Adaptive Score Normalization: A Novel Approach for Multimodal Biometric Systems

Anouar Ben Khalifa, Sami Gazzah, Najoua Essoukri BenAmara

Abstract—Multimodal biometric systems integrate the data presented by multiple biometric sources, hence offering a better performance than the systems based on a single biometric modality. Although the coupling of biometric systems can be done at different levels, the fusion at the scores level is the most common since it has been proven effective than the rest of the fusion levels. However, the scores from different modalities are generally heterogeneous. A step of normalizing the scores is needed to transform these scores into a common domain before combining them. In this paper, we study the performance of several normalization techniques with various fusion methods in a context relating to the merger of three unimodal systems based on the face, the palmprint and the fingerprint. We also propose a new adaptive normalization method that takes into account the distribution of client scores and impostor scores. Experiments conducted on a database of 100 people show that the performances of a multimodal system depend on the choice of the normalization method and the fusion technique. The proposed normalization method has given the best results.

Keywords—Multibiometrics, Fusion, Score level, Score normalization, Adaptive normalization.

I. INTRODUCTION

BIOMETRICS is an alternative to person identity verification, which has been proven effective compared to traditional techniques based on badges, cards, passwords [12]. Yet, biometric systems exploring only one biometric modality (called unimodal systems) have limitations that are: noisy sensor data, non universality or lack of distinctiveness of the biometric trait, spoof attacks, unacceptable error rates and use restriction [6], [21]. These limitations can be overcome by the use of other information from soft modalities [24] or by combining two or more unimodal systems. The latter type (called multimodal biometric system), which we are interested in, has the following advantages:

- Improving the performance of the multimodal system by increasing the quality of discriminative biometric data for each person.

- Solving the problem of non-universality; for example, a person who can not be enrolled in a fingerprint authentication system due to worn-out ridge details can still be authenticated using other biometric characteristics like face or voice.

- The presence of several sources of information significantly reduces the impact of noisy data. If a biometric sample obtained from one of the sources is not of a sufficient quality at a particular acquisition, the samples from other sources can still provide enough information to allow a reliable discrimination for decision making.

- Multimodal biometric systems are more resistant to attacks because it is possible to spoof a unimodal biometric system like imitating a voice or a signature; nevertheless, it is difficult to imagine that all biometric features can be falsified by the same impostor during the same session authentication.

- Multimodal biometric systems can also provide some flexibility. For example, a user can register with multiple biometric characteristics, but at the time of authentication, they can select a subset of features that suit him.

According to [6], [16], [19], multimodality can generally take five possible forms which are: the multiple sensor systems (e.g., information is obtained from different sensors for the same biometric feature), the multiple sample systems (e.g., the acquisition is done with several different samples of the same biometric feature), the multiple instance systems (e.g., the capture of the same biometric characteristic is repeated with the same acquisition system), the multiple algorithm systems (e.g., multiple algorithms working at different levels are applied to the same biometric signal) and the multiple biometric trait systems (e.g., the acquisition of biometric data is done on two or several different biometric traits). In addition, the five multimodality forms can also be combined. For instance, the identity verification by face and palmprint can be performed with different algorithms; multi-sensor acquisition can be performed on the face and the fingerprint [16]. These multimodality forms can therefore remedy the various limitations of unimodal systems; especially, multi-biometric trait systems present the advantage of treating multiple biometric modalities, which overcomes the problem of non-universality and the problem of resistance to attacks unlike the other four multimodal systems that operate only on a single modality. So, the development of biometric systems based on multiple biometric traits has received considerable attention from researchers.

In multimodal biometrics, the combination of two or more systems can be done at four different levels: at the signal level, at the feature extraction level, at the score level and at the decision level [5]. It is difficult to combine biometric information at the signal level and at the feature extraction...
level since this combination requires the homogeneity between data, which is not usually checked. Fusion at the decision level is too rigid because only a limited amount of information is available at this level. Fusion at the score level seems to be the most interesting; it is the most used in multimodal biometrics [1]-[3], [13]. However, the fusion at the score level can not be done on the raw scores from different unimodal systems. Indeed, the scores from unimodal systems are generally of different nature and scale. A step of normalizing the scores is therefore essential to combine the scores. In this work, we are particularly interested in the process of score normalization; we study the effect of different methods of the score normalization on the performance of a multimodal biometric system based on the face, the palmprint and the fingerprint (Fig. 1). We also propose a new score normalization method.

The rest of this paper is organized as follows. In Section II, we state the main methods of score normalizations and a selection of work carried out in this domain. In Section III, we present a new approach to normalizing scores. The achieved experiments and results are explained in section IV.

II. SCORE NORMALIZATION

Normalization of scores is a critical step in the design of a multimodal biometric system. Scores from unimodal systems are generally of different nature and scale. They can be heterogeneous; for example, a classifier output may give a measure of similarity (in this case, the more the similarity measure is important the more the test sample is close to the reference sample) while another classifier presents these responses as an distance (in this case, the more the distance is small the more the test sample is close to the reference sample). In addition, scores are generally of different statistical distributions and are not included in the same range of values, thus a biometric system with a high range of scores can eliminate the contribution of another system which has a range of lower scores. Different works show several methods of score normalization. In what follows, we detail the five most commonly used techniques in the literature. We illustrate the results of applying these techniques on synthetic scores corresponding to two virtual unimodal systems (Fig. 2). We can verify that these scores have different variation ranges and thus they can not be combined with methods, such as the maximum and average ones that will favour System 2 or the minimum method that will favour System 1.

For all those reasons a normalization step is required. Normalization consists in changing the location and scale of the scores to bring them back into a common domain. Once the scores are normalized, they will be combined. Several normalization techniques of the scores have been used in multimodal biometrics; the challenge lies in determining the normalization method that is the more robust and more suitable for a fusion technique.
presence of a single outlier makes most of the scores concentrated only in a small range. Fig. 3 shows the distribution of scores of the two unimodal systems after the MinMax normalization.

![Fig. 3 Distribution of genuine and impostor synthetic scores after MinMax normalization](image)

(a) System 1  
(b) System 2

Fig. 3 Distribution of genuine and impostor synthetic scores after MinMax normalization

B. Z-score Normalization

It is the most commonly used score normalization technique. The normalized scores $s'_N$ are calculated as follows:

$$s'_N = \frac{s_k - \mu(S)}{\sigma(S)}$$  \hspace{1cm} (2)

where $\mu(S)$ and $\sigma(S)$ are respectively the mean and the standard deviation of the set of scores. The Z-scores technique generates positive client scores and negative impostor ones. This method fails to transform the scores of both systems in a common interval, and it can keep the original distribution only if the distribution of raw scores is Gaussian. This technique is not robust because of its two parameters (mean and standard deviation) which are sensitive to outliers. It is called Z-Norm when the mean and standard deviation are calculated from impostor scores [20]. Fig. 4 shows the distribution of scores of the two unimodal systems after the Z-score normalization.

![Fig. 4 Distribution of genuine and impostor synthetic scores after Z-score normalization](image)

(a) System 1  
(b) System 2

Fig. 4 Distribution of genuine and impostor synthetic scores after Z-score normalization

C. Median-MAD Normalization

This technique is based on the calculation of the Median Absolute Deviation. The normalized scores $s'_N$ are calculated as follows:

$$s'_N = \frac{s_k - \text{median}}{\text{MAD}}$$  \hspace{1cm} (3)

where $\text{MAD} = \text{median}(|s_k - \text{median}|)$ is a robust estimator of the dispersion of the scores. The Median-MAD normalization is insensitive to the presence of aberrant scores, does not keep the input distribution and does not transform the scores in a common interval. Fig. 5 shows the score distribution of the two unimodal systems after the Median-MAD normalization.
Fig. 5 Distribution of genuine and impostor synthetic scores after Median-MAD normalization

**D. Tanh normalization**

This technique was introduced by Hampel et al. [4]. The normalized scores $s^N_k$ are calculated as follows:

$$s^N_k = 0.5\left\{\tanh(0.01\left(\frac{s_k - \mu(S_g)}{\sigma(S_g)}\right)) + 1\right\}$$  \hspace{1cm} (7)

where $\mu(S_g)$ and $\sigma(S_g)$ are respectively the mean and the standard deviation calculated from the client scores. The Tanh technique transforms the scores into a common interval. It is robust because it is insensitive to the presence of aberrant scores. We note that in [14, 17], the authors used the mean and the standard deviation of all the scores (impostors and clients) and they suggested that the performances were better considered when having only of the mean and the standard deviation of the client scores. Fig. 6 shows the score distribution of the two unimodal systems after the Tanh normalization.

Fig. 6 Distribution of genuine and impostor synthetic scores after Tanh normalization

**E. Double-Sigmoid Normalization**

This technique was proposed by Cappelli et al. [18]. The normalized scores $s^N_k$ are calculated as follows:

$$s^N_k = \begin{cases} 
\frac{1}{1 + \exp(-2((s_k - t)/r_1))} & \text{if } s_k < t \\
\frac{1}{1 + \exp(-2((s_k - t)/r_2))} & \text{if } s_k \geq t
\end{cases}$$  \hspace{1cm} (7)

where $t$ is the reference operating point and $r_1$ and $r_2$ denote the left and right edges of the region in which the function is linear. The Double-Sigmoid (DSig) normalization transforms the scores into a common interval equal to [0, 1] and it does not usually keep the score distribution. The choice of parameters $t$, $r_1$ and $r_2$ is conditioned by the variation range of the scores; they must be carefully selected to achieve a good efficiency. Fig. 7 shows the distribution of scores of the two unimodal systems after the DSig normalization.

Fig. 7 Distribution of genuine and impostor synthetic scores after DSig normalization

Several works have been developed in the context of multimodal biometric people authentication, but only a few papers have addressed the problem of score normalization. In what follows, we provide a selection of works that have focused on the normalization of scores in multimodal biometrics.

In [8], Shu et al. proposed a multimodal biometric system based on the fusion of face, iris, online signature and offline signature. The scores were normalized through the MinMax and Tanh techniques, then they were combined by the sum, the weighted sum and the product. On a database relating to 100 people, the Tanh and MinMax techniques gave similar results with the average and weighted average fusion methods.
In [14], the authors introduced the modified Tanh normalization where the mean and standard deviation were calculated from all scores (impostors and clients). This technique was compared with the Median-MAD, MinMax and Tanh techniques. The score fusion was made with several techniques such as the sum, the average, the product, the minimum and the maximum. The experiments were validated on the CASIA database composed of 5,502 palmprint images and 22,035 iris images. The obtained results showed that the combination of the Tanh normalization technique with the maximum fusion method gave the best results.

In a face-recognition framework, Wang et al. [7] considered scores from three different classifiers. A new method for score normalization based on the false acceptance was implemented and compared to the Z-score and Tanh normalizations. Several score fusion techniques such as the minimum, the maximum, the median and the sum were considered. On a database relating to 200 people, the best results were achieved with the fusion method of a minimum type through the normalization technique based on the false acceptance rate.

In [23], the authors described a system for person identification based on the hand geometry, palmprint and fingerprint. The fusion was operated at the score level through several methods with different techniques of score normalization. On a database relating to 2,000 images, the DSign normalization followed by a sum fusion method gave the best performance.

Table I summarizes a selection of works dealing with the normalization of scores in multimodal biometrics.

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Biometric traits</th>
<th>Database</th>
<th>Classification</th>
<th>Normalization</th>
<th>Fusion</th>
<th>Performance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[14]</td>
<td>• Palmprint • Iris</td>
<td>• CASIA-Palmprint 312 Sub/5502 Sam • CASIA-IrisV3: 700 Sub/22035 Sam</td>
<td>• Similarity measure • Hamming distance</td>
<td>Tanh Min-Max</td>
<td>Product Median-MAD</td>
<td>RR = 98.1 RR = 99.5 RR = 97.4</td>
</tr>
<tr>
<td>[14]</td>
<td>• Palmprint • Iris</td>
<td>• CASIA-Palmprint 312 Sub/5502 Sam • CASIA-IrisV3: 700 Sub/22035 Sam</td>
<td>• Similarity measure • Hamming distance</td>
<td>Tanh Min-Max</td>
<td>Mean Median-MAD</td>
<td>RR = 98.1 RR = 96.6 RR = 95.5</td>
</tr>
<tr>
<td>[21]</td>
<td>• Face • Fingerprint</td>
<td>NIST: 517 Sub/226772 Sam</td>
<td>SVM</td>
<td>Z-score</td>
<td>Sum</td>
<td>Min-Max FAR = 0.01 GAR = 97.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[20]</td>
<td>• Face • Speech - spectrum • Speech - prosody</td>
<td>• XM2VTS: 295 Sub / 33361 Sam • Switchboard-I: 543 Sub /2430 Sam</td>
<td>SVM</td>
<td>Min-Max Z-score</td>
<td>Weighting sum</td>
<td>EER = 1.09 EER = 1.01 EER = 1.01 EER = 0.869</td>
</tr>
<tr>
<td>[15]</td>
<td>• Hand vein • Iris • Fingerprint</td>
<td>• TJU-Hand vein • CASIA-Iris • CASIA-Fingerprint 108 Sub / 756 Sam</td>
<td>SVM</td>
<td>Min-Max Z-score</td>
<td>Simple Average</td>
<td>DUE EER = 0.0045 EER = 0.0062 EER = 0.0038 EER = 0.23</td>
</tr>
<tr>
<td>[4]</td>
<td>• Face • Fingerprint • Hand-geometry</td>
<td>50 Sub / 750 Sam</td>
<td>Euclidean distance</td>
<td>Min-Max</td>
<td>Median-MAD Min (FAR=0.1)</td>
<td>DSignoid Tanh GAR = 97.8 GAR = 98.6 GAR = 84.5 GAR = 96.5 GAR = 98.5</td>
</tr>
<tr>
<td>[22]</td>
<td>• Palmprint</td>
<td>• 241 Sub/1928 Sam</td>
<td>Similarity measure</td>
<td>Min-Max</td>
<td>Median-MAD</td>
<td>DSignoid Sum</td>
</tr>
</tbody>
</table>
III. THE PROPOSED NORMALIZATION SCHEMES

The technique we propose for the score normalization of multimodal biometric systems is adaptive and it depends on the distribution of client and impostor scores. The normalized scores \( S^N_k \) are calculated as follows:

\[
S^N_k = \begin{cases} 
\frac{2}{1 + \exp(-\beta(s_k - \alpha))} & \text{if } s_k < \alpha \\
\frac{s_k - \text{Min}(S)}{\text{Max}(S) - \text{Min}(S)} & \text{if } s_k \geq \alpha
\end{cases}
\]

where \( \alpha = \text{median}(\mu_G), \mu(S) \) and \( \mu(S) \) are respectively the mean of client scores and the mean of impostor scores, \( \beta = k\sigma(S) \), \( \sigma(S) \) is the standard deviation of impostor scores and \( k \) is a scaling constant.

Fig. 8 shows normalization examples where the scores are transformed from \([-10, 10]\) into \([-1, 1]\) for different values of \( \alpha \) and \( \beta \).

This normalization scheme can transform the scores into a common interval equal to \([-1, 1]\). It offers a simple linear transformation of client scores while the impostor scores are transformed in a nonlinear adaptive manner. The parameter \( \alpha \) is calculated from the average of the client scores and the average of the impostor scores. It is strictly chosen to have a value that belongs to the overlap zone between the two distributions. The parameter \( \beta \) is conditioned by the dispersion of the impostor scores: the higher the dispersion is, the more the extremity of the overlap region is shifted to the right. Fig. 9 shows the score distribution of the two unimodal systems after the proposed normalization technique.

IV. EXPERIMENTATION AND RESULTS

The experiments have been performed on an Intel Dual-Core PC, having 1.73GHz, 1GB RAM, with the Matlab R2007 and Visual C++ environment under the Windows XP platform. To confirm the validity of the proposed approach, we have implemented it on a biometric database relating to the face, the fingerprint and the palmprint.

A. Databases

**Face database:** The face images are obtained from the face94 database of the University of Essex (Dr L.Spacek - http://cswww.essex.ac.uk/mv/allfaces/index.html). The face database consists of 153 subjects with 20 face images available for each subject. The subjects sit at a fixed distance from the camera and are asked to speak. The speech is used to introduce facial expression variation. They are all RGB images, of a 180 \times 200 pixel resolution in a JPEG format [10]. Fig. 10 shows face image samples of 10 users.

**Palmprint database:** The palmprint images are obtained from the 2D_3D palmprint database of the Hong Kong Polytechnic University (http://www4.comp.polyu.edu.hk/~biometrics/2D_3D_Palmprint.htm). The database consists of 400 subjects with 20 palmprint images available for each subject. All palmprint are in greyscale images, with a 128 \times 128 resolution which contain the ROI of the palmprint of the right hand [9]. Fig. 11 shows palmprint image samples of 10 users.
Fingerprint database: The fingerprint images are obtained from the HRF database of the Hong Kong Polytechnic University (http://www4.comp.polyu.edu.hk/~biometrics/HRF/HRF.htm). The HFR database contains 1,480 fingerprint images from 148 fingers. All fingerprints are in greyscale images, of a 640 × 480 resolution [11]. Fig. 12 shows fingerprint image samples of 10 users.

Fig. 12 Fingerprint image samples

B. Performance Results

The three biometric authentication systems respond to the traditional model of a system of pattern recognition. They consist of the following steps: acquisition, characterizing, learning and decision.

In our work, the characterization is based on a Discrete Wavelet Transformation (DWT). The Daubechies9 at level 2 of decomposition has been selected for the face and fingerprint while the Symlet6 at level 2 of decomposition has been used for the palmprint modality. The features used for each modality are composed by the mean and standard deviation from an approximation image and the standard deviation of the vertical, horizontal and diagonal details. For learning, we have opted for a modular architecture based on the support vector machines with the RBF kernel.

Fig. 13 shows the initial distribution of client and impostor scores before the normalization step. We can see a significant overlap between these scores for the systems based on the fingerprint and the palmprint. This overlap is less important in the context of the system based on the face.

Fig. 13 Distribution of genuine and impostor scores before normalization: (a) face, (b) palmprint, (c), fingerprint.

The performance of the three unimodal systems with different normalization methods are given in Fig. 14. Accordingly, the performances of the authentication system based on the face are better than the other two unimodal systems with all the normalization techniques. Particularly, the normalization method we propose outperforms the performances recorded with the other normalization techniques.

Fig. 14 Performance of unimodal systems with different normalization techniques

The performance of multimodal biometric system has been studied with different normalization techniques. Four fusion techniques have been used for the fusion of the normalized scores. Fig. 15 illustrates the performance of the multimodal
system when the scores are fused by: (a) the mean, (b) the maximum, (c) the minimum and (d) the product. Table II summarizes the Equal Error Rates of the multimodal system obtained from the above ROC (Receiver Operating Characteristics) graphs for the different normalization and fusion techniques.

<table>
<thead>
<tr>
<th>Normalization techniques</th>
<th>Fusion techniques (EER %)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>DSig</td>
<td>0.49</td>
</tr>
<tr>
<td>MAD-Median</td>
<td>3.34</td>
</tr>
<tr>
<td>MinMax</td>
<td>0.69</td>
</tr>
<tr>
<td>Tanh</td>
<td>1.25</td>
</tr>
<tr>
<td>Zscore</td>
<td>1.40</td>
</tr>
<tr>
<td>Proposed</td>
<td>0.48</td>
</tr>
</tbody>
</table>

In Table II, we observe that the performances of the multimodal system depend on the choice of the normalization technique as well as the fusion method. For each fusion method there corresponds a specific normalization technique. The multimodal system using the proposed normalization technique provides the best performance, and this is with all the fusion techniques except the Minimum technique which exhibits a good performance with the Tanh normalization.

V. CONCLUSION

In this paper, we are interested in the process of the score normalization within the framework of data fusion in multimodal biometrics. We have examined the effect of a selection of score normalization techniques and fusion methods on the performance of a multimodal biometric system based on the coupling of the face, the palmprint and the fingerprint. We have also proposed a new technique for the score normalization that takes into account the distribution of the impostor scores and client scores. We have shown that the performance of a multimodal biometric system depends on the choice of the technical normalization; in particular, the proposed normalization technique provides an optimal performance when it is coupled with the fusion methods such as the mean and the product.

Fig. 15 ROC obtained for: (a) Mean fusion, (b) Max fusion, (c) Min fusion, (d) Product fusion, of different normalization methods.

REFERENCES