Biometric Fusion: Does Modeling Correlation Really Matter?

Karthik Nandakumar, Arun Ross and Anil K. Jain

Abstract— Sources of information in a multibiometric system are often assumed to be statistically independent in order to simplify the design of the fusion algorithm. However, the independence assumption may not be always valid. In this paper, we analyze whether modeling the dependence between match scores in a multibiometric system has any effect on the fusion performance. Our analysis is based on the likelihood ratio (LR) based fusion framework, which guarantees optimal performance if the match score densities are known. We show that the assumption of independence between matchers has a significant negative impact on the performance of the LR fusion scheme only when (i) the dependence characteristics among genuine match scores is different from that of the impostor scores and (ii) the individual matchers are not very accurate.

I. INTRODUCTION

Multibiometric systems [1] integrate evidence from different biometric sources and typically provide better recognition accuracy and increased population coverage compared to unibiometric systems. Multibiometric systems generally fuse information at the match score level because it offers the best trade-off in terms of the information content and the ease in fusion. Score level fusion schemes often assume that the component matchers are statistically independent to simplify the design of the fusion algorithm. For example, consider the likelihood ratio (LR) based fusion approach [2], which is a more principled approach in the context of biometrics because it directly minimizes the false reject rate (FRR) at the specified values of false accept rate (FAR), provided the match score densities are known. In LR fusion, the joint density of the match scores can be expressed as the product of the marginal densities if the matchers are assumed to be independent. Since marginal density estimation is generally more efficient than joint density estimation, the independence assumption is advantageous in LR fusion, particularly for small sample size situations. However, the designer of a multibiometric system needs to analyze two important issues, namely, (i) Are the biometric matchers under consideration truly independent? and (ii) If the matchers are *not* independent, is there a significant reduction in the accuracy of the fusion scheme due to assumption of independence?

While different biometric traits of an individual (e.g., face and fingerprint) tend to be independent, the independence assumption may not be true in a general multibiometric system. Examples of multibiometric systems involving correlated matchers include (i) systems using physically related traits (e.g., speech and lip movement), (ii) multiple matchers operating on the same biometric data (e.g., two different face matchers) and (iii) multiple samples of the same biometric trait (e.g., two impressions of a person's right index finger). Hence, it is important to study the role of dependence between the matchers in biometric fusion.

Many researchers have investigated whether fusion of independent matchers results in better performance compared to fusion of dependent matchers [3], [4], [5], [6], [7]. While the answer to this question is certainly important in designing a multibiometric system, the focus of this paper is to analyze a different issue: *Given that the matchers are dependent, does modeling the correlation between the matchers and incorporating this dependence information into the fusion scheme lead to any significant performance improvement compared to a simplified fusion model that is based on the independence assumption?* To the best of our knowledge, this issue has not been adequately addressed in the literature. There have been claims of marginal improvement (see [8], [9], [10]) in the fusion performance when the correlation between match scores is considered, but a systematic analysis of this problem has not been presented thus far.

The contributions of this paper are three-fold. First, we present a mathematical framework based on the likelihood ratio-based fusion approach to analyze the effect of modeling correlation on the fusion performance. The proposed framework is implicitly based on the assumption that there are no *user-specific* and/or *temporal* variations in the match score distributions. As a result, this analysis cannot be applied to scenarios like fusion of scores obtained from successive frames in a video sequence. Secondly, we consider a simple bivariate Gaussian distribution for the match scores and study the difference in fusion performance of the full model (includes correlation) and the independence model (ignores correlation) by systematically varying the strength of correlation between the match scores and the performance of the individual matchers. Finally, we verify whether the trends observed in the Gaussian example, also hold good in the case of real biometric data.

II. MATHEMATICAL FRAMEWORK

Let $S = [S_1, \cdots, S_K]$ be the random vector corresponding to the match scores of K different biometric matchers, where S_k is the random variable representing the match score provided by the kth matcher, $k = 1, \cdots, K$. Let $f_g(s)$ and $f_i(s)$ be the conditional joint density of the K match scores

given the genuine and impostor classes, respectively, where $s = [s_1, \cdots, s_K]$. Let Ψ be a statistical test for testing H_0: *score S corresponds to an impostor* vs. H_1: *score S corresponds to a genuine user*. Let $\Psi(s) = c$ imply that we decide in favor of H_c, where $c = 0, 1$. 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 [11] states that

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

$$P(\Psi(S) = 1 | H_0) = \alpha \qquad (1)$$

and

$$\Psi(s) = \begin{cases} 1, & \text{when } \frac{f_g(s)}{f_i(s)} \geq \eta, \\ 0, & \text{when } \frac{f_g(s)}{f_i(s)} < \eta. \end{cases} \qquad (2)$$

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

According to the Neyman-Pearson theorem, given the false accept rate (FAR) α, the *optimal* test for deciding whether a match score S 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) and there does not exist any other decision rule with a higher GAR. The optimality of the LR test is guaranteed only when the underlying densities are known. If $\lambda(s) = \log\left(\frac{f_g(s)}{f_i(s)}\right)$, the decision rule based on the LR test can be stated as,

$$\text{Assign } s \text{ to the genuine class if } \lambda(s) \geq \tau, \qquad (3)$$

where $\tau = \log \eta$. Since the decision rule in equation (3) requires knowledge of the true densities $f_g(s)$ and $f_i(s)$, we refer to it as LR fusion under the true model.

Given a FAR α, τ is computed such that $\int_{\lambda(S) \geq \tau} f_i(s)ds = \alpha$ and the corresponding GAR (power of the test) is given by $P(\alpha) = \int_{\lambda(S) \geq \tau} f_g(s)ds$. Equivalently, if the densities of the log-likelihood ratio under the genuine and impostor hypotheses are known (denoted by $h_g(\lambda)$ and $h_i(\lambda)$, respectively), $P(\alpha) = \int_\tau^\infty h_g(\lambda)d\lambda$, where $\int_\tau^\infty h_i(\lambda)d\lambda = \alpha$. Furthermore, if $H_g(\lambda)$ and $H_i(\lambda)$ are the cumulative distribution functions (cdf) corresponding to the densities $h_g(\lambda)$ and $h_i(\lambda)$, respectively, and if $H_g^{-1}(.)$ and $H_i^{-1}(.)$ are the corresponding inverse cdf, the GAR (or power) can be simplified as,

$$P(\alpha) = 1 - H_g\left(H_i^{-1}(1 - \alpha)\right). \qquad (4)$$

In practice, we only have a finite set of genuine and impostor match scores, so we need to reliably estimate the densities $f_g(s)$ and $f_i(s)$ before applying the likelihood ratio test. One approach is to directly estimate the K-dimensional densities $f_g(s)$ and $f_i(s)$ using non-parametric techniques like multivariate kernel density estimation [12] or semi-parametric techniques like mixture of Gaussians [13]. This approach does not make any assumption about the

independence between the matchers and hence, we refer to it as the full model. Let $\hat{f}_g(s)$ and $\hat{f}_i(s)$ be the density estimates of the genuine and impostor match scores obtained under the full model. If $\lambda_1(s) = \log\left(\frac{\hat{f}_g(s)}{\hat{f}_i(s)}\right)$, the LR-based decision rule under the full model is,

$$\text{Assign } s \text{ to the genuine class if } \lambda_1(s) \geq \tau_1, \qquad (5)$$

where τ_1 is computed such that $\int_{\lambda_1(s) \geq \tau_1} f_i(s)ds = \alpha$ and α is the given FAR. Let $m_g(\lambda_1)$ and $m_i(\lambda_1)$ be the densities, $M_g(\lambda_1)$ and $M_i(\lambda_1)$ be the distribution functions, and $M_g^{-1}(.)$ and $M_i^{-1}(.)$ be the inverse distribution functions of $\lambda_1(s)$ under the genuine and impostor hypotheses, respectively. The power of the LR test under the full model is given by,

$$\hat{P}(\alpha) = 1 - M_g\left(M_i^{-1}(1 - \alpha)\right). \qquad (6)$$

The Neyman-Pearson theorem guarantees that the LR test under the true model is uniformly most powerful. Therefore, $\{P(\alpha) - \hat{P}(\alpha)\} \geq 0, \forall \alpha$ and the loss in power due to density estimation under the full model can be defined as,

$$\Delta P_{tf} = \int_0^1 \left(P(\alpha) - \hat{P}(\alpha)\right) d\alpha. \qquad (7)$$

Note that $\Delta P_{tf} \geq 0$. However, an important characteristic of the LR fusion framework is that its performance is dependent only on the ratio of the genuine and impostor density estimates and not the individual density estimates. This distinction is critical because as Silverman observed (see page 124 in [12]), "*it may well be the case that even crude density estimates will give reasonable results* (when used to give a discrimination rule rather than as an end in themselves)." This was also demonstrated by Devroye et al. (see Theorem 10.1 in [14]), who showed that while the kernel density estimates themselves converge to the true densities only at the rate of $O(N^{-4/5})$ [15], where N is the number of training samples, the error rate of a classifier based on the likelihood ratio test converges to the true Bayes error at an exponential rate. Hence, when sufficient number of training samples are available, ΔP_{tf} tends to zero.

Another approach to estimate the joint densities of genuine and impostor match scores is to assume that the match scores of different matchers are statistically independent. Under the independence model, the joint density of the K match scores can be approximated as the product of the marginal densities. This greatly simplifies the density estimation process. Let $\tilde{f}_g(s) = \prod_{k=1}^K \tilde{f}_{g,k}(s_k)$ and $\tilde{f}_i(s) = \prod_{k=1}^K \tilde{f}_{i,k}(s_k)$ be the density estimates obtained under the independence model, where $\tilde{f}_{g,k}(s_k)$ and $\tilde{f}_{i,k}(s_k)$ are the marginal density estimates of the genuine and impostor match scores of the kth matcher, $k = 1, \cdots, K$. If $\lambda_2(s) = \log\left(\frac{\tilde{f}_g(s)}{\tilde{f}_i(s)}\right)$ and τ_2 is such that $\int_{\lambda_2(s) \geq \tau_2} f_i(s)ds = \alpha$, the LR-based decision rule under the independence model can be stated as,

$$\text{Assign } s \text{ to the genuine class if } \lambda_2(s) \geq \tau_2. \qquad (8)$$

Let $q_g(\lambda_2)$ and $q_i(\lambda_2)$ be the densities, $Q_g(\lambda_2)$ and $Q_i(\lambda_2)$ be the distribution functions, and $Q_g^{-1}(.)$ and $Q_i^{-1}(.)$ be the

inverse distribution functions of $\lambda_2(s)$ under the genuine and impostor hypotheses, respectively. The power of the LR test under the independence model is,

$$\tilde{P}(\alpha) = 1 - Q_g\left(Q_i^{-1}(1-\alpha)\right). \tag{9}$$

and the loss in power due to density estimation under the independence model can be defined as,

$$\Delta P_{ti} = \int_0^1 \left(P(\alpha) - \tilde{P}(\alpha)\right) d\alpha. \tag{10}$$

Again due to the NP theorem, $\Delta P_{ti} \geq 0$. However, ΔP_{ti} may not always tend to zero as the sample size increases, because if the match scores of different matchers are indeed dependent, LR fusion under independence model will always result in inferior performance compared to the true model.

Both ΔP_{tf} and ΔP_{ti} can be computed only if the true densities of the genuine and impostor match scores are known. However, if we assume that one of the densities is known (say impostor score density), then it can be shown the loss in power is equal to the Kullback-Leibler divergence between the true and estimated genuine score densities [16], i.e., $\Delta P_{tf} = D(f_g||\hat{f}_g)$ and $\Delta P_{ti} = D(f_g||\tilde{f}_g)$, where $D(p||q) = \int p(x) \log\left(\frac{p(x)}{q(x)}\right) dx$. Similarly, if we assume that either genuine or impostor match scores are correlated (but not both), it can be shown that the loss in power due to the independence assumption (ΔP_{ti}) is related to the multi-information between the scores [9]. As pointed out earlier, the performance of the LR fusion scheme depends on the likelihood ratio estimate and not the individual density estimates of the genuine or impostor match scores. Therefore, when both the genuine and impostor densities need to be estimated and the scores of different matchers exhibit dependence characteristics under both genuine and impostor hypotheses, it is very difficult to derive any theoretical results or bounds for ΔP.

In this paper, we focus on measuring the difference in the fusion performance between the full model and the independence model. The difference in power between the full and independence models can be defined as

$$\Delta P_{fi} = \int_0^1 \left(\hat{P}(\alpha) - \tilde{P}(\alpha)\right) d\alpha. \tag{11}$$

Note that ΔP_{fi} cannot be referred to as the "loss in power" because there is no guarantee that ΔP_{fi} will always be greater than zero. Typically, we can expect ΔP_{fi} to be positive, because the full model provides more accurate density estimates than the independence model. Positive values of ΔP_{fi} indicate a loss in performance due to the independence assumption. However, when the number of training samples is small and K is large, directly estimating the K dimensional joint densities is more difficult and error-prone compared to marginal density estimation. Hence, in extreme cases, it may be possible to have negative values of ΔP_{fi}.

III. THE GAUSSIAN EXAMPLE

Knowledge of the distributions of log-likelihood ratio is essential to estimate any possible loss in fusion performance

due to the independence assumption. Vuong [17] showed that it is possible to derive the asymptotic distribution of log-likelihood ratio ($h_g(\lambda)$ and $h_i(\lambda)$) under certain assumptions. When the genuine and impostor densities belong to different family of distributions that are non-nested (i.e., none of the individual distributions in one model may be obtained from the other model either by imposition of parameter restrictions or through a limiting process), the asymptotic distribution ($h_g(\lambda)$ or $h_i(\lambda)$) can be approximated by a normal distribution [17]. However, in the biometric fusion scenario, the genuine and impostor densities belong to the same family (e.g., mixture of Gaussians) and they overlap partially. In this case, Vuong showed that the distribution of the log-likelihood ratio may be a weighted sum of chi-square distributions, whose parameters are usually difficult to estimate in practice. Hence, we first consider a simple distribution model for the match scores and estimate the distribution of log-likelihood ratio empirically using a large number of samples from this known distribution.

In order to illustrate the relationship between the matcher correlation and the difference in power between the full and independence models (ΔP_{fi}), we consider a simple model involving two matchers ($K = 2$). Suppose the class conditional densities of match scores from the two matchers follow the bivariate normal distribution. Without loss of generality, the match score densities can be represented as $f_g(s) = \phi^2(s; \mu_g, \Sigma_g)$ and $f_i(s) = \phi^2(s, \mu_i, \Sigma_i)$, where $\phi^K(s; \mu, \Sigma)$ represents the K-variate Gaussian density with mean vector μ and covariance Σ, $\mu_g = [\mu_1 \ \mu_2]^T$, $\mu_i = [0 \ 0]^T$,

$$\Sigma_g = \begin{pmatrix} \sigma_1^2 & \rho_g \sigma_1 \sigma_2 \\ \rho_g \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix}, \text{and } \Sigma_i = \begin{pmatrix} 1 & \rho_i \\ \rho_i & 1 \end{pmatrix}.$$

Here, μ_k and σ_k are the mean and standard deviation of the genuine match scores from the kth matcher and ρ_c is the class conditional correlation between S_1 and S_2, $c \in \{g, i\}$. To further simplify our analysis, we assume that the performances of the two matchers are similar, and one way of imposing this constraint is to set $\mu_k = \mu$ and $\sigma_k = \sigma$, for $k = 1, 2$. The recognition performance of this simplified model can be characterized in terms of three parameters, namely, d', ρ_g and ρ_i, where the value of d' (d-prime) statistic is $\sqrt{2}\mu/\sqrt{1+\sigma^2}$. Smaller values of d' indicate large overlap between the genuine and impostor distributions and lower accuracy. On the other hand, larger values of d' indicate better separation between the genuine and impostor distributions and hence, higher accuracy.

Now, we consider the following three dependence scenarios under this simplified model. (i) $\rho_g = \rho$, $\rho_i = 0$, (ii) $\rho_g = \rho_i = \rho$, and (iii) $\rho_g = -\rho_i = \rho$. For each scenario, ρ is varied from -0.95 through 0.95 in steps of 0.05 and d' is varied from 0.5 through 8 in steps of 0.5. For each combination of ρ and d', 10^6 samples are generated from both the genuine and impostor distributions. Parameters μ_g, Σ_g, μ_i, Σ_i are estimated from these training samples under both the full model and the independence model (ρ_g and ρ_i are forcibly set to zero for the independence model). The

log-likelihood ratio of the genuine and impostor training samples are computed based on the estimated parameters. The empirical cdf of the genuine and impostor log-likelihood ratio are then obtained under the full model ($M_g^*(\lambda_1)$ and $M_i^*(\lambda_1)$, respectively) and the independence model ($Q_g^*(\lambda_1)$ and $Q_i^*(\lambda_1)$, respectively). Finally, ΔP_{fi} and ΔP_{tf} are computed using numerical integration.

First, we observed that $0 \leq \Delta P_{tf} \leq 10^{-4}$, for all the above experiments, indicating negligible loss in performance due to density estimation under the full model. This is not surprising because we use a large number of samples (10^6) to estimate simple densities with only 5 parameters (full model), leading to reliable density estimates. Next, the variations of ΔP_{fi} with respect to ρ and d' under the three dependence scenarios are shown in Figure 1. We observe that $\Delta P_{fi} \geq 0$ for all the experiments. In the first scenario, ΔP_{fi} is close to zero for large values of d' and small values of ρ (see Figure 1(a)). The loss in power due to the independence assumption is large only when the absolute value of correlation is very high ($|\rho| > 0.5$) and the accuracies of the individual matchers are low ($d' < 2$). The value of ΔP_{fi} is close to zero for all values of d' and ρ in the second scenario (see Figure 1(b)). The loss in power is the largest in the third scenario (see Figure 1(c)), where the genuine and impostor scores exhibit diametrically opposite correlation behavior.

These results indicate that when the accuracy of a biometric matcher is high ($d' > 4$, which roughly corresponds to $< 5\%$ equal error rate), the information loss due to the assumption of independence between matchers is low. In other words, correlation between matchers does not play a significant role in determining the fusion performance when the matchers have similar performance and are highly accurate. Contrarily, when the d' metric is very small (less than 2) and the difference between ρ_g and ρ_i is large, the loss in power is significant. The reason for this phenomenon could be that when the individual matchers are not very accurate, correlation between the two match scores contains more information about the class label than the individual match scores. Therefore, we believe that the difference between ρ_g and ρ_i is also one of the factors that affects the amount of discriminatory information lost due to the assumption of independence between the matchers. We also observe that the loss in power due to the independence assumption increases as the correlation between the two matchers increases.

IV. EXPERIMENTAL RESULTS

While the simple Gaussian example highlighted certain scenarios where the assumption of independence is likely to result in loss in fusion performance compared to the full model, we need to verify whether these broad trends hold good even in the case of real biometric data. For this purpose, we consider scores from three different databases, namely NIST-Face, XM2VTS-Benchmark and FVC2002-DB1 databases. The same density estimation and evaluation procedures described in [2] have been employed here.

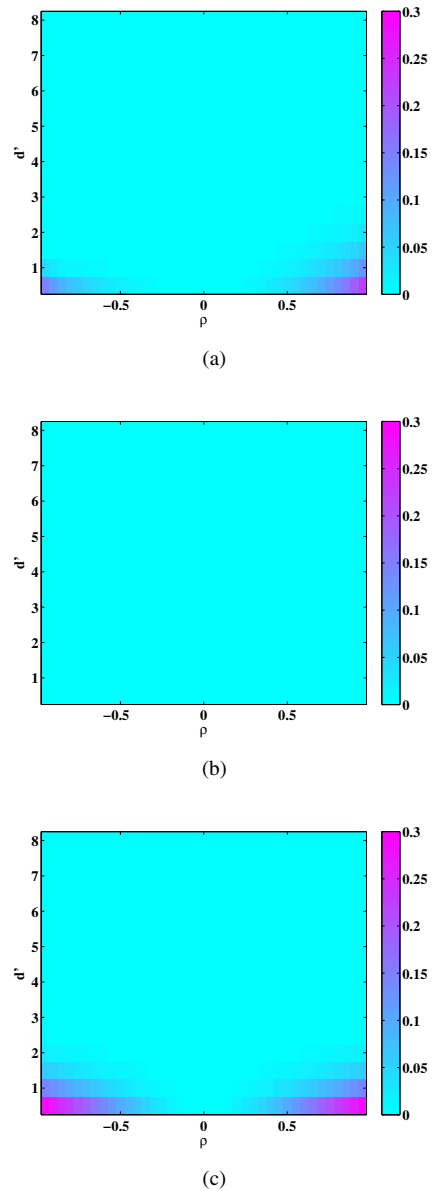The NIST-Face database is one of the partitions in the NIST Biometric Score Set Release 1 (BSSR1) [18]. This

(a)

(b)

(c)

Fig. 1. Variation of difference in power (ΔP_{fi}) between the full and independence models with respect to strength of correlation (ρ) and the accuracy of the individual matchers (d' metric) when the match scores follow a bivariate Gaussian distribution. (a) $\rho_g = \rho$, $\rho_i = 0$, (b) $\rho_g = \rho_i = \rho$, and (c) $\rho_g = -\rho_i = \rho$.

database consists of scores from two face matchers applied on three frontal face images from $3,000$ individuals. The reason for using the NIST-Face database is that we wanted to explicitly consider a situation where the match scores are highly correlated. Pearson's correlation coefficient for the genuine and impostor scores of the two face matchers are 0.7 and 0.3, respectively. Due to failures during feature extraction, the first matcher in the NIST-Face database may sometimes output a score of -1. Since this discrete score affects the correlation estimates significantly, we do not use those samples with a score of -1 in all our experiments.

The XM2VTS-Benchmark database [19] consists of five

face matchers and three speech matchers. In this database, we consider only the fusion of the two speech matchers, namely, LFCC-GMM and SSC-GMM, because this matcher pair had the highest correlation value among the different matcher pairs (0.8 for genuine and 0.7 for the impostor class).

The FVC2002-DB1 [20] is used to examine a scenario where match scores from multiple fingerprint impressions are fused. This database consists of 100 fingers with 8 impressions per finger. Among the 8 impressions, we treat the first two impressions as templates and the remaining six as queries. Each query is matched to both the templates using a minutiae matcher and the two resulting match scores are fused. In this scenario, the Pearson's correlation coefficient for the genuine and impostor scores corresponding to the two templates are 0.5 and 0.45, respectively.

The performances of the LR fusion scheme under the full and independence models are shown in Figure 2(a), which indicates that there is no significant difference in the performance of the two rules. We believe that this is because the difference between genuine and impostor correlations is not high and the two matchers in this database are reasonably accurate (the d' values for both the matchers in the NIST-Face database is ≈ 3.2). Now, we apply a linear transformation of the form $s'_k = (s_k - a)/b$ to the genuine match scores from the two matchers, where s_k is the original score of the kth matcher and s'_k is the modified score. The values of a and b are chosen such that the d' metric of the transformed scores is ≈ 2. This transformation does not affect the correlation between the genuine scores of the two matchers. We also remove the correlation between impostor scores by randomly permuting the impostor scores from one of the two matchers. Note that this permutation does not change the marginal distribution of the impostor scores. Thus, the d' value for the modified match scores is ≈ 2 and the correlation between them is 0.7 for the genuine class and 0 for the impostor class.

The performance of the LR decision rule based on the full vs. independence model on the modified scores in the NIST-Face database is shown in Figure 2(b). Since the separation between the genuine and impostor distributions was reduced by applying a linear transformation to the genuine scores, the accuracy of the individual matchers and hence the fusion performance is reduced substantially. However, in this case we observe that the LR fusion rule under the full model clearly outperforms the independence model. For example, at a FAR of 0.1%, the average improvement in the GAR is 2.7% and the 95% confidence interval for the difference in the GAR between the two rules is [2.5%, 2.9%]. This result indicates that modeling the correlation between the match scores, and hence the use of the full model is justified only if the both the matchers are of low accuracy and the difference between genuine and impostor correlation is large.

Similar results were also obtained in the case of correlated matcher pairs in the XM2VTS-Benchmark database. Figure 3(a) shows the ROC curves for the fusion of LFCC-GMM and SSC-GMM speech matchers in the XM2VTS database. The d' values for the LFCC-GMM and SSC-GMM matchers
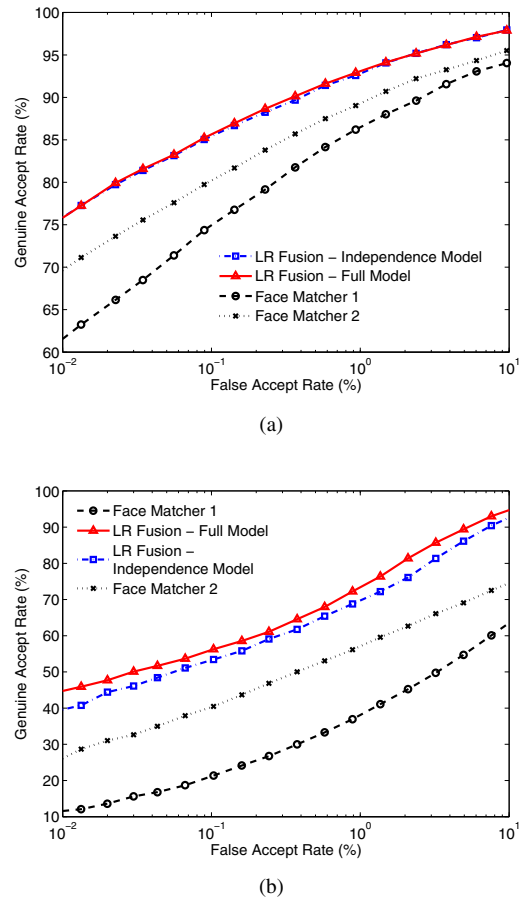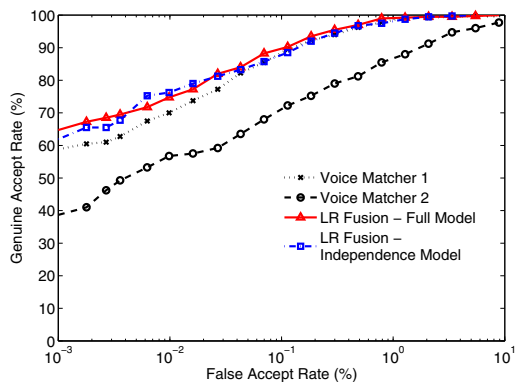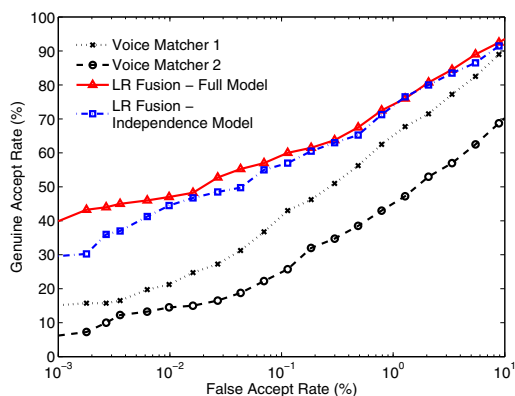

(a)

(b)

Fig. 2. Performance of likelihood ratio based fusion rules under the full and independence models for the two face matchers in the NIST-Face database. (a) Fusion of original match scores and (b) Fusion of match scores transformed such that the impostor correlation is zero and the d' metric after transformation is approximately 2 for both the matchers.

are ≈ 4 and 3, respectively. From Figure 3(a), we observe that the decision rules based on full and independence models perform equally well on this pair of matchers. However, if the d' values of the two matchers are reduced by applying a linear transformation to the genuine scores and if the impostor correlation is removed, we observe that the decision rule based on the full model provides better performance than the rule based on the independence model (see Figure 3(b)).

Finally the performance of LR fusion on the match scores from the FVC2002-DB1 database is shown in Figure 4. Even in this multiple impression scenario, we do not observe any significant difference in the performance of the full and independence models. Furthermore, we did not observe any significant difference between the performance of the full vs independence model even after transforming the scores based on the same procedure used in the other two databases. This may be because the correlation between the genuine scores corresponding to the two fingerprint templates is not as high as in the NIST-Face and XM2VTS databases.

(a)

(b)

Fig. 3. Performance of likelihood ratio based fusion rules under the full and independence models for the LFCC-GMM and SSC-GMM voice matchers in the XM2VTS database. (a) Fusion of original match scores and (b) Fusion of match scores transformed such that the impostor correlation is zero and the d' metric after transformation is approximately 2 for both the matchers.

V. CONCLUSIONS

We have studied the effect of modeling the correlation between matchers on the fusion performance of multibiometric systems. A mathematical framework based on the likelihood ratio-based fusion scheme has been presented. We also analyzed the difference between the fusion performance of the full model (directly estimating the multi-dimensional joint density) and the independence model (joint density is approximated using the product of the marginals) in a simple bivariate Gaussian example. This example indicates that the assumption of independence between matchers to be fused does not adversely affect the performance of the LR fusion scheme, especially when the individuals matchers are accurate and the difference between genuine and impostor correlation is not high. These trends have also been verified using real biometric data. Given that biometric matchers are typically quite accurate (EER is less than 5%) and the dependence structure between biometric matchers is usually similar for both genuine and impostor samples, LR fusion under the independence model is adequate to obtain good performance, especially for small sample size scenarios.
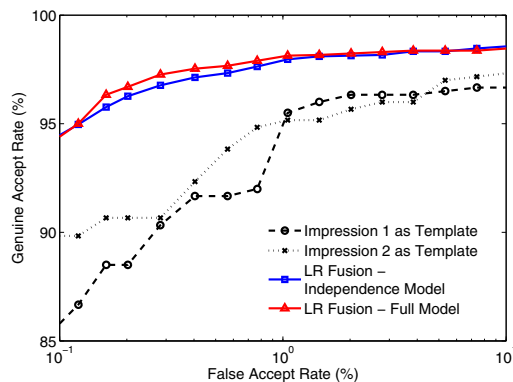

Fig. 4. Performance of likelihood ratio based fusion rules under the full and independence models for the FVC2002-DB1 database.

REFERENCES

[1] A. Ross, K. Nandakumar, and A. K. Jain, *Handbook of Multibiometrics*. Springer, 2006.

[2] K. Nandakumar, Y. Chen, S. C. Dass, and A. K. Jain, "Likelihood Ratio Based Biometric Score Fusion," *IEEE Trans. on PAMI*, vol. 30, no. 2, pp. 342–347, February 2008.

[3] L. I. Kuncheva, C. J. Whitaker, C. A. Shipp, and R. P. W. Duin, "Is Independence Good for Combining Classifiers?" in *Proc. of ICPR*, vol. 2, Barcelona, Spain, 2000, pp. 168–171.

[4] P. Ivatury, "A Study of Diversity Metrics For Biometric Fusion," West Virigina University, Tech. Rep., 2005.

[5] S. Prabhakar and A. K. Jain, "Decision-level Fusion in Fingerprint Verification," *Pattern Recognition*, vol. 35, no. 4, pp. 861–874, 2002.

[6] O. Koval, S. Voloshynovskiy, and T. Pun, "Analysis of Multimodal Binary Detection systems Based on Dependent/Independent Modalities," in *Proc. of Workshop on MMSP*, Crete, Greece, 2007, pp. 70–73.

[7] K. Kryszczuk and A. Drygajlo, "Impact of Feature Correlations on Separation between Bivariate Normal Distributions," in *Proc. of ICPR*, Tampa, USA, December 2008, pp. 1–4.

[8] O. Ushmaev and S. Novikov, "Biometric Fusion: Robust Approach," in *Proc. of MMUA*, Toulouse, France, 2006.

[9] S. G. Iyengar, P. K. Varshney, and T. Damarla, "A Parametric Copula-based Framework for Multimodal Signal Processing," in *Proc. of ICASSP*, Taiwan, 2009, pp. 1893–1896.

[10] S. C. Dass, K. Nandakumar, and A. K. Jain, "A Principled Approach to Score Level Fusion in Multimodal Biometric Systems," in *Proc. of AVBPA*, Rye Brook, USA, July 2005, pp. 1049–1058.

[11] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*. Springer, 2005.

[12] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, 1986.

[13] M. Figueiredo and A. K. Jain, "Unsupervised Learning of Finite Mixture Models," *IEEE Trans. on PAMI*, vol. 24, no. 3, pp. 381–396, March 2002.

[14] L. Devroye, L. Gyrfi, and G. Lugosi, *A Probabilistic Theory of Pattern Recognition*. New York: Springer-Verlag, 1996.

[15] M. P. Wand and M. C. Jones, *Kernel Smoothing*. Chapman & Hall, CRC Press, 1995.

[16] S. Eguchi and J. Copas, "Interpreting Kullback-Leibler Divergence with the Neyman-Pearson Lemma," *Journal of Multivariate Analysis*, vol. 97, no. 9, pp. 2034–2040, 2006.

[17] Q. H. Vuong, "Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses," *Econometrica*, vol. 57, no. 2, pp. 307–333, 1989.

[18] National Institute of Standards and Technology, "NIST Biometric Scores Set," 2004, available at http://http://www.itl.nist.gov/iad/894.03/biometricscores.

[19] 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.

[20] D. Maio, D. Maltoni, J. L. Wayman, and A. K. Jain, "FVC2002: Second Fingerprint Verification Competition," in *Proc. of ICPR*, Quebec City, Canada, August 2002, pp. 811–814.