

Quantitative Evaluation of Normalization Techniques of Matching Scores in Multimodal Biometric Systems

Y.N. Singh and P. Gupta

Department of Computer Science and Engineering,
Indian Institute of Technology Kanpur, Kanpur-208016, India
singhyn@iitk.ac.in
pg@iitk.ac.in

Abstract. This paper attempts to make an quantitative evaluation of available normalization techniques of matching scores in multimodal biometric systems. Two new normalization techniques Four Segments Piecewise Linear (FSPL) and Linear Tanh Linear (LTL) have been proposed in this paper. FSPL normalization techniques divides the region of genuine and impostor scores into four segments and maps each segment using piecewise linear function while LTL normalization techniques maps the non-overlap region of genuine and impostor score distributions to a constant function and overlap region using tanh estimator. The effectiveness of each technique is shown using EER and ROC curves on IITK database of having more than 600 people on following characteristics: face, fingerprint, and offline-signature. The proposed normalization techniques perform better and particularly, LTL normalization is efficient and robust.

1 Introduction

In the recent years biometric becomes popular due to automated identification of people based on their distinct physiological and/or behavioral characteristics [1]. Most of the practical biometric systems are *unimodal* (e.g., rely on the evidence of any single biometric information). Unimodal systems are usually, cost-efficient but may not achieve the desired performance because of, noisy data, non-universality, lack of uniqueness of the biometric trait, and spoofing attacks [2]. The performance of the biometric system can be improved by combining of multiple biometric characteristics. These systems are referred as *multimodal* biometric systems [3]. In multimodal biometric systems fusion at matching score level is commonly preferred because matching scores are easily available and contains a sufficient information to make decision about legitimate user and impostor.

Assume that $O_k^G = \{r_{k_1}^G, r_{k_2}^G, \dots, r_{k_N}^G\}$ is the set of genuine scores of N individuals and $O_k^I = \{r_{k_1}^I, r_{k_2}^I, \dots, r_{k_n}^I\}$ is the set of impostor scores of those individuals where, $n = NX(N - 1)$ for characteristic k . The complete set of matching scores is denoted as O_k where, $O_k = O_k^G \cup O_k^I$ and $|O_k^G \cup O_k^I| = N + n$.

Prior to combine the matching scores of different characteristics, scores are preprocessed and to make them homogeneous. The dissimilarity score (r'_{k_i}) of user i for characteristic k can be converted into similarity score in the common numerical range, let it be $[0, 1]$ using the formula, $r_{k_i} = \frac{\max(O_k^G, O_k^I) - r'_{k_i}}{\max(O_k^G, O_k^I) - \min(O_k^G, O_k^I)}$. Alternatively, if the raw scores are found in the range $[\min(O_k), \max(O_k)]$, then they are converted to similarity scores by simply subtracting them from $\max(O_k)$ (e.g., $\max(O_k) - r'_{k_i}$). In the rest of the paper the symbol r_{k_i} is used for similarity score of user i for characteristic k . Further, matching scores of different characteristics need not to be on same numerical scale. Using normalization technique scores of different characteristics are transformed to a common numerical scale. In this paper matching scores of face and fingerprint characteristics are obtained using Haar wavelet [4] and minutiae based technique [5], respectively while global and local features are used to compute the matching scores for offline-signature [6].

The rest of the paper is organized as follows: Section 2 presents the related work in the area of normalization techniques of matching score in multimodal biometric systems. Section 3 proposes two new normalization techniques of matching scores that improve the system performance. The performance of normalization techniques is evaluated using different fusion strategies. Normalization and fusion at matching score level are discussed in Section 4. Experimental results are given in Section 5. Finally, conclusions are presented in the last Section.

2 Related Work

Normalization of matching scores in the multimodal biometric systems is an important issue that leads to system performance. In [7] experiments on a database of 100 users for face, fingerprint and hand-geometry characteristics indicate that the performances of min-max, z-score, and tanh normalizations are found to be better than others. Also, min-max and z-score normalization techniques are sensitive to outliers. Hence, there is a need for a robust and efficient normalization procedure like the tanh normalization. A comprehensive study on normalization-fusion, permutations has been done in [8] where Snelick *et al.*, have proposed an adaptive normalization technique of matching scores. This technique is computationally intensive and suffered with parameters overhead.

Score Normalization

Score normalization refers to transformation of scores obtain from different matchers into a common numerical range. A number of normalization techniques such as min-max, z-score, double sigmoid, tanh, piecewise linear, adaptive normalization along with their evaluation are well studied in [7] and [8]. Assume n_{k_i} be the normalized score corresponding to the similarity score r_{k_i} .

Min-Max (MM) - MM normalization transforms the raw scores of O_k in the range of $[0, 1]$ using,

$$n_{k_i} = \frac{r_{k_i} - \min(O_k^G, O_k^I)}{\max(O_k^G, O_k^I) - \min(O_k^G, O_k^I)}$$

Z-Score (ZS) - ZS normalization transforms the scores to a distribution with mean 0 and standard deviation 1. Let μ_{O_k} , δ_{O_k} be the mean and standard deviation of the set O_k then ZS represents the distance between raw score r_{k_i} and μ_{O_k} in units of δ_{O_k} as,

$$n_{k_i} = \frac{r_{k_i} - \mu_{O_k}}{\delta_{O_k}}$$

Since μ_{O_k} , δ_{O_k} are sensitive to outliers, therefore z-score is not robust. Statistically using Grubbs' test [9] one can identify outliers and evaluate the performance of ZS.

Double-Sigmoid (DS) - DS normalization transforms the scores into the range of $[0, 1]$ using,

$$n_{k_i} = \begin{cases} \frac{1}{1 + \exp\left(-2\left(\frac{r_{k_i} - t_k}{t_{k_L}}\right)\right)} & \text{if } r_{k_i} < t_k, \\ \frac{1}{1 + \exp\left(-2\left(\frac{r_{k_i} - t_k}{t_{k_R}}\right)\right)} & \text{otherwise.} \end{cases}$$

where t_k is the reference point chosen some value falling in the region of genuine and impostor scores and the parameters t_{k_L} and t_{k_R} are chosen as, $t_{k_L} = t_k - \min(O_k^G)$ and $t_{k_R} = \max(O_k^I) - t_k$. DS exhibits a linear characteristic of scores in the overlap region of interval $[t_k - t_{k_L}, t_{k_R} - t_k]$ and nonlinear characteristic beyond to that.

Tanh - Tanh normalization is based on tanh estimator [10]. It maps the raw scores of O_k in the range of $[0, 1]$ as,

$$n_{k_i} = 0.5 * \left[\tanh \left\{ 0.01 * \left(\frac{r_{k_i} - \mu_{O_k^G}}{\sigma_{O_k^G}} \right) \right\} + 1 \right]$$

where, $\mu_{O_k^G}$ and $\sigma_{O_k^G}$ are the mean and standard deviation of the genuine matching scores of characteristic k , respectively.

Piecewise-Linear (PL) - Piecewise linear (PL) normalization technique transforms the scores of O_k in the range of $[0, 1]$. The normalization function of PL maps the raw scores using piecewise linear function as,

$$n_{k_i} = \begin{cases} 0 & \text{if } r_{k_i} \leq \min(O_k^G), \\ 1 & \text{if } r_{k_i} \geq \max(O_k^I), \\ \frac{r_{k_i} - \min(O_k^G)}{\max(O_k^I) - \min(O_k^G)} & \text{otherwise.} \end{cases}$$

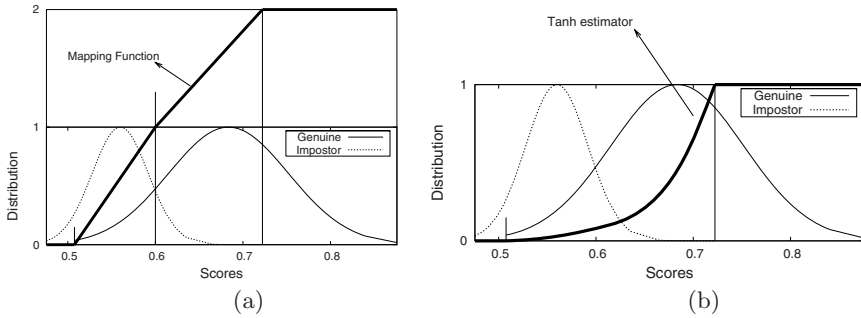


Fig. 1. Proposed Score Normalization Techniques (a) Four Segments Piecewise Linear (FSPL) (b) Linear Tanh Linear (LTL)

3 Proposed Score Normalization Techniques

This section proposes two new matching score normalization techniques: *Four-Segments-Piecewise-Linear* (FSPL) and *Linear-Tanh-Linear* (LTL) normalization, using qualitative combination of multiple normalization techniques. FSPL and LTL techniques take advantages of the characteristics resulted from the piecewise linear function and the tanh estimator for the separability of genuine and impostor scores distributions and robustness, respectively.

3.1 Four-Segments-Piecewise-Linear (FSPL)

FSPL normalization technique divides the regions of impostor and genuine scores into four segments and map each segment using piecewise linear functions (Fig. 1(a)). A reference point t_k is chosen in between the overlapping region of O_k^G and O_k^I . The scores between two extremities of the overlap region are mapped using two linear functions separately in range of [0, 1] and of [1, 2] towards left and right of t_k , respectively as,

$$n_{k_i} = \begin{cases} 0 & \text{if } r_{k_i} \leq \min(O_k^G), \\ \frac{r_{k_i} - \min(O_k^G)}{t_k - \min(O_k^G)} & \text{if } \min(O_k^G) < r_{k_i} \leq t_k, \\ 1 + \frac{r_{k_i} - t_k}{\max(O_k^I) - t_k} & \text{if } t_k < r_{k_i} \leq \max(O_k^I), \\ 2 & \text{if } r_{k_i} > \max(O_k^I). \end{cases}$$

3.2 Linear-Tanh-Linear (LTL)

LTL normalization technique takes the advantage of the characteristic resulted from *tanh* estimator. Normalization function of LTL maps the non overlap region of impostor scores to a constant value 0 and non overlap region of genuine scores

to a constant value 1 (Fig. 1(b)). The overlapped region between O_k^I and O_k^G is mapped to a nonlinear function using \tanh estimator as,

$$n_{k_i} = \begin{cases} 0 & \text{if } r_{k_i} \leq \min(O_k^G), \\ 1 & \text{if } r_{k_i} \geq \max(O_k^I), \\ 0.5 * \left[\tanh \left\{ 0.01 * \left(\frac{r_{k_i} - \mu_{O_k^G}}{\delta_{O_k^G}} \right) \right\} + 1.5 \right] & \text{otherwise.} \end{cases}$$

The effect of normalization techniques both discussed in the previous section and the proposed ones, are examined on system performance using the following fusion strategies. These fusion strategies take into account the performance of the individual characteristic in weighting their contributions [8].

I. Fusion Strategy A. (*Assignment of Weights based on EER*)

This fusion strategy assigns the weight to each characteristic based on their equal error rate (EER). Weights for more accurate characteristics are higher than those of less accurate characteristic. Thus, the weights are inversely proportional to the corresponding errors. Let e_k be the EER to characteristic k , then weight w_k associated to characteristic k can be computed by,

$$w_k = \left(\sum_{k=1}^t \frac{1}{e_k} \right)^{-1} * \frac{1}{e_k} \tag{1}$$

II. Fusion Strategy B. (*Assignment of Weights based on Score Distributions*)

Here weights are assigned to individual characteristic based on their impostor and genuine scores distributions. The means of these distribution are defined by $\mu_{O_k^I}$ and $\mu_{O_k^G}$ respectively, and standard deviations by $\sigma_{O_k^I}$ and $\sigma_{O_k^G}$ respectively. A parameter d_k [11] is used as a measure of the separation of these two distributions for characteristic k as,

$$d_k = \frac{\mu_{O_k^G} - \mu_{O_k^I}}{\sqrt{(\sigma_{O_k^G})^2 + (\sigma_{O_k^I})^2}}$$

If d_k is small, overlap region of two distributions is more, and if d_k is large, overlap region of two distributions is less. Therefore, weights are assigned to each characteristic proportional to this parameter as,

$$w_k = \left(\sum_{k=1}^t d_k \right)^{-1} * d_k \tag{2}$$

For both fusion strategies, $0 \leq w_k \leq 1, (\forall k); \sum_{k=1}^t w_k = 1$ and the fused score f_i for user i is computed as,

$$f_i = \sum_{k=1}^t w_k * n_{k_i}; (\forall i)$$

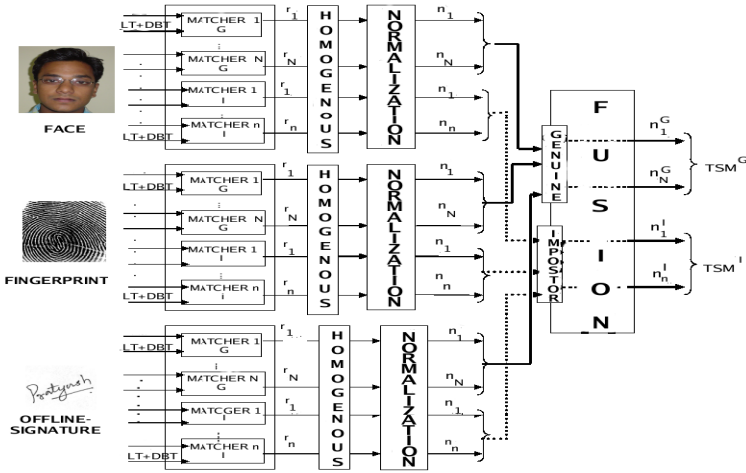


Fig. 2. Block Diagram of Multimodal System

4 System Description

Block diagram of the multimodal biometric verification system based on the fusion of face, fingerprint and signature information at matching score level is shown in Fig. 2. For each characteristic k , first N matchers generate genuine scores $\{r_{k_1}^G, r_{k_2}^G, \dots, r_{k_N}^G\}$ using the matching of live-template to the template of the same individual stored in the database. Next n matchers generate impostor scores $\{r_{k_1}^I, r_{k_2}^I, \dots, r_{k_n}^I\}$ using matching of live-template to the template of other individual stored in the database. Prior to transformation of scores to a common numerical range matching scores of different characteristics must be homogeneous. In the normalization phase scores obtained from different matchers (genuine and impostor) are scaled to a common numerical range. Finally to obtain the fused scores, genuine and impostor scores of each characteristic are combined separately using the weighted fusion strategies as follows,

$$(n_1, n_2, \dots, n_N)^G = \sum_{k=1}^t w_k * (n_{k_1}, n_{k_2}, \dots, n_{k_N})^G$$

and

$$(n_1, n_2, \dots, n_n)^I = \sum_{k=1}^t w_k * (n_{k_1}, n_{k_2}, \dots, n_{k_n})^I;$$

The fused matching scores $(n_1, n_2, \dots, n_N)^G \cup (n_1, n_2, \dots, n_n)^I$ are commonly referred as *total similarity measures (TSM)* of the biometric system. The performance of different normalization techniques for each fusion method is studied against EER values, number of false rejections for subjects and Receiver Operating Characteristics (ROC) curves.

5 Experimental Results

In this section the effect of different normalization techniques on system performance for a multimodal verification system based on face, fingerprint and offline-signature has been discussed using IITK database. For each of these characteristics of total 609 users, live-template is matched against database template, yielding 609 genuine scores and 609 (609x1) impostor scores. The EER values for raw scores for each characteristics are found to be 2.03%, 9.86%, 6.25% for face, fingerprint and signature respectively. The weights for different characteristics for both fusion strategies are calculated according to (1) and (2) which are found as (0.684, 0.123, 0.193) and (0.530, 0.297, 0.177) for face, fingerprint and signature respectively.

Table 1. EER Values for (Normalization, Fusion) Combinations (%)

<i>Normalizations</i>	<i>Fusion Strategy A</i>	<i>Fusion Strategy B</i>
MM	1.07	0.75
ZS	0.74	0.58
DS	0.77	1.08
Tanh	0.91	0.48
PL	1.08	0.91
FSPL	0.71	0.45
LTL	0.42	0.38

Table 1 shows the EER values against different normalization techniques under different fusion strategies. The best one is the lowest EER value in the individual column. As seen in Table 1, the proposed new normalization technique LTL leads to better performance of EER values 0.42% and 0.38% than any other normalization techniques under fusion strategy A and B, respectively. These two near EER values also lead to conclude that the performance of LTL normalization is least dependent upon the distribution of matching scores. The effect of

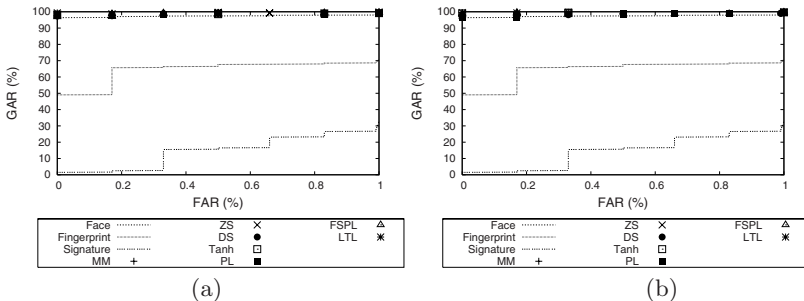
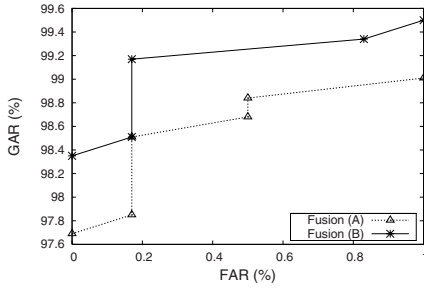
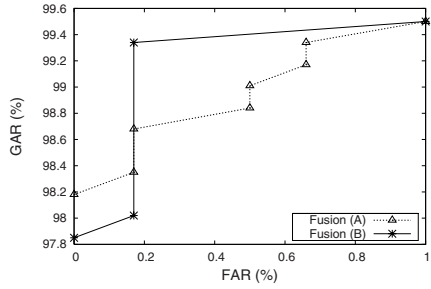


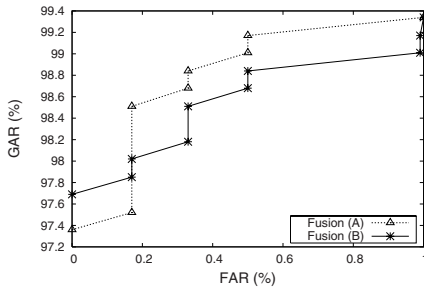
Fig. 3. Effect of Different Normalization Techniques on System Performance (a) Fusion Strategy A and (b) Fusion Strategy B



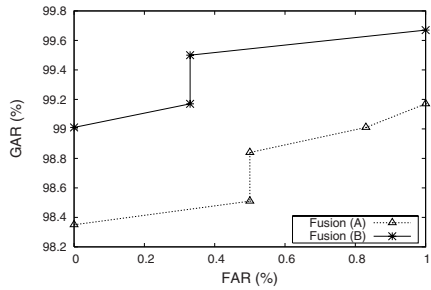
(a) MM



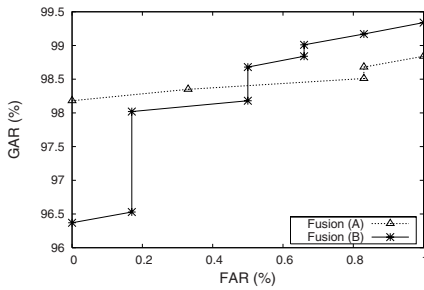
(b) ZS



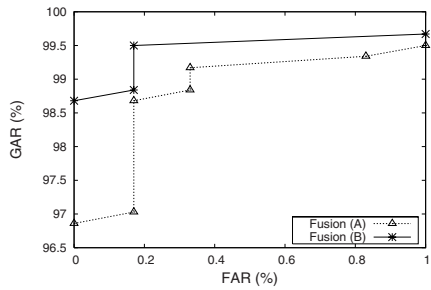
(c) DS



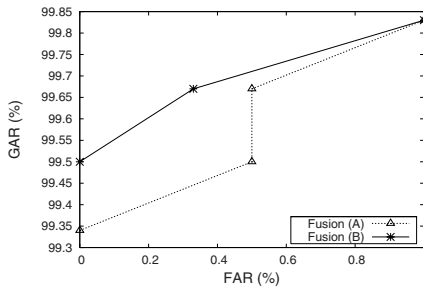
(d) Tanh



(e) PL



(f) FSPL



(g) LTL

Fig. 4. Performance of Different Normalization Techniques under Fusion Strategy A and B

different normalization techniques on system performance under fusion strategy A and B are shown in Fig. 3(a) and (b). ROC curves for the face, fingerprint and signature characteristics are also shown on these figures for comparison which shows the improvement in performance after fusion for all normalization techniques. The performance of different normalization techniques: MM, ZS, DS, Tanh, PL, FSPL, and LTL under both fusion strategies are shown in Fig. 4 (a) through (g).

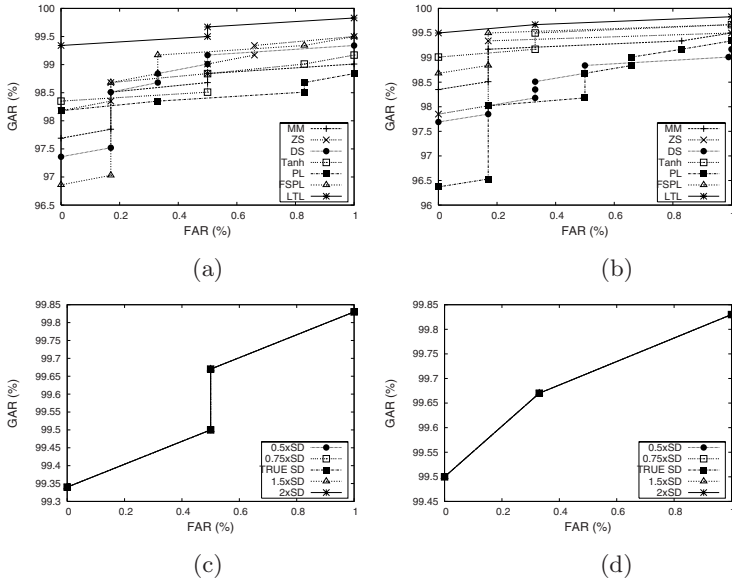


Fig. 5. (a), (b) shows Outperformance of LTL under Fusion Strategy A and Fusion Strategy B. (c), (d) shows Robustness Analysis of LTL normalization under Fusion Strategy A and Fusion Strategy B.

It has been observed from Fig. 4(a) that the performance of MM normalization technique under the fusion strategy B is better than that of fusion strategy A. Similarly it is the case with Tanh, FSPL and LTL normalization as shown in Fig. 4(d), Fig. 4(f) and Fig. 4(g), respectively. The performance of ZS normalization is better at lower FAR under fusion strategy A while at higher FAR, fusion strategy B gives better performance as shown in Fig. 4(b). From Fig. 4(c) the performance of DS normalization is better at higher FAR under fusion strategy A. PL normalization achieved better performance at higher FAR under fusion strategy B while at lower FAR, fusion strategy A gives better performance as shown in Fig. 4(e).

wPerformance of different normalization techniques under both fusion strategies are compared in Fig. 5(a) and (b). From these figures it is observed that the proposed normalization technique LTL outperform the other normalization techniques under both fusion strategies at lower FAR as well as at higher FAR.

The performance of another proposed normalization technique FSPL is better at higher FAR under both fusion strategies. Robustness behavior of the proposed normalization technique LTL is analyzed under fusion strategy A and fusion strategy B that are respectively shown in Fig. 5(c) and (d). From these figures it is found that the performance of the biometric system is completely invariant to the change in standard deviation (SD) of matching scores. In other words, LTL normalization is insensitive towards outliers and hence, it is a robust matching scores normalization technique.

6 Conclusion

This paper deals with the effect of normalization techniques of matching scores on the performance of multimodal biometric systems using face, fingerprint and offline-signature. The experimental results, obtained on a biometric database of IITK of more than 600 individuals, show that the proposed normalization technique Linear-Tanh-Linear (LTL) is *efficient* and *robust*. The performance of Four-Segments-Piecewise-Linear (FSPL) normalization technique is better at low FAR. This analysis of normalization techniques of matching scores suggests that an exhaustive testing of score normalization is needed to evaluate the performance of any multimodal biometric system.

References

1. Jain, A.K., Ross, A.: Information Fusion in Biometrics. Pattern Recognition Letters, Special Issues on Multimodal Biometrics, 2115–2125 (2003)
2. Snelick, R., Indovina, M., Yen, J., Mink, A.: Multimodal Biometrics: Issues in Design and Testing. In: ICMI'03, Canada, pp. 68–72 (2003)
3. Jain, A.K., Ross, A.: Learning User Specific Parameters in a Multimodal Biometric Systems. In: Proc. Inter. Conf. on Image Processing (ICIP), pp. 57–60 (2002)
4. Hassanien, A.E., Ali, J.M.: An Iris Recognition System to Enhance E-security Environment Based on Wavelet Theory. AMO 5(2), 93–104 (2003)
5. Raymond Thai: Fingerprint Image Enhancement and Minutiae Extraction. Technical Report, The University of Western Australia (2003)
6. Ismail, M.A., Gad, S.: Off-line Arabic Signature Recognition and Verification. Pattern Recognition 33, 1727–1740 (2000)
7. Jain, A.K., Nandakumara, K., Ross, A.: Score Normalization in Multimodal Biometric Systems. Pattern Recognition 38, 2270–2285 (2005)
8. Snelick, R., Uludag, U., Mink, A., Indovina, M., Jain, A.K.: Large-Scale Evaluation of Multimodal Biometric Authentication Using State-of-the-Art Systems. IEEE Trans. on PAMI 27(3), 450–455 (2005)
9. Grubbs, F.E.: Procedures for Detecting Outlying Observations in Samples. Technometrics 11(1), 121 (1969)
10. Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A.: Robust Statistics: The Approach Based on Influence Functions. Wiley, New York (2005)
11. Bolle, R.M., Pankanti, S., Ratha, N.K.: Evaluation Techniques for Biometrics-Based Authentication Systems (FRR). In: Proc. of 15th International Conference Pattern Recognition, vol. 2, pp. 831–837 (2000)