FVC2000: Fingerprint Verification Competition

Dario Maio, *Member*, *IEEE*, Davide Maltoni,
Raffaele Cappelli, J.L. Wayman, and
Anil K. Jain, *Fellow*, *IEEE*

Abstract—Reliable and accurate fingerprint recognition is a challenging pattern recognition problem, requiring algorithms robust in many contexts. FVC2000 competition attempted to establish the first common benchmark, allowing companies and academic institutions to unambiguously compare performance and track improvements in their fingerprint recognition algorithms. Three databases were created using different state-of-the-art sensors and a fourth database was artificially generated; 11 algorithms were extensively tested on the four data sets. We believe that FVC2000 protocol, databases, and results will be useful to all practitioners in the field not only as a benchmark for improving methods, but also for enabling an unbiased evaluation of algorithms.

Index Terms—Fingerprint verification, performance evaluation, biometric systems.

1 Introduction

In the last decade, interest in fingerprint-based biometric systems has grown significantly [9]. Activity on this topic increased in both academia and industry as several research groups and companies developed new algorithms and techniques for fingerprint recognition and as many new fingerprint acquisition sensors were launched into the marketplace.

Nevertheless, to date only a few benchmarks have been available for comparing developments in fingerprint verification. Developers usually perform internal tests over self-collected databases. In practice, the only public domain data sets are the US National Institute of Standards and Technology (NIST) CD-ROMs [20], [21] containing thousands of images scanned from paper cards where fingerprints were impressed by rolling "nail to nail" inked fingers. Since these images significantly differ from those acquired by optical or solid state sensors, they are not well-suited for testing "online" fingerprint systems [9], although they constitute an excellent benchmark for AFIS (Automated Fingerprint Identification Systems) developments [11] and fingerprint classification studies [4]. NIST recently released a database containing digital videos of live-scan fingerprint data [22]; since this database was specifically collected for studying plastic distortion affecting the online acquisition process [5], [6] and the impact of finger rotation, it models only certain fingerprint variations and it is not recommendable for a general evaluation of verification algorithms.

The lack of standards has unavoidably led to the dissemination of confusing, incomparable, and irreproducible results, sometimes embedded in research papers and sometimes enriching the commercial claims of marketing brochures.

The aim of this initiative was to take the first step towards the establishment of a common basis, both for academia and industry, to better understand the state-of-the-art and what can be expected from the fingerprint technology in the future. Analogous efforts have been recently carried out for other biometric characteristics (e.g., face [15], [12]) and, in general, for other classical pattern recognition tasks ([17], [1], [8], [18]). We decided to pose this effort as an international open competition to boost interest and give our results larger visibility. The 15th International Conference on Pattern Recognition (ICPR 2000) was ideal for this purpose. Starting in late spring 1999, when the FVC2000 Web site [7] was set up, we broadly publicized this event, inviting all companies and research groups we were aware of to take part.

From the beginning, we stated that the competition was not meant as an official performance certification of the participant biometric systems, as:

- The databases used in this contest have not been acquired in a real environment and according to a formal protocol [23], [16], [19], [2] (also refer to [24] for an example of performance evaluation on real applications).
- Only parts of the participants software are evaluated by using images from sensors not native to each system. In fact, fingerprint-based biometric systems often implement proprietary solutions to improve robustness and accuracy (e.g., quality control modules to reject poor quality fingerprints, visual feedback to help the user in optimally positioning his/her finger, using multiple fingerprint instances to build more reliable templates, etc.) and these contributions are here discounted.
- According to the definition reported in [16], [19], FVC2000 should be conceived as a *technology evaluation* (with some analogies with the FERET contest organized by Philips on face recognition [15]). In fact, quoting [2]:

> "The goal of a technology evaluation is to compare competing algorithms from a single technology. Testing of all algorithms is done on a standardized database collected by a "universal" sensor. Nonetheless, performance against this database will depend upon both the environment and the population in which it was collected. Consequently, the "three bears" rule might be applied, attempting to create a database that is neither too difficult nor too easy for the algorithms to be tested. Although sample or example data may be distributed for developmental or tuning purposes prior to the test, the actual testing must be done on data which has not been previously seen by algorithm developers. Testing is done using "offline" processing of the data. Because the database is fixed, results of technology tests are repeatable."

In FVC2000, the "universal" sensor is actually a collection of four different sensors/technologies to better cover the recent advances in fingerprint sensing techniques and to avoid favoring a particular algorithm through the choice of a specific sensor. In fact, databases 1 and 2 were collected by using two small-size and low-cost sensors (optical and capacitive, respectively). Database 3 was collected by using a higher quality (large area) optical sensor. Finally, images in database 4 were synthetically generated by using the approach described in [3]. Each of the four databases contained 880 fingerprints from 110 different fingers, collected using the "three bears rule" (not too easy, not too hard), based on our prior subjective experiences with fingerprint recognition algorithms; in particular, on the one hand, we discarded fingerprint images we considered completely intractable even for a human expert, on the other hand, we avoided collecting perfect fingerprints which will be very easy for a matching algorithm; some internally developed

- *D. Maio, D. Maltoni, and R. Cappelli are with Biometric System Lab (BIOLAB)—DEIS, University of Bologna, via Sacchi 3, 47023 Cesena, Italy. E-mail: {maio, maltoni, cappelli}@csr.unibo.it.*
- *J.L. Wyman is with the US National Biometric Test Center, College of Engineering, San Jose State University, San Jose, CA 95192. E-mail: jlwayman@aol.com.*
- *A.K. Jain is with the Pattern Recognition and Image Processing Laboratory, Michigan State University, East Lansing, MI 48824. E-mail: jain@cps.msu.edu*

Manuscript received 09 Oct. 2000; revised 27 Feb. 2001; accepted 15 May 2001.
Recommended for acceptance by D. Kriegman.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 112958.

TABLE 1
List of Participants

ID	Organization	Type
CETP	CEFET-PR / Antheus Technologia Ltda (Brasil)	Academic
CSPN	Centre for Signal Processing, Nanyang Technological University (Singapore)	Academic
CWAI	Centre for Wavelets, Approximation and Information Processing, Department of Mathematics, National University of Singapore (Singapore)	Academic
DITI	Ditto Information & Technology Inc. (Korea)	Commercial
FPIN	FingerPin AG (Switzerland)	Commercial
KRDL	Kent Ridge Digital Labs (Singapore)	Academic
NCMI	Natural Sciences and Mathematics, Institute of Informatics (Macedonia)	Academic
SAG1	SAGEM SA (France)	Commercial
SAG2	SAGEM SA (France)	Commercial
UINH	Inha University (Korea)	Academic
UTWE	University of Twente, Electical Engineering (Netherlands)	Academic

A four digit ID was assigned to each algorithm. (Sagem SA submitted two different algorithms).

TABLE 2
The Four FVC2000 Databases

	Sensor Type	Image Size	Set A (w×d)	Set B (w×d)	Resolution
DB1	Optical Sensor	300×300	100×8	10×8	500 dpi
DB2	Capacitive Sensor	256×364	100×8	10×8	500 dpi
DB3	Optical Sensor	448×478	100×8	10×8	500 dpi
DB4	Synthetic Generator	240×320	100×8	10×8	About 500 dpi†

† In the artifical generation, the resolution is controlled by the average ridge-line interdistance; this input parameter was estimated from a real 500 dpi fingerprint database.

algorithms helped us in accomplishing this task. Each database was split into a sequestered "test" set of 800 images (set A) and an open "training" set of 80 images (set B), made available to participants for algorithm tuning. The samples in each set B were chosen to be as much as possible representative of the variations and difficulties in the corresponding set A; to this purpose fingerprints were automatically sorted by quality as in [14] and samples covering the whole range of quality were included in set B. A final visual inspection of the obtained data sets was carried out to assure that "dry," "wet," "scratched," "distorted," and "markedly rotated" fingerprints were also adequately represented.

As initially specified in the call for participation *"FVC2000 competition focuses only on fingerprint verification (1-1 matching) and not on fingerprint identification (1-N matching)"* [9]. Each participant was required to submit two executable computer programs: the first *enrolling* a fingerprint image and producing the corresponding template, the second *matching* a fingerprint template against a fingerprint image. Participants were allowed to submit four distinct configuration files, to adjust the algorithms internal parameters according to each specific database; configuration files could also contain precomputed data, to save time during enrollment and matching. For practical testing reasons, the maximum response time of the algorithms was limited to 15 seconds for each enrollment and five seconds for each matching (on a Pentium III—450 MHz machine).

In March 2000, after several months of active promotion, we had 25 volunteering participants (about 50 percent from academia and 50 percent from industry), far more than our initial expectation. By the end of April, the training sets were released to the participants.

After the submission deadline (June 2000) for the executables, the number of participants decreased to 11 (most of the initially registered companies withdrew). In any case, the number of participants (see Table 1) was more than anticipated, so we started working on the submitted executables to complete their evaluation by August 2000.

Once all the executables were submitted, feedback was sent to the participants by providing them the results of their algorithms over training set B (the same data set they had previously been given) to allow them to verify that neither run time problems nor hardware-dependent misbehaviors were occurring on our side.

Section 2 describes the four databases used; in Section 3, we present the criteria and the procedures used for performance evaluation. Section 4 reports the overall performance of the participating algorithms on each database and concludes with a comparison of the average results. Finally, in Section 5, we draw some concluding remarks and discuss how we intend to continue supporting this initiative in the future.

2 DATABASES

Four different databases (hereinafter DB1, DB2, DB3, and DB4) were collected by using the following sensors/technologies [10]:

- DB1: optical sensor "Secure Desktop Scanner" by KeyTronic
- DB2: capacitive sensor "TouchChip" by ST Microelectronics
- DB3: optical sensor "DFR-90" by Identicator Technology[1]
- DB4: synthetically generated based on the method proposed in [3].

Each database is 110 fingers wide (**w**) and eight impressions per finger deep (**d**) (880 fingerprints in all); fingers from 101 to 110 (set B) were made available to the participants to allow parameter tuning before the submission of the algorithms; the benchmark is then constituted by fingers numbered from 1 to 100 (set A). For a system evaluation, the size of the above four databases is certainly not sufficient to estimate the performance with high confidence.

1. These sensors are identified in order to clearly specify the features of the databases. None of the authors have any proprietary interests in these companies or products.

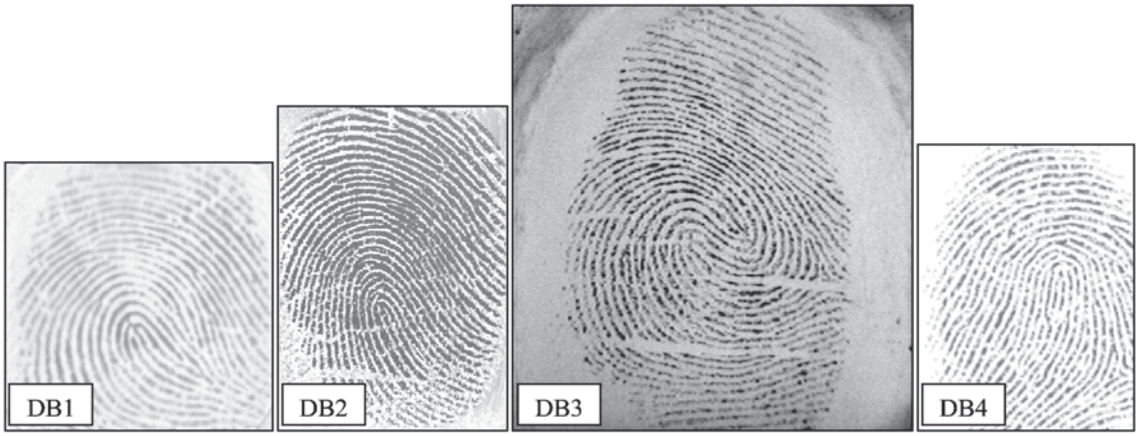

Fig. 1. Sample images taken from DB1, DB2, DB3, and DB4. In order to show the different image sizes of each database, the four images are displayed at the same scale factor.

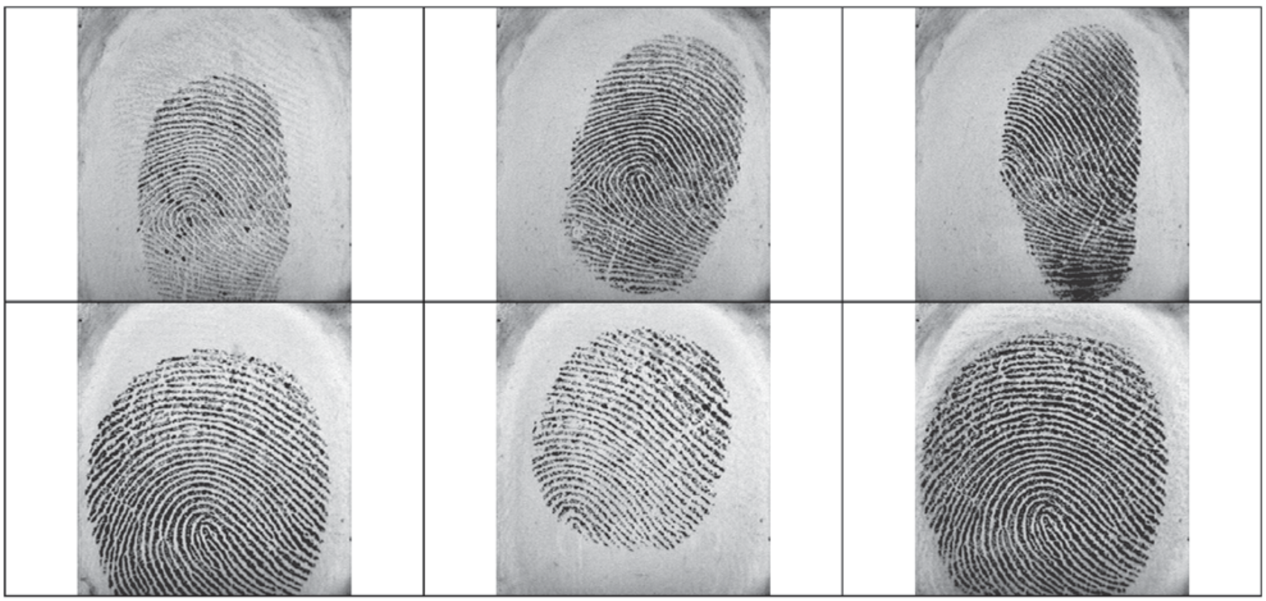

Fig. 2. Sample images from DB3; each row shows different impressions of the same finger.

However, in a technology evaluation (like, FVC2000), the aim is to capture the variability and the difficulties of the problem at hand and to investigate how the different algorithms deal with them. For this purpose, the size of our databases are adequate.

Table 2 summarizes the global features of the four databases, and Fig. 1 shows a sample image from each one of them.

It is worth emphasizing that the choice of providing more than one database is not aimed at comparing different acquisition technologies and devices; the results obtained by the algorithms on the different databases should not be conceived as a quality measure of the corresponding sensors, since the acquisition conditions and the volunteer crew of each database are different.

To summarize, DB1 and DB2 have the following features:

- The fingerprints are mainly from 20 to 30 year-old students (about 50 percent male).
- Up to four fingers were collected for each volunteer (forefinger and middle finger of both the hands).
- The images were taken from untrained people in two different sessions and no efforts were made to assure a minimum acquisition quality.
- All the images from the same individual were acquired by interleaving the acquisition of the different fingers (e.g., first sample of left forefinger, first sample of right fore finger, first sample of left middle, first sample of right middle, second sample of the left forefinger, ...).
- The presence of the fingerprint cores and deltas is not guaranteed since no attention was paid on checking the correct finger position on the sensor.
- The sensor platens were not systematically cleaned (as usually suggested by the vendors).
- The acquired fingerprints were manually analyzed to assure that the maximum rotation is approximately in the range $[-15°, 15°]$ and that each pair of impressions of the same finger has a nonnull overlapping area.

Database DB3 was collected as follows:

- The fingerprints are from 19 volunteers between the ages of five to 73 (55 percent male).
- One-third of the volunteers were over 55 years of age.
- One-third of the volunteers were under 18 years of age.
- One-sixth of the volunteers were under seven years of age (childrens fingerprints constitute an interesting case study, since the usable image area is small and the ridge-line density is high).
- Two images of up to six fingers (thumb, fore, and middle on left and right hands) were taken without interleaving

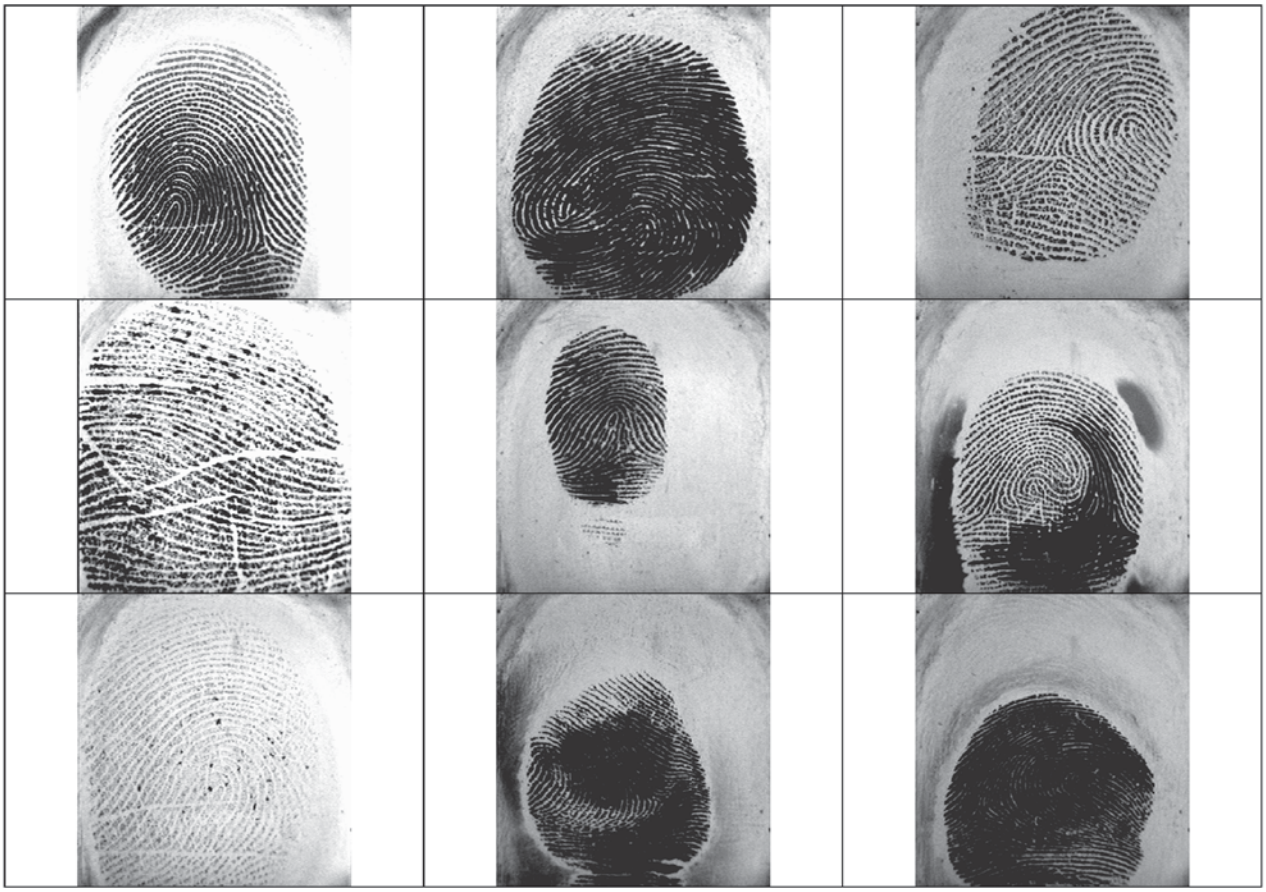

Fig. 3. Images from DB3; all the samples are from different fingers and are roughly ordered by quality (top-left: high quality, bottom-right: low quality).

from each volunteer at each session and no efforts were made to assure a minimum acquisition quality.

- Each volunteer was seen at four sessions, with no more than two sessions on any single day.
- The time gap between the first and last sessions was at least three days and as long as three months, depending upon volunteer.
- The sensor plate was systematically cleaned between image acquisitions.
- At one session with each volunteer, fingers were cleaned with rubbing alcohol and dried.
- Some part of the core was apparent in each image, but care was taken to avoid a complete overlap between consecutive images taken during a single session.
- The acquired fingerprints were manually analyzed to assure that the maximum rotation is approximately in the range $[-15°, 15°]$ and that each pair of impressions of the same finger has a nonnull overlapping area.

Figs. 2 and 3 show some sample images taken from DB3.

The collection of DB4 requires some explanation. In general, the use of artificial images for testing biometric systems is not considered to be the "best practice" [19]. Although, this may be the case for performance evaluation in real applications, we believe that in a technology evaluation event such as FVC2000, the use of synthetic images has three main advantages:

- It supplies images which are native to none of the participant algorithms, thus providing a fair comparison.
- Synthetic fingerprint databases can be created at a very low cost. Acquiring a large number of fingerprints for testing purposes may be problematic due to the great amount of time and resources required and to the privacy legislation which in some countries prohibits the diffusion of such personal information. Furthermore, once a database has been "used," its utility is limited since, for successive testing of algorithms, a new unknown database should be used.
- It is possible to adjust the database difficulty by tuning different kinds of perturbations (e.g., maximum amount of rotation and translation, and the amount of skin distortion).

If the generated artificial images were not a suitable simulation of real fingerprint patterns, the comparisons on the synthetic

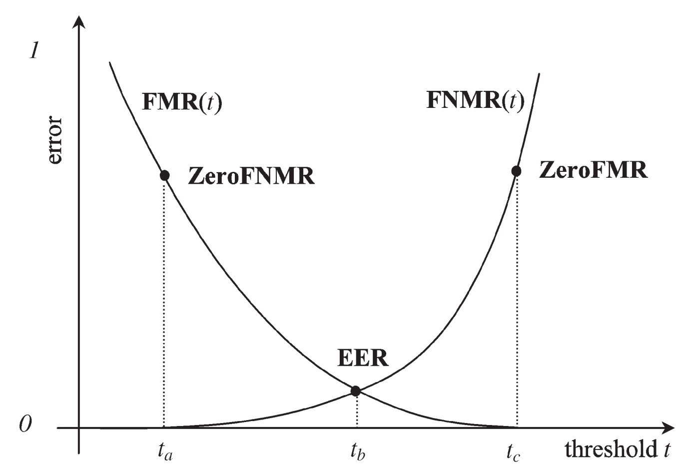

Fig. 4. An example of $\mathrm{FMR}/\mathrm{FNMR}$ curves, where the points corresponding to EER, $\mathrm{ZeroFMR}$, and $\mathrm{ZeroFNMR}$ are highlighted.

TABLE 3
Algorithm Performance over DB1 Sorted by $\mathbf{EER}$

Algorithm	EER (%)	REJ$_{\text{ENROLL}}$ (%)	REJ$_{\text{MATCH}}$ (%)	Avg Enroll Time (sec.)	Avg Match Time (sec.)
Sag1	0.67	0.00	0.00	2.48	0.96
Sag2	1.17	0.00	0.00	0.88	0.88
Cetp	5.06	0.00	0.00	0.81	0.89
Cwai	7.06	3.71	3.90	0.22	0.32
Cspn	7.60	0.00	0.00	0.17	0.17
Utwe	7.98	0.00	0.00	10.40	2.10
Krdl	10.66	6.43	6.59	1.00	1.06
Fpin	13.46	0.00	0.00	0.83	0.87
Uinh	21.02	1.71	5.08	0.53	0.56
Diti	23.63	0.00	0.00	0.65	0.72
Ncmi	49.11	0.00	0.12	1.13	1.34

TABLE 4
Algorithm Performance over DB2 Sorted by $\mathbf{EER}$

Algorithm	EER (%)	REJ$_{\text{ENROLL}}$ (%)	REJ$_{\text{MATCH}}$ (%)	Avg Enroll Time (sec.)	Avg Match Time (sec.)
Sag1	0.61	0.00	0.00	2.63	1.03
Sag2	0.82	0.00	0.00	0.93	0.93
Cspn	2.75	0.00	0.00	0.17	0.17
Cwai	3.01	1.29	1.29	0.23	0.30
Cetp	4.63	0.00	0.09	0.85	0.98
Krdl	8.83	3.29	4.41	1.16	2.88
Utwe	10.65	0.00	0.00	10.42	2.12
Fpin	11.14	0.00	0.00	1.16	1.24
Diti	13.83	0.00	0.00	1.21	1.28
Uinh	15.22	0.86	4.08	0.60	0.65
Ncmi	46.15	0.00	0.00	1.28	1.57

database would be misleading. Furthermore, in order to improve the performance, ad hoc algorithms could be designed/tuned according to the same assumptions which model the synthetic generation. However, the presence of three real databases in FVC2000 provides a natural way to check the validity of the results on DB4.

The parameters of the synthetic generator were tuned to emulate a low-cost sensor with a small acquisition area; the maximum rotation and displacement and skin-distortion are adjusted to roughly reproduce the perturbations in the three previous databases.

3 Performance Evaluation

For each database, we will refer to the *j*th fingerprint sample of the *i*th finger as F_{ij}, $i = 1 \dots 100$, and $j = 1 \dots 8$, and to the corresponding template (computed from F_{ij}) as T_{ij}.

For each database and for each algorithm:

- The templates T_{ij}, $i = 1 \dots 100$, and $j = 1 \dots 7$ are computed from the corresponding F_{ij} and stored on a disk; one of the following three kinds of rejection can happen for each image F_{ij}:
  1. **F** (Fail): the algorithm declares that it cannot enroll the fingerprint image.
  2. **T** (Timeout): the enrollment exceeds the maximum allowed time (15 seconds).
  3. **C** (Crash): the algorithm crashes during fingerprint processing.

  The three types of rejections are added and stored in $\mathbf{REJ}_{\mathrm{ENROLL}}$.
- Each fingerprint template T_{ij} is matched against the fingerprint images F_{ik} $(j < k \leq 8)$ and the corresponding *Genuine Matching Scores* $\mathbf{gms}_{ijk}$ are stored.[2] The number of matches (denoted as $\mathbf{NGRA}$ - Number of Genuine Recognition Attempts) is $((8 \times 7)/2) \times 100 = 2,800$ in case $\mathbf{REJ}_{\mathrm{ENROLL}} = 0$. The failed, timeout (five seconds) and crash rejections are accumulated into $\mathbf{REJ}_{\mathrm{NGRA}}$; no $\mathbf{gms}_{ijk}$ is stored in this case.
- Each fingerprint template T_{i1}, $i = 1 \dots 100$ is matched against the first fingerprint image from different fingers $F_{k1} (i < k \leq 100)$ and the corresponding *Impostor Matching Scores* $\mathbf{ims}_{ik}$ are stored. The number of matches (denoted as $\mathbf{NIRA}$ - Number of Impostor Recognition Attempts) is $((100 \times 99)/2) = 4,950$ in case $\mathbf{REJ}_{\mathrm{ENROLL}} = 0$. The failed, timeout (5 seconds) and crash rejections are accumulated into $\mathbf{REJ}_{\mathrm{NIRA}}$; no $\mathbf{ims}_{ik}$ is stored in this case.
- The genuine matching score distribution and the impostor matching score distribution are computed (actually, the

2. If g is matched with h, the symmetric match (i.e., h against g) is not executed.

TABLE 5
Algorithm Performance over DB3 Sorted by **EER**

Algorithm	EER (%)	REJ_{ENROLL} (%)	REJ_{MATCH} (%)	Avg Enroll Time (sec.)	Avg Match Time (sec.)
Sag1	3.64	0.00	0.00	5.70	2.13
Sag2	4.01	0.00	0.00	1.94	1.94
Cspn	5.36	0.57	1.24	0.35	0.36
Cetp	8.29	0.00	0.00	1.49	1.66
Cwai	11.94	12.86	8.00	0.46	0.57
Krdl	12.20	6.86	5.12	1.48	1.60
Uinh	16.32	10.29	7.64	1.28	1.36
Utwe	17.73	0.00	0.00	10.44	2.31
Diti	22.63	0.00	0.00	2.59	2.67
Fpin	23.18	0.00	0.00	2.13	2.19
Ncmi	47.43	0.00	0.01	2.25	2.75

TABLE 6
Algorithm Performance over DB4 Sorted by **EER**

Algorithm	EER (%)	REJ_{ENROLL} (%)	REJ_{MATCH} (%)	Avg Enroll Time (sec.)	Avg Match Time (sec.)
Sag1	1.99	0.00	0.00	1.90	0.77
Sag2	3.11	0.00	0.00	0.69	0.69
Cspn	5.04	0.00	0.00	0.11	0.11
Cwai	6.30	0.00	0.00	0.16	0.20
Cetp	7.29	0.00	0.00	0.65	0.72
Krdl	12.08	10.86	10.24	0.70	0.79
Fpin	16.00	0.00	0.00	0.77	0.80
Diti	23.80	0.00	0.00	0.52	0.60
Utwe	24.59	0.00	0.00	10.42	4.17
Uinh	24.77	2.14	4.28	0.42	0.45
Ncmi	48.67	0.00	0.25	1.08	1.19

term "distribution" denotes a histogram) and graphically reported to show how the algorithm "separates" the two classes. In fingerprint verification, higher scores are associated with more closely matching images.

- The $\mathbf{FMR}(t)$ (False Match Rate) and $\mathbf{FNMR}(t)$ (False NonMatch Rate) curves are computed from the above distributions for t ranging from 0 to 1.[3] Given a threshold t, $\mathbf{FMR}(t)$ denotes the percentage of $\mathbf{ims}_{ik} \geq t$, and $\mathbf{FNMR}(t)$ denotes the percentage of $\mathbf{gms}_{ijk} < t$. Actually, since **FMR** and **FNMR** are used in the contest to compare the performance of different algorithms, **FMR** and **FNMR** are "corrected" to keep into account rejections stored in $\mathbf{REJ_{NIRA}}$ and $\mathbf{REJ_{NGRA}}$:

$$\mathbf{FMR}(t) = \frac{card\{\mathbf{ims}_{ik} \mid \mathbf{ims}_{ik} \geq t\}}{\mathbf{NIRA}},$$

$$\mathbf{FNMR}(t) = \frac{card\{\mathbf{gms}_{ijk} \mid \mathbf{gms}_{ijk} < t\} + \mathbf{REJ_{NGRA}}}{\mathbf{NGRA}},$$

where *card* denote the cardinality of a given set. This correction assumes that a failure to match is always treated by the system as a "nonmatch" (matching score < 0).

- A **ROC** (Receiving Operating Curve) is obtained, where pairs ($\mathbf{FMR}(t)$, $\mathbf{FNMR}(t)$) are plotted for the same value of t; in particular, for $t = 0$, $\mathbf{FMR} = 1$, and $\mathbf{FNMR} = 0$, while for $t > 1$, $\mathbf{FMR} = 0$, and $\mathbf{FNMR} = 1$. The **ROC** curve is drawn in log-log scales for better comprehension.
- The Equal Error Rate **EER** is computed as the point where $\mathbf{FNMR}(t) = \mathbf{FMR}(t)$ (see Fig. 4); in practice, the matching score distributions (histograms) are not continuous and a crossover point might not exist. In this case, we report the interval [$\mathbf{EER}_{low}$, $\mathbf{EER}_{high}$]. In Appendix A, an operational definition of **EER** is given.
- **ZeroFMR** is defined as the lowest **FNMR** at which no False Matches occur and **ZeroFNMR** is defined as the lowest **FMR** at which no False NonMatches occur (Fig. 4):

$$\mathbf{ZeroFMR}(t) = \min_t t\{\mathbf{FNMR}(t) \mid \mathbf{FMR}(t) = 0\}$$

$$\mathbf{ZeroFNMR}(t) = \min_t t\{\mathbf{FMR}(t) \mid \mathbf{FNMR}(t) = 0\}$$

Both **ZeroFMR** and **ZeroFNMR** may not exist; in such a case, we assign to them the value 1.
- The **average enroll time** is calculated as the average CPU time for a single enrollment operation, and **average match time** as the average CPU time for a single match operation between a template and a test image.

3. **FMR** and **FNMR** are often referred as **FAR** (False Acceptance Rate) and **FRR** (False Rejection Rate) respectively, but the FAR/FRR notation is misleading in some applications. For example, in a welfare benefits system, which uses fingerprint identification to prevent multiple payments under false identity, the system "falsely accepts" an applicant if his/her fingerprint is "falsely rejected"; otherwise, a "false acceptance" causes a "false rejection."

4 RESULTS

This section reports the performance of the tested algorithms on each of the four databases (Tables 3, 4, 5, and 6) and the average

TABLE 7
Average Performance over the Four Databases Sorted by Avg EER

Algorithm	Avg EER (%)	Avg $\mathbf{REJ}_{\mathrm{ENROLL}}$ (%)	Avg $\mathbf{REJ}_{\mathrm{MATCH}}$ (%)	Avg Enroll Time (sec.)	Avg Match Time (sec.)
Sag1	1.73	0.00	0.00	3.18	1.22
Sag2	2.28	0.00	0.00	1.11	1.11
Cspn	5.19	0.14	0.31	0.20	0.20
Cetp	6.32	0.00	0.02	0.95	1.06
Cwai	7.08	4.46	3.14	0.27	0.35
Krdl	10.94	6.86	6.52	1.08	1.58
Utwe	15.24	0.00	0.00	10.42	2.67
Fpin	15.94	0.00	0.00	1.22	1.27
Uinh	19.33	3.75	5.23	0.71	0.76
Diti	20.97	0.00	0.00	1.24	1.32
Ncmi	47.84	0.00	0.09	1.44	1.71

results over the four databases (Table 7). Fig. 5 shows the $\mathbf{ROC}$ for DB3, which proved to be the most difficult data set. The notation introduced in Section 3 is used in both the graphics and tables, with the only exception of reporting $\mathbf{REJ}_{\mathrm{ENROLL}}$ as a percentage value and to collapse both $\mathbf{REJ}_{\mathrm{NGRA}}$ and $\mathbf{REJ}_{\mathrm{NIRA}}$ into a single value $\mathbf{REJ}_{\mathrm{MATCH}}$:

$$\mathbf{REJ}_{\mathrm{MATCH}} = \frac{\mathbf{NIRA} \cdot \mathbf{REJ}_{\mathrm{NIRA}} + \mathbf{NGRA} \cdot \mathbf{REJ}_{\mathrm{NGRA}}}{\mathbf{NIRA} + \mathbf{NGRA}}.$$

For a correct interpretation of the results, $\mathbf{EER}$ alone is not a sufficient metric; $\mathbf{REJ}_{\mathrm{ENROLL}}$ should be also taken into account. For each algorithm, detailed results (including genuine and impostor distributions, $\mathbf{FMR}$ and $\mathbf{FNMR}$ curves, $\mathbf{NGRA}$, $\mathbf{NIRA}$, ...) are reported in [13]. Due to lack of space, Appendix B of this paper presents only detailed results of the SAG1 algorithm which had the best accuracy in our competition.

5 Conclusions

Most of the algorithms submitted to the competition performed well, if we take into account the difficulty of adapting a given

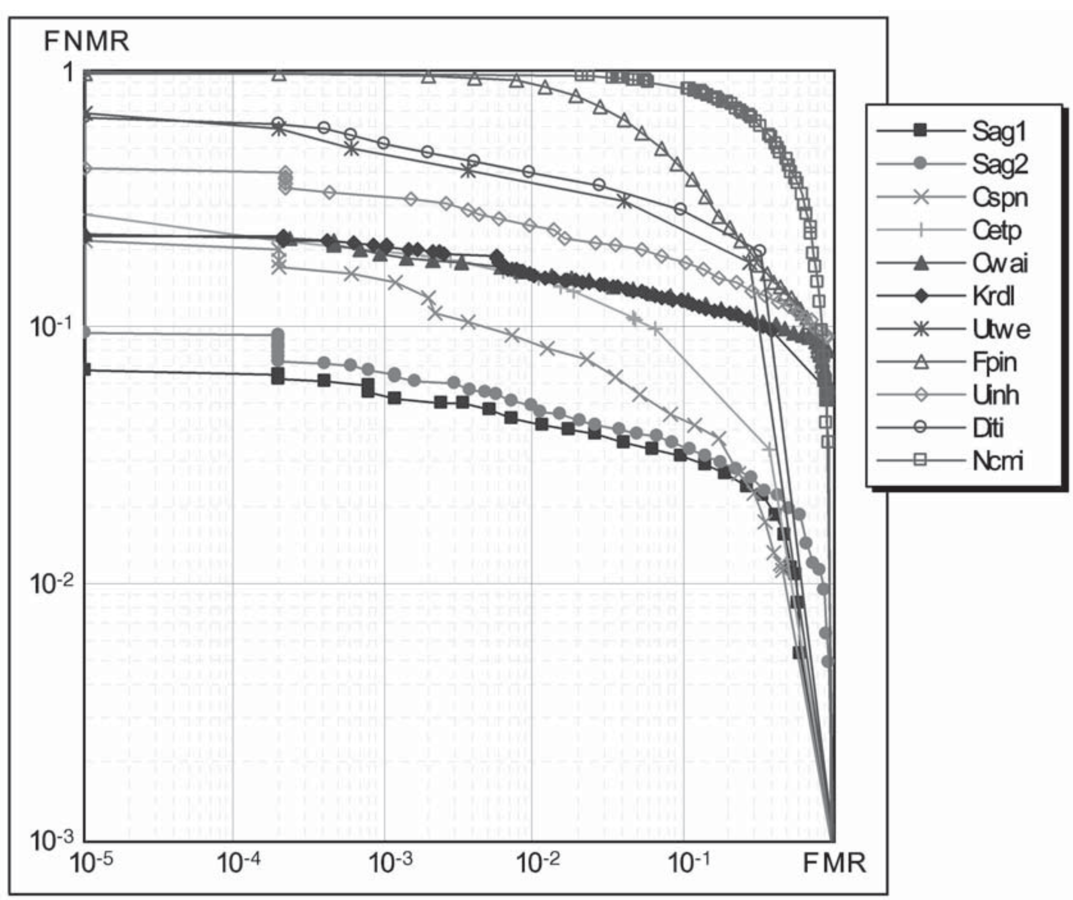

Fig. 5. $\mathbf{ROC}$ curves on DB3. Each point denotes a pair ($\mathbf{FMR}(t)$, $\mathbf{FNMR}(t)$) for a given value of t.

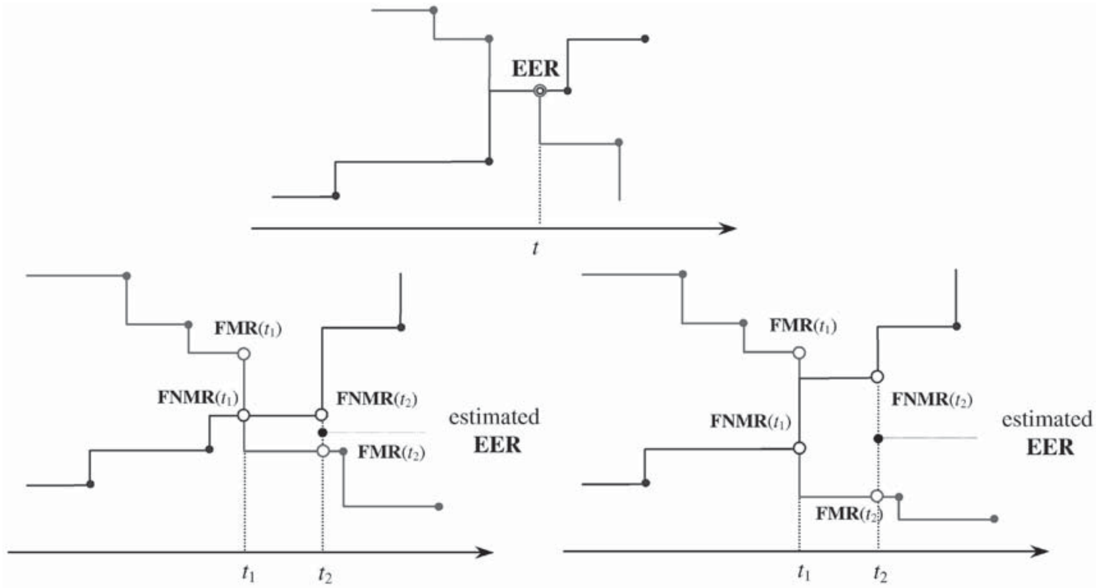

Fig. 6. Computing the $\mathbf{EER}$ interval. On the top an example is given where an $\mathbf{EER}$ point exists. On the bottom, two cases are shown where an $\mathbf{EER}$ point does not exist and the corresponding intervals are highlighted.

Average enroll time: 2.48 seconds
Average match time: 0.96 seconds

$\mathbf{REJ_{ENROLL}}$	**NGRA**	**NIRA**	$\mathbf{REJ_{NGRA}}$	$\mathbf{REJ_{NIRA}}$
0.00% (F:0 T:0 C:0)	2800	4950	0.00% (F:0 T:0 C:0)	0.00% (F:0 T:0 C:0)

EER	**EER***	**ZeroFMR**	**ZeroFNMR**
0.67% (0.67%-0.68%)	0.67% (0.67%-0.68%)	2.11%	53.13%

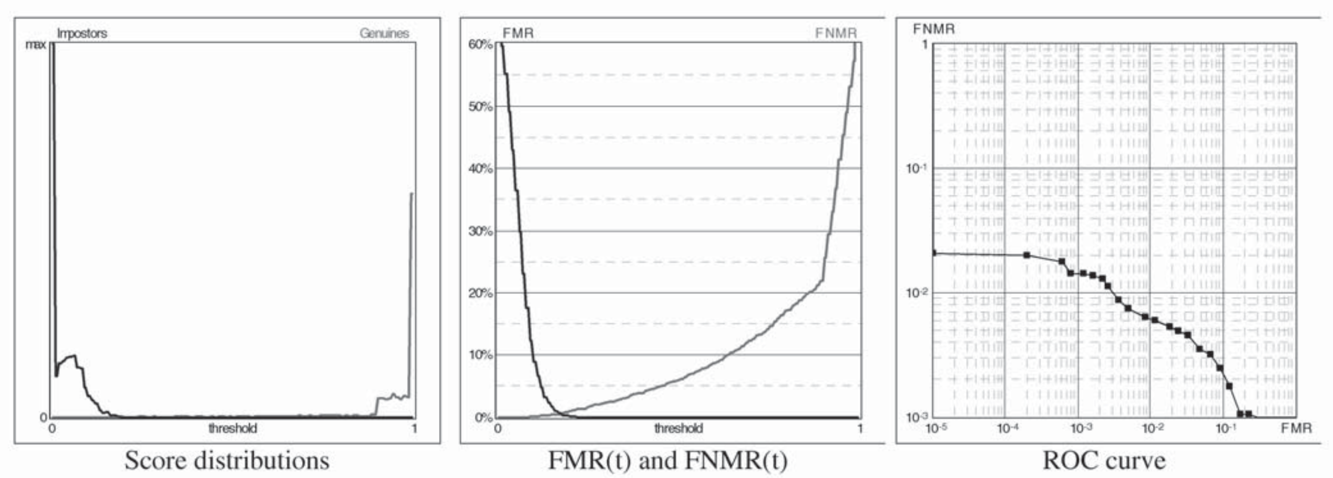

Fig. 7. Performance of algorithm $Sag1$ on database $DB1\_A$.

algorithm to new types of images. In particular, algorithms $Sag1$ and $Sag2$ showed the best accuracy and $Cspn$ exhibited a good trade-off between accuracy and efficiency.

Table 7 highlights a significant gap in the performance of the different algorithms and it would be extremely interesting to understand the reasons for such differences. To this purpose, after the presentation of the results, we asked the participants to provide some technical details about their methods, but only a few of them responded (the responses can be found at the FVC2000 Web site [7]); in any case, on the basis of the participant responses and on what we learned from this experience, we can make the following observations:

- A coarse analysis of the errors on genuine attempts showed that most of the errors were made by the algorithms on about 15-20 percent poor-quality fingerprints in each database. In other words, we could claim that a 20-80 rule is valid: that is, 20 percent of the database is responsible for 80 percent of the errors.
- The most accurate algorithm ($Sag1$) takes a lot of time for enrollment (3.18 sec with respect to a median enrollment time of 1.08 sec). This suggests that an accurate image

Average enroll time: 2.63 seconds
Average match time: 1.03 seconds

REJ_{ENROLL}	NGRA	NIRA	REJ_{NGRA}	REJ_{NIRA}
0.00% (F:0 T:0 C:0)	2800	4950	0.00% (F:0 T:0 C:0)	0.00% (F:0 T:0 C:0)

EER	EER*	ZeroFMR	ZeroFNMR
0.61%	0.61%	1.36%	50.69%

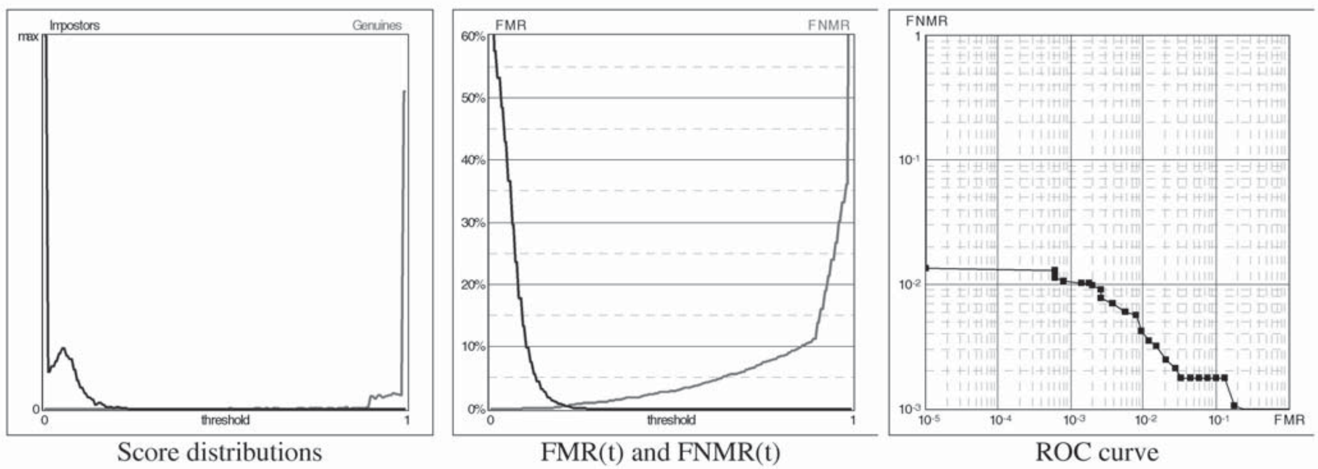

Fig. 8. Performance of algorithm $Sag1$ on database $DB2\_A$.

Average enroll time: 5.70 seconds
Average match time: 2.13 seconds

REJ_{ENROLL}	NGRA	NIRA	REJ_{NGRA}	REJ_{NIRA}
0.00% (F:0 T:0 C:0)	2800	4950	0.00% (F:0 T:0 C:0)	0.00% (F:0 T:0 C:0)

EER	EER*	ZeroFMR	ZeroFNMR
3.64%	3.64%	6.82%	100.00%

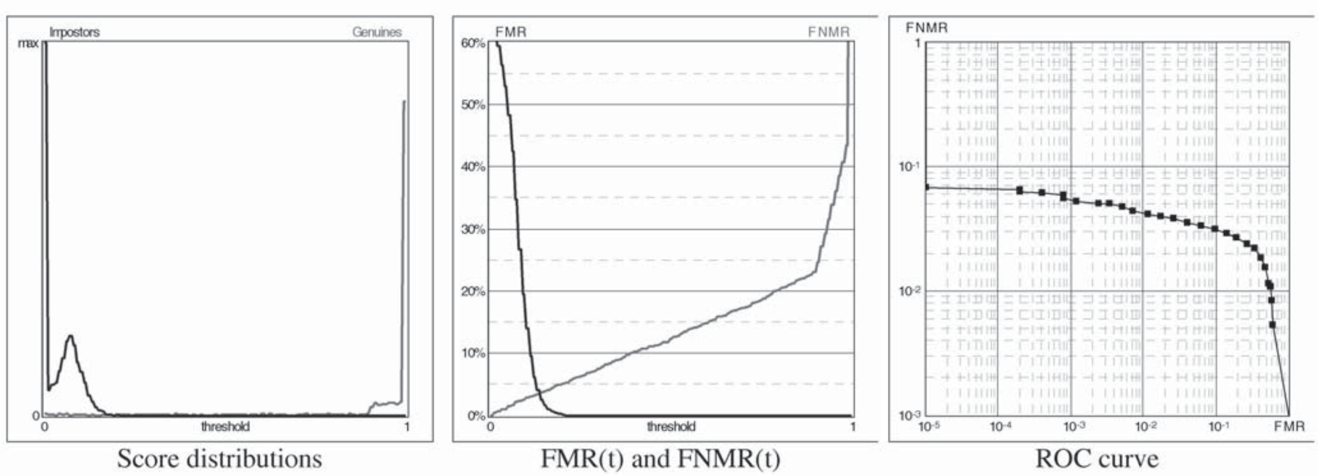

Fig. 9. Performance of algorithm $Sag1$ on database $DB3\_A$.

enhancement and feature extraction is really important for improving the matching accuracy. Furthermore, feature extraction seems to perform asymmetrically, since the average matching time (which also includes the feature extraction time for the test image) is substantially lower than a single enrollment time.

- The fastest algorithms ($Cspn$) extracts minutiae by an adaptive tracing of the gray-level ridges, without a priori binarization and thinning (which are time consuming tasks) [14], [25] and exploits local minutiae arrangement to speed-up the initial steps of minutiae matching [26].

Average enroll time: 1.90 seconds
Average match time: 0.77 seconds

REJ$_{ENROLL}$	NGRA	NIRA	REJ$_{NGRA}$	REJ$_{NIRA}$
0.00% (F:0 T:0 C:0)	2800	4950	0.00% (F:0 T:0 C:0)	0.00% (F:0 T:0 C:0)

EER	EER*	ZeroFMR	ZeroFNMR
1.99% (1.98%-2.00%)	1.99% (1.98%-2.00%)	6.71%	100.00%

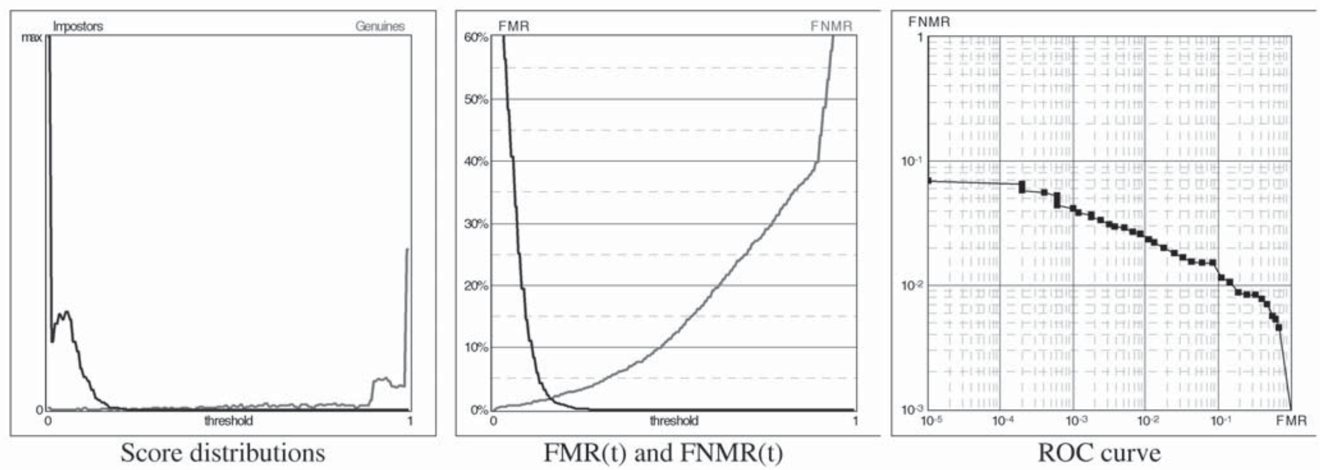

Fig. 10. Performance of algorithm $Sag1$ on database $DB4\_A$.

Databases DB1 and DB2 proved to be "easier" than DB3, even though the sensor used for DB3 is of higher quality. This means that the acquisition conditions and the volunteer population can have a stronger impact on the performance than sensor quality.

The synthetically-generated database (DB4) was demonstrated to be adequate for FVC2000 purposes: in particular, from Tables 3, 4, 5, and 6, it is evident that the algorithm ranking on DB4 is quite similar to the other databases, proving that no algorithm was favored or penalized by the synthetic images. In particular, if an algorithm performs well on real fingerprints, then it also performs well on synthetic fingerprints and vice versa. The visual analysis of impostor and genuine distributions (see [13]) definitely supports this claim, since no significant differences are seen between the DB4 graphics and the others.

Once again we would like to remark that the results reported here do not necessarily reflect the performance that the participating algorithms would achieve in a real environment or when embedded into a complete biometric system. In any event, we believe that FVC2000 results:

- provide a useful overview of the state-of-the-art in this field,
- allow researchers and companies to test their algorithms over common databases collected using state-of-the-art sensors, and
- provide guidance to the participants for improving their algorithms.

In future, we intend to continue supporting this initiative as follows:

- The existing FVC2000 Web site [7] will be maintained to diffuse FVC2000 results and to promote FVC2000 testing protocol as a standard for technological evaluations.
- Companies and academic research groups will be allowed to test new algorithms or improved versions of existing algorithms on the FVC2000 benchmark databases and to add their results to the FVC2000 Web site. New entries will be kept isolated from the original entries, since hereafter, the full databases are known, in advance, which could allow algorithm tuning to give unfair advantage to new participants.
- The second Fingerprint Verification Competition (FVC2002) has been scheduled and its results will be presented at the 16th International Conference of Pattern Recognition.
- Generating synthetic fingerprint databases for future evaluations will be further investigated.

APPENDIX A

An operational procedure for computing **EER** (interval), given a finite number of genuine and impostor matching scores, is reported in the following. Let

$$t_1 = \underset{t \in \{\mathbf{gms}_{ijk}\} \cup \{\mathbf{ims}_{ik}\}}{max} \{t \mid \mathbf{FNMR}(t) \leq \mathbf{FMR}(t)\},$$

and

$$t_2 = \underset{t \in \{\mathbf{gms}_{ijk}\} \cup \{\mathbf{ims}_{ik}\}}{min} \{t \mid \mathbf{FNMR}(t) \geq \mathbf{FMR}(t)\}.$$

The **EER** interval is defined as:

$$[\mathbf{EER}_{low}, \mathbf{EER}_{high}] = \begin{cases} [\mathbf{FNMR}(t), \mathbf{FMR}(t_1)] & if\ \mathbf{FNMR}(t_1) + \mathbf{FMR}(t_1) \leq \mathbf{FMR}(t_2) + \mathbf{FNMR}(t_2) \\ [\mathbf{FMR}(t_2), \mathbf{FNMR}(t_2)] & otherwise \end{cases}$$

and **EER** is estimated as $(\mathbf{EER}_{low} + \mathbf{EER}_{high})/2$ (see Fig. 6).

APPENDIX B

Please see Figs. 7, 8, 9, and 10.

REFERENCES

[1] J. Barron, D. Fleet, and S. Beauchermin, "Systems and Experiment: Performance of Optical Flow Techniques,," *Int'l J. Computer Vision,* vol. 12, no. 1, pp. 43-77, 1994.

[2] BIOTEST: European Esprit Project No: 21978 (http://www.npl.co.uk/npl/sections/this/biotest/index.html), 1998.

[3] R. Cappelli, A. Erol, D. Maio, and D. Maltoni, "Synthetic Fingerprint-Image Generation," *Proc. 15th Int'l Conf. Pattern Recognition (ICPR 2000),* pp. 475-478, Sept. 2000.

[4] R. Cappelli, D. Maio, and D. Maltoni, "Combining Fingerprint Classifiers," *Proc. First Int'l Workshop Multiple Classifier Systems (MCS2000),* pp.351-361, June 2000.

[5] R. Cappelli, D. Maio, and D. Maltoni, "Modelling Plastic Distortion in Fingerprint Images," *Proc. Second Int'l Conf. Advances in Pattern Recognition (ICAPR 2001),* pp. 369-376, Mar. 2001.

[6] C. Dorai, N.K. Ratha, and R.M. Bolle, "Detecting Dynamic Behaviour in Compressed Fingerprint Videos: Distortion," *Proc. Computer Vision and Pattern Recognition, (CVPR 2000),* vol. II, pp. 320-326, June 2000

[7] FVC2000 Web site: http://bias.csr.unibo.it/2000. Sept. 2000.

[8] A. Hoover et al., "An Experimental Comparison of Range Image Segmentation Algorithms," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 18, no. 7, pp. 673-689, July 1996.

[9] *Biometrics—Personal Identification in Networked Society.* A.K. Jain, R. Bolleand, S. Pankanti, eds. Kluwer Academic, 1999.

[10] A.K. Jain, S. Prabhakar, and A. Ross, "Fingerprint Matching: Data Acquisition and Performance Evaluation," MSU Technical Report TR99-14, 1999.

[11] H.C. Lee and R.E. Gaensslen, "Advances in Fingerprint Technology," 1991.

[12] J. Matas et al., "Comparison of Face Verification Results on the XM2VTS Database," *Proc. 15th Int'l Conf. Pattern Recognition,* vol. 4, pp. 858-863, Sept. 2000.

[13] D. Maio, D. Maltoni, R. Cappelli, J.L. Wayman, and A.K. Jain, "FVC2000: Fingerprint Verification Competition," DEIS Technical Report BL-09-2000, Univ. of Bologna; available online at: http://bias.csr.unibo.it/fvc. 2000.

[14] D. Maio and D. Maltoni, "Direct Gray-Scale Minutiae Detection in Fingerprints," *IEEE Trans. Pattern Analysis Machine Intelligence,* vol. 19, no. 1, pp. 27-40, 1997.

[15] P.J. Phillips, H. Moon, S.A. Rizvi, and P.J. Rauss, "The FERET Evaluation Methodology for Face-Recognition Algorithms," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 10, pp. 1090-1104, Oct. 2000.

[16] P.J. Phillips, A. Martin, C.L. Wilson, and M. Przybocky, "An Introduction to Evaluating Biometric Systems," *Computer,* Feb. 2000,

[17] P.J. Phillips and K.W. Bowyer, "Introduction to Special Section on Empirical Evaluation of Computer Vision Algorithms," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 21, no. 4, pp. 289-290, Apr. 1999.

[18] T. Randen and J.H. Husoy, "Filtering for Texture Classification: A Comparative Study," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 21, no. 4, pp. 291-310, Apr. 1999.

[19] "Biometrics Working Group, Best Practices in Testing and Reporting Performance of Biometric Devices,"UK Government's HYPERLINK http://www.afb.org.uk/bwg/index.html, Jan. 2000. (available online at: http://www.afb.org.uk/bwg/bestprac10.pdf).

[20] C.I. Watson and C.L. Wilson, *NIST Special Database 4, Fingerprint Database.* US Nat'l Institute of Standards and Technology, 1992.

[21] C.I. Watson, *NIST Special Database 14, Fingerprint Database.* US Nat'l Inst. Standards and Technology, 1993.

[22] C.I. Watson, "NIST Special Standard Reference Database 24, NIST Digital Video of Live-Scan Fingerprint Database," *US Nat'l Inst. Standards and Technology,* 1998.

[23] J.L. Wayman, "Technical Testing and Evaluation of Biometric Devices," *Biometrics Personal Identification in Networked Society,* A. Jain, et al. eds., Kluwer Academic, 1999.

[24] J.L. Wayman, "The Philippine AFIS Benchmark Test," *Nat'l Biometric Test Center Collected Works, 1997-2000,* Sept. 2000, (available online at: http://www.engr.sjsu.edu/biometrics/collected).

[25] X. Jiang et al., "Minutiae Extraction by Adaptive Tracing the Gray Level Ridge of the Fingerprint Image," *Proc. IEEE Int'I Conf. Image Processing (ICIP '99),* 1999.

[26] X. Jiang et al., "Fingerprint Minutiae Matching Based on the Local And Global Structures," *Proc. 15th Int'l Conf. Pattern Recognition,* Sept. 2000.

A Feature-Based Technique for Joint, Linear Estimation of High-Order Image-to-Mosaic Transformations: Mosaicing the Curved Human Retina

Ali Can, Charles V. Stewart, *Member, IEEE*, Badrinath Roysam, *Member, IEEE*, and Howard L. Tanenbaum

Abstract—An algorithm for constructing image mosaics from multiple, uncalibrated, weak-perspective views of the human retina is presented and analyzed. It builds on a previously described algorithm for registering pairs of retinal images using a noninvertible, 12-parameter, quadratic image transformation model and a hierarchical, robust estimation technique. The major innovation presented here is a linear, feature-based, noniterative method for jointly estimating consistent transformations of all images onto the mosaic "anchor image." Constraints for this estimation are derived from pairwise registration both directly with the anchor image and indirectly between pairs of nonanchor images. An incremental, graph-based technique constructs the set of registered image pairs used in the joint solution. The joint estimation technique allows images that do not overlap the anchor frame to be successfully mosaiced, a particularly valuable capability for mosaicing images of the retinal periphery. Experimental analysis of the algorithm on data sets from 16 eyes shows the average overall median transformation error in final mosaic construction to be 0.76 pixels. Overall, the technique is simpler, more accurate, and offers broader coverage than previously published methods.

Index Terms—Robust estimation, image mosaic, image montage, transformation estimation, retinal imaging, joint estimation.

1 INTRODUCTION

BUILDING a mosaic image from a sequence of partial views is a powerful means of obtaining a broader view of a scene than is available with a single view. Research on automated mosaic construction has been underway for at least 25 years (see early work in [17], [18], [20], [29]). The range of applications includes panoramic image formation [22], [21], virtual reality [8], [28], image compression [15], [14], change detection, superresolution [7], tracking [10], navigation [11], indexing and key frame identification [15], document compositing [31], and cartography [30]. One application domain in which mosaics are particularly valuable is in the diagnosis and treatment of diseases of the retina [1], [3], [9], [16]. A seamless mosaic formed from multiple fundus camera images aids in diagnosis, provides a means for monitoring the progression of diseases, and may be used as a *spatial map* during surgical treatment [1], [3].

Several issues must be addressed in designing a mosaic construction technique. First, the coordinate system in which to build the mosaic must be established. Second, an appropriate mathematical model of the image-to-mosaic transformation must be developed. Most work in mosaic construction uses low-order,

- *A. Can and B. Roysam are with the Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180-3590. E-mail: alican@alum.rpi.edu, roysab@rpi.edu.*
- *C.V. Stewart is with the Department of Computer Science, Rensselaer Polytechnic Institute, Troy, NY 12180-3590. E-mail: stewart@cs.rpi.edu.*
- *H.L. Tanenbaum is with The Center for Sight, 349 Northern Blvd., Albany, NY 12204. E-mail: how1@albany.net.*

Manuscript received 25 July 2000; revised 17 May 2001; accepted 25 July 2001.
Recommended for acceptance by Y.-F. Wang.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 112587.