

Likelihood Ratio Based Biometric Score Fusion

Karthik Nandakumar, *Student Member, IEEE*, Yi Chen, *Student Member, IEEE*,
Sarat C. Dass, *Member, IEEE*, and Anil K. Jain, *Fellow, IEEE*

Abstract—Multibiometric systems fuse information from different sources to compensate for the limitations in performance of individual matchers. We propose a framework for optimal combination of match scores that is based on the likelihood ratio test. The distributions of genuine and impostor match scores are modeled as finite Gaussian mixture model. The proposed fusion approach is general in its ability to handle (i) discrete values in biometric match score distributions, (ii) arbitrary scales and distributions of match scores, (iii) correlation between the scores of multiple matchers and (iv) sample quality of multiple biometric sources. Experiments on three multibiometric databases indicate that the proposed fusion framework achieves consistently high performance compared to commonly used score fusion techniques based on score transformation and classification.

Index Terms—Multibiometric systems, score level fusion, Neyman-Pearson theorem, likelihood ratio test, Gaussian mixture model, image quality

I. INTRODUCTION

Biometrics refers to the automatic identification of individuals based on their anatomical and behavioral characteristics. Biometric systems based on a single source of information (unibiometric systems) suffer from limitations such as the lack of uniqueness and non-universality of the chosen biometric trait, noisy data and spoof attacks [1]. Multibiometric systems fuse information from multiple biometric sources in order to achieve better recognition performance and to overcome other limitations of unibiometric systems [2]–[4]. Fusion can be performed at four different levels of information, namely, sensor, feature, match score, and decision levels. Score level fusion is generally preferred because it offers the best trade-off in terms of the

This research was supported by Army Research Office contract W911NF-06-1-0418.

Karthik Nandakumar, Yi Chen and Anil K. Jain are with the Department of Computer Science and Engineering, Michigan State University, 3115 Engineering Building, East Lansing, MI 48824. Email: {nandakum, chenyl, jain}@cse.msu.edu. Sarat C. Dass is with the Department of Statistics and Probability, Michigan State University, A430 Wells Hall, East Lansing, MI 48824. Email: sdass@stt.msu.edu

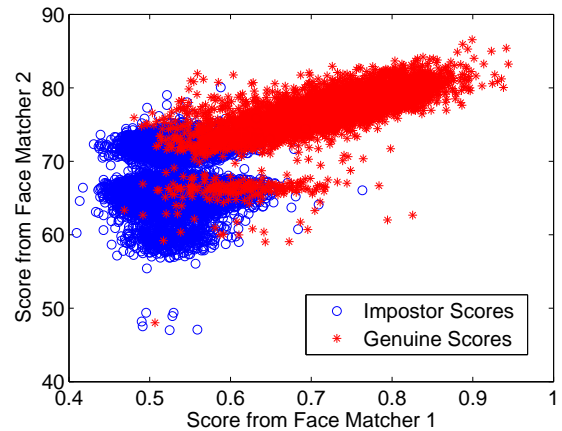


Fig. 1. Non-homogeneity in the match scores provided by the two face matchers in the NIST-Face database [5]. Note that about 0.2% of the scores output by matcher 1 are discrete scores with value -1 which are not shown in this plot.

information content and the ease in fusion. Combining match scores is a challenging task because the scores of different matchers can be either distance or similarity measure, may follow different probability distributions, may provide quite different accuracies and may be correlated. Consider the scores provided by the two face matchers in the NIST-Face database. The scores from these two matchers are in the range $[-1, 1]$ and $[0, 100]$ (see Figure 1) and the Pearson's correlation coefficient for the genuine and impostor scores of the two matchers are 0.7 and 0.3, respectively.

Score fusion techniques can be divided into the following three categories.

- *Transformation-based score fusion*: The match scores are first normalized (transformed) to a common domain and then combined. Choice of the normalization scheme and combination weights is data-dependent and requires extensive empirical evaluation [4], [6], [7].
- *Classifier-based score fusion*: Scores from multiple matchers are treated as a feature vector and a classifier is constructed to discriminate

genuine and impostor scores [2], [8], [9]. When biometric score fusion is considered as a classification problem, the following issues pose challenges. (i) Unbalanced training set: The number of genuine match scores available for training is $O(n)$, but the number of impostor scores is $O(n^2)$, where n is the number of users in the database. (ii) Cost of misclassification: Depending on the biometric application, the cost of accepting an impostor may be very different from the cost of rejecting a genuine user. For example, a biometric system in security applications typically requires the false accept rate (FAR) to be less than 0.1%. Therefore, the fusion classifier needs to minimize the false reject rate (FRR) at the specified FAR values rather than minimizing the total error rate (sum of FAR and FRR) [3]. (iii) Choice of classifier: Given a variety of admissible classifiers, selecting and training a classifier that gives the optimal performance (minimum FRR at a specified FAR) on a given data set is not trivial.

- *Density-based score fusion*: This approach is based on the likelihood ratio test and it requires explicit estimation of genuine and impostor match score densities [3], [10]. Density based approach has the advantage that it directly achieves optimal performance at any desired operating point (FAR), provided the score densities are estimated accurately. In fact, a comparison of eight biometric fusion techniques conducted by NIST [11] with data from 187,000 subjects concluded that “Product of Likelihood Ratios was consistently most accurate, but most complex to implement” and “complexity in this implementation is in the modeling of distributions, rather than fusion per se”. The statement in [11] about the complexity of density estimation was based on the use of kernel density estimator (KDE). The selection of kernel bandwidth and density estimation at the tails proved to be the most complex steps in estimating the score densities using KDE.

In this paper, we show that (i) finite Gaussian mixture model (GMM) is quite effective in modeling the genuine and impostor score densities and is easier to implement than KDE, (ii) fusion based on the resulting density estimates achieves consistently

high performance on three multibiometric databases involving face, fingerprint, iris, and speech modalities and (iii) biometric sample quality can be easily incorporated in the likelihood ratio based fusion framework.

II. LIKELIHOOD RATIO BASED SCORE FUSION

A. Likelihood ratio test

Let $\mathbf{X} = [X_1, X_2, \dots, X_K]$ denote the match scores of K different biometric matchers, where X_k is the random variable representing the match score of the k^{th} matcher, $k = 1, 2, \dots, K$. Let $f_{gen}(\mathbf{x})$ and $f_{imp}(\mathbf{x})$ be the conditional joint densities of the K match scores given the genuine and impostor classes, respectively, where $\mathbf{x} = [x_1, x_2, \dots, x_K]$. Suppose we need to assign the observed match score vector \mathbf{X} to genuine or impostor class. Let Ψ be a statistical test for testing $H_0: \mathbf{X}$ corresponds to an impostor against $H_1: \mathbf{X}$ corresponds to a genuine user. Let $\Psi(\mathbf{x}) = i$ imply that we decide in favor of H_i , $i = 0, 1$. The probability of rejecting H_0 when H_0 is true is known as the *false accept rate* (size or level of the test). The probability of correctly rejecting H_0 when H_1 is true is known as the *genuine accept rate* (power of the test). The Neyman-Pearson theorem [12] states that

- 1) For testing H_0 against H_1 , there exists a test Ψ and a constant η such that

$$P(\Psi(\mathbf{X}) = 1 | H_0) = \alpha \quad (1)$$

and

$$\Psi(\mathbf{x}) = \begin{cases} 1, & \text{when } \frac{f_{gen}(\mathbf{x})}{f_{imp}(\mathbf{x})} \geq \eta, \\ 0, & \text{when } \frac{f_{gen}(\mathbf{x})}{f_{imp}(\mathbf{x})} < \eta. \end{cases} \quad (2)$$

- 2) If a test satisfies equations (1) and (2) for some η , then it is the *most powerful test* for testing H_0 against H_1 at level α .

According to the Neyman-Pearson theorem, given the false accept rate (FAR) α , the *optimal* test for deciding whether a score vector \mathbf{X} corresponds to a genuine user or an impostor is the likelihood ratio test given by equation (2). For a fixed FAR, we can select a threshold η such that the likelihood ratio test maximizes the genuine accept rate (GAR). Based on the Neyman-Pearson theorem, we are guaranteed that *there does not exist any other*

decision rule with a higher GAR. However, this optimality of the likelihood ratio test is guaranteed only when the underlying densities are known. In practice, we estimate the densities $f_{gen}(\mathbf{x})$ and $f_{imp}(\mathbf{x})$ from the training set of genuine and impostor match scores, respectively and the performance of likelihood ratio test will depend on the accuracy of these estimates.

B. Estimation of Match Score Densities

It is well known that the Gaussian density is not appropriate for modeling biometric match scores because the score distributions generally have a long tail and may have more than one mode. Moreover, the presence of discrete score values and correlation among match scores of different biometric matchers makes density estimation a challenging task. Non-parametric techniques like density histogram and kernel density estimator require careful choice of histogram bin width or kernel bandwidth [11], [13] that is critical to the fusion performance. Gaussian mixture model (GMM) has been successfully used to estimate arbitrary densities and theoretical results in [14], [15] show that the density estimates obtained using finite mixture models indeed converge to the true density when a sufficient number of training samples are available. For these reasons, we use GMM for estimating the genuine and impostor score densities.

Let $\phi^K(\mathbf{x}; \boldsymbol{\mu}, \Sigma)$ be the K -variate Gaussian density with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ , i.e., $\phi^K(\mathbf{x}; \boldsymbol{\mu}, \Sigma) = (2\pi)^{-K/2} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right)$. The estimates of $f_{gen}(\mathbf{x})$ and $f_{imp}(\mathbf{x})$ are obtained as a mixture of Gaussians as follows.

$$\hat{f}_{gen}(\mathbf{x}) = \sum_{j=1}^{M_{gen}} p_{gen,j} \phi^K(\mathbf{x}; \boldsymbol{\mu}_{gen,j}, \Sigma_{gen,j}), \quad (3)$$

$$\hat{f}_{imp}(\mathbf{x}) = \sum_{j=1}^{M_{imp}} p_{imp,j} \phi^K(\mathbf{x}; \boldsymbol{\mu}_{imp,j}, \Sigma_{imp,j}), \quad (4)$$

where M_{gen} (M_{imp}) is the number of mixture components used to model the density of the genuine (impostor) scores, $p_{gen,j}$ ($p_{imp,j}$) is the weight assigned to the j^{th} mixture component in $\hat{f}_{gen}(\mathbf{x})$ ($\hat{f}_{imp}(\mathbf{x})$), $\sum_{j=1}^{M_{gen}} p_{gen,j} = \sum_{j=1}^{M_{imp}} p_{imp,j} = 1$.

Selecting the appropriate number of components is one of the most challenging issues in mixture density estimation; while a mixture with too many components may result in over-fitting, a mixture with too few components may not approximate the true density well. The GMM fitting algorithm proposed in [16]¹ automatically estimates the number of components and the component parameters using an EM algorithm and the minimum message length criterion. This algorithm is also robust to initialization of parameter values (mean vectors and covariance matrices) and can handle discrete components in the match score distribution by modeling the discrete scores as a mixture component with very small variance. This is achieved by adding a small value (regularization factor) to the diagonal of the covariance matrices. The actual value of this variance does not affect the performance as long as it is insignificant compared to the variance of the continuous components in the match score distribution. For example, the lowest value of variance in the match score data used in our experiments is of the order of 10^{-3} . Hence, we used the value of 10^{-5} as the lower bound for the variance. Our experiments indicate that a value smaller than 10^{-5} (say, 10^{-7} or 10^{-9}) does not change the performance of GMM. Since we do not place any restrictions on the component covariance matrices $\Sigma_{gen,j}$ and $\Sigma_{imp,j}$, the estimates of the joint densities $\hat{f}_{gen}(\mathbf{x})$ and $\hat{f}_{imp}(\mathbf{x})$ also take into account the correlation between the match scores. Figure 2 shows that Gaussian mixture model reliably estimates the 2-D genuine and impostor densities of the two face matchers in the NIST-Face database.

We now define the likelihood ratio (LR) fusion rule as follows. Given a vector of K match scores $\mathbf{x} = [x_1, \dots, x_K]$ and estimated densities $\hat{f}_{gen}(\mathbf{x})$ and $\hat{f}_{imp}(\mathbf{x})$, compute the likelihood ratio $LR(\mathbf{x}) = \hat{f}_{gen}(\mathbf{x}) / \hat{f}_{imp}(\mathbf{x})$.

Assign \mathbf{x} to the genuine class if $LR(\mathbf{x}) \geq \eta$, (5) where η is the decision threshold that is determined based on the specified FAR.

C. Incorporating Image Quality in Fusion

It is well known that the quality of biometric samples has a significant impact on the accuracy of a

¹The MATLAB code for this algorithm is available at <http://www.lx.it.pt/~mtf/mixturecode.zip>

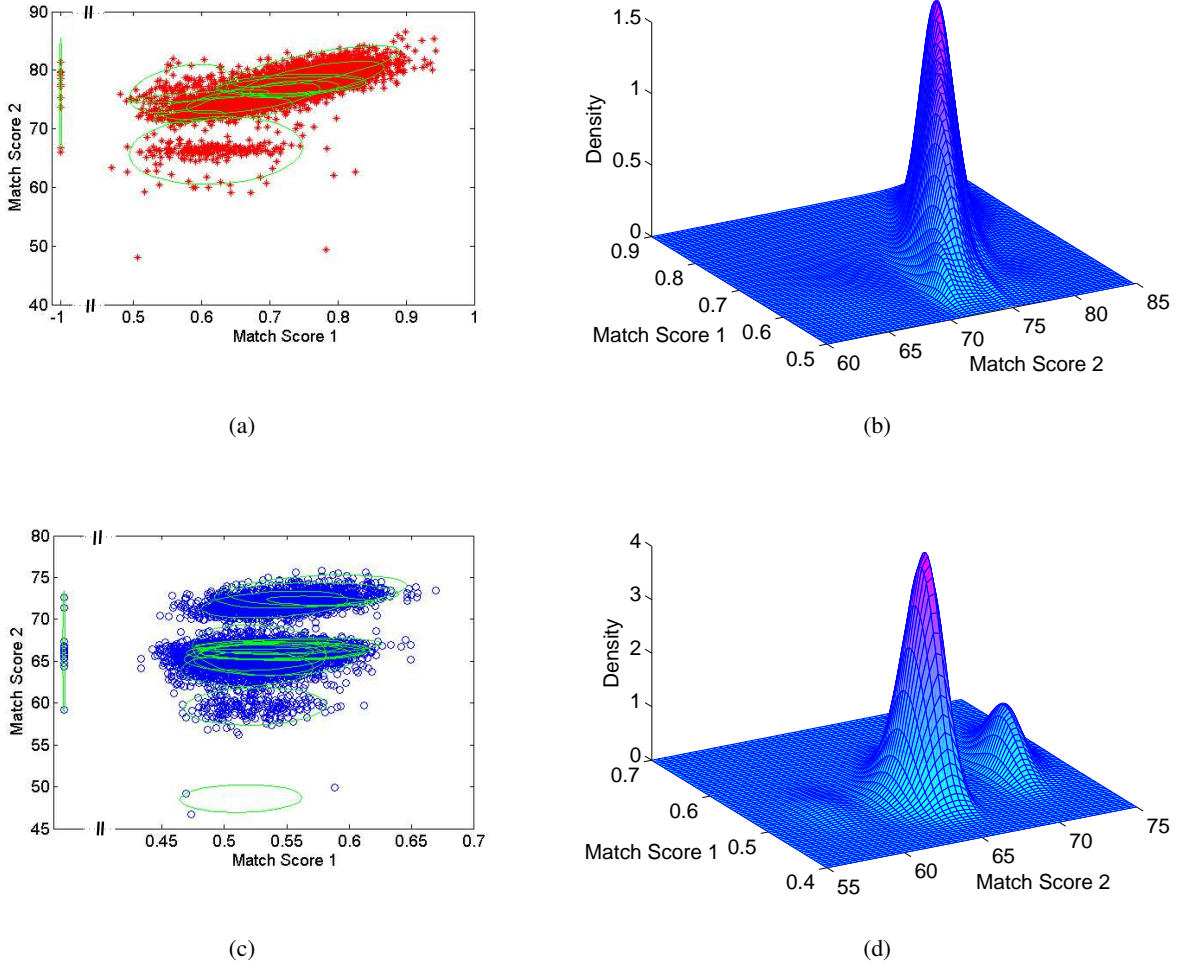


Fig. 2. Density estimation based on Gaussian mixture models for the NIST-Face database. (a) Scatter plot of the genuine scores along with the fitted mixture components, (b) density estimates of the genuine scores, (c) scatter plot of the impostor scores along with the fitted mixture components and (d) density estimates of the impostor scores. In this example, the estimated number of mixture components is 12 for the genuine density and 19 for the impostor density. The discrete score in the first matcher at value -1 is modeled as a separate mixture component as shown in (a) and (c).

matcher [17]. Several schemes have used biometric sample quality for improving the performance of a multibiometric system [8], [18]–[20] by assigning weights to individual match scores. To incorporate sample quality in the likelihood ratio framework, we first make the following observation. Since a poor quality sample will be difficult to classify as genuine or impostor (see Figure 3), the likelihood ratio for such a sample will be close to 1. On the other hand, for good quality samples, the likelihood ratio will be greater than 1 for genuine users and less than 1 for impostors. Hence, if we estimate the joint density of the match score and the associated quality, the resulting likelihood ratios will be implicitly weighted by the respective sample quality.

Let $\mathbf{Q} = [Q_1, Q_2, \dots, Q_K]$ be the quality vector, where Q_k is the quality of the match score provided by the k^{th} matcher, $k = 1, \dots, K$. Let $\hat{f}_{gen}(\mathbf{x}, \mathbf{q})$ and $\hat{f}_{imp}(\mathbf{x}, \mathbf{q})$ be the joint densities of the K -dimensional match score vector and the K -dimensional quality vector estimated from the genuine and impostor template-query pairs, respectively. The quality-based likelihood ratio, $QLR(\mathbf{x}, \mathbf{q})$, is given by

$$QLR(\mathbf{x}, \mathbf{q}) = \frac{\hat{f}_{gen}(\mathbf{x}, \mathbf{q})}{\hat{f}_{imp}(\mathbf{x}, \mathbf{q})}. \quad (6)$$

The decision rule based on $QLR(\mathbf{x}, \mathbf{q})$ is similar to the one used in equation (5). Note that the joint

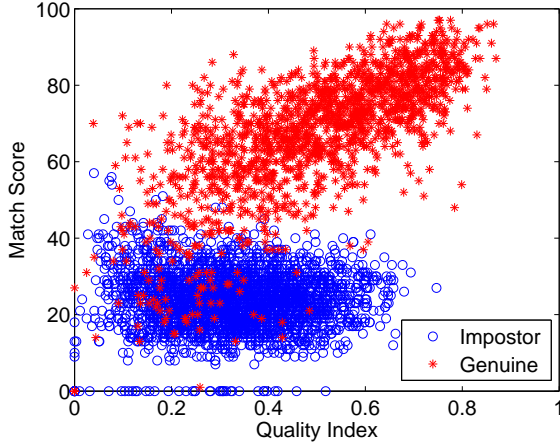


Fig. 3. Variation of match score with quality for fingerprint modality in the WVU-Multimodal database. We observe that the genuine and impostor match scores are well-separated only for good quality (with quality index > 0.5) samples.

density estimation of (\mathbf{X}, \mathbf{Q}) now involves $2K$ variables which may not be very reliable with limited training data. To avoid the curse of dimensionality, we assume independence of K matchers and write equation (6) as $\hat{f}(\mathbf{x}, \mathbf{q}) = \prod_{k=1}^K \hat{f}_k(x_k, q_k)$. So, now we estimate K 2-dimensional densities for both genuine and impostor classes. To perform quality-based fusion based on equation (6), we use quality metrics proposed in [21] for fingerprint and iris.

III. EXPERIMENTAL RESULTS

The performance of likelihood ratio based fusion rule was evaluated on two public-domain databases, namely, NIST-BSSR1 [5] and XM2VTS-Benchmark databases [22]. The performance of the quality-based product fusion rule was evaluated only on the WVU-Multimodal database since the two public databases did not contain raw images to estimate the quality. The three databases used in our study are summarized in Table I.

For each experiment, half of the genuine and half of the impostor match scores were randomly selected to form the training set for density estimation. This training-test partitioning was repeated m times ($m = 20$) and the reported ROC curves correspond to the mean GAR values over the m trials at different FAR values. We also report the 95% confidence interval for the improvement in GAR achieved by likelihood ratio-based fusion at

specific FAR values².

The receiver operating characteristic (ROC) curves of the individual matchers and the LR fusion rule for the three partitions of the NIST-BSSR1 and XM2VTS databases are shown in Figure 4. The LR fusion leads to significant improvement in the performance compared to the best single modality on all the four databases. At a false accept rate (FAR) of 0.01%, the improvement in the genuine accept rate (GAR) achieved due to LR fusion is presented in Table II. We observe that the 95% confidence intervals estimated in Table II are fairly tight which indicates that the performance improvement is consistent across the 20 cross-validation trials. We also observe that multimodal fusion (face and two fingers) in NIST-BSSR1 results in larger improvement in GAR than two-finger fusion or multi-algorithm fusion (two face matchers).

The performance of LR fusion rule is first compared to fusion based on Support Vector Machine (SVM) classifier. While the performance of SVM based fusion is comparable to LR fusion on the NIST-Fingerprint and XM2VTS-Benchmark databases (see Figures 4(b) and Figures 4(d)), it is inferior to LR fusion on the NIST-Multimodal and NIST-Face databases (see Figures 4(a) and 4(c)). Moreover, the kernel function and the associated parameters for SVM must be carefully chosen in order to achieve this performance. For example, while linear SVM gave good performance on the NIST-Multimodal and XM2VTS-Benchmark databases, we had to use a radial basis function kernel with different parameter values for the NIST-Fingerprint ($\gamma = 0.005$) and NIST-Face ($\gamma = 0.1$) databases to obtain the results reported in Figure 4.

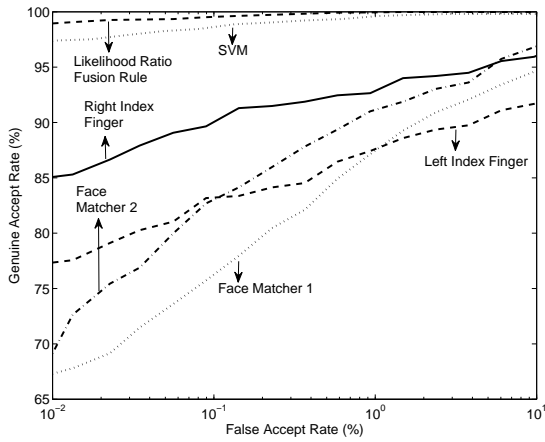
We also compared the performance of LR fusion rule with a commonly used transformation-based score fusion technique, namely, the sum of scores fusion method, which can be derived as an approximation to LR fusion rule under a set of conditions [23]. However, in order to use the sum of scores fusion method, we need to choose a score normalization method. After an empirical evaluation, we found that the min-max normalization [7] is the best for the datasets used here. The ROC curves for the

²For experiments on XM2VTS-Benchmark database, we do not randomly partition the scores into training and test sets because this partitioning is already defined in [22]. Hence, confidence intervals are not estimated for this database.

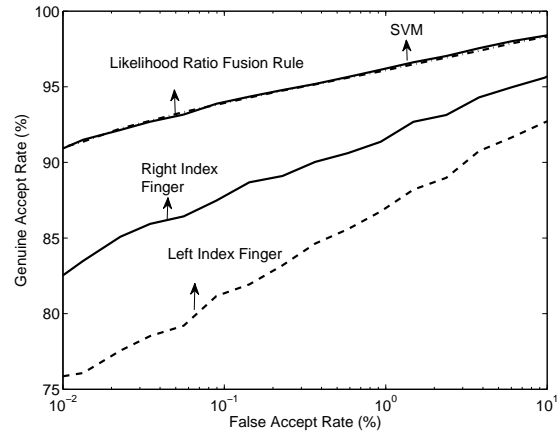
TABLE I

SUMMARY OF MULTIBIOMETRIC DATABASES. NOTE THAT THE NIST-MULTIMODAL, NIST-FINGERPRINT AND NIST-FACE DATABASES ARE DIFFERENT PARTITIONS OF THE NIST BIOMETRIC SCORE SET RELEASE-1 (BSSR1).

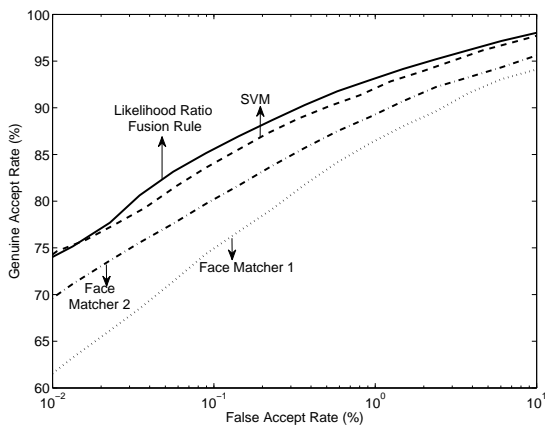
Database	Biometric Traits	No. of matchers (K)	No. of users
NIST-Multimodal	Fingerprint (Two fingers) Face (Two matchers)	4	517
NIST-Fingerprint	Fingerprint (Two fingers)	2	6,000
NIST-Face	Face (Two matchers)	2	3,000
XM2VTS-Benchmark	Face (Five matchers) Speech (Three matchers)	8	295
WVU-Multimodal	Fingerprint, Iris	2	320



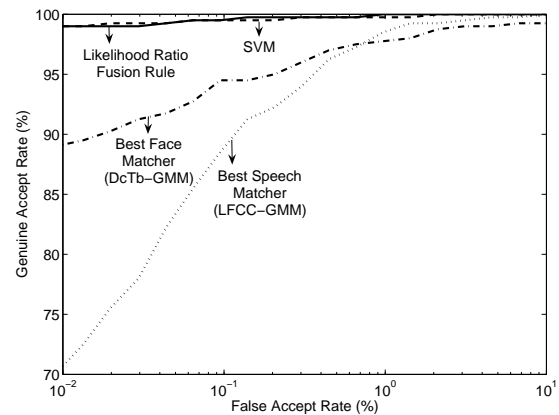
(a)



(b)



(c)



(d)

Fig. 4. Performance of likelihood ratio based fusion rule and SVM-based fusion on (a) NIST-Multimodal, (b) NIST-Fingerprint, (c) NIST-Face and (d) XM2VTS-Benchmark databases. Although there are 8 matchers in the XM2VTS-Benchmark database, only the ROC curves of the best face matcher (DcTb-GMM) and the best speech matcher (LFCC-GMM) are shown in (d) for clarity. A linear SVM was used in (a) and (d) and a SVM with radial basis function kernel was used in (b) and (c) (with $\gamma = 0.005$ and $\gamma = 0.1$, respectively).

TABLE II

PERFORMANCE IMPROVEMENT ACHIEVED BY LIKELIHOOD RATIO BASED FUSION.

Database	Best Single Matcher	Mean GAR at 0.01% FAR		95% Confidence Interval on increase in GAR at 0.01% FAR
		Best Single Matcher	Likelihood Ratio based Fusion	
NIST-Multimodal	Right Index Finger	85.3%	99.1%	[13.5%, 14%]
NIST-Fingerprint	Right Index Finger	83.5%	91.4%	[7.6%, 8.2%]
NIST-Face	Matcher 1	71.2%	77.2%	[4.7%, 7.3%]
XM2VTS-Benchmark	DCTb-GMM Face Matcher	89.5%	98.7%	N/A

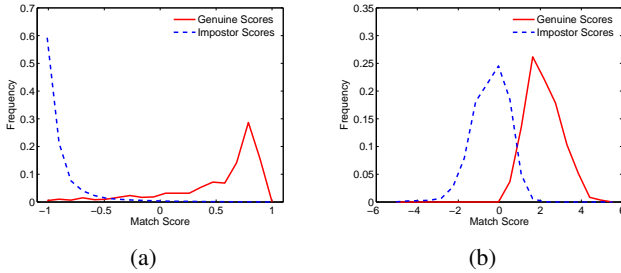


Fig. 6. Distribution of genuine and impostor match scores in the XM2VTS-Benchmark database for (a) MLP-face classifier and (b) speech.

LR rule and sum of scores fusion rule with min-max normalization are shown in Figure 5. While the sum rule is quite comparable to LR rule for NIST-Multimodal database, it does not perform well on the XM2VTS-Benchmark database (see Figure 5(b)).

The reason for the inferior performance of sum rule in the case of XM2VTS-Benchmark database is that the score distributions of the face and speech matchers are very different (see Figure 6). The min-max normalization is not effective for the face scores in Figure 6(a) because the genuine and impostor scores are peaked around 1 and -1 , respectively. This is because the face match scores were the output of a multi-layer perceptron classifier that used a tanh non-linearity function. However, if we first transform the distribution in Figure 6(a) by applying an inverse tangent function to these scores followed by min-max normalization, then the performance of the sum rule improves and it is now comparable to the LR fusion as observed in

Figure 5(b). These results demonstrate that while it is possible to achieve good fusion performance for a specific database using the simple sum rule by carefully choosing the normalization scheme, the proposed LR fusion rule is a general approach that consistently provides good performance.

Figure 7 shows the performance of the LR and QLR fusion rules on the WVU-Multimodal database. For this data, the LR fusion rule improves the GAR compared to the best single modality (iris) and the quality-based fusion (QLR) rule further improves the GAR. For example, at a FAR of 0.001%, the mean GAR of the iris modality is 66.7%, while the GAR values of the LR and QLR fusion rules are 85.3% and 90%, respectively. The 95% confidence interval for the improvement in GAR obtained by using QLR fusion instead of LR fusion is [4.1%, 5.3%].

IV. CONCLUSIONS

We have proposed a framework for fusion of match scores in a multibiometric system based on the likelihood ratio test. This approach is general and is able to minimize the FRR at a specified FAR. In practice, one needs to reliably estimate the genuine and impostor match score densities from the available training set of match scores. Due to the availability of relatively large multi-biometric databases and the low dimensionality of the score vectors, the density estimation problem in the proposed LR fusion rule is quite tractable. We show that densities estimated using a mixture of Gaussian model lead to good performance on

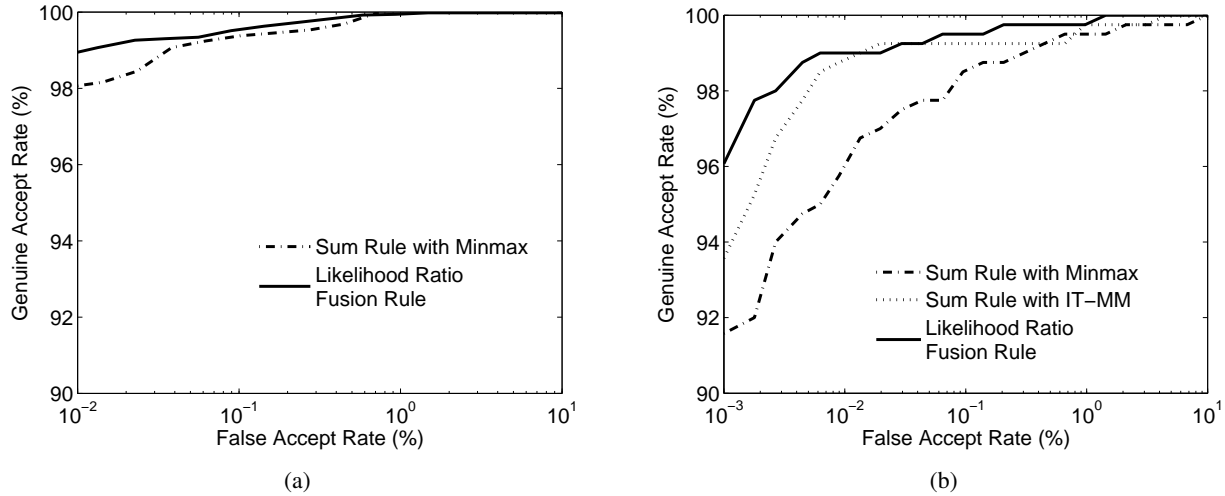


Fig. 5. Performance of likelihood ratio based fusion rule and sum of scores fusion rule with min-max normalization on (a) NIST-Multimodal database and (b) XM2VTS-Benchmark database. In (b), IT-MM denotes that an inverse tangent function is applied only to the match scores of the MLP classifiers prior to normalizing all the match scores using min-max normalization.

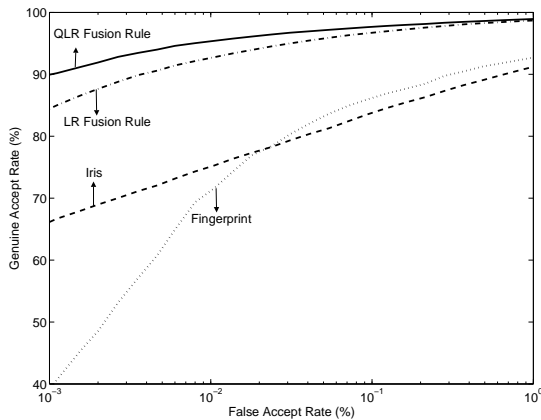


Fig. 7. Performance of LR and QLR fusion rules on the WVU-Multimodal database.

several multibiometric databases. Based on these experiments, our conclusions are as follows.

- The likelihood ratio based fusion rule with GMM-based density estimation achieves consistently high recognition rates without the need for parameter tuning by the system designer.
- While other fusion schemes such as sum rule and SVM can provide performance comparable to that of LR fusion, these approaches require careful selection of parameters (e.g., score normalization and fusion weights in sum rule, type of kernel and kernel parameters in SVM) on a case-by-case basis. The LR rule does not need

to make these choices.

- Biometric sample quality information can be incorporated within the likelihood ratio based fusion framework leading to improvements in the performance of multibiometric systems.

REFERENCES

- [1] A. Ross, K. Nandakumar, and A. K. Jain, *Handbook of Multi-biometrics*. Springer-Verlag, 2006.
- [2] R. Brunelli and D. Falavigna, "Person Identification Using Multiple Cues," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 10, pp. 955–966, October 1995.
- [3] S. Prabhakar and A. K. Jain, "Decision-level Fusion in Fingerprint Verification," *Pattern Recognition*, vol. 35, no. 4, pp. 861–874, April 2002.
- [4] K.-A. Toh, X. Jiang, and W.-Y. Yau, "Exploiting Global and Local Decisions for Multimodal Biometrics Verification," *IEEE Transactions on Signal Processing, (Supplement on Secure Media)*, vol. 52, no. 10, pp. 3059–3072, October 2004.
- [5] National Institute of Standards and Technology, "NIST Biometric Scores Set - release 1," 2004, Available at <http://www.itl.nist.gov/iad/894.03/biometricscores>.
- [6] R. Snelick, U. Uludag, A. Mink, M. Indovina, and A. K. Jain, "Large Scale Evaluation of Multimodal Biometric Authentication Using State-of-the-Art Systems," *IEEE Transactions on PAMI*, vol. 27, no. 3, pp. 450–455, March 2005.
- [7] A. K. Jain, K. Nandakumar, and A. Ross, "Score Normalization in Multimodal Biometric Systems," *Pattern Recognition*, vol. 38, no. 12, pp. 2270–2285, December 2005.
- [8] 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, May 2005.
- [9] Y. Ma, B. Cukic, and H. Singh, "A Classification Approach to Multi-biometric Score Fusion," in *Proceedings of Fifth International Conference on AVBPA*, Rye Brook, USA, July 2005, pp. 484–493.

- [10] P. Griffin, "Optimal Biometric Fusion for Identity Verification," Identix Research, Tech. Rep. RDNJ-03-0064, 2004.
- [11] B. Ulery, A. R. Hicklin, C. Watson, W. Fellner, and P. Hallinan, "Studies of Biometric Fusion," NIST, Tech. Rep. IR 7346, September 2006.
- [12] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*. Springer, 2005.
- [13] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, 1986.
- [14] J. Q. Li and A. Barron, "Mixture Density Estimation," in *Advances in Neural Information Processings Systems 12*, S. A. Solla, T. K. Leen, and K.-R. Muller, Eds. San Mateo, USA: Morgan Kaufmann Publishers, 1999.
- [15] A. Rakhlin, D. Panchenko, and S. Mukherjee, "Risk Bounds for Mixture Density Estimation," *ESAIM: Probability and Statistics*, vol. 9, pp. 220–229, June 2005.
- [16] M. Figueiredo and A. K. Jain, "Unsupervised Learning of Finite Mixture Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 3, pp. 381–396, March 2002.
- [17] C. Wilson, A. R. Hicklin, M. Bone, H. Korves, P. Grother, B. Ulery, R. Micheals, M. Zoepfl, S. Otto, and C. Watson, "Fingerprint Vendor Technology Evaluation 2003: Summary of Results," NIST, Tech. Rep. IR 7123, June 2004.
- [18] K.-A. Toh, W.-Y. Yau, E. Lim, L. Chen, and C.-H. Ng, "Fusion of Auxiliary Information for Multi-modal Biometrics Authentication," in *Proceedings of ICBA*, Hong Kong, July 2004, pp. 678–685.
- [19] N. Poh and S. Bengio, "Improving Fusion with Margin-Derived Confidence in Biometric Authentication Tasks," in *Proceedings of Fifth International Conference on AVBPA*, Rye Brook, USA, July 2005, pp. 474–483.
- [20] J. P. Baker and D. E. Maurer, "Fusion of Biometric Data with Quality Estimates via a Bayesian Belief Network," in *Proceedings of Biometrics Symposium (BC2005)*, Arlington, USA, September 2005, pp. 21–22.
- [21] K. Nandakumar, Y. Chen, A. K. Jain, and S. Dass, "Quality-based Score Level Fusion in Multibiometric Systems," in *Proceedings of International Conference on Pattern Recognition (ICPR)*, Hong Kong, China, August 2006, pp. 473–476.
- [22] N. Poh and S. Bengio, "Database, Protocol and Tools for Evaluating Score-Level Fusion Algorithms in Biometric Authentication," *Pattern Recognition*, vol. 39, no. 2, pp. 223–233, February 2006.
- [23] J. Kittler, M. Hatef, R. P. Duin, and J. G. Matas, "On Combining Classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226–239, March 1998.