FINGERPRINT VENDOR TECHNOLOGY EVALUATION 2003

APPENDIX F

PARTICIPANT RESPONSES

A draft of the FpVTE 2003 Analysis Report was made available to the Participants on April 26, 2004. At that time, Participants were invited to provide comments, with the following instructions:

> *The FpVTE 2003 Analysis Report and five Appendices are now available for review by Participants. These have* **not** *been approved for public release.*
>
> *Participants will have 2 weeks to provide comments and feedback to NIST. Comments must be received by 12:00 Noon EDT, Monday 10 May 2004. Comments may be no longer than two pages, and must be in MS Word format. The Comments will be included in an appendix to the final report.*
>
> *As a reminder, these are the sections of the Application to Participate in FpVTE 2003 that are relevant to the release of results and participant responses:*
>
> *7. Release of Evaluation Results*
>
> *7.1. After the completion of the evaluations, the Government will combine all results into a Final Report. The FpVTE 2003 Final Report will contain, at a minimum, descriptive information concerning FpVTE 2003, descriptions of each experiment, evaluation results, and each Participant's five-page system description document.*
>
> *7.2. A pre-release version of the FpVTE 2003 Final Report will be made available to Participants. Participants will be invited to provide comments which will be included as an appendix to the FpVTE 2003 Final Report. More specific guidance concerning the report and Participant comments will be provided at a later date.*
>
> *7.3. Participants shall not comment publicly on the pre-release version of the FpVTE 2003 Final Report until it has been released to the public.*
>
> *7.4. After the release of the FpVTE 2003 Final Report, Participants may decide to use results of these evaluations for their own purposes. Such results shall be accompanied by the following phrase: "Results shown from the Fingerprint Vendor Technology Evaluation 2003 do not constitute endorsement of any particular system by the Government." Such results shall also be accompanied by an Internet hyperlink (URL) to the FpVTE 2003 Final Report on the FpVTE 2003 website.*
>
> *8. Additional Information*
>
> *8.1. Any data obtained during these evaluations, as well as any documentation required by the Government from the participants, becomes the property of the Government. Participants will not possess a proprietary interest in the data and/or submitted documentation.*
>
> *8.2. With the signing of this form, Tentative Participants and Participants attest that they will not file any FpVTE-related claim against FpVTE 2003 Sponsors, Supporters, staff, contractors, or agency of the U.S. Government, or otherwise seek compensation for any equipment, materials, supplies, information, travel, labor and/or other participant provided services.*
>
> *8.3. The Government is not bound or obligated to follow any recommendations that may be submitted by the Participant. The United States Government, or any individual agency, is not bound, nor is it obligated, in any way to give any special consideration to FpVTE 2003 Participants on future contracts.*

This Appendix includes those participant responses, in alphabetical order by company name.

Contents

General Comments 4
123 ID 6
Avalon 8
Bioscrypt 10
Cogent 13
Dermalog 15
Griaule 17
Identix 19
Motorola 20
NEC 22
Neurotechnologija 24
Phoenix Group 26
SAGEM 28
Technoimagia 31

The following Participants did not provide responses to the FpVTE Draft Analysis Report:

- Antheus
- Biolink
- Golden Finger
- Raytheon
- UltraScan

General Comments

The FpVTE team would like to thank the Participants for their cooperation and all of their hard work in the evaluations.

While the FpVTE team has no intention of responding to every point brought up in the Participant Responses, several general comments can be made that apply across responses.

A) Several Participants stated that their systems were designed or optimized for different types of fingerprints than were used in FpVTE. Some systems were designed for images at resolutions other than 500 ppi, or were tuned for images from specific models of scanners. Several systems showed a particular sensitivity to poor-quality images. It should be noted that the measured accuracy of such systems might be higher if different sources or types of fingerprints had been included in FpVTE. The sources and types of fingerprints used in FpVTE were selected to represent a broad range of real-world operational government fingerprint data, in order to measure the capabilities of each system on existing data.

B) Some Participants expressed doubts about whether other Participants might have taken advantage of (or "gamed") the evaluation.

   1) NIST has corroborated the results for several of the Participants (particularly several accurate ones) in the SDK tests, which use a different testing protocol and different datasets. Participants who wish to have their systems independently retested should contact the FpVTE Liaison for information about the SDK tests.

   2) FpVTE 2003 was designed with a variety of measures to prevent gaming or make it more difficult. We believe these measures were effective.

   3) Analysis showed no evidence of gaming, and raised no suspicions.

   4) One Participant proposed methods of gaming the test. We believe that such methods would have been as likely to reduce accuracy as improve it. The consistent performance of the most accurate systems, and the corroborated provided through the SDK tests, gives us high confidence in the test results.

C) Several comments addressed comparisons at data points other than FAR=10^{-4}. The results would have differed little if at all had the systems been compared at lower values of FAR: the ranks of the 4 most accurate systems would be the same for 10^{-6} as 10^{-4} for LST or MST. The LST and MST charts in Appendix D have been updated to include lower values of FAR.

D) Some comments dealt with misunderstandings of how the data partitions used to compare systems were selected. The text in Appendix E, Section 2.2 was updated to clarify this process.

E) Normalization was not performed during FpVTE analysis. Operational fingerprint systems ***do*** use normalization, both for verification and identification, so permitting normalized results does ***not*** move away from real-world applications.

F) In the histograms in Appendix C, in some cases where the mate (red) and non-mate (green) bars are superimposed they are difficult to differentiate. Please see (for example) the NIST VTB histograms for comparison (Figure C-74).

G) Images that were listed as "F" quality in MST or SST were considered poor by every system that submitted quality/FTE information, ***or*** were noted as unusually poor quality by a fingerprint examiner. F Quality did not necessarily mean that the fingerprints were unusable or unmatchable.

A limited number of changes were made to the Analysis Report and appendices after the Draft Report was reviewed by the Participants. Some of these changes were in response to the Participant Responses in this Appendix.

- Analysis Report
  - Parts of the Abstract, Executive Summary, and Conclusions were reworded.
  - Text was reworded in a few other locations for clarification.
  - Figure 12 (Effect of Image Quality (MST)) was updated, with only trivial differences.
- Appendix C
  - Text was added or reworded in a few locations for clarification.
- Appendix D
  - Table D-2 included some numbers that misleadingly rounded to 1.00; the number of digits has been increased.
  - The LST and MST results include data to show that results do not substantially differ at FARs lower than 10^{-4}.
  - A table of Equal Error Rates was added for MST
- Appendix E
  - Section 2.2 has been updated to clarify how the data partitions used to compare LST systems were selected.
  - A section entitled "Methods of Differentiation between Systems" was accidentally included in the 23 April draft that referred to an analysis methodology that was considered, but was ***not*** used in FpVTE. That section has been removed.

123 ID

123ID, Inc.

Comments about 123ID Participation in the NIST fpVTE2003

We want to thank NIST for the opportunity to measure our matching algorithm in the NIST (FBI) standard for the fist time.

Although our matching and filter/repair systems were created for live scanners in the civil (non-FBI) environment, we felt the necessity to rate our matching engine to the traditional AFIS-NIST standard.

Our processing and matching algorithms were created and tuned to be used with live fingerprint scanner images in a different ppi resolution. We had a very short amount of time available before the test to tune the algorithms such that to work with slapped and rolled fingerprints in 500 ppi resolution.

Even so the 123ID CVT fingerprint matching engine achieved a very good separation between customers and impostor scores as can be seen in the NIST detailed performance report. A clear and distinct separation that was not achieved by any of the top 3 engines in the competition. This is verified by the fact that in our deployments using live fingerprint scanners we achieve FA rates equal to zero and very low FR rates. See System Specific Results, Appendix C, Figure 4. 123ID (LST) histogram of Match vs Non-Match Distribution for BxA subsets, on page 10.

We believe that most of the errors were due to inconsistencies in dealing with slapped and rolled type print impressions and we feel confident that our rating will substantively improve in future NIST testing competitions.

This competition has served 123ID to confirm the discrimination superiority of a Vector based algorithm (CVT) and that the filtering and repairing technology needs to be designed specifically for the image source and resolution of operation.

Comments and requests about the report documents

1. LST (Overtime)
As you know the reason for exceeding the time allowed were two (2) hard disk failures at different times and the lose of processed data in such nodes. The lose of such processed data forced us to restart the process for the compromised matrixes. It would help to indicate these facts so that our system as a software solution is separated from hardware problems that anyone can have.[1]

2. Relationship of Accuracy to Score discrimination capability
We feel that measuring the separation between True accepts and False accepts are an important measure of accuracy and fundamental for projecting results to much larger data sets. We would like to request that the discrimination capability of each algorithm be added as an accuracy measure in the Detail Results. An algorithm that shows the most scoring separation will clearly

[1] Noted in Appendix C, Section 2.1

hold the claimed error rates when the solution climbs to massive volumes. This analysis will also help discern and separate the filtering capabilities of a system versus the discrimination and FR/FA separation capability of a system.

It is clear that each algorithm has a score cut-off explicitly defined or defaulted by the distribution and therefore this analysis is feasible. Perhaps a further breakdown of the discrimination capability as a function of Quality will ultimately separate the discrimination capability from the image repairing/filtering capability of a matching system.

Author: Roger Quint, Chief Technology Officer, 123ID, Inc.

AVALON

semantic system

Analysis of the FpVTE Report by the FpVTE Liaison

Explanation about the UltraMatch technology

UltraMatch is as total new approach for matching biometric data, such as fingerprints. Currently every matcher is fix programmed and connected with a dedicated encoder. UltraMatch has the ability to recognise generically any patter in any data format. The result is that UltraMatch can be connected to practically any encoder with any kind of biometric data. Of course, if these encoded and proprietary data have some special calibration besides the intrinsic fingerprint pattern, UltraMatch needs to be adjusted for these rules. Beside this, UltraMatch does not need any further calibration or changes. *UltraMatch works with a new Mathematics (Algorithm) with enables UltraMatch to detect and recognises any intrinsic information (i.e. fingerprint pattern) in any dataset.*

Scope of attendance for Avalon with the UltraMatch technology

Our first Scope was, to test whether UltraMatch*/one* is capable to work with the NIST ANSI T9 standard.

Our second scope was, to test whether UltraMatch*/one* fulfils the rules how the matching logistic must be fulfilled.

Our second scope was, to test which UltraMatch will reach by using any 3rd party encoder. In our case we used an encoder of Antheus, Brazil.

Conclusion of the results at the FP VTE Results

We are very much satisfied with the results described in the Report. Our tree scopes have been 100% reached. Unfortunately the ranking regarding the quality is not as good as it could be. However, the Quality represented is limited by the encoder and not by UltraMatch! Because we received encoded images we only can match as accurate as the quality of the encoded files will allow.

The match Quality of UltraMatch

Since UltraMatch*/one* works from already encoded data. Therefore the quality standard always is given by the quality of the templates.

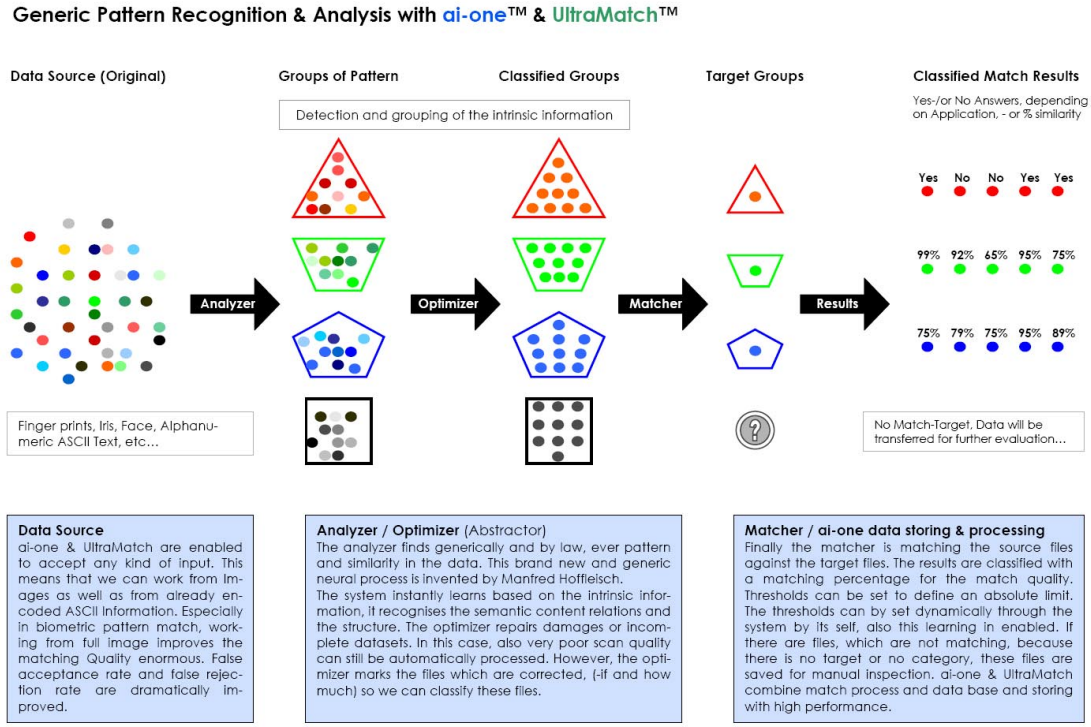
The speed of UltraMatch

During the MST UltraMatch/one finished the whole MST in less than 8 hors! The net matching time was 6h. 2 hours where used for all back office operations like import, encoding export of the result file.

Scope for UltraMatch*/one* & UltraMatch*/two*

As explained earlier in this document, UltraMatch is as total new approach for matching biometric data, such as fingerprints.

UltraMatch*/one* works with 3rd party Templates (NIST T9 & proprietary formats)
UltraMatch*/two* works directly with the original raster image.
This UltraMatch version has the image encoder directly embedded
(not tested at the FpVTE Contest, because this is a brand new version of UltraMatch technology).



Ultra fast back-office operation

In the today's world of Biometrics, we are over floated with various proprietary standards. It is also a fact, that the implementation of Front end systems has already reached a very high level of quality and penetration.

Interoperability and communication between the systems becomes as the most important killer factor for very large application projects. We appreciate very much the effort of NIST to create standards for templates and image information. However we believe, that it will take some time, to integrate NIST standards and exchange NIST standards in all the various internal processes of the proprietary systems worldwide. UltraMatch*/one* offers the solution straight forward and today! UltraMatch*/one* is enabled to be interfaces with all the proprietary formats in order to connect all front ends together. Furthermore UltraMatch*/one* also directly works with the NIST T9 standard. The 100Mio. matches, finished in 6 hours on a Dell Latitude Notebook talks a very clear language about the power of this match solution.

Final conclusion of UltraMatch in cooperation with the FpVTE Contest

The test has clearly proofed, that UltraMatch concept works perfect. Of course UltraMatch/one depends in quality of the 3rd party encoder. However we proofed that UltraMatch probably is the fastest and comprehensive format independent back office matcher currently on the marked.

For further Information please contact:
semantic system ag, Switzerland, info@semanticsystem.com +4143 8284533

BIOSCRYPT

NIST FpVTE results - Bioscrypt Response

The image databases selected for the FpVTE test were typical of AFIS background or criminal check applications (Analysis Report, Page 19, Table 6). These types of images are quite different from those captured under commercial conditions (cooperative users, small- or slide-sensors, daily usage etc.), for which Bioscrypt's algorithm has been optimized and demonstrated to exhibit exemplary performance. Despite this, Bioscrypt is pleased that our algorithm competed in the FpVTE tests as part of the NIST statutory mandate to certify biometric technologies that may be used in the U.S. VISIT program. In particular, we believe that the Bioscrypt algorithm offers a very suitable option for use in a multi-algorithm approach in a large-scale system such as the U.S. VISIT program: providing a complementary method to AFIS optimized algorithms; or for use in the verification applications which will likely be deployed using commercial equipment and conditions as described above. Furthermore, Bioscrypt is pleased that NIST chose to openly disclose all Participant's results, rather than offer Participants the opportunity to remove their affiliation following review of their results compared with the average of the top ten Participants as was done in the FVC 2004 test.

Bioscrypt has carefully reviewed the FpVTE Analysis Report and has identified the following areas where we respectfully request clarification:

1) As described above, the nature of the images used in this test tended to highlight the ranking of the participants, with the top AFIS vendors obtaining the best results. In fact, some of the vendors performed extraordinarily well; scoring 0% Errors on one of the most difficult databases (Analysis Report, Page 40, Table 19), specifically DHS2, which contained fingerprints deemed unusable by all seven image quality metrics or by a human examiner. In the detailed presentation of one example of these results, it was indicated (Appendix C, page 42, Section 7.3) that a Participant's scores for the DHS2 database for non-match cases were all zero. Why do these data points (i.e., as a spike at 0) not appear in the histogram in Figure 40 on page 45 of Appendix C, where the non-match scores appear to be distributed as "expected" between 0 and ~300)? Bioscrypt requests clarification of this apparent inconsistency.[2]

2) Given that the databases were gleaned from US Government sources, to which some Participants may have had legitimate prior access, were steps taken to ensure that no a priori information was stored on the Participant's systems prior to commencement of the tests?

3) We request that anti-gaming measures as indicated in the FAQ section of the FpVTE website are included in the Analysis Report, i.e., along with a clarification of the NIST's policy on this issue.

   (Extract from the FAQ section)
   It is suggested to prepare anti-gaming means to "statistically normalized, modified, or adjusted score" such as score comparison of 1:1 matching, e.g. asking participants to provide matching scores for randomly selected 1:1 pairings. These pairings are contained in LST and MST with different ID. If a participant provided a normalized score, this score differed from 1:1 matching score.
   A variety of anti-gaming measures are designed into the test.

[2] This was corrected for the final report.

We request this clarification as we believe that the test methodology, whereby each participant's system was treated as a black box receiving a database of N fingerprints and returning a NxN matrix of scores, can be exploited in manners that are unrelated to actual deployed systems for one-to-one or true one-to-many matching (where a comparison is achieved purely on the basis of a comparison between a candidate and a reference print). One example of a gaming method, using an unsupervised classification (clustering) algorithm, is described below:

> Consider that there are 4 impressions of the same fingerprint: A, B, C, D. The participants are not supposed to know that they all belong to one finger, but have to return 12 = 4x3 scores, that is AB, AC, AD, BA, BC, BD, CA, CB, CD, DA, DB, DC. The auto scores AA, BB, CC, and DD are not used. Let us assume that all the scores are low except AB, BC, and CD. It would result in 9 false rejections out of 12. However, if the scores AB, BC, and CD are high, the unsupervised classification algorithm will conclude that all the images A, B, C, D belong to the same finger, i.e. form a cluster of data. Thus, all the remaining combinations will be artificially assigned a high score, even though the actual scores are low. The modified False Rejection will be 0 out of 12.

Alternatively, if the unsupervised classification algorithm believes that the vast majority, if not all, of fingerprints have either one mate only or no mates, then the following step can be taken:

> Suppose that the score AB is high, AC is marginal, and BC is low. The classification algorithm concludes that AB is a true match while AC is a false accept (otherwise, there would be three mates A, B, and C). The score AC will be artificially set to a low value, thus modifying the False Acceptance rate.

Prior to commencement of the FpVTE test (in an email 1st October 2003), we had expressed concern that gaming methods using full knowledge of the gallery and probe image analysis, could be used to exploit the test methodology in a manner that is inconsistent with commercial systems, but we don't know if this issue was addressed and, indeed, it appears to be accepted in the report (Page 50, second last paragraph). Further, we had expected that the scores generated for the trivial data would be used as an anti-gaming measure, but this appears not to be the case (Appendix D, pages 31-32, Section 5.3) as some Participants had scores different for the same pairs of images in the trivial and the main data sets. Also, it would have been straightforward for a Participant to retain knowledge of the scores and other data obtained during the trivial and main comparisons, thus rendering this potential anti-gaming measure as irrelevant.

4) We note that in grouping the MST and SST results together in the Analysis Report ignores the fact that algorithms and systems provided for these components of the tests had different objectives. Specifically, our understanding was that the SST was designed for low cost commercial systems.

5) We would like to ask NIST to disclose more precise numbers of quality distribution in SST for DHS2 and DOS-BCC databases (Analysis Report, Page 40, Table 19, "Quality Distribution in SST"). This will help industry understand more clearly the relationship between the exact number of prints that were deemed unusable by all seven image quality metrics or by a human examiner, and the perfect performance of some algorithms on database DHS2.

Finally, please remove the following two sections:

(Analysis Report Page 67, last line)

"; the most accurate systems did not. For example, in SST Cogent had a true accept rate of 0.996 for BCC data and 1.000 for DHS2, at a false accept rate of 10^{-3}. Bioscrypt had a true accept rate of 0.972 for BCC data and 0.668 for DHS2".[3]

(Appendix C, Page 34, last line)

"Compare these results with Cogent's SST results in Figure 41."[4]

We believe that these statements are unnecessary and pejorative.

[3] Reworded.
[4] Removed.

COGENT

COGENT SYSTEMS

Cogent Comments on FpVTE 2003 Benchmark Results

Cogent Systems, Inc. requests that NIST update the final report to reflect the operational test points published in the FpVTE announcement. Cogent submitted three algorithms for the FpVTE 2003 benchmark tests: one each for the specified LST, MST, and SST categories. For each of these categories NIST defined the operational points that would be used as the basis for system evaluations. The Cogent algorithms were, therefore, configured as defined in *Section 5.3 Operational Points* of the FpVTE Test Overview (http://fpvte.nist.gov) in accordance with the public NIST announcement. While the Cogent algorithms performed exceptionally well in all three tests, the current report is inconsistent with NIST's previously specified operational points. [5]

As NIST stated in *Section 5.3*:

> *"For LST, the relationship between TAR and FAR will be analyzed, focusing solely on very low FAR values (down to about 10^{-8})."* – The current report address results at a much higher FAR, 10^{-4} and would imply a much higher rate of false positives than a FAR of 10^{-8}.

Cogent's Operational Points

An LST algorithm is utilized for applications that require identification of a subject from a **very large sized database** per the FpVTE test plan. Each transaction searches against the entire database collection with databases greater than 1 million records and upwards of tens of millions of records. These systems require algorithms and computing resource to provide a response time, which for applications in today's world, is in seconds. Accuracy for these systems is characterized by:

- Low False Rejection Rate (FRR) - Low rate of not finding a candidate in the database, also known as Type 1 error
- Very low False Acceptance Rate (FAR) - Very low rates of false hits

In practice, systems containing 1 million or more records would require a LST algorithm that would allow from 1 to 5 FAR error(s) for every 100 transaction. **A system would therefore be required to operate at a FAR of $0.5*10^{-7}$ through $10^{-8,}$ which is what was published as the objective of the test for the FPvVTE. However, the test report provides results at a much less stringent evaluation point of $10^{—4}$.** Cogent therefore built the LST test **to conform to the FpVTE guideline** whereby the LST algorithm evaluation would be **solely at a very low FAR**. We configured our system to operate at the required very low false acceptance rate (FAR=$0.5*10^{-7}$ through 10^{-8}) and to maximize system response time. This resulted in some accuracy loss (FRR) at a high false acceptance rate (i.e., FAR =$10^{—4}$). However, the algorithm was designed such that **this accuracy loss at a high FAR is eliminated for a very low FAR. For applications with large databases, achieving a very low false acceptance rate is a critical measure.** The Cogent algorithms performed exceedingly well in this regard and **the NIST report should be updated to reflect how the other vendors performed in the same context which is based on the published FpvTE operational test points.**

For the MST test, because the database size is much smaller, throughput of a matching engine is less of a factor than for databases of 1 million or more records. **For the MST test, Cogent therefore developed a configuration whereby the accuracy (FRR) is maintained at a consistently high level across a full range of false acceptance rate values (FAR).** This can be seen from the FpVTE Analysis Report for a FAR of 10^{-4} where the MST had a higher accuracy rate (FRR) than LST at the same evaluation point (FAR=10^{-4}). This difference can be easily

[5] The LST and MST charts in Appendix D have been updated to include lower values of FAR.

explained due to the different nature of the tests: the LST test was designed for large databases, fast response time and very low false acceptance rates (FAR=$0.5*10^{-7}$ through 10^{-8}) vs. the MST for a much smaller database. **Cogent's approach for the LST was to achieve the best possible false acceptance rate (FAR) performance (per the test guidelines) while achieving the high throughput rates required to search extremely large databases.** Although search time was not considered as a factor in the benchmarks, Cogent reflected performance considerations since they are reflective of what is required to deploy a large scale system to meet throughput demands required by today's customer community. This strategy did not impact the performance of our LST algorithm at very low FAR's, which is what is required for real world applications.

Database Size and Thresholds

Database size is a critical component of test design. Vendors set thresholds at different levels depending upon the database size and the stated FRR and FAR levels. A valid test to compare vendors should be set to specific parameters. Cogent followed the NIST instructions to set the LST FAR at very low levels, but NIST did not conduct the test consistent with their published test operational points. Cogent maintains that as the database would increase in size, the difference between the Cogent results and the vendor in the first position would get closer and closer and could reach the point of reversal such that Cogent would have been the top performer.

Resources Needed

Although FpVTE did not intend to measure the resources required to achieve FRR and FAR levels (e.g., equipment, response time, etc.), these factors will have a significant bearing on cost to develop, deploy, and operate a system for an end user. For the FpVTE benchmark, the resources needed for searching with the Cogent algorithms were significantly less than some other vendors within the group of most accurate systems. Also, Cogent conducted the LST in eight days while it took the vendor in the first position nineteen days. When considering the time and the computing resources, the differential between the vendor in the first position and Cogent is about 4 times. Cogent completed the LST faster and with fewer resources than the vendor in the first position. Time and resources are real world issues that were not controlled in the test.

The range of accuracy performance and response time are critical factors users are faced with in deciding which vendor to choose to implement a system. As demonstrated in the analysis regarding the FRR and FAR performance over a range of FAR levels, the Cogent algorithms can provide an extremely accurate system AND provide that accuracy with the highest levels of system throughput than can be scaled to databases from 1 million to tens of millions of subjects.

Conclusions

The following key performance characteristics were demonstrated by Cogent:

- Cogent's LST algorithm was designed to minimize false hits (FAR) and achieved a very low false acceptance rate of $0.5*10^{-7}$ through 10^{-8}; consistent with the original NIST Operational Points.
- Cogent's algorithms and solutions for large scale systems can be scaled to provide the lowest level of false acceptance and also provide the response time necessary for searching large databases in near real-time.
- The Cogent solutions benchmarked are not engineering models, but are based on commercial products that are available today and can be scaled to meet both database size and response time requirements.

DERMALOG

DERMALOG

Abstract

We would like to thank the FpVTE team for preparing and conducting such a comprehensive technology evaluation. The evaluation is a great contribution to the fingerprint community as a whole. It provides everyone in this field with a valuable comprehensive summary regarding such issues as single fingerprint vs. multiple fingerprints, importance of good enrolment quality, as well as the influence of many other variables on fingerprinting systems.

We were very disappointed that only DERMALOG's contribution to the LST could be evaluated, we look forward to submitting results for all parts of future evaluations. Unfortunately there were two aspects of the evaluation that proved to be a little disappointing in light of the high standard of the work.

In several places the vendor rankings are presented as results of the evaluation, e.g. in the document abstract, however fundamental differences in the evaluation conditions of the vendors are not mentioned, e.g. the relationship between computational power used and the accuracy obtained.

Attempting to remove the throughput from the purpose of the evaluation and nonetheless present vendor rankings as general results of the evaluation is like trying to make an omelette without breaking eggs. Additionally the rankings are presented as though they were transferable into the real world although the evaluation conditions were significantly different from those normally encountered.

Hardware Used

The FpVTE 2003 was based on equal conditions for all participants. In terms of throughput, the prerequisites were moderate enough to allow system providers new to the field of fingerprinting to take part. The openness of the requirements left the selection of hardware to the choice of the vendors. This led to a situation where some of the vendors used sophisticated, high performance hardware, while others decided to go with more affordable equipment. The costs of the hardware used by the different vendors in the LST differed by a factor greater than thirty, i.e. ranging from $3,080 up to $98,000. Unfortunately, there is no mention of these cost differences and the resulting differences in processing power in the rankings.

Software based acceleration of the matching process increases the error rates. To what extent this happens, differs from vendor to vendor and thus cannot be exactly estimated. Comparing matching accuracy without normalizing or taking into account the computational power used leads to biased results.

The report itself states: "For any operational system, throughput is a key parameter" (p. 71). This parameter had been specifically omitted from the evaluation. By setting almost no limitations for the hardware biased results have been obtained. We presume this was done with the intention of keeping the prerequisites on the vendors low, ensuring high participation levels, and to reduce the complexity of this first Fingerprint Vendor Technology Evaluation (FpVTE).

The report presents however very clear ranking of the vendors, and it mentions the consistency of the rankings throughout the tests. This consistency is not surprising as all tests took place under the same conditions, all influenced by the computational power used. By leaving out this hardware factor, an important key variable has not been evaluated.

When the report mentions accuracy, it should have mentioned that the FpVTE 2003 could judge accuracy only in the context of its test requirements. For other systems, e.g. real-world AFIS systems, different accuracy requirements apply. We would have been pleased to see this more clearly stated in the presentation of the rankings.

The following simplified considerations could illustrate the impact of the used hardware on the ranking: We assume a linear correlation between error rate (= 1-TrueAcceptRate) and hardware cost (given in appendix B of the report). Since the throughput requirements of LST were very low (in comparison to real-world AFIS requirements), we took the cheapest hardware in the test as a reference, and scaled the error rates of the other systems by the hardware cost factor. Even though this calculation is very inaccurate, it shows a tendency: After such normalization, there would be a leading group of only four vendors close to each other. The error rates of the other vendors, after normalization by the hardware cost factor, would be significantly higher.

Realistic Applications

The approach taken by the FpVTE 2003 makes it difficult to judge the performance of a real-world system. Even the LST does not come close to a real-world AFIS system regarding the required throughput figures and FAR settings.

Apart from the vendor rankings, all results presented in this report can be generalized. The vendor rankings, however, cannot. For this reason it is very unfortunate that both have been presented as if they had the same significance without further comment. This leaves a false impression of an objective and transferable judgment of the vendor capabilities.

In the report there is no mention that even the Large Scale Test only covers some parameters of a real world AFIS, which is a very complex system. The FpVTE 2003 tests only reflect some of the important features that should be considered when choosing an AFIS vendor or AFIS technology. Other issues that also play an important role (open system, scalability, system architecture, workflow, GUI, interfaces, price-performance etc.) were not part of the test and should be considered when selecting a vendor.

Future Evaluations

For future tests, DERMALOG would like to suggest a move towards real-world scenarios. This would allow for better comparisons between the vendors under more realistic conditions. It would also give an idea of the performance level that could be expected from real applications. Scenarios like 1:1 verification, 1:N access-control, or 1:LargeN AFIS should replace the abstract SST, MST, and LST. For each of these test scenarios, adequate hardware, throughput, and FAR requirements should be defined. Since all vendors, except NIST VTB, used Microsoft operating systems it would be a good idea to have the equipment provided by NIST. This would create equal test conditions for all vendors.

Conclusion

We hope this evaluation will help everyone in the fingerprinting community to continue to improve the systems and their acceptance through better understanding of current fingerprinting technology requirements. We are certain that the Fingerprint Vendor Technology Evaluation has proved it’s worth and sincerely hope that it will be continued.

In future evaluations we would like to see the inclusion of throughput and scalability evaluations, as well as other important features, as described in Section 7 of the report, to make the evaluation even more useful to the continually growing number of people interested in real world applications of fingerprinting systems.

GRIAULE

Centro de Tecnologia da UNICAMP
Rua Bernardo Sayão, 100/209, Cidade Universitária
13083-866 Campinas - SP – Brasil
info@griaule.com (55 19) 3788-4998

May 10, 2004

Griaule is a leading AFIS (Automated Fingerprint Identification System) software components supplier in Brazil, with customers in 6 states, some of them with more than 1.000.000 fingerprints enrolled.

We are very pleased with Griaule performance compared to the other participants. These excellent results prove that our technology accomplishes its goal of combating fraud by identifying people on the basis of their unique fingerprint features.

Frauds are usually related to the duplicate enrollment of someone under two different identities (names). As we were able to identify most of fingerprints, this can be translated as combating almost all frauds, inducing a systemic fraud diminution: **would someone try to enroll twice in a system that could identify up till 99, 8%[6] of these fraud attempts?**

Based on this we can expect to win an in depth, systemic, cost/benefits analysis.

Our software includes specific features that were designed for optimized performance on latent, single and n-print data. Unfortunately, due to a configuration issue, discovered only after reviewing our performance results, we only processed up till 2 fingers, even if more were available. It affected the performance of the tests where more than 2 fingers were available. Despite this we still had good results. FpVTE team notice in “Appendix C - System-specific results” that "*Griaule (LST) does not, in general, show an increase in accuracy as the number of fingers increases beyond 2. (…) As can be seen throughout the main body of the report, on 1-finger and 2-finger tests Griaule (LST) is generally about as accurate as Identix (LST) and consistently more accurate than either Biolink (LST) or NIST VTB (LST). This relationship changes when more than 2 fingers are considered.*"

We have been developing this technology since 1992 and permanently make R&D efforts to improve its recognition speed and accuracy. Since the evaluated release many improvements have already been implemented.

[6] Result for “GxC Ohio LST Partition”, FpVTE 2004.

Griaule thanks the organizers and sponsors of FpVTE for the opportunity to participate and for the excellent and independent work they did. Their work contributes to demystify AFIS technology and broadening its use.

IDENTIX

Comments on FpVTE 2003

The results of the Fingerprint Vendor Technology Evaluation (FpVTE) 2003 validate Identix BioEngine® as a strong and reliable fingerprint matching algorithm, that works well with fingerprint images of varying quality. With more than 100 million templates issued, BioEngine technology has been deployed around the world, for both 1:1 verification and 1:N civil identification programs.

Identix' active research program is continually improving the performance of the BioEngine fingerprint technology. Identix is a multi-biometric company with over 20 years of real world experience, established partnerships with a large number of systems integrators, card manufacturers and other biometric solution providers. We have worldwide fingerprint and facial biometric deployments for national ID, drivers' license, voter registration, background checking, physical and logical access control, social security payments and airport security. Identix applauds the momentous effort by the evaluating body and we value our continued strong relationship with the agencies and personnel involved in carrying out this evaluation.

MOTOROLA

Motorola Comments on NIST FpVTE 2003 Test
05/09/2004

Motorola/Printrak (Motorola) was very pleased to participate in the first FpVTE evaluation process. Over time this iterative testing process enables the identification and verification industry to greatly improve their ability to develop unbiased testing techniques that will greatly improve civil and criminal solutions that utilize fingerprints. The results of the first FpVTE test, taken in context and with recognition of its limitations, will be helpful to the industry.

Since Motorola's AFIS business is targeted to "identification solutions" (one to many) we are especially pleased to see that when multiple fingers were utilized, as happens in real world operations, performances between the top contenders became increasingly insignificant. This is shown within the Analysis Report on page 45 (figure 16) which clearly indicates Motorola's match rate for the 4, 8 and 10 fingers at better than 0.995 for the controlled data, and in Figure 17 a match rate of better than 0.985 for the operational data. These data points are indistinguishable from the designated "top tier" AFIS vendors. Another data point shows Motorola's average match rate to be at 0.9999 for tenprint matching at a False Accept Rate of 10^{-4} (Appendix D for the LST, Large Scale Test 17 partitions).

The true test for any AFIS solution happens during the customer acceptance testing phase of an implementation project. During the most recent such test conducted by Motorola, the "production" based Omnitrak solution searched and matched at a hit rate that exceeded the results of the FpVTE test. The customer's test included a background database of 800,000 tenprint records, and involved 1000 search tenprint records which is significantly larger than the tests included within the FpVTE. As a result, it is important to acknowledge that the FpVTE tests, while appropriate to a laboratory environment, are severely limited with respect to applicability in real world fingerprint identification applications. The FpVTE provides only a snapshot of the overall verification and identification performance picture.

The FpVTE testing limitation was further exacerbated in that only the results of 1:1 "verifications" were reported. By limiting the scope of the reported results, the ability to select the right candidate from a group, which is the most fundamental requirement for a successful real world AFIS operation, is not shown. Additional limitation considerations include:

- ***Test Results Include NO Identification (1:N) Results.*** By the very design of the test the 1:N identification search of a database, which represents the majority of the uses of an AFIS, was not performed, or evaluated in any way – apparently for the reason stated in the Appendix E Analysis Issues (p.8):
  "Large fingerprint matchers such as automated fingerprint identification systems (AFISs) typically perform identifications using 1:N strategies, often filtering out many of the candidates quickly to minimize resource expenditures. Following this approach on a verification test can be disadvantageous."
  We agree with this comment that acknowledges that there are fundamental differences in approach from the "laboratory" environment of the 1:1 tests which are reported vs. "real world" operations on 1:N identifications. We believe that it is these very differences that

severely limit the ability to draw some of the hard and fast conclusions stated in the reports Executive Summary.

- ***The majority of LST partitions are not considered in the report***. The decision by NIST to not report on the results of 51 of the 95 partitions of the LST findings, selecting only 44 for detailed analysis is surprising and results in reporting that at best brings much subjectivity into the process.[7]

- ***Use of normalized scores move results away from "real world"***. An earlier publication by the NIST (NISTIR 6965), entitled 'The FRVT Evaluation Report' makes an interesting remark about score normalization: (p. 13, 2nd paragraph)

  "FRVT 2002 allowed participants to submit normalization procedures. Normalization is a post processing function that adjusts similarity scores based on a specific gallery (file prints) …The input to the normalization routine is the set of all similarity scores between a probe (a search print) and a gallery….Normalization requires that a probe be compared with a gallery. When normalization is applied, is verification still 'one to one' matching?"

  In real life, of course, one would have only one search and one file print – there would be no recourse to the normalization process that is allowed in both the FRVT and FpVTE. Therefore, tests that allow normalization do not correctly reflect the accuracy of true, real-life 1:1 matching. It should be noted that Motorola had to modify our Omintrak production software in order to normalize the results.

- ***Test results were limited to a single accuracy metric***. The test employed an average value of TAR at an arbitrary value of FAR = 10^{-4}, as the main criterion for the evaluation. Alternative criteria such as the median, or the value of the TAR at other FAR values, for example at zero FAR which reflects true "lights out" operation were not given their due consideration in the final evaluation. In a real world environment, careful consideration must also be given to the trade-off between the false accept rate and the false reject rate. Such a trade-off is made within the context of a given application, and is critical to the successful operation of identification systems.

In summary, Motorola found the test results both informative and instructive, but we suggest that due consideration of test limitations be considered in attempting to draw conclusions regarding real world performance. As a result, we look forward to the opportunity to participate in any future FpVTE testing which we are hopeful will be designed provide a broader picture of the verification and identification solutions.

[7] This is addressed by a clarification in Appendix E, Section 2.2.

NEC

FpVTE 2003 Report - NEC Comments – May 10, 2004

NEC Solutions (America), Inc.

1. Introduction

NEC appreciates the opportunity that the FpVTE 2003 Team provided for the demonstration of our latest Fingerprint Matching technology. NEC is very pleased with our exceptional results, of achieving the highest accuracy of all of the participants. These results prove that NEC's matching algorithms, originally developed for latent matching, can also successfully be applied to non-law enforcement applications.

The FpVTE 2003 Report illustrates valuable and important statistical accuracy information on the operation of fingerprint matching systems. NEC would like to suggest an additional method that can be utilized to evaluate and compare the test results. The method described below also seems to clearly demonstrate NEC's superior matching technology.

2. Accuracy Differences among Top Three Participants

The FpVTE2003 Report concluded that the three most accurate systems are developed by **NEC**, Cogent, and Sagem (refer to Abstract, page 2). However, we would like to emphasize as discussed below, that NEC's matching technology greatly out-performs Cogent and Sagem. Per the Table 20 (page 52), "MST Identification Rate at Rank 1" the top three participants are shown below:

1st	**NEC**	**99.4%**
2nd	Cogent	98.9% (-0.5% to NEC)
3rd	Sagem	98.4% (-1.0% to NEC)

At a glance, these small differences (0.5% or 1.0%) do not seem significant. If the target accuracy range is around 50% or 60%, such small differences are not statistically significant. However, when the target accuracy range is close to 100%, even a small difference can be very significant. The rate of NEC's mis-identification (0.6%) is almost one half of Cogent's mis-identification (1.1%), which is a significant difference.

In order to properly compare performance of two algorithms, NEC adopts a simple method called "*Improvement Percentage or %imp*". When we compare Algorithm 2 (alg2) over Algorithm 1 (alg1), the **%imp** is calculated as follows:

%imp = (%alg2 – %alg1)*100/ (100 – MIN(%alg1, %alg2))

*Note: The **%imp** conceptually shows how the better algorithm reduced errors of the worse algorithm.*

NEC applied the **%imp** method on the most difficult of 10 partitions out of the LST 27 Partitions of Operational Data. The figures in the following table (columns 2, 3, 4 and 7) are "TARs where the FAR = 10^{-4} ". They were stripped from Appendix D page 5 "**Comparison of Performance on LST Partitions**".

LST Partitions (Tough 10)	1st NEC	2nd Sagem	3rd Cogent	NEC's %imp over Sagem	NEC's %imp over Cogent	4th Dermalog	Cogent's %imp over Dermalog
BxA/BCC	0.991	0.979	0.971	**57%**	**69%**	0.944	**48%**
BxA/DHS2	0.988	0.963	0.973	**68%**	**56%**	0.946	**50%**
BxA/IdentIafis	0.996	0.986	0.985	**71%**	**73%**	0.964	**58%**
HxC/IdentIafis	0.987	0.970	0.966	**57%**	**62%**	0.937	**46%**
HxD/12k	0.983	0.961	0.959	**56%**	**59%**	0.901	**59%**
HxI/IdentIafis	0.992	0.950	0.953	**84%**	**83%**	0.905	**51%**
HxJ/12k	0.990	0.974	0.952	**62%**	**79%**	0.923	**38%**
AxA/BCC	0.996	0.990	0.987	**60%**	**69%**	0.972	**54%**
GxC/IdentIafis	0.997	0.994	0.992	**50%**	**63%**	0.985	**47%**
GxI/IdentIafis	0.997	0.980	0.982	**85%**	**83%**	0.953	**62%**
Average	0.992	0.975	0.972	**65%**	**70%**	0.943	**51%**

Note: NEC excluded all partitions where one or more Participants achieved 0.999 or better (accuracy saturated) because too easy data sets (very good quality images) are not useful to compare potential performance of each algorithm.

NEC selected the LST results for review and analysis because the LST test data size is more than 10 times larger than the MST or SST test data size, therefore they are the most reliable. NEC's **%imp** over Sagem is 65% (average) and 70% over Cogent. These figures suggest that NEC can reduce mis-identification by 65% or 70% (more than half). Moreover, these **%imp** figures are consistent over most of the 10 Partitions. This consistency supports that the %**imp** is reliable and applicable to many operational cases.

The last column illustrates Cogent's **%imp** over Dermalog (4th ranked participant). Please note that Cogent's %**imp** (51%) is much less than NEC's %**imp** (65% or 70%). This suggested analysis method demonstrates that, *NEC's greater accuracy over Sagem and Cogent, is more significant than Cogent's greater accuracy over Dermalog*. This is the reason NEC believes the FpVTE 2003 results demonstrate NEC's much greater degree of accuracy over the 2nd and 3rd ranked participants, Sagem and Cogent.

3. Accuracy Robustness against Data Base Growth

It is generally known that accuracy drops when data base size grows. Even when the accuracy difference is small between the best and worst algorithm on a small data base size, the accuracy difference will become larger when data base size grows. Figure 22 (page 51) shows the Rank-based Identification performance of MST. Top three accuracies are as shown below:

Participants	**1st Rank**	**Up to 10th**
NEC	99.4%	99.67%
Cogent	98.9% (-0.5% to NEC)	99.33% (-0.34% to NEC)
Sagem	98.4% (-1.0% to NEC)	99.00% (-0.67% to NEC)

*Note: The numerical values of the **Up to 10th** column were manually estimated from the graph in Figure 22.*

The accuracy differences in the **1st Rank** are larger than the ones in the **Up to 10th**. The similar behavior of differences (from the **Up to 10th** to the **1st Rank**) is often observed when the database size grows. Therefore, the insignificant difference in a small data base should be carefully analyzed. NEC applied the **%imp** method on these accuracies.

	1st Rank	**Up to 10th**	**Difference**
NEC's %imp over Cogent	**45%**	**51%**	**6%**
NEC's %**imp** over Sagem	**63%**	**67%**	**4%**

Another advantage of the **%imp** method is the fact that this value is consistent regardless of the database size. Please note that the %**imp** over Cogent and Sagem is almost the same between the **Up to 10th** and the **1st Rank**. This demonstrates that NEC's superior accuracy is more critical when the data base size becomes large.

4. Cost Performance Issues

The FpVTE2003 conditions clearly state that it will only evaluate accuracy and not cost-effectiveness, matching cost or speed (refer to Section 1.2, page 7). Therefore, NEC opted to utilize an accuracy-oriented system for the test and not use Pre-selection or Adaptive Finger Selection to improve performance. NEC used the maximum combination of matching algorithms, including a prolonged algorithm to further improve accuracy. However, this may not be the same strategy used if input data quality is relatively good. Because NEC has a wide range of matching algorithms and additional core technologies to improve system performance, we are confident we can provide the most accurate solution under the specific conditions required by each customer. Please reference NEC's System Description Document for detail of NEC Core Technologies.

5. Final Comments

NEC would like to thank the FpVTE 2003 Team for conducting such a comprehensive evaluation utilizing a sizable amount of reliable test data without any duplicates or mating errors. Without reliable test data, it would be impossible to target a high level (99%) of matching accuracy. NEC looks forward to our continued work with the user community to allow us to further enhance and contribute our core technologies.

NEUROTECHNOLOGIJA

Ateities 10, Vilnius 08303, Lithuania, Phone: +370 52 773 315, Fax: +370 52 773 316,
http://www.neurotechnologija.com

Comments on the FpVTE 2003 Report

Neurotechnologija Ltd. would like to thank the FpVTE 2003 team for giving us the opportunity to participate in the evaluation. We are also thankful for the help that was provided before and during the tests and for the detailed FpVTE 2003 results report. We are pleased with the results of the evaluation and the fact that the compact, fast and utilising small templates VeriFinger algorithm was shown to provide one of best reliability results in the evaluation.

We ran the entire FpVTE MST test series on a single desktop PC with common components. For optimal reliability results, we set the algorithm in slowest mode. Even in this mode, VeriFinger algorithm was able to complete the entire MST series in only 2 days. These facts need to be taken into consideration by individuals that are considering investing in a fingerprint identification technology since they have strong implications on cost and systems complexity.

The FpVTE protocol did not allow the use of some of our advanced algorithm features, which, in a real world application, would further increase the recognition quality. Particularly, the MST set contained images from different scanners, but each certain image scanner model was not disclosed. In a real world scenario, specific parameters would be set for each specific scanner type. This would allow the algorithm to perform at an even higher accuracy level.

Another such real world example that was not simulated in the FpVTE protocol is the ability to generate globalized or generalized features templates by capturing several images from the same finger and combining the templates into a single features set. Using a generalized feature set can significantly improve the algorithms reliability and produces much improved matching scores. In the FpVTE MST set such a method could not be used as only two matched fingerprints information were allowed to consider.

During the test we experienced some software problems related to WSQ file decoding. During the FpVTE Evaluation period our original software did not contain WSQ reading functions and so we used third party software for this. Unfortunately the third party software was unable to decode more than 30 WSQ files used in the FpVTE MST live dataset and all these fingerprints matching results was equate to zero.

Since the FpVTE 2003 Neurotechnologija has developed some algorithm improvements on the version tested in the contest. New fingerprint filtration functions were developed, allowing better filtration of low quality images. Additionally by using feature set optimisation, the generated templates size has been decreased from 300 - 600 bytes to 150 - 300 bytes per fingerprint. Identification speed has also been increased from 5% to 100%, depending on the number of

fingerprint minutiae. All these improvements give us hope of even better results in the next evaluation.

PHOENIX GROUP

The Phoenix Group, Inc.
205 N. Walnut
Pittsburg, KS

Response to the NIST FpVTE report.

We are very pleased that we were invited and then made the decision to participate in the NIST FpVTE test. The test results have proven invaluable to us in identifying areas for improvement in our fingerprint and palmprint matching products. Hundreds of law enforcement agencies around the world who use our software in crime scene investigation will benefit from our participation in the test.

Absolutely no optimization of our software was made specifically for this test. Our intention was to test the exact algorithms that examiners and investigators in the field are applying when they use our products. What have we learned?

The extract and match results reported from the Ohio (Slaps) (689 pairs) was as expected. We also expected the results from the lower quality image data in the test to fall below the Ohio data; as the image quality decreased so would the matching proficiency of our software. However, we were surprised to see the amount of disparity from the higher quality data to the medium and lower quality images that the test revealed.

Feedback from users in the field has typically been that our software is very proficient in matching extremely poor quality crime scene latents with record prints in the database. Often matches are found with as few as 5 or 6 points plotted. Of course, minutiae on crime scene latents are plotted manually by the latent print examiner or investigator entering the crime scene data into the system. With high quality record prints scanned into the system from ten-print cards, or imported directly from live-scan systems, and then automatically extracted by the system, these matches would parallel more closely the Ohio slap match data in the FpTVE test. Thus, we feel the field reports of the success of our software in crime scene investigation, where a relatively high level of user intervention in plotting minutiae on poor quality latent fragments, matched against a database of reasonably good quality record prints, agrees with the results we see in the FpVTE test results of our software.

Based on the FpTVE report, we conclude that our software could use improvement in several key areas, especially the auto-extraction process as it is applied to poorer quality record prints without any operator intervention. When plotting minutiae automatically, our software uses several methods in an attempt to find quality level 2 details in fingerprint images that have less than ideal clarity. We apply some filtration

and some background adjustments depending on what the extraction algorithm sees when attempting to extract the data. We also employ an automatic thresholding system, based on other data gathered, in an attempt to eliminate as many false points as possible while not eliminating any valid points. While our review is only preliminary, we have identified three specific features of our extraction algorithms and two specific areas of our matching algorithms that have become the focus of intense scrutiny. Improvements are already being developed. An update will be tested and in users' hands within a few months.

Again, we sincerely thank NIST for allowing us to participate in the FpTVE test. We appreciate all the hard work that goes into preparing, promoting and conducting a formal test of this magnitude. We also want to thank the other vendors that participated. The FpTVE has provided us with information that will benefit law enforcement agencies here in the U.S. and around the world.

SAGEM

SAGEM MORPHO, Inc.

SAGEM Morpho, Inc

Comments on the FpVTE 2003 Report

May 2004

Introduction

SAGEM was pleased to be able to participate in the Fingerprint Vendor Technology Evaluation (FpVTE) conducted by NIST. This type of objective benchmark is critical to distinguish between vendors who have convincing marketing hype from those with truly high-performance technology.

SAGEM would like to commend the FpVTE test team on their planning, thoroughness, fairness, objectivity and discipline to be able to conduct such an ambitious test successfully, and offer the following comments for consideration regarding the evaluation.

Scope of Test – Large AFIS Systems

The FpVTE certainly met the objective of evaluating fingerprint identification systems in how well they match different combinations of types and qualities of fingerprints, including flat/rolled prints, single/multiple fingers, inked/livescan input devices and varying quality images.

Given the limited time and resources available for FpVTE, it was not practical to create a test database of sufficient size to evaluate the performance of each participant in matching very large-scale AFIS databases. Such a test would include evaluation of the accuracy of techniques, including dynamic thresholding and multi-stage matching algorithms, which are required when searching databases of tens or hundreds of millions of prints. In LST, even though there were more that one billion set to set comparisons, the largest target set contained only 9,000 records. It would be advisable for agencies that have applications requiring one-to-many matching of very large databases to consider additional benchmark testing of a limited number of vendors using datasets at least two orders of magnitude larger. In addition, an agency should review and evaluate the performance of existing operational identification systems that contain millions or tens of million of records.

Factors Impacting Accuracy

SAGEM strongly agrees with the conclusion of FpVTE that "additional fingers greatly improve accuracy" and that "poor quality fingerprints can greatly reduce accuracy". Agencies implementing large-scale systems should take these factors into account when defining requirements. A system that will grow to tens of million of records will need to use more than one or two flat fingerprints per record in order to sustain a high performance level. Also efficient quality control functions at the point of capture and during back-file conversion are essential to successful operational performance.

Cost versus Accuracy Trade-off

SAGEM recognizes that the primary purpose of FpVTE was to evaluate accuracy rather than system cost but it is interesting to note that even though the accuracy of the top three vendors was very similar, the hardware configuration and time required to process the more than one billion comparisons in LST was significantly different between the vendors. Section 5 of Appendix D discussed only the lack of correlations of accuracy to system size and accuracy to processing time in the MST. However there was a significant difference in the processing power required to achieve a similar level of accuracy between the top three vendors in LST as discussed below.

SAGEM is proud of the fact that our MetaMatcher™ algorithms are optimized and in commercial production operating on commercial-off-the-shelf (COTS) processors rather than using proprietary hardware accelerators. This is highlighted by the clearly superior processing efficiency demonstrated during the FpVTE LST evaluation.

As an example, it is very revealing to calculate the "CPU-Days" (processing days multiplied by the number of CPU's) required by the top three vendors in LST. The standard operational SAGEM algorithm testbed (SAGEM L2) used only **12** CPU-Days to complete the LST testing, while the other two leading vendors required **96** and **342** CPU-Days to perform the same billion comparisons. It is significant that SAGEM achieved comparable accuracies to these other vendors who used **8** times and **28** times as much processing power as the SAGEM testbed. It is both a demonstration of the optimization that SAGEM has achieved with its software-based feature extraction and matching libraries and is also a strong indication of the potential lower system cost of the SAGEM system compared to other vendor systems built with their FpVTE libraries that achieve the same accuracy operating point.

SAGEM chose to include in the testing our commercially available algorithm (L2) that is integrated into operational systems rather than only testing experimental (non-commercial) algorithms. New SAGEM algorithms integrate improvements based on the latest operational site feedback acquired on very large-scale systems. These same MetaMatcher™ software libraries are being integrated into all AFIS products currently being deployed by SAGEM, whether they are used for person identification as in FpVTE or for crime solving matching of latents of fingerprint or palmprint friction ridges. SAGEM's platform-independent software approach (rather than a dependence on proprietary hardware) also allows our customers to immediately benefit from both algorithm improvements and from commercial processor hardware improvements in a cost-effective way.

Conclusion

FpVTE has confirmed SAGEM as a world-wide leader in the fingerprint identification industry. We are pleased with the results and the dominant processing efficiency of the SAGEM MetaMatcher™ algorithms. We welcome the opportunity to demonstrate the continual improvements in our technology as a result of our on-going algorithm research and development in future tests such as large database evaluations, quality level correlation testing, minutiae-based interoperability testing, etc.

SAGEM would like to thank the sponsors, supporters and especially NIST and the rest of the FpVTE test team who designed and conducted this important evaluation.

Contacts
International: SAGEM, SA: www.sagem.com
U.S. and Canada: SAGEM Morpho, Inc: www.morpho.com

1145 Broadway Plaza, Suite 200, Tacoma, WA 98402
info@morpho.com, 800-346-2674

TECHNOIMAGIA

Final Comment on FpVTE2003

Technoimagia Co., Ltd.

Overview

FpVTE2003 is the first fingerprint evaluation that was conducted globally with a specific requirement: establishing a measure of identifying terrorists to prevent our daily life from their attack. Therefore, the ultimate purpose of the evaluation is to investigate the current technological levels of fingerprint authentication systems for "the negative identification."

Technoimagia has been devoting itself to develop commercial fingerprint authentication systems for "the positive identification." The positive identification is the kernel of commercial applications. However, its idea is philosophically different from what FpVTE2003 is seeking because the former is a measure of securing one's identity. For the positive identification, a user voluntarily and willingly provides his fingerprint images to an authentication system so that he/she may secure a data communication environment through the Internet and other digital media.

Technoimagia offers positive authentication systems with the following advantages:

1. **Nimble authentication engine:** a complete program size 50 Kbytes or less to run the *match on card* function for a smart IC card application.
2. **Smallest data size to store characteristics points of a fingerprint:** template required 100 bytes or less to work with the *match on card* function.
3. **High-speed verification and identification:** less than 1 second for *1:N* identification, where $N = 50{,}000$.
4. **Optimal set of parameters appropriate to any customer's preferred sensor type.**

Thus, it was our challenge to determine how to tailor our existing authentication algorithm in order for participating in FpVTE2003. We did a minimum modification so that we may utilize the advantages of our algorithm. Our most concern was the processing time.

At first, the result of FpVTE2003 totally shocked us. However, careful examination of the result convinced us of proper functionalities that our algorithm has achieved for positive commercial applications. This final comment describes how we have come to the conclusion.

The analysis result pointed out that capture devices alone do not determine fingerprint quality. It is true. Then, what factor mostly determines the quality of a captured image? The result also pointed out that our algorithm vividly showed a dramatic difference in performance among various fingerprint images used in the evaluation. It led us to conclude that the analysis result indeed suggested us what degree of image quality and operational condition we had to use for accurate and robust fingerprint identification. The negative identification needs to deal with wide variety of image qualities due to unpredictable conditions while capturing fingerprints. If you would setup a fingerprint identification system with specially dedicated hardware and perform complicated preprocesses such as image enhancement prior to extracting characteristic points, you would be able to overcome the issue. However, for any commercial positive identification, we believe that a massive system is not practical nor unbeneficial to commercial needs because the raw image quality is the most crucial factor to achieve required authentication performance. Thus, we attempted to challenge FpVTE2003 with a modest hardware configuration.

The analysis of FpVTE2003 also revealed importance of optimizing fingerprint parameters to take an account of different sensor types and image acquisition conditions. In addition, the analysis guided us to discover an alternative method for improving authentication accuracy without knowing the various conditions on image acquisition sensors if our propriety optimization is not allowed.

Comment 1: Characteristics feature of our algorithm

Because each sensor type for fingerprint image is different, our authentication algorithm offers optimization for a sensor type. For the task, we evaluate a statistically meaningful accuracy based on several images acquired by the sensor. In this way, we can prepare an appropriate template. Better a template is, more accurate fingerprint authentication can be. This is a propriety procedure that is used in "Fingerprint Authentication System Tool 21 (*FAST21*)" for the positive identification. Since we were unable to perform the preprocess routine during the evaluation, we determined to use a set of parameters that were optimized to our own optical fingerprint sensor, *FP-50LSC,* assuming optical sensors adopted by NIST would be equal or better than ours in a similar operational environment. With *FP-50LSC*, we have achieved a commercially available system with typically FAR $\leq$ 0.0002%, and FRR $\leq$ 0.05%.

Figure **6** of the report shows a dramatic difference between the BCC and the DHS2 results. We attributed the difference to the selectively optimized set of parameters because DHS2 has a greater proportion of poor-quality fingerprints that were obtained in a stressful condition to the subjects. We are confident that once we optimize parameters of each of the sensors, we would be able to record much better performance.

We could develop an "automatic adjustment function" for the parameter optimization. On the other hand, we must consider the throughput as another important factor that determines the matching performance. We decided not to add any automatic adjustment routine for this reason. As the result, we were successfully complete the task of FpVTE2003 in 5 days with a single PC that has a commercially available hardware configuration. That process was actually *shorter* than we estimated.

Comment 2: Poor quality image and IQM score

FpVTE2003 did not allow us to reject any fingerprint image even it has poor quality. For any commercial application, *FAST21* requires a user to re-enter a fingerprint image if the previous image is unacceptable. The algorithm issues this request immediately after receiving a bad image. We believe that this real time procedure would be the best solution to construct an acceptable template. With well-prepared templates, the authentication process will be accurate and robust for commercial applications while keeping the template size minimum. For example, *Fast21* can process *1:N* identification with 50,000 fingers within a second. Such high-speed process can be attained with a template whose data size is 256 bytes or less. Furthermore, we only require 100 bytes or less to have the *match on card* function for smart IC card applications.

Fast21 simply rejects a fingerprint image if we could extract the required number of characteristics points in a commercial application. The number of characteristic points to be extracted is different: 12 or 15. It is determined by on each country's judicial criterion, and we set the number of characteristic points in our algorithm following this rule. As a strategy to FpVTE2003, because the minimum number of the characteristics points required to determine a coordinate system is 3, we cerate a special rule for fingerprints that provide at most 2 characteristics points as follows:

A. If one fingerprint of a pair has zero characteristic point, the pair unconditionally has the similarity score 0: unmatched;

B. If each fingerprint of a pair has 1 characteristic point, the pair unconditionally has the similarity score 1: matched; or

C. If one fingerprint has 1 characteristic point whereas the other has 2 characteristic points in a pair, the similarity score will be around 0.5.

This is the *sole* modification that we made for FpVTE2003. Figures **104** and **105** showed the effect of this modification. In particular, figure **104** showed peculiar fingerprint images from which our matching routine concluded two extreme results of coexistence of matched and non-matched scores at a same score. No other vendor showed this behavior to this extent. We also attributed the curve of figure **103** to this modification. As the analysis pointed out, figure **105** shows noticeable peaks in non-match scores at 0.5 and 1.0, which may account for the distinct shape of our BCC ROC. The rule indicates that there were substantial amount of fingerprint image with characteristics points 2 or less that affected the statistical significance. From figure **106**, it is clear that what degree of image quality we must use for accurate authentication.

We provided the image quality data. Analysis reveled that %FET is 8.55 %. From our experience, the number is unacceptable; it should be less than 6 % for a realistic commercial fingerprint authentication. The high %FET value is a consequence of the wide variety of fingerprint images provided by FpVTE2003. However, more importantly, we have discovered a method for improving the accuracy for the same image set by adjusting one parameter. Since we are not allowed to reproduce the image set of the test, we are unable to prove it. However, our conjecture is conclusive.

Acknowledgement

FpVTE2003 provided us with an opportunity to expose ourselves to an unchallenged domain of fingerprint authentication. We experienced much excitement from preparation to test. It is our honor to participate in FpVTE2003 and we would like to express our sincere gratitude to the test and the analysis teams of FpVTE2003.