On the relation between biometric quality and user-dependent score distributions in fingerprint verification

Fernando Alonso-Fernandez [a], Raymond N. J. Veldhuis[b], Asker M. Bazen[b]
Julian Fierrez-Aguilar[a], Javier Ortega-Garcia[a]

[a]Biometrics Research Lab.- ATVS, Escuela Politecnica Superior - Universidad Autonoma de Madrid
Avda. Francisco Tomas y Valiente, 11 - Campus de Cantoblanco - 28049 Madrid, Spain
{fernando.alonso, julian.fierrez, javier.ortega}@uam.es

[b]University of Twente, 7500 AE Enschede, The Netherlands
{r.n.j.veldhuis, a.m.bazen}@utwente.nl

Abstract

The lack of robustness against image quality degradation is a open issue in fingerprint verification. It has been found in previous studies that the behavior of a fingerprint verification system may vary depending on the quality of the fingerprints. In this paper, we study the performance for individual users under varying image conditions using a multisensor database acquired with three different fingerprint sensors. We propose a user-dependent score normalization scheme that exploits quality information, reaching an EER improvement of $\sim 15\%$ *in one particular sensor. We have also included the proposed score normalization scheme in a multisensor fingerprint verification system that combines the three sensors, obtaining an EER improvement of* $\sim 13\%$ *in the best case* [1].

1. Introduction

In the current networked society, personal identification is becoming a crucial issue in several business sectors such as access or border control, government, finance, health care, etc. Reliable personal recognition, often remotely, and by means of automatic systems is necessary nowadays [10]. This has given rise to a research field known as biometrics [13], in which identification is based on distinctive anatomical (e.g., face, fingerprint, iris) or behavioral (e.g., signature, gait) characteristics. Within the field of biometrics, fingerprint recognition is widely used in many personal identification systems due to its permanence and uniqueness [15]. Due to the low cost and reduced size of new fingerprint sensors, several devices of daily use already include fingerprint sensors embedded (e.g. mobile telephones, PC peripherals). But contrary to the common belief, automatic fingerprint recognition is still an open issue [15].

One of the open issues in fingerprint verification is the lack of robustness against image quality degradation [19]. Our first objective in this work is to investigate the effects of image quality in the performance of individual users. This is motivated by previous studies [5, 4] in which different behavior of different approaches to fingerprint recognition under varying image quality has been observed. In this work, we focus on the performance for individual users using a minutiae-based approach. A score normalization scheme adapted to the quality of individual users is presented. To the best of our knowledge, no previous work on effects of fingerprint image quality in the performance of individual users has been found in the literature.

The second objective in this work is to exploit the quality information of fingerprint images in a multisensor environment. Several results related to information fusion for fingerprint verification have been presented [6, 17, 4]. However, few papers have been focused on sensor fusion [16]. In this paper, we incorporate the quality-based score normalization scheme mentioned in a verification system that fuses the information provided by different fingerprint sensors.

The rest of the paper is organized as follows. Sensor fusion and user-dependent score normalization topics are briefly addressed in Sects. 2 and 3, respectively. The fingerprint verification system used in our experiments is described in Sect. 4. The database and protocol are described in Sect. 5. Experiments and results are described in Sect. 6.

[1]Part of this work has been carried out while F. A.-F. was guest scientist at University of Twente.

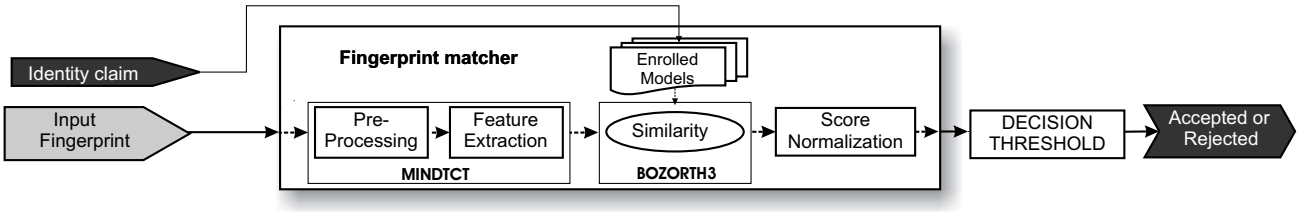

Figure 1. Architecture of the proposed fingerprint verification system.

Conclusions are finally drawn in Sect. 7.

2. Fusion of sensors

Multibiometric systems refer to biometric systems based on the combination of a number of instances, sensors, representations, units and/or traits [12]. Several approaches for combining the information provided by these sources have been proposed in the literature [14, 8]. However, fusion of sensor data has not been extensively analyzed (e.g. [1] and the references therein).

Fusion of sensors offers some important potentialities in biometric verification systems [16]: $i)$ the performance of a verification system can be improved substantially, $ii)$ population coverage can be improved by reducing enrollment and verification failures [18] and $iii)$ it may discourage fraudulent attempts to deceive biometric systems, since deceiving a multisensor system by submitting fake fingers would require different kinds of fake fingers for each sensor. But there are some drawbacks as well: the cost of the system may be higher and more user cooperation is needed. However, these drawbacks are also observed in multibiometric systems that incorporate multiple traits [16].

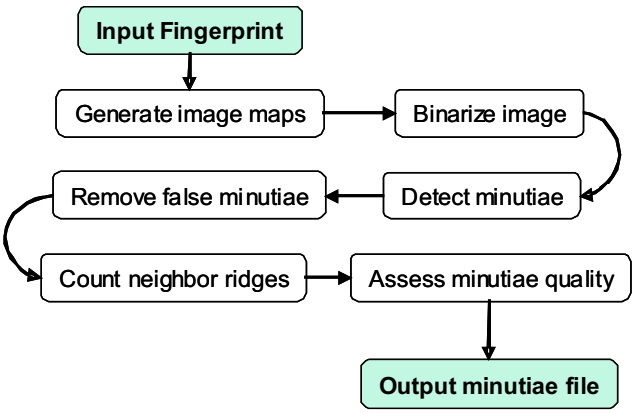

Figure 2. Processing steps of the MINDTCT package of the NIST Fingerprint Image Software 2 (NFIS2).

3. User-dependent score normalization

Score normalization refers to changing the location and scale parameters of the matching score distributions at the outputs of individual matchers, so that the matching scores are transformed into a common domain [9]. In *fixed score normalization*, the normalization follows a fixed rule, whereas in *adaptive score normalization*, the rule can be varied depending on particular characteristics of the input data. It has been shown that the performance of a biometric verification system can be improved exploiting user-dependent information in the score normalization stage (e.g. [7] and the references therein). Previous studies have also shown that using user-dependent decision thresholds (which can be viewed as a particular case of user-dependent score normalization) can improve the performance of a verification system. Multibiometric systems that include user-specific threshold learning has been also reported in previous studies [11]. However, no previous work on score normalization using fingerprint quality measures has been found in the literature.

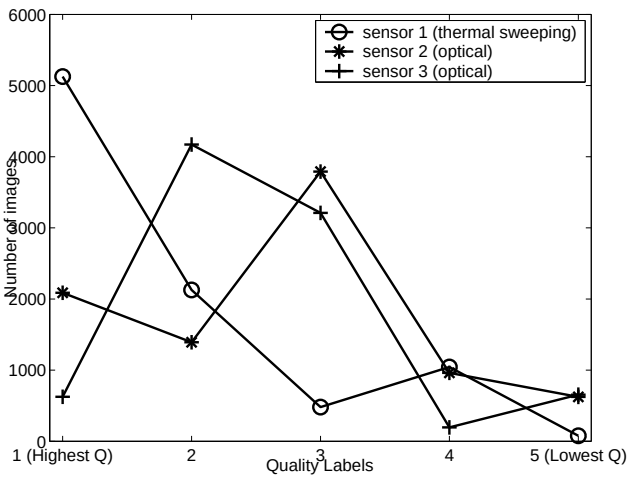

Figure 3. Quality distribution of the datasets used for the experiments provided by the NFIS2 software.

4. Fingerprint verification system

In the experiments reported in this paper, we use the minutiae-based verification system included in the NIST Fingerprint Image Software 2 (NFIS2) [21]. The system architecture of our fingerprint verification system using NFIS2 is depicted in Fig. 1. The NIST Fingerprint Image Software 2 (NFIS2) contains software technology, developed for the Federal Bureau of Investigation (FBI), designed to facilitate and support the automated manipulation and processing of fingerprint images. For our evaluation and tests with NFIS2, we have used the following packages: i) MINDTCT for minutiae extraction; and ii) BOZORTH3 for fingerprint matching.

MINDTCT takes a fingerprint image and locates all minutiae in the image, assigning to each minutia point its location, orientation, type, and quality. The architecture of MINDTCT is shown in Fig. 2 and it can be divided into the following phases: i) generation of image quality map; ii) binarization; iii) minutiae detection; iv) removal of false minutiae, including islands, lakes, holes, minutiae in regions of poor image quality, side minutiae, hooks, overlaps, minutiae that are too wide, and minutiae that are too narrow (pores); v) counting of ridges between a minutia point and its nearest neighbors; and vi) minutiae quality assessment. The BOZORTH3 matching algorithm computes a match score between the minutiae from a template and a test fingerprint. The BOZORTH3 matcher uses only the locations and orientations of the minutia points to match the fingerprints. It is rotation and translation invariant. BOZORTH3 constructs a compatibility table which consists of a list of compatibility association between two pairs of potentially corresponding minutiae, one pair from the template fingerprint and the other pair from the test fingerprint. These associations represent single links in a *compatibility graph*. The matching algorithm then traverses and links table entries into clusters, combining compatible clusters and accumulating a similarity match score s_m. The larger the number of linked compatibility associations, the higher the match score, and the more likely the two fingerprints originate from the same person. For detailed information of MINDTCT and BOZORTH3, we refer the reader to [21]. The similarity match score s_m is normalized into the $[0, 1]$ range by $\tanh(s_m/c_m)$, where c_m is a normalization parameter chosen heuristically.

We have also used the automatic quality assessment software included in the NIST Fingerprint Image Software 2 [20]. This software computes the quality of a given fingerprint based on the minutiae extracted by MINDTCT. The quality is defined as the degree of separation between the match and non-match distributions of a given fingerprint and it is computed using a neural network. This quality measure can be seen as a prediction of the matcher performance. A fingerprint is assigned one of the following quality values: 5 (poor), 4 (fair), 3 (good), 2 (very good) and 1 (excellent). In Fig. 3 we can see the quality distribution of the database used in this paper (see Sect. 5). In our experiments, these quality values are normalized into the $[0, 1]$ range, with 0 corresponding to the worst quality and 1 corresponding to the best quality.

5. Database and protocol

A database with 26568 fingerprint images from 123 participants has been acquired at the University of Twente using three different fingerprint sensors, namely: i) thermal sensor Atmel Sweeping, with an image size of 360 pixels width and 800 pixels height; ii) optical sensor Digital Persona U.are.U, with an image size of 500 pixels width and 550 pixels height; and iii) optical sensor Polaroid, with an image size of 300 pixels width and 302 pixels height. The three sensors have a resolution of 500 dpi. From now on, the three sensors will be referred as *sensor 1* (Atmel Sweeping), *sensor 2* (Digital Persona) and *sensor 3* (Polaroid). The next 6 fingers have been acquired per participant: right index, left index, right middle, left middle, right ring and left ring. For each finger, 12 prints with each sensor have been acquired. This results in 738 different fingers with 36 impressions per finger. The prints were collected from untrained users under supervised conditions, so if the acquired image was not of reasonable quality, it was taken again. However, the quality remained poor for some of the prints, and those are then included in the database. In Fig. 3 it is depicted the quality distribution of the database provided by the quality assessment software described in Sect. 4. Some example fingerprints from this database are shown in Fig. 4. We consider the different fingers as different users enrolled in the system. Data from each sensor are then divided into a *training set* and a *test set* as follows.

For the *training set*, we choose the first four impressions of each user. Each fingerprint image is considered as an enrollment fingerprint and it is compared to the remaining images of the same finger, but avoiding symmetric matches, resulting in $738 \times 4 \times 3/2 = 4.428$ genuine matching scores per sensor. The second fingerprint image of each finger is also compared with the third fingerprint of the remaining fingers, resulting in $738 \times 737 = 543.906$ impostor matching scores per sensor.

For the *test set*, we consider the remaining 8 impressions of each user. One fingerprint image of the training set is considered as the enrollment fingerprint and it is compared to the 8 impressions of the test set, resulting in $738 \times 8 = 5.904$ genuine matching scores per sensor. Each enrolment fingerprint is also compared with two fingerprints from the test set of the remaining fingers, resulting in $738 \times 737 \times 2 = 1.087.812$ impostor matching scores per sensor.

(a) (b)

Figure 4. Fingerprint samples of two different users of the database. Fingerprint images are plotted for the same finger for i) Atmel thermal (left), ii) Digital Persona optical (upper right) and iii) Polaroid optical (lower right).

A quality value is also assigned to each user in the database based on the quality measure described in Sect. 4. We first define the quality of a matching score as $Q_{score} = \sqrt{Q_{enroll} \times Q_{input}}$, where Q_{enroll} and Q_{input} are the image qualities of the enrolled and input fingerprints respectively. The quality of a user is then computed as the average quality of their genuine matching scores from the training set. This process is repeated for the three sensors, thus resulting in three different quality values per user.

6. Experiments and results

6.1 User-dependent score normalization exploiting quality measures

We first analyze the effects of image quality in the performance of individual users. A ranking of users is carried out based on the user quality values described in Sect. 5. We then consider the matching scores of the training set and compute the verification performance of each user separately, obtaining an EER value and a threshold value t_{EER} for each user. In Fig. 5, threshold values t_{EER} for all the users are depicted. We can see that quality values and threshold values are highly correlated for the optical sensors; as user quality value increases, the threshold value t_{EER} is also increased. These results suggest that there is misalignment in the score distributions for the different users due to differences in the quality of the fingerprints. This behavior is not found in the thermal sensor; this could be because quality of users is higher in this sensor, as can be seen in the solid black line of Fig. 5.

To prevent such misalignments in the optical sensors, we propose to normalize the scores based on the quality of each particular user. Given a set of scores $\{s_{i,j}\}$ from user i in sensor j, a normalization constant value $C_{i,j}$ is computed so that the normalized scores are given by $\{s_{i,j} - C_{i,j}\}$. We calculate the value $C_{i,j}$ as

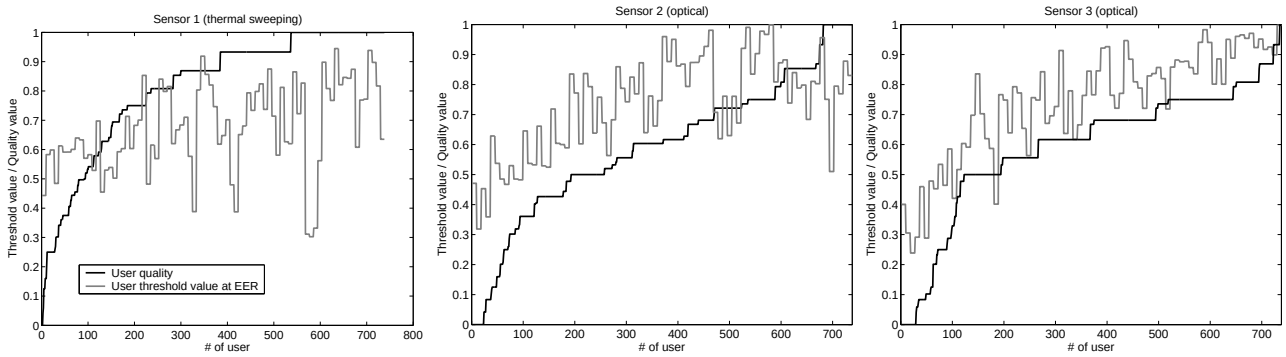

Figure 5. Threshold value t_{EER} of each user of the training set. Users are ranked by quality.

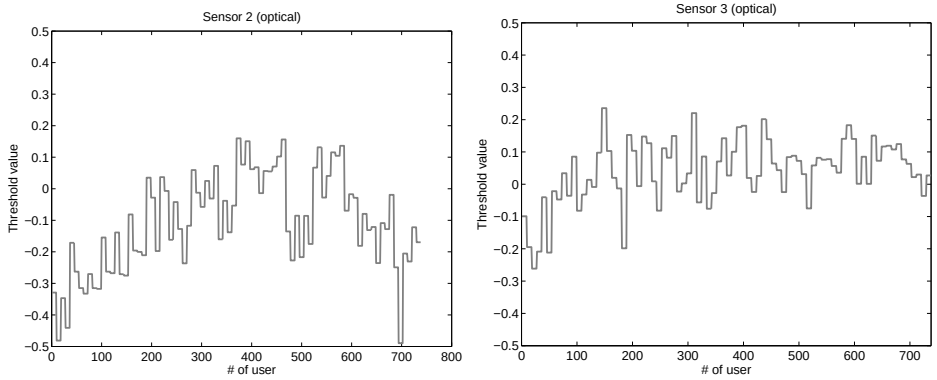

Figure 6. Threshold value t_{EER} of each user of the training set with the proposed score normalization scheme. Users are ranked by quality.

$$C_{i,j} = \begin{cases} Q_{i,j} + (1 - Q_{i,j}) \times K_j & Q_{i,j} > Q_{MIN,j} \\ C_{MIN,j} & otherwise \end{cases} \quad (1)$$

where $Q_{i,j}$ is the quality of user i for sensor j. K_j, $Q_{MIN,j}$ and $C_{MIN,j}$ are experimental constants. It is observed in Fig. 5 that for low quality values, the threshold value t_{EER} is not dramatically decreased, so if $Q_{i,j}$ falls below a certain threshold $Q_{MIN,j}$, $C_{i,j}$ is set to the constant value $C_{MIN,j}$. In our experiments, we have set K_j, $Q_{MIN,j}$ and $C_{MIN,j}$ so as to minimize the EER value of each sensor on the training set. In Fig. 6 threshold values t_{EER} for all the users with this normalization scheme are plotted. It can be seen that the correlation between t_{EER} and user quality has been removed for $sensor3$. This is not true for $sensor2$, in which we still have some correlation. In Fig. 7, we can see the verification performance of the optical sensors on the training set before and after normalizing the scores with this scheme. As can be seen, the proposed normalization scheme results in an EER reduction of $\sim 17\%$ for $sensor3$; only at low FAR values, the proposed scheme results in worse performance. Normalizing the matching score of $sensor2$ does not result in improved performance, maybe because the correlation between quality values and threshold values has not been removed with this normalization, as explained above. The proposed normalization scheme exploits the quality information using a linear function (see Eq. 1). For $sensor2$, a non-linear function could result in improved performance and will be the source of future work.

To validate the proposed normalization scheme, we now normalize the scores of the test set using the parameters computed from the training set. In an operational environment, this means that we compute the user-dependent normalization parameters from a set of fingerprint images provided at the enrolment stage (in our experiments, the fingerprints of the training set) and later, at the operational stage, we use the parameters computed at the enrolment stage to normalize the scores of new incoming fingerprints (in our experiments, the fingerprints of the test set). In Fig. 8 we can see the verification performance on the test set. We can observe that the proposed normalization scheme also results in better performance at an operational stage for $sensor$ 3.

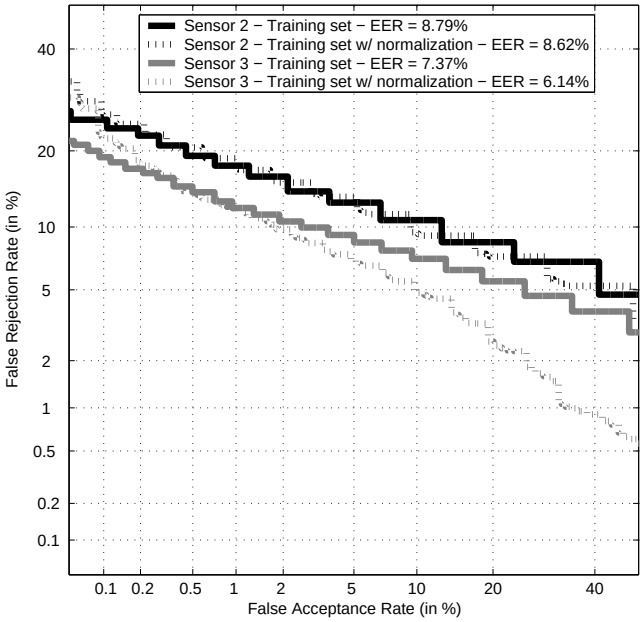

Figure 7. Verification performance on the training set.

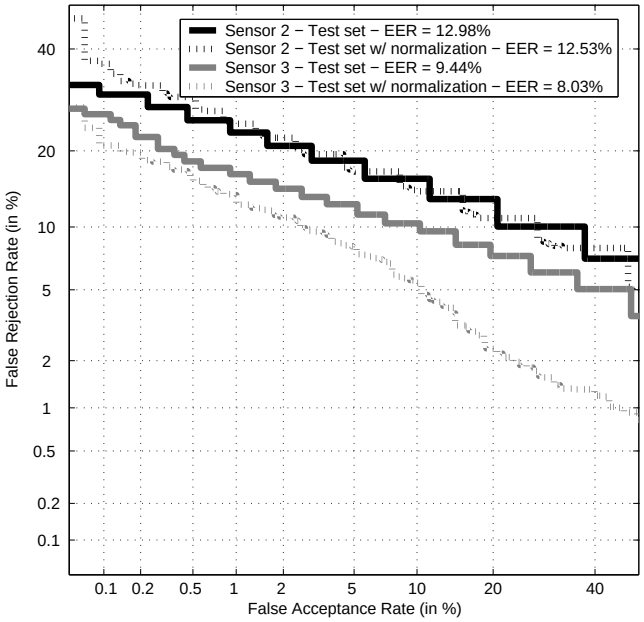

Figure 8. Verification performance on the test set.

In our experiments, a reduction of $\sim 15\%$ in the EER value is obtained. In addition, the proposed normalization scheme results in better performance at any FAR/FRR value. As on the training set, $sensor$ 2 does not result in improved performance with this normalization scheme.

6.2 Sensor fusion experiments

We now exploits quality information to improve the verification performance in a multisensor environment. We incorporate the score normalization scheme proposed to enhance the performance of a multisensor fingerprint verification system.

In this work, we have evaluated a simple fusion approach based on the sum rule. This scheme has been used to combine multiple classifiers in biometric authentication with good results reported [3, 14]. The motivation to use this simple approach comes from the fact that complex trained fusion rules do not clearly outperform simple fusion rules, e.g. see [6].

For the fusion experiments, we have considered all the available scores from the test set resulting from the experimental protocol defined in Sect. 5. In Table 1, we can see the verification performance results. It can be seen that including the normalization scheme proposed always results in improved performance. An EER improvement of $\sim 7.5\%$ and $\sim 13\%$ is obtained when fusing $sensor3$ with $sensor1$ and $sensor2$, respectively, using our normalization scheme.

Regarding absolute EER values, the fusion of $sensor3$-$sensor1$ outperforms the fusion of $sensor3$-$sensor2$, although $sensor2$ has better individual performance than $sensor1$. This could be because the fusion of $sensor3$-$sensor1$ involves sensors of different technology. $Sensor3$ and $sensor2$ are both of optical technology, showing more statistical correlation in their output scores, as can be seen in Fig. 5. This reveals an important source of complementarity between different sensors.

7. Conclusions

The effects of image quality in the performance of individual users have been studied using a multisensor database on a minutiae-based fingerprint verification approach. It has been found for two particular sensors that as user quality value increases, the threshold value at EER is also increased. Worth noting, both sensors have the same technology. We propose a linear quality-based score normalization scheme that exploits this correlation, reaching an EER improvement of $\sim 15\%$ in one sensor. For the other sensor, the normalization scheme proposed does not result in improved performance. Non-linear normalization schemes may be able to improve the performance on this sensor and will be the source of future work. It must be emphasized that we have used a multisensor database, thus containing the same individuals acquired with different sensors. Be-

fusion	*EER value (%)*
$s1$	15.09 %
$s2$	12.98 %
$s3$	9.44 %
$s3_N$ (with normalization)	8.03 %
$s1$ - $s3$	4.87 %
$s1$ - $s3_N$	4.50 % (-7.57 %)
$s2$ - $s3$	5.75 %
$s2$ - $s3_N$	5.00 % (-13.04 %)

Table 1. Error rates in terms of EER for the experiments evaluating fusion of sensors. $s1$, $s2$ and $s3$ stand for *sensor*1 (thermal), *sensor*2 (optical) and *sensor*3 (optical), respectively. The relative performance gain including the normalization scheme proposed is also given.

cause of that, we could consider the above-mentioned correlation as a particular property of each sensor, although this evidence is based on particular implementations of well-known approaches for fingerprint verification and quality assessment. Other implementations of the same approaches may lead to different behavior and should be deeply studied. Future work includes extending this study to approaches for fingerprint verification that does not use minutiae features (e.g. ridge-based [5] or correlation-based [2]).

We have also included the proposed score normalization scheme in a multisensor fingerprint verification system. In our experiments, including the normalization scheme proposed always results in improved performance. An EER improvement of $\sim 13\%$ is obtained in the best case. We have also observed that the best EER value is obtained when combining sensors of different technology, revealing an important source of complementarity.

Acknowledgements

This work has been supported by BioSecure NoE and the TIC2003-08382-C05-01 project of the Spanish Ministry of Science and Technology. F. A.-F. and J. F.-A. are supported by a FPI scholarship from Comunidad de Madrid.

References

[1] F. Alonso-Fernandez, J. Fierrez-Aguilar, and J. Ortega-Garcia. Sensor interoperability and fusion in signature verification: a case study using Tablet PC. *Proc. Intl. Workshop on Biometric Recognition Systems 2005, IWBRS, Springer LNCS-3781*, pages 180–187, 2005.

[2] A. M. Bazen, G. T. B. Verwaaijen, S. H. Gerez, L. P. J. Veelenturf, and B. J. van der Zwaag. A correlation-based fingerprint verification system. *Proc. Workshop on Circuits System and Signal Processing, ProRISC*, pages 205–213, 2000.

[3] E. Bigun, J. Bigun, B. Duc, and S. Fischer. Expert conciliation for multimodal person authentication systems by bayesian statistics. *Proc. Intl. Conf. on Audio- and Video-Based Biometric Person Authentication, AVBPA, Springer LNCS-1206*, pages 291–300, 1997.

[4] J. Fierrez-Aguilar, Y. Chen, J. Ortega-Garcia, and A.-K. Jain. Incorporating image quality in multi-algorithm fingerprint verification. *Proc. IAPR Intl. Conf. on Biometrics, ICB, Springer LNCS-3832*, pages 213–220, 2006.

[5] J. Fierrez-Aguilar, L. Munoz-Serrano, F. Alonso-Fernandez, and J. Ortega-Garcia. On the effects of image quality degradation on minutiae- and ridge-based automatic fingerprint recognition. *Proc. IEEE Intl. Carnahan Conference on Security Techonology, ICCST*, pages 79–82, 2005.

[6] J. Fierrez-Aguilar, L. Nanni, J. Ortega-Garcia, R. Capelli, and D. Maltoni. Combining multiple matchers for fingerprint verification: A case study in FVC2004. *Proc. Intl. Conf. on Image Analysis and Processing, ICIAP, Springer-LNCS 3617*, pages 1035–1042, 2005.

[7] J. Fierrez-Aguilar, J. Ortega-Garcia, and J. Gonzalez-Rodriguez. Target dependent score normalization techniques and their application to signature verification. *IEEE Trans. on Systems, Man and Cybernetics-Part C, Special Issue on Biometric Systems*, 35(3), 2005.

[8] J. Fierrez-Aguilar, J. Ortega-Garcia, J. Gonzalez-Rodriguez, and J. Bigun. Discriminative multimodal biometric authentication based on quality measures. *Pattern Recognition*, 38(5):777–779, 2005.

[9] A. K. Jain, K. Nandakumar, and A. Ross. Score normalization in multimodal biometric systems. *Pattern Recognition*, 38(12):2270–2285, 2005.

[10] A. K. Jain, S. Pankanti, S. Prabhakar, L. Hong, A. Ross, and J. L. Wayman. Biometrics: A grand challenge. *Proc. Intl. Conf. on Pattern Recognition, ICPR*, 2:935–942, 2004.

[11] A. K. Jain and A. Ross. Learning user-specific parameters in a multibiometric system. *Proc. Intl. Conf. on Image Processing, ICIP*, pages 57–60, 2002.

[12] A. K. Jain and A. Ross. Multibiometric systems. *Communications of the ACM, Special Issue on Multimodal Interfaces*, 47(1):34–40, 2004.

[13] A. K. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):4–20, 2004.

[14] J. Kittler, M. Hatef, R. Duin, and J. Matas. On combining classifiers. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(3):226–239, 1998.

[15] D. Maltoni, D. Maio, A. Jain, and S. Prabhakar. *Handbook of Fingerprint Recognition*. Springer, New York, 2003.

[16] G. Marcialis and F. Roli. Fingerprint verification by fusion of optical and capacitive sensors. *Pattern Recognition Letters*, 25:1315–1322, 2004.

[17] G. Marcialis and F. Roli. Fusion of multiple fingerprint matchers by single-layer perceptron with class-separation loss function. *Pattern Recognition Letters*, 26:1830–1839, 2005.

[18] A. Ross and A. K. Jain. Information fusion in biometrics. *Pattern Recognition Letters*, 24(13):2115–2125, 2003.

[19] D. Simon-Zorita, J. Ortega-Garcia, J. Fierrez-Aguilar, and J. Gonzalez-Rodriguez. Image quality and position variability assessment in minutiae-based fingerprint verification. *IEE Proceedings - Vision Image Signal Processing*, 150(6):402–408, December 2003.

[20] E. Tabassi and C. L. Wilson. A novel approach to fingerprint image quality. *Proc. IEEE Intl. Conf. on Image Processing, ICIP*, 2:37–40, 2005.

[21] C. Watson, M. Garris, E. Tabassi, C. Wilson, R. McCabe, and S.Janet. *User's Guide to Fingerprint Image Software 2 - NFIS2 (http://fingerprint.nist.gov/NFIS)*. NIST, 2004.