

FpVTE 2003 Evaluation Plan

Note: The FpVTE Test Overview serves as an introduction to this document.

Table of Contents

FpVTE 2003 Evaluation Plan	1
1 Terminology Used in This Document	3
Flat fingerprint	3
Segmented slap.....	3
Rolled fingerprint.....	3
ANSI/NIST.....	3
WSQ.....	3
Fingerprint set.....	3
Dataset	3
Subject.....	3
Target Set.....	3
Query Set.....	4
Similarity matrix.....	4
Mate	4
Self-ident.....	4
Preprocessing	4
Finger number	4
2 Overview of Tests	5
2.1 Small-Scale Test (SST).....	5
2.2 Medium-Scale Test (MST).....	5
2.3 Large-Scale Test (LST).....	6
3 Evaluation Data	7
3.1 Mate Relationships	7
3.2 Individual Flat Fingerprint Images (SST, MST, and LST)	7
3.3 Segmented Slap Fingerprint Images (MST and LST).....	8
3.3.1 Rotation of Slap Images.....	9
3.3.2 Masking of Extraneous Ridge Detail in Slap Images	10
3.4 Rolled fingerprints (LST only)	10
4 Similarity Scores and Matrices	12
4.1 Reporting Similarity Scores for Multi-Stage Systems	13
5 Large-Scale Test (LST) Details	14

5.1 Evaluation Datasets Used in LST	14
LST Dataset A: 2F.....	14
LST Dataset B: 1F.....	14
LST Dataset C: 10S-L.....	14
LST Dataset D: 10S-P.....	14
LST Dataset E: 8S-L.....	14
LST Dataset F: 4S-L.....	14
LST Dataset G: 2S-L.....	15
LST Dataset H: 1S-L.....	15
LST Dataset I: 10R-L.....	15
LST Dataset J: 10R-P.....	15
5.2 LST Subtests	15
6 Advanced Topics	17
6.1 Normalization	17
6.2 Failure to Enroll and Fingerprint Quality	17
7 Test Procedures	19
Test Preparation.....	19
Conduct of Test.....	20
Post-Test Procedures.....	20

1 Terminology Used in This Document

Flat fingerprint

A fingerprint image collected from a single-finger livescan device, resulting from the touching of a finger to a platen without any rolling motion. Also known as a single-finger plain impression.

Segmented slap

An image of a single fingerprint that was segmented (cropped) from an image of a 4-finger slap (4-finger simultaneous impression), such as found at the bottom of a fingerprint card. Slaps may be from livescan devices or scanned from paper fingerprint cards. FpVTE segmented slaps have been segmented using automatic and/or manual processes; all segmentation has been human verified.

Rolled fingerprint

A fingerprint image collected by rolling the finger edge to edge across the livescan platen (or paper) from nail to nail. Rolls may be from livescan devices or scanned from paper fingerprint cards.

ANSI/NIST

A file format for fingerprint files compliant with NIST Special Publication 500-245, Data Format for the Interchange of Fingerprint, Facial, & Scar Mark & Tattoo (SMT) Information (http://www.nist.gov/customcf/get_pdf.cfm?pub_id=151453) The FBI's Electronic Fingerprint Transmission Specification (EFTS) is based on ANSI/NIST. Fingerprint files that are EFTS compliant are necessarily ANSI/NIST compliant. In FpVTE, all images embedded in ANSI/NIST files use WSQ compression.

WSQ

Wavelet Scalar Quantization. The image compression method used for fingerprint images.

Fingerprint set

A single ANSI/NIST file containing multiple fingerprint images from a single individual, collected at one time. The fingerprint positions (finger numbers) are noted in the file. The finger positions are not repeated in the file: no more than one fingerprint per position is included.

Dataset

A collection of multiple fingerprint sets.

Subject

An individual person

Target Set

The dataset being searched against in a given test or subtest: an experiment searches a Query Set against a Target Set. Also known as a File set or just fingerprint database.

Query Set

The dataset containing the searches for a given test or subtest: an experiment searches a Query Set against a Target Set. Also known as a Search set.

Similarity matrix

A matrix of Participant-specific matcher scores, which compare each member of a Query Set against each member of a Target Set. The file format for a similarity matrix is defined in the Data Format Specification.

Mate

Different fingerprint sets are mates if they came from the same person. The tests (of course) do not note which fingerprint sets are mates.

Self-ident

The special case of a fingerprint set (or an individual fingerprint) being compared against itself. Self-idents are ignored during analysis. When a dataset is compared against itself and a square matrix of scores is generated, the scores on the diagonal are self-idents.

Preprocessing

Also known as Characterization or Feature Extraction. The process of creating a machine representation of a fingerprint image. A few matchers do not perform preprocessing.

Finger number

Finger 01 Right thumb
Finger 02 Right index
Finger 03 Right middle
Finger 04 Right ring
Finger 05 Right little
Finger 06 Left thumb
Finger 07 Left index
Finger 08 Left middle
Finger 09 Left ring
Finger 10 Left little

2 Overview of Tests

2.1 Small-Scale Test (SST)

The SST is designed for Participants whose throughput rates will not allow them to complete the MST. This test will evaluate matching accuracy using individual fingerprint images, not sets of multiple fingerprint images.

The test will consist of a single SST dataset containing 1,000 ANSI/NIST files. The SST dataset will be used as both the Query Set and the Target Set — in other words, all fingerprints in the dataset will be compared against all other fingerprints in the dataset.

The images in the dataset may have zero, one, or more mates in the dataset (disregarding self-idents). The SST dataset will consist exclusively of single-finger flat images (not segmented slap or rolled images). The images will be images of the right index finger (Finger 02). No other fingers will be included in the test.

The SST must be completed in a period of no more than two weeks (1,209,600 seconds), not including setup and checkout.

SST Participants may ***optionally*** provide normalization code for post-processing of the SST Similarity Matrix. See the [Normalization](#) section below for additional information.

2.2 Medium-Scale Test (MST)

The MST is designed to evaluate matching accuracy using individual fingerprint images, not sets of fingerprint images.

The test will consist of a single MST dataset containing 10,000 files.

The MST dataset will be used as both the Query Set and the Target Set — in other words, all fingerprints in the dataset will be compared against all other fingerprints in the dataset. The MST dataset will be larger than the SST dataset. The images in the dataset may have zero, one, or more mates in the dataset (disregarding self-idents).

The MST dataset will consist exclusively of single-finger flat images and segmented slap images (not rolled images). 50-60% of the images will be single-finger flat; 40-50% of the images will be segmented slaps. All of the images will be from livescan devices. Most of the images will be images of the right index finger (Finger 02) but some may be of the right middle finger (Finger 03). No other fingers will be included in the test. The finger number will ***not*** be identified for each image.

The MST is designed so that the SST is a subset of the MST. That is, all images in the SST dataset will be scattered within the first quarter of the MST dataset. The MST will be sized with the expectation that all MST Participants will complete the test within the allotted time. However, if an MST Participant only partially completes the MST for any reason, the likelihood is high that the SST comparisons within the MST would have been completed. In this case, the FpVTE 2003 personnel may be able to meaningfully analyze the partially completed MST results in the context of the SST.

The MST must be completed in a period of no more than two weeks (1,209,600 seconds), not including setup and checkout.

MST Participants may ***optionally*** provide normalization code for post-processing of the MST Similarity Matrix. See the [Normalization](#) section below for additional information.

2.3 Large-Scale Test (LST)

The LST is designed primarily for AFIS Participants. The test is composed of a series of subtests to measure:

- Performance of rolled fingerprint sets against rolled fingerprint sets. Subtests of 10 fingers per set will be conducted.
- Performance of segmented slap fingerprint sets against rolled fingerprint sets. Subtests of 1,2,4,8, and 10 fingers per set will be conducted.
- Performance of segmented slap fingerprint sets against segmented slap fingerprint sets. Subtests of 1,2,4,8, and 10 fingers per set will be conducted.
- Performance of flat fingerprint sets against flat fingerprint sets. Subtests of 1 and 2 fingers per set will be conducted. All fingers are index fingers (fingers 02 and 07).

The rolled and segmented slap images come from livescan devices, or from paper fingerprint cards that were scanned on flatbed scanners. Images from paper cards may include pencil marks, or printed lines and text from the card itself.

The LST must be completed in a period of no more than three weeks (1,814,400 seconds), not including setup and checkout.

More information about the LST is included in [Section 5, Large-Scale Test \(LST\) Details](#) .

LST Participants do not provide normalization code for post-processing of similarity matrices. See the [Normalization](#) section below for additional information.

3 Evaluation Data

The fingerprints in FpVTE have been collected from a range of U.S. Government sources. Some of the fingerprints are representative of current operational data, while others are representative of legacy data. In practice, this means that the test data will range from good to poor quality.

It is critical to note that FpVTE will report multiple results based on multiple types or sources of fingerprints. It is assumed that an individual matcher will have very different performance using pristine, carefully collected data instead of legacy operational data. The results of FpVTE will not be a single data point or a single chart, but a series of charts describing the performance for the different types of data.

3.1 Mate Relationships

Most datasets are searched against themselves. Obviously such searches generate self-idents, which are ignored.

Each individual dataset will contain some mated subjects. Many subjects will not have mates within a single dataset, while some subjects will have more than one mate within a single dataset.

When searching a Query set against a different Target set, subjects in the Query set may have zero, one, or more mates in the Target set.

A dataset occasionally may contain duplicate images, which may be either the same livescan image or the same paper image scanned twice. These self-idents will be ignored during analysis, but similarity scores should be provided.

3.2 Individual Flat Fingerprint Images (SST, MST, and LST)

In FpVTE, a flat fingerprint is a fingerprint image collected from a single-finger livescan device, resulting from the touching of one finger to a platen without any rolling motion. A flat fingerprint is also known as a single-finger plain impression. In FpVTE, the term “flat” fingerprint always means an individual flat fingerprint and should not be confused with a “segmented slap.”

All of the fingerprints in the SST Evaluation Dataset will be individual flat fingerprint images. Some of the fingerprints in the MST, and LST Evaluation Datasets will be individual flat fingerprint images.

The flat fingerprints in the Evaluation Datasets were collected by the following single-finger optical livescan devices:

- Identix/Identicator DFR-90
- Cross Match Verifier Model 300
- Identix TV555

The type of scanner used to acquire each fingerprint will **not** be provided in the tests.

(The listing of makes and models does not imply a recommendation by NIST or FpVTE personnel, but simply recognizes the actual devices used by the variety of Federal agencies that contributed data to FpVTE.)

All images are 500 pixels per inch, 8-bit grayscale images.

Most of the flat fingerprints are 368 pixels by 368 pixels; the size may vary from 368x368 to 420x480 (width x height).

All of the flat fingerprints are from the index fingers. In SST and MST, only the right index fingers are used. In LST, both index fingers are used, and the finger position is always noted in the ANSI/NIST file.

The images are usually upright, but are sometimes rotated up to about ± 25 degrees, and rarely up to about ± 45 degrees. The core is usually (but not always) centered in each image.



Figure 1: Sample Flat Fingerprint

3.3 Segmented Slap Fingerprint Images (MST and LST)

In FpVTE, a segmented slap is an image of a single fingerprint that was segmented (cropped) from an image of a 4-finger slap (4-finger simultaneous impression), such as found at the bottom of a fingerprint card. Slaps may be from livescan devices or scanned from paper fingerprint cards. FpVTE segmented slaps have been segmented using automatic and/or manual processes; all segmentation has been human verified.

Some of the fingerprints in the MST and LST Evaluation Datasets will be segmented slap fingerprint images.

The slap fingerprints in the Evaluation Datasets were collected by the following multi-finger optical livescan devices:

- Identix TP-600
- CrossMatch ID-1000
- DBI Tenprinter
- Heimann LS2 Check

The type of scanner used to acquire each fingerprint will ***not*** be provided in the tests.

(The listing of makes and models does not imply a recommendation by NIST or FpVTE personnel, but simply recognizes the actual devices used by the variety of Federal agencies that contributed data to FpVTE.)

All of the slap images in MST are from livescan devices.

Some of the slap images in LST were scanned from paper fingerprint cards, using FBI EFTS Appendix F-certified flatbed scanners. Fingerprints scanned from paper fingerprint cards may contain extraneous text, lines, or other marks.

All images are 500 pixels per inch, 8-bit grayscale images.

The size of segmented slap fingerprints varies. It may be as small as 150 pixels by 150 pixels, and may be as large as 500x600 (width x height).

In MST, slap fingerprints may be from the right index or right middle fingers, but the finger position is not noted in the ANSI/NIST file.

In LST, a variety of finger combinations is used, and the finger position is always noted in the ANSI/NIST file.

[Figure 2](#) and [Figure 3](#) show an example of a good-quality livescan slap image before and after segmentation. Note that part of the little finger was not included in the slap image: incomplete fingerprints such as this can sometimes occur with any finger in slap images.



Figure 2: Unsegmented 4-finger slap image (Unsegmented images are not used in FpVTE)



Figure 3: Segmented slap images as used in FpVTE — Note rotation

3.3.1 Rotation of Slap Images

Slap images (except for thumbs) are usually rotated, as shown in [Figure 3](#). The average rotation is 20-25 degrees. Few images are rotated more than 45 degrees. Fingers from the left hand are usually rotated clockwise, and those from the right hand are usually rotated counterclockwise.

3.3.2 Masking of Extraneous Ridge Detail in Slap Images

In some cases, the rectangles used to segment images include extraneous ridge detail, as shown in [Figure 4](#). This extraneous ridge detail is **excluded** in FpVTE by use of a white mask, as can be seen by comparing [Figure 3](#) (with mask) and [Figure 4](#) (without mask). Note that ridge detail below the crease is excluded (as shown in the ring finger image).



Figure 4: Segmented slap images showing extraneous ridge detail excluded in FpVTE

3.4 Rolled fingerprints (LST only)

A rolled fingerprint is a fingerprint image collected by rolling the finger edge to edge across the livescan platen (or paper) from nail to nail. Rolls may be from livescan devices or scanned from paper fingerprint cards.

Some of the fingerprints in the LST Evaluation Datasets will be rolled fingerprint images.

The rolled fingerprints in the Evaluation Datasets were collected by the following multi-finger optical livescan devices:

- Identix TP-600
- CrossMatch ID-1000
- DBI Tenprinter
- Heimann LS2 Check

The type of scanner used to acquire each fingerprint will **not** be provided in the tests.

(The listing of makes and models does not imply a recommendation by NIST or FpVTE personnel, but simply recognizes the actual devices used by the variety of agencies that contributed data to FpVTE.)

All images are 500 pixels per inch, 8-bit grayscale images.

Rolled fingerprints can vary in size from 500 pixels by 500 pixels up to 800x750 (width x height).

The images are usually upright. The core is usually (but not always) centered in each image.



Figure 5: Sample Rolled Fingerprint (Scanned from Paper)

4 Similarity Scores and Matrices

The output from FpVTE tests is a Participant-specific measure of similarity known as a **similarity score**. For most Participants, this corresponds to a matcher score. Note that the systems **do not** return Boolean Match and Non-match determinations: for analysis to be meaningful, a continuous distribution of degrees of similarity must be present.

Each similarity score compares the similarity of the fingerprints in an ANSI/NIST file to those in another ANSI/NIST file:

- In SST and MST, an ANSI/NIST file always contains a single fingerprint, so a similarity score is a determination of the similarity of an **individual** fingerprint to another **individual** fingerprint.
- In LST, an ANSI/NIST file may contain from one to ten fingerprints (collected at one time from an individual), so a similarity score is a determination of the similarity of a **set** of fingerprints to another **set** of fingerprints. Put another way, in LST, when comparing a set of ten fingerprints to another set of ten fingerprints, **one** similarity score is put in the similarity matrix, **not ten** scores.

The scale used for similarity scores is entirely up to the Participant: one Participant may use a scale of 0.0 (no similarity) to 1.00 (identical), while another may use a scale of -1,000,000 to 1,000,000. It is also permissible for Participants to use distance scores, in which increasing values indicate dissimilarity rather than similarity — for simplicity, only similarity scores are discussed in this document.

Scores should not be quantized to a limited range. If internal matcher scores are reported on a 0 to 1000 range, quantizing (or grouping) them to a 1 to 5 scale will have a negative effect of the FpVTE results for that matcher.

A **similarity matrix** is the array of scores produced by a matcher when a Query set is searched against a Target set. If a Query set contains 1,000 ANSI/NIST files and a Target set contains 3,000 ANSI/NIST files, the resulting similarity matrix contains 1,000 columns, each of which contains 3,000 scores. The exact format of a similarity matrix is defined in the *Data Format Specification*. In short, the values in a similarity matrix are stored as floating point numbers, and each column of values is stored as a separate file, with a minimal header.

- SST and MST Participants will produce only one similarity matrix.
- LST Participants will produce multiple similarity matrices, one for each subtest.

FpVTE analyses are based on distributions of similarity scores for mate vs. non-mate comparisons. Within a single similarity matrix (of scores generated in a single subtest), all the scores **must** be comparable: a higher score **must** mean a higher degree of similarity.

For the different subtests in LST, scores do **not** have to be comparable between different subtests (and the associated similarity matrices). For example, different subtests can use different scoring methods: the scores when comparing 2-flats against 2-flats do not have to correspond to the scores comparing 10-rolls against 10-rolls.

Systems will sometimes modify the distribution of scores based on the Target set being used, a process known in FpVTE as Normalization. Please see the Normalization section for details.

4.1 Reporting Similarity Scores for Multi-Stage Systems

Since a multi-stage AFIS generally filters out many non-mates before the final matcher assigns scores, the system will generally not calculate similarity scores for every comparison. Participants nevertheless will be expected to generate a fully-populated matrix of scores, but may fill the majority of the matrix with one or more default values. Participants are advised that the choice of true scores vs. default values will affect evaluation results, and are encouraged to fill the matrix as completely as possible with true scores.

There are several considerations guiding how similarity scores are reported for multi-stage systems:

- An AFIS will often only assign a matcher score to a small portion of searches, and all other comparisons default to a value indicating no similarity. For FpVTE a Participant may consider using some method that differentiates between borderline comparisons, and comparisons that were dropped at different matcher stages.
- Scores in different similarity matrices do not have to be comparable. For example, scores comparing flats to flats may have an entirely different scale than would roll to roll scores.
- FpVTE analyses assume that all scores above a Participant-defined level of similarity will be returned. Please do ***not*** return a fixed, limited number of scores for every search regardless of the level of similarity: this would negatively affect the analysis of matcher performance.

5 Large-Scale Test (LST) Details

5.1 Evaluation Datasets Used in LST

All datasets contain ANSI/NIST files. Every file in a given dataset contains the same number of images, in type-4 records, WSQ-compressed. In LST, all finger positions are noted (correctly) in the type-4 record headers. A single dataset will not commingle rolled, slap, and flat images. A single dataset will not commingle images from Livescan and Paper sources. The following list identifies the types and sizes of datasets that will be used in LST.

Sizes of individual datasets may vary slightly, but by no more than +/- 15%. The overall size of the test will not change substantially from the numbers stated here.

LST Dataset A: 2F

Set of 8,000 ANSI/NIST files, each containing 2 Flat fingerprint images, one for each index finger (fingers 02 and 07). All images are from livescan devices.

LST Dataset B: 1F

Set of 3,000 ANSI/NIST files, each containing 1 Flat fingerprint image, for either index finger (finger 02 or 07). All images are from livescan devices.

LST Dataset C: 10S-L

Set of 9,000 ANSI/NIST files, each containing 10 segmented Slap fingerprint images. Every file contains images for all 10 fingers. All images were collected from Livescan devices.

LST Dataset D: 10S-P

Set of 4,000 ANSI/NIST files, each containing 10 segmented Slap fingerprint images. Every file contains images for all 10 fingers. All images were scanned from Paper fingerprint cards. (Same as 10S-L, but contains only images scanned from Paper fingerprint cards.)

LST Dataset E: 8S-L

Set of 7,000 ANSI/NIST files, each containing 8 segmented Slap fingerprint images. Every file contains images for 8 fingers, excluding thumbs. All images were collected from Livescan devices.

LST Dataset F: 4S-L

Set of 7,000 ANSI/NIST files, each containing 4 segmented Slap fingerprint images. Every file contains images for 4 fingers. All images were collected from Livescan devices. LST Dataset F (4S-L) will contain approximately equal numbers of the following finger groups, which will be noted in the Dataset Definition File's Metadata attribute:

- 4S-L-TI (Thumb-Index: fingers 01,02,06,07)
- 4S-L-IM (Index-Middle: fingers 02,03,07,08)
- 4S-L-Right (Right: fingers 02,03,04,05)

In the dataset definition file, all of the 4S-L-TI files will be listed first, then the 4S-L-IM, then the 4S-L-Right.

LST Dataset G: 2S-L

Set of 7,000 ANSI/NIST files, each containing 2 segmented Slap fingerprint images. All images were collected from Livescan devices. LST Dataset G (2S-L) will contain approximately equal numbers of the following finger groups, which will be noted in the Dataset Definition File's Metadata attribute:

- 2S-L-T (Thumb: fingers 01,06)
- 2S-L-I (Index: fingers 02,07)

In the dataset definition file, all of the 2S-L-T files will be listed first, then the 2S-L-I.

LST Dataset H: 1S-L

Set of 3,000 ANSI/NIST files, each containing one single segmented Slap fingerprint image. All images were collected from Livescan devices. The dataset will include examples from all ten individual fingers, labeled in the Dataset Definition File's Metadata attribute as 1S-L-01 through 1S-L-10.

LST Dataset I: 10R-L

Set of 8,000 ANSI/NIST files, each containing 10 Rolled fingerprint images. Every file contains images for all 10 fingers. All images were collected from Livescan devices.

LST Dataset J: 10R-P

Set of 8,000 ANSI/NIST files, each containing 10 Rolled fingerprint images. Every file contains images for all 10 fingers. All images were scanned from Paper fingerprint cards. (Same as 10R-L, but contains only images scanned from Paper fingerprint cards.)

5.2 LST Subtests

The structure of the LST subtests is shown in [Figure 6](#). This structure includes ten LST datasets. Five of the datasets will be used as Target Sets, and all of the datasets will be used as Query sets. A total of 31 similarity matrices will be generated.

The subtests must be performed in row order: all of the tests using A as a target set must be performed first, then the tests using C as a target set, etc.

LST Subtests				Query Sets									
				A	B	C	D	E	F	G	H	I	J
				2F	1F	10S-L	10S-P	8S-L	4S-L	2S-L	1S-L	10R-L	10R-P
				8,000	3,000	9,000	4,000	7,000	7,000	7,000	3,000	8,000	8,000
Target Sets	A	2F	8,000	AxA	BxA	-	-	-	-	-	-	-	-
	C	10S-L	9,000	-	BxC	CxC	DxC	ExC	FxC	GxC	HxC	-	-
	D	10S-P	4,000	-	-	CxD	DxD	ExD	FxD	GxD	HxD	-	-
	I	10R-L	8,000	-	-	CxI	DxI	ExI	FxI	GxI	HxI	IxI	JxI
	J	10R-P	8,000	-	-	CxJ	DxJ	ExJ	FxJ	GxJ	HxJ	IxJ	JxJ

Figure 6: LST Subtests

Each dataset will be used in multiple subtests. It is assumed that any preprocessing for a dataset will be performed only once, not each time it is used.

Failure to complete the entire test during the allotted time will be noted in the FpVTE 2003 Final Report. FpVTE personnel will decide, after conclusion of the test, if partial test results will be evaluated and reported in the FpVTE 2003 Final Report.

Note also that all subtests should run without human administration: starting or completion of a subtest cannot require operator intervention.

6 Advanced Topics

The following Advanced Topics and the implementations discussed are *optional*.

6.1 Normalization

When an individual Query is searched against a Target set (database), a raw similarity score is computed for each pair-wise comparison. Normalization refers to a function that adjusts these scores. A normalization function simply maps raw scores to “normalized” scores, and it operates on the entire list of scores generated by each Query. For instance, the function might determine the mean and standard deviation of a set of raw scores, then adjust each score such that the resulting, normalized distribution has a mean of 0.0 and a standard deviation of 1.0.

One useful oversimplification is that normalization is the training of a system for the specific contents of each Target set.

A common misconception is that systems that perform 1:1 matches cannot use normalization. Verification systems can and do normalize based on the aggregate distribution of actual data.

In FpVTE, normalization is handled differently for the different tests:

- In MST and SST, Participants may optionally provide a compiled software object that performs normalization for use in post-test analysis. The format of this is described in the FpVTE Normalization Specification. The purpose for this is to allow FpVTE analysis on subsets of the similarity matrices, while still allowing Participants to control normalization. SST and MST results will be analyzed in three ways:
 - Participant-provided similarity scores will be analyzed
 - All scores will be normalized using FpVTE normalization methods, and analyzed
 - All scores will be normalized using Participant-specific normalization (if provided), and analyzed
- LST participants may perform normalization, but do *not* provide normalization software for post-test analysis. LST results will be analyzed based on subsets of Queries, but Target sets will not be partitioned during analysis. The rationale for this approach is based on the fact that a multi-stage AFIS generally returns true scores for only a portion of candidates from the Target set, and any post-test partitioning of the Target set might conflict with how those candidates were selected, thereby preventing statistically meaningful analyses. Analysis reports for LST will be conducted in two ways in terms of normalization:
 - Participant-provided similarity scores will be analyzed
 - All scores will be normalized using FpVTE normalization methods, and analyzed

It is important to differentiate between normalization, which is often used to great effect in real-world systems, and “gaming”, or attempting to take advantage of the test structure. A variety of anti-gaming measures have been designed into FpVTE, which include (but are not limited to) use of a mix of true imposters (subjects in the Query but not Target sets), background subjects (subjects in the Target but not Query sets), duplicate images, and multiple subjects in Query and Target sets.

6.2 Failure to Enroll and Fingerprint Quality

Some systems are designed to reject some fingerprints due to poor image quality. This is generally known as the Failure to Enroll (FTE) rate. FpVTE datasets include some poor

quality fingerprints. All fingerprints in FpVTE must be compared and similarity scores generated: fingerprints should not be ignored due to poor quality.

In addition to analysis of accuracy, FpVTE may include a secondary analysis of FTE and quality metrics. FpVTE provides an ***optional*** method for Participants to indicate which fingerprints would have been rejected as FTE in an operational system. Often this capability is used operationally to alert the fingerprint taker to attempt to obtain a better image

If a Participant's system uses quality metrics to reject poor quality fingerprints before attempting to match them under normal operations, then those quality metrics, and the thresholds used for rejection, can ***optionally*** be provided for FpVTE analysis. Whether or not a Participant provides quality metrics or FTE thresholds will not affect analysis of accuracy in any way.

The format for fingerprint quality vectors is defined in the Data Format Specification.

7 Test Procedures

Note: a detailed Test Procedures document will be provided to Participants when their evaluations are scheduled.

Test Preparation

FpVTE 2003 evaluations will be conducted at the NIST facilities at Gaithersburg, MD, no earlier than September 29 and no later than December 31, 2003.

Each FpVTE 2003 Participant will be assigned a date to arrive for the evaluation. Participants may ship equipment to NIST to arrive up to one week prior to their assigned date. NIST will provide suitable storage for shipped equipment, but will not provide personnel to set up or test equipment. Such set up and test activities are the responsibility of each Participant. Participants will be allotted a certain amount of time to set up and test their equipment. The System Throughput Specification requests an estimate of setup time.

Not all Participants will start the test on the same day.

Participants will be required to submit a five-page (maximum) System Description Document, in electronic form, on the first day of testing. This document will be included in the final FpVTE 2003 report that will be released to the public. This document must adequately address the following topics:

- Overview of the evaluated system(s).
- Component list for the evaluated system(s).
- Detailed cost breakdown of the submitted system(s) (commercial vendors only).
- Details of any modifications required to take FpVTE 2003.

FpVTE 2003 will serve as part of NIST's statutory mandate under section 403c of the USA PATRIOT Act to certify biometric technologies. This certification is for a specific system configuration. To define that specific system configuration, Participants will be required to submit a Configuration Management Document on the first day of testing. This document will be treated as Proprietary by NIST, and will not be disclosed without the permission of the Participant. The Configuration Management Document will contain sufficient information to enable the Participant to precisely recreate, at some later date, the system(s) evaluated in FpVTE 2003. In addition, future Government evaluations or interested agencies may request that a Participant use precisely the same system as was used in FpVTE and certified by NIST. Some of the FpVTE Participants' systems are expected to be custom configurations that may be difficult to recreate without sufficient documentation, as provided in the configuration management document. The configuration management document will be archived by NIST and will not be included in the final FpVTE 2003 report. A copy will be provided to the Participant upon request, or to interested Government entities with permission of the Participant. Participants with custom or 1-of-a-kind systems should be especially careful to delineate every hardware and software component, and all modifications, in the Configuration Management Document so that the system certified by NIST can be precisely recreated in the future.

FpVTE 2003 Participants will be given the FpVTE Evaluation Datasets after their equipment is set up.

The agencies that have provided Evaluation Datasets have done so with the restriction that the data shall not be retained in any way or form whatsoever by Participants. The FpVTE 2003 Evaluation Datasets will be protected under the Freedom of Information Act (5 U.S.C 552) and the Privacy Act (5 U.S.C. 552a) to the extent permitted by law, and will bear the legend "Notice: May contain Privacy Act or FOIA Protected Information."

The FpVTE Evaluation Datasets will be provided on a CD for SST and MST, and on a Universal Serial Bus (USB) Hard Drive for LST. (If needed, Participants may request SST or MST Evaluation Datasets on USB.)

The USB drive will be an IDE hard drive, formatted either NTFS or EXT3 (as requested by the Participant), with one partition, in an external enclosure connected to host computer via a USB port (compatible with both USB 1.0 and USB 2.0). If required, a separate computer (provided by the Participant) may be used to facilitate transfer to and from the USB hard drive. This computer would be considered part of the overall system, and would have its drives expunged at the conclusion of the test.

Conduct of Test

Testing activities will be recorded using video cameras for documentation of the evaluations. Footage from this documentation will not be made available to the public without review and comment from any participant that is named in the video.

Systems being evaluated in FpVTE 2003 shall not be accessible from outside the room in which the evaluation is being conducted. Modem, Internet, or wireless access is expressly prohibited. After the Evaluation Datasets have been given to Participants, all removable media (such as CDs, DVDs, Zip disks, Jaz drives, USB memory sticks, etc.) and all devices connected to the system (such as additional computers, laptops, PDAs or other handheld devices, etc.) are considered part of the system, and shall not leave the room without express Government approval. Offenders will be subject to criminal penalties.

The FpVTE Evaluation is designed to test systems running continuously (24 hours per day) over the test period with no substantial user/operator intervention. Participants shall have very limited contact with their systems during the test: three minutes of supervised and videotaped direct operator access per hour, during normal work hours Monday through Friday, will be permitted for system administration. Greater interaction with the systems during the test will only be permitted for system administration reasons by the express permission of the FpVTE Lead Test Agent, with the following restrictions:

- A written explanation for the need of system administration (such as a system crash) will be signed by the Participant and the FpVTE Lead Test Agent;
- The explanation and the amount of time required will be included in the FpVTE 2003 Final Report;
- All activity will be supervised by and explained to an FpVTE Test Agent;
- All activity will be videotaped.

Failure to complete the test during the allotted time will be noted in the FpVTE 2003 Final Report. FpVTE personnel will decide, after conclusion of the test, if the partial test results will be reported in the FpVTE 2003 Final Report.

Post-Test Procedures

At the completion of the Evaluation, the Participants will transfer all required output files from their system to the storage medium used for the Evaluation Datasets. SST and MST Participants who received CDs will be required to burn a CD with their output files. The output CD and the Evaluation Dataset CD will be returned to the Government. LST and other Participants who received Evaluation Datasets on a USB will transfer all the required output files to the original USB hard drive. The USB drive will then be returned to the Government.

FpVTE 2003 Evaluation Datasets and data derived from the datasets shall not be retained in any way or form whatsoever by Participants after completion of the evaluation. FpVTE 2003 Evaluation Datasets and data derived from the datasets shall not be distributed, published, copied, or disseminated in any way or form whatsoever by Participants. Participants shall track all copies of the FpVTE 2003 Evaluation Datasets and return or destroy all copies at

the end of the test, prior to leaving NIST. Failure to observe the restrictions on use of the FpVTE 2003 Evaluation Datasets is a violation of Federal law. Offenders will be subject to criminal penalties.

The Government will assure that none of the FpVTE fingerprints, or data derived from the fingerprints, are still resident on the Participant's system after the completion of the test. Participants will allow the Government to inspect all disks and other storage media on the system to verify compliance. This inspection will involve, at a minimum, the Government deleting files generated during testing and wiping free space on all disk drives and other storage media. The Government may choose to remove all files or format all disks, including system disks. The Government may also choose to remove and destroy certain storage media that cannot effectively be expunged of data.

It is recommended that Participants use separate drives or drive partitions for working space, including database management system (DBMS) data. This is so the areas for the operating system (OS) and fingerprint algorithms are clearly separated from the areas for working space. At the completion of the test, the Participant, under supervision, will perform a low-level format on the working space partition of their hard drive. If the working space is not clearly separated by drive or partition, all drives will be formatted. The Government will inspect all disks on the system to verify compliance.