Multi-Modal Biometric Verification Based on FAR-score Normalization

Chengbo Wang, Yongping Li, Hongzhou Zhang, Lin Wang

Shanghai Institute of Applied Physics, Chinese Academy of Sciences, 2019 Jialuo Road., Jiading, Shanghai, P.R. CHINA

Summary
Fusion different biometrics is an effective way to design a biometric system with robust performance. To do this, normalization functions are employed. However, these functions can not follow the distributions of scores from distinct classifiers. Consequently different normalization errors are introduced. In this paper, the scores from different classifiers are converted into the corresponding false accept rate (FAR), which introduces smaller normalization error than traditional methods, and makes the fusion more operable. To further enhance the fusion result, a dynamic selection of fusion rule is implemented based on the discrepancy between scores of different classifiers. Experiments conducted on a multi-modal biometric system composed of face and fingerprint verification system show our methods are superior to conventional approaches.

Key words:
Biometrics; multi-modal biometric fusion; normalization; fusion method.

1. Introduction

Due to the various applications in verifying a claimed identity, many important verification algorithms of different biometrics [1-4] have been proposed during the past few years. Because of the noise introduced in both acquiring and processing procedures, the constraint of the environment and the robustness of the algorithm, a sole biometric usually can not be able to provide a satisfying result [5]. A possible way to alleviate this problem is to design multi-model recognition system which combines results of different biometrics together.

There are three kinds of multi-modal biometrics verification system: The first is the multi-algorithm system which employs different algorithms to verify a single biometric trait. The second is the multi-biometric system [5] that involves two or more distinct modals of biometric traits. The third is the hybrid system wherein the multi-algorithm and multi-biometric systems are integrated together.

The matching scores are the outputs of the classifiers measuring the similarities of the testing samples to the claimed class. they are different in numerical domains and have distinct distribution. Before combining these scores, they should be normalized to the homogeneous domain. Min-Max, z-score and *tanh* methods are adopted in [6] to normalize the scores at first and then the normalized scores were combined with the sum, min, med and max rules. Although such methods have achieved some satisfying fusion results, there are some deficiencies. Firstly, different normalization errors will be introduced, because distinct scores are normalized by a single form of function which could not keep the distributions. Secondly, when the normalized scores are combined, the conventional fusion rules overlook the discrepancy of the outputs of different classifiers, which is essential to the multi-modal biometric fusion.

The function between FAR and threshold, called as FAR-score curve in this paper (fig.1), is proposed to be adopted as the normalization function. In training stage, the FAR-curve of each classifier can be properly obtained because there are always enough negative instances to compute FARs. Thus every classifier has its own normalization function that follows the distribution of the scores and FARs. When score is normalized by the FAR-score curves, the normalized score is the probability of the classifier to accept a negative instance. After all scores from different classifiers are normalized, the conventional fusion rules, such as sum, min, med and max, can be used to compute a single scalar to make the final decision. In order to enhance the performance of the combination of the normalized scores, the method of dynamic selection fusion rule (DSFR) is put forward to select fusion rule according to the discrepancy of the outputs from different biometric systems.

The rest of the paper is organized as follows: Section 2 presents the method of normalizations with the FAR-score curve. In Section 3, DSFR are described. The experimental results are shown in the Section 4. At last we draw the conclusion in section 5.

2. Convert the Matching Score into the False Acceptance Rate

Score normalization for multi-classifier fusion refers to transform the various scores obtained by different classifiers into a common domain. Distinct matcher produce score diversely in numerical range and meaning, so the evaluation standards vary accordingly. It is

Manuscript received April 5, 2008
Manuscript revised April 20, 2008

necessary to normalize the scores into the homogeneous domain before combination. When normalizing scores of different classifiers, two factors should be considered. One is that the normalized scores should keep the discriminating information as much as possible; the other is that the normalization method should introduce equivalent error to different distributed scores.

In practice, classifier outputs a matching score s to reflect the similarity between the testing sample Z and the claimed class. In general, s can be modeled as [7]:

$$s = f[P(genuine \mid Z)] + \eta(Z) \qquad (1)$$

Where $f(\bullet)$ is a monotonic function and $\eta(\bullet)$ is the bias of the classifier and often be supposed to be zero.

Jain et. al recommend normalizing scores with a certain functions such as z-score and tanh functions as below [5]:

$$n = \frac{s - mean(S)}{std(S)} \qquad (2)$$

$$n = \frac{1}{2}[\tanh(0.01\frac{(s - mean(S))}{std(S)}) + 1] \qquad (3)$$

n is the normalized score, $mean(\bullet)$ and $std(\bullet)$ denote the arithmetic mean and standard deviation operators respectively, **S** is the set composing of scores from the classifier. Although these functions use the statistic some statistical characters of scores such as the means and variances, they do not follow the distributions of the scores from different classifiers. When they are used to normalize distinct distributed scores, diverse normalization errors are introduced into different scores due to the deviation between the score distributions and the function.

A novel normalization method is introduced here, which converts scores into false acceptance rate. When studying a typical Receiver Operating Characteristic (ROC) curve of a classifier, two kinds of probabilities relate to the scores: the false accept rate and false reject rate. From Fig.1, the false acceptance rates and the false rejection rates are functions of threshold (denoted as h). The two functions can be written as below:

$$f_{far}(h) = P(genuine| impostor, s < h) = \frac{\text{false positives}}{\text{negative instances}} \qquad (4)$$

$$f_{frr}(h) = P(impostor \mid genuine, s > h) = \frac{\text{false negatives}}{\text{positive instances}} \qquad (5)$$

When training a classifier, we can not get the exact forms of $f_{far}()$ and $f_{frr}()$, but a series of h can be used to calculate $f_{far}(h)$ and $f_{frr}(h)$, then both of the functions can be computed by interpolation. Once these two functions are available, a score can be transformed into FAR or FRR. The two curves are functions of monotonically with range of [0, 1] and only depend on the distribution between the error rates and the scores. Both of the two functions can be used to normalize the scores of the classifier.

However, only the FAR-score curve is used in the paper. Given a training set of K classes, (K>>1), m samples for each class, we have $mK(K-1)$ negative instances for computing the FAR, while the instances for FRR are mK which is usually much less than $mK(K-1)$. Since there are far more samples to compute the FAR than those for the FRR, the computation of FAR introduces less error than that of FRR, which can be seen from Fig.1.

The FAR-score curve, which represents the function between the false accept rate and the matching score of a classifier, can be learned especially through experiments. When it is obtained, scores can be transformed into FARs to reflect the global probability of a negative instance normalized scores obtained from different classifiers by a single function, which assumed that scores follow Gaussian distribution. Whilst our FAR-score normalization method compute the posteriori of P(genuine|imposter,$s<h$) through experiments, without any assumption of original score distribution but only suppose that samples follow a certain distribution as proposed in [8].

To learn the FAR-score curve, the FARs of a series threshold h should be calculated beforehand, as mentioned before hand. For a training set of K classes, each class has m samples, so there are $mK(K-1)$ imposter samples in total. For the jth classifier, at the ith threshold of h^j_i, if there are e impostor samples are accepted, the false acceptation rate is $far^j_i = e/(K(K-1)m)$. Using a set of thresholds ($h^j_{i-1} < h^j_i < h^j_{i+1}$), the FAR-score curve can be calculated. When a testing sample Z comes with a claim, the score s^j from the jth matcher can be normalized by the curve. If FAR monotonically increase with h^j_i, s^j is normalized by (6):

$$n^j = far_i^j(h_i^j) + \frac{dfar_i^j}{dh^j}\Big|_{h^j = h_i^j}(s_i^j - h_i^j) \qquad (6)$$

for $h^j_{i-1} \le s^j_i \le h^j_{i+1}$, otherwise, (7) is used to normalize s:

$$n^j = far_i^j(h_{i+1}^j) + \frac{dfar_i^j}{dh^j}\Big|_{h^j = h_i^j}(s_i^j - h_{i+1}^j) \qquad (7)$$

for $h^j_{i-1} \ge s^j_i \ge h^j_{i+1}$. When scores from all classifiers are normalized into FARs, the common fusion rules such as sum, min, med and max can be adopted to compute a single scalar to make a final decision.

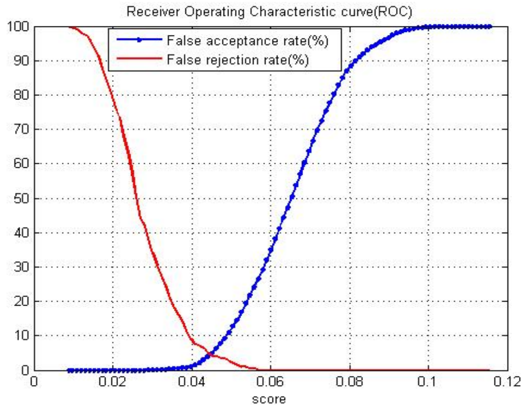

Figure 1: A typical ROC curve

3. Dynamic Selection of Fusion Rules

In the multi-biometric systems, the classifiers of distinct modal biometrics give independent outputs, and the discrepancy among the outputs from distinct classifiers are quite different. The fusion rules of min, med, max and sum [5] only make a comparison or compute the sum of the outputs, and ignore the discrepancy of the outputs from distinct classifiers. In this case, all of the conventional fusion rules do not make full use of the information of the normalized scores.

The dynamic selection fusion rules (DSFR) first computes the discrepancy of the scores from different classifiers, then chooses a suitable fusion rule to combine the scores together. When the discrepancy is smaller than a threshold, the scores from two classifiers are consistent and confirm to each other, thus the better result of the two classifiers could be chosen as the fusion result. If the discrepancy is bigger than a threshold, it is hard to judge which classifier gives a reasonable result, in this case the average rule is adopted to alleviate the risk. For a multi-biometric system as above, when the scores are normalize by FAR-score curve, suppose normalized scores are s_1 and s_2 given by face verification subsystem and fingerprint verification subsystem respectively, s is the fusion result, the realization of the DSFR is: first compute the difference of the two scores: $t=| s_1- s_2|$; then if $t<th$ $s=\min(s_1, s_2)$; else $s=\text{mean}(s_1, s_2)$ (th is a threshold).

For a multi-biometric system comprised by a face and a fingerprint verification subsystem, face image and fingerprint image of a probe sample are verified respectively with the scores denoted as (s_f, s_p). Now suppose two testing samples verified by the system and normalized by FAR-score algorithm, the normalized scores are: (5%, 3%), (20%, 3%). That is for the first sample, the subsystems takes the risk of 3% and 5% probabilities to accept a spoof attack; for the second one, the two traits with 20% and 3% probabilities from an impostor. It can be seen that the discrepancy between the outputs of the two classifiers is very little for the first sample, which means the two subsystems output consistent results; while there is great discrepancy for the other sample, which indicates that one of the two subsystems must lead to a wrong decision. In this case, the conventional fusion rules will lead to different fusion result. If the min rule is applied to combine the two outputs, it will focus on the classifier that gives better result. For the first sample, the 3% is acceptable, but the system takes greater risk to accept a false claim for the second sample. If the max rule is adopted, the fusion result are 5% and 20%, apparently this rule only takes account of the worse result of the two classifier. As to the sum rule, it sums results from all classifiers, so it alleviates effect of the worse result, at the same time it also depresses the better one. In this case, the first sample takes the conservative fusion result, but for the second sample, the fusion result is more reasonable than that from min and max rules. From this example, it can be seen that none of the fusion rules can achieve a reasonable result when used solely.

4 Experiments

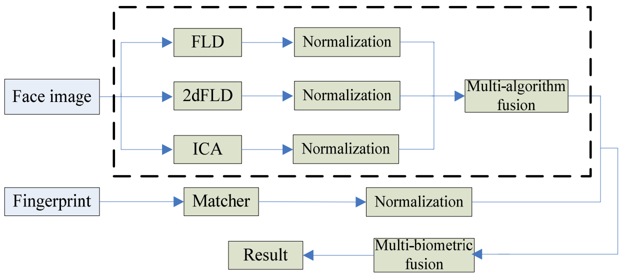

Fig.2: A combined biometric system

The experiments adopt the hybrid system in Fig2, which consists of a fingerprint verification subsystem and a multi-algorithm face verification subsystem. The fingerprint verification subsystem uses the algorithm in [1] and takes samples from the FVC2002 [10], Db1a, which enrolls images from 100 different fingerprints. Each fingerprint has 4 fingerprint images to train the system and 2 fingerprints as testing samples in the experiments. The multi-algorithm face verification subsystem employs LDA[11], ICA[12] and 2d-LDA[13] as three individual classifiers, and selects face images of 100 objects from XM2VTS database [9], each object has three face images used as training samples and two as testing samples. In the experiments, each face is corresponding to a unique fingerprint. A pair of testing sample contains a face image and a fingerprint image. Thus in the testing stage, there are 200 pairs of positive samples and 19800 negative samples.

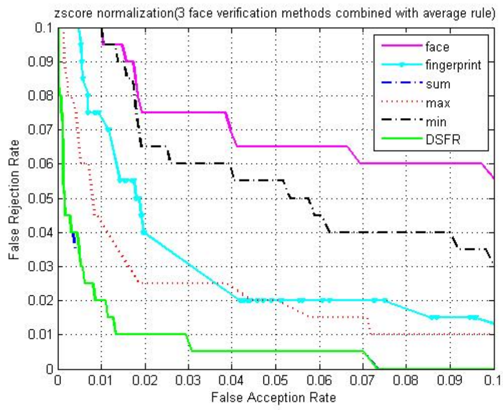

Fig.3a: zscore normalization (3 face verification methods combined with average rule)

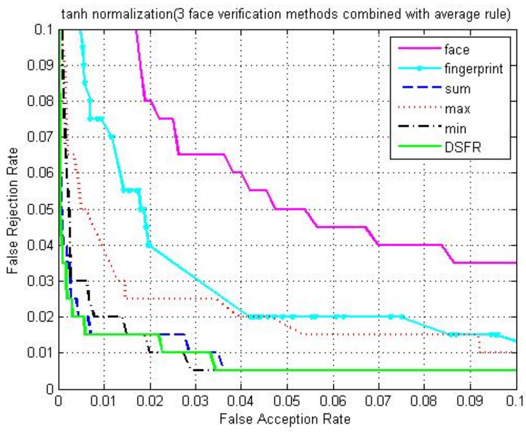

Fig.3b: tanh normalization (3 face verification methods combined with average rule)

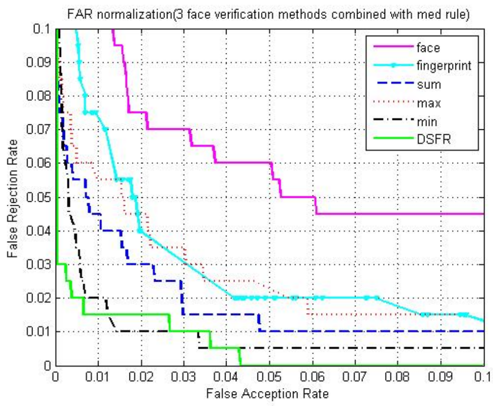

Fig.3c: FAR normalization (3 face verification method combined with med rule)

Fig.3a-Fig.3c are ROCs for the Hybrid system in Fig.2.; Table1 gives the important points on ROCs. Fig.3a and Fig.3b are the results of the z-score and tanh normalization, the face verification subsystem is combined with sum rule; in Fig.3c, scores are normalized by FAR-score curve and the face verification subsystem combined with med rule.

Studying the three figures, when the DSFR are not used, we find that the sum rule for z-score and tanh normalization methods gets the best fusion results and the min rule achieves the best fusion result if the scores are converted to FARs. Comparing the three methods, when FAR<1, tanh normalization method gives the best fusion result which FRR=4% when FAR=0.1%. But z-score and FAR normalization methods achieve EER (Equal Error Rate) of 1.35%, which is better than that from the tanh normalization method. As the normalized scores are combined with DSFR, although there is no improvement for the z-score normalization, the FRR decreases to 3.5% from 4% when FAR=0.1% for the tanh normalization method, and the system gives 3% FAR at the similar false accept rate with the FAR-score normalization method. From Fig.3c, we can even see that if the FAR decrease to 0.01%, the DSFR gets 3% FRR, which is by far better than any other fusion methods mentioned in this paper. Jain and Ross achieved the similar fusion result with a system involved face, fingerprint and hand geometry traits [6].

In the experiments, the DSFR perform better than other fusion rules, which shows that the discrepancy of score from different classifiers should be considered when designing a multi-classifier system. From the experiments above, the FAR-score normalization combined with DSFR gives the best fusion results. The functions of z-score or tanh do not follow the distribution of the scores from different classifier. Different normalization errors are introduced as a single form of function is used to normalize distinct distributed scores. When the DSFR are applied, the normalized scores are compared on different ground. As the scores are normalized by the FAR-score curves, the transformed scores are the probability of the classifier to accept a negative instance, which can be compared fairly.

5 Conclusions

The FAR-score curve of each classifier is computed without assumptions of observing any distributions, and then scores from all classifiers can be normalized by its own FAR-score curve. Therefore, the method can be adapted to scores from any classifiers. When the scores are normalized with FAR-score curve, the normalized scores show the probabilities of accepting an impostor. This makes the normalized scores are easer to compare. To achieve better fusion result for the multi-biometric system, dynamic selection fusion rule according to the discrepancy of the outputs from different classifiers is proposed. Experimental results for a combined multi-modal biometric verification system show our approaches' efficiency and effectiveness compared to conventional methods.

The false rejection information of the matching scores has been discarded in our current work because there are usually not enough samples to calculate FRR-score curve precisely. When the scores are normalized by the FRR-score curve, more normalization errors will be introduced that may lead to even worse fusion results. However, matching scores normalized by FRR-score are another part of important information of the matching scores that might be made use of by second-order combination. These will be a big part of our further research.

Table1:Key error rate data (%) in Fig3a, Fig3b and Fig3c.

		EER	FRR FAR=0.1	FRR FAR=1	FAR FRR=1
fingerprint		3	>10	7.5	>10
zscore	Face	5.5	>10	9.5	>10
	sum	1.35	7	2	1.33
	max	2.5	>10	4.3	7.12
	min	5.2	>10	>10	>10
	DSFR	1.35	7	2	1.33
tanh	Face	5	>10	10	>10
	sum	1.5	4	1.5	2
	max	1.5	8.5	2	2
	min	2.5	9.5	4	9.2
	DSFR	1.5	3.5	1.5	2.3
FAR-score curve normalizeation	Face	5.2	>10	>10	>10
	sum	2.5	7.5	4	4.76
	max	3	8.5	5.5	>10
	min	1.35	7.5	2	1.41
	DSFR	1.5	3	1.5	2.67

References

[1] He Y., Tian J., Luo X., Zhang T.: "Image enhancement and minutiae matching in fingerprint verification", Pattern Recognition Letters 24 (2003) 1349-1360.

[2] Belhumeur P. N., Hespanha J. P., and Kriegman D. J.: "Eigenface vs. Fisherfaces: Recognition using class specific linear projection," IEEE Trans on PAMI 19(7) (1997)711-720.

[3] Bartlett M.S., Movellan J.R. and Sejnowski T.J.: "Face recognition by independent component analysis", IEEE Trans. on Neural Networks, vol. 13, no.6, pp. 1450-1464, 2002.

[4] Li M., Yuan B., "2D-LDA: A statistical linear discriminant analysis for image matrix", Pattern Recognition Letters, vol. 26, no.5, pp. 527-532, 2005.

[5] Ross A. and Jain A.: "Information fusion in biometrics", Pattern Recognition Letters, 24(2003) 2115-2125.

[6] Jain A., Nandakumar K. and Ross A.: "Score normalization in multimodal biometric system", Pattern Recognition 38(2005) 2270-2285.

[7] Kittler J., Hatef M., Duin R. P.W. and Matas J.: "On combining classifiers", IEEE Trans. on PAMI vol. 20, No.3 pp.226-239, 1997.

[8] Duda R.O., Hart P.E. and Stork D.G.: Pattern Classification, Wiley, New York, 2001.

[9] Messer K., Matas J., Kittler J., Luettin J., Maitre G., "XM2VTSDB: the extended M2VTS database", Proceeding of Audio and Video-based Biometric Person Authentication, Washington DC, USA, pp.72-77, 1999.

[10] http://bias.csr.unibo.it/fvc2002/

[11] P. N.Belhumeur, J. P.Hespanha, and D. J. Kriegman. Eigenface vs. Fisherfaces: Recognition using class specific linear projection. IEEE Trans on PAMI 1997, 19(7): 711-720.

[12] M.S.Bartlett, J.R.Movellan and T.J.Sejnowski. Face recognition by independent component analysis [J]. IEEE Trans. On Neural Networks, 2002, 13(6). 1450-1464,.

[13] M. Li, B. Yuan. 2D-LDA: A statistical linear discriminant analysis for image matrix. Pattern Recognition Letters, 2005, 26 (5): 527-532.

Chengbo Wang received MS degree in Signal and Information Processing in 2005 from Dalian University of Technology. He is a PhD candidate of Graduate school of Chinese Academy of Sciences now. His research is focused on digital image processing and pattern recognition.

Yongping Li received MS degree in the Electrical Engineering in 1989 from the Graduate University of Chinese Academy of Sciences. He worked as research fellow at Shanghai Institute of Applied Physics for various research projects on computerized instrumentations from 1989 to 1995. He joined the Center for Vision Speech and Signal Processing, Univ. of Surrey, UK in 1996 with a Royal Society fellowship, and received PhD in Pattern Recognition, with emphasis on Biometrics, in 2000 under the supervision of Professor Josef Kittler. He worked on speech signal processing at a startup company in Silicon Valley, California from 2001 to 2002. From 2003 to present, he is a research professor and PhD student supervisor at Shanghai Institute of Applied Physics, and a part-time PhD student supervisor at the Center for Biometrics and Security Research (CBSR), Institute of Automation, Chinese Academy of Sciences. His current research interests include biometric algorithms for face and fingerprint recognition, biometric solution integration, biometrics system performance evaluation, testing, and standardization. Dr. Li is a member of the IEEE, IEEE computer society and IEEE instrumentation and measurement society.