Performance evaluation of fingerprint open-set identification algorithms

Elham Tabassi
tabassi@nist.gov

Craig Watson
craig.watson@nist.gov

Gregory Fiumara
greg@nist.gov

Wayne Salamon
wayne.salamon@nist.gov

Patricia Flanagan
patricia.flanagan@nist.gov

Su Lan Cheng
su.cheng@nist.gov

National Institute of Standards and Technology
100 Bureau Drive, Gaithersburg, MD 20899-8940 USA

Abstract

*We report performance of one-to-many fingerprint identification algorithms using one, two, four, eight or ten fingers for recognition. Performance is quantified in terms of recognition accuracy (false positive and false negative identification rate), robustness of algorithms (failure to process images) and computational efficiency (time to execute and size of the generated templates). We measured performance on a large dataset of fingerprint images of operational quality using varying enrollment sizes. The main contributions of this paper are two fold: a) open-set identification performance measures and b) an assessment of core capability of current one-to-many fingerprint recognition algorithms. This is accomplished using a subset of algorithms reported in the Fingerprint Vendor Technology Evaluation (*FpVTE*2012) that show the range and bounds of performance seen in that evaluation.*

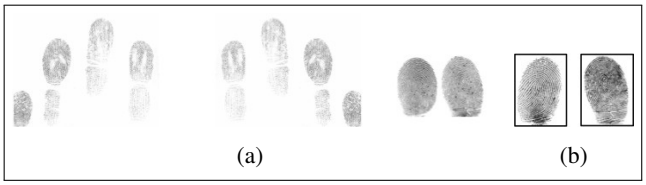
1. Introduction

Biometric identification is defined as the "process of searching against a biometric enrollment database to find and return the biometric reference identifier(s) attributable to a single individual [4]." Biometric identification is broadly categorized into closed-set and open-set identification.

Closed-set identification refers to cases where all searches have a corresponding enrolled mate in the biometric enrollment database. An example of a closed-set identification application is a cruise ship on which all passengers are enrolled. The outcome of a closed-set identification subsystem is a candidate list which contains the identity of one or more of the enrolled individuals whose enrolled samples are most similar to the search (query) sample. Ideally, the correct mate appears in the first rank. As such, the primary accuracy metric for closed-set identification is hit rate (or its complement, miss rate $= 1.0 -$ hit rate), which is the fraction of times the system returns the correct identity within the specified top ranks.

Closed-set identification has very limited applications because in the majority of real-world applications not all the individuals are or can be enrolled. Most real-world biometric identification applications are open-set, where some searches do not have corresponding enrolled mates in the biometric enrollment database. One example of an open-set identification application is searches against a watch-list. The expected outcome of an open-set identification subsystem is a candidate list of L closest (or most similar) enrolled identities when the search sample is from an enrolled individual, or a signal that the search sample is from an individual not in the biometric enrollment database. Therefore, the primary accuracy metrics for an open-set identification are false positive identification (false alarm or Type I error) rate and false negative identification (miss or Type II error) rate. These metrics are described in section 4.

Biometric performance evaluation is an important task, particularly for fingerprint identification, given its widespread applications. Large-scale evaluations of core accuracy and functionality of biometric recognition algorithms using operational data will not only reveal the capabilities of the current state of the art but can also identify the limitations and gaps of the current algorithms. The former sets realistic operational expectations and the latter could direct future research to improve and enhance the current status. The National Institute of Standards and Technology (NIST) in the United States [8] and the University of Bologna in Italy [2] run ongoing evaluations of one-to-one fingerprint recognition algorithms. Except for the one-to-many test of fingerprint recognition technologies that the National Institute of Standards and Technology ran in 2003 [13], and the Unique Identification Authority of India's re-



(a) (b)

Figure 1: Example of multi-finger and single-finger plain: (a) left slap (4L) + right slap (4R) and thumbs, (b) left index (1L) + right index (1R)

port on its enrollment proof of concept [11], the authors are not aware of any recent evaluations of one-to-many fingerprint identification.

The purpose of this paper is to review metrics for a performance evaluation of open-set identification algorithms and apply them to report performance of fingerprint recognition algorithms that participated in NIST FpVTE2012 [7], using different numbers of fingers (one, two, four, eight or ten). While recognition error rates are the most important and widely reported performance metrics, computational resources required by algorithms are a significant aspect of performance, especially for large-scale operations. To that end, we report the computation time, storage requirements and their accuracy tradeoffs for each of the algorithms under test.

The rest of the paper is organized as follows: section 2 overviews the data used for the evaluation, section 3 explains the method, section 4 details the metrics we used followed by results and discussion in section 5. We summarize and conclude in section 6.

Providers of the algorithms under test remain unnamed, and algorithms are identified by roman numerals.

2. Data

The FpVTE2012 evaluation dataset is a representation of fingerprints collected from operational biometric deployments. It is a larger sample of data used in previous NIST evaluations, such as PFT [1], MINEX [3] and FpVTE 2003 [13].

The evaluation dataset is comprised of several fingers and fingerprint impression types, including single-finger and multi-finger plains. While FpVTE2012 evaluated roll impression, we do not report performance of roll fingerprints in this paper.

Single finger plains are individual captures of the left and right index fingers on a single finger capture device. Examples of these image types are shown in Figure 1. The single-finger capture and identification flat fingerprint images (all live-scan) were provided by the Department of Homeland Security (DHS).

Multi-finger plains consist of capturing the four left fingers, the four right fingers and the capture of the two thumbs in a single transaction such that four fingers (or slap) of the right hand, the left hand and the two thumbs are captured in three separate images. Multi-finger plain images are passed to the algorithms under test as a single image — in other words, we did not segment multiple finger plain images into individual fingerprints prior to passing them to the algorithms. Whether or not algorithms under test performed any segmentation prior to search was not apparent to us.

The multi-finger plain (slap) fingerprint images are 70% live-scan and 30% re-scanned ink. These images came from law enforcement and other government agencies. The live-scan data was captured between 2000–2008, and the re-scanned ink are much older.

Table 1 lists the number of individuals in our evaluation dataset whose fingerprint images are used for enrollment and for search sets.

Evaluation images are of varying sizes, but all are 8-bit grayscale at 500 pixels per inch resolution. Images were originally compressed using Wavelet Scalar Quantization (WSQ) compression.

We used the NIST Fingerprint Image Quality (NFIQ) algorithm [9, 10] to assess the diversity of the quality of evaluation images. NFIQ scores are in the range [1,5], where NFIQ $= 1$ means highest quality and NFIQ $= 5$ means lowest quality. Distribution of NFIQ scores of evaluation data is shown in Figure 3.

To ensure data integrity and ground truth, we visually inspected dubious results, corrected ground truth errors and removed blank or non-fingerprint images from our evaluation set. Ground truth of identities could be erroneous in two ways: the same individual is in the database under two different identifiers or two different individuals are in the database using the same identifier. The first category wrongly elevates the false positives. We resolved this issue by vertically flipping fingerprint images used for nonmated searches. To correct topology, left/right labelings are reversed after flipping. An example is shown in Figure 2. The second category of ground truth errors would wrongly elevate false negatives. This was mitigated by visual inspection of most of the low scoring mates. Additionally, cases with more than one strong hit on a candidate list were manually inspected to resolve duplicated enrollments.

3. Experiments

A one-to-many identification has two main phases, enrollment and search or identification. We employed a two-step enrollment: first, all the templates used in the enrollment set are extracted from the fingerprint images, and second, a process is run that finalizes the enrollment templates used during identification.

Search or identification consists of generating templates

Enrolled Size	Search Size	Num Fingers	Finger Positions	Capture Type
100 000	30 000	1	Left Index (1L)	Single
100 000	30 000	1	Right Index (1R)	Single
1 600 000	30 000	2	Left+Right Index (2LR)	Single
3 000 000	30 000	4	Left Slap (4L)	Multi
3 000 000	30 000	4	Right Slap (4R)	Multi
3 000 000	30 000	8	Left+Right Slap (8LR)	Multi
3 000 000	30 000	10	L/R Slap, Thumbs (10LR)	Multi

Table 1: Evaluation dataset. Number of images for enrolled and search sets. The search set in all cases is a disjoint set of 10 000 images for mated searches (where there is an enrolled template for the search image) and 20 000 images for nonmated searches (where there is no enrolled template for the search image). Different enrollment sizes are due to limitation on data availability — we opted for the largest possible enrollment size for each case of single or multi-finger capture types.

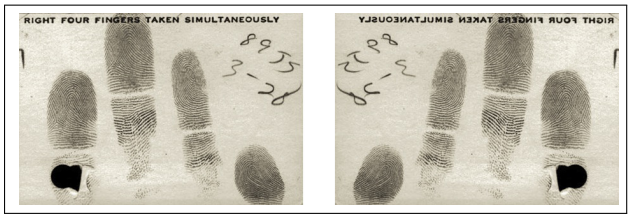


Figure 2: Example of flipping an image. The image on the left is the original image of a right slap and its reflection across the y-axis is shown on the right. We flipped images to resolve ground truth error cases where the same individual is in the evaluation set under two different identities. Right/left labelings are reversed for the flipped images (*i.e.*, the image on the right is considered a left slap).

from search (query) images and performing a one-to-many search against the enrollment set. We performed two types of searches on two disjoint subsets of test images: mated and nonmated searches. Mated searches are those where the search sample has an enrolled mate in the enrollment dataset. Mated searches are used to compute false negative identification or miss rates. Nonmated searches are those for which no enrolled mate exists. Nonmated searches are used to compute false positive identification or false alarm rates. The nonmated searches are necessary to preclude any gaming possibility.

The FpVTE2012 testing protocol [7] required recognition algorithms to produce, for each search, a ranked list of $L \leq 100$ most similar candidates along with the corresponding comparison scores. Rankings were to be based on

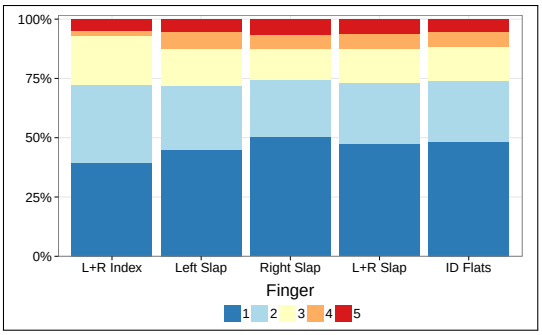


Figure 3: Distribution of NFIQ scores for the evaluation data. NFIQ $=$ 1 means highest quality and NFIQ $=$ 5 means lows quality.

the comparison scores, in descending order (*e.g.*, rank one means the most similar or most probable match). As shown in Table 1, 10 000 images were used in mated searches and 20 000 for nonmated searches. Therefore, algorithms are required to produce 10 000 candidate lists for mated searches and 20 000 for nonmated searches. We treated failure to produce a candidate list as a miss — this treatment (correctly) increases false negative identification rate but (slightly) decreases false positive identification rate.

All accuracy measurements and reporting are based on the candidate lists produced by the algorithms. Additionally, we measured and reported template generation time, search time and template sizes as an indication of computational resources needed by the algorithms.

Enrollment sizes are shown in Table 1. We aimed for the largest possible enrollment sizes. Due to limitations on data availability, we opted for different enrollment sizes for each case of single or multi-finger capture types.

FpVTE2012 tested implementations from eighteen providers, but this paper reports performance of six (three top performers and three others) due to space limitation. Thus we keep the identity of algorithm providers (participants) anonymous. Participants are labeled with roman numerals. The evaluation considers an algorithm as a "black box" in order to protect the provider's intellectual property. No information on the internal workings of the algorithms, (*e.g.*, indexing, pre-selection techniques) were provided.

4. Performance metrics

Details on our quantification of accuracy, efficiency and robustness of algorithms follows.

4.1. Accuracy

Open-set identification algorithms can make two types of recognition error: a) search of a biometric sample of an individual *not* enrolled in the biometric enrollment database

(a nonmated search) returns the biometric reference identifier(s) attributable to one or more enrolled person (Type I or false alarm because it returns a false identity); or b) search of a biometric sample of an enrolled individual (a mated search) returns the *incorrect* enrolled identity (Type II or miss because it misses the correct identity).

Occasionally, recognition algorithms fail to process biometric samples to generate templates or fail to execute one-to-many searches to produce comparison scores. The result is that a valid candidate list is not produced. Such failures might be voluntary (*e.g.*, refusal to process a poor quality image) or involuntary (*e.g.*, software crashes). Either way, it is an undesirable behavior and should be included in computation of recognition errors particularly to allow for fair comparison of algorithms. We treated such failure cases as a miss and added them to the Type II errors.

To summarize, we quantified the accuracy of the open-set identification algorithms as follows:

- False positive identification rate (FPIR) or Type I error rate is the fraction of the nonmated searches where one or more enrolled identities are returned at or above threshold, T. FPIR is a function of: the size of the enrollment database, N, length of candidate lists, L and score threshold, T.

- False negative identification rate (FNIR) or Type II error rate is the fraction of the mated searches where the enrolled mate is outside the top R ranks or comparison score is below threshold, T. FNIR is a function of: the size of the enrollment database, N, length of candidate lists, L, score threshold, T and the number of top candidates being considered, R.

Note that FNIR computation does not care about the cause of a miss: failure to correctly identify a sample (due to poor quality), failure to extract a template, failure to generate a comparison score or software crashes are all dealt with similarly.

The terms "hit rate," "reliability" and "sensitivity" that have been mentioned in some literature on automated fingerprint identification systems (AFIS) [5, 12] are just the complement of FNIR, computed as $1.0-$FNIR .

As it is conventional, we plotted detection error tradeoff curves (DET) [6]. In a DET curve, Type I and II error rates are plotted on both axes, giving uniform treatment to both types of error. Both axes use log scale, which spreads out the plot and better distinguishes different well-performing systems.

Another widely used accuracy metric is cumulative match characteristic (CMC), which is the fraction of the mated searches where the enrolled mate is at rank R or better — regardless of its comparison score. CMC is a special case of FNIR, or more precisely, hit rate, when the constraint on threshold is removed: CMC(N, L, R) = 1 − FNIR(N, L, T=0, R).

Rank-one hit rate CMC(N, L, R=1) is the most common accuracy metric reported in academic and AFIS-related literatures. While we report CMC for the tested algorithms, we believe it is an inadequate accuracy metric. Because similarity scores are ignored, CMC makes strong or weak hits indistinguishable. Additionally, with CMC, Type I errors remain unreported.

4.2. Failure to extract a template

Failure to extract is the fraction of images for which a template is not generated. Template generation can fail for the enrollment sample or the search sample. In both cases, failure to extract a template is included as a miss in the computation of FNIR (see section 4.1).

4.3. Computational efficiency

Another aspect of performance is the computational resources required by an algorithm. We report template generation time, one-to-many search time and template sizes for the algorithms we tested along with their tradeoff between accuracy and the required computational resources. Timing computations were performed on a Dell M610 with two Intel X5690 3.47 GHz processors and 192 GB of RAM. The algorithm under the test was the only process being run on the machine (except for a minimal set of required Linux daemons).

5. Results

This section reports the results of comparative analysis of six fingerprint identification algorithms using the performance metrics described in section 4. FpVTE2012 tested implementations from eighteen providers, but we report results for the top three performers (I, IV, VI) and another three (II, III, V) selected to exhibit the range and bounds of performance of current one-to-many fingerprint identifications. The main reason is space limitation, but also that this selection provides sufficient information for study and examination of open-set fingerprint identification performance metrics and evaluation results. Algorithms are labeled with roman numerals: I, II, III, IV, V, and VI. The rest of this section details performance reporting of these six algorithms. Table 2 summarizes the results.

5.1. Detection error tradeoff

Figure 4 shows detection error tradeoff (DET) curves. X-axis and y-axis show, in log scale, false negative identification and false positive identification rates respectively. The lower curves demonstrate lower FNIR, and the bottom left corner displays the lowest error rates. In general, flat DET curves are desired as they indicate a more stable impostor

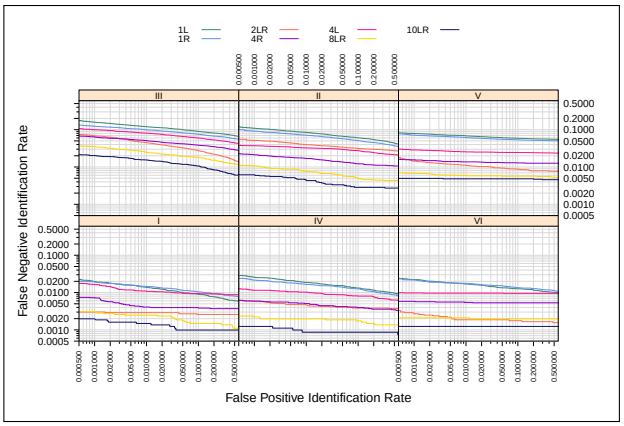


Figure 4: Detection error tradeoff (DET) curves. Panels show results for each algorithm (identified by a roman numeral on the top of the panel). In each panel, there are seven DET curves for the different number of fingers tested as shown in the figure legend: one finger searches comprise of left index (1L) or right index (1R); two finger searches used left and right index (2LR); four finger searches include left slap (4L) or right slap (4R); eight finger searches used left and right slaps (8LR); and finally ten finger searches used all ten fingers in left/right slaps and the two thumbs (10LR). The size of the enrollment set (N) is: for one finger identification (1L and 1R) $N = 100\,000$, for two finger identification (2LR) $N = 1\,600\,000$, for four (4L, 4R), eight (8LR) and ten finger (10LR) identification searches $N = 3\,000\,000$. The performance of the most accurate participants are shown in the bottom row.

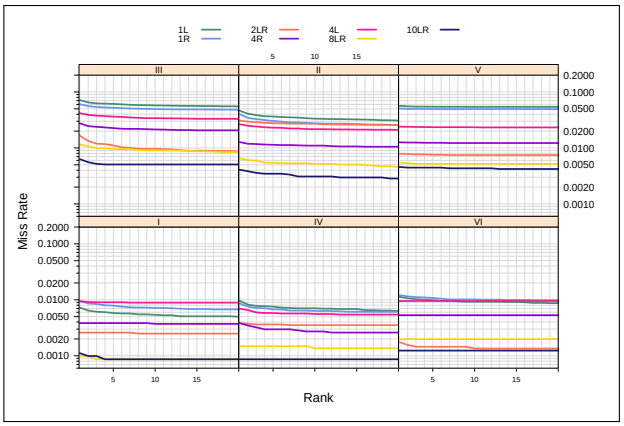


Figure 5: Cumulative match characteristic (CMC). Panels show results for each algorithm (identified by a roman numeral on the top of the panel). In each panel, there are seven CMC curves for the different number of fingers tested as shown in the figure legend: one finger searches comprise of left index (1L) or right index (1R); two finger searches used left and right index (2LR); four finger searches include left slap (4L) or right slap (4R); eight finger searches used left and right slaps (8LR); and finally ten finger searches used all ten fingers in left/right slaps and the two thumbs (10LR). The size of the enrollment set (N) is: for one finger identification (1L and 1R) $N = 100\,000$, for two finger identification (2LR) $N = 1\,600\,000$, for four (4L, 4R), eight (8LR) and ten finger (10LR) identification searches $N = 3\,000\,000$. The performance of the most accurate participants are shown in the bottom row.

distribution, meaning change of operating threshold would have little or no effect on FPIR. Top performers are I, IV and VI. For all the algorithms, the lowest error rates are achieved using ten finger (10LR), followed by eight finger searches (8LR). This is expected, as more information is passed to the algorithms. Right hand (1R or 4R) performs (sometimes just slightly) better than left hand (1L or 4L). This might be explained by usability issues such as handedness, and we plan to study this later.

An interesting observation is that for all algorithms except II, two finger (2LR) searches perform better than four finger (4L or 4R). We will further investigate its cause. To the best of our knowledge, the better performance of the two finger searches can be explained by a) correlation of the quality of fingers in the same hand and b) segmentation error. As previously mentioned, four finger searches are performed on a slap image (Figure 2). The algorithms under test had to first segment the image and then perform the search. Therefore, any error in the segmentation process would propagate through the whole search process.

5.2. Cumulative match characteristic

Figure 5 shows CMC or miss rate as a function of rank for the six algorithms tested. Note that almost always, the correct mate is identified at ranks 1 or 2. This has implications for human adjudication — it is sufficient to examine only the first five candidates in a candidate list.

Similar to results in section 5.1, algorithms I, IV and VI produce the lowest error rates. For all algorithms, ten and eight finger searches are the most accurate (except VI, where two finger outperforms eight finger searches) and right hand yields better accuracy than left hand.

5.3. Template size

Size of the templates is an important parameter in design of a biometric identification system for two reasons: a) the size of the storage required for the enrolled templates is considered a cost to the overall system; and b) for mobile applications, network bandwidth places a constraint on the size of search templates. Figure 6 shows the tradeoff of accuracy versus median search template size. While we report

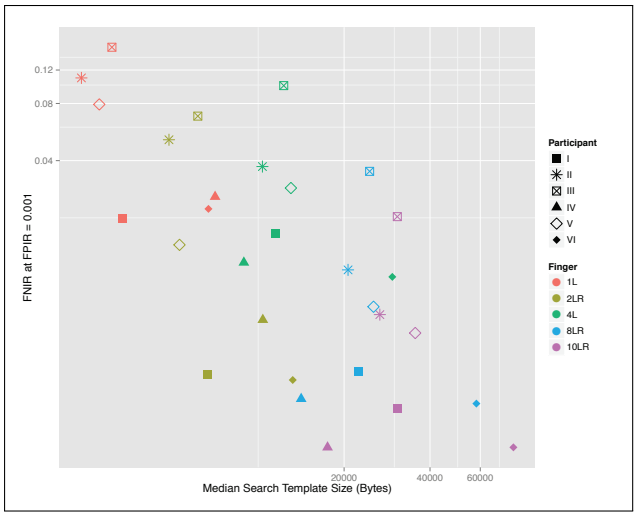


Figure 6: False negative identification rate versus template size in bytes. Symbols represent algorithms and colors represent different fingers. Template size is quantified by median of search templates sizes. FNIR is computed at the score threshold that gives FPIR $= 0.001$.

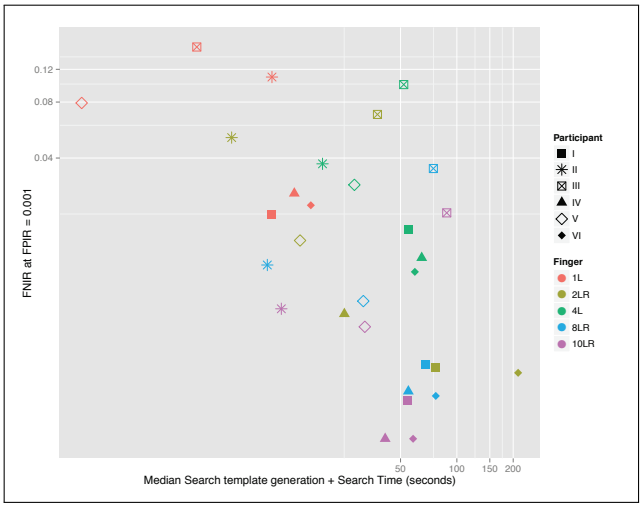


Figure 7: False negative identification rate versus computation time in seconds. Symbols represent algorithms and colors represent different fingers. Computation time is quantified as the median of the sum of search template generation times and search times. FNIR is computed at the score threshold that gives FPIR $= 0.001$.

search template size here, we note that enrollment templates are of similar sizes. Accuracy is quantified by FNIR at the score threshold that gives FPIR $= 0.001$.

Not surprisingly, template sizes correlate with the number of fingers — for each algorithm, templates generated from one finger are the smallest in size and templates generated from ten fingers are the largest in size.

Generally speaking, the larger template sizes provide better accuracy. This result is somewhat expected since it is assumed that the size of a template correlates with the amount of information encoded in the template. The exception is algorithm IV, which has lowest template sizes and lowest FNIR for four, eight and ten fingers (although algorithm IV has the largest one finger templates). Algorithms VI and IV have comparable accuracy, but VI has larger templates by a factor of 4. In fact, algorithm VI produces the largest templates for (almost) any number of fingers used for the search. Algorithm II is the least costly algorithm among the test algorithms, but its accuracy is among the worst.

5.4. Timing Statistics

Figure 7 shows the tradeoff of accuracy versus computation time. We quantified computation time as the median of total computation time. Total computation time is measured as the sum of search template generation time and one-to-many search time. This is motivated by real-world operational scenarios that give enrollment more relaxed time constraints than the search phase. As before, accuracy is quantified by FNIR at the score threshold that gives FPIR $= 0.001$.

One finger searches are the fastest and least accurate. Ten finger searches take slightly less time than four or eight finger searches. Computation time of four and eight fingers are comparable, but their error rates differ significantly — eight finger searches produce roughly twice as many false negatives as four finger searches.

Another interesting observation is the relatively tight clustering of the three top performers (I, IV and VI) for one, four, eight and ten fingers searches.

6. Summary and Conclusion

We reviewed performance metrics for evaluation of core capability of open-set one-to-many fingerprint recognition algorithms. We considered and reported algorithm efficiency and robustness along with accuracy metrics. Algorithm efficiency is measured by computational resources required by the algorithm. Algorithm robustness is quantified by failure to extract templates or failure to compute comparison score. Employing these metrics, we reported performance of six open-set one-to-many fingerprint recognition algorithms who participated in NIST FpVTE2012. We showed that lowest error rates are achieved using ten fingers followed by eight fingers (right and left slaps). However, both of these cases require higher computational resources than one or two finger searches.

Using an enrollment size of three million individuals, the

Enrollment Size	Search Size	Finger(s)	Participant	FNIR @ FPIR = 0.001	Median Template Size (bytes)	Mean Template Size (bytes)	Computation Time (seconds)
100 000	30 000	1L	I	0.0198	3 343	3 368	10.2091
1 600 000	30 000	2LR	I	0.0030	6 649	6 692	77.3341
3 000 000	30 000	4L	I	0.0165	11 502	11 457	54.9344
3 000 000	30 000	8LR	I	0.0031	22 538	22 480	68.0086
3 000 000	30 000	10LR	I	0.0020	30 725	30 758	54.522
100 000	30 000	1L	IV	0.0258	7 063	7 042	13.5339
1 600 000	30 000	2LR	IV	0.0058	10 380	10 383	25.0500
3 000 000	30 000	4L	IV	0.0116	8 908	8 885	64.8854
3 000 000	30 000	8LR	IV	0.0022	14 144	14 038	55.0444
3 000 000	30 000	10LR	IV	0.0012	17 497	17 372	41.3894
100 000	30 000	1L	VI	0.0222	6 694	6 750	16.5853
1 600 000	30 000	2LR	VI	0.0028	13 215	13 322	212.5220
3 000 000	30 000	4L	VI	0.0098	29 509	29 506	59.6385
3 000 000	30 000	8LR	VI	0.0021	58 147	58 220	77.1740
3 000 000	30 000	10LR	VI	0.0012	78 387	78 576	58.3660
100 000	30 000	1L	II	0.1090	2 402	2 546	10.2799
1 600 000	30 000	2LR	II	0.0515	4 868	5 072	6.2604
3 000 000	30 000	4L	II	0.0372	10 347	10 714	19.1478
3 000 000	30 000	8LR	II	0.0106	20 644	21 202	9.7079
3 000 000	30 000	10LR	II	0.0062	26 684	27 346	11.5341
100 000	30 000	1L	V	0.0791	2 775	2 816	0.9892
1 600 000	30 000	2LR	V	0.0144	5 301	5 372	14.5267
3 000 000	30 000	4L	V	0.0287	13 023	13 071	28.3454
3 000 000	30 000	8LR	V	0.0068	25 359	25 460	31.6330
3 000 000	30 000	10LR	V	0.0049	35 497	35 540	32.2278
100 000	30 000	1L	III	0.1582	3 072	3 068	4.0766
1 600 000	30 000	2LR	III	0.0683	6 144	6 137	37.6436
3 000 000	30 000	4L	III	0.0993	12 288	12 139	52.0084
3 000 000	30 000	8LR	III	0.0351	24 576	24 299	74.8950
3 000 000	30 000	10LR	III	0.0202	30 720	30 262	88.3038

Table 2: Tabulation of results presented in this paper. The best performance in each category is shaded in green and the worst in pink. Reported template sizes are the median and mean of search templates sizes. Reported computation times are the median of the sum of search template generation times and one-to-many search times. FNIR is computed at the score threshold that gives FPIR $= 0.001$.

lowest error rates are FNIR $= 0.0012$ for ten fingers and FNIR $= 0.0021$ for eight fingers (both computed at FPIR $= 0.001$). The most accurate algorithm (VI) is also the most expensive both in terms of template size and computation time.

Two finger searches (left and right index) outperformed four finger (right or left slap) searches. We will further investigate the cause, but for now, we speculate that it is due to either segmentation errors or correlation of fingers in the same hand. Current and future works include examination of the effect of enrollment size, quality, time lapse between the enrollment and search images on performance, and investigating the cause/source of effect of handedness on accuracy.

Disclaimer: Certain commercial equipment, instruments, or materials are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

References

[1] S. L. Cheng, G. Fiumara, and C. Watson. PFT II: Plain and Rolled Fingerprint Matching with Proprietary Templates. *NIST Interagency Report 7821*, 2011. 2

[2] B. Dorizzi, R. Cappelli, M. Ferrara, D. Maio, D. Maltoni, N. Houmani, S. Garcia-Salicetti, and A. Mayoue. Fingerprint and On-Line Signature Verification Competitions at ICB 2009. In *Proceedings International Conference on Biometrics (ICB), Alghero, Italy*, pages 725–732, June 2009. 1

[3] P. Grother, M. McCabe, C. Watson, M. Indovina, W. Salamon, P. Flanagan, E. Tabassi, E. Newton, and C. Wilson. Performance and Interoperability of the INCITS 378 Fingerprint Template. *NIST Interagency Report 7296*, 2006. 2

[4] ISO/IEC SubCommittee 37 Working Group 1. ISO/IEC 2382-37 Harmonized Biometric Vocabulary, 2012. 1

[5] P. Komarinski. *Automated Fingerprint Identification Systems (AFIS)*. Elsevier, 1st edition, 2004. ISBN 9780080475981. 4

[6] A. F. Martin, G. R. Doddington, T. Kamm, M. Ordowski, and M. A. Przybocki. The DET curve in assessment of detection task performance. In *EUROSPEECH*, 1997. 4

[7] National Institute of Standards and Technology. Fingerprint Vendor Technology Evaluation. nist.gov/itl/iad/ig/fpvte2012.cfm. [Accessed: Jul 15, 2014]. 2, 3

[8] National Institute of Standards and Technology. Image Group Fingerprint Overview. nist.gov/itl/iad/ig/fingerprint.cfm. [Accessed: Jul 15, 2014]. 1

[9] E. Tabassi and C. Wilson. A novel approach to fingerprint image quality. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, volume 2, pages II–37–40, Sept 2005. 2

[10] E. Tabassi, C. L. Wilson, and C. I. Watson. Fingerprint Image Quality. *NIST Interagency Report 7151*, 2004. 2

[11] Unique Identification Authority of India. The UID Enrolment Proof-of-Concept Report. uidai.gov.in/images/FrontPageUpdates/uid_enrolment_poc_report.pdf. [Accessed: Jul 15, 2014]. 2

[12] J. L. Wayman, A. K. Jain, D. Maltoni, and D. Mai. *Biometric Systems: Technology, Design and Performance Evaluation*. Springer, 2005. ISBN 1852335963. 4

[13] C. Wilson, R. A. Hicklin, M. Bone, H. Korves, P. Grother, B. Ulery, R. Micheals, M. Zoepfl, S. Otto, and C. Watson. Fingerprint Vendor Technology Evaluation 2003. *NIST Interagency Report 7123*, 2004. 1, 2