Quasi-convex Optimization of Metrics in Biometric Score Fusion

Yanmin Gong, Jiansheng Chen, Guangda Su
Department of Electronic Engineering
Tsinghua University
Beijing, China
Email: {gym09@mails., jschenthu@mail., susu@}tsinghua.edu.cn

***Abstract*—In this paper, we address the problem of score fusion in biometric authentication. Single valued metrics related to the receiver operating characteristics (ROC) curve, such as Equal Error Rate (EER) and False Rejection Rate (FRR) when False Acceptance Rate equals zero , are extensively used for evaluating biometric authentication performances. Various requirements and preferences, for example, lower EER, or smaller FRR, may be imposed on biometric authentication systems in different application scenarios. We propose a novel method of score fusion based on quasi-convex optimization to directly improve biometric authentication metrics. Experiments based on a face recognition system demonstrate the effectiveness of the proposed method.**

***Keywords*-quasi-convex optimization; biometric authentication; score fusion**

I. Introduction

Previous research has demonstrated that information fusion is of vital importance for improving the performance of biometric authentication systems [1]. According to different inputs of the fusion process, biometric fusion can be divided into four levels: (i) sensor level, (ii) feature level, (iii) matching score level, and (iv) decision level. Among all these levels, the score level fusion is commonly preferred because matching scores are easily available and contain sufficient information to distinguish genuine and imposter matching cases.

There are three major approaches for consolidating biometric matching scores [2]. One approach is to construct the output scores from individual matchers into a feature vector, so that the score fusion problem can be formulated as a two-class classification problem, in which the matching score is classified as either "Genuine" or "Imposter" [3], [4]. Another approach, or the combination approach, combines individual matching scores into a single-scalar overall matching score, which is then used for the final decision [5]–[9]. Maximizing the area under the ROC curve (AUC) [8], for example, is a combination approach in which the area under the ROC curve is optimized as a polynomial of individual matching scores. The third approach, density-based score fusion, is based on the likelihood ratio test. It requires explicit estimation of probability densities of genuine and imposter matching score [2].

Under different situations, we have various requirements and preferences for the authentication system, for example, lower EER, or smaller FRR_0. However, these performance requirements are usually not directly addressed in previous score fusion methods. For example, the likelihood approach may not be suitable for the biometric systems in which EER is a primary concern. In the AUC method, the relationship between the performance metrics and the optimization objective is only implicit. Kumar et. al [9] considers optimizing a global cost E defined as a weighted sum of FAR and FRR. However, only heuristic optimization methods such as PSO are used to approximately solve it. To solve this problem, we propose a novel score fusion method in which the score fusion problem can be formulated into a quasi-convex optimization, which can be solved efficiently. Experiments show that our method is able to achieve better authentication accuracy in comparison with the AUC method in terms of certain metrics.

The rest of this paper is organized as follows. Section II elaborates the mathematical backgrounds and formulations of our method. Experimental results are presented in Section III. Conclusions are drawn in Section IV.

II. Optimization Metrics in Score Fusion

We follow the problem formulation proposed by Kruger [10]. Let $\vec{l_1}, \vec{l_2}$ be the feature vectors of two biometric identifiers for matching. Suppose that $\vec{l_1}$ and $\vec{l_2}$ share the same length N, as is a common practice in biometric authentication. The weighted matching score between $\vec{l_1}$ and $\vec{l_2}$ is defined by the following equation, in which $s_i(\vec{l_1}, \vec{l_2})$ is the similarity score of the i^{th} feature and w_i is the weight assigned to the i^{th} feature,

$$s(\vec{l_1}, \vec{l_2}) = \sum_{i=1}^{N} w_i * s_i(\vec{l_1}, \vec{l_2}). \quad (1)$$

The matching score vector $\vec{s} = (s_1, s_2, \ldots, s_N)^T$ consists of all N feature similarity scores between two feature vectors. Suppose that $s_i > 0$ and larger s_i indicates higher similarity for the i^{th} feature. The weight vector $\vec{w} = (w_1, w_2, \ldots, w_N)^T$ is constrained by two conditions: $w_i \geq 0$ for $i = 1, 2, \ldots, N$, i.e., $\vec{w} \succcurlyeq 0$, and $\sum_{i=1}^{N} w_i = 1$.

Given the matching score vector $\vec{s} = (s_1, s_2, \ldots, s_N)^T$, the score fusion problem becomes how to assign weights to these matching scores. In our approach, we achieve this

978-0-7695-4541-7/11 $26.00 © 2011 IEEE
DOI 10.1109/ICIG.2011.50

task by using quasi-convex optimization of authentication metrics (QOM). Our main idea is that by utilizing domain specific knowledge on feature similarity score distributions, the problem becomes quasi-convex in nature. One of the most attractive properties about a convex optimization problem is that it has no local minimums that are not global. Thus the quasi-convex optimization can then be efficiently solved.

To start with, we define a genuine matching score matrix G. Each column of G is a matching score vector from one genuine matching on a training set. Obviously, $G^T * \vec{w}$ is a vector consisting of the weighted matching scores defined in Equation (1) of all the genuine matching. We further define μ_G and σ_G as the mean and standard deviation of $G^T * \vec{w}$ respectively. Moreover, we define a imposter matching score matrix I. Each column of I is a matching score vector from one imposter matching on a training set. Similarly, $I^T * \vec{w}$ is a vector consisting of the weighted matching scores defined in Equation (1) of all the imposter matching. μ_I and σ_I are the mean and standard deviation of $I^T * \vec{w}$ respectively.

A. Minimizing EER ($QOM - EER$)

EER is defined as the value of the error rate at the point where FAR equals FRR. Poh et al. [11] introduce the $F - ratio$ factor to measure the separation of matching score distributions between genuine and imposter classes. $F - ratio$ is defined as follows

$$F - ratio = (\mu_G - \mu_I)/(\sigma_G + \sigma_I), \tag{2}$$

which is a function of the weight vector $\vec{w}$ as μ_G, μ_I, σ_G, and σ_I are all functions of $\vec{w}$. Moreover, it has been proven in [12] that if the genuine and imposter matching score distribution are both Gaussian, EER can be explicitly expressed by the following equation

$$EER = (1 - erf(F - ratio/\sqrt{2}))/2 \tag{3}$$

where $erf()$ is the Gauss error function.

Since $erf()$ is monotonically increasing, minimizing EER is equivalent to maximizing $F - ratio$, or minimizing $1/F - ratio$. Actually, $1/F - ratio$ is a quasi-convex function of $\vec{w}$. The denominator of $1/F - ratio$, $\mu_G - \mu_I$, is the sum of two affine functions and is therefore affine itself. To be more specific, it is concave. Under our problem formulation where larger matching score indicates higher similarity, the mean genuine matching score should always be larger than the mean imposter matching score in a reasonable biometric authentication application scenario, indicating that $\mu_G - \mu_I$ is positive. In the numerator, σ_G is convex because it is the composition of a convex function (the standard deviation) with an affine expression $G^T * \vec{w}$. The same argument holds for σ_I. Also, both σ_G and σ_I are non-negative. Therefore, $\sigma_G + \sigma_I$ is non-negative and convex. $1/F - ratio$ is a quasi-convex function of $\vec{w}$ because it is a fractional function of a non-negative convex function over a positive concave function. Therefore, the score fusion problem of EER minimization can be modeled as the following quasi-convex optimization, called $QOM - EER$:

$$\min_{\vec{w}} \frac{1}{F - ratio} \tag{4}$$

s.t.

$$\mu_G - \mu_I \geq 0,$$

$$\vec{w} \succcurlyeq 0,$$

$$\sum_{i=1}^{N} w_i = 1$$

where all the equality and inequality constraints are affine.

There is another issue that needs further discussion. A sufficient condition for Equation (3) is the Gaussianity of genuine and imposter matching scores [11]. This widely used assumption has been proved to be effective in theoretical analysis as well as in practical application for biometrics [5], [13]. In our approach, elements in the weighted score vector $G^T * \vec{w}$ are required to follow a Gaussian distribution for all feasible values of $\vec{w}$. Therefore, we make a stronger assumption that the similarity scores of different features are independently and normally distributed. The same argument holds for the imposter score matrix I.

1) Minimizing FRR_0 ($QOM - FRR_0$): In biometric applications for security purpose, false acceptances of unauthorized users are usually regarded as serious threats. More seriously, in-field measurements of the probability of such a system failure, or FAR, after launching a biometric system is very difficult or even practically infeasible. Therefore, FRR_0, the smallest possible value of FRR at zero FAR, is considered as an important performance metric for evaluating biometric authentication systems for security purpose.

Geometrically, FRR_0 equals the area of the shadowed region under the genuine matching score probability distribution curve, as shown in Figure 1. If we impose the Gaussianity assumption to the genuine matching score, FRR_0 can be expressed by the following equations, which can be achieved by integrating a Gaussian function,

$$FRR_0 = (1 - erf(1/\sqrt{2}\eta))/2, \tag{5}$$

$$\eta = \sigma_G/(\mu_G - \max(I^T * \vec{w})), \tag{6}$$

where $\max(\vec{x})$ means the largest element of $\vec{x}$. According to the monotonicity of $erf()$, minimizing FRR_0 is equivalent to minimizing η. As explained before, the numerator σ_G is convex and non-negative, and the denominator $\mu_G - \max(I^T * \vec{w})$ is concave. Hence, when $\mu_G - \max(I^T * G)$ is positive, η is a quasi-convex function which can be minimized efficiently.

The positivity of $\mu_G - \max(I^T * \vec{w})$ can be guaranteed by simply adding $\mu_G - \max(I^T * \vec{w}) > 0$ as a constraint of the optimization problem. The possibility does exist for $\mu_G - \max(I^T * \vec{w})$ to remain non-positive for all feasible values of

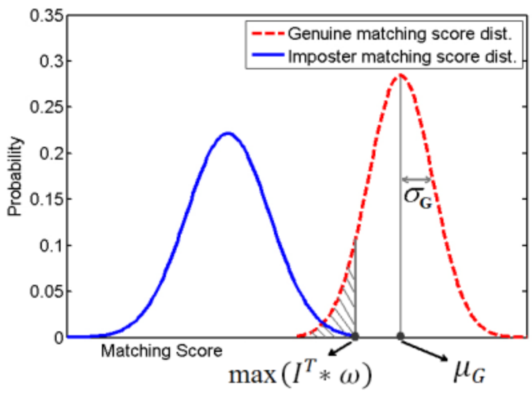


Figure 1. Illustration of FRR_0

$\vec{w}$. Under such a circumstance, FRR_0 is always larger than or equal to 50%, and modifications much more significant than adjusting the feature weights should be considered for the biometric authentication system under evaluation.

Another problem is that there may be major outliers in the training set that may cause overtraining. According to the definition (6) of η, only one imposter score vector will be chosen by the function $\max(I^T * \vec{w})$, leading to possible instability of the resultant weight vector $\vec{w}$. To solve this problem, we use $\Psi(I^T * \vec{w}, r)/r$ as a proper estimation of $\max(I^T * \vec{w})$, where $\Psi(\vec{v}, r)$ is the sum of r largest elements of vector $\vec{v}$. The quasi-convexity of η is still preserved since $\Psi()$ is convex. Hence, we redefine η as follows

$$\eta = \sigma_G/(\mu_G - \Psi(I^T * \vec{w}, r)/r), \tag{7}$$

which generalize the definition (6). Then, the score fusion problem of FRR_0 minimization can be modeled as the following quasi-convex optimization, called $QOM-FFR_0$:

$$\min_{\vec{w}} \eta \tag{8}$$

s.t.

$$\Psi(I^T * \vec{w}, r)/r - \mu_G < 0,$$

$$\vec{w} \succcurlyeq 0,$$

$$\sum_{i=1}^{N} w_i = 1$$

where the first inequality constraint is convex and the other two are affine. Strictly speaking, when $r \neq 1$, the solution to the problem $QOM - FRR_0$ does not correspond to the optimum FRR_0. In $QOM - FFR_0$, FRR is optimized at the point where the rejection score threshold equals $\Psi(I^T * \vec{w}, r)/r$ rather than $\max(I^T * \vec{w})$. However, in biometric system evaluation, the number of imposter matching is usually a large number [14], which means that r is typically far smaller than the total number of imposter matching. Thus, FAR at this point is very small. In our experiments, we use 249500 imposter matching and choose the largest value for r as 40 so that our method results in a feature weight vector that minimizes the FRR at a point where FAR is smaller than or equal to $1.6 \times 10^{-4} (40/249500)$.

Figure 2. Illustration of face graph

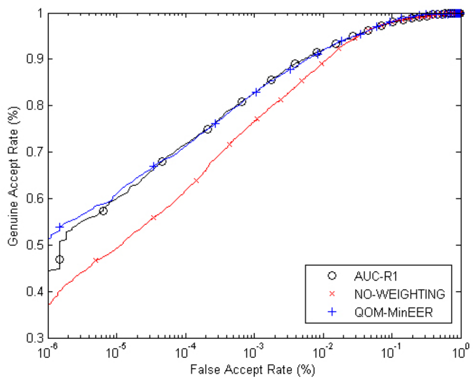
The above two metrics are most commonly used in practical authentication systems. Optimization of other metrics can also be formulated in a similar way.

III. Experiments and Results

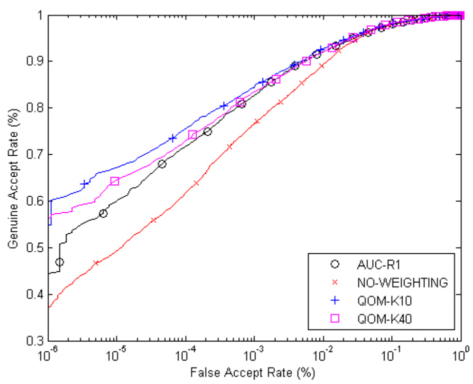
We apply our method to feature scores extracted from two face databases: the FERET face database [15], [16] and CAS-PEAL-R1 database [17]. We choose 1000 subjects, each with two frontal images, from the FERET database. These 1000 subjects are divided into training set and testing set. As the size of training set changes, experimental results vary accordingly. In the experiment, we choose the size of training set as 500, which means 500 subjects for training and 500 for testing. A complete one-to-one verification experiment on each set consists of 500 genuine matches and 249500 imposter matches. For the CAS-PEAL-R1 database, we choose two expression subsets, "laugh" and "frown", of 320 subjects. We compare both subsets with the corresponding normal images, getting experimental results of two different expressions. For each expression subset, we get 160 genuine matches and 25440 imposter matches. Experiments for each approach are repeated for at least 12 times, in which 500 training samples are randomly selected each time.

We use the CSU public version of the baseline elastic bunch graph matching (EBGM) [18] for feature extraction. In EBGM, descriptors for local facial features are extracted at nodes of a sparse graph suited to the face. For each face image, 25 landmarks as well as other 55 interpolated points among these landmarks are combined as the face template. Hence, the dimensionality of the feature vector is 80. An illustration of the face graph is shown in Figure **??**. There are 25 nodes in the graph which represent 25 landmarks where features are extracted. The remaining 55 landmarks are located at midpoints of lines in the graph.

The experimental results on both FERET and CAS-PEAL-R1 databases show that our method has a better performance over both equal weighting method and the AUC method. By referring to equal weighting method, we mean the method that simply uses average value of elements in each matching score vector $\vec{s}$ as the overall matching score. The results



(a)



(b)

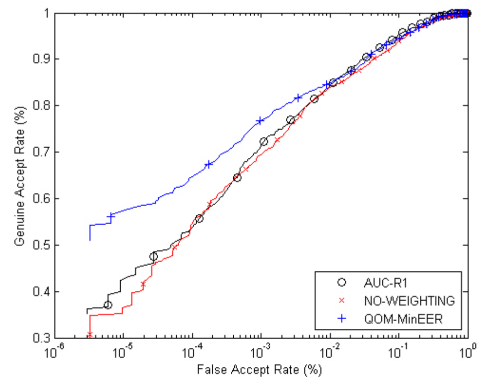
Figure 3. Comparison of ROC curves on FERET: (a) $QOM-EER$, AUC, and equal weighting; (b) $QOM-FRR_0$, AUC, and equal weighting.

Table I
COMPARISON OF THREE APPROACHES

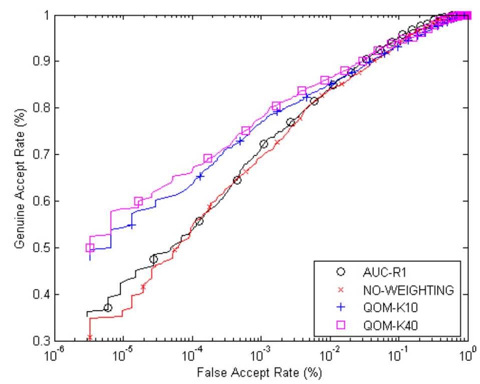
Database	Metric	Equal Weights	AUC	QOM
FERET	EER	0.0412	0.0384	**0.0363**
	FRR_0	0.4852	0.3862	**0.3152**
PEAL (laugh)	EER	0.0358	0.0171	**0.0034**
	FRR_0	0.2214	0.1200	**0.0717**
PEAL (frown)	EER	0.0725	0.0716	**0.0682**
	FRR_0	0.3227	0.3133	**0.0644**

from FERET and CAS-PEAL-R1 database are depicted in Figure 3, Figure 4, and Figure 5.

From these figures, we see that our method improves the ROC curve significantly in comparison with the equal weighting approach. Since AUC method [8] is originally proposed under the reduced polynomial model which is nonlinear when the order of the polynomial is greater than 1, for fair comparison, we only adopt the first order polynomial model of AUC approach. This corresponds to AUC-R1 in these figures. As to $QOM-K10$ and $QOM-K40$, we mean $QOM-FFR_0$ method when $r=10$ and $r=40$ respectively, where r is the parameter to approximately solve FRR_0 as defined in Equation (7). Since AUC aims to maximize the area under ROC curve, we observe that



(a)



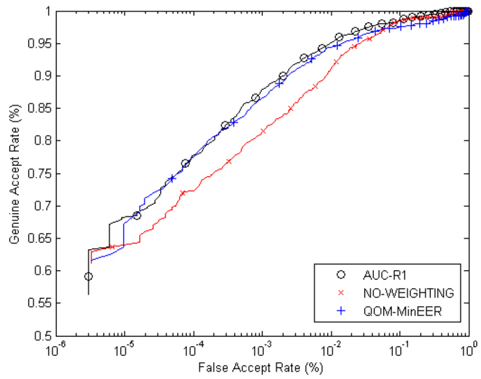
(b)

Figure 4. Comparison of ROC curves on expression subset 'laugh' of PEAL: (a) $QOM-EER$, AUC, and equal weighting; (b) $QOM-FRR_0$, AUC, and equal weighting.

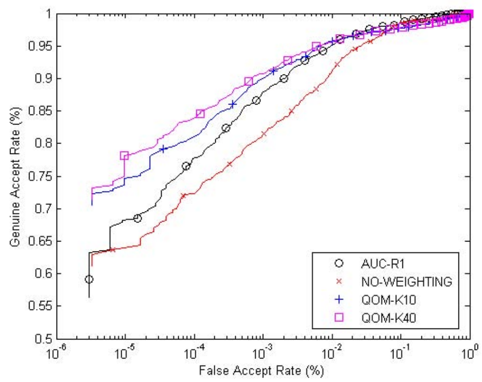
ROC curves of our approach are not always better than AUC. When FAR approximates 1, ROC curve of AUC-R1 is usually above that of our approach, which indicates a better performance. However, as FAR approaches zero, which is common in practical applications, our methods outperform the AUC-R1 method. We then conclude that under linear weighting conditions, our method works better than the AUC approach.

To further demonstrate the advantage of the QOM method on metric-oriented optimization, we present the results in Table I. Note that in all three approaches, we calculate FRR when FAR equals a small value 10^{-5} as an approximation of FRR_0. The results of QOM for EER and FRR_0 in the table are generated by $QOM-EER$ and $QOM-FRR_0$ respectively. From the table, it is clear that for most of the experiments, the performance of our proposed methods are better than the AUC method, especially for optimization of the metric FFR_0.

We mentioned earlier that we formulate the score fusion problem into a linear model. We choose this model for several reasons. For one reason, approaches to solve linear problems are extensive and mature, compared with nonlinear models. Mathematical tools offered by convex optimization,



(a)



(b)

Figure 5. Comparison of ROC curves on expression subset 'frown' of PEAL: (a) $QOM-EER$, AUC, and equal weighting; (b) $QOM-FRR_0$, AUC, and equal weighting.

for example, can be applied to linear problems easily, while solutions to nonlinear problems remain unsolved with high computational cost or instability, especially when the size of training database is hugh. The computation cost of our method is illustrated in Figure 6. The above experiments are performed on a desktop PC equipped with an Intel Pentium D 3.4GHz CPU and 3GB physical memory. In a MATLAB based implementation, the "minimization process of EER" in $QOM-EER$ consumes 3375 seconds and the "minimization process of FRR_0" in $QOM-FRR_0$ consumes 1671 seconds. The time consumption is comparable to that of the feature extraction and the one-to-one matching. We further observe that the running time is apparently affected by the total number of imposter matching used as training samples. Experiments are performed to discover their relationship. The linearity of the resultant curves indicates the scalability of our method in terms of the time complexity.

There is another advantage of linear model for our research in face recognition. In our experiments, We generate matching score vectors with EBGM baseline [18], which would be used as matching score vector $\vec{s}$ in score fusion. Besides the improvements of performance, we can observe

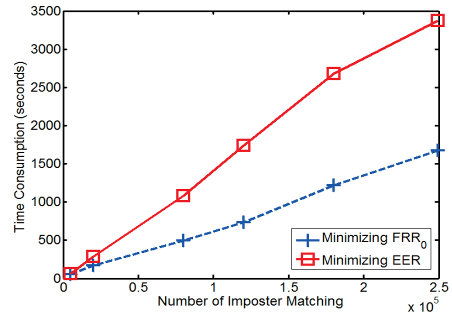


Figure 6. Running time vs the number of imposter matching

Table II
TOP 10 AND LAST 10 WEIGHTS SORTED FROM THE SCORE FUSION

Top 10	6	10	28	80	4	5	18	20	27	43
Last 10	14	49	50	11	37	39	66	67	78	79

other interesting phenomena. Table II shows landmarks which are assigned the top and last ten largest weights. Numbers in the table are index numbers of landmarks ranging from 1 to 80. Note that the order of landmarks is based on average weights generated in different database configurations. Although specific order of landmarks in each experimental result changes, the average order does present a general tendency that which landmark means more to face recognition (FR). The feature weights actually reflect the discriminative power and stability of different facial points across expression changes. Comparing experimental results of different expression configurations, some landmarks are sensitive to expressions, while other landmarks are important for all the expressions. This coincides with our original guess that noses are stable during expression changes and eyes contribute a lot in face recognition. It has been revealed that conceivable expression changes are actually driven by the movements of different Action Units (AU) consisting of muscles [19]. The number of AU in nose is only one, the smallest among all facial organs. In contrast, the mouth contains more than 15 AUs. Although the forehead and the hair are also robust to expression changes, facial points inside these areas are also associated with relative low weights because they have much lower interclass variation compared to other facial landmarks.

IV. CONCLUSION

In this paper, we propose a novel score fusion method, which can be easily adapted to different biometric authentication scenarios. We validate our method by making extensive verification experiments on two face databases. Under linear weighting formulation, we gain a significant improvement in the authentication accuracy. In terms of certain metrics, the proposed method performs better than some previous score fusion methods. In the future, we plan to incorporate the reduced polynomial model with nonlinear items as in [8] into the problem formulation so as to further improve the overall performance.

ACKNOWLEDGMENT

Portions of the research in this paper use the FERET database of facial images collected under the FERET program, sponsored by the DOD Counterdrug Technology Development Program Office [14], [15]. The research in this paper use the CAS-PEAL-R1 face database collected under the sponsor of the Chinese National Hi-Tech Program and ISVISION Tech. Co. Ltd [17].This work was partially supported by a National 973 Project: Fundamental Research on Multi-Domain Collaboration for Broadband Wireless Communications (Project No. 2007CB310600), and a Tsinghua Chuan Xin Fundation Project: Robust Face Recognition Based on Illumination Transformation (Project No. 110107001).

REFERENCES

[1] A. K. Jain and A. Ross, "Multibiometric systems," *Commun. ACM*, vol. 47, pp. 34–40, January 2004.

[2] K. Nandakumar, Y. Chen, S. Dass, and A. Jain, "Likelihood ratio-based biometric score fusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 342–347, 2007.

[3] Y. Ma, B. Cukic, and H. Singh, "A classification approach to multi-biometric score fusion," in *Proc. Fifth Intl Conf. Audio Video-based Biometric Person Authentication*. Springer, 2005, pp. 484–493.

[4] J. Fierrez-Aguilar, J. Ortega-Garcia, J. Gonzalez-Rodriguez, and J. Bigun, "Discriminative multimodal biometric authentication based on quality measures," *Pattern Recognition*, vol. 38, no. 5, pp. 777–779, 2005.

[5] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern recognition*, vol. 38, no. 12, pp. 2270–2285, 2005.

[6] K. Toh, X. Jiang, and W. Yau, "Exploiting global and local decisions for multimodal biometrics verification," *Signal Processing, IEEE Transactions on*, vol. 52, no. 10, pp. 3059–3072, 2004.

[7] R. Snelick, U. Uludag, A. Mink, M. Indovina, and A. Jain, "Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 450–455, 2005.

[8] K. Toh, J. Kim, and S. Lee, "Maximizing area under roc curve for biometric scores fusion," *Pattern Recognition*, vol. 41, no. 11, pp. 3373–3392, 2008.

[9] A. Kumar, V. Kanhangad, and D. Zhang, "Multimodal biometrics management using adaptive score-level combination," in *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*. IEEE, 2008, pp. 1–4.

[10] N. Kruger, "An algorithm for the learning of weights in discrimination functions using a priori constraints," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 7, pp. 764–768, 1997.

[11] N.Poh and S.Bengio, "An investigation of f-ratio client-dependent normalisation on biometric authentication tasks," in *IEEE Int'l Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, vol. I. IEEE, 2005.

[12] M. Abramowitz and I. Stegun, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Dover publications, 1964.

[13] R. Snelick, M. Indovina, J. Yen, and A. Mink, "Multimodal biometrics: issues in design and testing," in *Proceedings of the 5th international Conference on Multimodal interfaces*. ACM, 2003, pp. 68–72.

[14] P. Phillips, W. Scruggs, A. OToole, P. Flynn, K. Bowyer, C. Schott, and M. Sharpe, "Frvt 2006 and ice 2006 large-scale results," *National Institute of Standards and Technology, NISTIR*, vol. 7408, 2007.

[15] P. P. S.A. Rizvi and H. Moon, "The feret verification testing protocol for face recognition algorithm," in *Proceedings of Intl. Conf. on Face and Gesture Recognition*, 1998, pp. 48–53.

[16] P. Phillips, H. Moon, S. Rizvi, and P. Rauss, "The feret evaluation methodology for face-recognition algorithms," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 10, pp. 1090–1104, 2000.

[17] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, and D. Zhao, "The cas-peal large-scale chinese face database and baseline evaluations," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, vol. 38, no. 1, pp. 149–161, 2008.

[18] D. Bolme, J. Ross Beveridge, M. Teixeira, and B. Draper, "The csu face identification evaluation system: its purpose, features, and structure," *Computer Vision Systems*, pp. 304–313, 2003.

[19] P. E. J.C. Hager and W. Friesen, *Facial action Coding system*. A Human Face, 2002.