Robust Biometric Score Fusion by Naive Likelihood Ratio via Receiver Operating Characteristics

Qian Tao, *Member, IEEE*, and Raymond Veldhuis

***Abstract*—This paper presents a novel method of fusing multiple biometrics on the matching score level. We estimate the likelihood ratios of the fused biometric scores, via individual receiver operating characteristics (ROC) which construct the Naive Bayes classifier. Using a limited number of operation points on the ROC, we are able to realize reliable and robust estimation of the Naive Bayes probability without explicit estimation of the genuine and impostor score distributions. Different from previous work, the method takes into consideration a particular characteristic of the matching score: its quantitative value is already an indication of the sample's likelihood of being genuine. This characteristic is integrated into the proposed method to improve the fusion performance while reducing the inherent algorithmic complexity. We demonstrate by experiments that the proposed method is reliable and robust, suitable for a wide range of matching score distributions in realistic data and public databases.**

***Index Terms*—Biometric fusion, logarithm likelihood ratio, Naive Bayes classifier, ROC.**

I. Introduction

BIOMETRIC fusion is an important way to improve the reliability of biometric systems by combining information from multiple biometric modalities such as face, speech, fingerprint, iris, etc. [7], [8], [12], [20], [26], [31]. Biometric fusion can be performed at four different levels: sensor level, feature level, matching score level, and decision level. Among all levels, fusion at the score level is the most popular, offering the best trade-off between information content and ease of fusion [26].

In the literature the fusion methods at the matching score level are categorized into three large groups [19], [26]. The first group of methods is transformation-based: all the component matching scores are firstly transformed onto a comparable scale and then simple fusion rules are applied on the transformed scores. Examples of the fusion rule are product, sum, mean, max, etc. [15]. The second group of methods are density-based: the underlying probability density function of multiple scores is firstly estimated, and then the joint probability is computed as the fused score [3], [13], [21]. The third group of methods is classifier-based: the component matching scores are concatenated as feature vectors, on which new classifiers are trained, and produces the corresponding classification score for each feature vector. Examples of classifiers are neural networks (NN) [34], support vector machines (SVM) [27], decision trees [25], [29].

By formulating score fusion as a probability or pattern recognition problem, the second and third categories of methods are very flexible. The formulation makes it possible to explore the rich literature resources in relevant domains. In doing so, however, we point out that the intrinsic characteristic of the matching score has not been taken into consideration. *A matching score differs from an arbitrary feature in that its quantitative value is already an indication of the sample's likelihood of being genuine* [26]. For example, if all component scores of input A are higher than all component scores of input B, A is by definition more likely to be genuine than B. This implies that the fused score of A should be larger than the fused score of B. Such requirement, however, is not always met by the probability- or pattern-based fusion methods. Taking probability-based fusion methods for example, assume all component classifiers produce scores ranging from zero to infinity, it is conceivable that the input A with extremely high component scores (e.g. all infinite) will actually have low fused score by probability-based fusion because the probability density at infinite always approaches zero. In this respect, utilizing the intrinsic characteristic of matching score would avoid such errors, accurately shape the problem, and lead to improved results. In this paper, we propose such a fusion method that explicitly takes into consideration the specific score characteristics. This is realized by exploiting the receiver operating characteristics (ROC), which, computed from the matching scores, is inherently associated with the score characteristics as well. Although the ROC has been commonly used to characterize the classification performance, only very few utilized it for the purpose of classification [10], [28], [30]. The limited work in literature, furthermore, mostly concentrated on optimizing the decisions instead of fusing on the score level in real number domain.

The Naive Bayes classifier is a probabilistic classifier based on the Bayes' theorem with independence assumptions on individual features [6]. Despite the assumption that can be unrealistic in many cases, the Naive Bayes classifier is surprisingly successful in practice, often competing with more sophisticated classification techniques. In fact, the optimality

Manuscript received January 10, 2012; revised November 24, 2012; accepted November 28, 2012. Date of publication December 05, 2012; date of current version January 07, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Arun Ross.

Q. Tao is with the Division of Image Processing, Department of Radiology, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands (e-mail: q.tao@lumc.nl).

R. Veldhuis is with the Signals and Systems Group, Department of Electrical Engineering, Mathematics and Computer Science, University of Twente, 7522 NB Enschede, The Netherlands (e-mail: r.n.j.veldhuis@utwente.nl).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIFS.2012.2231862

of the Naive Bayes classifier in dependent cases has been investigated from a theoretical point of view [4], [5], [35]. The Naive Bayes classifier has proven to be effective in many practical applications including text classification, medical diagnosis, etc. [1], [18], and biometric score fusion [16]. A biometric score fusion method based on standard computation of the likelihood ratio has demonstrated excellent performance [19]. In [19], the probability density functions of the genuine and the impostor scores are firstly estimated by the Gaussian Mixture Model (GMM). The algorithm integrates parameter estimation and model selection for density estimation in an integral optimization framework, producing a globally optimal choice of cluster modeling under the minimum description length (MDL) criterion. Although optimal from the pattern classification point of view, the method still regards the score as common features (as defined in pattern recognition terminology), and the algorithm is computationally expensive with the MDL optimization. In the proposed method, in comparison, the score-specific characteristics will be integrated, to improve the fusion performance while reducing the inherent algorithmic complexity.

The introduction of the unique score property contributions to the fusion performance in three senses: firstly, it improves the fusion accuracy, secondly, it increases the robustness of fusion to small sample size, and thirdly, it significantly reduces the computational complexity. In the following, Section II introduces the theory of the biometric score fusion by Naive likelihood ratio. Section III presents the implementation aspects of the method. Section IV shows the experimental results of the proposed fusion on realistic biometric score data. Section V is the conclusion.

II. Naive Likelihood Ratio Estimation via ROC: Theory

A. Naive Bayes Classifier

We consider two classes in the biometrics scenario: the genuine user class $\omega_{\rm g}$ and the impostor class $\omega_{\rm i}$. The probability density functions of the two classes are expressed as $p(\mathbf{s}|\omega_{\rm g})$ and $p(\mathbf{s}|\omega_{\rm i})$, where $\mathbf{s}$ is the concatenated score vector, with each element s_i $(i = 1, \ldots, N)$ the matching score from one of the N individual biometrics.

A Naive Bayes classifier is a simple probabilistic classifier based on the Bayes' theorem with independence assumptions on individual features [6]. The original form of Bayesian rule, in the two-class case, is expressed as

$$\begin{aligned}\omega^* &= \arg \max_{\omega \in \{\omega_{\rm g}, \omega_{\rm i}\}} p(\omega|\mathbf{s}) \\ &= \arg \max_{\omega \in \{\omega_{\rm g}, \omega_{\rm i}\}} p(\omega) p(\mathbf{s}|\omega). \end{aligned} \tag{1}$$

where ω^* is the class estimation.

Suppose $\mathbf{s}$ is the input score vector composed of N independent scores, $\mathbf{s} = [s_1, \ldots, s_N]$, then (1) can be rewritten under the independency assumption

$$\omega^* = \arg \max_{\omega \in \{\omega_{\rm g}, \omega_{\rm i}\}} p(\omega \prod_{i=1}^{N} p(s_i|\omega), \tag{2}$$

which is the Naive Bayes classifier.

B. Naive Likelihood Ratio

The likelihood ratio is a statistic defined as

$$\lambda(\mathbf{s}) = \frac{p(\mathbf{s}|\omega_{\rm g})}{p(\mathbf{s}|\omega_{\rm i})}. \tag{3}$$

The likelihood ratio test is defined as follows

$$\begin{aligned} &\text{if} \quad \lambda(\mathbf{s}) \geq t, \qquad \omega^* = \omega_{\rm g}, \\ &\text{if} \quad \lambda(\mathbf{s}) < t, \qquad \omega^* = \omega_{\rm i}, \end{aligned} \tag{4}$$

where t is the application-specific threshold of the likelihood ratio.

The likelihood ratio test is optimal in the Neyman–Pearson sense: at a given false acceptance rate, the obtained false rejection rate reaches minimal, and likewise, at a give false rejection rate, the obtained false acceptance rate reaches minimal [33].

If the same independency assumption between features is introduced as in (2), (3) is rewritten

$$\begin{aligned}\lambda_{\rm Naive}(\mathbf{s}) &= \frac{\prod_{i=1}^{N} p(s_i|\omega_{\rm g})}{\prod_{i=1}^{N} p(s_i|\omega_{\rm i})} \\ &= \prod_{i=1}^{N} \lambda(s_i), \end{aligned} \tag{5}$$

where we call $\lambda_{\rm Naive}(\mathbf{s})$ the *Naive likelihood ratio*. For ease of computation, the logarithm of the likelihood ratio is taken, $\rho = \log(\lambda)$, and we have

$$\rho_{\rm Naive}(\mathbf{s}) = \sum_{i=1}^{N} \rho(s_i). \tag{6}$$

It follows by comparing (1) and (4) that in the two-class problem, the likelihood ratio test is equivalent to the Bayes rule of (1), if and only if the threshold in (4) is set as $t = p(\omega_{\rm i})/p(\omega_{\rm g})$. In fact, this represents a single operation point of the likelihood ratio test. In the likelihood ratio test, by varying t, the entire ROC curve of the likelihood ratio test can be acquired, describing all situations with different combination of class priors and misclassification costs.

C. ROC and Likelihood Ratio

The ROC is widely-used to characterize the classification performance. It describes the trade-off between the detection rate (or correct acceptance rate) $p_{\rm d}$ and the false acceptance rate α, with the change of the threshold on the matching score. The threshold t could be any scalar value, $t \in [-\infty, \infty]$, covering the entire matching score space.

Theoretically, $p_{\rm d}$ and α can be computed from the distributions of the matching scores in the genuine and impostor classes, $p(s|\omega_{\rm g})$ and $p(s|\omega_{\rm i})$

$$p_{\rm d}(t) = \int_t^{\infty} p(s|\omega_{\rm g}) {\rm d}s, \tag{7}$$

$$\alpha(t) = \int_t^{\infty} p(s|\omega_{\rm i}) {\rm d}s, \tag{8}$$

where t is the operating threshold.

Taking the derivative (slope) of the ROC, and referring to (8) and (7), we have

$$\frac{\mathrm{d}p_\mathrm{d}}{\mathrm{d}\alpha} = \frac{\frac{\mathrm{d}p_\mathrm{d}}{\mathrm{d}t}}{\frac{\mathrm{d}\alpha}{\mathrm{d}t}} = \frac{p(t|\omega_\mathrm{g})}{p(t|\omega_\mathrm{i})} = \lambda(t), \tag{9}$$

which is, by definition, the likelihood ratio of the matching score. This means that by taking the slope of the ROC curve, a direct mapping from the score t to its corresponding likelihood ratio value is realized without the need to estimate the two probability density functions $p(t|\omega_\mathrm{g})$ and $p(t|\omega_\mathrm{i})$.

D. Biometric Fusion by Naive Likelihood Ratio

Suppose we have N sets of biometric scores, denoted by s_i, $i = 1, \ldots, N$, coming from different biometric modalities or different classifiers. We concatenate the individual matching scores to construct a new feature vector $\mathbf{s} = [s_1, \ldots, s_N]$, and classify the feature vector by a Naive Bayes classifier. The Naive likelihood ratio is a summation of individual log-likelihood ratios, which can be derived from the slope of the component ROCs as introduced in Section II-C

$$\lambda_\mathrm{Naive}(\mathbf{s}) = \prod_{i=1}^{N} \frac{\mathrm{d}_{\mathrm{d},i}}{\mathrm{d}\alpha_i},$$
$$\rho_\mathrm{Naive}(\mathbf{s}) = \sum_{i=1}^{N} \log\left(\frac{\mathrm{d}p_{\mathrm{d},i}}{\mathrm{d}\alpha_i}\right) = \sum_{i=1}^{N} \rho(s_i), \tag{10}$$

where $p_{\mathrm{d},i}$ α_i are the operation points on the ith component ROC.

With s being a quantative measure of the underlying biometric being genuine, the derived new score $\rho(s)$ should keep the same property. This implies that $\rho(s)$ needs to be a monotonic function of s, and the fusion process should integrate particular steps to ensure this property.

III. Naive Likelihood Ratio Estimation via ROC: Implementation

A. ROC Estimation and Processing

In practice, α and p_d in (10) can only be estimated in discrete manner from a number of given matching scores:

$$\alpha(t) = \frac{\| \{s|s > t, s \in \omega_\mathrm{i}\} \|}{\| \{s|s \in \omega_\mathrm{i}\} \|},$$
$$p_\mathrm{d}(t) = \frac{\| \{s|s > t, s \in \omega_\mathrm{g}\} \|}{\| \{s|s \in \omega_\mathrm{g}\} \|}, \tag{11}$$

where t denotes the varying score, and $\| \cdot \|$ denotes the length of the set.

The accuracy of the estimated operation points largely relies on the number of training samples. Fig. 1 shows how sample size influences the estimation. In the two examples, the genuine and impostor score distributions are both modeled as Gaussian, by $N(1, 1)$ and $N(-1, 2)$, respectively. The two distributions overlap with each other such that the ROC, $p_\mathrm{d}(\alpha)$, is not trivial. In one example, large number of samples of 100,000 per class are generated, while in the other, only 200 samples per class are generated. The theoretical likelihood ratio can be obtained as

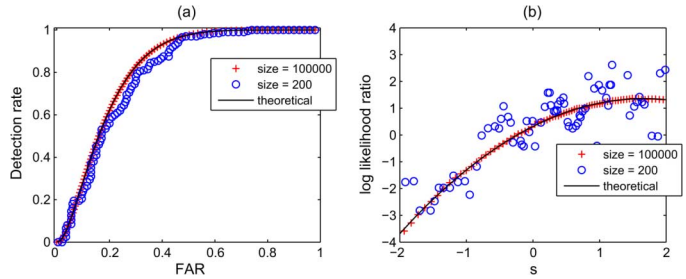


Fig. 1. (a) Estimated ROC; (b) estimated log likelihood ratio versus matching score. The large sample case (10,000), small sample case (200), and theoretical reference are all plotted.

$\log(p(s|\omega_\mathrm{g})/p(s|\omega_\mathrm{i})) = \log(2) - ((x-1)^2/2) + ((x+1)^2/8)$. Fig. 1(a) shows the estimated ROC of the two cases computed by (11). Theoretical values of the ROC and log likelihood ratio are also plotted for reference. It can be observed that for the smaller sample size, the estimated ROC operation points and log likelihood ratio become less accurate.

One practical parameter is the number of operation points on the ROC. Generally speaking, too densely sampled operation points tends to produce a noisy ROC due to the limited number of score samples, whereas too sparsely distributed operation points may under-represent the true ROC. In our work, we have chosen 100 as an empirically good number, with evenly distributed percentiles (1%, ..., 100%) in the score set. Consequently, the input scores of arbitrary size (which can be orders higher than 100) are transformed into a fixed number (100) of operation points.

The accuracy of ROC approximation is important for the fusion performance. In large sample case, the ROC can be reliably computed given the accuracy of estimated $\alpha(t)$ and $p_\mathrm{d}(t)$. In small sample case, some processing on the operation points is necessary and beneficial. The ROC curve $p_\mathrm{d}(\alpha)$ can be effectively refined by simultaneously smoothing $\alpha(t)$ and $p_\mathrm{d}(t)$. This is equivalent to smoothing the probability density estimation with simple linear filters, since both $\alpha(t)$ and $p_\mathrm{d}(t)$ are smooth and monotonic curves given their accumulative nature. In our work we used an 8th order Chebyshev lowpass filter, with the cutoff frequency 1/4 of the sampling rate. Imposing the monotonicity of score mapping, furthermore, is equivalent to ensuring monotonicity of the ROC slope, which can be realized by taking the convex hull of the operation points on the resulting ROC. The properties of $\alpha(t)$, $p_\mathrm{d}(t)$, and $p_\mathrm{d}(\alpha)$, therefore, are effectively incorporated into the fusion process.

Consequently, the computation of the likelihood ratios can be accomplished by simple manipulations on a small number of operation points, provided that the operation points are accurately approximated. Fig. 2 shows the results of the ROC processing on data of Fig. 1. More accurate estimation of the ROC, as well as the ROC slope, can be observed in Fig. 2(c), (d).

B. Parameterization and Score Field Visualization

The proposed method transforms the complex joint probability density estimation problem into multiple one-dimensional fitting of simple monotonic $\rho(s)$ functions. Given a number of pairs $[s_i, \rho_i], i = 1, 2, \ldots, N$, parameterization of the $\rho(s)$ function enables score mapping in continuous domain. In realistic

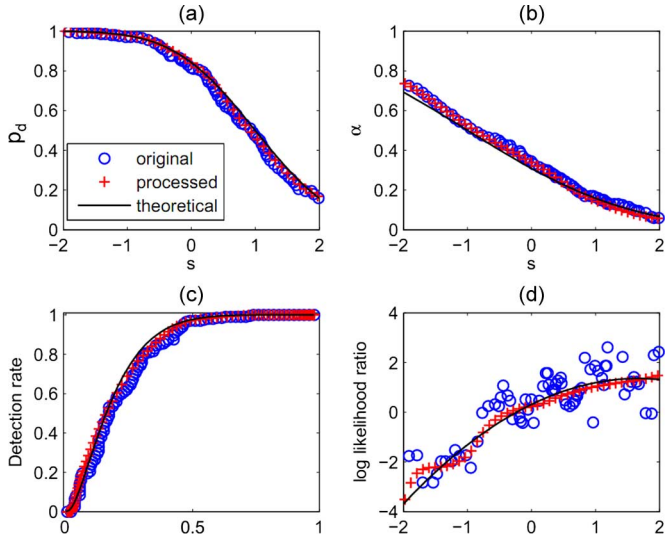


Fig. 2. (a) p_{d} varying with t; (b) α varying with t; (c) ROC and the processed ROC; (d) the original and processed log likelihood ratio estimation, compared to the theoretical value.

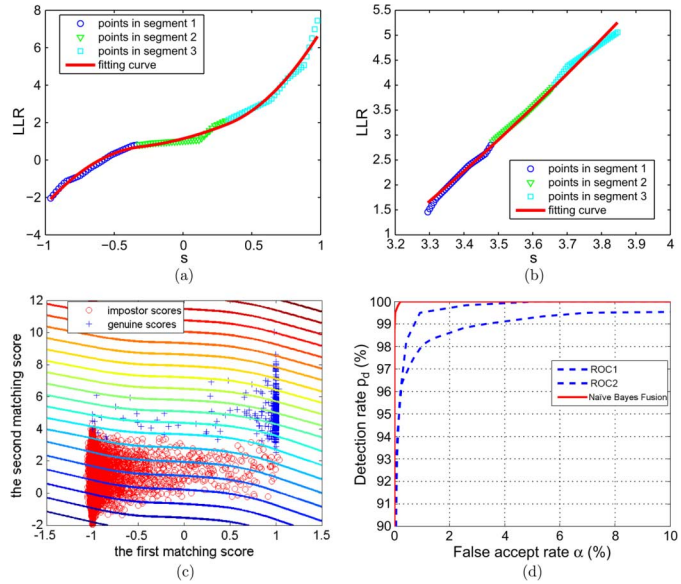


Fig. 3. (a) Fitting s_1 using piecewise polynomial. (b) Fitting s_2 using piecewise polynomial. (c) Score distribution and the decision boundaries. (d) ROCs of the individual scores and of the Naive likelihood ratio fused scores.

situations, there are a number of frequently encountered distributions, which consist exponential terms, like Gaussian, Exponential, Rayleigh, Gamma, etc., and as such the log likelihood ratio can indeed assume a simple form. In our work, we fit $\rho(s)$ with polynomial functions under the minimum square error (MSE) criterion.

To better follow the curvature of $\rho(s)$ and meanwhile maintain robustness, piecewise polynomial fitting can be applied. We have observed that most often 1–3 polynomial functions suffice to fit $\rho(s)$. The method fits $\rho(s)$ with piecewise smooth low-order polynomials, where "smooth" implies continuity of the value as well as derivatives. In this way the local smoothness and global flexibility of the curve are satisfied simultaneously. The problem can be formulated as a standard quadratic programming problem. Mathematics of piecewise polynomial fitting can be found in the Appendix A. An example of piecewise polynomial fitting will be given in Fig. 3.

Non-parametric approximation of $\rho(s)$ is also possible by interpolation between a number of reliably estimated discrete points $\rho(s_i)$. This includes among others the linear or spline interpolation. However, in most cases the parametric fitting suffices to achieve accurate and robust results.

We further visualize the decision boundary resulting from the proposed method. By calculating the likelihood ratio values in the score space using (9) and (10), the likelihood ratio field can be obtained, from which the boundaries can be derived as the iso-value contours of this field. The decision boundaries are illustrative for the properties of fusion methods such as monotonicity and smoothness. Suppose we have two mapping functions $\rho_1(s_1)$ and $\rho_2(s_2)$, then the contours can be written in the mathematical form

$$\rho(s_1, s_2) = \rho(s_1) + \rho(s_2) = c \qquad c = \{c_1, \ldots, c_K\}, \quad (12)$$

where c is a set of K constants corresponding to different performance requirements (i.e. FAR or FRR). The visualization helps illustrate how the processing in the fusion method influences the final decision boundaries in the score space. For example, the shape of the decision boundaries is determined by the parametric forms of $\rho_1(s_1)$ and $\rho_2(s_2)$. As a consequence, the monotonicity, smoothness, and robustness of the $\rho(s)$ are inherited by the decision boundaries. As an example, we show in Fig. 3 the decision boundaries of piecewise polynomial fitting, in which 3 equidistance pieces are taken and polynomials with order up to 2 are used.

IV. Experiments and Discussion

A. Databases

The first database contains the matching score of the two-dimensional and three-dimensional face-recognition data from the Face Recognition Grand Challenge face database [23]. The scores were derived from a diversity of classifiers, including the hierarchical graph matching (HGM), Gabor filters, principle component analysis (PCA), linear discriminant analysis (LDA). For details, please refer to [30]. The database contains data of 465 subjects and has in total 4007 samples. The classifiers that produce the matching scores are trained on 309 subjects in the database. The large sample set in this step ensures that the classifiers are well-trained, and that the classification and fusion altogether lead to a high-performance biometric system. To train fusion, another 100 subjects are taken to obtain the matching scores from the trained classifier, resulting in 25,520 genuine scores and 2,568,190 impostor scores. The remaining 56 subjects are used for evaluation, resulting in 12,270 genuine scores and 700,910 impostor scores.

The second database is the publicly available Biometric Authentication Fusion Benchmark Database (BA-fusion) [24], developed from the XM2VTS database [17], which contains the matching scores from synchronized face video and speech data of 295 subjects. The matching scores are derived from various

baseline systems, with different feature and classifier combinations (5 for face and 3 for speech), as described in detail in [24]). The fusion is trained on the development set, and evaluated on the independent evaluation set, both provided by the database. The databases consist of scores computed from a variety of feature extraction and pattern classification methods (for details see [24]), covering a large range of realistic score distributions.

To evaluate the proposed method, we also implemented four other well-established fusion methods for comparison:

(i) Sum Rule: This method firstly transforms the individual scores by Z-normalization [15], which normalizes the genuine scores to unit variance, and then sum them up.

(ii) SVM: This method maximizes the (soft) margin between classes [2]. The flexible radius basis function (RBF) kernel was used. The tuning of the SVM parameters are done by cross validation and grid search as described in [11]. We used the SVM-light package [14] to realize fast core implementation of the SVM.

(iii) LLR-GMM: This method is theoretically optimal from a classification point of view. The joint probability density functions of the genuine and impostor scores are estimated using the GMM model. The implementation integrates parameter estimation and model selection in a single algorithm [9], producing a globally optimal choice of cluster modeling under the MDL criterion. Then the likelihood ratio is calculated from the estimated genuine and impostor score distributions.

(iv) Naive-LLR-GMM: This method computes the Naive likelihood ratio as the proposed method, but instead of estimating the ratio directly, it first estimates the genuine and impostor distributions and then computes their ratio. One-dimensional probability density function estimation using GMM model is performed for each component biometric score. The same method as in (iii) is used.

B. Results

All above score-level fusion methods eventually produce the fused scores, with which the ROC curve can be derived to give an overall evaluation of the fusion performance. Visualizing the resulting fused score (Naive LLR) field help demonstrate the different fusion mechanisms. To evaluate the robustness of the fusion method to sample size, the same Naive likelihood ratio fusion trained on a small subset of the score data, can be drawn for comparison. The first database containing large sample set was used for this purpose, and in the subset case, only 1% of the original data is taken.

Fig. 4 shows the results from the five different methods, in cases of large and small sample sizes. With the same underlying distribution but different sample size, the influence of sample size on the final estimated Naive LLR field can be observed. The monotonicity in the proposed method is clearly shown, while more flexible boundaries of the other two LLR methods can also be observed.

Fig. 5 compares the ROC curves of the five different methods on the independent evaluation data. The ROC curves showed the overall performance of the five fusion methods. The three methods that work with LLR, namely, LLR-GMM, Naive-LLR-GMM, Naive-LLR-ROC, exhibited high performances. In the small sample case (Fig. 5(b)), the proposed method yielded slightly better performance. The mean and standard deviation of the Equal Error Rate (EER) were also computed and reported in Fig. 6. To reduce the statistical variance of the subset sampling, 50 random subsets were taken and the results were averaged over all subsets to provide a reliable performance estimation. Paired T-test showed that the

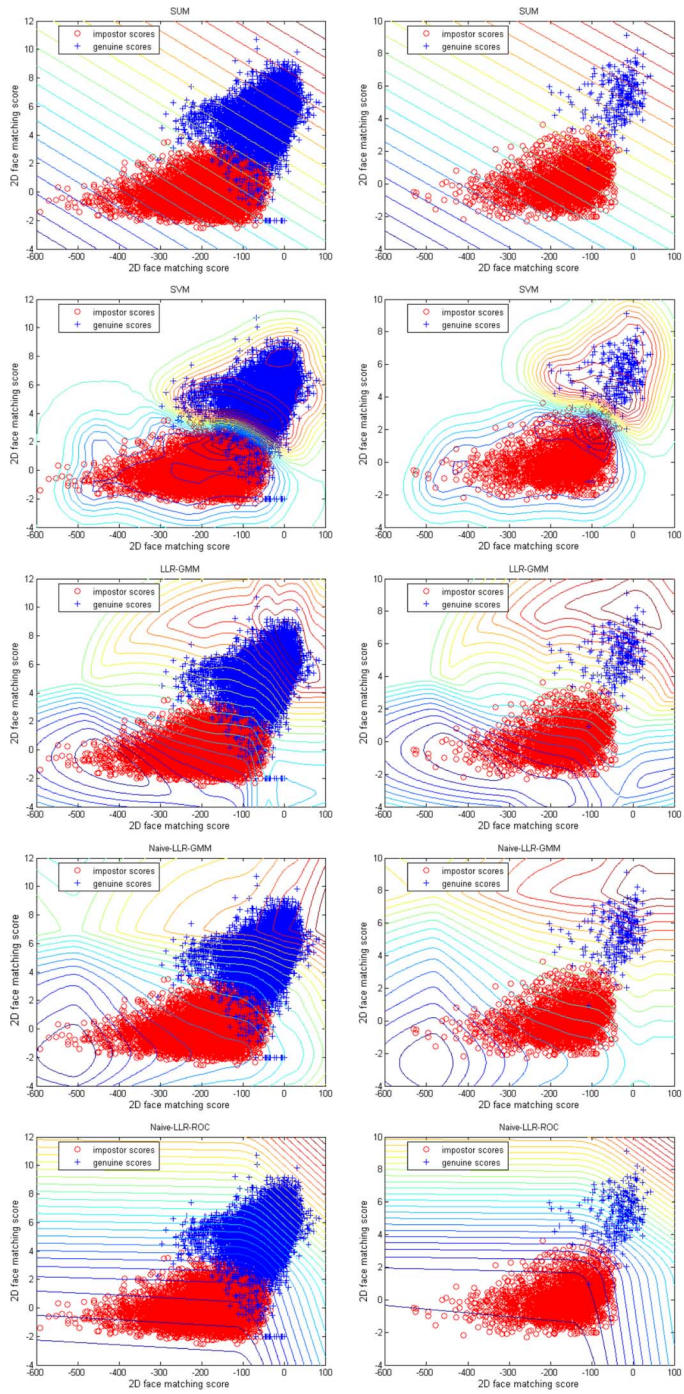


Fig. 4. Visualization of the final score field of different fusion methods on the FRGC database [32]. From top down: sum rule, SVM, LLR-GMM, Naive-LLR-GMM, Naive-LLR-ROC. The left column shows the results on the full dataset, while the right shows the results on the 1% data subset. The field is visualized by the iso-value contours, with red indicating high values and blue low.

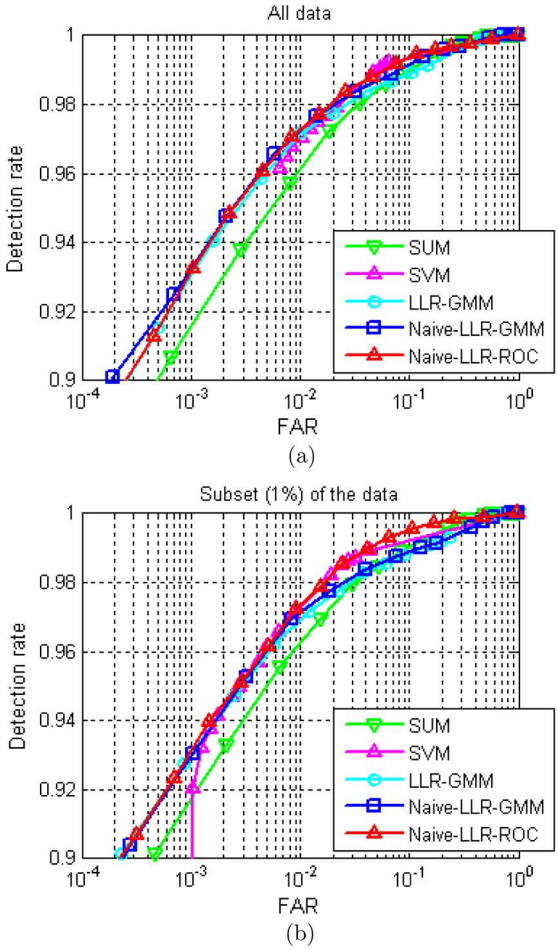


Fig. 5. Comparison of the fusion results on the FRGC database [32] in terms of ROC. The fusion method is trained on the development set and tested on the evaluation set. (a) Results on the complete dataset; (b) results on the 1% data subset.

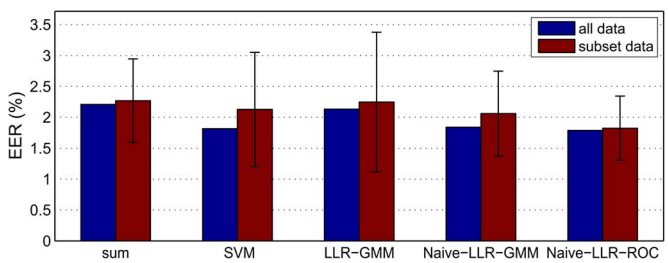


Fig. 6. Comparison of the fusion results on the FRGC database [32] in terms of EER. The results from both the entire dataset and 1% dataset are shown.

proposed method yielded significantly lower EER than other methods when performing fusion on the 1% subset ($p < 0.05$).

We have further carried out a larger-scale validation on the second database, which contains a wide variety of score distributions. Two types of evaluation, namely *intermodality* and *intramodality* fusion, was performed. The intermodality fusion work on scores coming from two distinct biometric modalities: face and speech, which are in principle independent of each other. In contrast, the intramodality fusion works on two scores coming from the same biometric, but computed by different classifiers. Obviously the component scores are dependent. The performance of Naive classifier in nonoptimal situations can therefore be examined in the intramodality scenario.

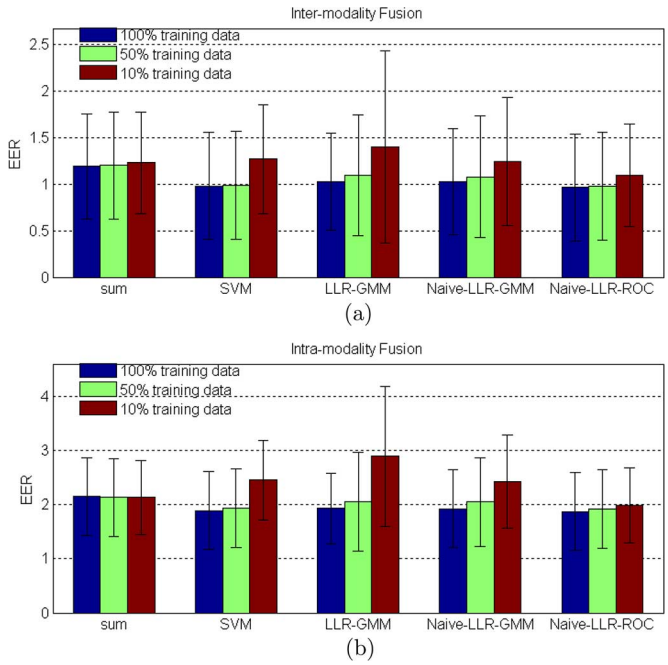


Fig. 7. Bar plot of EER comparison for the BA-fusion database [24]: (a) intermodality fusion; (b) intramodality fusion. EER is plotted for three different sample sizes: 100%, 50%, and 10%.

With 5 scores for face modality and 3 scores for speech modality, there are in total 15 combinations for intermodality fusion and 10 combinations for intramodality fusion of the face modality. We further evaluated the robustness of the five methods in small sample size case, by comparing the results on fusing scores of different sample sizes, namely, 100%, 50%, and 10% from the original set. In Fig. 7 the mean and standard deviation of EER is displayed for the five fusion methods, at different sample sizes.

It can be observed from the figures that the sum rule, despite its high robustness, exhibited in general poorer performance compared to the advanced methods. The two methods that are based on LRR and separate estimation of the genuine and impostor probability density functions (LLR-GMM-Naive and LLR-GMM) showed comparable performance in both intermodality and intramodality fusion. This indicates that the Naive assumption can be indeed suitable for the biometric fusion scenario. Both methods showed a trend of degradation in small sample size case, and the same trend is true for the SVM fusion method using RBF kernels. This can be explained by the high algorithmic complexity, which demands a large number of the training samples. Overall the proposed method shows a good compromise between accuracy and robustness, as can be observed from Fig. 7. Using the paired T-test, the EER values from different methods did not show significant difference when applied on the full dataset, but with the 10% subset, the proposed method yielded significant lower EER than the SVM and GMM method ($p < 0.05$).

The computational load of different fusion methods differ significantly. In the current computer of 2.16 GHz, 2 GB RAM, using Matlab for computation, the fusion experiments

on the BA-fusion database take 0.33 sec for the sum rule fusion, 836.45 sec for the SVM fusion with tuning of parameters, 43.26 sec for the LLR-GMM, 12.24 sec for the LLR-GMM-Naive, and 0.72 sec for the Naive likelihood ratio fusion. For the larger score database from the FRGC database, the difference between computation time is more substantial. The fusion takes 0.32 sec for the sum rule fusion, 9239.32 sec for the SVM fusion with tuning of parameters, 543.16 sec for the LLR-GMM, 122.39 sec for the LLR-GMM-Naive, and 0.81 sec for the Naive likelihood ratio fusion.

The proposed method is uniformly light in computation for any input sample set. Any number of input scores are firstly converted into a limited number (100) of operation points, for all the rest computation. For SVM or GMM, in contrast, the computation is still based on the original data, whose size can be orders higher, and involves more complicated steps like expectation maximization (EM) in the GMM case and quadratic programming (QP) in the SVM case. From practical concerns, the proposed method is favorable in the sense it can achieve competing performance with much more sophisticated methods, at substantially lower hardware cost.

V. Conclusion

This paper presents a new method of fusing multiple biometrics, using the likelihood ratio under the Naive Bayes assumption. The method achieves the likelihood ratio estimation via a limited number of operation points on the ROC, avoiding explicit estimation of the genuine and impostor score distributions. We have extensively validated the proposed method on the FRGC 2D-3D face database and the public BA-fusion database. The method demonstrated competitive fusion performance when compared to more sophisticated fusion methods like SVM and GMM, with much lower computation complexity. In addition, the method exhibited high robustness to small training sample size, demonstrating higher fusion performance compared to the SVM and GMM methods when only a small subset is taken.

Appendix

In the following, we will provide mathematical solutions to the piecewise polynomial fitting problem. Polynomials of order up to 2 will be smoothly connected. Higher order piecewise-polynomials are also possible with similar mathematics derivation, but not recommended in this case for the risk of overfitting.

Assume the matching scores range from $s_{\min}$ to $s_{\max}$, and three pieces are taken from this range: $[s_{\min}, t_1]$, $[t_1, t_2]$, $[t_2, s_{\max}]$, where $s_{\min} < t_1 < t_2 < s_{\max}$. Let the three polynomial function be $F_1(s)$, $F_2(s)$, and $F_3(s)$, then the fitting error is to be minimized

$$\mathcal{F} = \sum_{s_{\min} \le s_i \le t_1} (F_1(s_i) - l_i)^2 + \sum_{t_1 < s_j \le t_2} (F_2(s_j) - l_j)^2 + \sum_{t_2 < s_k \le s_{\min}} (F_3(s_k) - l_k)^2 \tag{13}$$

As defined in Section III-B, smooth piecewise polynomial fitting means continuity of the value as well as derivatives. Then the following equations should be satisfied at the conjoining points t_1 and t_2

$$F_1(t_1) = F_2(t_1), \qquad F_2(t_2) = F_3(t_2) \tag{14}$$

$$F_1'(t_1) = F_2'(t_1), \qquad F_2'(t_2) = F_3'(t_2) \tag{15}$$

This is a second-order optimization problem with constraints, and can be formulated into a standard quadratic programming problem with respect to the polynomial coefficients of the three piecewise functions. A standard quadratic programming problem is written as follows [22]

$$x = \arg\min\left(\frac{1}{2}x^T H x + fx\right), \quad \text{subject to}: A_1 x \le b_1, \quad \text{or} \quad A_0 x = b_0 \tag{16}$$

The unknowns are the 9 polynomial coefficients of the 3 functions F_1, F_2, and F_3. Let us put them into one coefficient vector $\mathbf{x} = [c_{12}, c_{11}, c_{10}, c_{22}, c_{21}, c_{20}, c_{32}, c_{31}, c_{30}]^T$, where for each c_{ij}, the first subscript i denotes the function index, and the second subscript j denotes the coefficient order. It is easy to transform the 9-coefficient vector back to the 3-coefficient vector of the three individual functions via a 3×9 matrix

$$\begin{aligned} [c_{12},\ c_{11},\ c_{10}]^T &= \mathbf{P}_1\mathbf{x}, \\ [c_{22},\ c_{21},\ c_{20}]^T &= \mathbf{P}_2\mathbf{x}, \\ [c_{12},\ c_{11},\ c_{10}]^T &= \mathbf{P}_3\mathbf{x} \end{aligned} \tag{17}$$

where

$$\mathbf{P}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\mathbf{P}_2 = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

$$\mathbf{P}_3 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{18}$$

To deal with the three piecewise polynomials, we partition the input matching scores into three ranges $[s_{\min}, t_1]$, $[t_1, t_2]$, $[t_2, s_{\max}]$, and assume the scores in the first range be $\{s_1, \ldots, s_{N_1}\}$, in the second range be $\{s_{N_1+1}, \ldots, s_{N_1+N_2}\}$, in the third range be $\{s_{N_1+N_2+1}, \ldots, s_{N_1+N_2+N_3}\}$, where $N_1 + N_2 + N_3 = N$. Define three matrices as follows

$$\mathbf{S}_1 = \begin{pmatrix} s_1^2 & s_1 & 1 \\ \vdots & \vdots & \vdots \\ s_{N_1}^2 & s_{N_1} & \end{pmatrix}$$

$$\mathbf{S}_2 = \begin{pmatrix} s_{N_1+1}^2 & s_{N_1+1} & 1 \\ \vdots & \vdots & \vdots \\ s_{N_1+N_2}^2 & s_{N_1+N_2} & 1 \end{pmatrix}$$

$$\mathbf{S}_3 = \begin{pmatrix} s_{N_1+N_2+1}^2 & s_{N_1+N_2+1} & 1 \\ \vdots & \vdots & \vdots \\ s_N^2 & s_N & 1 \end{pmatrix}$$

Likewise, define three vector of the LLR values as follows

$$\mathbf{l}_1 = \begin{pmatrix} l_1 \\ \vdots \\ l_{N_1} \end{pmatrix}$$
$$\mathbf{l}_2 = \begin{pmatrix} l_{N_1} \\ \vdots \\ l_{N_1+N_2} \end{pmatrix}$$
$$\mathbf{l}_3 = \begin{pmatrix} l_{N_1+N_2+1} \\ \vdots \\ l_N \end{pmatrix}$$

The function to minimize, (13), therefore, can be rewritten as

$$\|\mathbf{S}_1\mathbf{P}_1\mathbf{x} - \mathbf{l}_1\|^2 + \|\mathbf{S}_2\mathbf{P}_2\mathbf{x} - \mathbf{l}_2\|^2 + \|\mathbf{S}_3\mathbf{P}_3\mathbf{x} - \mathbf{l}_3\|^2$$

Let $\mathbf{Q}_1 = \mathbf{S}_1\mathbf{P}_1$, $\mathbf{Q}_2 = \mathbf{S}_2\mathbf{T}_2$, $\mathbf{Q}_3 = \mathbf{S}_3\mathbf{T}_3$. Extend this function, and refer to (16), we have

$$H = \mathbf{Q}_1^T\mathbf{Q}_1 + \mathbf{Q}_2^T\mathbf{Q}_2 + \mathbf{Q}_3^T\mathbf{Q}_3 \tag{19}$$
$$f = -\mathbf{Q}_1^T\mathbf{l}_1 - \mathbf{Q}_2^T\mathbf{l}_2 - \mathbf{Q}_3^T\mathbf{l}_3 \tag{20}$$

The smoothness at the conjoining points can be formulated as the equality constraints. Put the two conjoining points t_1 and t_2 into two vectors $\mathbf{t}_1 = [t_1^2\ t_1\ 1]^T$ $\mathbf{t}_2 = [t_2^2\ t_2\ 1]^T$, then (14) can be written as

$$\mathbf{t}_1^T\mathbf{P_1}\mathbf{x} = \mathbf{t}_1^T\mathbf{P_2}\mathbf{x}, \qquad \mathbf{t}_2^T\mathbf{P}_2\mathbf{x} = \mathbf{t}_2^T\mathbf{P}_3$$

The first order derivative can be obtained via the following matrix

$$\mathbf{D} = \begin{pmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

Using D, (15) is then written as

$$\mathbf{t}_1^T\mathbf{D}\mathbf{P_1}\mathbf{x} = \mathbf{t}_1^T\mathbf{D}\mathbf{P}_2\mathbf{x}, \qquad \mathbf{t}_2^T\mathbf{D}\mathbf{P}_2\mathbf{x} = \mathbf{t}_2^T\mathbf{D}\mathbf{P_3}\mathbf{x}$$

Referring to (16), we have

$$A_0 = \begin{pmatrix} \mathbf{t}_1^T\mathbf{P_1} - \mathbf{t}_1^T\mathbf{P_2} \\ \mathbf{t}_2^T\mathbf{P}_3 - \mathbf{t}_2^T\mathbf{P_3} \\ {}_1^T\mathbf{D}\mathbf{P}_1 - \mathbf{t}_1^T\mathbf{D}\mathbf{P_2} \\ \mathbf{t}_2^T\mathbf{D}\mathbf{P}_2 - \mathbf{t}_2^T\mathbf{D}\mathbf{P}_3 \end{pmatrix} \tag{21}$$

$$b_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{22}$$

With A_0 and b_0 established, we have completely converted the smooth piecewise polynomial fitting problem into a standard quadratic programming problem as in (16), with H, f, A_0, b_0 calculated by (19), (20), (21), (22). The problem is then solved using the standard quadratic programming methods. Further reference of the quadratic programming problem can be found in [22].

REFERENCES

[1] R. Abraham, J. Simha, and S Iyengar, "Medical datamining with a new algorithm for feature selection and naive bayesian classifier," in *Proc. 10th Int. Conf. Information Technology (ICIT '07)*, Washington, DC, 2007, pp. 44–49, IEEE Computer Society.

[2] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[3] S. Dass, K. Nandakumar, and A. Jain, "A principled approach to score level fusion in multimodal biometric systems," in *Proc. Audio- and Video-Based Biometric Person Authentication*, 2005, pp. 1049–1058.

[4] P. Domingos and M. Pazzani, "On the optimality of the Simple Bayesian Classifier under zero-one loss," *Mach. Learning*, vol. 29, no. 2–3, pp. 103–130, 1997.

[5] P. Domingos and M. Pazzani, "On the optimality of the simple bayesian classifier under zero-one loss," *Machine Learning*, vol. 29, no. 2–3, pp. 103–130, 1997.

[6] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.

[7] J. F. Aguilar, L. Nanni, J. O. Garcia, R. Cappelli, and D. Maltoni, Combining Multiple Matchers for Fingerprint Verification: A Case Study in fvc2004, 2005, pp. 1035–1042.

[8] J. Fierrez-aguilar, Y. Chen, J. Ortega-garcia, and A. K. Jain, "Incorporating image quality in multi-algorithm fingerprint verification," in *Proc. IAPR Int. Conf. Biometrics (ICB)*, 2006, pp. 213–220, Springer LNCS-3832.

[9] M. Figueiredo and A. Jain, "Unsupervised learning of finite mixture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 3, pp. 381–396, Mar. 2002.

[10] S. Haker, W. Wells, S. Warfield, I. Talos, J. Bhagwat, D. Goldberg-Zimring, A. Mian, L. Ohno-Machado, and K. Zou, "Combining classifiers using their receiver operating characteristics and maximum likelihood estimation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, pp. 506–514, 10 2005.

[11] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, A Practical Guide to Support Vector Classification, Department of Computer Science, National Taiwan University, Tech. Rep., 2003.

[12] A. Jain, R. Bolle, and S. Pankanti, *Biometrics, Personal Identification in Networked Society*. Norwell, MA: Kluwer, 1998.

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

[14] T. Joachims, *Making Large-Scale Support Vector Machine Learning Practical*. Cambridge, MA: MIT Press, 1999.

[15] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 226–239, Mar. 1998.

[16] D. Maurer and P. Baker, "Fusing multimodal biometrics with quality estimates via a bayesian belief network," *Pattern Recognit.*, vol. 41, no. 5, pp. 821–832, 2008.

[17] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, "XM2VTsbd: The extended m2vts database," in *Proc. 2nd Conf. Audio and Video-base Biometric Person Authentication*, 1999, pp. 72–77.

[18] T. Mitchell, *Machine Learning*. New York: McGraw-Hill, 1997.

[19] K. Nandakumar, Y. Chen, S. Dass, and A. Jain, "Likelihood ratio-based biometric score fusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 342–347, Feb. 2008.

[20] L. Nanni and A. Lumini, "Ensemble of parzen window classifiers for on-line signature verification," *Neurocomputing*, vol. 68, pp. 217–224, 2005.

[21] L. Nanni, A. Lumini, and S. Brahnam, "Likelihood ratio based features for a trained biometric score fusion," *Expert Syst. With Applicat.*, vol. 38, no. 1, pp. 58–63, 2011.

[22] J. Nocedal and S. Wright, *Numerical Optimization*. New York: Springer-Verlag, 2006.

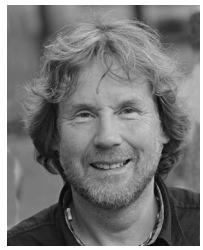
[23] P. Phillips, P. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *Proc. Computer Vision and Pattern Recognition*, 2005, pp. 947–954.

[24] N. Poh and S. Bengio, "Database, protocols and tools for evaluating score-level fusion algorithms in biometric authentication," *Pattern Recognit.*, vol. 39, no. 2, pp. 223–233, 2006.

[25] A. Ross, K. Nandakumar, and A. Jain, *Handbook of Multibiometrics*, ser. International Series on Biometrics. Secaucus, NJ: Springer-Verlag, 2006.
[26] A. Ross, K. Nandakumar, and A. Jain, *Handbook of Multibiometrics (International Series on Biometrics)*. Secaucus, NJ: Springer-Verlag, 2006.
[27] C. Sanderson and K. Paliwal, Information Fusion and Person Verification Using Speech and Face Information IDIAP, Switzerland, Tech. Rep., Sep. 2002.
[28] Q. Tao and R. Veldhuis, "Optimal decision fusion for a face verification system," in *Proc. The 2nd Int. Conf. Biometrics*, Seoul, Korea, 2007, pp. 958–967.
[29] Q. Tao and R. Veldhuis, "Hybrid fusion for biometrics: Combining score-level and decision-level fusion," in *Proc. Workshop on Biometrics, IEEE Conferene on Computer Vision and Pattern Recognition*, Alaska, 2008.
[30] Q. Tao and R. Veldhuis, "Threshold-optimized decision-level fusion and its application to biometrics," *Pattern Recognit.*, vol. 42, no. 5, pp. 823–836, 2009.
[31] B. Ulery, A. Hicklin, C. Watson, W. Fellner, and P. Hallinan, Studies of Biometric Fusion NIST Tech. Rep. IR 7346, 2006.
[32] FRGC Face Database [Online]. Available: http://face.nist.gov/frgc/
[33] H. Van Trees, *Detection, Estimation, and Modulation Theory*. New York: Wiley, 1969.
[34] Y. Wang, T. Tan, and A. K. Jain, "Combining face and iris biometrics for identity verification," in *Proc. 4th Int. Conf. AVBPA*, 2003, pp. 805–813.
[35] H. Zhang, "The optimality of Naive Bayes," in *Proc. FLAIRS Conf.*, 2004, pp. 562–567.

Qian Tao (M'11) received the B.Sc. degree in electrical engineering from Fudan University, Shanghai, China, in 2001. She received the M.Sc. degree in biomedical engineering from Fudan University in 2004, with her master thesis on Doppler ultrasound blood flow signal simulation and characterization. She received the Ph.D. degree from University of Twente, the Netherlands, in 2008, with a thesis titled "Face Verification for Mobile Personal Devices."

Currently she is a scientific researcher working at the Division of Image Processing, Department of Radiology, Leiden University Medical Center, The Netherlands. Her research interest includes pattern classification with applications to biomedical data, biomedical image, and signal processing.



Raymond Veldhuis received the M.Eng. degree in electrical engineering from the University of Twente, Enschede, The Netherlands, in 1981, and the Ph.D. degree from Nijmegen University, Nijmegen, The Netherlands, in 1988, with a thesis titled, "Adaptive restoration of lost samples in discrete-time signals and digital images."

From 1982 to 1992, he was a Researcher with Philips Research Laboratories, Eindhoven, The Netherlands, in various areas of digital signal processing, such as audio and video signal restoration and audio source coding. From 1992 to 2001, he was with the Institute of Perception Research (IPO), Eindhoven, working on speech signal processing and speech synthesis. From 1998 to 2001, he was a Program Manager of the Spoken Language Interfaces research program. He is currently an Associate Professor with the Signals and Systems Group, Department of Electrical Engineering, Mathematics and Computer Science, University of Twente, working in the fields of biometrics and signal processing. He is the author of over 120 papers published in international conference proceedings and journals. He is also a coauthor of the book *An Introduction to Source Coding* (Prentice-Hall) and the author of the book *Restoration of Lost Samples in Digital Signals* (Prentice-Hall). He is the holder of 21 patents in the fields of signal processing. His expertise involves digital signal processing for audio, images and speech; statistical pattern recognition and biometrics.