JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, July 2012, Vol. 57, No. 4
doi: 10.1111/j.1556-4029.2012.02114.x
Available online at: onlinelibrary.wiley.com

TECHNICAL NOTE

CRIMINALISTICS

Arent de Jongh,[1] Ph.D. and Crystal M. Rodriguez,[1] M.Sc.

Performance Evaluation of Automated Fingerprint Identification Systems for Specific Conditions Observed in Casework Using Simulated Fingermarks

ABSTRACT: Few studies have been reported on the performance evaluation of automated fingerprint identification systems (AFIS) for fingermark-to-fingerprint comparisons. This paper aims to illustrate to fingerprint examiners the relevance of evaluating the AFIS performance under specific conditions by carrying out five types of performance tests. The conditions addressed are the number of minutiae assigned to a fingermark, manual and automatic assignment of the minutiae, the finger region from which the fingermark originates, the degree of distortion in the fingermark, and the difference in orientation between fingermarks and fingerprints. In these tests, the magnitude of the influence for each condition was quantified. The comparisons were performed using a research AFIS technology with simulated fingermarks. Simulated fingermarks provide a practical way to create fingermarks for specific conditions in large quantities. The results showed that each condition influences the performance significantly, emphasizing the relevance of developing, and applying performance tests for specific conditions.

KEYWORDS: forensic science, automated fingerprint identification systems, matching performance, simulated data, fingermark, latent fingerprint, performance evaluation, benchmark

In the fingerprint identification field, automated fingerprint identification systems (AFIS) are used by law enforcement among others to automatically search for candidate sources of crime scene fingermarks in large fingerprint databases (1). Successfully finding the source of a fingermark depends largely on the matching performance of the technology. Therefore, in the procurement phase of an AFIS, the system with a high matching performance is often preferred.

The matching performance of these technologies can be evaluated with different types of performance tests. These tests are largely used in benchmarking to compare the performance of different AFIS technologies. Through the years, several studies have been performed to develop tests to evaluate the matching performance of AFIS technologies for *fingerprint-to-fingerprint* comparisons (1–5). However, for forensic applications, more sophisticated performance tests should be developed to evaluate the matching performance of AFIS technology for *fingermark-to-fingerprint* comparisons. Forensic fingermarks are fingermarks that are found at a crime scene are very different from fingerprints as they usually have a much lower quality, smaller area size, and more distortion resulting in a much larger variability among fingermarks than among fingerprints (6).

Performance tests for fingermark-to-fingerprint comparisons can be created for the general evaluation of an AFIS matching performance or for the evaluation of the performance for specific conditions. Until now, few studies have been published on the evaluation of the general matching performance for fingermark-to-fingerprint comparisons and even less for specific conditions (1,7). Evaluation of the matching performance for specific conditions is relevant to the daily practice of fingerprint examiners. Fingerprint examiners usually know that AFIS technologies are influenced by certain conditions; however, they are often not informed about the magnitude of this influence. Tests could be performed to inform the fingerprint examiners on among others how much the AFIS matching performance is influenced by the number of minutiae assigned to the fingermark and by replacing the manual assignment of minutiae with the automatic assignment.

In order to assess the matching performance accurately for fingermark-to-fingerprint comparisons, large numbers of fingermarks and corresponding fingerprints are required. The large quantity of fingermarks is needed to model the different aspects contributing to the variability in fingermarks observed in forensic casework. Unfortunately, forensic fingermarks are not publicly available in large quantities for research purposes owing to among others privacy legislations (8). The main fingermark–fingerprint database publicly available for research purposes is the NIST Special Database 27 (9). This database consists of only 258 forensic fingermark images with corresponding fingerprint images and their minutiae data. Alternatively, fingermarks could be created in a laboratory setting. This procedure is however very labor intensive, especially when large quantities are required.

Several groups have studied different alternatives to collect large sets of fingermarks. These groups have developed different methodologies to create large quantities of simulated fingermarks that can be used as test samples in performance tests. These are further

[1]Digital Technology & Biometry Department, The Netherlands Forensic Institute, Laan van Ypenburg 6, 2497GB The Hague, The Netherlands.

Received 18 Jan. 2011; and in revised form 21 April 2011; accepted 25 June 2011.

denoted in this paper as simulated fingermarks. One of the advantages of these methods is that research groups or institutes can create their own test samples and therefore no privacy concerns apply. Maltoni and Capelli described a method to create simulated fingermarks by applying artificial displacement, rotation, distortion, skin conditions, and noise to so-called master fingerprints (8,10). These master fingerprints are created based on a statistical model of fingerprints and therefore do not involve fingerprint data of specific individuals. Bazen and Gerez created simulated fingermarks by applying a statistical model describing distributions of elastic skin deformations to fingerprints (11). Rodriguez et al. (12) described a method to create simulated fingermarks by selecting local clusters of minutiae with a specific number of minutiae from a complete minutiae set of flat fingerprint images. These fingerprint images are acquired by capturing a movie sequence while a finger is making sequentially a set of movements resulting in a wide variety of elastic skin deformations. For fingermark-to-fingerprint comparisons, no studies have been reported until now in which simulated fingermarks have been used for the performance evaluation of AFIS technologies.

This paper aims to illustrate to fingerprint examiners the relevance of evaluating the matching performance of AFIS technologies under specific conditions observed in casework. The relevance is illustrated with five examples of tests that can be performed for fingermark-to-fingerprint comparisons. Each test quantifies the magnitude of the influence on the performance for a different condition. The tests were performed using a research AFIS technology and with a large set of simulated fingermarks under specific conditions. Simulated fingermarks provide a practical way to create test samples for specific conditions in large quantities.

The first test focuses on the influence of the number of minutiae assigned to a fingermark. In this test, the performance of a second research AFIS algorithm is also evaluated to illustrate the use of this test in benchmarking. The second test focuses on the influence of replacing the manual assignment of minutiae by the automatic assignment of minutiae for both fingermarks and fingerprints. The third test investigates the influence of the finger region from which the fingermarks originates. The fourth test evaluates the performance for fingermarks with different degrees of distortion. The fifth test focuses on the influence of the orientation of the fingermarks. The performances are expressed using analysis methods that are commonly applied in the technology evaluation of AFIS systems. The first test is illustrated using a detection error trade-off (DET) curve analysis with equal error rates (EERs) (13,14), to further illustrate the use in benchmarking. The other four tests are illustrated using a cumulative match characteristic (CMC) curve analysis (15) to illustrate the use in scenario evaluation. Both methods are commonly used in performance testing.

To achieve an accurate evaluation of the performance for forensic application, the simulated fingermarks created for these tests are required to model realistically the aspects contributing to the variability in forensic fingermarks that are relevant for the AFIS algorithms. For this study, the method described by Rodriguez et al. (12) was used to create the simulated fingermarks. This method takes into account several aspects contributing to the variability in fingermarks such as the number of minutiae, the finger region and distortions in ridge flow, and details resulting from elastic skin deformation. Rodriguez et al. (12) compared the similarity scores obtained from comparisons between fingerprints and fingermarks from the same and different sources for both simulated and forensic fingermarks. The study showed that the score distributions for both types of fingermarks were very similar, thereby suggesting that the simulated fingermarks created with this method are good substitutes for forensic fingermarks and can be used to test the performance of AFIS algorithms. For the methods described by Maltoni, Capelli, and Bazen, studies have not yet been reported on the similarity between the simulated fingermarks and forensic fingermarks.

Test 1: Fingermarks with Different Numbers of Minutiae

Forensic fingermarks are observed with a wide range of number of minutiae. In the forensic fingerprint field, fingerprint examiners usually assign minutiae observed on the mark manually. These minutiae sets are then used to perform database searches using an AFIS system to search for potential source candidates of the questioned fingermark. These candidates are presented in a list of a predetermined size. The probability to find the source is influenced by among others the distinctiveness of the minutiae sets. Fingermarks with a low number of minutiae are less distinctive than fingermarks with a high number of minutiae. Therefore, the matching performance is expected to be influenced by the number of minutiae observed in fingermarks. In this test, the matching performance is evaluated as a function of the number of minutiae assigned to a fingermark.

Both a fingermark and a fingerprint data set were prepared. The fingermark data set consisted of 11 sets of simulated fingermarks with 5–15 minutiae. Each set consisted of fingermarks with a specific number of minutiae; the first set consisted of fingermarks with five minutiae and the last set consisted of fingermarks with 15 minutiae. This is the average range of minutiae observed in forensic casework. Each set consisted of *c.* 25,000 simulated minutiae clusters. These clusters were created from movies of six fingers from separate donors as described in Rodriguez et al. (12). The common general patterns, different distortions, and finger regions were represented in the data set. From the six movies, 865 still images were extracted. The minutiae clusters were created using the multiple marks approach from the complete manually assigned minutiae sets on the still images (12). The fingerprint data set consisted of the images and the manually assigned minutiae configurations for six rolled fingerprints images from the six fingers in the fingermark data set.

Two research AFIS algorithms, hereafter called research algorithms 1 and 2, were used in this test for the comparison between fingermarks and fingerprints. Research algorithm 1 is optimized for fingermark-to-fingerprint comparisons and research algorithm 2 is optimized for fingerprint-to-fingerprint comparisons. The performances of these algorithms were compared to illustrate the difference in performance between AFIS technologies optimized for different types of comparisons.

The matching performance was determined by measuring the ability of the AFIS to discriminate between true (same source) and false (different source) comparisons of fingermarks with fingerprints. In each of the 11 simulated fingermark sets, the fingermarks were compared first to the corresponding fingerprints (true source comparisons). The scores resulting from these comparisons were plotted in distributions called true score distributions. Next, the fingermarks were compared to the noncorresponding fingerprints (false source comparisons). The scores resulting from these comparisons were plotted in distributions called false score distributions. From these distributions, DET curves were created where the false acceptance rate is plotted against the false rejection rate as a function of the score threshold. The performance metric used in this test was the EER. The equal error is the point on the DET curve where the false acceptance rate and the false rejection rate are equal.

Figure 1 shows three DET curves for research algorithms 1 and 2 for the comparison of fingermarks with 7, 10, and 13 minutiae to

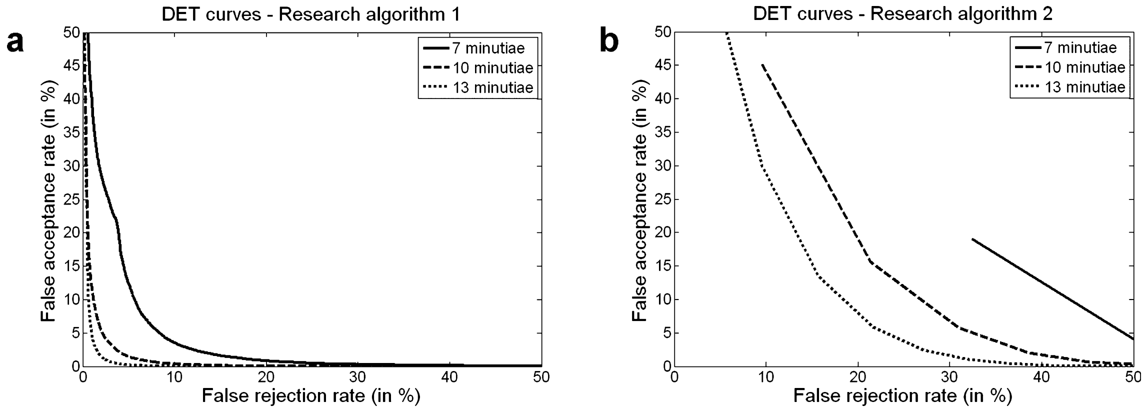

FIG. 1—*Detection error trade-off curves for fingermarks with 7, 10, and 13 minutiae for (a) research algorithm 1 and (b) research algorithm 2.*

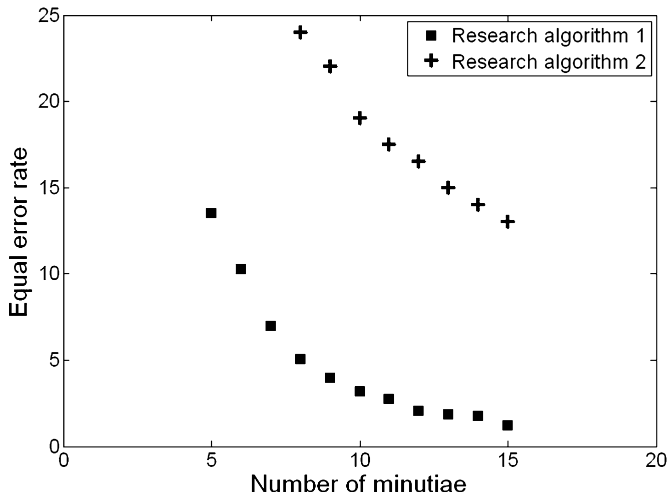

FIG. 2—*Equal error rates for fingermarks with 5–15 minutiae for research algorithms 1 and 2.*

fingerprints. Figure 2 shows a plot with the EERs for research algorithms 1 and 2 observed for the 11 sets of fingermarks with 5–15 minutiae. Both figures show that the performance varies for fingermarks with different numbers of minutiae; the performance decreases sharply with a decrease in the number of minutiae. For research algorithm 2, the EERs could not be determined for fingermarks with 5–7 minutiae as a large number of early out scores were observed for these fingermarks. Early out scores are scores obtained for pairs of fingermarks and fingerprints that have a lower degree of similarity than a predefined threshold and are discarded in the complete matching procedure. The performance for research algorithm 2 is much lower than for research algorithm 1. This result suggests that for forensic fingermarks, AFIS technologies optimized for fingermark–fingerprint comparisons show a significant better performance than those optimized for fingerprint–fingerprint comparisons. Similar tests can be performed by applying another performance metric than the EER. Another option is to perform a scenario evaluation with a metric based on the CMC analysis.

Test 2: Manual and Automatic Assignment of Minutiae for Fingermarks and Fingerprints

In recent years, a new approach has been developed to perform data base searches with a minimum amount of effort (1). In this approach, the initial search is performed by automatic assignment of minutiae with a feature extraction algorithm. This approach is based on replacing the manual assignment of minutiae by the automatic assignment of minutiae for fingermarks using an AFIS feature extraction algorithm. A similar approach can also be applied for the minutiae assignment of fingerprints. This test provides a way to compare the performance of manual and automatic minutiae assignment for both fingermarks and fingerprints. This test consists of two subtests. In test 2a, the performance is compared for manually and automatically assigned minutiae on *fingermarks*. In test 2b, the performance is compared for manually and automatically assigned minutiae on *fingerprints* in the fingerprint database. With this test, fingerprint examiners can weigh the loss in performance against the increase of the searching efficiency when fingermarks or fingerprints are assigned automatically.

In test 2a, the fingermark data set consisted of two sets of 865 simulated fingermarks, fingermarks with manually and fingermarks with automatically assigned minutiae. The first set consisted of minutiae clusters of 12 minutiae. The minutiae clusters were created using the single marks approach from the complete manually assigned minutiae sets on the still images (12). The single marks approach was chosen in the current test, because performing a CMC analysis with large sets of marks created with the multiple marks approach would simply take too much processing time. The second set consisted of fingermark images with forensic backgrounds. Each fingermark image was created based on the minutiae clusters in the first set following the method described by Rodriguez et al. (12). Figure 3 shows an example of a simulated fingermark. The forensic background is an image of the surface where the fingermark was placed on and contains typically the residue of the development technique and patterns of the surface such as text. The forensic backgrounds were obtained from six forensic fingermark images in the NIST SD27 database by removing the fingermark on the image using a clone tool. The forensic backgrounds were randomly added to the images in the first fingermark data set to create fingermark images with a forensic background. Subsequently, the minutiae on the fingermark images were assigned automatically using the feature extraction AFIS algorithm. The fingerprint data set consisted of the manually assigned minutiae configurations from 1,000,006 rolled fingerprints. Six fingerprint images were from the six fingers in the fingermark data set and 1,000,000 fingerprints were from the Dutch criminal fingerprint database.

The performance was determined by measuring the ability of the AFIS system to return the true source (fingerprint) of the questioned fingermark in a candidate list of specific size. In this test, the comparisons between fingermarks and fingerprints were performed using research AFIS algorithm 1. The fingermarks were

FIG. 3—*Example of a simulated fingermark image with a forensic background.*

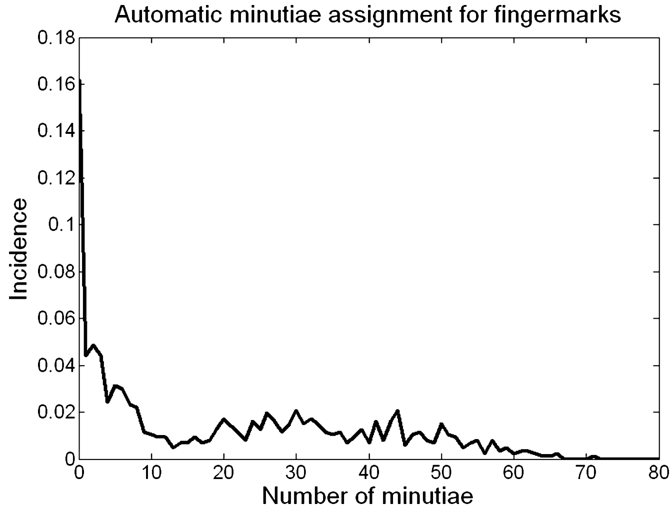

FIG. 5—*Distribution of the number of assigned minutiae for fingermarks with the feature extraction algorithm.*

compared to their corresponding fingerprints and to the 1,000,000 rolled fingerprints. From the resulting scores, CMC curves were created where the probability of returning the true source was plotted against the size of the candidates list.

Figure 4a shows two CMC curves expressing the performance for both fingermarks with manually and automatically assigned minutiae. The performance for fingermarks with manually assigned minutiae is significantly better than for fingermarks with automatically assigned minutiae. This finding can be a result of the relatively large numbers of falsely assigned and/or missed minutiae on the fingermark images. Ideally, the algorithm should assign 12 minutiae on each fingermark. Figure 5 shows the distribution for the actual number of minutiae automatically assigned on the fingermark images. The width of the distribution for the number of minutiae assigned is very large and very few marks with exactly 12 minutiae were observed. Many minutiae were either falsely assigned or missed owing to the forensic background present in the image. A more sophisticated evaluation of the performance can be achieved by creating more variation in the simulated fingermark images. For example, in this study, the ridge details of the simulated fingermarks have a constant quality throughout the entire image. Simulated fingermark images can be created whereby the quality in ridge details varies throughout the image. Simulated fingermarks with an even more "forensic" look can be created by including the appearance associated with certain detection techniques and smudges (8).

In test 2b, the fingermark data set consisted of the 865 manually assigned minutiae clusters of 12 minutiae in test 2a. The fingerprint data set consisted of the manually and automatically assigned minutiae configurations from 1,000,006 rolled fingerprints. Six fingerprint images were from the six fingers in the fingermark data set and 1,000,000 fingerprints were from the Dutch criminal fingerprint data base.

The fingermarks were compared to their corresponding fingerprints and to the 1,000,000 rolled fingerprints. From the resulting scores, CMC curves were created where the probability of returning the true source was plotted against the size of the candidates list.

Figure 4b shows two CMC curves expressing the performance for both sets of fingerprints with manually and automatically assigned minutiae. These results show that the performances for comparisons between fingermarks and fingerprints with manually and automatic assigned minutiae are approximately the same.

Test 3: Fingermarks Originating from Different Finger Regions

Fingermarks originate from different regions of the finger. In general, a well-trained fingerprint examiner should be able to

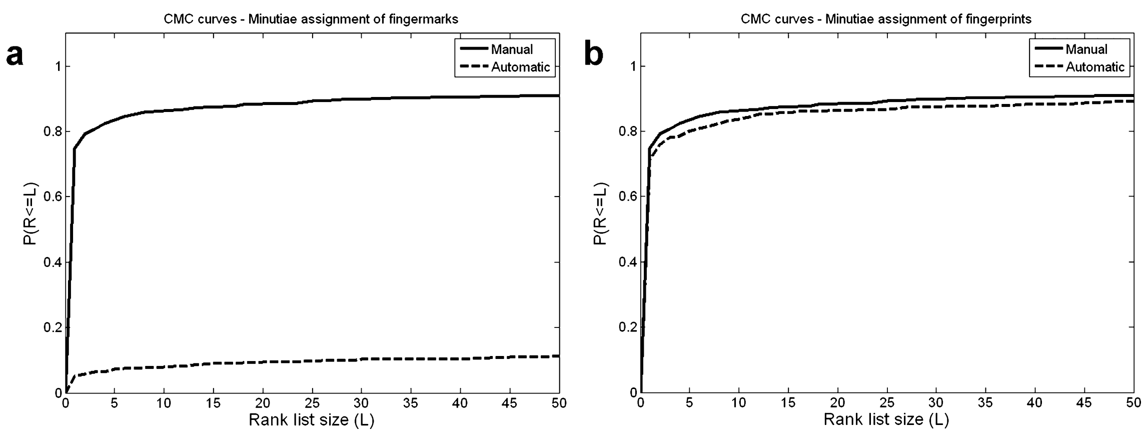

FIG. 4—*CMC curves for fingermarks (a) and fingerprints (b) with manually and automatically assigned minutiae.*

determine from which finger region a mark originates. Yet, fingerprint examiners are usually not informed on how much the performance is influenced by the selection of the finger region in the minutiae assigned for database searching. There may be two reasons why the matching performance may depend on the finger region. First, the ridge flow varies in different finger regions, for example, in the top area of a finger, the ridge flow can be rather parallel, and in the delta area, the ridge flow is multidirectional. Second, certain minutiae configurations might be related to their location in a general pattern and may therefore occur more often than other configurations, the so-called pattern-related minutiae. This test aims to evaluate the effect of fingermarks originating from specific regions in the finger on the matching performance. To evaluate the performance in this study, a simple region classification was made to compare the performance for two different regions: the top half and the bottom half of the finger tip (Fig. 6). The regions were defined by the core position. Simulated fingermarks were created for this test originating from these two regions.

The fingermark data set consisted of two sets of simulated fingermarks based on minutiae clusters of eight minutiae. The minutiae clusters were created from the minutiae sets assigned manually on six plain fingerprint images. The plain fingerprint images were captured using a livescan device by placing the finger flat on the livescan sensor. The first set of fingermarks consisted of 63 minutiae clusters originating from the region above the core in the fingerprint, this region is called the top region. The minutiae clusters were created from the minutiae in the top region of the fingerprint minutiae sets using the multiple marks approach. The second set of fingermarks consisted of 165 minutiae clusters originating from the region below the core in the fingerprint, this region is called the bottom region. The minutiae clusters were created from the minutiae in the bottom region of the fingerprint minutiae sets using the multiple marks approach. The size of the data sets differs, because less minutiae were observed in the top region than in the bottom region. The fingerprint data set consisted of the manually assigned minutiae configurations from 1,000,006 rolled fingerprints. Six fingerprint images were from the six fingers in the fingermark data set, and 1,000,000 fingerprints were from the Dutch criminal fingerprint database.

The fingermarks were compared to their corresponding fingerprints and to the 1,000,000 rolled fingerprints. The matching performance was determined for each set of fingermarks from the resulting scores by measuring the rank of each fingermark in a certain candidate list size.

FIG. 6—*Plain fingerprint image showing two different regions: the top and bottom region.*

Figure 7 shows two CMC curves for the fingermarks originating from the top and bottom area. A higher performance was observed for fingermarks originating from the top than from the bottom area of the finger. This finding indicates that it would be valuable to develop tests with sets of fingermarks that originate from the regions that are thought to be relevant by fingerprint examiners, such as the core, delta, ulnar, and the radial region. Such tests may help to inform the fingerprint examiners whether to select minutiae from certain regions on a fingermark for the searching in a fingerprint database.

Test 4: Fingermarks with Different Degrees of Distortion

The skin deforms each time it makes contact with a surface owing to elasticity of the skin. These deformations are called elastic skin deformations. Nonlinear distortions are observed in the ridge flow of fingermarks resulting from these elastic skin deformations. Because of this distortion, the minutiae configuration in a fingermark is distorted compared to the minutiae configuration in the corresponding fingerprint. The type and the degree of the distortion depend on the direction in which the finger is pressed or displaced and the amount of force applied during contact with the surface (16). Fingerprint examiners are presented with fingermarks for which the degree of distortion cannot always be determined. The matching performance for fingermarks having large distortions in their ridge details is expected to be lower than for fingermarks with less distortion. This test aims to evaluate the effect of the degree of distortion in fingermarks on the matching performance. In this test, the performance obtained for fingermarks with a minimal amount of distortion is compared to the performance obtained for fingermarks with different degrees of more substantial distortions.

The fingermark data set consisted of two sets of simulated fingermarks based on minutiae clusters of eight minutiae. The first set consisted of 254 minutiae clusters with a minimal amount of distortion. The minutiae clusters were created from the minutiae sets assigned manually on six plain fingerprint images using the multiple marks approach described in (12). The plain fingerprint images were captured by placing the finger flat on the livescan sensor. These fingerprints are expected to have a relative small variability in distortions. The second set consisted of 867 minutiae clusters with different degrees of variability in distortions. The minutiae

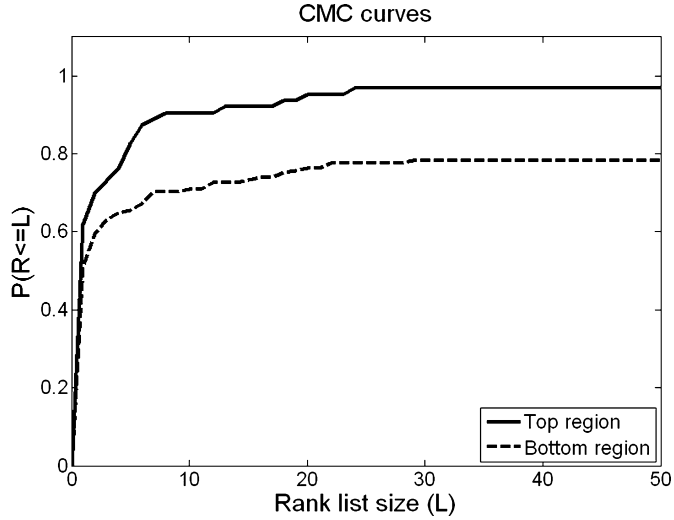

FIG. 7—*CMC curves for fingermarks originating from different finger regions.*

clusters were created from the minutiae sets assigned manually on the still images extracted from the six movies in test 1 using the single mark approach. In the movies, each finger started in the same flat position as in the first set. Subsequently, the fingers moved in several directions to gradually increase the extent of the distortion with respect to a flat scan (12). The fingerprint data set consisted of the manually assigned minutiae configurations from 1,000,006 rolled fingerprints. Six fingerprint images were from the six fingers in the fingermark data set, and 1,000,000 fingerprints were from the Dutch criminal fingerprint database.

The fingermarks were compared to their corresponding fingerprints and to the 1,000,000 rolled fingerprints. The matching performance was determined for each set of fingermarks from the resulting scores by measuring the rank of each fingermark in a certain candidate list size.

Figure 8 shows the CMC curves for both sets of fingermarks. The performance observed for the fingermarks with minimal distortion is higher than for the fingermarks with substantial distortion and therefore suggests that the matching performance is sensitive to the degree of the distortion. Besides the degree of distortion, sensitivity to the type of distortions may also play a role. For developers of AFIS technology, it is interesting to test their technology with sets of fingermarks with specific types of distortion. With the method developed by Rodriguez et al. (12), it is possible to capture specific types of distortion owing to elastic skin deformation.

Test 5: Fingermarks with Different Orientations

In order to determine whether there is much similarity between a fingermark and a fingerprint, a matching algorithm has to determine both the optimal location and orientation of the minutiae features on the fingermark with respect to the minutiae features on the fingerprint. It is generally known that the AFIS matching performance can be sensitive to the difference in orientation between fingermarks and their corresponding fingerprints. Fingerprint examiners are required to estimate and correct the orientation of the questioned fingermark before database searching is performed. The orientation of the fingermark can usually be determined by among others the ridge flow and if present the core and or delta (general pattern). For cases in which fingerprint examiners cannot or only roughly determine the orientation of the fingermark, they would like to know whether this diminishes the chance to find the source in a candidate list. If this is the case, the examiner could perform multiple searches by changing each time the orientation of the fingermark. This test provides a way to determine the influence of differences in orientation between a fingermark and the corresponding fingerprint on the matching performance. The results of this test are relevant for cases in which the orientation of a fingermark cannot be determined exactly. The performance is evaluated for fingermarks in which the orientation differs 0, 15, 45, and 90 degrees compared to the corresponding fingerprints.

The fingermark data set consisted of four sets of simulated fingermarks with different angles of orientation compared to their corresponding fingerprints. Each set consisted of 867 simulated minutiae clusters of eight minutiae. These minutiae clusters were created from the minutiae sets assigned on the still images that were extracted from the six movies in test 1 using the single mark approach. To create the simulated minutiae clusters with specific orientations, image registration (12) was applied to align the still images and their minutiae sets to the corresponding fingerprint images. After image registration, minutiae clusters of eight minutiae were selected from the complete minutiae set. These minutiae clusters constitute the first set of fingermarks with the same orientation as the corresponding fingerprints, denoted as orientation angle of 0 degrees. The remaining sets of fingermarks were created by rotating these minutiae clusters with 15, 45, and 90 degrees. The fingerprint data set consisted of the manually assigned minutiae configurations from 1,000,006 rolled fingerprints. Six fingerprint images were from the six fingers in the fingermark data set, and 1,000,000 fingerprints were from the Dutch criminal fingerprint database.

The fingermarks were compared to their corresponding fingerprints and to the 1,000,000 rolled fingerprints. The matching performance was determined for each set of fingermarks from the resulting scores by measuring the rank of each fingermark in a certain candidate list size.

Figure 9 shows four CMC curves, each for a different orientation of the fingermarks. The curves show that a small difference in orientation between the fingerprint and the fingermarks may already lead to a significant decrease in performance. For fingermarks rotated with 45 degrees, the performance is about 50% less than for fingermarks that were rotated 0–15 degrees. For fingermarks rotated with 90 degrees, no true sources were observed in the

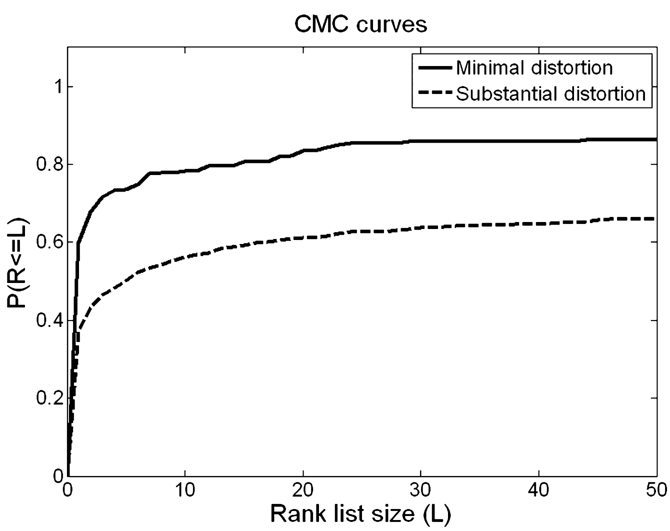

FIG. 8—*CMC curves for fingermarks with different degrees of distortion.*

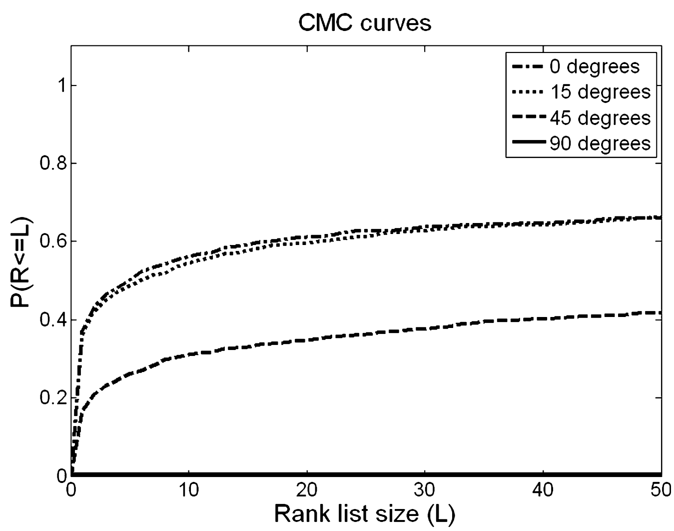

FIG. 9—*CMC curves for fingermarks with increasing differences in orientation with respect to the fingerprints.*

candidate list. For a more continuous evaluation of the performance, the orientations can be increased with smaller intervals.

Discussion

This paper illustrates to fingerprint examiners the relevance of evaluating the performance of AFIS technologies under specific conditions by carrying out five different types of performance tests. The results showed that differences in the performances for each condition were rather significant, emphasizing the relevance of developing and applying performance tests for specific conditions. The tests were based on large numbers of simulated fingermarks created from six fingers. In order to improve the reliability of the performance measurements, fingermarks should be created for a larger population of fingers.

For future studies on performance testing, it would be good to have a larger involvement of fingerprint examiners in the creation of the tests. Fingerprint examiners should consider what kind of tests are relevant to their daily practice as such tests can give them a good sense of what the opportunities and limits are to the system used at their institutions. Such specific tests can be applied in different settings. For example, they can be incorporated in general benchmark tests to make a more detailed comparison of the performance of different AFIS technologies. Another setting would be to use these tests to make an accurate evaluation of the actual performance of an already acquired AFIS technology; the performance can be determined for the actual operational conditions and data.

Acknowledgment

The authors would like to thank Anko Lubach, fingerprint examiner of the National Police Services Agency in the Netherlands, for reviewing this paper.

References

1. Dvornychenko VN, Cochran B, Grother PJ, Indovina MD, Watson CI. Concepts of Operations (CONOPS) for the Evaluation of Latent Fingerprint Technologies (ELFT). Gaithersburg, MD: National Institute of Standards and Technology, 2007.
2. Khanna R, Weicheng S. Automated Fingerprint Identification System (AFIS) benchmarking using the National Institute of Standards and Technology (NIST) Special Database 4. In: Proceedings of the Institute of Electrical and Electronics Engineers 28th Annual International Carnahan Conference on Security Technology; 1994 Oct 12–14; Albuquerque, NM. Piscataway, NJ: Institute of Electrical and Electronics Engineers, 1994;188–94, DOI 10.1109/CCST.1994.363768.
3. Jain K, Prabhakar A, Ross A. Fingerprint matching: data acquisition and performance evaluation. East Lansing, MI: Michigan State University, Technical Report TR99-14, 1999.
4. Garris MD, Wilson CL. NIST biometric evaluations and developments. Gaithersburg, MD: National Institute of Standards and Technology, Technical Report No. NISTIR 7204, 2005.
5. Wilson CL, Watson CI, Garris MD, Hicklin A. Studies of fingerprint matching using the NIST verification test bed. Gaithersburg, MD: National Institute of Standards and Technology, Technical Report No. NISTIR 7020, 2003.
6. Champod C, Lennard CJ, Margot PA, Stoilovic M. Fingerprints and other ridge skin impressions. Boca Raton, FL: CRC Press, 2004.
7. Indovina M, Hicklin A. NIST evaluation of latent fingerprint technologies: extended feature sets. Gaithersburg, MD: National Institute of Standards and Technology, Preliminary Report, 2010.
8. Capelli R. Synthetic fingerprint generation. In: Maltoni D, Maio D, Jain AK, Prabhakar S, editors. Handbook of fingerprint recognition. New York, NY: Springer-Verlag, 2003;203–31.
9. Garris MD, McCabe RM. NIST special database 27: fingerprint minutiae from latent and matching tenprint images. Gaithersburg, MD: National Institute of Standards and Technology, Technical Report No. NISTIR 6534, 2000.
10. Maltoni D, Cappelli R. Advances in fingerprint modeling. Image Vis Comput 2009;27(3):258–68.
11. Bazen A, Gerez S. Thin-plate spline modeling of elastic deformations in fingerprints. Proceedings of the 3rd IEEE Benelux Signal Processing Symposium (SPS-2002); 2002 March 21–22; Leuven, Belgium: IEEE Similar Benelux 2002;1–4, DOI: 10.1.1.112.1931.
12. Rodriguez C, de Jongh A, Meuwly D. Introducing a semi-automatic method to simulate large numbers of forensic fingermarks for research on fingerprint identification. J Forensic Sci 2011; e-pub ahead of print. DOI: 10.1111/j.1556-4029.2011.01950.x
13. Martin A, Doddington G, Kamm T, Ordowski M, Przybocki M. The DET curve in assessment of detection task performance. Proceedings of the 5th European Conference on Speech Communication and Technology; 1997 Sept 22–25; Rhodes, Greece: Eurospeech 1997;1895–8, http://www.isca-speech.org/archive/eurospeech_1997/e97_1895.html (accessed March 8, 2012).
14. Mansfield AJ, Wayman JL. Best practices in testing and reporting performance of biometric devices. Centre for Mathematics and Scientific Computing, NPL Report CMSC 14/02, 2002.
15. Bolle RM, Connell JH, Pankanti A, Ratha NK, Senior AW. The relation between the ROC curve and the CMC. Fourth IEEE Workshop on Automatic Identification Advanced Technologies 2005;15–20, DOI: 10.1109/AUTOID.2005.48.
16. Maceo A. Qualitative assessment of skin deformation: a pilot study. J Forensic Identification 2009;59:390–440.

Additional information and reprint requests:
Arent de Jongh, Ph.D.
Digital Technology & Biometry Department
The Netherlands Forensic Institute
Laan van Ypenburg 6
2497GB The Hague
The Netherlands
E-mail: a.de.jongh@nfi.minjus.nl