Proceedings of the 2002 IEEE
Workshop on Information Assurance and Security
United States Military Academy, West Point, NY, 17–19 June 2002

Investigating Performance and Impacts on Fingerprint Recognition Systems

Michael Arnold, Christoph Busch, Heinrich Ihmor

Abstract—
This paper presents a comparative study on fingerprint recognition systems. The goal of this study was to investigate the capability characteristics of biometric systems regarding integration of biometric features in personnel documents such as IDcards and Visa application documents. Thus the designed test has the focus on performance testing of selected algorithms and systems with dedicated investigations on side effects such as independence of matching rates and results from the scanning device or the impacts of ageing effects on the received operator characteristics. The study was carried out in close collaboration between German Federal Criminal Police Office (Bundeskriminalamt, BKA), the German Federal Office for Information Security (Bundesamt fuer Sicherheit in der Informationstechnik, BSI) and the Fraunhofer-IGD.

I. Introduction

As a biological characteristics, analysis of fingerprints have had a long tradition [1] and are a synonym for automated recognition of individuals. Up until recently, it was only the resulting fingerprint image that was exclusively used as an identification [2] feature in criminal investigation; Human fingerprints were almost solely used for forensic purposes in dactyloscopy. Dactyloscopists examine fingerprints with regard to details that can be used to identify people. Evidence of a fingerprint found at a scene of a crime can thus be linked to a person as the one who left that trace. Classification of fingerprints can be based on the categorization of patterns into various finger classes exploiting the fact that due to the ridge flow so-called patterns (loops, arches, whorls) are formed and furthermore extracting ridge end points of ridge bifurcation points as anatomic characteristics (minutiae). Thanks to the rich entropy of fingerprints individual prints provide large information to a dactyloscopic expert to distinguish, whether individual fingerprints stem from the same source or not.

Nowadays international criminal investigation offices are using standardized data formats [3] to support the exchange of fingerprint images among distributed Automated Fingerprint Identification Systems (AFIS). Furthermore biometric fingerprints recognition sensors are well suited for convenience applications in the consumer market in mobile phones or Personal Digital Assistants (PDAs). The sensors are in general extremely small in their size and inexpensive to produce and thus well suited for mass production.

In the discussion on anti-terror measures the German legislation [4] paved the roaded for the inclusion of further biometrical features in German ID cards and passports, in order to verify the identity of an ID card holder. In the meantime the International Civil Aviation Organization (ICAO) and the International Organization for Standardization (ISO) have established standards to allow interoperability of such ePassports. For the member states of the European Union ePassports will include both facial images [5] and fingerprint images [6] in interoperable formats [7]. The integration of fingerprints in European ePassports shall be used to verify the authenticity of such documents.This is the background of the study *BioFinger - Evaluation of Fingerprint Recognition Technologies*, which analyzed the characteristics of fingerprint recognition systems. Hence, the very aim of the BioFinger is the verification, i.e. the examination of the identity claimed by the person (1 : 1 comparison). On the other hand, the identification application, (1 : n-comparison), with which a person is to be identified by comparing him/her with n number of people in a database, is out of the scope of the project.

Within this context, a number of investigations are carried out, which are to clarify the suitability of some chosen products. The question is this: Using today's systems or components, what are the achievable verification accuracies, or can they be increased by assembling of the different components to a fingerprint recognition system? As such the independence of matching rates and results from the scanning device is of high importance. For the potential operator as replacement of components in large numbers is cost-intensive and thus not feasible. In contrary to International Competitions for Fingerprint Verification Algorithms ([8],[9]) only those system were considered in this study that are commercially available on the German market and do provide adequate product support..

Furthermore due to the specific demands on passport, i.e. current period of validity of ten years, the longevity of fingerprints with regard to their characteristic to verify people, is very significant. This was investigated by examining the influence of the ageing process on the algorithm performance. This research was supported by a special

M. Arnold: Fraunhofer IGD, Darmstadt, Germany
C. Busch: Fraunhofer IGD, Darmstadt, Germany
H. Ihmor: Federal Office for Information Security BSI, Bonn, Germany

ISBN 0-xxxx-xxxx-x/$xx.xx

database compiled by the German Federal Criminal Police Office (BKA).

The paper is structured as follows. The second section describes the collection process for the different databases of fingerprints of the different systems. The third section elaborates on objective evaluation criteria of biometric systems as well as the relation to application scenarios. The fourth section summarizes the results of the above-mentioned examinations for various sensors and algorithms. The fifth section describes the research performed regarding ageing using the fingerprint image database provided by the German Federal Criminal Police Office (BKA). The conclusion summarizes the investigations performed and provides an outlook for future research directions.

II. Databases of Fingerprints

A. Data Collection

The data collection in this study was conducted on three independent sessions. For all three segments of the data base no synthetic fingerprints were accepted [10], [11]. In order to address the investigation target - as described in the previous section - volunteers are expected to provide the impression to all available sensors.

A total of $N_P = 30$ people took part in the test. The conditions for taking fingerprints are found below:

- Each of the N_P test persons had fingerprints taken for each of the N_S =11 test sensors.
- Four fingerprints of both hands of each test person were taken (all fingers of one hand with the exception of the small finger), $N_{FP} = 8$.
- The fingerprints of the test persons were taken in a total of $N_D = 3$ sessions.
- For each sensor $N_A = 3$ transactions were taken per session, i.e. there was a total of $N_{AF} = 9$ images per finger.

Every single database for the various sensors thus contains $N_{FDB} = N_P \times N_{FP}$ (e.g. $N_{FDB} = 240$) different fingers. In each procedure, a total of $N_{Session} = N_{FDB} \times N_A$ (e.g. $N_{Session} = 720$) fingerprints were acquired. In turn each database contains a total of $N_{FADB} = N_{Session} \times N_D$ (e.g. $N_{FA} = 2160$) fingerprints.

The data acquisition was conducted at various milestones in the study period thus statistical independence of the probes is improved and potential impacts of varying environmental conditions is included.

B. Data Acquisition Errors and Database Cleanup

Naturally, when acquiring a large number of fingerprint images, a range of errors can occur. Firstly, if the acquisition software is not handled correctly, low-quality images are generated, for example if an image is acquired even though the finger has not been fully placed on the sensor.

Following the enrollment procedure, this kind of errors can be detected if such images, which prove to have a Failure to Enroll (FTE) error for the algorithms Nos. 1 to 7 (see Table II-B), are examined by experienced analysts.

Vendor	Used in connection with sensor
Dermalog	Only algorithm
IKENDI	AT77C101B
IDENCOM	BACU-100
IdentTechnologies	AT77C101B
NEC	Only algorithm
Neurotechnologija	TST BiRD IIi
Siemens	IDMouse, TouchChip

TABLE I
Algorithms evaluated in BioFinger

In some cases the sensor could not acquire images of the finger of a certain person, which contributes to the Failure to Acquire (FTA) rate.

For this purpose the matching of identical fingers was performed with all algorithms. Fingerprints with very low scores for matching identical fingers were examined and corresponding images sorted out.

Technology	Name	Image size	dpi
pressure	BACU-100	256×384	440
capacitive	TouchChip(TM)	256×360	508
capacitive	ID Mouse	224×288	513
capacitive	AES4000	193×193	250
optical	TFS 050	576×744	500
optical	TST BiRD IIi	320×384	500
optical	ACCO USB	376×472	500
optical	FX 2000	316×376	569
optical	LS2(TM)/F	740×580	500
optical	MorphoSmart(TM) MSO100	416×416	500
thermal	FingerChip(TM) AT77C101B	280×320	280

TABLE II
Sensors evaluated in BioFinger

Other than sensors Nos. 1 to 11 (see Table II-B), fingerprint images of the database entitled *Sensor 13* were provided by the Federal Office of Criminal Investigation (BKA). This database will be described in more detail in section V-A.

ISBN 0-xxxx-xxxx-x/$xx.xx

III. Objective Evaluation Metrics

A. Description of the Evaluation Criteria

Different types of error rates are used as metrics for the operative capability of biometric authentication systems in general and for fingerprint image recognition systems [12] in particular[11].

The result of a comparison in the feature matcher within a fingerprint recognition system is called Matching Score s. It measures a quantified similarity between the fingerprint image and the stored template and is normalized in this evaluation to the intervall $s \in [0,1]$, where 1 indicate that both fingerprints originate from the same finger and 0 corresponds to the comparison of fingerprints from different fingers. As usual in a decision systems the decision is made based on a threshold T. In connection with this, the two erroneous decisions, that can be made by biometric systems are:

- False Match - Two fingerprint images of different fingers are categorized as being identical.
- False Non-Match - Two fingerprints of the same finger are categorized as being different.

The corresponding False Match $FMR(T)$ and False Non-Match $FNMR(T)$ error rates depend on the probability densities of the score values for imposter $p(s|H_i)$ and genuines $p(s|H_g)$ and are a function of the chosen threshold $T \in [0,1]$. In contrast to the FAR and FRR, which are often used metrics in literature, the FMR and $FNMR$ are calculated by the enrolled template through a number of comparisons. In contrast to it, the FAR and FRR are calculated via transactions and include, for example, the Failure to Acquire (FTA) and Failure to Enroll (FTE) rates as well (see section below).

B. Generalised Error Rates - Interdependencies

If we look at a biometric recognition system from the outside as a Black Box, it does not matter where the FAR and FRR error rates stem from. They consist of (1) errors resulting from the acquisition of FTA images, (2) errors from enrolling fingerprints, and (3) $FNMR$ and FMR errors resulting from the actual comparison of fingerprints:

- The FTA rate is equivalent to the proportion of attempts where fingerprint images could not be recorded and therefore will not be enrolled. A higher FTA increases FRR and, on the other hand, decreases FAR. The portion of fingerprints that could be recorded therefore is $(1-FTA)$.
- The FTE rate describes the percentage of fingerprints, which could not be enrolled by their respective algorithms. Higher FTEs increases FRR and, consequently, reduces FAR. The portion of fingerprints that could be enrolled is $(1-FTE)$.

Consequently, this results in the following generalized error rates, where

- $(1-FTA)\times FTE$: is the proportion of fingerprints, which could be acquired but not enrolled.
- $(1-FTA)\times(1-FTE)$: is the proportion of fingerprints, which could be both acquired and enrolled.

Contributions to the false acceptance rate can only come from the portion which could be acquired and enrolled. Moreover contributions to the false rejection rate can stem from all portions.

Together, various error rates have been integrated into a verification system as follows:

$$FAR(T) = (1-FTA)\times(1-FTE)\times FMR(T) \quad (1)$$

$$FRR(T) = FTA + (1-FTA)\times FTE \quad (2)$$

$$+ (1-FTA)\times(1-FTE)\times FNMR(T) \quad (3)$$

Additionally we have the following boundary conditions for the match rates:

$$FMR(0) = 1 \qquad FMR(1) = 0 \quad (4)$$

$$FNMR(0) = 0 \qquad FNMR(1) = 1 \quad (5)$$

C. Objective Comparison in Application Scenarios

The system's performance in different operating points (threshold T) can be shown in a Detection Error Tradeoff (DET) curve (see Figure 1). This curve juxtaposes the error rates thus eliminating the graph's dependence on threshold T.

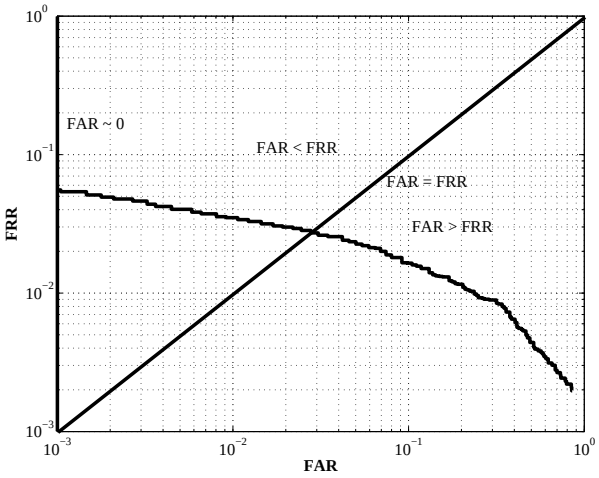

Fig. 1. Regions in the DET graph for different scenarios.

DET curves either show the detecting rate $(1-FRR)$ or they show FRR as an FAR function. The graph chosen in this paper are $FRR = f(FAR)$. To compare the different fingerprint recognition system one has to choose specific regions, operating lines and operating points which can be identified in the DET graph according to the envisaged application scenario (see Figure 1). For example in police-related application scenarios of biometrics the following requirements on the error rates can be identified.

ISBN 0-xxxx-xxxx-x/$xx.xx

1. Using different biometric systems for identity check-ups, e.g. as part of a border control process (Auto-control track, 1 to 1 comparison).

$$FAR < FRR(FAR \approx 0) \tag{6}$$

2. Mobile use of fingerprint scanners (local identity check-ups, $1 : n$).
3. Forensic processing using biometric fingerprint scans and AFIS-comparisons ($1 : n$).

$$FRR < FAR \tag{7}$$

While in the first scenario, there usually is no further examination of the result of biometric matching, the results of biometric systems in scenarios nos. 2 and 3 (identification scenarios) simply serve as approaches for investigations or hints which would definitely have to be verified, e.g. by using forensic experts.

IV. Comparison of Various Systems

A. Approaches to Compare Biometric Systems

As the DET curve is the basis for comparing different fingerprint systems a matrix of all possible DET curves for all sensors and algorithms can be constructed (see Table IV-A).

Sensor	1	...	k	...	7
DB_1	...	...	...	...	...
...	...	...	...	...	...
DB_i	...	...	DET_{ik}	...	...
...	...	...	...	...	...
DB_{11}	...	...	...	...	...

TABLE III
Matrix of the DET curves

One column of the matrix contains all DET curves for a fixed algorithm and all sensors if they are supported by the algorithm. The rows contain DET curves for a fixed sensor and different algorithms. In order to generate the curves, all fingerprint samples of all sensors were evaluated by using all algorithms. These processing results are stored in this matrix as a curve.

Alongside one column or row of the matrix, all DET curves can be inserted into a graph. Thus, the question can be answered which algorithm generates the best biometric performance for a given sensor and vice versa.

In order to compare different systems, a DET curve can be chosen for each sensor; then it can be inserted into a plot. This results in the combination of various systems sorted according to their performance, i.e. according to the lowest error rates.

Besides comparing the whole DET curve, one can move along fixed working lines for various application scenarios or in certain areas of the DET curve (see Section III-C). Therefore, the evaluation below was carried out for the working lines $EER(FAR = FRR)$ and a fixed $FAR = 0.01$ as well.

B. Test Results

The results for various sensors and algorithms are documented in the graphs below (see Figures 2, 3). In connection with this not all algorithms for all sensors could be tested since the algorithm manufacturer could not adapt the fingerprint scanner on time.

B.1 Comparing Individual Systems

A few combinations of sensors and algorithms led either to a notably higher error rate or were not compatible at all.

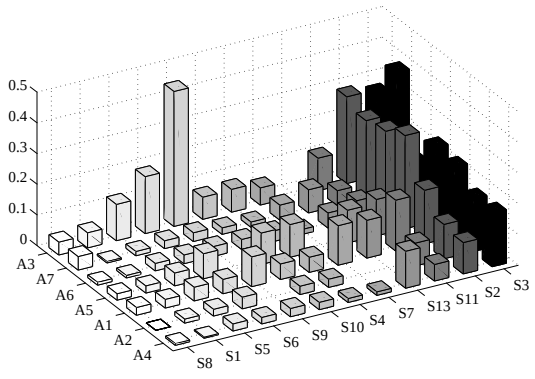

Fig. 2. EER for different combinations (1 = 100%)

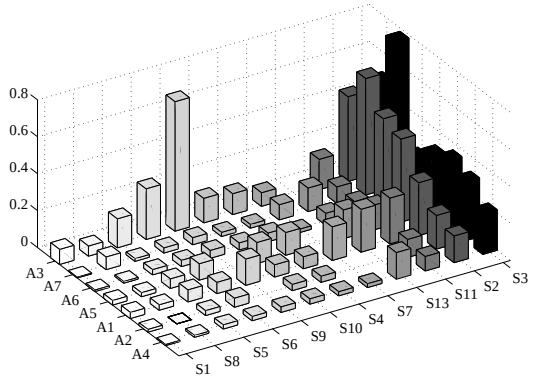

Fig. 3. $FRR(FAR = 0.01)$ for different combinations (1 = 100%)

Comparing results of individual sensors-algorithm combination showed significant differences. The best sensor

achieved an error rate that was ten times lower than the worst one. The best algorithm achieved error rates that were three times lower than those of the worst algorithms.

B.2 Achievable Recognition Performance

The examination has shown what kind of performance today's technology can achieve. The result was that half of the tested systems had an $EER < 5\%$. One third achieved $EER < 3\%$. In the top range ($EER < 1\%$), there are 8% of the tested systems.

As far as the verification of ePassport card holders is concerned, the recognition system will probably be run in such a way as to have an FAR that is better than the EER, e.g. $FAR = 0.1\%$. Even though this leads to a worse FRR, half of the tested systems still generates $FRR < 10\%$ for this operational mode. About 23% of the tested systems can still reach $FRR < 3\%$.

This means that, if mutually compatible components (scanner and algorithm) are carefully chosen, only one out of 1,000 persons with a false ID card would be accepted by the biometric system. However, the probability of wrongly rejecting a person with a correct ID card would be about 1:50. Thus, this technology shows an effective improvement to people comparing faces with ID card pictures.

V. Template Ageing of Fingerprints

From a theoretical point of view it is common sense that aging may not impact the characteristics of fingerprints [1]. However for practical purposes, scaling effects of minutiae based matching algorithms may render older templates useless. In order to investigate those effects in detail a dedicated database was compiled from fingerprints taken of a long period of time.

A. Database

This dedicated database is composed with fingerprints of a total of $N_P = 183$ persons gathered over an period of approx. 40 years.

The number of acquisitions as a function of the respective year can be seen in the graph (see Figure 4).

Furthermore, the number of acquisitions for the respective time period can also be seen. If the number decreases depending on the time difference between forensic processing, the result is that less fingerprint images for matching identical fingers are available. Consequently, if the age of fingerprints exceeds 30 years, then the statistic significance of a comparison will be much lower than if the age is less than 10 years.

In addition, fingerprint images that are 10, 20, or 30 years old do not exist for every person.

There are fingerprint images taken at intervals of 10, 20, or 30 years for the number of persons indicated in the table below (see Table V-A):

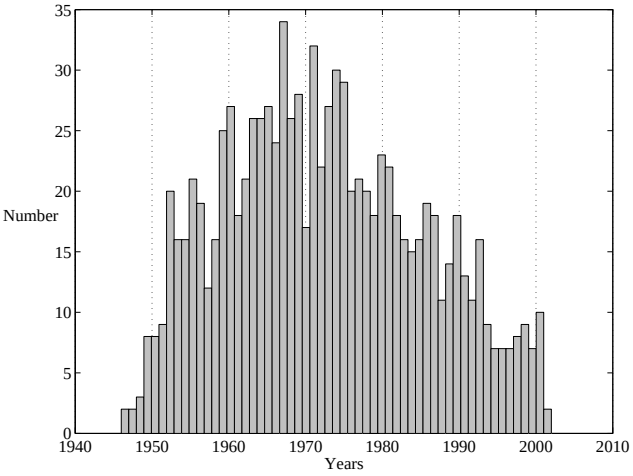

Fig. 4. Number of acquired fingerprints per year

Time intervall	Number of persons
10	65
20	55
30	26

TABLE IV
Number of persons per time intervall

B. Examination of the Ageing Characteristics

The examination described in this section focuses on fingerprint images coming from the same finger. In addition, it is only the ageing effect of the fingerprints that is to be examined.

In most cases, the ageing process does not change the structure of the fingerprint image. The ridges in the epidermis (dead dry skin) always show the same pattern since the information thereof is stored in the lower layer of the finger (dermis - live skin). If an injury of only the upper skin is sustained, after a certain time the same ridges are formed as before. Even the ageing process cannot change the paths of the ridges. The fingerprint may be a little larger, the ridges may be lower (if they were worn due to working), and the finger may show some wounds. However, the pattern always remains the same. Therefore, it should not be difficult, for the different verification algorithms, to identify fingerprints of the same finger, which only differ in the date of their acquisition, as being identical.

To test the impact of ageing, fingerprint images of the same finger were selected from the database in order to determine the probability density $p_g^{\Delta t}(s|H_g)$ of the similarity values for the purpose of comparing fingerprint images of the same finger (index g) as a function of the time interval (index). For determining the probability density $p_i^{\Delta t}(s|H_i)$

of the similarity values of different fingerprint images (index i), the year in which most test persons were enrolled was chosen from the data record in order to keep the time intervals between the acquisitions of different persons as small as possible (see Figure 4).

The probability density $p_i(s|H_i) \approx p_i^{\Delta t}(s|H_i)$ of the similarity values of different fingers (index i) does not change. Therefore, if the function $p_g(s|H_g) \neq p_g^{\Delta t}(s|H_g)$ is changed when comparing fingerprints of the same finger taken at different points in time, a shift in the DET curve along the FRR-axis is expected (see Figure 5).

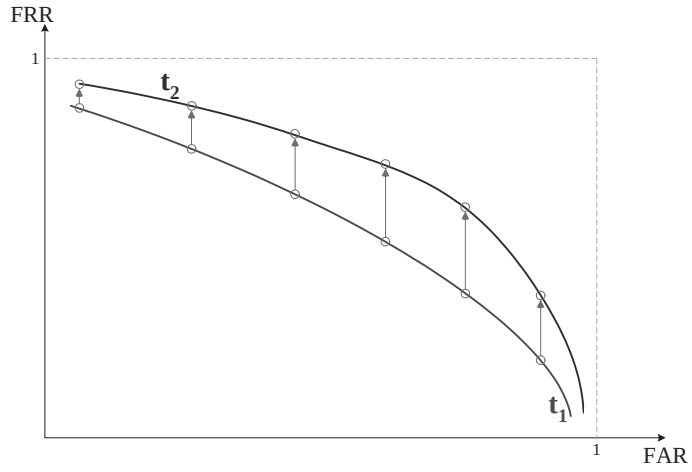

Fig. 5. Expected shift in the DET curve due to ageing

Thus, changes in the probability densities $p_g^{\Delta t}(s|H_g)$ for comparing identical fingers can be determined for different algorithms as a function of the time interval at which the fingerprints were taken. Furthermore, the change in the DET curves and thus the error rates can be examined for every algorithm. In addition, different algorithms can be compared with regard to their robustness towards ageing of identical fingerprint images if all DET curves are incorporated into a graph.

C. Test Results

The results of the experiment are shown below using a subset of the algorithms indicated above. The tests confirmed the general hypothesis of the section above for all tested algorithms. When comparing similar fingerprint images that came from the same fingers, but were acquired at different times, the biometric performance of all algorithms was reduced (see Figure 6).

In order to measure the relative increase in error rates, the FRR ratios for different time intervals are indicated in the figures below in dependence on the FAR (for identical FAR values).

$$q(\Delta t, FAR) = \frac{FRR_{\Delta t=10}}{FRR_{\Delta t=3}}, \text{ for fixed } FAR \qquad (8)$$

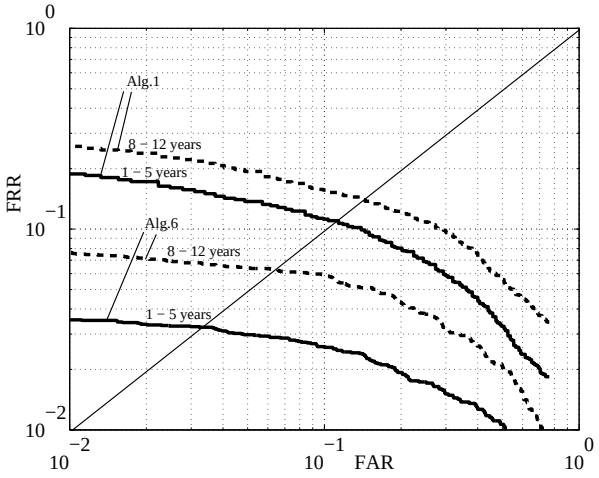

Fig. 6. Shift in the DET curve for algorithms 1 and 6

The quotient hence provides for a statement about the relative change in the FRR when comparing two DET curves for different time intervals (see Figure 7).

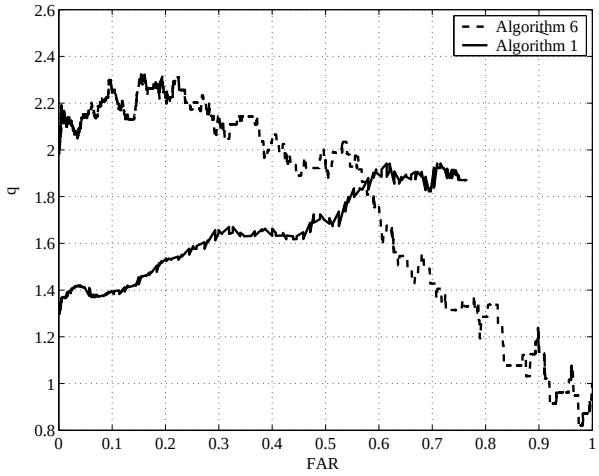

Fig. 7. Relative change of the FRR with regard to Algorithm 6

In conclusion, when comparing fingerprint images with a time difference interval $\Delta t_1 = 10$ years, it can be estimated that the FRR degrades by factor $q \approx 2$ upwards (effects at and beyond the margin $FAR > 0.1$ are not relevant for practical purposes).

VI. Conclusions and Outlook

This paper presents a comparative study of fingerprint recognition systems with special emphasis on investigating the performance with regard to different combinations of fingerprint sensors and algorithms and the impact of ageing.

The results show a great range of error rates for different

sensor-algorithm combinations. Systems performing good at EER also perform good at $FRR(FAR = 1\%)$. Optical sensors operating with the method of frustrated total internal reflection achieved the best of results. Differences between algorithms were notably less pronounced.

The integration of fingerprint images in European ePassports raises the question of whether recognition ability remains if reference and verification data were not recorded within a certain period of time but rather at large intervals. Based on the examinations that were carried out, it can be estimated that the FRR doubles if the time period reaches ten years.

The main benefits that can be expected from these considerations are a) development of an approach for an objective comparision of different fingerprint systems by linking the various error rates during aquisition, enrollment and matching of fingerprints; b) linking of police related application scenarios to different regions in the DET curves enables a clear evaluation of the suitability of different systems to the envisaged application; c) an investigation of the impact of the ageing of fingerprints on the recognition performance relevant to the ePassport scenario.

Since this examination was carried out with all fingers except for the small finger, further improvement can be expected if only fingers with a large area (thumb, index finger) are used. Since the best fingerprints were not singled out but rather all images were analyzed, an improved recognition performance can be expected, if for example in case of a wrong rejection further verification attempts are allowed for or if a sample quality control is carried out at enrollment. The follow-up study, BioFinger II, shall show what kind of improvement can be reached if several finger are used for verification purposes.

VII. Acknowledgment

The research presented in this paper has been founded by the German Federal Office for Information Security (BSI) and was supported by the German Federal Criminal Police Office (Bundeskriminalamt, BKA).

References

[1] Robert Heindl, *System und Praxis der Daktyloskopie und der sonstigen technischen Methoden der Kriminalpolizei*. Berlin and Leipzig, Germany: Walter de Gruyter & Co., 1st ed., 1927.

[2] A. K. Jain, S. Prabhakar, and S. Pankanti, "On the Individuality of Fingerprints," in *Proceedings IEEE CVPR*, (Hawaii), pp. 805–812, dec 2001.

[3] The Interpol AFIS Expert Group, "Data format for the interchange of fingerprint, facial & smt information," Dec. 2003.

[4] "Gesetz zur Bekämpfung des internationalen Terrorismus," Jan 2002.

[5] "FDIS 19794-5, Biometric Data Interchange Formats: Part5: Face Image Data," Jan. 2005.

[6] "FDIS 19794-4, Biometric Data Interchange Formats: Part4: Finger Image Data," Dec. 2004.

[7] "Council Regulation (EC) No 2252/2004 of 13 December 2004 on standards for security features and biometrics in passport and travel documents issued by Member States," Dec 2004.

[8] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain, "Fingerprint verification competition 2000," Sept. 2000.

[9] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain, "Fingerprint verification competition 2002," Apr. 2002.

[10] R. Cappelli, A. Erol, D. Maio, and D. Maltoni, "Synthetic Fingerprint-Image Generation," in *International Conference on Pattern Recognition (ICPR)*, vol. 3, (Barcelona, Spain), pp. 475–478, sep 2000.

[11] "FCD 19795-1, Biometric Performance Testing and Reporting - Part1: Principles and Framework," Jan. 2005.

[12] D. Maltoni and D. Maio and A.K. Jain and S. Prabhakar, *Handbook of Fingerprint Recognition*. Springer, 1st ed., 2003.