	MINEX	Minutiae Interoperability Exchange Test
37	IAFIS	The FBI's Integrated Automatic Fingerprint Identification System (IAFIS)
38	FRVT	Face Recognition Vendor Test
39	FVC	Fingerprint Verification Competition

Table 1: Glossary of MINEX related terms

3 Introduction

It has been generally acknowledged that the interchange of fingerprint image data provides the greatest interoperability between dissimilar fingerprint recognition systems. However, standards now exist that specify the location and formatting of processed minutiae locations data, or templates, for matching purposes [4, 2]. For many applications, minutiae templates offer a more space-efficient, less resource intensive, and more cost effective alternative to raw images. However, there is limited information regarding the interoperability, accuracy, and matching accuracy of fingerprint matching systems that exchange minutiae extracted using different methods.

While interoperability can be achieved through the use of images as the method for recording fingerprints data, it could also be achieved by identifying a standard template with which multiple vendors' matching algorithms can achieve high accuracy. In the telecommunications and cable industries, the challenge is preventing a monopoly in the last mile (i.e. the connection from the communications provider to individual customers). Similarly, in biometric authentication, both the industry and users benefit from an interoperable representation or record of biometric data so that a particular vendor is prevented from gaining monopolistic control over the matching systems and thereby blocking entry into the market and controlling price. On the other hand, the use of strict regulatory controls might inhibit innovation and prevent technological improvement.

The Minutiae Interoperability Exchange Test (MINEX) was performed to determine the feasibility of using standard minutiae templates as the interchange medium for fingerprint information between different fingerprint matching systems. This higher order dependency is a feature of interoperability testing that distinguishes this evaluation from linear ones such as FVC [12, 7] and FRVT [14]. This test is not specifically intended to rank vendors but rather to determine whether various subsets of vendors can produce and successfully match each other's standards-conformant templates.

A verification system includes a matcher that compares a submitted sample with an enrolled template to produce a measure of the similarity between the two templates. For fingerprint minutiae templates, accuracy and interoperability are affected by the minutiae detection and extraction algorithms. These include proprietary approaches, the FBI-IAFIS encoding approach that incorporates the number of ridge-crossings to the eight nearest neighbors, and the method specified by the M1 biometrics committee in INCITS 378 [4]. Performance is also dependent on how the matcher is able to process authentication and enrollment templates.

3.1 MINEX Objectives

The MINEX evaluation was intended to assess the viability of the INCITS 378 templates as the interchange medium for fingerprint data. Three specific objectives were

1. To determine whether standardized minutiae enrollment templates can be subsequently matched against an authentication template from another vendor;
2. To estimate the verification accuracy when INCITS 378 templates are compared relative to existing proprietary formats;
3. To compare the INCITS 378 template enhanced with ridge count "extended" data (MIN:B) with the standard's base template (MIN:A).

Letter	Vendor Name	MIN:B
A	Cogent Systems Incorporated	Yes
B	Dermalog Identification Systems GMBH	
C	Bioscrypt Incorporated	
D	Sagem Morpho Incorporated	Yes
E	Neurotechnologija	Yes
F	Innovatrics	Yes
G	NEC Corporation	
H	Technoimagia Corporation	
I	Identix Incorporated	Yes
J	Biologica Sistemas	
K	SPEX Forensics	Yes
L	Secugen Corporation	
M	NITGen Corporation	
N	Cross Match Technologies	

Table 2: Vendors Participating in MINEX

These objectives resulted in a test that is by some measures the largest biometric test ever conducted. It involved testing the core template handling competency of fourteen fingerprint vendors using fingerprint images from a quarter of a million people, and executing in excess of 4.4 billion comparisons, in the production of at least of 23,792 detection error tradeoff (DET) characteristics. This report distills these results into a more tractable number of interoperability tables. The MINEX effort involved four full time equivalents, a ten month period of computation employing greater than thirty dual processor PCs, and a one and a half month analysis and reporting period.

The companies participating in MINEX are identified by their full name in Table 2, and by a letter code and an abbreviated name in the running footer of each page.

3.2 Impact

The results of MINEX have implications that may affect planning decisions for projects such as Personal Identity Verification (PIV) and the interoperability of DHS' IDENT and FBI's IAFIS systems.

- PIV was initiated by Homeland Security Presidential Directive 12[1]. This mandated the establishment of a common identification standard for federal employees and contractors. It required interoperable use of identity credentials to control physical and logical access to federal government locations and systems. In response, NIST released FIPS 201[2] in February 2005, which defines the structure of an identity credential. It specified the inclusion of data from two fingerprints as a third authentication factor. The format for this information was finalized in February 2006, when NIST *Special Publication 800-76* specified essentially the MINEX MIN:A template as a profile of the INCITS 378 standard.

  The result of this program will be the presence of INCITS 378 templates in PIV cards carried by all employees and contractors of federal agencies. Other programs may adopt this specificaton, and together these might number in the millions. One such program is TSA's Transportation Worker Identification Credential (TWIC) program. This is likely to use biometric templates. Another, the Registered Traveler program, will base its biometric content on the PIV specification.

- Initially mandated by the USA Patriot Act[3], Congress requested a "cross-agency, cross-platform electronic system that is a cost-effective, efficient, fully integrated means to share law enforcement and intelligence information necessary to confirm the identity of persons applying for a United States visa." The Department of Justice, Department of Homeland Security, Department of State, and NIST are jointly tasked with the request to "develop and certify a technology standard that can be used to verify the identify of persons applying for a United States visa or such persons seeking to enter the United States pursuant to a visa." In other words, US-VISIT and the FBI each need to be able to exchange fingerprint data with the other to run queries against their respective databases. At present, this exchange is based entirely on image data. If template data could be used with sufficient accuracy in a multi-vendor system, then bandwidth, storage space, and number of template extractions would all be substantially reduced.

4 Background

4.1 Previous NIST Tests

In response to the Patriot Act legislation, NIST initiated extensive testing programs on different fingerprint matching verification systems and algorithms. In 2003, NIST and the Justice Management Division[5] sponsored the Fingerprint Vendor Technology Evaluation (FpVTE [6]) to determine the state-of-the-art of fingerprint matching, identification, and verification systems. It was designed to assess the capability of vendors' systems to meet requirements for both large-scale and small-scale real world applications. Each of the 14 vendors setup their hardware and software systems in NIST laboratory space

[1] The text is here: HSPD-12

[2] Federal Information Processing Standards Publication 201 *Personal Identity Verification for Federal Employees and Contractors.*

[3] USA Patriot Act, Pubic Law 107-56, Section 403(c)(2), was enacted on October 26, 2001

[5] Part of the U.S. Department of Justice

No.	Period	Event
1	August 25 1986	Minutiae standardization begins: ANSI/NBS[4]-ICST 1-1986 Data Format for Fingerprint Information Interchange standard.
2	December 12 2003	Initial discussions for MINEX at NIST
3	January 2004	MINEX sponsorship in place
4	March 8 2004	INCITS 378 Finalized
5	April 20 2004	First ISO Interoperability Testing Contribution
6	June 13 2004	First Working Draft of ISO/IEC 19795-4
7	August 27 2004	Homeland Security Presidential Directive 12 is signed
8	September 21 2004	MINEX is announced publicly
9	January 1 2005	First Committe Draft of 19795-4 (N0910)
10	March 7 - March 15, 2005	Conformance testing starts on submitted sample templates. See Appendix A.1
11	March 15 2005	Deadline for submission of SDK libraries.
12	March 10 - March 18, 2005	Pre-acceptance product validation: See Appendix A.2
13	March 21 - April 30, 2005	Pre-test product validation: See Appendix A.3
14	August 24 2005	Second Committe Draft of 19795-4 (N1252)
15	September 15 2005	Publication of ISO/IEC 19794-2 Biometric Data Interchange Format - Finger minutiae data
16	December 2005	Amendments to INCITS 378 discussed in Toronto meeting of M1
17	April 30 - December 31, 2005	In-test product validation: See Appendix A.4
18	February 20 2006	Third Committe Draft of 19795-4

Table 3: MINEX chronology and related events.

and executed the required tests specified by the FpVTE specification. This program was followed by the fingerprint *matcher Software Development Kit* (mSDK) testing program, which remains open to new entrants and new versions. The mSDK tests are performed by NIST staff using a wider range of larger datasets, thus allowing NIST to perform a wider variety of tests. It also removes much of the burden from the vendor who wants to participate in these testing programs but does not want to devote their company's manpower or equipment for the testing.

Both the FpVTE and the mSDK programs relied on interoperable fingerprint image data obtained from diverse sources and formatted in accord with the ANSI/NIST standard [13] as input to their algorithms. The results derived from these evaluations have demonstrated that excellent matching accuracy can be obtained using interoperable fingerprint images. Furthermore, accuracy rates from the mSDK testing program have confirmed the results derived from the FpVTE program, illustrating that both tests produce equivalent results even though the mSDK is more cost-effective.

4.2 Role of Minutiae Templates

For operational systems, the use of interoperable image data can be costly in terms of additional resources required. A fingerprint image requires a large amount of memory for storage. This can be a considerable burden for applications that store data in a limited-size memory chip on a card. The image data size is also a consideration effecting bandwidth and transmission times. Additional processing time for repeated image compression and decompression, minutiae extraction, and other processing functions required for minutiae matching is also a drain on resources.

Minutiae templates are an alternative to the use of images for fingerprint matching. The template is a list of specific characteristic data processed from a fingerprint image. These characteristics are known as minutiae. In simplest terms a minutiae describes the location on a fingerprint where a friction skin ridge begins, terminates, or splits into two or more ridges. Position and orientation are generally used as the basic attributes to describe a minutia. Minutiae templates have the advantage of being only a fraction of the size of the fingerprint images, using less bandwidth for transmission. But the value of minutiae templates as the medium for fingerprint interchange has been unquantified. Different vendors use different coordinate systems, and location and angle definitions to describe the same minutia. Only a limited amount of information was previously available regarding the effects of these differences on the interoperability, performance, and matching accuracy available when interoperable templates are exchanged. This shortfall pointed to a large scale testing effort.

(a) B (b) I

Figure 1: Examples of Minutiae Placement Variation
A *NIST Special Database 29* image annotated with the $(x, y, \theta, \text{type})$ minutiae points of the MIN:A template generators. Red indicates type "other", green indicates "ridge ending", and blue labels "bifurcation".

4.3 Factors Effecting Template Interoperability

Interoperability of templates is affected by the method used for detection, location, extraction, and formatting of minutiae. Many vendors detect and locate a bifurcation at the same place on the ridge. But this is not the case for ridge endings that may be placed anywhere from a few pixels leading up to the end of a ridge to several pixels into the valley in front of the ridge. The minutiae angle may also vary considerably depending on the intrinsic angular quantization of the implementation, the number of bits of the encoding, and rounding effects. Also some systems rely on minutiae type (ridge ending, bifurcation or other), while others use work better with core and delta information. Other systems, such as the FBI's IAFIS, use the number of ridge-crossings or distances to neighboring minutiae to enhance matching. Still other encoding schemes include unique system proprietary approaches.

Therefore, due to the many factors associated with the creation of minutiae templates, it was necessary in MINEX to limit the representations to two, named MIN:A and MIN:B . These are essentially application profiles of the INCITS 378 standard. Interoperability of the templates is driven by the "evenness" with which vendors generate them. For the MIN:A templates interoperability will be driven by variations in the selection and placement of minutiae:

1. **Detection:** Each template generator must find enough of the same minutiae to support matching. The success of this stage rests on the standardized guidance on locating ridge endings and birfurcations given in clauses 5.3.{2,3} of [4] and on the knowledge accrued by implementers. The general case is that the intersection of the sets of minutiae detected by products X and Y is neither empty nor equal to the union of the two.

2. **Coding:** Each generator must compute the location, orientation and type of the minutiae equivalently.

An example of the variation in $(x, y, \theta, type)$ is shown for two example template generators in the images of Figure 1.

4.4 Review of Minutiae Standardization

The earliest minutiae standard dates back to 1986 when the FBI and NIST (formerly the National Bureau of Standards) developed the minutiae-based ANSI/NBS-ICST 1-1986 Data Format for Fingerprint Information Interchange standard. Many of the requirements from this original standard are still contained in the current ANSI/NIST-ITL 1-2000 version of the standard. This series of standards were developed primarily by law-enforcement agencies, major AFIS vendors and users.

In January 2002, the INCITS M1 Biometrics committee was formed. It is driven by commercial verification rather than law-enforcement identification needs. One of its first projects was the development of a standard for a finger minutiae data interchange format. This standard was progressed in cooperation with developers the of BioAPI and CBEFF standards. It contains provision for formatting data from several presentations or views of the same finger thus accommodating systems that rely on several readings of the same finger to construct a good average template. Published as INCITS 378 Finger Minutiae Format for Data Interchange, this standard was based on the ANSI/NIST-ITL 1-2000 standard and the FBI's Electronic Fingerprint Transmission Specification (EFTS 7.0).

The standard provides guidance on how ridge-endings and bifurcations are to be located, and on how the minutia angle is to be calculated. In addition the format uses the upper left corner of the image as the origin. A minutia's angle is stated in increments of two degrees. Minutia type and quality are also recorded. The standard also has provision for an open format defined for the optional inclusion of common extended data fields. These include core and delta information, ridge count information for either four-neighbor quadrants or eight-neighbor octants, and vendor-defined information.

In December of 2002 subcommittee 37 (SC37) was formed by the Joint Technical Committee of ISO and IEC (JTC1). Essentially, SC37 was created as an international analog to M1 with substantially similar work items. ISO standardization garners more technical input and sets a higher bar for achieving consensus. A newly assembled data formats working group adopted the INCITS 378 draft in its development of what was ultimately published, in 2005, as ISO/IEC 19795-2 standard [2]. The most significant difference between the ISO standard and the INCITS 378 is the representation of minutiae angle is in 2 degrees increments in INCITS 378 and 1.40625 degrees in the ISO version. As different vendors quantize to different values before mapping to 2 degree increments, this change in representation may not be significant. Additional card formats were introduced, which added to the complexity of the standard.

4.5 MINEX Overview

To satisfy its objectives, MINEX followed the testing strategy used for the mSDK tests. That is, each vendor provided NIST with their SDK that contained the following callable functions to:

- create an INCITS 378 MIN:A template from an image
- create an INCITS 378 MIN:B template from an image (optional)
- create an proprietary template from an image
- produce a comparison score from two MIN:A templates
- produce a comparison score from two MIN:B templates (optional)
- produce a comparison score from two proprietary templates

NIST staff developed test harnesses around these functions, and scripts around those compiled programs. For this test to be meaningful, a very large number of samples were required. The data chosen originated from four datasets that represented a range of operational image qualities. All images selected for use in MINEX had been gathered from subjects using live-scan devices.

In order to quantify any loss in matching accuracy associated with the use of interoperable minutiae templates, it was first necessary to establish a baseline set of performance statistics. Each vendor created their own proprietary minutiae templates from the fingerprint images belonging to each of the authentication and enrollment datasets. That vendor's proprietary matcher was then used to compare the authentication templates to the enrollment templates for the datasets used. The match results from these comparisons can be expected to give the best performance that the vendor could attain using their submitted matcher.

Once this baseline performance for a vendor was known, these results could be compared to those achieved when templates from dissimilar systems were interchanged and matched. An authentication template, generated by one vendor in conformance to one scheme of encoding, was compared to an similarly encoded enrollment template generated by a second vendor using a fingerprint matcher from a third vendor. This "mix and match" of authentication template, enrollment template, and matcher provider produced a three-dimensional matrix of all possible interoperable accuracies.

Although MINEX was the first large-scale test of minutiae interoperability, it was not the first test. In 2003, the International Labour Organization (ILO) conducted a scenario test of minutiae-based fingerprint templates[9]. The MINEX test benefited from the lessons learned during that project and passed on information derived from this study to vendors prior to the deadline for their SDK submission. First, over-reliance on the minutiae type field may have degraded performance. Participants should regard minutiae type simply as an extra piece of information. Secondly, due to variances in angular quantization used by particular algorithms, an accurate determination of minutia direction was encouraged using the smallest amount of quantization possible so that it will be within the two degree reporting limit. Thirdly, since the ILO study used one of the *card* formats from the ISO standard it was necessary to limit the number of minutiae in the record to 52. MINEX provided for up to 128 minutiae to avoid this problem.

4.5.1 MINEX Templates

The INCITS 378 standard is being specified for commercial and other applications aimed at identity verification. Vendors are also building to this standard and are marketing products compatible with INCITS 378 . Therefore, in addition to the proprietary template generation and matching functions, each MINEX vendor's SDK was required to encode and match a single view MIN:A template as well. The minutiae quality field required by INCITS 378 was set to zero in all cases, as no universally accepted definition for it exists.

Traditional AFIS vendors do not universally use ridge crossing data as part of the matching algorithms. But for the past ten years, states and other law-enforcement agencies have been submitting images to the FBI which are converted to minutiae templates that include eight-neighbor octant ridge crossing data. In order to determine if performance would be significantly improved by the extended data fields allowed by INCITS 378 , vendors were also encouraged, but not required, to submit a MIN:B template generator function. This would encode eight-neighbor octant ridge crossing data surrounding each minutiae with core and delta data if available. The MINEX specification provided detailed instructions on determining the minutia angle. Although the INCITS 378 specification allows purely proprietary data to be included in the extended data section this was specifically disallowed in MINEX .

Each image presented to an SDK for template generation included the NIST Fingerprint Image Quality (NFIQ) value. To conform to the common meaning of the word quality, NFIQ values $\{5, 4, 3, 2, 1\}$ were remapped to $\{1, 25, 50, 75, 100\}$ respectively. The NFIQ values were demonstrated [16, 17] to be predictive of the normalized genuine scores. The actual use of the input NFIQ value by the receiving library function was not and could not be required.

5 Results

The following subsections present the results of the MINEX evaluation. The material is organized as follows.

Data.	Temp.	Fing.	Total	A	B	C	D	E	F	G	H	I	J	K	L	M	N
POEBVA	MIN:A	Right	246435	516	1	1	32	200	1	0	57	2	25	4619	12	14	17
		Left	245994	881	2	3	35	264	3	0	67	2	32	6830	11	6	15
POEBVA	MIN:B	Right	246712	513			32	95	1			0		4619			
		Left	245994	881			37	108	2			0		6830			
POEBVA	PROP	Right	245994	0	1	1	0	0	0	0	0	0	14	0	0	0	0
		Left	245994	0	2	3	0	0	0	0	0	0	18	0	0	0	0
DHS2	MIN:A	Right	245978	506	98	415	102	4239	415	1472	5139	12797	226	1657	7689	26245	103
		Left	245978	419	39	284	60	3346	284	1347	3900	10759	122	1306	6200	23474	40
DHS2	MIN:B	Right	245978	508			97	3917	256			0	0	1657			
		Left	245978	420			58	3112	168			0	1	1306			
DHS2	PROP	Right	246063	0	98	415	0	0	0	0	0	0	148	0	0	0	0
		Left	245978	0	39	284	0	0	0	0	0	0	80	0	0	0	0
POE	MIN:A	Right	245991	514	1	0	36	191	0	0	57	2	13	4731	19	10	13
		Left	245991	862	6	0	46	282	0	0	97	5	21	7011	19	24	15
POE	MIN:B	Right	245991	516			35	94	0			1		4731			
		Left	245991	863			46	117	0			3		7011			
POE	PROP	Right	245991	0	1	0	0	0	0	0	0	0	9	0	0	0	0
		Left	245991	0	6	0	0	0	0	0	0	0	16	0	0	0	0
DOS	MIN:A	Right	245852	103	3	5	8	21	5	4	8	16	11	1360	14	19	6
		Left	245852	210	2	10	7	50	9	4	19	25	16	2352	22	45	5
DOS	MIN:B	Right	245852	104			8	12	5			1		1360			
		Left	245852	210			7	28	7			1		2352			
DOS	PROP	Right	245852	0	3	5	0	0	0	0	0	0	2	0	0	0	0
		Left	245852	0	2	9	0	0	0	0	0	0	8	0	0	0	0

Table 4: Fraction of templates with size ≤ 38 bytes (i.e. the size of a single minutia MIN:A record). Blank cells occur for vendors who elected not to supply a MIN:B generator.

5.1 Datasets

Four datasets were used in MINEX testing. They are referred to as POEBVA , DHS2 , POE and DOS and are described in detail in Appendix B. All of these are operational data sets gathered in on-going US Government operations, and have been sequestered at NIST for testing. MINEX uses randomly selected extracts of those databases. The integrity of the ground truth of the datasets was assured by human inspection. This process is described Appendix B.1. The quality composition of the datasets is tabulated using the NIST Fingerprint Image Quality (NFIQ [16, 17]) method in appendix B.2. Unless otherwise stated, the results from POEBVA are featured in the tables of this report. Results for the other datasets are covered exhaustively in the appendices.

The POEBVA dataset is distinguished from many testing corpora in two valuable ways. First, the enrollment and authentication images are collected at separate locations in different environments with different sensors. Second, the authentication images were collected without human intervention in a process referred to as *autocapture*. This has considerable positive implications for the expected relevance of the MINEX results to fielded performance. These aspects enhance the operational relevance of the MINEX results.

5.2 Performance Measures

This report makes extensive reference to the false non-match and false match rates, FNMR and FMR . These are the fundamental error rates produced in offline testing. The FNMR is the fraction of genuine comparisons that result in a score less than or equal to the operating threshold of the matcher. FNMR is a measure of inconvenience i.e. the fraction of genuine transactions that result in failure[6] Likewise the FMR is the fraction of impostor comparisons that result in a score greater than the operating threshold. FMR is regarded as a measure of security, i.e. the fraction of illegitmate matching attempts that result in success. As is typical in offline testing [3], this report does not fix an operating threshold but instead uses all the scores from a matcher as thresholds that could be used in actual operation. This contrasts with scenario testing which often uses a device configured with one fixed operating threshold. The output is then a *decision* and not a score, and this precludes investigation of performance at other thresholds. The advantage of requiring matchers to produce integral or real-valued scores is that it allows a survey over *all* operating points, t, to be used in the production of a DET characteristic. This is a plot of FNMR(t) against FMR(t) [7] and is the primary output of a biometric performance test. DET characteristics are vital in establishing the balance between the inconvenience associated with the incorrect rejection of legitimate users (as quantified by FNMR), and the incorrect acceptance of fraudulent users (as quantified by FMR).

Setting the threshold may be a sensitive issue and is always application specific. Although this report makes no recommendations on threshold setting, it necessarily adopts "default" performance figures of merit in support of comparison

[6]Inconvenience is not the appropriate term for negative identification systems for which FNMR expresses the fraction of transactions in which an enrolled entry is not returned. As discussed in section 5.5 identification is not explicitly covered in this report.

[7]DET characteristics plot false non-match rate (FNMR) against false match rate (FMR). This differs trivially from the more common Receiver Operating Characteristic (ROC) in that it plots 1 - FNMR on the y-axis.

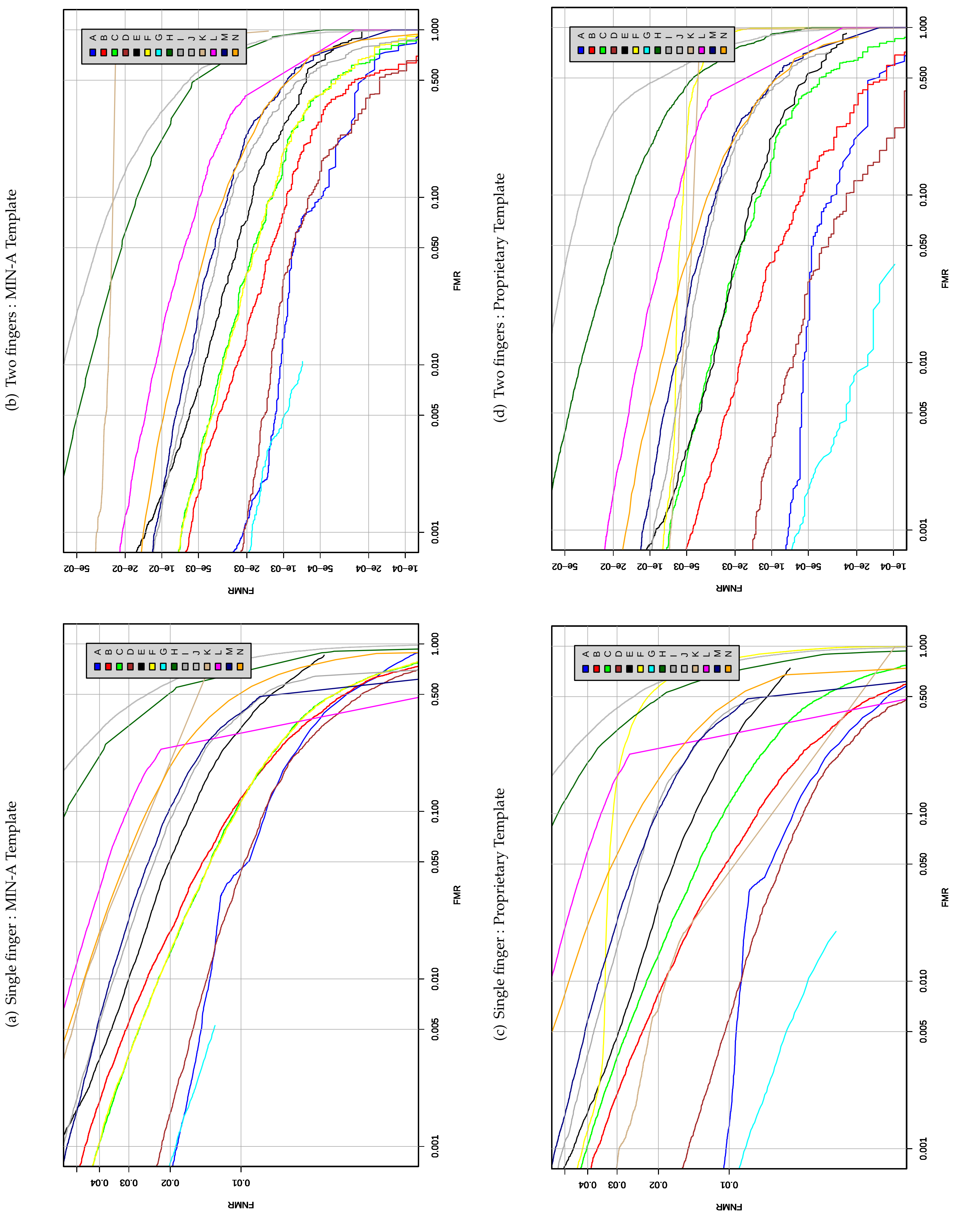


Table 5: DET characteristics for the POEBVA dataset.

A = Cogent	B = Dermalog	C = Bioscrypt	D = Sagem Morpho	E = Neurotechnologija	F = Innovatrics	G = NEC
H = Technoimagia	I = Identix	J = Biologica	K = SPEX Forensics	L = Secugen	M = NITGen	N = Cross Match

objectives. Unless stated otherwise the results in this report correspond to the threshold that produces a FMR of 0.01. The figure of merit is the FNMR at that point. The value 0.01 should not be construed as a recommended operating point but as a value at which error rate differences may be readily observed. Note however that operationally thresholds may be set to produce a desired FNMR , and the appropriate figure of merit would then be FMR at FNMR of 0.01. Results for both approaches are contained within the DETs of Figure 5 and are discussed later in section 5.6.

5.3 Handling Failure to Enroll

In many operations a fraction of the population is unable to enroll in a biometric system. This may arise because a biometric system is forced to reject samples on the basis that it simply cannot find the needed signal (e.g., finding a face in a photograph) or that the signal is in some sense unsuitable (e.g., the fingerprint image is poor). Such a determination would likely initiate a request to the user to try again. If each repeated attempt fails the individual contributes to the *failure to enroll* rate.

It is often said that evaluations which are conducted offline (MINEX being one example) that employ archived image databases are by defintion incapable of measuring failure to enroll rates. This is correct in that an offline database will not include samples from the fraction of the population who simply do not possess the biometric (in this case, those with missing fingers). However a database may well contain samples that, if presented to the front-end image processing, feature detection, or quality assessment algorithms, would cause a failure to enroll. This will be particularly the case in databases that were originally collected without a mechanism for declaring failures to enroll, or for which the acquisition policy demanded that some raw image sample *must* result from a transaction. This is the case with the POEBVA database used in MINEX .

The MINEX test protocol required template generators to produce a template in all cases. This applied to both proprietary and standard templates, and the latter were required to be conformant to the standard. Thus if a template generator was presented with an image of such poor quality that it would ordinarily reject it, the result in MINEX is nevertheless a template that is a valid input to the matcher and as such will produce a low similarity score. The MINEX specification required this score to be -1. For the standard MIN:A and MIN:B templates, this would have resulted in a template containing zero minutiae that still conforms to the standard. A similar, but not explcitly measureable effect, is likely present in proprietary templates. The frequency of template creation “failure” is shown in Table 4, which gives the numbers of templates whose size is less than or equal to 38 bytes (i.e. the size of a MIN:A template with *one* minutia). The same limit is used for proprietary templates, although this is of unclear meaning.

Thus, for MINEX , all images resulted in a template. The rationale for this constraint on vendors is that it allows failure to enroll to be accounted for correctly. This is important in an interoperablility test because *all* matchers will have to deal with empty templates. Of course these are unmatchable and the MINEX matcher specification required implementations to report a similarity of -1 if either of the input templates was empty. Thus the effect for an empty template involved in an impostor transaction is a correct rejection. For a genuine transaction, the result is a false reject. The important aspect of this treatment is that it gives a simple, uniform and fair accounting of failed template generation.

5.4 One and Two Finger Matching

This report contains performance estimates for one and two-finger authentication. Unless stated otherwise, the single-finger results are obtained by pooling the scores from the left and right index finger comparisons as though they were from different individuals. The performance estimates are therefore representative of single-finger verification applications in which users choose to use either their left or right index finger for authentication. This report does not assess the effect of multiple verification attempts because it uses archived datasets with only two impressions per finger.

The two-finger similarities were produced in a score-level fusion process. The fused score is simply the sum of the left and right comparison scores:

$$s_{ij} = s_{ij}^{(R)} + s_{ij}^{(L)} \tag{1}$$

where i and j denote the i-th enrolled image and the j-th authentication sample and s is the scalar output of a matcher. This *sum-rule* is a simple yet powerful method for multi-sample fusion, is ubiquitous in the literature [15, 10], and has long had theoretical recommendation [11].

(a) FNMR at FMR = 0.01

POEBVA	Single finger							Two fingers					
	Proprietary		MIN:A		MIN:B			Proprietary		MIN:A		MIN:B	
	FNMR	Rank	FNMR	Rank	FNMR	Rank		FNMR	Rank	FNMR	Rank	FNMR	Rank
A	0.0089	2	0.0136	2	0.0135	1	A	0.0006	2	0.0011	2	0.0010	1
B	0.0189	5	0.0251	6			B	0.0018	4	0.0024	4		
C	0.0225	6	0.0225	5			C	0.0032	6	0.0032	6		
D	0.0089	3	0.0140	3	0.0140	2	D	0.0007	3	0.0013	3	0.0013	2
E	0.0251	7	0.0301	7	0.0296	4	E	0.0030	5	0.0045	7	0.0043	4
F	0.0337	9	0.0224	4	0.0199	3	F	0.0061	10	0.0031	5	0.0028	3
G	0.0047	1	0.0129	1			G	0.0002	1	0.0007	1		
H	0.1004	13	0.1027	13			H	0.0367	13	0.0422	13		
I	0.0329	8	0.0348	8	0.0336	5	I	0.0051	7	0.0056	8	0.0054	5
J	0.1503	14	0.1505	14			J	0.0704	14	0.0640	14		
K	0.0186	4	0.0461	10	0.0634	6	K	0.0054	8	0.0275	12	0.0305	6
L	0.0575	12	0.0524	12			L	0.0126	12	0.0113	11		
M	0.0358	10	0.0359	9			M	0.0060	9	0.0061	9		
N	0.0481	11	0.0486	11			N	0.0082	11	0.0081	10		

(b) FNMR at FMR = 0.001

POEBVA	Single finger							Two fingers					
	Proprietary		MIN:A		MIN:B			Proprietary		MIN:A		MIN:B	
	FNMR	Rank	FNMR	Rank	FNMR	Rank		FNMR	Rank	FNMR	Rank	FNMR	Rank
A	0.0103	2	0.0187	1	0.0180	1	A	0.0007	2	0.0021	3	0.0015	1
B	0.0365	5	0.0462	6			B	0.0044	4	0.0059	4		
C	0.0403	6	0.0403	4			C	0.0069	6	0.0069	5		
D	0.0149	3	0.0218	3	0.0218	2	D	0.0013	3	0.0021	2	0.0021	2
E	0.0461	8	0.0600	9	0.0602	5	E	0.0089	8	0.0136	9	0.0143	5
F	0.0420	7	0.0407	5	0.0348	3	F	0.0076	7	0.0069	6	0.0057	3
G	0.0086	1	0.0190	2			G	0.0006	1	0.0018	1		
H	0.1618	13	0.1665	13			H	0.0737	13	0.0783	13		
I	0.0510	9	0.0575	8	0.0522	4	I	0.0093	9	0.0110	7	0.0094	4
J	0.2372	14	0.2372	14			J	0.1114	14	0.1192	14		
K	0.0294	4	0.0689	10	0.0634	6	K	0.0068	5	0.0345	12	0.0305	6
L	0.0844	12	0.0793	12			L	0.0224	12	0.0211	11		
M	0.0549	10	0.0550	7			M	0.0114	10	0.0114	8		
N	0.0837	11	0.0764	11			N	0.0159	11	0.0142	10		

Table 6: Proprietary vs. Native FNMR at fixed FMR .
Summary of performance for all vendors' proprietary implementations, and their native MIN:A and MIN:B perfomance, on POEBVA data. The tables give FNMR at FMR values of 0.01(a) and 0.001(b).

We note that by executing fusion after matching NIST may have usurped what would ordinarily be a function residing in a dedicated two-finger matcher. The proper method for conducting a two-finger test is to require vendors to provide a matcher capable of comparing *multi-templates*. A multi-template, in this context, would be a single, conformant INCITS 378 record containing the minutiae records from the left and right index fingers, or some analagous bundling of those fingers' proprietary templates. This approach would have delegated the responsibility of conducting fusion to the vendors (rather than NIST) who should retain the right to develop and implement their own intellectual property in this area. Such encapsulation was not provided for in the MINEX API because of the additional complexity and the fact that summing of scores is a potent method of fusion.

The use of fusion, however, has significant implications. Error rates drop substantially but there is the attendant requirement to *always* acquire samples from both fingers. This is not equivalent to access control implementations that grant access if *either* the left *or* right finger can be authenticated in any of say three attempts (a practice used in the ILO trials[9]). This is because use of two-finger authentication in a system whose threshold is set for single-finger operation will decrease false non-matches markedly but not without attendant increase in the false match rate. Instead the threshold must be set to reflect whatever authentication protocol is instantiated. Research into the use of fusion where second samples are only *sometimes* acquired is markedly smaller than the traditional "always on" decision and score-level fusion literatures.

Results comparing single and two-finger matching are presented in Tables 6. The situation is similar for both proprietary and standard templates. Comprehensive results for all datasets in the form of Tables and per-matcher DET characteristics

(a) FMR at FNMR = 0.02

POEBVA	Single finger							Two fingers					
	Proprietary		MIN:A		MIN:B			Proprietary		MIN:A		MIN:B	
	FMR	Rank	FMR	Rank	FMR	Rank		FMR	Rank	FMR	Rank	FMR	Rank
A	0.0000	2	0.0006	1	0.0004	2	A	0.0000	3	0.0000	3	0.0000	2
B	0.0084	5	0.0192	6			B	0.0000	6	0.0000	6		
C	0.0143	6	0.0143	4			C	0.0000	3	0.0000	3		
D	0.0003	3	0.0016	3	0.0016	3	D	0.0000	3	0.0000	3	0.0000	2
E	0.0289	7	0.0507	8	0.0464	5	E	0.0003	10	0.0005	10	0.0006	5
F	0.5714	13	0.0146	5	0.0096	4	F	0.0000	7	0.0000	3	0.0000	2
G	0.0000	1	0.0008	2			G	0.0000	3	0.0000	3		
H	0.4643	12	0.2521	13			H	0.0484	13	0.0554	12		
I	0.1144	9	0.0897	9	0.1159	6	I	0.0000	7	0.0000	7	0.0000	4
J	0.6045	14	0.5923	14			J	0.3077	14	0.1418	13		
K	0.0069	4	0.0243	7	0.0004	1	K	0.0000	3	0.9283	14	0.9283	6
L	0.2266	11	0.2351	12			L	0.0016	12	0.0012	11		
M	0.1042	8	0.1042	10			M	0.0000	9	0.0000	8		
N	0.1573	10	0.1778	11			N	0.0004	11	0.0002	9		

(b) FMR at FNMR = 0.01

POEBVA	Single finger							Two fingers					
	Proprietary		MIN:A		MIN:B			Proprietary		MIN:A		MIN:B	
	FMR	Rank	FMR	Rank	FMR	Rank		FMR	Rank	FMR	Rank	FMR	Rank
A	0.0013	2	0.0448	3	0.0448	2	A	0.0000	2	0.0000	2	0.0000	1
B	0.0535	5	0.1208	6			B	0.0001	4	0.0002	4		
C	0.1154	6	0.1154	5			C	0.0002	7	0.0002	6		
D	0.0060	3	0.0437	2	0.0441	1	D	0.0000	2	0.0000	2	0.0000	1
E	0.2389	8	0.3104	8	0.3165	4	E	0.0008	9	0.0017	9	0.0017	5
F	0.8368	14	0.1142	4	0.1021	3	F	0.0002	6	0.0002	5	0.0001	3
G	0.0005	1	0.0053	1			G	0.0000	2	0.0000	2		
H	0.6568	12	0.5500	12			H	0.1920	13	0.1722	12		
I	0.4053	10	0.3730	9	0.4151	5	I	0.0006	8	0.0014	7	0.0008	4
J	0.8046	13	0.8046	13			J	0.4668	14	0.3250	13		
K	0.0196	4	0.9552	14	0.9552	6	K	0.0001	5	0.9283	14	0.9283	6
L	0.2266	7	0.2351	7			L	0.0227	12	0.0151	11		
M	0.3874	9	0.3875	10			M	0.0016	10	0.0016	8		
N	0.4076	11	0.4584	11			N	0.0046	11	0.0039	10		

Table 7: Proprietary vs. Native FMR at fixed FNMR .
Summary of performance for all vendors' proprietrary implementations, and their native MIN:A and MIN:B perfomance, on POEBVA data. The tables give FMR at FNMR of 0.02 (a) and 0.01 (b).

may be found in the accompanying document, MINEX Supplement A - Native Matching.

5.5 Identification

This report does not report identification system performance. Testing of multifinger AFIS systems is more specialized than that for verification, primarily because it includes an explicit enrollment of a population because it employs multiple fingers and additional techniques for expediting searching. However, this study is relevant to identification system implementers because it quantifies the raw biometric capability of generators and matchers that implement the standard template. It is likely that the relationship between verification and identification accuracy for proprietary implementations is the same for standard templates. Further any interoperability problems present in verification trials will almost certainly manifest themselves in identification scenarios. One aspect peculiar to identification is the searching of an enrolled population made up of templates from more than one source.

5.6 Proprietary vs. Native Standard Template Performance

When a biometric interchange format is standardized, two performance related questions arise. The first is whether the format embeds sufficient information for instances of the format to be matched with low error rates. It is clear that a minutiae interchange format that coded only the x-coordinates of minutiae would perform poorly, and that one that coded only x and y coordinates would perform better. By defining a richer template there is potential for elevated accuracy. The question inevitably becomes whether a new standard offers error rates comparable with proprietary templates. This concept is variously termed[8] *sufficiency* and *performance* of a data interchange format. This concept is distinct from the second issue, interoperability, which involves the exchange of samples between vendors' implementations of the standard. Instead, sufficiency is quantified by considering whether vendors can generate and successfully match their own standard templates. If leading vendors can match their "native" standard templates as accurately as their own proprietary templates, then it can be said that the new format is sufficient.

Results comparing native MIN:A and MIN:B performance with proprietary performance are presented in Tables 6 and 7. For three reasons, these results are at least as important as the interoperabilty results that constitute the bulk of this report. First, the results publicly quantify the relative efficacy of the basic $(x, y, \theta, type, quality)$ representation. Second, many applications built on interoperable templates will likely execute many native comparisons. This will depend on the number of vendors producing templates and on how organizations procure and deploy products. Third, some applications might allow entirely proprietary data to be co-located within the standard minutiae record[8].

Table 6 gives FNMR for FMR values of 0.01 and 0.001, while Table 7 gives FMR for FNMR values of 0.02 and 0.01. Broadly similar patterns present themselves:

1. Proprietary templates are superior to native MIN:A templates in all cases except F and L. The number of errors is sometimes substantially lower. For the three most accurate systems (A,D and G), there are broadly two times as many rejections using standard templates vs. proprietary ones (e.g. FNMR = 0.013 vs 0.005, single finger, vendor G). When FNMR is fixed at 0.01, the situation is worse: For the leading matchers, there is about a factor of 10 more false matches.
2. These performance reductions can be offset by using two fingers in all authentication attempts. Two finger performance using native MIN:A templates is an order of magnitude superior to single finger proprietary performance but again inferior to the proprietary two-finger operation. Vendor F is an exception.
3. Differences between vendors are significant. Some systems match two-finger proprietary templates less accurately than the leading matcher does with its single-finger MIN:A template.
4. Performance from a native MIN:B template offers little performance improvement. Generally, it does not approach that of the same vendor's proprietary template. In native operation, the MIN:B template from vendors A, E, F and I outperforms their MIN:A template, but the accuracy is still inferior to that obtained from their proprietary templates.

[8] Storage of opaque proprietary data is provided for in the extended data fields of the INCITS 378 template, per the type field of clause 6.6.1.2. MINEX disallowed such data. An application may store proprietary data out-of-band.

Indeed, proprietary templates from vendors A, D and E produce half as many false non-matches as their MIN:B templates, while vendor K produces a sixth as many. Vendor F is an exception because it does produce lower FNMR . The performance of vendor D with MIN:A and MIN:B templates is equal.

5. Vendor C's MIN:A performance is identical to that of its proprietary implementation because the company adopted the MIN:A templates as its proprietary implementation.
6. While vendor F matches its MIN:B templates with lower error rate than its MIN:A template, the proprietary template unexpectedly underperforms both MIN:A and MIN:B .
7. Vendor K has a higher error rate with its MIN:B template than with MIN:A .

The MIN:B template contains more information than the MIN:A template but still codes substantially less information than is typically represented proprietary templates. Based solely on information content, one would expect that the accuracy of the templates, as measured by the FNMR at some fixed FMR, to be best for proprietary templates, intermediate for the MIN:B templates, and least accurate for MIN:A templates. This expectation is upheld in the data in that four of six do show better performance. The conclusion is that the MIN:B template provides only a minimal increase in accuracy over the MIN:A template and is substantially less accurate than the proprietary templates of the most accurate three vendors.

5.7 Interoperabilty

The success of a proprietary biometric system rests on its ability to consistently extract uniquely identifying information from repeated time-separated acquistions of a subject's biometric. For fingerprints, minutiae based systems are but one means of doing this. Systems must find minutiae in an enrollment sample, store them, and compare that later with minutiae from an authentication sample.

The task becomes harder when interoperability is required because interchange is now mediated by a standardized format. Difficulty may arise because template generation products may systematically interpret a common input differently. For MINEX interoperability will be achieved if two separately developed implementations can locate minutiae in two separate imagings of a finger to produce records similar enough for matching. This problem is a step beyond the traditional interpretation of different patterns, which combines fingerprint reacquisition with interpretation errors. It is implicit that the generators must select at least some of the same minutiae because even if they place the minutiae and determine their angles and type equivalently, then interoperability might still not be achieved. This report examines three interoperable verification scenarios.

5.7.1 Scenario 1

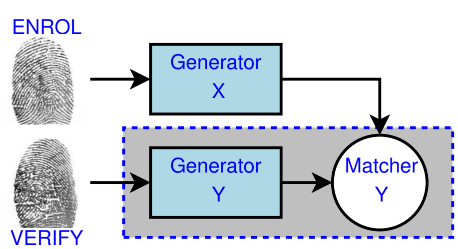


Figure 2: Scenario 1

In the scenario depicted in Figure 2, the enrollment template is prepared by product X and later used in a verification transaction in which the authentication template is prepared and matched by product Y. This is the most relevant scenario because it reflects the typical access control situation in which product Y's generator and matcher products are bundled together. The top half of the figure covers enrollment, while the grey box in the lower half indicates the coupling of the authentication template generator and the matcher.

Scenario 1 results for single and two-finger matching on the POEBVA dataset are presented in Tables 8 and 9, which present FNMR results at a fixed FMR of 0.01 for the proprietary, MIN:A , and MIN:B templates. Analogous tables for the other datasets are included in in the accompanying MINEX Supplement B - Scenario 1 Interoperability document. The cells in the scenario 1 tables are colored green when performance of the matcher on its own template is improved by using another generator's template. Such occurences are rare, indicating some intrinsic advantage to native generation and comparison. In scenario 1, the authentication process involves comparison of two standard templates (e.g. two MIN:A templates). But commercially, the authentication template need not conform to a standard because it exists

(a) Proprietary

NF = 1	A	B	C	D	E	F	G	H	I	J	K	L	M	N
FNMR	0.0089	0.0189	0.0225	0.0089	0.0251	0.0337	0.0047	0.1004	0.0329	0.1503	0.0186	0.0575	0.0358	0.0481
Rank	2	5	6	3	7	9	1	13	8	14	4	12	10	11

(b) MIN:A

NF = 1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	Mean	Rank	Med.	Rank
A	0.0136	0.0549	0.0458	0.0225	0.0641	0.0459	0.0417	0.0834	0.0334	0.1707	0.0747	0.0659	0.0792	0.0966	0.0637	6	0.0595	10
B	0.0218	0.0251	0.0385	0.0173	0.0402	0.0382	0.0192	0.1136	0.0336	0.1501	0.1599	0.0506	0.0442	0.0561	0.0578	2	0.0394	2
C	0.0357	0.0428	0.0225	0.0204	0.0519	0.0225	0.0348	0.1969	0.0484	0.3034	0.2451	0.0743	0.0493	0.0691	0.0869	8	0.0489	6
D	0.0207	0.0357	0.0301	0.0140	0.0485	0.0303	0.0316	0.0945	0.0392	0.2013	0.1218	0.0655	0.0551	0.0582	0.0605	4	0.0438	3
E	0.0236	0.0365	0.0340	0.0225	0.0301	0.0341	0.0286	0.0874	0.0476	0.1885	0.0896	0.0600	0.0557	0.0397	0.0556	1	0.0381	1
F	0.0359	0.0430	0.0222	0.0206	0.0522	0.0224	0.0345	0.1967	0.0485	0.3038	0.2456	0.0743	0.0493	0.0686	0.0870	9	0.0489	7
G	0.0300	0.0291	0.0447	0.0205	0.0390	0.0441	0.0129	0.0905	0.0441	0.1747	0.1632	0.0559	0.0419	0.0526	0.0602	3	0.0441	4
H	0.0437	0.1336	0.1212	0.0656	0.1860	0.1215	0.1339	0.1027	0.0796	1.0000	0.1748	0.1181	0.1818	0.2296	0.1923	14	0.1276	14
I	0.0397	0.0806	0.0830	0.0518	0.1062	0.0828	0.0548	0.2227	0.0348	0.2470	0.2756	0.0791	0.1030	0.1383	0.1142	12	0.0829	12
J	0.0403	0.0602	0.0939	0.0455	0.0987	0.0943	0.0489	0.2852	0.0542	0.1505	0.7314	0.0773	0.1026	0.1169	0.1428	13	0.0941	13
K	0.0188	0.0593	0.0476	0.0280	0.0661	0.0467	0.0428	0.0790	0.0400	0.1920	0.0461	0.0770	0.0885	0.1015	0.0667	7	0.0534	9
L	0.0467	0.0558	0.0704	0.0428	0.0822	0.0708	0.0432	0.1640	0.0485	0.1901	0.2375	0.0524	0.0866	0.0938	0.0918	10	0.0706	11
M	0.0496	0.0493	0.0455	0.0307	0.0545	0.0454	0.0327	0.2066	0.0616	0.3022	0.3929	0.0868	0.0359	0.0855	0.1057	11	0.0520	8
N	0.0368	0.0436	0.0428	0.0293	0.0458	0.0428	0.0393	0.1019	0.0497	0.1945	0.0865	0.0682	0.0621	0.0486	0.0637	5	0.0472	5
Mean	0.0326	0.0535	0.0530	0.0308	0.0690	0.0530	0.0428	0.1447	0.0474	0.2692	0.2175	0.0718	0.0740	0.0896				
Rank	2	7	6	1	8	5	3	12	4	14	13	9	10	11				
Med.	0.0358	0.0465	0.0451	0.0253	0.0533	0.0447	0.0370	0.1082	0.0480	0.1932	0.1690	0.0712	0.0589	0.0773				
Rank	2	6	5	1	8	4	3	12	7	14	13	10	9	11				

(c) MIN:B

NF = 1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	Mean	Rank	Med.	Rank
A	0.0135			0.0225	0.0685	0.0398			0.0333		0.2097				0.0646	2	0.0366	3
B																		
C																		
D	0.0206			0.0140	0.0554	0.0264			0.0417		0.3480				0.0844	3	0.0341	2
E	0.0234			0.0225	0.0296	0.0298			0.0466		0.2027				0.0591	1	0.0297	1
F	0.0354			0.0207	0.0885	0.0199			0.0486		0.8705				0.1806	5	0.0420	4
G																		
H																		
I	0.0389			0.0518	0.1073	0.0765			0.0336		0.4839				0.1320	4	0.0641	6
J																		
K	0.0189			0.0281	0.6535	0.0463			0.8615		0.0634				0.2786	6	0.0549	5
L																		
M																		
N																		
Mean	0.0251			0.0266	0.1671	0.0398			0.1776		0.3630							
Rank	1			2	4	3			5		6							
Med.	0.0220			0.0225	0.0785	0.0348			0.0442		0.2788							
Rank	1			2	5	3			4		6							

Table 8: Scenario 1 Interoperability for Single-finger Authentication
The FNMR at FMR of 0.01 for the POEBVA data. The vendor identified in the row produced the enrollment template. The vendor identified in each column produced the authentication template and performed the comparison. Cells are colored green when $F_{ij} < F_{jj}$. The tables refer, from top to bottom, to the proprietary, MIN:A and MIN:B templates.

(a) Proprietary

NF = 2	A	B	C	D	E	F	G	H	I	J	K	L	M	N
FNMR	0.0006	0.0018	0.0032	0.0007	0.0030	0.0061	0.0002	0.0367	0.0051	0.0704	0.0054	0.0126	0.0060	0.0082
Rank	2	4	6	3	5	10	1	13	7	14	8	12	9	11

(b) MIN:A

NF = 2	A	B	C	D	E	F	G	H	I	J	K	L	M	N	Mean	Rank	Med.	Rank
A	0.0011	0.0092	0.0080	0.0021	0.0130	0.0079	0.0049	0.0248	0.0049	0.0755	0.0292	0.0161	0.0213	0.0210	0.0171	4	0.0111	8
B	0.0027	0.0024	0.0072	0.0018	0.0073	0.0071	0.0017	0.0456	0.0049	0.0589	0.0588	0.0105	0.0083	0.0096	0.0162	1	0.0072	2
C	0.0052	0.0057	0.0032	0.0025	0.0104	0.0031	0.0039	0.0851	0.0081	0.1571	0.1048	0.0177	0.0099	0.0133	0.0307	8	0.0090	6
D	0.0021	0.0046	0.0045	0.0013	0.0097	0.0044	0.0035	0.0325	0.0062	0.0877	0.0442	0.0154	0.0126	0.0103	0.0171	5	0.0080	4
E	0.0025	0.0061	0.0056	0.0028	0.0045	0.0054	0.0035	0.0280	0.0102	0.0900	0.0362	0.0163	0.0147	0.0053	0.0165	2	0.0059	1
F	0.0054	0.0060	0.0032	0.0025	0.0103	0.0031	0.0038	0.0855	0.0081	0.1597	0.1058	0.0177	0.0097	0.0131	0.0310	9	0.0089	5
G	0.0040	0.0032	0.0085	0.0022	0.0061	0.0085	0.0007	0.0308	0.0068	0.0715	0.0693	0.0116	0.0074	0.0084	0.0171	3	0.0079	3
H	0.0084	0.0421	0.0393	0.0143	0.0767	0.0395	0.0384	0.0422	0.0210	0.9999	0.0724	0.0413	0.0753	0.0831	0.1139	14	0.0417	14
I	0.0073	0.0184	0.0252	0.0100	0.0333	0.0249	0.0083	0.1137	0.0056	0.1206	0.1170	0.0207	0.0313	0.0415	0.0413	12	0.0250	12
J	0.0077	0.0119	0.0259	0.0082	0.0276	0.0257	0.0070	0.1572	0.0103	0.0640	0.5736	0.0198	0.0296	0.0297	0.0713	13	0.0258	13
K	0.0018	0.0130	0.0108	0.0051	0.0134	0.0109	0.0049	0.0280	0.0068	0.0929	0.0275	0.0270	0.0313	0.0234	0.0212	7	0.0132	10
L	0.0115	0.0109	0.0218	0.0097	0.0258	0.0213	0.0066	0.0795	0.0105	0.0861	0.1123	0.0113	0.0267	0.0254	0.0328	10	0.0216	11
M	0.0099	0.0096	0.0106	0.0049	0.0116	0.0108	0.0039	0.1007	0.0134	0.1573	0.1929	0.0247	0.0061	0.0211	0.0413	11	0.0112	9
N	0.0063	0.0077	0.0086	0.0042	0.0094	0.0087	0.0056	0.0368	0.0104	0.0862	0.0353	0.0169	0.0157	0.0081	0.0186	6	0.0090	7
Mean	0.0054	0.0108	0.0130	0.0051	0.0185	0.0129	0.0069	0.0636	0.0091	0.1648	0.1128	0.0191	0.0214	0.0224				
Rank	2	5	7	1	8	6	3	12	4	14	13	9	10	11				
Med.	0.0053	0.0084	0.0086	0.0035	0.0110	0.0086	0.0044	0.0439	0.0081	0.0889	0.0709	0.0173	0.0152	0.0171				
Rank	3	5	6	1	8	7	2	12	4	14	13	11	9	10				

(c) MIN:B

NF = 2	A	B	C	D	E	F	G	H	I	J	K	L	M	N	Mean	Rank	Med.	Rank
A	0.0010			0.0021	0.0152	0.0062			0.0046		0.0758				0.0175	2	0.0054	3
B																		
C																		
D	0.0021			0.0013	0.0124	0.0039			0.0067		0.1586				0.0308	3	0.0053	2
E	0.0024			0.0028	0.0043	0.0048			0.0099		0.0787				0.0172	1	0.0046	1
F	0.0055			0.0026	0.0249	0.0028			0.0076		0.7646				0.1347	5	0.0065	4
G																		
H																		
I	0.0071			0.0100	0.0350	0.0233			0.0054		0.2717				0.0587	4	0.0167	5
J																		
K	0.0017			0.0051	0.4804	0.0107			0.8070		0.0305				0.2226	6	0.0206	6
L																		
M																		
N																		
Mean	0.0033			0.0040	0.0954	0.0086			0.1402		0.2300							
Rank	1			2	4	3			5		6							
Med.	0.0023			0.0027	0.0200	0.0055			0.0071		0.1187							
Rank	1			2	5	3			4		6							

Table 9: Scenario 1 Interoperability for Two-finger Authentication

The FNMR at FMR of 0.01 for the POEBVA data. The vendor identified in the row produced the enrollment template. The vendor identified in each column produced the authentication template and performed the comparison. Cells are colored green when $F_{ij} < F_{jj}$, i.e. when the matcher performs better on the foreign enrollment templates than on its own ones. The tables refer, from top to bottom, to the proprietary, MIN:A and MIN:B templates.

ephemerally and the only requirement is that it should be *matchable* against a standard template. This constraint is likely to be quite limiting and it is unclear to what extent vendors can and do supplement the core $(x, y, \theta, type, quality)$ information of the authentication template. Further, the relative benefit of matching such a proprietary template against a standard template has not been independently reported. In any case, the MINEX results establish an upper bound on the error rates inherent in the use of standard templates from one or two fingers.

Note that this report does not include data to substantiate the assertion that better performance will be obtained if the template generation is tailored to a particular sensor, or class of sensor. The means for achieving this lies in the potential for the generator to invoke different image processing algorithms, specifically for the imaging and noise properties of the sensor and less easily for the capture environment. Although the MINEX API specification did not provide a sensor identifier to the generator, there is provision for such information in fingerprint image records (e.g. INCITS 381 [5] clauses 7.1.4 and 7.1.5). In MINEX , all the input images were captured and stored at a resolution of 500 pixels/inch (197 pixels/cm) using single-finger optical sensors typical of existing government systems. Performance available from other classes of sensor (e.g. capacitative) are likely to be different than those reported here. Particularly, comparison of templates derived from optically and non-optically sensed images would add another dimension to the interoperability space, and may therefore add to error rates.

5.7.2 Threshold Setting for Interoperable Templates

The FNMR and FMR can be traded off against each other by setting the matcher threshold. The threshold implements the performance requirement. For verification applications, this usually entails some balancing of the inconvenience associated with false rejections and the security implications of false acceptances. The threshold is set by consulting a table of the points of the DET characteristic. For example, to achieve an expected FMR of 0.01, the threshold must be set to a specific value, t_0. The scale of this value is arbitrary and depends on the underlying matching algorithm and on any internal transformation of the raw value. Some implementations "pre-normalize" their raw scores onto an output range of $[0, 1]$, and operators and implementers regard these as estimates of the likelihood of a false match. In either case (arbitrary or normalized), the error rates that are ultimately observed will depend on a number of variables. These include the kind of sensor used, the imaging environment, subject behaviour, and the subjects themselves. When the verification process is mediated by a standardized template, another variable enters the mix, namely the generator of the template. Thus, for any of the matchers in the scenario 1 study (i.e. any column of Table 8), the performance values for each row were achieved at generally different thresholds.

The dependence of matching performance on the template generator should not be unexpected. A template generator that in some sense fails to faithfully represent the input image (e.g. by misplacement of minutiae) would yield degraded performance. The INCITS 378 standard neither requires nor defines any part of the image processing and minutiae detection chain that govern *how* a template should be formed. Instead, it gives normative requirements on *what* a minutiae is. The MINEX API did regulate the angle computation in terms of skeletonization of ridges (which INCITS 378 did not), but vendors were free to implement this internally in any way they saw fit. The result is that the format is well defined, but that *how* that format is instantiated is largely (and rightly) not, and the result is a compromise between performance and innovation on the one side and varying interoperability on the other.

It therefore should be expected that the matching performance should depend on the "quality" of the input templates and that this dependence on the template generator presents a threshold setting policy problem in interoperable scenarios. Two approaches to the problem are as follows:

- **Fixed threshold:** The effect of fixing the threshold to one permanent value is that both error rates will float. This situation is represented in Table 10. There, the threshold is set such that the performance of the matcher on its own templates gives FMR of 0.01 [9] as shown on the diagonal of the matrix. The two values in each cell are the FMR on top and the associated FNMR directly below. The off-diagonal error rates are computed at that same threshold, and both FMR and FNMR vary. The red color coding indicates a degradation of FMR and/or FNMR relative to the native performance. The prevalence of this occurence makes it clear that rejection increases both for impostors, as quantified by FMR (which goes down), and for genuine users, as stated by FNMR (which goes up). This holds for single-finger 9(a) and two-finger authentication 9(b). The consequence of this is that to preserve *overall* FNMR performance the threshold will have to decrease.

[9] The observant reader will notice that in some cells (such as NN) the FMR of 0.01 could not be precisely achieved. This is a consequence of the matcher reporting discrete scores (usually integral values) such that the quantile function of the impostor distribution when evaluated at $Q(1-0.01)$ gives a threshold t for which the number of false matches *higher* than this is fewer than 0.01 times the total.

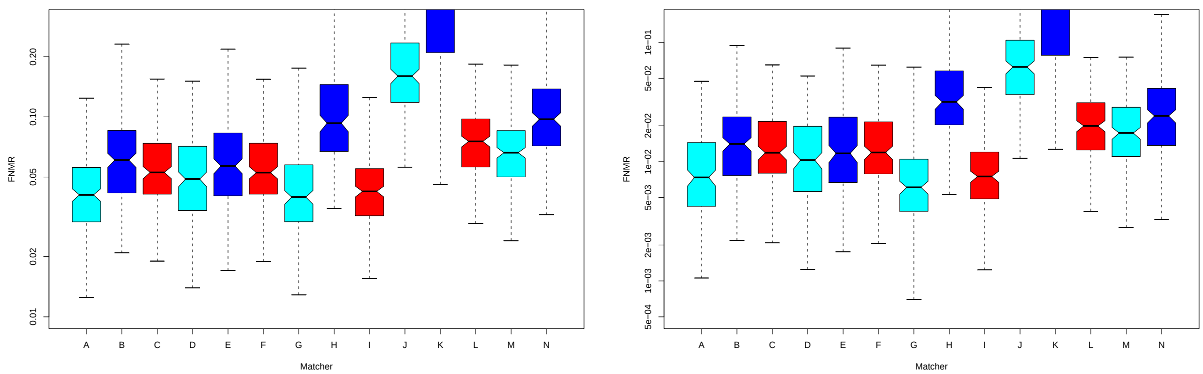
(a) Single finger

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	0.010	0.007	0.008	0.008	0.007	0.008	0.003	0.012	0.010	0.002	0.006	0.009	0.007	0.006
	0.014	0.061	0.049	0.024	0.071	0.049	0.042	0.081	0.034	0.244	0.088	0.067	0.084	0.109
B	0.005	0.010	0.009	0.009	0.008	0.009	0.004	0.008	0.009	0.001	0.005	0.010	0.010	0.007
	0.024	0.025	0.040	0.018	0.044	0.039	0.019	0.120	0.035	0.230	0.191	0.050	0.044	0.060
C	0.003	0.007	0.010	0.008	0.007	0.010	0.003	0.004	0.006	0.001	0.003	0.007	0.011	0.007
	0.043	0.047	0.022	0.021	0.057	0.022	0.035	0.224	0.056	0.463	0.300	0.079	0.049	0.074
D	0.005	0.005	0.007	0.010	0.007	0.007	0.002	0.007	0.006	0.001	0.003	0.006	0.008	0.005
	0.023	0.043	0.033	0.014	0.053	0.033	0.032	0.104	0.044	0.318	0.152	0.071	0.058	0.065
E	0.007	0.007	0.017	0.010	0.010	0.017	0.003	0.011	0.008	0.001	0.005	0.008	0.008	0.009
	0.025	0.040	0.029	0.023	0.030	0.029	0.029	0.086	0.050	0.315	0.111	0.062	0.059	0.040
F	0.003	0.007	0.010	0.008	0.007	0.010	0.003	0.004	0.005	0.001	0.003	0.007	0.011	0.007
	0.043	0.048	0.022	0.021	0.057	0.022	0.034	0.224	0.056	0.464	0.300	0.079	0.048	0.073
G	0.005	0.008	0.012	0.009	0.008	0.012	0.005	0.012	0.009	0.001	0.005	0.009	0.008	0.007
	0.033	0.031	0.042	0.021	0.041	0.041	0.013	0.088	0.045	0.282	0.195	0.057	0.044	0.056
H	0.008	0.002	0.006	0.006	0.006	0.006	0.001	0.010	0.006	0.000	0.002	0.004	0.002	0.003
	0.046	0.195	0.134	0.073	0.218	0.134	0.134	0.103	0.088	1.000	0.222	0.138	0.233	0.269
I	0.004	0.005	0.007	0.007	0.007	0.007	0.004	0.008	0.010	0.002	0.002	0.009	0.006	0.007
	0.045	0.096	0.090	0.056	0.120	0.090	0.055	0.228	0.035	0.346	0.346	0.082	0.113	0.148
J	0.002	0.009	0.001	0.005	0.006	0.001	0.002	0.001	0.006	0.010	0.000	0.009	0.010	0.005
	0.053	0.062	0.149	0.051	0.114	0.149	0.049	0.339	0.060	0.150	0.755	0.079	0.102	0.130
K	0.008	0.005	0.009	0.007	0.006	0.009	0.003	0.010	0.007	0.001	0.010	0.008	0.006	0.006
	0.019	0.070	0.048	0.030	0.074	0.048	0.043	0.078	0.042	0.296	0.046	0.081	0.097	0.114
L	0.004	0.006	0.007	0.006	0.006	0.007	0.004	0.008	0.008	0.002	0.003	0.010	0.008	0.006
	0.055	0.063	0.077	0.046	0.092	0.077	0.043	0.171	0.051	0.268	0.306	0.052	0.090	0.104
M	0.002	0.005	0.006	0.008	0.007	0.006	0.003	0.004	0.004	0.001	0.001	0.006	0.010	0.005
	0.062	0.059	0.051	0.032	0.061	0.050	0.033	0.236	0.075	0.445	0.469	0.095	0.036	0.094
N	0.005	0.007	0.013	0.009	0.009	0.013	0.003	0.009	0.008	0.001	0.006	0.009	0.009	0.008
	0.041	0.048	0.040	0.030	0.047	0.040	0.039	0.104	0.052	0.311	0.101	0.069	0.063	0.049

(b) Two fingers

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	0.010	0.007	0.007	0.008	0.007	0.007	0.005	0.014	0.010	0.001	0.005	0.009	0.007	0.006
	0.001	0.011	0.009	0.002	0.016	0.009	0.005	0.022	0.005	0.135	0.033	0.016	0.023	0.025
B	0.004	0.010	0.009	0.008	0.008	0.009	0.008	0.010	0.009	0.000	0.004	0.010	0.010	0.008
	0.003	0.002	0.007	0.002	0.008	0.007	0.002	0.046	0.005	0.122	0.072	0.011	0.008	0.010
C	0.003	0.007	0.010	0.008	0.008	0.010	0.005	0.003	0.005	0.000	0.003	0.007	0.011	0.007
	0.007	0.007	0.003	0.003	0.012	0.003	0.004	0.109	0.010	0.323	0.142	0.020	0.010	0.015
D	0.004	0.005	0.007	0.010	0.007	0.007	0.004	0.008	0.006	0.000	0.003	0.006	0.008	0.005
	0.003	0.006	0.005	0.001	0.011	0.005	0.004	0.035	0.008	0.195	0.057	0.018	0.013	0.013
E	0.007	0.007	0.019	0.009	0.010	0.019	0.006	0.012	0.008	0.000	0.004	0.008	0.007	0.010
	0.003	0.007	0.004	0.003	0.005	0.004	0.004	0.026	0.011	0.197	0.044	0.017	0.016	0.005
F	0.003	0.007	0.010	0.008	0.007	0.010	0.005	0.003	0.005	0.000	0.003	0.007	0.011	0.007
	0.008	0.007	0.003	0.003	0.012	0.003	0.004	0.108	0.010	0.323	0.141	0.020	0.009	0.014
G	0.005	0.008	0.013	0.008	0.008	0.013	0.010	0.016	0.009	0.000	0.005	0.009	0.008	0.008
	0.005	0.004	0.008	0.002	0.007	0.008	0.001	0.026	0.007	0.160	0.085	0.012	0.008	0.008
H	0.008	0.001	0.006	0.006	0.005	0.006	0.002	0.008	0.005	0.000	0.001	0.003	0.001	0.003
	0.009	0.077	0.046	0.017	0.096	0.047	0.039	0.042	0.026	1.000	0.114	0.054	0.112	0.112
I	0.003	0.005	0.007	0.005	0.007	0.007	0.007	0.010	0.010	0.001	0.002	0.008	0.006	0.007
	0.009	0.024	0.029	0.012	0.040	0.028	0.009	0.113	0.006	0.211	0.181	0.022	0.036	0.044
J	0.002	0.010	0.001	0.004	0.006	0.001	0.004	0.001	0.006	0.010	0.000	0.008	0.010	0.005
	0.012	0.012	0.056	0.010	0.034	0.055	0.007	0.201	0.012	0.064	0.622	0.021	0.029	0.035
K	0.008	0.005	0.009	0.006	0.006	0.009	0.005	0.011	0.007	0.000	0.010	0.007	0.005	0.006
	0.002	0.016	0.011	0.006	0.017	0.011	0.005	0.027	0.007	0.186	0.027	0.029	0.034	0.027
L	0.003	0.006	0.006	0.005	0.006	0.006	0.007	0.009	0.007	0.001	0.002	0.010	0.008	0.006
	0.015	0.013	0.025	0.011	0.031	0.024	0.007	0.082	0.012	0.150	0.164	0.011	0.028	0.028
M	0.002	0.005	0.006	0.007	0.007	0.006	0.006	0.003	0.003	0.000	0.001	0.005	0.010	0.004
	0.015	0.012	0.013	0.005	0.014	0.013	0.004	0.125	0.018	0.308	0.297	0.030	0.006	0.025
N	0.005	0.007	0.014	0.008	0.008	0.014	0.006	0.010	0.008	0.000	0.005	0.008	0.009	0.009
	0.008	0.008	0.008	0.004	0.010	0.008	0.006	0.037	0.011	0.188	0.041	0.018	0.016	0.008

Table 10: The effect of a fixed threshold on Scenario 1 Interoperability
For MIN:A the values in each cell are the FMR (above) and FNMR (below) at a fixed threshold. The threshold on each matcher is set to achieve a FMR of 0.01 on its own templates. This is scenario 1, so the vendor identified in the row produced the enrollment template, and the vendor identified in the column produced the authentication template and performed the comparison. The red color indicates performance that is worse than intended FMR (i.e. > 0.01) or FNMR worse than the native value which is marked in yellow.



(a) Single finger (b) Two fingers

Figure 3: Variation in Scenario 2 Matcher Accuracy.
Boxplots of FNMR at a fixed FMR of 0.01 for fourteen matchers processing MIN:A templates from all 14^2 possible pairs of templates generators including their own. The dataset is POEBVA . The colors are used cyclically for clarity, but otherwise have no meaning.

- **Source-varying threshold:** If a mechanism is instituted to dynamically tailor the threshold to the input templates, then an operational performance target such as FMR = 0.01 may be maintained. This possibility is supported by the recording an identifier for the product in the template generator field of the INCITS 378 header (see *Product Identifier*, INCITS 378 *clause 6.4.4*). In the general case, the matcher would apply a threshold tailored to *both* the enrollment and authentication template generators. The latter may well be supplied by the matcher vendor. In any case, while the product identifier field would be necessary, it would not be sufficient. Additional "calibration" information would be required. This would be, in essence, a table of threshold and false match rate for each interoperable template generator pair.

5.7.3 Interoperability of the MIN:B Template

Results for the MIN:B template on the POEBVA dataset are given as interoperabilty matrices in Tables 8 and 9.

Comparisons of interoperability between the MIN:A and MIN:B templates are easier if vendor K is ignored in the MIN:B template interoperability table. Vendor K gives very poor performance on templates from generators F and I, and performs poorly on templates from E and I. When K is ignored in two-finger matching (Table 9) the three most accurate template generators (A, D, and E) have mean error rates that are very similar for MIN:A and MIN:B templates. In most cases the MIN:B template is less interoperable than the MIN:A template. The results for the POE dataset follow the same trend.

The results presented here are for verification and are not directly applicable to one-to-many applications. One-to-many tasks usually have higher error rates than one-to-one applications with comparable quality of biometric input because algorithms are optimized for speed, and must handle more stringent false matching requirements. In various potential scenarios for the use of interoperable identification systems such as IDENT and IAFIS, the use of templates similar to MIN:B has been proposed. However, one-to-many systems augment the matching process with a number of enhancements, such as ridge flow information. As yet, there is no standardized format for such data. It is unlikely therefore that even though interoperability of the MIN:A and MIN:B templates is a necessary condition for large-scale one-to-many operations, it is not sufficient unless the images themselves are available to the matching process.

The MIN:B template was advanced as a compromise between the simplicity of MIN:A and the power of proprietary templates. However, the conclusion is that MIN:B does not currently offer a middle-ground because false non-match rates are two or three times higher than those available from proprietary templates, and this will necessitate a change in threshold to contain FMR .

(a) Single finger

D	A	B	C	D	E	F	G	H	I	J	K	L	M	N	Median	Rank
A	0.019	0.034	0.038	0.023	0.047	0.038	0.054	0.059	0.041	0.061	0.026	0.044	0.058	0.038	0.040	6
B	0.028	0.021	0.029	0.017	0.034	0.028	0.028	0.071	0.053	0.060	0.040	0.038	0.036	0.031	0.032	1
C	0.039	0.034	0.019	0.020	0.040	0.019	0.047	0.106	0.072	0.110	0.052	0.055	0.037	0.034	0.039	5
D	0.026	0.025	0.023	0.014	0.035	0.023	0.034	0.070	0.049	0.069	0.033	0.038	0.034	0.028	0.033	2
E	0.034	0.030	0.029	0.023	0.030	0.030	0.037	0.083	0.058	0.070	0.042	0.042	0.034	0.026	0.034	4
F	0.040	0.035	0.019	0.021	0.040	0.019	0.047	0.107	0.072	0.110	0.052	0.055	0.037	0.033	0.040	7
G	0.037	0.024	0.033	0.021	0.034	0.033	0.024	0.087	0.068	0.067	0.049	0.047	0.034	0.032	0.034	3
H	0.056	0.085	0.108	0.066	0.119	0.109	0.127	0.081	0.084	0.109	0.071	0.078	0.151	0.104	0.094	14
I	0.046	0.070	0.078	0.052	0.085	0.078	0.100	0.097	0.053	0.113	0.059	0.072	0.114	0.073	0.076	13
J	0.044	0.050	0.079	0.045	0.067	0.079	0.065	0.086	0.076	0.073	0.060	0.063	0.096	0.063	0.066	12
K	0.025	0.044	0.048	0.028	0.057	0.049	0.066	0.071	0.054	0.078	0.028	0.060	0.072	0.046	0.051	10
L	0.050	0.053	0.062	0.043	0.067	0.062	0.073	0.089	0.071	0.099	0.068	0.045	0.087	0.058	0.064	11
M	0.055	0.042	0.034	0.031	0.044	0.035	0.046	0.136	0.102	0.127	0.073	0.071	0.030	0.041	0.045	8
N	0.044	0.042	0.039	0.029	0.041	0.039	0.051	0.102	0.070	0.094	0.054	0.053	0.050	0.034	0.047	9
Median	0.039	0.038	0.036	0.025	0.043	0.037	0.049	0.087	0.069	0.086	0.052	0.054	0.043	0.036		
Rank	6	5	3	1	7	4	9	14	12	13	10	11	8	2		0.049

(b) Two fingers

D	A	B	C	D	E	F	G	H	I	J	K	L	M	N	Median	Rank
A	0.002	0.004	0.005	0.002	0.008	0.005	0.009	0.013	0.007	0.014	0.006	0.008	0.011	0.005	0.006	5
B	0.004	0.002	0.004	0.002	0.006	0.004	0.004	0.017	0.010	0.015	0.009	0.008	0.006	0.005	0.005	2
C	0.006	0.006	0.003	0.003	0.008	0.002	0.009	0.033	0.019	0.040	0.014	0.014	0.007	0.006	0.007	6
D	0.003	0.003	0.003	0.001	0.006	0.003	0.004	0.016	0.009	0.017	0.008	0.007	0.005	0.004	0.005	1
E	0.005	0.005	0.005	0.003	0.005	0.005	0.006	0.023	0.013	0.018	0.012	0.009	0.006	0.004	0.006	4
F	0.006	0.005	0.003	0.003	0.007	0.002	0.009	0.033	0.019	0.040	0.014	0.014	0.007	0.006	0.007	6
G	0.005	0.003	0.005	0.002	0.006	0.005	0.003	0.024	0.016	0.018	0.013	0.011	0.006	0.005	0.005	3
H	0.012	0.020	0.030	0.014	0.038	0.031	0.037	0.022	0.021	0.035	0.023	0.021	0.051	0.030	0.026	14
I	0.009	0.018	0.022	0.010	0.025	0.022	0.031	0.031	0.013	0.042	0.017	0.022	0.040	0.020	0.022	13
J	0.008	0.009	0.022	0.008	0.015	0.022	0.015	0.025	0.021	0.020	0.016	0.017	0.031	0.015	0.017	11
K	0.004	0.009	0.011	0.005	0.014	0.010	0.016	0.020	0.012	0.022	0.007	0.016	0.019	0.010	0.011	10
L	0.012	0.013	0.019	0.010	0.020	0.019	0.022	0.029	0.023	0.038	0.023	0.012	0.031	0.017	0.020	12
M	0.012	0.009	0.006	0.005	0.009	0.006	0.010	0.049	0.034	0.052	0.023	0.023	0.006	0.009	0.010	9
N	0.009	0.008	0.008	0.004	0.009	0.008	0.010	0.032	0.019	0.030	0.016	0.014	0.011	0.006	0.010	8
Median	0.006	0.007	0.006	0.004	0.009	0.006	0.009	0.024	0.018	0.026	0.014	0.014	0.009	0.006		
Rank	5	6	3	1	7	2	9	13	12	14	11	10	8	4		0.010

Table 11: Scenario 2 Interoperability for Matcher D.
All values are FNMR at a fixed FMR of 0.01 for one and two-finger matching of MIN:A templates from POEBVA images. The vendor identified in each row makes the enrollment template; the vendor identified in each column makes the authentication template. The top table is for single-finger matching, with two-finger matching below. In all cases the MIN:A template and the standard POEBVA dataset are used. The value in the far bottom righthand corner is the median of the whole matrix.

(a) Matcher G

G	A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	0.042 0.018	0.042 0.030	0.042 0.038	0.042 0.025	0.042 0.043	0.042 0.038	0.042 0.042	0.042 0.074	0.042 0.033	0.042 0.064	0.042 0.023	0.042 0.038	0.042 0.054	0.042 0.036
B	0.019 0.024	0.019 0.016	0.019 0.025	0.019 0.018	0.019 0.030	0.019 0.025	0.019 0.019	0.019 0.084	0.019 0.035	0.019 0.055	0.019 0.030	0.019 0.028	0.019 0.029	0.019 0.025
C	0.035 0.038	0.035 0.027	0.035 0.015	0.035 0.021	0.035 0.040	0.035 0.015	0.035 0.035	0.035 0.123	0.035 0.057	0.035 0.105	0.035 0.043	0.035 0.043	0.035 0.030	0.035 0.031
D	0.032 0.030	0.032 0.026	0.032 0.026	0.032 0.017	0.032 0.040	0.032 0.026	0.032 0.032	0.032 0.097	0.032 0.045	0.032 0.083	0.032 0.032	0.032 0.039	0.032 0.036	0.032 0.031
E	0.029 0.031	0.029 0.027	0.029 0.030	0.029 0.025	0.029 0.028	0.029 0.030	0.029 0.029	0.029 0.101	0.029 0.045	0.029 0.070	0.029 0.036	0.029 0.036	0.029 0.032	0.029 0.025
F	0.034 0.038	0.034 0.027	0.034 0.015	0.034 0.021	0.034 0.040	0.034 0.015	0.034 0.034	0.034 0.124	0.034 0.057	0.034 0.105	0.034 0.043	0.034 0.043	0.034 0.030	0.034 0.031
G	0.013 0.027	0.013 0.015	0.013 0.026	0.013 0.017	0.013 0.026	0.013 0.026	0.013 0.013	0.013 0.090	0.013 0.038	0.013 0.052	0.013 0.033	0.013 0.030	0.013 0.023	0.013 0.022
H	0.134 0.072	0.134 0.102	0.134 0.127	0.134 0.092	0.134 0.136	0.134 0.129	0.134 0.134	0.134 0.121	0.134 0.094	0.134 0.147	0.134 0.080	0.134 0.091	0.134 0.175	0.134 0.123
I	0.055 0.033	0.055 0.042	0.055 0.058	0.055 0.040	0.055 0.060	0.055 0.058	0.055 0.055	0.055 0.095	0.055 0.031	0.055 0.080	0.055 0.039	0.055 0.043	0.055 0.078	0.055 0.049
J	0.049 0.043	0.049 0.043	0.049 0.078	0.049 0.051	0.049 0.063	0.049 0.078	0.049 0.049	0.049 0.110	0.049 0.058	0.049 0.062	0.049 0.052	0.049 0.049	0.049 0.083	0.049 0.057
K	0.043 0.022	0.043 0.032	0.043 0.042	0.043 0.026	0.043 0.046	0.043 0.042	0.043 0.043	0.043 0.080	0.043 0.038	0.043 0.068	0.043 0.021	0.043 0.045	0.043 0.055	0.043 0.037
L	0.043 0.038	0.043 0.035	0.043 0.045	0.043 0.036	0.043 0.051	0.043 0.046	0.043 0.043	0.043 0.091	0.043 0.043	0.043 0.072	0.043 0.049	0.043 0.027	0.043 0.061	0.043 0.041
M	0.033 0.047	0.033 0.032	0.033 0.029	0.033 0.028	0.033 0.040	0.033 0.029	0.033 0.033	0.033 0.147	0.033 0.073	0.033 0.110	0.033 0.054	0.033 0.054	0.033 0.022	0.033 0.034
N	0.039 0.042	0.039 0.036	0.039 0.037	0.039 0.032	0.039 0.039	0.039 0.036	0.039 0.039	0.039 0.120	0.039 0.053	0.039 0.088	0.039 0.045	0.039 0.043	0.039 0.044	0.039 0.031

(b) Matcher I

I	A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	0.033 0.021	0.033 0.031	0.033 0.035	0.033 0.023	0.033 0.048	0.033 0.035	0.033 0.047	0.033 0.059	0.033 0.033	0.033 0.056	0.033 0.028	0.033 0.041	0.033 0.047	0.033 0.035
B	0.034 0.028	0.034 0.016	0.034 0.025	0.034 0.017	0.034 0.036	0.034 0.025	0.034 0.023	0.034 0.062	0.034 0.034	0.034 0.045	0.034 0.037	0.034 0.030	0.034 0.027	0.034 0.027
C	0.048 0.039	0.048 0.028	0.048 0.016	0.048 0.019	0.048 0.048	0.048 0.017	0.048 0.042	0.048 0.095	0.048 0.048	0.048 0.079	0.048 0.046	0.048 0.043	0.048 0.030	0.048 0.032
D	0.039 0.029	0.039 0.024	0.039 0.024	0.039 0.016	0.039 0.044	0.039 0.024	0.039 0.036	0.039 0.069	0.039 0.039	0.039 0.062	0.039 0.035	0.039 0.038	0.039 0.033	0.039 0.031
E	0.048 0.035	0.048 0.030	0.048 0.034	0.048 0.027	0.048 0.033	0.048 0.034	0.048 0.036	0.048 0.081	0.048 0.048	0.048 0.065	0.048 0.041	0.048 0.042	0.048 0.034	0.048 0.027
F	0.048 0.039	0.048 0.028	0.048 0.016	0.048 0.019	0.048 0.048	0.048 0.017	0.048 0.042	0.048 0.094	0.048 0.048	0.048 0.079	0.048 0.046	0.048 0.043	0.048 0.030	0.048 0.032
G	0.044 0.035	0.044 0.018	0.044 0.030	0.044 0.020	0.044 0.036	0.044 0.030	0.044 0.019	0.044 0.071	0.044 0.044	0.044 0.047	0.044 0.044	0.044 0.035	0.044 0.027	0.044 0.028
H	0.080 0.063	0.080 0.078	0.080 0.104	0.080 0.067	0.080 0.119	0.080 0.104	0.080 0.112	0.080 0.089	0.080 0.080	0.080 0.101	0.080 0.073	0.080 0.080	0.080 0.124	0.080 0.096
I	0.035 0.039	0.035 0.042	0.035 0.053	0.035 0.035	0.035 0.068	0.035 0.053	0.035 0.062	0.035 0.079	0.035 0.035	0.035 0.070	0.035 0.045	0.035 0.048	0.035 0.066	0.035 0.050
J	0.054 0.044	0.054 0.037	0.054 0.065	0.054 0.042	0.054 0.063	0.054 0.065	0.054 0.047	0.054 0.073	0.054 0.054	0.054 0.051	0.054 0.054	0.054 0.047	0.054 0.062	0.054 0.051
K	0.040 0.027	0.040 0.037	0.040 0.042	0.040 0.028	0.040 0.053	0.040 0.042	0.040 0.054	0.040 0.067	0.040 0.040	0.040 0.064	0.040 0.026	0.040 0.050	0.040 0.054	0.040 0.041
L	0.048 0.047	0.048 0.039	0.048 0.050	0.048 0.037	0.048 0.064	0.048 0.050	0.048 0.054	0.048 0.078	0.048 0.048	0.048 0.065	0.048 0.058	0.048 0.033	0.048 0.058	0.048 0.047
M	0.062 0.047	0.062 0.032	0.062 0.030	0.062 0.027	0.062 0.046	0.062 0.030	0.062 0.038	0.062 0.105	0.062 0.062	0.062 0.079	0.062 0.058	0.062 0.051	0.062 0.024	0.062 0.036
N	0.050 0.043	0.050 0.036	0.050 0.037	0.050 0.030	0.050 0.043	0.050 0.037	0.050 0.046	0.050 0.089	0.050 0.050	0.050 0.073	0.050 0.049	0.050 0.045	0.050 0.042	0.050 0.032

Table 12: Scenario 2 - Benefit of Template Generator Substitution.
Values of FNMR at a fixed FMR of 0.01 for single-finger matching of MIN:A templates from POEBVA images. In each cell the top value is for scenario 1 - the matcher compares an authentication template of its own against an enrollment template from the row generator. The second value applies to scenario 2 - the matcher compares templates from the row *and* column generators. Cells are colored green when the generator identified in the column is substituted in place of the matcher vendor's own template generator.

5.7.4 Scenario 2

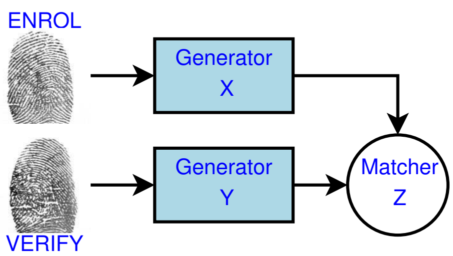


Figure 4: Scenario 2

This scenario represents the general case in which the enrollment template is prepared by vendor X, the verification template is prepared by vendor Y, and they are compared using vendor Z's matcher. This full three dimensional interoperability space supports the investigation of the extent to which the template generation and matching functions can be regarded as logically separated activities. Note that this scenario is the general cubic interoperability space and that scenario 1 is a special partition of it.

Thus, the scenario 2 interoperability matrix gives the performance of one matcher on all pairs of enrollment and authentication template generators (i.e. element F_{rc} is the FNMR at FMR = 0.01 for the comparison by the matcher of templates from enrollment generator r and column authentication generator c, where r denotes the row and c denotes the column). Example matrices, for matcher D, are presented in Table 11, and a full set is included in the accompanying document, MINEX Supplement C - Scenario 2 Interoperability. Scenario 2 matrices are not symmetric because the enrollment and authentication datasets are necessarily disjoint. Indeed, for the POEBVA set the two sets of images come from different sources.

The Scenario 2 tables are summarized in the boxplots of Figure 3. Each box-and-whisker shows the median, quartiles and extrema for the 14^2 possible pairs of templates that the given matcher may encounter. Thus, in Figure 3(a), it is apparent from the leftmost box that matcher A processed templates with median FNMR of about 0.04 and with best and worst cases of 0.014 and 0.13 where these figures apply to single-finger matching of MIN:A templates derived from the POEBVA dataset.

The matrices of Table 12 show the effect on accuracy of substituting the generator provided by the matcher vendor (i.e. the scenario 1 case) with one from another vendor. Table 11(a), for vendor G, and Table 11(b), for vendor I, are included here as examples and are typical: The prevalence of green in columns B, D, N and some others shows that replacement of the authentication template generator is often worthwhile. The reductions in FNMR can be substantial even off the diagonal.

A full set of tables is given in the accompanying document: MINEX Supplement D - Scenario 2 Template Generator Substitution.

5.7.5 Scenario 3

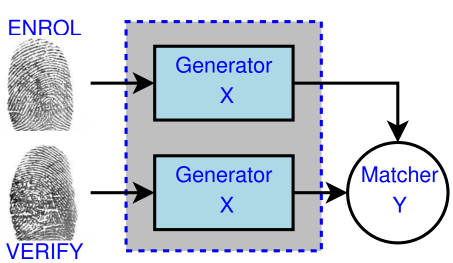


Figure 5: Scenario 3

Scenario 3 is the special case of interoperability for which the enrollment and authentication templates are generated by the *same* product, X, and ultimately matched by product Y. It is included here to examine if a matcher is better able to deal with "two-of-a-kind", than it does with one of its own and one other. Again this is commercially atypical because with a fully interoperable minutiae standard it should not be necessary to bind the enrollment template generation to that of the verification template. Note that this case could be deployed: it would necessitate the access controller possessing all possible template generators and, at verification time, invoking that generator that corresponds to the one identified in the header of the submitted enrollment template. Although this is technically viable, it is unlikely to be economically so. Note that it is not a fully proprietary concept because the matcher could still be procured from another vendor.

Scenario 3 is included in this report to support analysis of the extent to which a good template generator could improve accuracy. The result, for the POEBVA set in Table 13, is that in a large majority of cases "two-of-a-kind" error rates are better. These are colored green. The notable exception to this is matcher D which successfully compares templates without regard to their origin. This also means that matcher D performs better if at least one of the templates originates from generator D.

A further result from the table is that in some cases (DA, AC and many others marked in pink) the result of using both templates from vendor X with matcher Y offers better accuracy than matcher Y on its own templates. An extreme case here is matcher I: It's native performance is beaten in 9 of 13 cases in which it is given like-pairs. Perhaps more notable is that matcher I becomes the second best (the rank of the mean is 2).

NF = 2	A	B	C	D	E	F	G	H	I	J	K	L	M	N	Median	Rank
A	0.0011	0.0037	0.0021	0.0022	0.0018	0.0021	0.0016	0.0068	0.0025	0.0222	0.3409	0.0067	0.0056	0.0033	0.0029	2
	0.0011	0.0092	0.0080	0.0021	0.0130	0.0079	0.0049	0.0248	0.0049	0.0755	0.0292	0.0161	0.0213	0.0210		
B	0.0014	0.0024	0.0037	0.0025	0.0023	0.0038	0.0017	0.0103	0.0012	0.0107	0.5179	0.0039	0.0028	0.0057	0.0033	5
	0.0027	0.0024	0.0072	0.0018	0.0073	0.0071	0.0017	0.0456	0.0049	0.0589	0.0588	0.0105	0.0083	0.0096		
C	0.0022	0.0023	0.0032	0.0025	0.0026	0.0032	0.0015	0.0202	0.0015	0.0110	0.4577	0.0045	0.0032	0.0046	0.0032	4
	0.0052	0.0057	0.0032	0.0025	0.0104	0.0031	0.0039	0.0851	0.0081	0.1571	0.1048	0.0177	0.0099	0.0133		
D	0.0011	0.0022	0.0021	0.0013	0.0018	0.0021	0.0019	0.0053	0.0015	0.0114	0.2807	0.0039	0.0031	0.0038	0.0021	1
	0.0021	0.0046	0.0045	0.0013	0.0097	0.0044	0.0035	0.0325	0.0062	0.0877	0.0442	0.0154	0.0126	0.0103		
E	0.0026	0.0068	0.0034	0.0052	0.0045	0.0035	0.0037	0.0107	0.0059	0.0259	0.0170	0.0120	0.0106	0.0057	0.0058	7
	0.0025	0.0061	0.0056	0.0028	0.0045	0.0054	0.0035	0.0280	0.0102	0.0900	0.0362	0.0163	0.0147	0.0053		
F	0.0021	0.0024	0.0031	0.0023	0.0027	0.0031	0.0014	0.0204	0.0015	0.0113	0.4572	0.0045	0.0032	0.0047	0.0031	3
	0.0054	0.0060	0.0032	0.0025	0.0103	0.0031	0.0038	0.0855	0.0081	0.1597	0.1058	0.0177	0.0097	0.0131		
G	0.0021	0.0027	0.0036	0.0029	0.0030	0.0036	0.0007	0.0084	0.0018	0.0168	0.0411	0.0046	0.0034	0.0052	0.0035	6
	0.0040	0.0032	0.0085	0.0022	0.0061	0.0085	0.0007	0.0308	0.0068	0.0715	0.0693	0.0116	0.0074	0.0084		
H	0.0195	0.0279	0.0316	0.0218	0.0370	0.0318	0.0407	0.0422	0.0282	1.0000	0.8470	0.0446	0.0441	0.0358	0.0364	14
	0.0084	0.0421	0.0393	0.0143	0.0767	0.0395	0.0384	0.0422	0.0210	0.9999	0.0724	0.0413	0.0753	0.0831		
I	0.0092	0.0100	0.0139	0.0129	0.0116	0.0139	0.0041	0.0702	0.0056	0.0384	0.4521	0.0142	0.0117	0.0199	0.0134	12
	0.0073	0.0184	0.0252	0.0100	0.0333	0.0249	0.0083	0.1137	0.0056	0.1206	0.1170	0.0207	0.0313	0.0415		
J	0.0160	0.0195	0.0205	0.0204	0.0166	0.0208	0.0116	0.0542	0.0103	0.0640	0.9921	0.0219	0.0203	0.0213	0.0205	13
	0.0077	0.0119	0.0259	0.0082	0.0276	0.0257	0.0070	0.1572	0.0103	0.0640	0.5736	0.0198	0.0296	0.0297		
K	0.0022	0.0078	0.0062	0.0074	0.0027	0.0062	0.0023	0.0265	0.0036	0.0282	0.0275	0.0264	0.0257	0.0049	0.0068	9
	0.0018	0.0130	0.0108	0.0051	0.0134	0.0109	0.0049	0.0280	0.0068	0.0929	0.0275	0.0270	0.0313	0.0234		
L	0.0079	0.0078	0.0105	0.0120	0.0114	0.0104	0.0035	0.0317	0.0059	0.0246	0.5553	0.0113	0.0093	0.0141	0.0109	11
	0.0115	0.0109	0.0218	0.0097	0.0258	0.0213	0.0066	0.0795	0.0105	0.0861	0.1123	0.0113	0.0267	0.0254		
M	0.0052	0.0045	0.0081	0.0056	0.0064	0.0082	0.0026	0.0223	0.0030	0.0158	0.6544	0.0085	0.0061	0.0110	0.0072	10
	0.0099	0.0096	0.0106	0.0049	0.0116	0.0108	0.0039	0.1007	0.0134	0.1573	0.1929	0.0247	0.0061	0.0211		
N	0.0042	0.0055	0.0055	0.0062	0.0062	0.0055	0.0044	0.0128	0.0055	0.0188	0.0168	0.0113	0.0094	0.0081	0.0062	8
	0.0063	0.0077	0.0086	0.0042	0.0094	0.0087	0.0056	0.0368	0.0104	0.0862	0.0353	0.0169	0.0157	0.0081		
Median	0.0024	0.0050	0.0046	0.0054	0.0037	0.0047	0.0024	0.0203	0.0033	0.0205	0.4546	0.0099	0.0077	0.0057		
Rank	1	7	5	8	4	6	2	12	3	13	14	11	10	9		

Table 13: Scenario 3 vs. Scenario 1.

The FNMR at a FMR of 0.01, for two-finger matching. The top value in each cell corresponds to scenario 3 for which the row vendor produced both enrollment *and* authentication templates, while the column vendor performed just the comparison. The bottom value in each cell corresponds to scenario 1: the row vendor produced the enrollment template while the column vendor produced the authentication template and performed the comparison. Cells are colored green when $F_{ij}^{upper} < F_{ij}^{lower}$, i.e. there is a benefit from using same-source templates. Cells are colored red when the performance is better than the native performance of the matcher. This applies to the MIN:A template throughout, and the POEBVA dataset.

A full set of tables is given in the accompanying document: MINEX Supplement E - Scenario 3.

5.7.6 Scenario 4

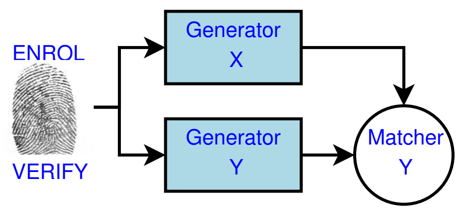


Figure 6: Scenario 4

In this case two generators X and Y are applied to the *same* input image, and the resulting templates are matched using product Z. This is depicted in Figure 6. Such operations, same-image comparisons, are irrelevant commercially and notoriously, useless yet not unknown, in biometric performance testing. However, because the verification process is mediated by a standard template, this scenario is useful in a test of a new biometric interchange standard because it allows the effect of fingerprint recapture to be separated from that of template preparation. That is, the typical biometric application suffers false rejects when the second sample is insufficiently similar to the original enrolled sample. But here, this is absent and he effect of different image processing and minutiae detection algorithms can be isolated. Differences here are at the core of interoperability failure.

The result of this operation is that when an images's template from vendor X's generator is matched with the corresponding template from vendor Y's the score is usually extremely high. When vendor X equals Y, matcher Z produces its maximum scores. But otherwise there are some template pairs that produce a low score. This is quantified for vendor A in Table 14 which quantifies such anomalous behvaiour as the fraction of images that would be falsely non-matched at a threshold that in normal operation would give a FMR of 0.01 and 0.001. Tables for all matchers are present in the accompanying document, MINEX Supplement F - Matching Same-image Templates, which only presents data for the POEBVA dataset. The computations used comparisons from 60000 right index finger images only.

The rejection rates in the tables are small, typically one tenth of the false non-match errors reported, for the same matcher (D), in the scenario 2 matrix of 10(a).

5.7.7 Interoperable Product Groups

The MINEX evaluation is one instance of the testing standard *Biometric Performance Testing and Reporting - Part 4: Performance and Interoperability testing of data interchange formats*[8] currently being developed in Working Group 5 of Subcommittee 37 of ISO's Joint Technical Committee 1. That standard introduces two definitions of interoperability:

basic interoperability ability of a vendor's generator to create biometric data blocks (BDBs) that can be processed by other vendor' comparison subsystems, and the ability of a vendor's comparison subsystem to process BDBs from other vendor's generators.

performance interoperability ability of biometric subsystems from different vendor to generate and compare samples and meet a specified level of performance.

NOTE basic interoperability is a necessary precondition for performance interoperability

The latter definition references a specified level of performance, and the standard gives a testing laboratory the latitude to select suitable figures of merit and thresholds to be applied. The standard advances the idea that if some subset of generators and matchers mutually operate at a low error rate then they can be deemed interoperable. For MINEX we adopt the FNMR at a fixed FMR as the figure of merit. If this FMR is set to a high enough value then all products are interoperable. As the FMR is reduced, the size of the interoperable group decreases. Ultimately, one and then zero products will meet the criterion. The general case for some given accuracy requirement is that some subset of the template generators, $\mathcal{T}$, will interoperate with a subset of the template matchers, $\mathcal{M}$.

Beyond specification of an acceptable performance threshold, it is also necessary to specify a criterion for *group performance*. The ISO standard suggests the maximum

$$\text{MAX}_{\mathcal{T}\mathcal{M}} = F_{mn}, \qquad \{m, n\} = \arg\max_{ij} F_{ij} \tag{2}$$

(a) D

D	A	B	C	D	E	F	G	H	I	J	K	L	M	N	Mean	Rank
A	0.000	0.002	0.002	0.001	0.004	0.002	0.003	0.004	0.003	0.004	0.005	0.004	0.004	0.003	0.003	3
B	0.002	0.000	0.001	0.000	0.003	0.001	0.000	0.007	0.006	0.004	0.006	0.003	0.002	0.003	0.003	2
C	0.002	0.001	0.000	0.001	0.004	0.000	0.003	0.014	0.008	0.014	0.007	0.006	0.002	0.002	0.005	7
D	0.001	0.000	0.001	0.000	0.003	0.001	0.001	0.006	0.004	0.005	0.005	0.002	0.002	0.002	0.002	1
E	0.004	0.003	0.004	0.003	0.000	0.004	0.003	0.013	0.009	0.008	0.009	0.005	0.004	0.001	0.005	8
F	0.002	0.001	0.000	0.001	0.004	0.000	0.002	0.014	0.008	0.014	0.007	0.006	0.002	0.002	0.005	6
G	0.003	0.000	0.003	0.001	0.003	0.002	0.000	0.010	0.008	0.006	0.008	0.006	0.002	0.003	0.004	4
H	0.004	0.007	0.014	0.006	0.013	0.014	0.010	0.000	0.012	0.008	0.012	0.006	0.019	0.011	0.010	14
I	0.003	0.006	0.008	0.004	0.009	0.008	0.008	0.012	0.000	0.015	0.010	0.008	0.017	0.006	0.008	11
J	0.004	0.004	0.014	0.005	0.008	0.014	0.006	0.008	0.015	0.000	0.011	0.007	0.019	0.008	0.009	13
K	0.005	0.006	0.007	0.005	0.009	0.007	0.008	0.012	0.010	0.011	0.004	0.013	0.013	0.009	0.009	12
L	0.004	0.003	0.006	0.002	0.005	0.006	0.006	0.006	0.008	0.007	0.013	0.000	0.011	0.003	0.006	9
M	0.004	0.002	0.002	0.002	0.004	0.002	0.002	0.019	0.017	0.019	0.013	0.011	0.000	0.004	0.007	10
N	0.003	0.003	0.002	0.002	0.001	0.002	0.003	0.011	0.006	0.008	0.009	0.003	0.004	0.000	0.004	5
Mean	0.003	0.003	0.005	0.002	0.005	0.005	0.004	0.010	0.008	0.009	0.009	0.006	0.007	0.004		
Rank	3	2	7	1	8	6	4	14	11	13	12	9	10	5		

(b) D

D	A	B	C	D	E	F	G	H	I	J	K	L	M	N	Mean	Rank
A	0.000	0.002	0.004	0.002	0.006	0.004	0.005	0.008	0.006	0.006	0.005	0.006	0.007	0.005	0.005	3
B	0.002	0.000	0.002	0.001	0.005	0.002	0.001	0.014	0.009	0.007	0.008	0.005	0.004	0.004	0.005	2
C	0.004	0.002	0.000	0.001	0.006	0.000	0.004	0.027	0.013	0.023	0.010	0.009	0.003	0.004	0.008	7
D	0.002	0.001	0.001	0.000	0.005	0.001	0.001	0.011	0.006	0.008	0.006	0.003	0.003	0.003	0.004	1
E	0.006	0.005	0.006	0.005	0.001	0.006	0.004	0.022	0.015	0.013	0.014	0.007	0.005	0.001	0.008	8
F	0.004	0.002	0.000	0.001	0.006	0.000	0.004	0.027	0.013	0.023	0.011	0.009	0.003	0.004	0.007	6
G	0.005	0.001	0.004	0.001	0.004	0.004	0.000	0.018	0.014	0.009	0.012	0.009	0.004	0.004	0.006	4
H	0.008	0.014	0.027	0.011	0.022	0.027	0.018	0.000	0.021	0.013	0.019	0.011	0.035	0.021	0.018	14
I	0.006	0.009	0.013	0.006	0.015	0.013	0.014	0.021	0.000	0.023	0.014	0.012	0.026	0.010	0.013	12
J	0.006	0.007	0.023	0.008	0.013	0.023	0.009	0.013	0.023	0.000	0.015	0.011	0.028	0.013	0.014	13
K	0.005	0.008	0.010	0.006	0.014	0.011	0.012	0.019	0.014	0.015	0.004	0.018	0.019	0.012	0.012	11
L	0.006	0.005	0.009	0.003	0.007	0.009	0.009	0.011	0.012	0.011	0.018	0.000	0.016	0.004	0.009	9
M	0.007	0.004	0.003	0.003	0.005	0.003	0.004	0.035	0.026	0.028	0.019	0.016	0.000	0.006	0.011	10
N	0.005	0.004	0.004	0.003	0.001	0.004	0.004	0.021	0.010	0.013	0.012	0.004	0.006	0.000	0.006	5
Mean	0.005	0.005	0.008	0.004	0.008	0.007	0.006	0.018	0.013	0.014	0.012	0.009	0.011	0.006		
Rank	3	2	7	1	8	6	4	14	12	13	11	9	10	5		

Table 14: Scenario 4
The fraction of same-image templates produced by companies X and Y that fail to exceed the threshold that produced a FMR of 0.001 on matcher D.

and the mean

$$\text{MEAN}_{\mathcal{TM}} = \frac{\sum_{i \in \mathcal{T}} \sum_{j \in \mathcal{M}} F_{ij}}{|\{\mathcal{T}\}| \quad |\{\mathcal{M}\}|} \tag{3}$$

error rates as possible grouping criteria, but advises that adoption of the *max*, *mean* or some other criterion should be application dependent. The use of the mean has the negative consequence that when those products are deployed the actual mean performance will be worse than specification if the least capable products in the group are deployed in locations where traffic is high (i.e. many authentication transactions). This might occur if the less capable performers are the least expensive.

The search for interoperable groups involves computing group error accuracy for all possible submatrices of the full N x N matrix and comparing it with the requirement. There are K such candidates:

$$K = \sum_{i=2}^{N} \sum_{j=2}^{N} C(N,i)C(N,j) \tag{4}$$

where the number of ways of choosing i elements from N is $C(N,i) = N!/i!(N-i)!$. For $N = 14$ an exhaustive search necessitates in excess of 268 million extractions. These are performed on just $N^2 = 196$ values and the task is entirely CPU bound and can be completed in a few hours. More efficient searches are possible, of course. For example for the max criterion, any subgroup of an interoperable subgroup is itself interoperable.

Table 15 summarizes the results of applying the two approaches to the two-finger scenario 1 interoperability matrix of Table 9. This was generated by naïve search of the matrix for core interoperable groups whose mean or maximum FNMR is less than or equal to 0.01 at the usual FMR of 0.01. The outcomes are substantially different for the two methods. Most importantly the *max*-based groups are substantially smaller and fewer. This is because the maximum criterion is intolerant of product pairs that give large error rates, while the mean criterion subsidizes poor product pairs with the low error rates of others.

For the easier datasets, such as POEBVA, there are many interoperable subgroups generated against the mean criterion, and it is often the case that for any specific sizes of the sets $\mathcal{T}$ and $\mathcal{M}$ multiple disjoint subgroups are found. Each vendor in general is present in only a fraction of those subgroups. For example if generators {U,V,X} interoperate with matchers {P,Q}, as do {U,V,Y}, but not {U,V,X,Y} then X and Y are only in half of the groups.

For MINEX , only the largest groups are reported, i.e. those with largest value of $|\mathcal{T}|.|\mathcal{M}|$. In cases where the search found several groups of that maximum size the one with the lowest error rate was reported. Subgroups of size 1 were not considered, although such a group may well be appropriate, for example in the case of a centralized matching service.

This kind of procedure for determining interoperable groups requires specification of

1. a figure of merit (e.g. FNMR at a fixed FMR);
2. an operating point (e.g. FMR = 0.005);
3. the acceptable value for the figure of merit (e.g. FNMR ≤ 0.01);
4. the grouping criterion (e.g. maximum);
5. a minimum number of generators that will be considered (e.g. 2)
6. a minimum number of matchers that will be considered (e.g. 3)
7. a policy for handling disjoint groups of the same size (e.g. minimum error criterion)
8. a policy for dealing with the case in which no products are found to be interoperable (e.g. relax the FMR).

The ILO [9] elected to certify products for its seafarers identity credential that could interoperate with a mean FNMR less than 0.01 at a fixed FMR of 0.01. From seven products tested three were certified in early 2005.

Neither of the criteria reported here are specifically recommended or deprecated. The median is unreasonable as it is entirely insensitive to poor performers, the mean is somewhat sensitive and the maximum is from the poorest performing

Dataset	Criterion	Value	No.		Template Generators	Template Matchers
POEBVA	group max FNMR $\leq$ 0.01 at FMR = 0.01	max = 0.0092	8	6	ABCDEFG------N	ABCD-FG-------
POEBVA	group mean FNMR $\leq$ 0.01 at FMR = 0.01	mean = 0.0100	12	11	ABCDEFG--JKLMN	ABCDEFG-I--LMN
DHS2	group max FNMR $\leq$ 0.01 at FMR = 0.01	max = 0.0088	3	3	-B-D---------N	AB-D----------
DHS2	group mean FNMR $\leq$ 0.01 at FMR = 0.01	mean = 0.0099	5	5	-B-D-F----K--N	ABCD-F--------
POE	group max FNMR $\leq$ 0.01 at FMR = 0.01	max = 0.0095	7	7	-BCDEFG------N	ABCDEFG-------
POE	group mean FNMR $\leq$ 0.01 at FMR = 0.01	mean = 0.0093	11	11	ABCDEFG---KLMN	ABCDEFG-I--LMN
DOS	group max FNMR $\leq$ 0.01 at FMR = 0.01	max = 0.0093	7	2	-BCDEFG------N	---D--G-------
DOS	group mean FNMR $\leq$ 0.01 at FMR = 0.01	mean = 0.0098	10	4	ABCDEFG---K-MN	A--DE-G-------
POEBVA	group max FNMR $\leq$ 0.01 at FMR = 0.003	max = 0.0084	9	3	ABCDEFG---K--N	A--D--G-------
POEBVA	group mean FNMR $\leq$ 0.01 at FMR = 0.003	mean = 0.0099	12	7	ABCDEFG--JKLMN	ABCD-FG-I-----
DHS2	group max FNMR $\leq$ 0.01 at FMR = 0.003	max = 0.0096	2	2	-B-D----------	A--D----------
DHS2	group mean FNMR $\leq$ 0.01 at FMR = 0.003	mean = 0.0093	5	2	AB-D------K--N	A--D----------
POE	group max FNMR $\leq$ 0.01 at FMR = 0.003	max = 0.0093	9	3	ABCDEFG----L-N	A--D--G-------
POE	group mean FNMR $\leq$ 0.01 at FMR = 0.003	mean = 0.0098	10	9	ABCDEFG---K-MN	ABCDEFG-I----N
DOS	group max FNMR $\leq$ 0.01 at FMR = 0.003	max = 0.0081	2	2	-B-D----------	---D--G-------
DOS	group mean FNMR $\leq$ 0.01 at FMR = 0.003	mean = 0.0097	9	2	ABCDEFG---K--N	---D--G-------
POEBVA	group max FNMR $\leq$ 0.01 at FMR = 0.001	max = 0.0099	10	2	ABCDEFG---K-MN	---D--G-------
POEBVA	group mean FNMR $\leq$ 0.01 at FMR = 0.001	mean = 0.0099	9	6	ABCDEFG---K--N	A-CD-FG-I-----
DHS2	group max FNMR $\leq$ 0.01 at FMR = 0.001		0	0	--------------	--------------
DHS2	group mean FNMR $\leq$ 0.01 at FMR = 0.001		0	0	--------------	--------------
POE	group max FNMR $\leq$ 0.01 at FMR = 0.001	max = 0.0090	9	2	ABCDEFG-----MN	---D--G-------
POE	group mean FNMR $\leq$ 0.01 at FMR = 0.001	mean = 0.0097	9	6	ABCDEFG-----MN	ABCD-FG-------
DOS	group max FNMR $\leq$ 0.01 at FMR = 0.001		0	0	--------------	--------------
DOS	group mean FNMR $\leq$ 0.01 at FMR = 0.001	mean = 0.0095	3	2	-B-D--G-------	---D--G-------
POEBVA	group max FMR $\leq$ 0.01 at FNMR = 0.01	max = 0.0082	8	6	ABCDEFG------N	ABCD-FG-------
POEBVA	group mean FMR $\leq$ 0.01 at FNMR = 0.01	mean = 0.0070	10	9	ABCDEFG---K-MN	ABCDEFG-I----N
DHS2	group max FMR $\leq$ 0.01 at FNMR = 0.01	max = 0.0079	3	4	-B-D---------N	AB-D--G-------
DHS2	group mean FMR $\leq$ 0.01 at FNMR = 0.01	mean = 0.0100	4	6	-B-D------K--N	ABCD-FG-------
POE	group max FMR $\leq$ 0.01 at FNMR = 0.01	max = 0.0087	7	7	-BCDEFG------N	ABCDEFG-------
POE	group mean FMR $\leq$ 0.01 at FNMR = 0.01	mean = 0.0086	11	9	ABCDEFG---KLMN	ABCDEFG-I----N
DOS	group max FMR $\leq$ 0.01 at FNMR = 0.01	max = 0.0068	8	2	ABCDEFG------N	---D--G-------
DOS	group mean FMR $\leq$ 0.01 at FNMR = 0.01	mean = 0.0090	7	4	ABCDEF-------N	A--DE-G-------

Table 15: Interoperable Subgroups for Two-finger Matching.
The last two columns give interoperable generators and matchers for the dataset and grouping criterion listed in the second and third columns. These are the largest subgroups. Where several groups of the same number of vendors exist, the one group with minimum error is reported.

pair. This issue remains under discussion in SC37/WG5 as the ISO standard [8] is developed. The reader is cautioned that the subgroups enumerated in the table should not be considered as definitively interoperable groups. A number of factors should be included in such a determination. Indeed other criteria may be appropriate too, particularly those that consider the fixed-threshold effects discussed in section 5.7.2.

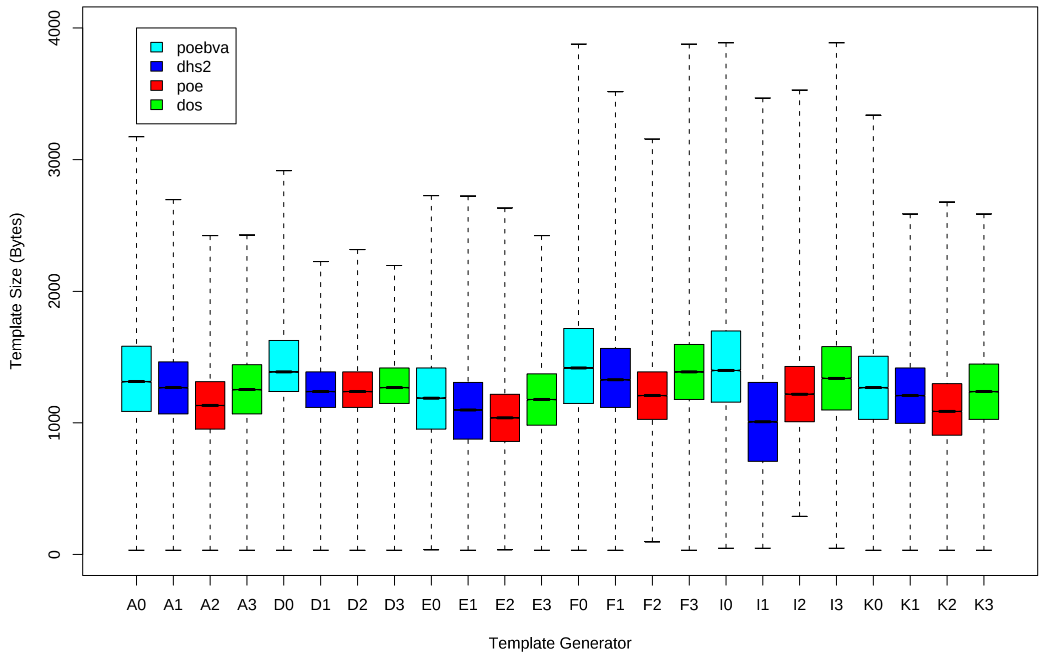
5.8 Template Sizes

The size of biometric templates is important operationally for storage and throughput reasons. The MINEX specification limited the number of minutiae that generators were allowed to include in the template records to 128. INCITS 378 itself imposes a cap of 255 minutiae per finger image. For a single image the size of an N-minutiae record is $32 + 6N$ bytes. Thus for MINEX the MIN:A templates cannot exceed 800 bytes. Median values for each template are aggregated over all vendors and data sets in Table 16. The sizes are broken out by vendor and dataset in the boxplots of Figures 7 and 8.

1. The MIN:A templates are around four times smaller than the MIN:B templates.
2. Aside from vendor J the interquartile ranges overlap considerably.
3. The vendors of the most accurately matched templates, B, N and D never produce templates with the maximum number of allowed minutiae.

Template	Number of Images	Median Left Index	Median Right Index
MIN:A	13793231	278	272
MIN:B	5904178	1237	1207

Table 16: Median Template Sizes
The sizes are in bytes and include the standard headers.



(a) MIN:B

Figure 7: Variation in MIN:B Template Sizes.
The center of the box gives the median size, the box itself gives the interquartile range, and the whiskers give the minimum and maximum values.

4. Vendors D, E, H, L and M produced fixed-size proprietary templates.
5. The sizes of proprietary templates are not reported here because: the MINEX API placed only a loose upper limit on this size; some vendors compressed their templates, while some padded to the maximum allowed size; and vendors may well be able to tailor template sizes to specific requirements.

5.9 Processing Times

The processing time is often an important performance parameter. Template generation times are dependent largely on the size of the input area. The template matching times are much lower and depend in a complicated way on the numbers of minutiae in the enrollment and authentication templates and on the relationship between the two.

Tables 17 and 18 summarize mean generation and matching times for each of the four datasets and the three templates. These tests were conducted 3 gigahertz i386 machines running the operating systems are listed in 5.11. The matching times should be viewed only in the context of the MINEX trial which did not specifically intend to evaluate throughput.

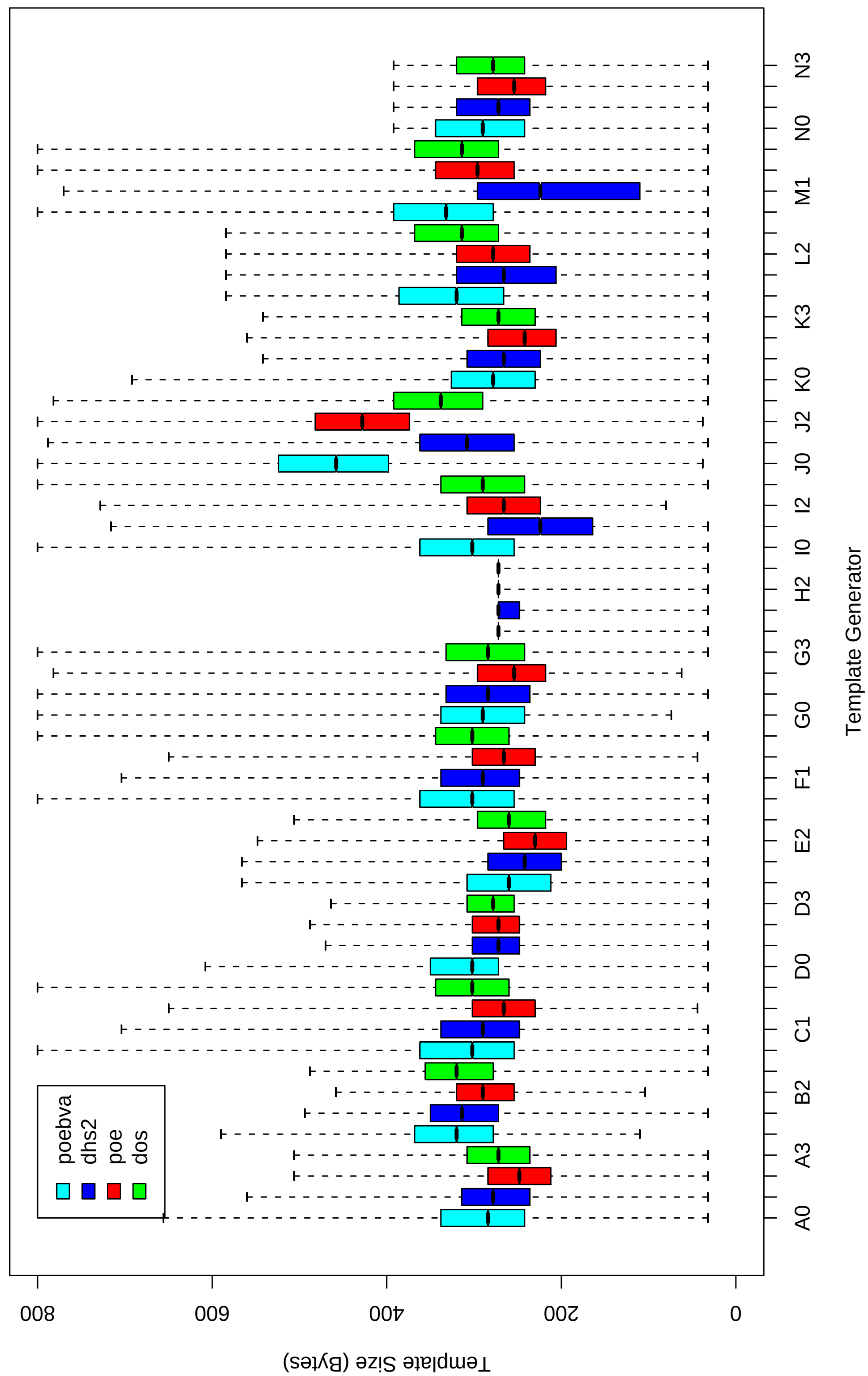


Figure 8: Variation in MIN:A Template Sizes.

Boxplots of template sizes of the MIN:A templates. The center of the box gives the median size. The whiskers at the ends give the minimum and maximum sizes. These values were capped in MINEX to 32 and 800 bytes respectively. The templates from vendor H usually have a fixed size.

Dataset / Template			A	B	C	D	E	F	G	H	I	J	K	L	M	N
poebva	A	Time	507	311	452	108	169	340	133	182	480	1350	694	138	167	195
		Rank	12	8	10	1	5	9	2	6	11	14	13	3	4	7
poebva	B	Time	507			380	169	397			480		695			
		Rank	5			2	1	3			4		6			
poebva	P	Time	737	310	451	129	169	373	442	182	480	1386	699	140	167	197
		Rank	13	7	10	1	4	8	9	5	11	14	12	2	3	6
dhs2	A	Time	596	359	481	110	155	365	135	170	478	672	586	138	157	192
		Rank	13	8	11	1	4	9	2	6	10	14	12	3	5	7
dhs2	B	Time	596			380	155	424			479		584			
		Rank	6			2	1	3			4		5			
dhs2	P	Time	849	360	480	130	155	399	537	169	479	704	584	140	156	195
		Rank	14	7	10	1	3	8	11	5	9	13	12	2	4	6
poe	A	Time	517	328	454	109	174	342	133	187	480	1401	754	139	169	195
		Rank	12	8	10	1	5	9	2	6	11	14	13	3	4	7
poe	B	Time	517			378	173	401			481		753			
		Rank	5			2	1	3			4		6			
poe	P	Time	753	328	453	130	173	376	449	187	481	1438	756	141	168	198
		Rank	12	7	10	1	4	8	9	5	11	14	13	2	3	6
dos	A	Time	544	348	476	109	197	360	136	206	481	763	620	144	174	199
		Rank	12	8	10	1	5	9	2	7	11	14	13	3	4	6
dos	B	Time	543			381	197	420			482		620			
		Rank	5			2	1	3			4		6			
dos	P	Time	782	346	474	131	197	394	510	205	482	800	617	146	174	202
		Rank	13	7	9	1	4	8	11	6	10	14	12	2	3	5

Table 17: Template Generation Times
The mean times, in milliseconds, for generation of the three kinds of templates from each of the four databases.

Dataset		A	B	C	D	E	F	G	H	I	J	K	L	M	N
poebva	Time	38.8	2.0	10.1	3.0	0.7	10.1	6.0	21.4	8.4	73.9	8.4	4.5	6.2	0.9
	Rank	13	3	10	4	1	11	6	12	9	14	8	5	7	2
dhs2	Time	40.3	2.1	9.0	2.9	0.7	9.8	6.2	21.1	6.5	33.1	8.4	3.4	2.4	0.9
	Rank	14	3	10	5	1	11	7	12	8	13	9	6	4	2
poe	Time	38.0	1.7	7.5	2.7	0.6	9.3	5.0	24.4	7.8	66.1	7.6	3.8	4.9	0.8
	Rank	13	3	8	4	1	11	7	12	10	14	9	5	6	2
dos	Time	43.6	2.0	9.7	2.7	0.7	10.5	6.4	23.5	8.7	38.7	9.4	4.4	5.6	0.9
	Rank	14	3	10	4	1	11	7	12	8	13	9	5	6	2

Table 18: Template Matching Times
Mean times, in milliseconds, for matching of proprietary templates from each of the four databases.

5.10 Does Poor Quality Degrade Interoperablity

The question of whether low quality images yield less interoperable templates is open. One means of assessing this is to apply an image quality assessment algorithm and observe the effect on verification accuracy. Table 19 shows an interoperability matrix that results from exclusion of transactions in which enrollment templates were derived from images whose NFIQ [17] values were poor, specifically $q \in \{4, 5\}$. This operation corresponds to the rejection of low quality enrollment prints and is representative of operational reality only if subsequent images can be captured with quality $q \in \{1, 2, 3\}$. The table shows that error rates decline across the board. This is simply an indication that NFIQ is rejecting images likely to perform poorly, just as it was designed to do. The question is really whether whether some measure of variance diminishes. One way of addressing this is to repeat the section 5.7 search for interoperable sets of vendors. The result shown in Table 20 qualifies groups as interoperable if their mutual *range* of FNMR values is within 0.01. The sets of generators actually contracts as quality of the images is raised. This result warrants further investigation, not least because the two template matchers, C and F, offer nearly identical error rates in MIN:A tests.

5.11 Implementation Sizes

The MINEX evaluation was conducted by making many calls to functions provided by vendors in the form of a compiled library. The implementations conformed to the specifications given in the MINEX API . The sizes of the submitted libraries are listed in Table 21. The reader is cautioned that these sizes are likely to differ substantially from those used in actual deployments. The reader should note also that the vendor's choice of the host operating system should not be held to imply availability or lack of availability of commercial products for any particular OS. The vendor should be contacted for further information.

5.12 Algorithm Fusion

Vendors' implementations embed considerable intellectual property and use a variety of algorithms. As such different matchers will succeed on many of the same image comparisons, but they will not always fail on the same samples. It is thus the case that multi-algorithmic fusion will enhance accuracy. It is implemented here as a demonstration of its efficacy and of its limitations. Systems are fused in pairs, at the score-level. This was done for left and right-finger pairs by simple summation of the raw matcher scores. Here it is neccessary to normalize the scores because different matchers have their own native scores.

If a first matcher compares the i-th pair of samples to produce a score $s_i^{(1)}$ and a second matcher likewise outputs a score $s_i^{(2)}$ for the same pair, then fusion proceeds as follows

$$s_i = 1 - (1 - N^{(1)}(s_i^{(1)}))(1 - N^{(2)}(s_i^{(2)})) \tag{5}$$

where $N^{(k)}(s)$ is the cumulative distribution function of the impostor scores of the k-th matcher, and $1 - N^{(k)}$ is then the expected fraction of impostor scores that are falsely accepted. This equation thereby renders the resulting fused score, s_i, an estimate of the likelihood that the sample is accepted by one or the other system. This is subject to the usual DET analysis and the results, for single-finger matches of proprietary and natively generated MIN:A templates, are presented in Table 22.

The results show that vendor G performance is rarely improved by any combination of it with another vendor. Further fusion of algorithms X and Y seldom produces better performance than G alone. Fusion offers larger performance improvements for the MIN:A template than for the proprietary. Pairwise matcher fusion offers much small benefits than the two-finger fusion described in 5.4.

(a) Enrollment templates have NFIQ $\in \{1, 2, 3\}$

NF = 1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	Mean	Rank	Med.	Rank
A	0.0074	0.0438	0.0337	0.0144	0.0484	0.0337	0.0299	0.0662	0.0215	0.1527	0.0585	0.0491	0.0602	0.0819	0.0501	6	0.0461	10
B	0.0138	0.0187	0.0269	0.0109	0.0278	0.0264	0.0116	0.0924	0.0238	0.1348	0.1456	0.0375	0.0312	0.0440	0.0461	2	0.0273	2
C	0.0229	0.0334	0.0135	0.0124	0.0372	0.0137	0.0247	0.1719	0.0364	0.2853	0.2288	0.0582	0.0357	0.0556	0.0735	8	0.0361	7
D	0.0131	0.0268	0.0203	0.0085	0.0349	0.0204	0.0212	0.0785	0.0277	0.1834	0.1063	0.0498	0.0397	0.0464	0.0484	3	0.0313	4
E	0.0132	0.0255	0.0222	0.0121	0.0178	0.0222	0.0162	0.0660	0.0308	0.1677	0.0701	0.0416	0.0356	0.0286	0.0407	1	0.0270	1
F	0.0232	0.0335	0.0134	0.0128	0.0374	0.0135	0.0244	0.1716	0.0366	0.2861	0.2291	0.0583	0.0356	0.0550	0.0736	9	0.0361	8
G	0.0195	0.0226	0.0315	0.0132	0.0268	0.0310	0.0086	0.0713	0.0332	0.1589	0.1482	0.0417	0.0300	0.0440	0.0486	5	0.0312	3
H	0.0273	0.1132	0.0972	0.0466	0.1608	0.0978	0.1079	0.1029	0.0587	1.0000	0.1474	0.0942	0.1551	0.2060	0.1725	14	0.1054	14
I	0.0247	0.0656	0.0625	0.0365	0.0857	0.0625	0.0406	0.1964	0.0232	0.2284	0.2548	0.0605	0.0825	0.1198	0.0960	12	0.0640	12
J	0.0267	0.0462	0.0752	0.0316	0.0781	0.0757	0.0340	0.2637	0.0395	0.1343	0.7207	0.0593	0.0814	0.0987	0.1261	13	0.0754	13
K	0.0099	0.0471	0.0345	0.0174	0.0505	0.0336	0.0304	0.0603	0.0261	0.1717	0.0309	0.0583	0.0683	0.0874	0.0519	7	0.0408	9
L	0.0303	0.0419	0.0525	0.0286	0.0617	0.0527	0.0292	0.1395	0.0339	0.1707	0.2138	0.0384	0.0660	0.0768	0.0740	10	0.0526	11
M	0.0311	0.0353	0.0283	0.0178	0.0358	0.0282	0.0199	0.1771	0.0445	0.2801	0.3724	0.0656	0.0229	0.0656	0.0875	11	0.0355	6
N	0.0237	0.0315	0.0292	0.0183	0.0308	0.0292	0.0264	0.0803	0.0350	0.1752	0.0672	0.0511	0.0434	0.0362	0.0484	4	0.0333	5
Med	0.0230	0.0344	0.0303	0.0159	0.0373	0.0301	0.0255	0.0977	0.0335	0.1735	0.1478	0.0546	0.0416	0.0606				
Rank	2	7	5	1	8	4	3	12	6	14	13	10	9	11				

(b) Enrollment templates are unrestricted

NF = 1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	Mean	Rank	Med.	Rank
A	0.0136	0.0549	0.0458	0.0225	0.0641	0.0459	0.0417	0.0834	0.0334	0.1707	0.0747	0.0659	0.0792	0.0966	0.0637	6	0.0595	10
B	0.0218	0.0251	0.0385	0.0173	0.0402	0.0382	0.0192	0.1136	0.0336	0.1501	0.1599	0.0506	0.0442	0.0561	0.0578	2	0.0394	2
C	0.0357	0.0428	0.0225	0.0204	0.0519	0.0225	0.0348	0.1969	0.0484	0.3034	0.2451	0.0743	0.0493	0.0691	0.0869	8	0.0489	6
D	0.0207	0.0357	0.0301	0.0140	0.0485	0.0303	0.0316	0.0945	0.0392	0.2013	0.1218	0.0655	0.0551	0.0582	0.0605	4	0.0438	3
E	0.0236	0.0365	0.0340	0.0225	0.0301	0.0341	0.0286	0.0874	0.0476	0.1885	0.0896	0.0600	0.0557	0.0397	0.0556	1	0.0381	1
F	0.0359	0.0430	0.0222	0.0206	0.0522	0.0224	0.0345	0.1967	0.0485	0.3038	0.2456	0.0743	0.0493	0.0686	0.0870	9	0.0489	7
G	0.0300	0.0291	0.0447	0.0205	0.0390	0.0441	0.0129	0.0905	0.0441	0.1747	0.1632	0.0559	0.0419	0.0526	0.0602	3	0.0441	4
H	0.0437	0.1336	0.1212	0.0656	0.1860	0.1215	0.1339	0.1027	0.0796	1.0000	0.1748	0.1181	0.1818	0.2296	0.1923	14	0.1276	14
I	0.0397	0.0806	0.0830	0.0518	0.1062	0.0828	0.0548	0.2227	0.0348	0.2470	0.2756	0.0791	0.1030	0.1383	0.1142	12	0.0829	12
J	0.0403	0.0602	0.0939	0.0455	0.0987	0.0943	0.0489	0.2852	0.0542	0.1505	0.7314	0.0773	0.1026	0.1169	0.1428	13	0.0941	13
K	0.0188	0.0593	0.0476	0.0280	0.0661	0.0467	0.0428	0.0790	0.0400	0.1920	0.0461	0.0770	0.0885	0.1015	0.0667	7	0.0534	9
L	0.0467	0.0558	0.0704	0.0428	0.0822	0.0708	0.0432	0.1640	0.0485	0.1901	0.2375	0.0524	0.0866	0.0938	0.0918	10	0.0706	11
M	0.0496	0.0493	0.0455	0.0307	0.0545	0.0454	0.0327	0.2066	0.0616	0.3022	0.3929	0.0868	0.0359	0.0855	0.1057	11	0.0520	8
N	0.0368	0.0436	0.0428	0.0293	0.0458	0.0428	0.0393	0.1019	0.0497	0.1945	0.0865	0.0682	0.0621	0.0486	0.0637	5	0.0472	5
Mean	0.0326	0.0535	0.0530	0.0308	0.0690	0.0530	0.0428	0.1447	0.0474	0.2692	0.2175	0.0718	0.0740	0.0896				
Rank	2	7	6	1	8	5	3	12	4	14	13	9	10	11				
Med.	0.0358	0.0465	0.0451	0.0253	0.0533	0.0447	0.0370	0.1082	0.0480	0.1932	0.1690	0.0712	0.0589	0.0773				
Rank	2	6	5	1	8	4	3	12	7	14	13	10	9	11				

Table 19: The Effect of Excluding Low Quality Images.
The FNMR at FMR of 0.01, for single-finger matching. The top table is the result of exluding the worst quality enrollment images (i.e. NFIQ $\in \{4, 5\}$) from the error rate computation. The bottom table, as a replica of scenario 1 table 7(b), includes all such images. The vendor identified in the row produced the enrollment template, and the vendor identified in each column produced the authentication template *and* performed the comparison. This applies to the MIN:A template throughout.

Dataset	NFIQ Range	Criterion	Value	No.		Template Generators	Template Matchers
POEBVA	[1-3]	group range FNMR $\leq$ 0.01 at FMR = 0.01	median = 0.0287	4	3	-B----G-----MN	--C-EF--------
POEBVA	[1-4]	group range FNMR $\leq$ 0.01 at FMR = 0.01	median = 0.0394	6	2	AB----G---K-MN	--C--F--------
POEBVA	[1-5]	group range FNMR $\leq$ 0.01 at FMR = 0.01	median = 0.0451	6	2	AB----G---K-MN	--C--F--------
POE	[1-3]	group range FNMR $\leq$ 0.01 at FMR = 0.01	median = 0.0319	6	2	AB--E-G---K-M-	--C--F--------
POE	[1-4]	group range FNMR $\leq$ 0.01 at FMR = 0.01	median = 0.0361	6	2	AB--E-G---K-M-	--C--F--------
POE	[1-5]	group range FNMR $\leq$ 0.01 at FMR = 0.01	median = 0.0443	6	2	AB--E-G---K-M-	--C--F--------

Table 20: Effect of Quality Restriction
The table gives interoperable generators and matchers when enrollment templates obtained from images of levels 4 and 5 are excluded.

Vendor	Operating System	Size of library (bytes)
A	Windows	2,849,342
B	Windows	5,198,308
C	Windows	122,880
D	Windows	7,024,118
E	Windows	176,128
F	Windows	335,872
G	Windows	3,420,032
H	Windows	75,204
I	Windows	402,664
J	Windows	272,752
K	Linux	822,556
L	Linux	221,142
M	Linux	386,888
N	Linux	635,724

Table 21: The sizes in bytes of the libraries delivered to NIST. These sizes represent compiled code able to generate and match both proprietary and MIN:A templates. These sizes include for vendors A, D, E, F, I and K support for the MIN:B template also.

(a) Proprietary

	Pairwise Matcher Fusion															
NF = 1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	Mean	Rank
A	0.009	0.007	0.008	0.006	0.007	0.007	0.005	0.009	0.008	0.009	0.007	0.008	0.008	0.008	0.008	2
B	0.007	0.019	0.013	0.007	0.011	0.014	0.006	0.022	0.014	0.027	0.010	0.016	0.013	0.016	0.014	5
C	0.008	0.013	0.022	0.008	0.013	0.018	0.006	0.028	0.016	0.027	0.012	0.018	0.016	0.019	0.016	7
D	0.006	0.007	0.008	0.009	0.007	0.008	0.005	0.011	0.008	0.011	0.007	0.009	0.008	0.009	0.008	3
E	0.007	0.011	0.013	0.007	0.025	0.013	0.006	0.025	0.016	0.024	0.013	0.018	0.017	0.020	0.015	6
F	0.007	0.014	0.018	0.008	0.013	0.034	0.007	0.025	0.014	0.027	0.013	0.017	0.014	0.021	0.017	8
G	0.005	0.006	0.006	0.005	0.006	0.007	0.005	0.009	0.007	0.009	0.005	0.007	0.007	0.007	0.006	1
H	0.009	0.022	0.028	0.011	0.025	0.025	0.009	0.100	0.031	0.073	0.022	0.043	0.032	0.040	0.034	13
I	0.008	0.014	0.016	0.008	0.016	0.014	0.007	0.031	0.033	0.032	0.015	0.025	0.021	0.020	0.019	9
J	0.009	0.027	0.027	0.011	0.024	0.027	0.009	0.073	0.032	0.150	0.023	0.046	0.033	0.039	0.038	14
K	0.007	0.010	0.012	0.007	0.013	0.013	0.005	0.022	0.015	0.023	0.019	0.017	0.016	0.015	0.014	4
L	0.008	0.016	0.018	0.009	0.018	0.017	0.007	0.043	0.025	0.046	0.017	0.058	0.025	0.025	0.024	12
M	0.008	0.013	0.016	0.008	0.017	0.014	0.007	0.032	0.021	0.033	0.016	0.025	0.036	0.020	0.019	10
N	0.008	0.016	0.019	0.009	0.020	0.021	0.007	0.040	0.020	0.039	0.015	0.025	0.020	0.048	0.022	11
Mean	0.008	0.014	0.016	0.008	0.015	0.017	0.006	0.034	0.019	0.038	0.014	0.024	0.019	0.022		
Rank	2	5	7	3	6	8	1	13	9	14	4	12	10	11		

(b) MIN:A

	Pairwise Matcher Fusion															
NF = 1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	Mean	Rank
A	0.014	0.010	0.010	0.008	0.011	0.010	0.009	0.015	0.011	0.015	0.013	0.012	0.012	0.012	0.011	2
B	0.010	0.025	0.014	0.010	0.013	0.014	0.012	0.027	0.016	0.031	0.019	0.017	0.015	0.019	0.017	4
C	0.010	0.014	0.022	0.010	0.015	0.022	0.011	0.029	0.017	0.026	0.019	0.018	0.016	0.020	0.018	5
D	0.008	0.010	0.010	0.014	0.010	0.010	0.009	0.015	0.011	0.015	0.014	0.012	0.011	0.011	0.011	1
E	0.011	0.013	0.015	0.010	0.030	0.014	0.013	0.031	0.019	0.028	0.023	0.021	0.019	0.023	0.019	7
F	0.010	0.014	0.022	0.010	0.014	0.022	0.011	0.029	0.016	0.027	0.019	0.018	0.016	0.020	0.018	6
G	0.009	0.012	0.011	0.009	0.013	0.011	0.013	0.023	0.014	0.021	0.014	0.015	0.014	0.015	0.014	3
H	0.015	0.027	0.029	0.015	0.031	0.029	0.023	0.103	0.035	0.077	0.047	0.045	0.034	0.041	0.039	13
I	0.011	0.016	0.017	0.011	0.019	0.016	0.014	0.035	0.035	0.034	0.026	0.026	0.021	0.022	0.022	9
J	0.015	0.031	0.026	0.015	0.028	0.027	0.021	0.077	0.034	0.150	0.046	0.044	0.033	0.041	0.042	14
K	0.013	0.019	0.019	0.014	0.023	0.019	0.014	0.047	0.026	0.046	0.046	0.031	0.026	0.027	0.026	12
L	0.012	0.017	0.018	0.012	0.021	0.018	0.015	0.045	0.026	0.044	0.031	0.052	0.025	0.027	0.026	11
M	0.012	0.015	0.016	0.011	0.019	0.016	0.014	0.034	0.021	0.033	0.026	0.025	0.036	0.022	0.021	8
N	0.012	0.019	0.020	0.011	0.023	0.020	0.015	0.041	0.022	0.041	0.027	0.027	0.022	0.049	0.025	10
Mean	0.011	0.017	0.018	0.011	0.019	0.018	0.014	0.039	0.022	0.042	0.026	0.026	0.021	0.025		
Rank	2	4	5	1	7	6	3	13	9	14	12	11	8	10		

Table 22: Pairwise Algorithm Fusion

The FNMR at a FMR of 0.01 after score-level fusion of matching algorithms identified in the rows and columns. The computation is for single-finger matching. The matrices are symmetric. The cells are colored green when the error rate F_{ij} is lower than *both* native performances, i.e. F_{ii} and F_{jj}.

6 References

References

[1] Working Group 1. Standing Document 2 Harmonized Biometric Vocabulary. Technical report, ISO/IEC JTC1 SC37 N1248, November 2005.

[2] Working Group 3. *ISO/IEC 19794-2 Information Technology - Biometric Data Interchange Formats - Part 2: Finger minutia data*. JTC1 :: SC37, international standard edition, 2005. http://isotc.iso.org/isotcportal.

[3] Working Group 5. *ISO/IEC 19795-1 Biometric Performance Testing and Reporting: Principles and Framework*. JTC1 :: SC37, international standard edition, August 2005. http://isotc.iso.org/isotcportal.

[4] ANSI-INCITS 378-2004, Washington. *Fingerprint Minutiae Format for Data Interchange*, 2004. American National Standard.

[5] ANSI-INCITS 381-2004, Washington. *Finger Image-Based Data Interchange Format*, 2004. American National Standard.

[6] A C. Wilson, Hicklin, H.Korves, B. Ulery, M. Zoepfl, M. Bone, P. Grother, R. Micheals, S. Otto, and C. Watson. *Fingerprint Vendor Technology Evaluation, FpVTE*. National Institute of Standards and Technology, NISTIR 7123 edition, June 2004.

[7] R. Capelli, D. Maio, D. Maltoni, J. L. Wayman, and A. K. Jain. Performance evaluation of fingerprint verification systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(1):3–18, January 2006. Relates to FVC 2004.

[8] P. J. Grother. *ISO/IEC 19795-4 Biometric Performance Testing and Reporting: Performance and Interoperability Testing of Data Interchange Formats*. JTC1 :: SC37 :: Working Group 5, committee draft 3 edition, January 2006. http://isotc.iso.org/isotcportal.

[9] International Labour Organization, Geneva. *ILO Seafarers' Identity Documents Biometric Testing Campaign Report*, 2005.

[10] A. K. Jain, S. Prabhakar, and S. Chen. Combining combining multiple matchers for a high security fingerprint verification system. *Pattern Recognition Letters*, 20(3):1371–1379, March 1999.

[11] J. Kittler, M. Hatef, R. Duin, and J. Matas. On combining classifiers. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 20(3), March 1998.

[12] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain. Fvc2000: Fingerprint verification competition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(3):402–412, March 2002.

[13] R. Michael McCabe. Data format for the interchange of fingerprint, facial, scar mark and tattoo information. Technical report, ANSI-NIST, 2005.

[14] P.J. Phillips, P. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, and M. Bone. Face recognition vendor test 2002. Evaluation Report IR 6965, National Institute of Standards and Technology, www.itl.nist.gov/iad/894.03/face/face.html or www.frvt.org, March 2003.

[15] R. Snelick, U. Uludag, A. Mink, M. Indovina, and A. Jain. Large scale evaluation of multimodal biometric authentication using state-of-the-art systems. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 27(3):450–455, March 2005.

[16] Elham Tabassi. *Fingerprint Image Quality, NFIQ*. National Institute of Standards and Technology, NISTIR 7151 edition, 2004.

[17] Elham Tabassi. Fingerprint image quality, nfiq. In *IEEE International Conference on Image Processing ICIP-05*, Genoa, Italy, September 2005.

A Procedures for Validating Conformance with the MINEX Specification

All submitted SDK libraries were tested for conformance with the MINEX specification. These conformance tests were carried out in several stages. These are listed in the chronology of Table 3. Testing continued throughout the MINEX test. The stages are described in the following four sub-sections.

Note that all problems detected during a particular validation stage were promptly reported to the participant via email, along with any known details about the problem which could be released without divulging sensitive information. Participants were encouraged to fix the reported problem(s) as quickly as possible (e.g. in the case of Stage 1, prior to the SDK submission deadline) and resubmit their data or SDK as necessary via email. During stages 2 through 4, any resubmitted SDKs were retested with all prior validation stage tests up through the current stage to ensure both continued compliance (i.e. regression testing) and to ensure that no other changes, for example improvements, were made to the SDK outside of fixing the specific problems(s) reported by NIST. As was often the case, resubmitted SDKs were sometimes rejected (and another resubmission was called for) for a multitude of reasons such as the failure of the participant to fix the reported problem(s); the re-introduction of previously corrected problems; the detection of new problems; or changes to previously validated code functionality such as feature extraction.

A.1 Stage 1: Pre-Submission Sample Template Validation

NIST provided 10 images to prospective test participants and required submission of proprietary, conformant MIN:A , and optionally MIN:B templates. These were validated at NIST, and success in this stage was a necessary condition for continued participation. The purpose here was to detect and fix SDKs which had clear conformance problems. Along with each set of templates, participants were required to submit a log named "RtnCodes.txt" containing one line for each call made to create_template(), along with the arguments passed and the return code from the call.

Compliance with the MINEX API specification was checked using both automated and manual procedures. Collectively these procedures are referred to as template validation. This relied primarily on two programs developed by NIST, *prfmr* and *minexv* which are both supported by NIST's *libfmr* library which creates, reads, parses, writes and validates the core data block of INCITS 378 records. The *prfmr* program attempts to print out the contents of an INCITS 378 [2] compliant template in human readable form. The *minexv* program attempts to automatically check for compliance to the MINEX API specification. (Note that severely malformed templates had the ability to "break" the parser underlying both *prfmr* and *minexv*, and thus visual byte-by-byte inspection was sometimes required). These programs were run only on MIN:A and MIN:B type templates; proprietary templates were checked for adherence to the length constraints defined in the MINEX API only.

The checks performed attempt to detect non-compliance with MINEX API specification, which defines templates in terms of INCITS 378 (using a set of constraints and modifications specified in section 2.4). Conformance testing was limited to inspection of the template, and to its consistency with image metadata such as height, width. quality and position. This did not involve testing whether the minutiae recorded was faithful to the original image content. For example, although minutiae (x, y) locations were checked to ensure that they lie within the bounds of the associated image they were not assessed for correct placement with respect to the image itself. Similarly, all minutiae angles (θ) were checked to ensure that they were encoded in the range 0-179, but not checked for correct angle determination with respect to the associated ridge bifurcation in the image.

Nearly all participants returned their sample template results ahead of the March 15th deadline. Numerous problems were detected with most but not all of the participants. These were resolved iteratively using email in several submit-report-fix-resubmit iterations.

Of the problems detected one is notable here: For a blank image included in the ten trial images many participants produced improper "NULL templates." One of the 10 sample images provided by NIST was a blank image, and for this image the SDKs were expected (but not required) to return a nonzero error code (e.g. 3 - Failed to Extract Minutiae) from create_template(). The MINEX specification defines the return of any nonzero value from create_template() as being a Failure-to-Enroll (FTE) case, and requires the output of a NULL template (as defined by the MINEX specification) from the create_template() function. Most, but not all, SDKs returned a nonzero return code for the blank image, as was their option. Others returned 0 indicating successful minutiae extraction. Such templates contain very few or zero minutiae.

This is permissible under the MINEX specification, which had no way of specifying exactly when an SDK should generate a FTE. However, those SDKs which did return a nonzero error code and thus were required to return a NULL template often either failed to do so, or returned a template which didn't meet the NULL template formatting requirements. (The latter cases were most probably attributable to confusion over wording in the MINEX specification.) For example, several SDKs which signaled FTE by returning a nonzero status code returned templates containing a single or very few minutiae. Others appeared to be attempting to follow the MINEX specification of NULL template, except for missing some required bytes, or having additional bytes. As an example of the latter, some SDKs supporting MIN:B output NULL templates which contained an additional zero-length Ridge Count table. A significant portion of the effort expended during the sample template validation stage was spent in getting the participants to generate NULL templates in a compliant and uniform way.

A.2 Stage 2: Pre-SDK Acceptance Validation

In Stage 2 vendors submitted a SDK to NIST for Pre-SDK Acceptance Validation. Sucessful completion of Stage 2 validation was necessary for MINEX participation. The purpose here was to ensure that each SDK could be successfully integrated (i.e. linked) with the test driver program, and operate correctly (for both template generation and matching) in accordance with the MINEX API specification. The driver program would be used later to perform the complete MINEX tests. The focus here was on duplicating the results of the previous Stage 1 sample template validation by running the SDKs on NIST hardware platforms under the control of the NIST, using the test NIST driver.

Once compiled and linked with the test driver, each SDK was used to prepare templates of each supported template type (MIN:A etc) from the 10 sample data images. A binary comparison to the Stage 1 templates was made. In addition, the return codes from each call to the create_template() function were recorded and compared to the RtnCodes.txt file submitted in Stage 1 for that SDK. If identical results were not obtained, the participant was contacted with the details (often being given the templates generated by NIST using their SDK), and an opportunity to fix the problem(s) by resubmitting their SDK was given.

During these tests, mismatches were encountered in the case of several of the submitted SDKs. Some were eventually traced to platform issues, such as differences in library behavior or linker versions (between NIST's and participant's platform). In some of these cases the templates generated were technically conformant, but a few of the minutia differed between the corresponding templates generated by NIST and those submitted in Stage 1 by the participant. In other cases, the templates generated on the NIST platform failed to pass the automated (e.g. minexv) validations entirely. The latter were usually attributable to latent bugs in the code, or unanticipated dependencies on platform specific behavior.

A test of each SDK's matcher was performed as well, by generating a 10x10 similarity score matrix using the templates from the previous step. The focus here was primarily on checking that the matcher produced valid scores and expected return codes for all comparisons (especially in the case of NULL templates). If results were obtained that didn't meet the MINEX specification (e.g. failure to generate a failure code when one or both of the templates was NULL), this behavior was reported to the participant, and an opportunity to fix the problem was given by resubmitting their SDK(s). During these tests, problems such as the improper handling of NULL templates, and the improper returning of other error codes were detected in several of the SDKs. Also, differences in the scores produced when matching NULL templates (permissible under the MINEX specification) led to a change in the test driver software, which now automatically outputs a score of -1 when any matching errors occur, such as in the case of matching NULL templates.

A.3 Stage 3: Pre-Test SDK Validation

Prior to the full-scale MINEX test, additional conformance checking was performed. This was supplemented with stress testing. The purpose here was to ensure that, prior to starting full-scale MINEX testing, all SDKs had been run on larger and representative data sets to detect any remaining template generation problems, and cross-matching of templates (between SDKs) had been performed to test software robustness.

Template generation and validation testing was conducted in similar to that in stage 2, but this time using a different (and slightly larger) set of 20 images. At this time several more template problems were detected, with some of these being of the same type that were checked for (but not detected) before in Stage 1 using the previous 10 image data set. Thus, some SDKs which had passed Stage 2 validation successfully began generating non-conformant templates on this additional validation test data. These were primarily attributable to latent problems in the code being touched upon by the use of a different and large data set. As in the prior validation stages, participants whose SDKs generated non-conformant templates were contacted with the details, and an opportunity to fix the problems by resubmitting their SDK was given.

Other problems detected here (and not in Stage 2). These arose when NIST *minexv* conformance testing program was upgraded to include additional conformance checks. These improvements were often the result of decisions made by NIST regarding ambiguous or erroneous text found in INCITS 378 . Differing interpretations of INCITS 378 sometimes resulted in variations in template formatting across many of the participant's SDKs, particularly those implementing MIN:B . Ambiguities and mistakes in INCITS 378 were later reported to the committee.

A highly scaled down version MINEX test concluded Stage 3. Fewer problems were detected here than previously. These included problems with generation non-conformant templates, and some minor out-of-bounds issues. These were primarily attributable to data-dependent problems in the implementation that are inevitably discovered as more data is used. One more serious problem, hanging and occasional crashing of one participant's SDK, occurred when large numbers of template pairs were matched. At first this was thought to be a platform-related issue, but was ultimately traced by the participant to a bug in their code which caused it to enter an infinite loop during matching.

A.4 Stage 4: In-Test Validation

After all SDKs had successfully passed the validation checks of stage 3, they were declared ready for the full-scale MINEX testing. The actual testing of the SDKs began on May 1, 2005, and completed on December 30, 2005. This phase involved production of all templates to be used in matching, validation of those, and then execution of the matching itself. A very small number of often esoteric template problems were encountered.

The purpose of Stage 4 validation was primarily to check the conformance of all templates created during the template generation phase of MINEX before passing these templates through to the matching phase of MINEX . This was accomplished by the use of *minexv* as an in-line test during the template generation phase. A few more latent problems with non-compliant templates being generated were detected here (usually simply out-of-bounds type issues), and the SDKs were corrected as before.

In addition, the purpose of this stage was to monitor for and correct any issues which arose such as crashing or non-conformant behaviors during either template generation or matching. Fortunately no such issues arose.

B MINEX Datasets

All datasets used were left and right index fingers only using Live-scan plain impressions. The original images were given to NIST already WSQ compressed at approximately 15:1. The images were given to the template extraction algorithms as decompressed (using NIST's WSQ decoder) "raw" pixel data. The original target sample sizes were 62,000 mates and 122,000 non-mates. These totals were reduced after consolidations (see section B.1) and a few WSQ decompression failures were taken into account.

The U.S.-VISIT dataset is split into two datasets. The first dataset used POE data for both the authentication and enrollment images. This occurs in practice when a subject gets enrolled at a POE station and then reenters through a POE station at a later date. The second dataset used BVA images for the enrollment and POE images for the authentication samples. This occurs when a subject is enrolled at the BVA station and then enters through at a POE. When randomly selecting the datasets none of the POE images selected for the POE vs. POE dataset were reused in the POE vs. BVA dataset.

The testing was performed by using the second instance of the mates as the enrollment image and the first instance as the authentication image. So for each dataset there were a little under 62,000 mate scores. The non-mate scores were generated by comparing the non-mate authentication samples to the same enrollment images used with the mates, so for non-mate scores most enrollment images were used twice. This generated a little under 122,000 non-mate scores for a total of just under 184,000 scores per finger/dataset.

Testing was performed by first extracting all the templates for each vendor and all the datasets. These templates were then distributed across several machines to perform inter-vendor matching. Matching was done by randomly mixing the mate and non-mate pairs before passing the pairs to the matchers one at a time.

B.1 Consolidation and Ground Truthing

The MINEX analysis phase included a process for detecting ground truth errors in the data. An error in this context is categorized into one of two types. The first occurs when a pair of images are thought to be of the same person but are actually not. Such Type I errors erroneously increase the false non-match rate. The second kind of error occurs when the two images are thought to be from different people but are actually from the same person. Such Type II errors give erroneously high false match rates. Both kinds of errors exist to some extent in virtually all large datasets. The errors are due to flaws in the metadata associated with the samples in the database.

In an evaluation the errors manifest themselves as low scoring matches and high-scoring non-matches. The procedure used for handling such errors in MINEX is to submit the images that produce candidate "problem" comparisons for human inspection and resolution. Such candidate lists were formed from the images that produced anomalously low and high scores in matching trials *from multiple matchers*. This consensus aspect is key to separated legitimately low or high scores from aberrant ones.

In MINEX , no Type II errors were observed. A few hundred pairs of images that produced high scores were submitted for human examination. None were determined to truly match. This finding is consistent with an operational database in which there are persons who are multiply enrolled.

A number of Type I errors were detected, however. The incidence of these in each of the four MINEX datasets is recorded in Table 24. The images that were found to produce consistently low match scores were submitted for examination. Only those images that were clearly not a matched pair were excluded. Pairs were not removed if one of the images was of low enough quality to make if difficult for the inspector to determine ground truth.

These errors arose for two reasons: First, one of the two images was either blank or so low in contrast that only a few ridges could be seen. Second, the left and right fingers of a person were swapped. This caused a first impression of a right finger to be compared with an impression of the left finger rather than a second impression of the right. In all cases the detection of an error was handled by deleting the genuine scores associated with comparisons involving that person's left and right fingers. This preserved the strict left and right pairings which were employed to support the two-finger matching results obtained using the fusion technique of section 5.4. scoring, when an error occured for either finger that "person" was removed from the matches score files. Impostor scores from these fingers were not excised as their impostor status

POEBVA	
Enrollment Description	Data from U.S. Consulate Offices captured when people applied for a U.S. VISA.
Authentication Description	Data from U.S. VISIT captured from persons entering the U.S. at airport ports of entry (POE)
Verification Environment	Airports (POE)
Enrollment Environment	Consular Offices (BVA)
Enrollment Capture Device	Smiths-Heimann ACCO 1394
Verification Capture Device	Cross Match 300A
Number of genuine comparisons	61,531 (BVA)
Number of impostor comparisons	121,994 (BVA)
Enrollment Image Size	500x500
Verification Image Size	368x368
Selection	The mates were randomly selected from the 290,000 subjects that had a mate in the POE dataset. The non-mate data was randomly selected from the remaing POE subjects with only 1 instance. This non-mate data did not include any of the non-mate data from the POE vs POE testing.
Sex (Male, Female, Unknown) (Genuine users)	51.5% 48.4% 0.1%
Sex (Male, Female, Unknown) (Impostors)	50.6% 48.5% 0.1%
DHS2	
Description	DHS recidivist cases, the majority of which are border crossing cases with Mexico.
Environment	border patrol field operations.
Capture	These prints were captured in varying environments that include both indoors and outdoors. Some fraction of the population was uncooperative.
Capture Device	Identix DFR-90
Image Size	368x368
Number of genuine comparisons	61,561
Number of impostor comparisons	121,994
Sex (Male, Female, Unknown) (Genuine users)	86.9% 13.0% 0.1%
Sex (Male, Female, Unknown) (Impostors)	85.2% 10.5% 4.3%
POE	
Description	Data from U.S. VISIT captured from persons entering the U.S. at airport ports of entry (POE)
Environment	Indoors with auto capture being used to acquire the best possible quality image from the subject in the operational allowed time frame (3 or 5 secs). Subjects are generally cooperative. This data was captured during the first months of the current two-finger US-VISIT process.
Capture Device	Cross Match 300A (POE)
Number of genuine comparisons	61,751 mates (POE)
Number of impostor comparisons	121,999 non-mates (POE)
Image Size	368x368 (POE)
Relationship to POEBVA	The POE dataset is drawn from the same population as the POE half of the POEBVA dataset. The two extracts are disjoint (i.e sampled from the parent without replacement). The subjects used in genuine comparisons were drawn from the population with two or sets of images, corresponding to two visits to POEs. The subjects used in impostor comparisons were drawn from the population with only one set of images.
Sex (Male, Female, Unknown) (Genuine users)	59.6% 40.0% 0.4%
Sex (Male, Female, Unknown) (Impostors)	50.5% 48.6% 0.9%
DOS	
Description	DOS non-immigrant visa cases. These are persons applying for border crossing cards. Collection is indoors in a high volume operation.
Environment	US Consular Offices in Mexico
Capture Device	Identix DFR-90
Image Size	368x368
Number of genuine comparisons	61,866
Number of impostor comparisons	121,958
Sampling	Sampling was done by randomly selecting the 62,000 mates from the approximately 274,000 subjects with multiple instances. The non-mates were randomly selected from the 5.5 million subjects with only one instance per subject.

Table 23: Summary properties of the MINEX datasets.

Dataset	Number Reviewed	Number Removed
POEBVA	923	469
DHS2	2989	431
POE	812	245
DOS	737	81

Table 24: Type I Consolidation Activity

		1 Best		2		3		4		5 Worst		
Dataset	Finger	Num.	Prop.	Num.	Prop.	Num.	Prop.	Num.	Prop.	Num.	Prop.	Mean
poebva	R	104273	0.424	77241	0.314	50594	0.206	6336	0.026	7550	0.031	1.93
	L	108709	0.442	66005	0.268	52085	0.212	8443	0.034	10752	0.044	1.97
dhs2	R	107547	0.437	83106	0.338	38699	0.157	1684	0.007	14942	0.061	1.92
	L	114969	0.467	77844	0.316	38567	0.157	1498	0.006	13100	0.053	1.86
poe	R	77499	0.315	92017	0.374	62778	0.255	6158	0.025	7539	0.031	2.08
	L	85528	0.348	80219	0.326	62210	0.253	6977	0.028	11057	0.045	2.10
dos	R	112733	0.459	99310	0.404	25868	0.105	3876	0.016	4065	0.017	1.73
	L	106122	0.432	92267	0.375	35226	0.143	5227	0.021	7010	0.029	1.84

Table 25: Summary of NIST Fingerprint Image Quality values for the four data sets.

remained correct.

B.2 Database Quality

Table 25 gives the composition of the four databases by NFIQ value.