Fusion of Face and Voice: An improvement

Y. M. Fouda
College of Computer Science and Information Technology King Faisal University, Al-Ahsa, Saudi Arabia

Abstract:
Multimodal biometric systems provide a better recognition performance compared to systems based on a single biometric modality. The fusion technique is necessary and effective for combing information in multimodal biometric system. In this paper, we propose a new fusion technique based on simple sum rule and product rule. The double sigmoid normalization technique is used after adjusting the parameter t to give us better results. The proposed fusion scheme is compared with some fusion schemes such as sum rule, product rule, maximum rule and minimum rule. Experimental are carried out on two different build multimodal biometric databases. Experimental results indicate that proposed fusion scheme achieves higher performance as compared with other fusion techniques.
Keywords:
Multimodal biometric, score fusion, normalization, face, voice.

1. Introduction

A biometric system is a pattern recognition system that recognizes a person based on a feature vector derived from a specific physiological or behavioral characteristic the person possesses [1]. Biometric offers a natural and reliable solution to the problem of identity determination by establishing the identity of a person based on "who he is", rather than "what he knows". Biometric systems are being increasingly adopted in a number of governments and civilian applications either as a replacement for or to complement existing knowledge and token-based mechanisms. A number of anatomical and behavioral body traits can be used for biometric recognition. Examples of anatomical traits include face, voice, fingerprint, iris, palm print, hand geometry and ear shape.
Biometric systems based on a single source of information (unibiometric systems) suffer from limitations such as the lack of uniqueness and non-universality of the chosen biometric trait, noisy data and spoof attacks [2]. To overcome these types of the problems multimodal biometric systems are used. These systems promise significant improvements over biometric system using a single characteristic, in terms of higher accuracy and most resistance to spoofing.
In literature different approaches for multimodal biometric systems have been proposed. These systems are based on different biometric features or introduce different fusion algorithm of these features. Brunelli and Falavigna [10] used hyperbolic tangent (tanh) for normalization and weighted geometric average for fusion of face and voice biometrics. They also proposed a hierarchical combination scheme for a multimodal identification system. Kittler et al. [11] have experimented with several fusion techniques for face and voice biometrics, including sum, product, minimum, median, and maximum rules and they have found that the sum rule outperformed others. Kittler et al. [11] note that the sum rule is not significantly affected by the probability estimation errors and this explains its superiority.
Conti et. al. [17] has presented a multimodal biometric system using two different fingerprints. The authors used a fuzzy logic based approach, in order to consider the effect of external conditions on the system. With more details they have implemented fuzzy logic module to calculate the weights for each recognition subsystem to realize the weight sum rule. Hong and Jain [8] proposed an identification system based on face and fingerprint, where fingerprint matching is applied after pruning the database via face matching. Ben-Yacoub et. al. [14] considered several fusion strategies, such as support vector machine, tree classifiers and multi-layer perceptrons, for face and voice biometrics. The Bayes classifier is found to be the best method. Ross and Jain [15] combined face, fingerprint and hand geometry biometrics with sum, decision tree and linear discriminant-based methods. The authors reports that sum rule outperform others.
In this paper a multimodal biometric system, combing two matchers face and voice, is proposed. With the proposed approach some biometric monomodal authentication systems limitations has been reduced. The simple sum and product rule are the best fusion techniques compared with the others, so the used fusion technique designed to based on sum and product rule fusions. Further, we modify the double sigmoid normalization method to improve the multimodal system performance. The developed approach can be adopted with general multimodal authentication systems involving different biometric features.
The rest of the paper is organized as follows: Section 2 presents a brief overview for some of the important fusion and normalization techniques. Section 3 introduces the proposed system. In section 4 the experimental results are presented and finally we have outlined our conclusions.

Manuscript received April 5, 2012
Manuscript revised April 20, 2012

2. Scores Normalization and Fusion in Biometrics

Fusion at the match score level is the most popular due to the ease in accessing and consolidating matching scores. As in [15] the three possible levels of fusion are: *First*, *fusion at the feature extraction level*: The data obtained from each sensor is used to compute a feature vector. As the features extracted from one biometric trait are independent of those extracted from the other, it is reasonable to concatenate the two vectors into a single new vector. The new feature vector now has a higher dimensionality and represents a person's identity in a different hyperspace. Feature reduction techniques may be employed to extract useful features from the large set of features. *Secondly*, *Fusion at the matching scores level*: Each system provides a matching score indicating the proximity of the feature vector with the template vector. These scores can be combined to assert the veracity of the claimed identity. Techniques such as (Brute force search, logistic regression, and Support vector machines) may be used to combine the scores reported by the two sensors. These techniques attempt to minimize the false reject rate (FRR) for a given false accept rate (FAR). *Third: Fusion at the decision level:* Each sensor can be capture multiple biometric data and the resulting feature vectors individually classified into the two classes – accept or reject. A majority vote scheme, such as employed in [16] can be used to make the final decision.

2.1 Scores normalization

The matching scores at the output of the individual matcher may not be homogeneous. For example, one matcher may output a distance (dissimilarity) measure while another may be a proximity (similarity) measure. Furthermore, the outputs of the individual matchers need not be on the same numerical scale (range). Finally, the matching scores at the output of the matchers may follow different statistical distributions [3]. Due to these reasons, score normalization is essential to transform the scores of the individual matchers into a common domain prior to combining them. Score normalization is a critical part in the design of a combination scheme for matching score level fusion.

Score normalization refers to changing the location and scale parameters of the matching score distributions at the outputs of the individual matchers, so that the matching scores of different matchers are transformed into a common domain. For a good normalization scheme, the estimates of the location and scale parameters of the matching score distribution must be robust and efficient. Robustness refers to insensitivity to the presence of outliers. Efficiency refers to the proximity of the obtained estimate to the optimal estimate when the distribution of the data is known [3].

In this work, we will use four well-known normalization methods, and the double sigmoid method which is used as a part in the proposed method after we modified it. We denote a raw matching score as s from the set S of all scores for that matcher, and the corresponding normalized score as n.

2.1.1 Min-Max (MM) normalization

MM normalization is best suited for the case where the bounds (maximum and minimum values) of the scores produced by a matcher are known. In this case, we can easily shift the minimum score to 0 and the maximum score to 1. Given that Max(S) and Min(S) are the maximum and minimum values of the raw matching scores, respectively, the normalized score is calculated as:

$$n = \frac{s - Min(S)}{\mathrm{Max(S)} - \mathrm{Min(S)}} \qquad (1)$$

This method is highly sensitive to outliers in the data used for estimation [3], and therefore it is not robust. MM normalization retains the original distribution of scores except for a scaling factor and transforms all the scores into a common range [0, 1].

2.1.2 Z-Score (ZS) normalization

ZS normalization calculates normalized scores using arithmetic mean and standard deviation of the given data. This method transforms the scores to a distribution with mean of 0 and standard deviation of 1. Let Mean(S) denote the arithmetic mean of S and STD(S) denote the standard deviation of S, and then the normalized scores are given by:

$$n = \frac{s - Mean(S)}{STD(S)} \qquad (2)$$

Both mean and standard deviation are sensitive to outliers and, hence, this method is not robust. ZS does not guarantee a common numerical range for the normalized scores of the different matchers. This method retains the original distribution of matching scores only if the distribution of the input scores is Gaussian [4].

2.1.3 Median and Median Absolute Deviation (MAD) normalization

MAD estimates the spread of score distribution, similar to the standard deviation, but it is preferable because of its much better robustness with respect to outliers. Let Median(S) denote the median of S, and MAD be the median of the absolute deviation of all scores from the median. The formula for MAD normalization is given by:

$$n = \frac{s - Median \quad (S)}{MAD} \qquad (3)$$

where MAD = median (|s-Median(S)|).

MAD is insensitive to outliers and the points in the extreme tails of the distribution [3]. This normalization technique does not retain the input distribution and does not transform the scores into a common numerical range.

2.1.4 Tanh normalization

This method is a robust statistical technique [6] and is not sensitive to outliers. It maps the raw scores to the (0, 1) range and is given by:

$$n = 0.5\left\{\tanh\left(0.01\frac{(s - Mean\ (S))}{STD\ (S)}\right) + 1\right\} \qquad (4)$$

To make this method more robust, estimated value for Mean(S) and STD(S) based on Hampel estimators can be used [5].

2.1.5 Double-Sigmoid (DS) normalization

DS normalization used by [7] in a multimodal biometric system that combines different fingerprint classifiers. The normalized score n is given by:

$$n = \begin{cases} \dfrac{1}{1+\exp(-2(s-t)/r_1)}, & s < t \\ \dfrac{1}{1+\exp(-2(s-t)/r_2)}, & otherwise \end{cases} \qquad (5)$$

where t is the reference operating point and r1 and r2 denote the left and right edges of the region in which the function is linear. This scheme transforms the scores into the [0, 1] interval, but it requires careful tuning of the parameters t, r1, r2 to obtain good efficiency.

2.1.6 Modify Double-Sigmoid (MDS) normalization

There are many reasons to use normalizations in multimodal biometric. One of these reasons reducing the overlapping between genuine and imposter scores. Fig. 1 shows the overlapping regions between genuine and imposter scores for the two matcher (face and voice) used in this paper. The double sigmoid technique Eq. (5) is used to reduce the overlapping between the genuine and imposter scores for any module. Choosing the value of parameter t in double sigmoid is a critical point. A. Jain et al. [3] suggested the center of the overlapping region to the parameter t when he used the fingerprint, face, and hand-geometry modules. When we use this choice of the parameter t we obtain 84.16 success percentages. Also S. Ribaric and I. Fratric [18] had chosen the parameter t to be the mean value of the minimum value of genuine scores and the maximum value of imposter scores. When we applied this choice in our modules (Face and Voice) we obtain 95.7 success percentages.
My suggestion is to use the brute force technique to find the best value of the parameter t. The brute force is based on searching the value of parameter t in the overlapping region which makes the overlapping between genuine and imposter scores is minimum. When we applied brute force technique we obtain 96.19 success percentages.

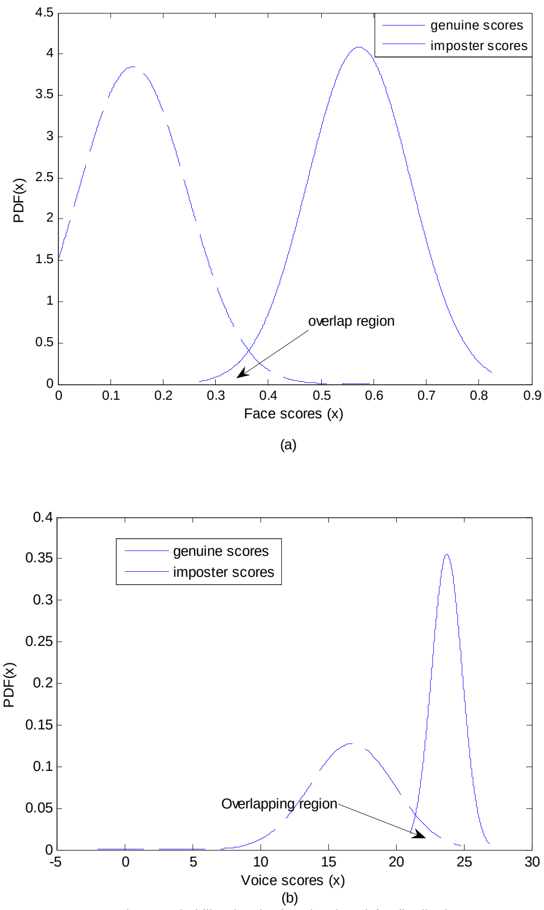

Fig. 1 Probability density function (PDF) for distribution for imposter and genuine scores (a)Face (b)Voice

2.2 Fusion biometrics

In this section we review the most popular score fusion technique in multimodal biometrics (e.g. Maximum rule, Minimum rule, Sum rule, Product rule). For the fusion rules presented in this paper, f is the fused score, x_m is the score of the m^{th} matcher, m=1,2,…, M.

Simple Sum Rule (SSR)

In SSR rule, the fused score is computed by adding the scores for all modalities involved. The computation here is defined as [6, 9]:

$$f = \sum_{m=1}^{M} x_m \tag{6}$$

Maximum Rule (MAR)

Maximum rule method selects the score having the largest value amongst the modalities involved. It is defined mathematically as [6, 9]:

$$f = \max(x_1, x_2, ..., x_M) \tag{7}$$

Minimum Rule (MIR)

In the Minimum rule, the match-score x_m represents the distance score. Minimum rule method chooses the score having the least value of the modalities involved. It is defined as [6, 9]:

$$f = \min(x_1, x_2, ..., x_M) \tag{8}$$

Product Rule (PRR)

In Product rule, the fused score is calculated by multiplying the scores for all modalities involved. It is mathematically defined as [10]:

$$f = \prod_{m=1}^{M} x_m \tag{9}$$

3. The proposed technique (SPR)

The new fusion technique uses the SSR and PRR rules so we call it by SPR. For face and voice matcher, experimental results in [11] showed that sum rule fusion was outperformed other rules. So, to improve the performance of the multimodal system, we try to increase the efficiency of fusion by using sum rule. By summing two scores the resultant sum scores will have smaller variance than the average variance of the individual scores. Also we try to reduce spoof attacks by using the product rule. Finally we combine scores with less variance (sum rule) with reducing spoof attacks scores (product rule). In this case the fusion rule based on these combinations between SSR and PRR is given by the mathematical formula:

$$f = \sum_{m=1}^{M} x_m + \prod_{m=1}^{M} x_m \tag{10}$$

The errors of individual biometric matchers stem from the overlap of the genuine and impostor score distributions. To decrease the effect of this overlap on the fusion rule, we suggest to use MDS normalization procedure that aims to increase the separation of the genuine and impostor distributions, while still mapping the scores to [0,1] range. Adjusting the parameter *t* by MDS normalization increase the separation between the genuine and imposter scores. The following is the procedure for our proposed method:

1. Prepare the training data and test data for each matcher in a similarity matrix. The training data contains the set of pattern that are known to the system (the biometric database). The test data contains the subjects that are to be compared against the training data.
2. Normalize the similarity matrix in step (1) by MDS normalization see eq. (5).
3. Fuse the set of normalization similarity matrices into a single fusion similarity based on SPR see Eq. (10).
4. Performance statistics for verification are computed from the genuine and imposter scores. Genuine scores are those that result from comparing elements in the training and test sets of the same subject. Imposter scores are those resulting from comparisons of different subjects.
5. Use a suitable fusion score as a threshold and compute the (FAR) and (FRR) by selecting those imposter and genuine scores, respectively, on the wrong side of this threshold and divide by the total number of scores used in the test. A mapping table of the threshold values and the corresponding error rates (FAR and FRR) are stored. The complement of the FRR (1 – FRR) is the genuine accept-rate (GAR).
6. The GAR and the FAR are plotted against each other in a curve called Receiver Operating Characteristic (ROC) curve, a common system performance measure. In practice, one chooses a desired operational point on the ROC curve and uses the FAR of that point to determine the corresponding threshold from the mapping table.

4. Experimental Results

To evaluate the performance of the system, based on testing the previously described fusion techniques, a database containing face and voice samples was required. The datasets considered for the face and voice modalities were extracted from the XM2VTS (clean images) [12] and from 1-speaker detection task of the TIMIT Speaker Recognition Evaluation (clean speech) databases, respectively [13]. Using these datasets, a total number of 140 client tests and (140×[140-1])=19460 imposter tests were used from the development data for investigating the performance of the proposed scheme.

4.1 Experimental design and performance results

For the evaluation of the normalization and fusion techniques, the average of GAR values for each pre-specified FAR value is reported. In order to compute FAR and GAR, first we need to generate all possible genuine and imposter matching scores and then set a threshold for deciding whether to accept or to reject a match. A genuine matching score is obtained when two feature vectors correspond to the same individual are compared, and an imposter matching score is obtained when feature vectors from two different individuals are compared. The FAR value is the fraction of the number of falsely accepted imposter scores divided by the total number of impostor scores. The GAR value is the fraction of the number of correctly accepted genuine scores divided by the total number of genuine scores. These two factors are brought together in a (ROC) curve that plots the GAR against each value of FAR. The recognition performance of the face and voice systems when operated as unimodal systems is shown in figure 2. We see that the performance of face module is better than the voice module.

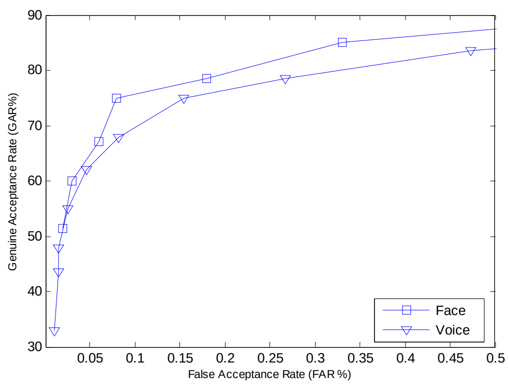

Fig. 2 ROC curves for individual matcher.

The performance of the multimodal system has been studied under some of different normalization and fusion techniques. The simple sum of scores see Eq. (6), the max score see Eq. (7), and the product-score see Eq. (9) fusion methods were applied on the normalized scores. The normalized scores were obtained by using one of the following techniques: MM normalization Eq. (1), ZS normalization Eq. (2), MAD normalization Eq. (3), Tanh normalization Eq. (4), and MDS normalization Eq. (5). A MATLAB codes was written to evaluate the effect of each normalization method on a system performance for different fusion methods. Table 1 summarizes the average GAR of the multimodal system for different normalization and fusion schemes, at a FAR of 0.5%. Note that GAR values for the two individual matchers face and voice are found to be 46.43 and 25.14 respectively. The average value for each fusion technique on all normalization methods is computed. Also the average value for each normalization method on all fusion technique is computed.

Table 1 Genuine acceptance rate (GAR) (%) of different normalization and fusion techniques at the 0.5% false acceptance rate (FAR)

Normalization Method	Fusion Method				Normalization average
	SSR	MAR	PRR	SPR	
MM	78.33	31.78	74.85	76.61	65.39
ZS	85.00	46.78	56.54	95.00	70.83
Tanh	84.28	46.78	83.43	85.89	75.09
MAD	86.94	47.14	67.5	94.52	74.03
MDS	96.19	46.78	73.93	93.57	77.62
Fusion average	86.15	43.85	71.25	89.12	

As seen from Table 1, all of the fusion methods lead to better performance than the voice matcher. But with respect to the matcher face approximately all fusion methods is better than this matcher expect MAR/MM. In the average, we note that, the normalization technique MDS is 77.62 which is the best one for all normalization methods. Also, in the average the SPR fusion is 89.12 which is the best one for all fusion techniques.

The significant improvement achieved by multimodal biometric system has also been highlighted by Ross et al. [5]. They showed that the system based on face, fingerprint, and hand geometry modalities can achieve a GAR of 98.6% (with FAR=0.1%) while ZS normalization combined with SSR-based fusion is used. In their experiments the best unimodal biometric system was fingerprint module, which has a GAR 83.6% and FAR of 0.1%. In Roee et al. [5], ZS normalization performed the best followed by Tanh normalization and MM normalization. In this paper, for the fusion between XM2VTS database and TIMIT database, MDS normalization performed the best, followed by DS normalization, MAD normalization, Tanh normalization, ZS normalization, and MM normalization.

Figure 3 shows the recognition performance of the system when the scores are combined using the SSR rule. We observe that a multimodal system employing the sum of scores provides better performance than the best unimodal system (face matcher) for all normalization techniques. For example, at a FAR of 0.1% the GAR of the face system is about 76%, while that of the multimodal system is high than 93% when MM normalization is used. This improvement in performance is significant and it underscores the benefit of multimodal systems. Among the various normalizations techniques, we observe that the MDS normalization technique outperform other techniques. This is because we control the parameter t in DS normalization to give us the overlapping region minimum.

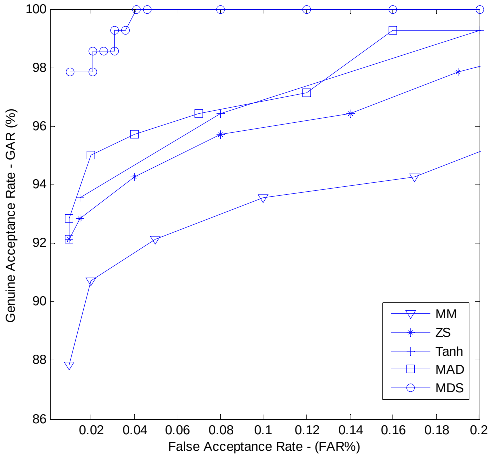

Fig. 3 ROC curves for sum of scores rule.

When we differentiate experimentally between the fusions techniques we founded that, the sum rule is the best followed by product rule, maximum rule, and minimum rule. So, we combined between the best two (sum rule and product rule) to obtain sum-product rule which give us a best result. Table 2 summarize the average of GAR of the multimodal system for the six normalizations methods and the three fusion techniques sum rule, product rule, and sum-product rule at a FAR of 0.5%. The parameters of the double sigmoid normalization were choose as follows: t is chosen to make the overlapping region is minimum, r_1 is the difference between t and the minimum of genuine scores, while r_2 is the difference between the maximum of imposter scores and t. In the face module we founded that $t = 0.16$, $r_1 = 0.0601$, $r_2 = 0.4068$ and in the voice module we founded that $t = 26$, $r_1 = 4.1392$, $r_2 = 0.7585$. The performance of the multimodal system using sum-product fusion is shown in fig. 4. Here MDS normalization provides better recognition performance compared to the other normalizations techniques. Also we observe that MAD normalization outperform ZS at low FARs, whereas they are the same performance at higher FARs.

Table 2 Genuine acceptance rate (GAR) (%) of different normalization and fusion techniques at the 0.5% false acceptance rate (FAR)

Normalization Method	Fusion Method		
	SSR	PRR	SPR
MM	78.33	74.85	76.61
ZS	85.00	56.54	95.00
Tanh	84.28	83.43	85.89
MAD	86.94	67.5	94.52
DS	95.7	76.43	92.62
MDS	96.19	73.93	93.57

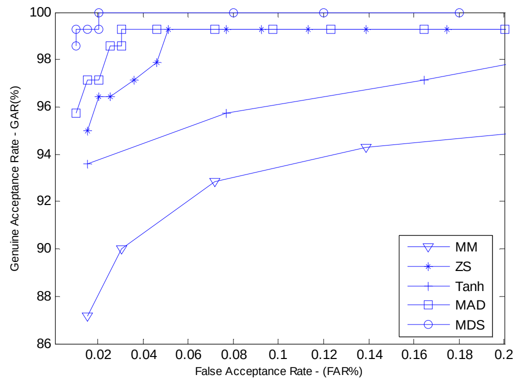

Fig. 4 ROC curves for sum-product rule

The new fusion technique sum-produced is compared with the four fusion considered in this paper. Sum-produced give us better results and outperform the three fusion technique (max, min, and product) for all normalization techniques. As in [6] and other reference, sum rule was outperforming many of fusion techniques, so we choose the sum rule to comparing. Fig. 5 shows the performance of sum-product compared with sum rule when we use ZS-Scores normalization technique.

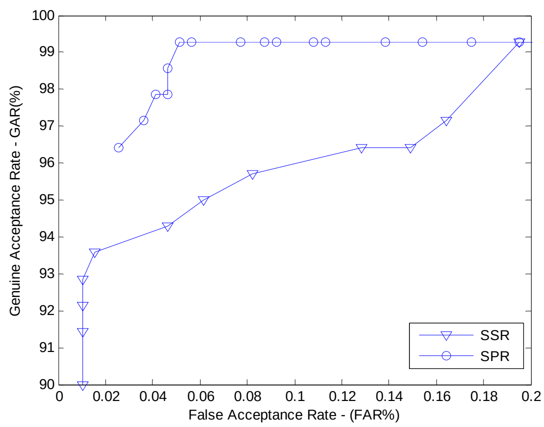

Fig. 5 ROC curves for SSR and SPR by ZS

5. Conclusion

This paper examines the performance of multimodal biometric system using face and voice matchers on 140 individuals. The performance of sum rule, product rule, minimum rule, and maximum rule fusion on multimodal system (face and voice) has been evaluated. Our experimental results showed that multimodal biometric system which combines multiple biometric data can achieve significantly better performance compared to a single biometric system. Also we have demonstrated that the normalization of scores prior to combining them improves the recognition performance of a multimodal biometric system that uses the face and voice traits for user authentication.

We have introduced new fusion method to accomplish matching scores level fusion of multimodal biometrics. This new fusion technique is preceded by double sigmoid normalization after we modified it. Our work shows that the proposed method can achieve better performance than the most popular fusion method such as (simple sum, product rule, maximum rule, and minimum rule). Future work will investigate alternative normalization and fusion methods, while honing our proposed testing methodology. Also alternative traits such as fingerprint and hand geometry can be used.

Reference

[1] S. Prabhakar, S. Pankanti,and A. K. Jain" Biometric recognition: security and privacy concerns" IEEE Security and Privacy, pp. 33-42, 2003.

[2] A. Ross, K. Nandakumar, and A. K. Jain" Handbook of Multibiometrics" Springer-Verlag, 2006.

[3] A. K. Jain, K. Nandakuman, and A. Ross "Score normalization in multimodal biometric systems" Pattern Recognition 38, pp. 2270-2285, 2005.

[4] M. He, S. Horng, P. Fan, R. Run, R. Chen, J. Lai, M. K. Khan, and K. O. Sentosa "Performance evaluation of score level fusion in multimodal biometric systems" Pattern Recognition 43, pp. 1776-1788, 2010.

[5] F. R. Hampel, P. J. Rousseeuw, E. M. Ronchetti, and W. A. Stahel "Robust Statistics: The approach based on influence functions" Wiley, New York, 1986.

[6] R. Snelick, U. Uludag, A. Mink, M. Indovina, and A. K. Jain "Large scale evaluation of multimodal biometric authentication using state-of-the-art systems" IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, No. 3, pp. 450-455, 2005.

[7] R. Cappelli, D. Maio, and D. Maltoni "Combining fingerprint classifier" Proceedings of First International Workshop on Multiple Classifier Systems, pp. 351-361, 2000.

[8] L. Hong and A. K. Jain, "Integrating faces and fingerprints for personal identification" IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 12, pp. 1295-1307, 1998.

[9] R. Snelick, M. Indovina, J. Yen, and A. Mink, " Multimodal Biometrics: Issues in Design and Testing " Proceedings of the Fifth International Conference on Multimodal Interfaces, pp. 68-72, 2003.

[10] R. Brunelli and D. Falavigna, "Person identification using Multiple Cues," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, pp. 955-966, 1995.

[11] J. Kittler, M. Hatef, R. P. W. Duin, and J. Mates, "On Combining Classifiers," IEEE, Transactions on Pattern Analysis and Machine Intelligence vol. 20, pp. 226-239,1998.

[12] S. Zafeiriou, A. Tefas, I. Buciu, and I. Pitas "Exploiting discriminant information in non-negative matrix factorization with application to frontal face verification" IEEE Transactions on Neural Networks, 17, pp. 683-695, 2006.

[13] J. Fortuna, P. Sivakumaran, A. Ariyaeeinia and A. Malegaonkar "Relative effectiveness of score normalization methods in open set speaker identification" Proceedings of the ODYSSEY 2004 the Speaker and language Recognition Workshop, May. 31-Jun. 3, Canon Research Centre Europe Ltd., Bracknell, Berkshire, UK, pp. 369-376, 2004.

[14] S. Ben-Yacoub, Y. Abdeljaoued, and E. Mayoraz, "Fusion of face and speech data for person identity verification" IEEE Transactions on Neural Networks, vol. 10, no. 5, pp. 1065-1075, 1999.

[15] A. Ross and A. K. Jain "Information fusion in biometrics" Pattern Recognition Letters 24, pp. 2115-2125, 2003.

[16] Y. Zuev and S. Ivanon "The voting as a way to increase the decision reliability" Foundations of Information/Decision fusion with applications to engineering problems, Washington, DC, USA, pp. 206-210, 1996.

[17] V. Conti, G. Milici, P. Ribino, F. Sorbello, and S. Vitabile "Fuzzy fusion in multimodal biometric systems" Proceedings of the 11th international conference, KES 2007 and XVII Italian workshop on neural networks conference on Knowledge-based intelligent information and engineering systems, pp. 108-115, 2007.

[18] S. Ribaric and I. Fratric "Experimental evaluation of matching-scores normalization techniques on different multimodal biometric systems" IEEE MELECON, Benalmadena (Malaga), Spain, May 16-19, 2006.