Pattern Recognition 42 (2009) 823–836

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Threshold-optimized decision-level fusion and its application to biometrics

Qian Tao*, Raymond Veldhuis

Signals and Systems Group, University of Twente, P.O. Box 217, 7500AE Enschede, The Netherlands

ARTICLE INFO

Article history:
Received 4 December 2007
Received in revised form 11 September 2008
Accepted 17 September 2008

Keywords:
Fusion
Matching score level
Decision level
Threshold-optimized decision-level fusion

ABSTRACT

Fusion is a popular practice to increase the reliability of biometric verification. In this paper, we propose an optimal fusion scheme at decision level by the AND or OR rule, based on optimizing matching score thresholds. The proposed fusion scheme will always give an improvement in the Neyman–Pearson sense over the component classifiers that are fused. The theory of the threshold-optimized decision-level fusion is presented, and the applications are discussed. Fusion experiments are done on the FRGC database which contains 2D texture data and 3D shape data. The proposed decision fusion improves the system performance, in a way comparable to or better than the conventional score-level fusion. It is noteworthy that in practice, the threshold-optimized decision-level fusion by the OR rule is especially useful in presence of outliers.

1. Introduction

Fusion is a popular practice to increase the reliability of the biometric verification by combining the information of multiple classifiers [1–3]. Combining information from different biometrics, such as face, fingerprint, palmprint, iris, etc., has been a trend in the biometrics society [4–8]. Generally speaking, fusion can be done at four different levels: sensor level, feature level, matching score level, and decision level [1,3], as illustrated in Fig. 1.

Fusion at sensor level is closely related to the specific sensor types and the corresponding signal/image processing methods. For a more compact review, we will concentrate on the last three levels, which are closely related to a classifier. At feature level, for each classifier, the feature vector is in a high dimensional space: $x_i \in \mathbb{R}^{m_i}, m_i \geqslant 1$, $i = 1, 2, \ldots, N$. Note that the dimensionalities m_i and m_j could be different for $i \neq j$. At matching score level, the feature vector is reduced to a scalar value, $s_i \in \mathbb{R}$, $i = 1, 2, \ldots, N$. At decision level, the matching scores s_i are compared to the thresholds T_i, and the outputs are binary decisions $d_i \in \{1, 0\}$, $i = 1, 2, \ldots, N$.

Fusion at the feature level combines the features before applying the respective classifiers of different biometrics [8–12]. Feature-level fusion has certain difficulties. Firstly, the feature sets of different modalities can be incompatible, for example some feature values might be locations (e.g. of the minutiae set of the fingerprint) while some might be gray values (e.g. of the face images), which makes it infeasible to combine them on the same ground. Secondly, even if a combination rule could be designed, the size of the resulting feature vector will often increase. This, in turn, increases the complexity of the system, making it more difficult to design and train the classifier.

Fusion at matching score level is the most popular way of fusion, offering the best tradeoff between information content and ease of fusion [1], and has been extensively studied in literature [13–18]. There are basically three categories of fusion schemes at the matching score level. The first category of fusion scheme is *transformation-based*. Firstly, all the component matching scores are transformed/normalized so that they are on a comparable scale. Then simple scalar functions are applied on the transformed matching scores, resulting in a new matching score. Examples of the functions are product, sum, mean, max, etc. [13,19,20]. It can be proved that under certain ideal situations, for example, taking the product of independent likelihood ratios can achieve the statistically optimal performance in the Neyman–Pearson sense [21]. The second category of fusion scheme is *density-based*. It relies on the estimation of the joint densities of the matching scores, and the fusion is done by statistical tests, like the likelihood ratio test. Examples of this category of work include Refs. [2,22–26][1] according to the estimated score distributions. This category of fusion scheme achieves optimal performance if the densities could be accurately learnt, under the situation when a large number of representative training matching scores are available. The third category of fusion scheme is *classifier-based*. It

* Corresponding author. Tel.: +31 53 289 2897.
E-mail addresses: q.tao@utwente.nl (Q. Tao), r.n.j.velhuis@utwente.nl (R. Veldhuis).

[1] Note in Ref. [26], the method is called "decision-level", nevertheless the fusion is at the score level according to our definition in Fig. 1, while the "decision" is about the pre-selection of a combination of the component classifiers.

0031-3203/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2008.09.036

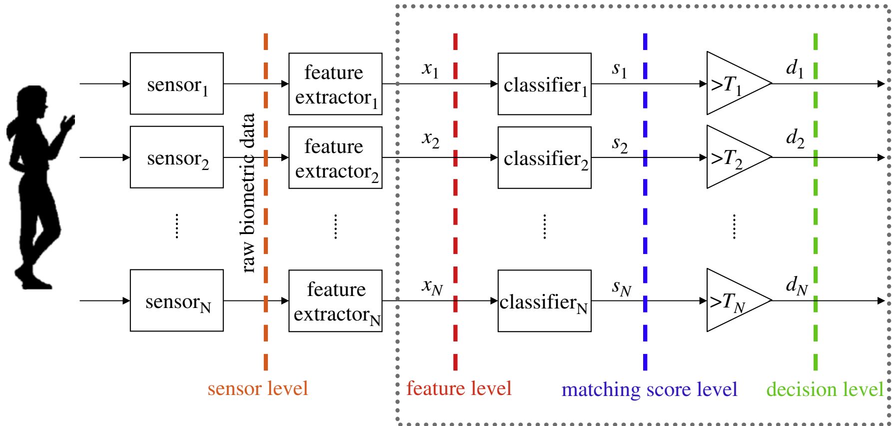

Fig. 1. Different levels of fusion: sensor level, feature level, matching score level, and decision level.

concatenates the component matching scores as a new feature vector, and trains additional classifiers on them [7,14–16,27]. Examples are neural networks, support vector machines, decision trees, etc. This type of fusion needs to train the relevant classification parameters.

Matching score normalization is necessary for the first type of matching score-level fusion, especially when the fusion is done between different classifiers or modalities, with the output matching scores defined in their own different ways. Such normalization is also important for the remaining two fusion schemes, as it can essentially affect the matching score densities.

Fusion at decision level is less studied in literature, as it is often considered inferior to matching score-level fusion, on the basis that decisions are too "hard" and have less information content compared to "soft" matching scores. One example of fusion on decision level is the majority vote [13,28], which counts the number of decisions d from the component classifiers, and chooses the majority of the decision as the final decision. Its derivative, weighted majority voting [29], assigns different weights according to different performances of the component classifiers. This consequently transforms the output value from logical numbers to continuous numbers. More of such examples are the Bayesian decision fusion [30], the Dempster–Shafer theory of evidence [31], which also convert the decisions into scores, with the converting parameters learned from a training score set.

"Soft" measures with information of the confidence level have always been preferred in fusion. It has been shown by Daugman [32] that the combination of two matchers using AND or OR rule might actually degrade the overall performance when the performances of component classifiers are significantly different. Due to this phenomenon, AND and OR rules are rarely recommended in practice [1]. In this paper, however, we propose a decision-level fusion scheme, by the AND and OR rule, in an optimal way such that it always gives an improvement in terms of error rates over the classifiers that are fused. Here, optimal is in the Neyman–Pearson sense [21]: at a given false-acceptance rate (FAR) α, the decision-fused classifier has a false-rejection rate (FRR) β that is minimal, and never larger than the FRR of the classifiers that are fused at the same α; or at a given β, the decision-fused classifier reaches a minimal α.

Some scenarios exist in which the proposed decision-level fusion is preferable to score-level fusion. For example, in template-protected biometrics, an accept or reject decision is based on the equality between a binary string extracted from the biometric data and a reference binary string [33]. This means the matching score is not available for fusion. If such a system is based on a fuzzy commitment [34] or a fuzzy vault scheme [35], the error correction is a part of the extraction of the binary string. The error correction can, within limits, be tuned to correct a certain maximum number of errors. This determines the point of operation on the receiver operating characteristic (ROC), and is equivalent to tuning a matching score threshold. Therefore, the proposed optimized decision-level fusion can be used to fuse two template protected biometric systems, and the optimal fusion can be achieved by tuning the number or corrected errors. Another scenario is that when the outliers are present in the biometric data. In that case, as we will discuss in Section 3, the proposed OR rule fusion often outperforms the conventional score-level fusion methods.

This paper is organized as follows. Section 2 presents the threshold-optimized decision-level theory on statistically independent and dependent classifiers. Section 3 discusses the application of the proposed method in the outlier scenario. Section 4 presents the experimental results of the fusion between two face modalities and different algorithms. Section 5 gives the conclusion.

2. Threshold-optimized decision-level fusion theory

A decision can be denoted by a logical number $d \in \{1, 0\}$, where 1 is for "accept" and 0 for "reject". From a classifier point of view, any decision d_i is obtained by comparing the matching scores s_i with a certain threshold T_i (see Fig. 1). In the proposed decision-level fusion with optimized thresholds, we do not pre-fix the thresholds T_i of the individual component classifiers as is common in conventional decision-level fusion [32], instead, we optimize the combination of these thresholds, according to their joint behavior in the AND or OR rule fusion.

Before discussing the optimization process in detail, let us first look at the characterization of individual classifiers. Each decision d of a classifier is characterized by two error probabilities: the first is the probability of a false acceptance, the FAR, α, and the second is the probability of a false rejection, FRR, β. Obviously, FAR and FRR are both functions of T. When T varies, the FRR can be seen as a function of the FAR, $\beta(\alpha)$, known as the detection error trade-off (DET) characteristic [36]. DET is an indication of classification performance, revealing the inherent separability of the two opposite classes. An equivalent measure is the ROC, in which the detection rate $p_{\mathrm{d}} = 1 - \beta$ is expressed as a function of α, $p_{\mathrm{d}}(lpha)$ [37]. We will use ROC for illustration throughout this paper. However, as we deal with the OR rule most of the time, it is more convenient to use $p_{\mathrm{r}}(\beta)$ ($p_{\mathrm{r}} = 1 - \alpha$ is the correct-rejection rate), a classifier's *rejection characteristic* equivalent to the ROC, in all the mathematical derivations.

Depending on statistical properties of the component decisions, two different situations are identified. First, the multiple decisions d_i, $i = 1, 2, \ldots, N$, are statistically independent. This is desirable in fusion, as it has been observed that fusion works better when the fused components are independent [13,28] or negatively dependent [38]. This situation occurs in many multi-modal biometric fusion cases, and facilitates a fast training based on ROC, as will be shown in Section 2.2. Second, the multiple decisions $d_i, i=1,2,\ldots,N$, possess some dependencies. Threshold-optimized decision-level fusion can also be solved for dependent decisions in a non-parametric manner, but the training is much slower and the optimized thresholds are more sensitive to the training set. Actually, the ROC-based training for independent decisions suffices for most fusion applications, even when some dependency exists. This is analog to the naive Bayes classifier [28], which also assumes independency between different features, but whose good performance in dependency cases has been acknowledged in a wide range of applications [39,40]. In the following, we will concentrate on the solution for the independent decisions, while the solution for the dependent decisions will be addressed in Appendix B.

In all the following derivations, we will mainly focus on the OR rule as it is of more interest in practice. The AND rule fusion can be derived in a similar way.

2.1. Problem definition

Suppose we have N statistically independent decisions d_i, $i = 1,2,\ldots,N$. To analyze the OR rule we have to work with the false-rejection rates, FRR β and the correct-rejection rate p_{r}. After application of the OR rule to decisions $d_i, i=1,\ldots,N$, we have, under the assumption that all decisions are statistically independent, that

$$\beta = \prod_{i=1}^{N} \beta_i, \quad p_{\mathrm{r}}(\beta) = \prod_{i=1}^{N} p_{\mathrm{r},i}(\beta_i) \tag{1}$$

with β the FRR and p_{r} the correct-rejection rate of the final fused decision, respectively. The optimized OR rule decision fusion can then be formally defined by finding

$$\hat{p}_{\mathrm{r}}(\beta) = \max_{\beta_i | \prod \beta_i = \beta} \prod_{i=1}^{N} p_{\mathrm{r},i}(\beta_i) \tag{2}$$

where $\hat{p}_{\mathrm{r}}$ is the maximal correct-rejection rate at β. In other words, the β_i's of the component classifiers are tuned during this optimization, so that the fused classifier can give maximal p_{r} at a fixed $\beta = \prod_{i=1}^{N} \beta_i$.

Likewise, the optimized AND rule decision fusion can also be formulated, based on the false accept rate α and detection rate p_{d}:

$$\hat{p}_{\mathrm{d}}(\alpha) = \max_{\alpha_i | \prod \alpha_i = \alpha} \prod_{i=1}^{N} p_{\mathrm{d},i}(\alpha_i) \tag{3}$$

It is easily proved that the optimized correct-rejection rate $\hat{p}_{\mathrm{r}}(\beta)$ is never smaller than any of the $p_{\mathrm{r},i}$'s at the same β:

$$\hat{p}_{\mathrm{r}}(\beta) \geqslant p_{\mathrm{r},i}(\beta), \quad i = 1,\ldots,N \tag{4}$$

Because, by definition

$$\hat{p}_{\mathrm{r}}(\beta) = \max_{\beta_i | \prod \beta_i = \beta} \prod_{i=1}^{N} p_{\mathrm{r},i}(\beta_i) \geqslant \prod_{j=1}^{N} p_{\mathrm{r},j}(\beta_j) \Big|_{\prod_{i=1}^{N} \beta_i = \beta} \tag{5}$$

As it holds for any classifier that $p_{\mathrm{r},i}(1) = 1$, Eq. (4) readily follows by setting $\beta_j = \beta$ and $\beta_i = 1$ for all $i \neq j$. Similarly, it can be proved for the AND rule that $\hat{p}_{\mathrm{d}}(\alpha) \geqslant p_{\mathrm{d},i}(\alpha)$, for $i = 1,\ldots,N$.

By solving the optimization problem in Eqs. (2) and (3), the optimal operation points for every component classifiers are obtained.

2.2. Problem solution

In the work of Zhang et al. [41], a similar optimization problem as in Eq. (3) is reformulated in a logarithmic domain. Under the assumption that $\log(p_{\mathrm{r},i}(\beta_i))$ is a concave function of $\log(\beta_i)$, it is proposed to find the optimal operation points by solving the unconstrained Lagrange optimization problem:

$$\begin{aligned}\max\{\log p_{\mathrm{r}} - \lambda \log \beta\} &= \max\left\{\sum_{i=1}^{N} \log(p_{\mathrm{r},i}(\beta_i)) - \lambda\left(\sum_{i=1}^{N} \log(\beta_i)\right)\right\} \\ &= \sum_{i=1}^{N} \max\{\log(p_{\mathrm{r},i}(\beta_i)) - \lambda \log(\beta_i)\}\end{aligned} \tag{6}$$

Due to the log-concavity assumption of each individual ROC, this optimization can be done by maximizing the value of $\log(p_{\mathrm{r},i}(\beta_i)) - \lambda \log(\beta_i)$ for each ROC individually, and thus avoiding exhaustive search. For more details, see Refs. [41,42]. One drawback of this method is that it does introduce a possibly too restrictive assumption on the ROC. The concavity of $p_{\mathrm{r}}\beta$ and $p_{\mathrm{d}}(\alpha)$ always holds in the original domain, but it does not always apply in the logarithmic domain. To avoid this drawback, we present an alternative approach, without any additional assumption or approximation. We propose that the optimization problem (2) and (3) be solved in a recursive manner: first fuse two arbitrary classifiers from the set of component classifiers, compute the ROC of the fused classifier, and then fuse the resulting ROC with the next arbitrary component ROC, and so on. The proof of optimality is put in Appendix A. This means that every time we only have to fuse two classifiers, thus avoiding the exponential explosion in computational complexity in combining multiple classifiers. We summarize the solution in the following:

(1) Given N component classifiers, each characterized by $p_{\mathrm{d},i}(\alpha_i)$ or $p_{\mathrm{r},i}(\beta_i)$, $i = 1,\ldots,N$. Each operation point corresponds to a threshold.
(2) Take any two ROCs and do threshold-optimized decision fusion.
(3) Replace the two ROCs with the optimally fused ROC. Note that for a single operation point on the already fused ROC, there are now multiple thresholds coming from the component classifiers.
(4) Repeat steps (2) and (3) until all the classifiers have been combined.
(5) A final ROC $p_{\mathrm{d}}(\alpha)$ or $p_{\mathrm{r}}(\beta)$ is obtained, with each operation point corresponding to N thresholds from the N component classifiers.

The only problem left now is the fusion of two ROCs in step (2). In real situations, $\hat{p}_{\mathrm{d}}(\alpha)$ is not available in its analytical form, but instead characterized by a set of discrete operation points. Therefore, we solve the fusion of two ROCs in a brute-force manner. Suppose we have two ROCs, denoted by N_1 and N_2 discrete operation points, respectively: $\mathrm{ROC}_1 = \{(\beta_1^i, p_{\mathrm{r},1}^i)\}$, $\mathrm{ROC}_2 = \{(\beta_2^j, p_{\mathrm{r},2}^j)\}$, where $i=1,2,\ldots,N_1$, $j=1,2,\ldots,N_2$. The fusion of these two classifiers, under the independent assumption, can have in total $N_1 \cdot N_2$ possible combinations after OR rule fusion: $\{(\beta_1^i \beta_2^j, p_{\mathrm{r},1}^i p_{\mathrm{r},2}^j)\}$. The AND rule fusion can be derived similarly by using α and p_{d}. Obviously, each pair of operation points corresponds to a pair of thresholds (T_1, T_2) with T_1 from the first classifier and T_2 from the second classifier. To get the optimized fusion, we select those operation points which form a concave hull of all the possible combinations. Fig. 2 illustrates this optimization process. In this example, we have generated the genuine and impostor scores independently for two classifiers. The genuine scores of the two classifiers has a multivariate Gaussian distribution

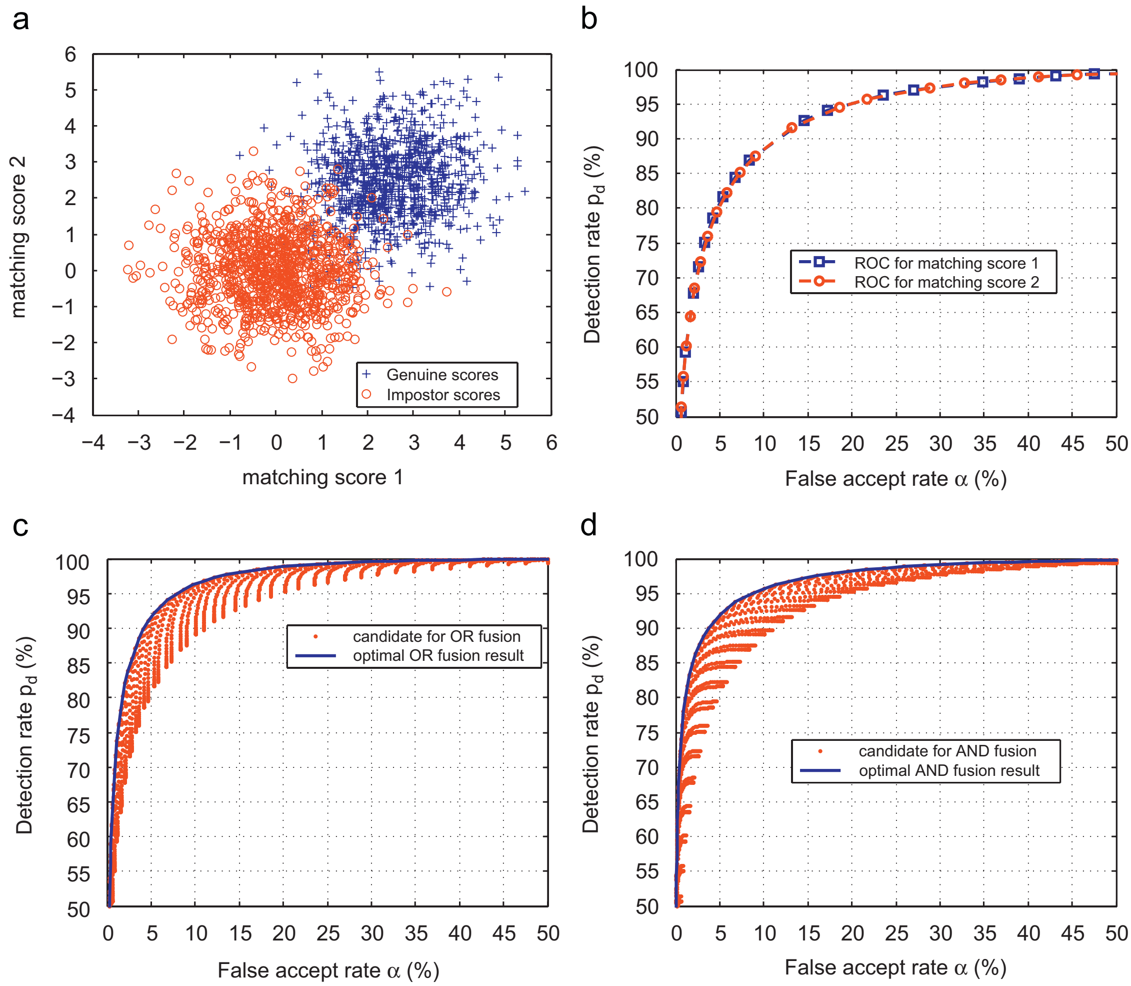

Fig. 2. Threshold-optimized decision fusion in the independent case: (a) the scatter plot of two matching scores, (b) two ROCs of the two matching scores, respectively, (c) all the possible OR fused points and the optimal ROC selected, and (d) all the possible AND fused points and the optimal ROC selected.

of $N\left((2.5, 2.5), \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$, while the impostor scores of the two classifiers have a multivariate Gaussian distribution of $N\left((0, 0), \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$. In Fig. 2(c) and (d), the dots denote all the possible combinations for the OR rule and the AND rule fusion, and the solid line marks the concave hull, which is optimal in the Neyman–Pearson sense. The optimized thresholds for the decision-level fusion are, therefore, obtained as the thresholds corresponding to the selected points of operation. It can be seen that both the OR rule and the AND rule fusion result in a better ROC than the original two ROCs.

Note that the optimality of the solution is only true in independent cases, and the ROCs in Fig. 2(c) and (d) are the estimation of the fused ROCs under the independency assumption. When the matching scores have some dependencies, as we will show in Appendix B, the ROC improvement is smaller compared to that of the independent case.

2.3. Additional remarks

The ROC is useful indirect indication of the score distributions. A highlight of the proposed decision-level fusion method is that it works on the operation points on the ROC, instead of on the matching scores as many other conventional fusion methods do. In practice, the number of the training matching scores could be enormous, but after calculating the ROC from the set, the number of ROC operation points is usually much smaller. On the other hand, when the number of the training matching scores is very small, the ROC points can even be interpolated and smoothed to produce a robust estimation. This simplifies the problem by converting any number of training scores to a manageable number of operation points on ROC. The optimization of the proposed decision-level fusion is again very simple. This makes the algorithm very efficient with training data sets of any size.

The computation involved in the training stage is the estimation of the ROC and the selection of the optimal ROC points. Given the training score set, it is very easy to calculate the ROC by comparing the scores with a number of thresholds, and estimate the FAR and FRR. The optimization, as in Eqs. (2) and (3), is achieved simply by taking the outer boundary points in the α–p_d plane. In the verification stage, the calculation is extremely fast: for N classifiers, only N comparisons and $N-1$ AND or OR operations are required. Both the training and the fusion are simpler compared with advanced score-level fusion methods, such as support vector machines or likelihood ratio methods based on Gaussian mixtures.

Score normalization is important in matching score-level fusion [1,5]. From the Neyman–Pearson point of view, it is most desirable that the matching score $s(x)$ be normalized in such a way that it is equal, or proportional to, the likelihood ratio of the feature vector x: $F(s(x)) = (p(x|\omega_{\text{gen}}))/(p(x|\omega_{\text{imp}}))$, where $F(\cdot)$ is a monotonic normalization function. Different normalization functions result in different decision boundaries in matching score-level fusion. In comparison, an advantage of threshold-optimized decision-level fusion is that the optimization is invariant to any monotonic transformation of the original matching scores. A monotonic function changes the absolute value of the matching scores, but does not alter the relative relationship between the matching scores. The operation points on the ROC, therefore, cannot be changed. As a result, the optimized operation points are invariant to any monotonic normalization. This implies that the final performance remains identical for any kind of score normalization function $F(\cdot)$.

There is always certain discrepancy between training and testing scores, which is one of the causes of overtraining. In many score-level fusion methods, such as the likelihood ratio method, SVM, or NN, there are a number of parameters to be estimated from the training data. The more parameters needed for characterization, the more flexible the boundary is in the score space, and the more sensitive it is to overtraining. In our decision-level method, we expect that, due to the coarser partitioning of the score space, the proposed fusion is more robust to model deviations between the training and testing data. This will be supported by results of the fusion experiments in Section 5.

3. OR fusion in presence of outliers

In this section, we will discuss the situation when the proposed OR rule decision-level fusion is especially favorable. Outliers, in biometric verification, refer to the biometric data which belong to the genuine user, but deviate from the genuine user distribution. Taking face for example, outliers can be caused by extraordinary expressions, poses, illuminations, or mis-registrations. Some examples are given in Fig. 3. Outliers cause false rejections most of the time.

Suppose the outlier scores have a probability density function of $\Psi_{\text{out}}(s)$. This function could be approximated by the impostor distribution $\Psi_{\text{imp}}(s)$, based on the fact the outlier scores have values that could otherwise be taken as impostors. Suppose the genuine score has a probability density function of $\Psi_{\text{gen}}(s)$, and the prior probability of outliers occurring in the genuine score is p_{o}. Taking into account the outliers, the probability of the genuine score s is

$$\Psi'_{\text{gen}}(s) = (1 - p_{\text{o}}) \cdot \Psi_{\text{gen}}(s) + p_{\text{o}} \cdot \Psi_{\text{imp}}(s) \tag{7}$$

Suppose we are fusing two independent classifiers, both with outliers in the genuine score. The joint probability of two independent samples s_1 and s_2 is

$$\begin{aligned}\Psi(s_1, s_2) = &(1 - p_{\text{o},1})(1 - p_{\text{o},2}) \cdot \Psi_{\text{gen},1}(s_1)\Psi_{\text{gen},2}(s_2)\\ &+ p_{\text{o},1}(1 - p_{\text{o},2}) \cdot \Psi_{\text{gen},1}(s_1)\Psi_{\text{imp},2}(s_2)\\ &+ (1 - p_{\text{o},1})p_{\text{o},2} \cdot \Psi_{\text{imp},1}(s_1)\Psi_{\text{gen},2}(s_2)\\ &+ p_{\text{o},1}p_{\text{o},2} \cdot \Psi_{\text{imp},1}(s_1)\Psi_{\text{imp},2}(s_2)\end{aligned} \tag{8}$$

where the subscripts 1 and 2 indicate the first and the second classifier, respectively.

For example in Fig. 4, for the first classifier, $p_{\text{o},1} = 0.03$, $\Psi_{\text{gen},1}(s_1) \sim N(1.5, 1)$, $\Psi_{\text{imp},1}(s_1) \sim N(-1.5, 1)$, while for the second classifier, $p_{\text{o},2} = 0.10$, $\Psi_{\text{gen},2}(s_2) \sim N(2, 1)$, $\Psi_{\text{imp},2}(s_2) \sim N(-2, 1)$. Fig. 4(a)–(c) shows the boundaries of AND rule decision fusion, OR rule decision fusion, sum rule matching score fusion, respectively, at the fixed FAR $\alpha = 0.01$. Fig. 4(d) compares the resulting ROC by different fusion schemes. Under the given situations with outliers, OR rule decision fusion achieves the best performance in a large range, for $\alpha > 0.005$. The AND rule fusion, in comparison, is not suitable for the given score distributions as it only results in the better of the two ROCs.

It is interesting to notice that in Eq. (8), the OR rule boundary accepts all the terms except the last one, which is negligible because of the small value of $p_{\text{o},1}p_{\text{o},2}$. This explains why OR rule decision fusion is suitable for this kind of problem.

The type of matching score distribution as simulated in Fig. 4 is not a rare scenario. It is very often the case that a number of outliers occur in the genuine class, thus making the genuine distribution extend to the impostor class. The impostor class, however, is less likely to produce such a comparable proportion of "outliers". Such phenomenon can be explained by the great discriminating power of a high-dimensional space [43], which makes a classifier in it more ready to reject than to accept. Realistic examples will follow in the next section.

Other fusion methods could also be applied to the fusion problem with outliers, such as density-based fusion, e.g. likelihood ratio test, or classifier-based fusion, e.g. SVM, NN, which also takes care of the outliers during training. However, the resulting decision boundary is more dependent on the training data. To accommodate the outliers, for example, the outliers should be included in the training set. In comparison, the OR rule fusion always has good tolerance with outliers no matter if they are included in the training set or not. The advantage of the proposed OR rule decision fusion, moreover, is its simplicity. First, a normalization step is not required; second, the calculation is faster, as only a limited number of operation points are involved in the calculation. Third, there is potentially less

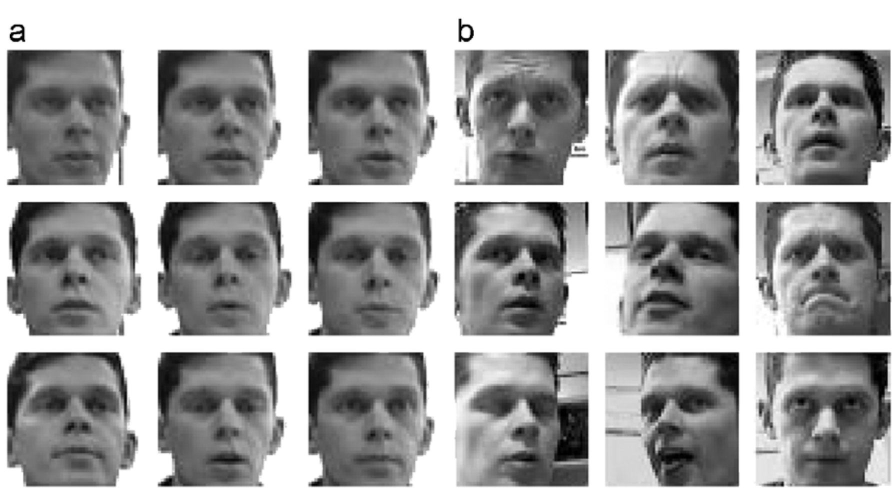

Fig. 3. (a) Normal samples of the user data and (b) outliers in the user data.

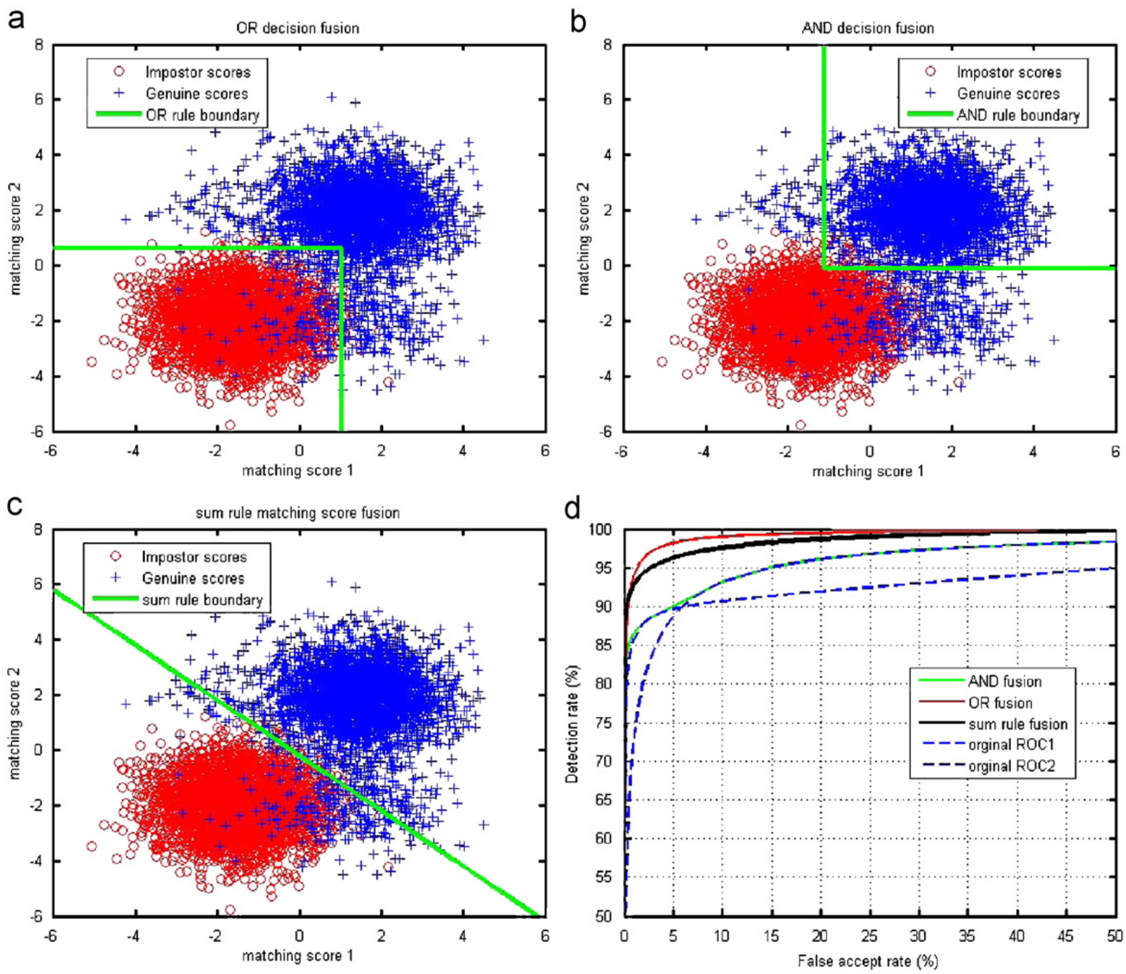

Fig. 4. (a) Scatter plot of the scores and OR rule boundary, (b) scatter plot of the scores and AND rule boundary, (c) scatter plot and sum rule boundary, and (d) comparison of the ROCs.

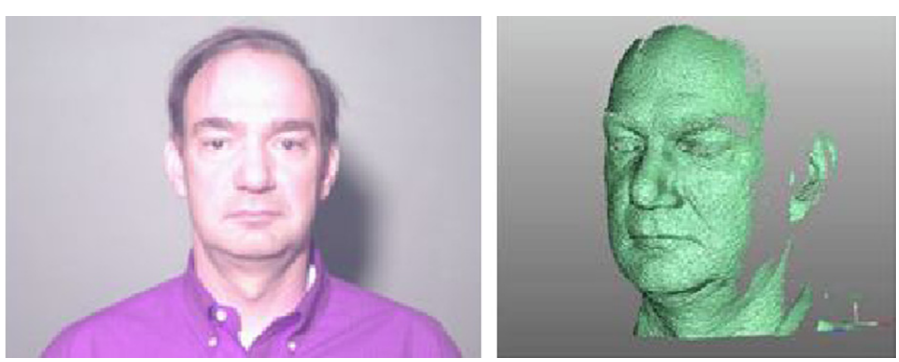

Fig. 5. Example from the FRGC database: the 2D texture and the 3D shape recorded simultaneously recorded.

overtraining possibilities, as the decision boundaries is much simpler than those of the SVM or NN.

4. Experiments and results

The larger context of this work is the EU FP6 3D-face project [44], which aims to use 3D facial shape data in combination with 2D texture data for reliable passport identification. The first database that the algorithms were developed on is the FRGC database [45], which contains the 2D face texture and 3D face shape data collected simultaneously. An example of the two modalities is shown in Fig. 5. 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. 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.

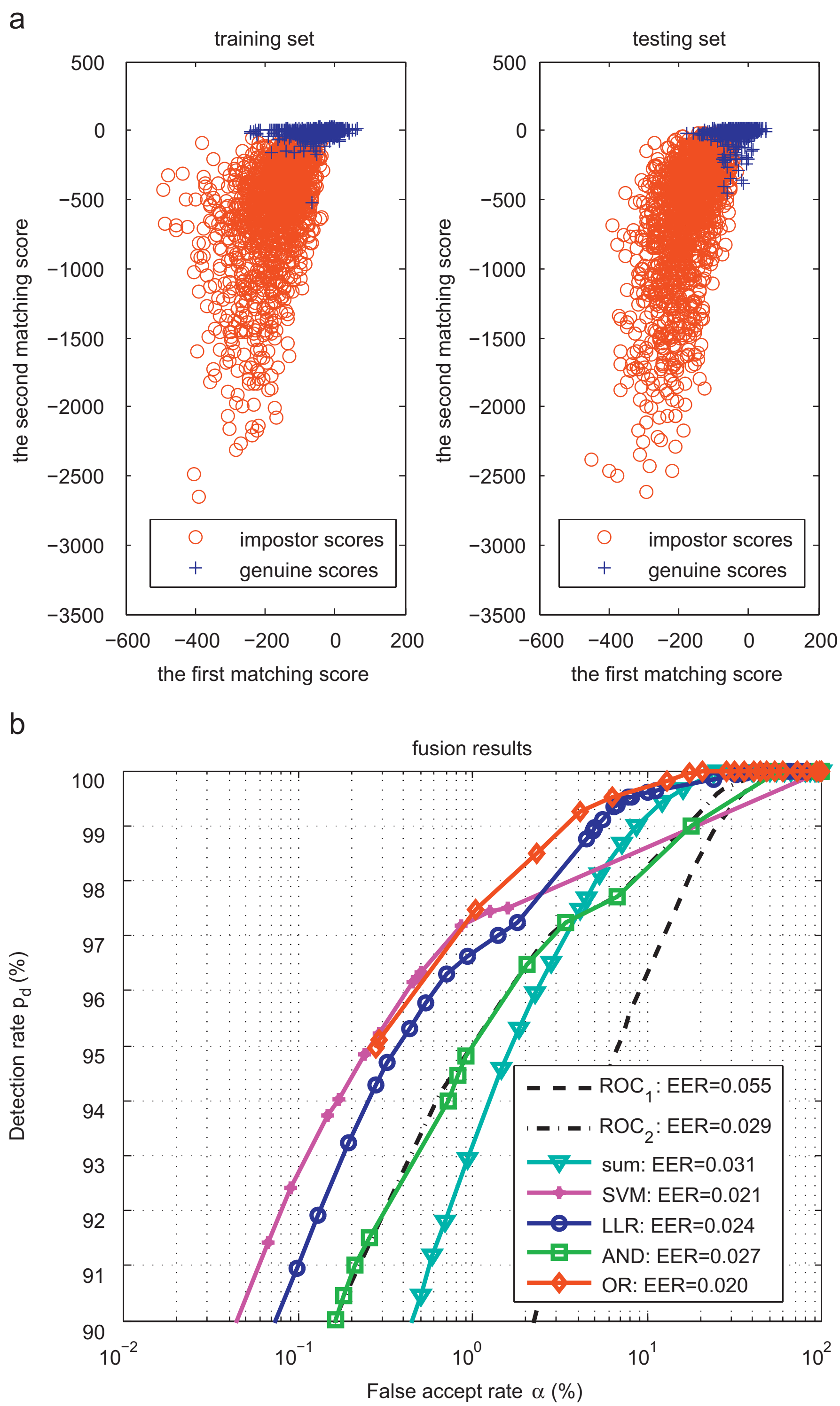

Fig. 6. Fusion between the UTW texture and UTW shape data: scatter plot and the fusion results.

For either modality, the matching scores are derived and provided by L-1 Identity Solutions (L1), Cognitec Systems (COG), and the University of Twente (UTW). In the L-1 method, the matching scores are computed using the hierarchical graph matching (HGM) methods [46], which represents the facial geometry by means of a flexible grid. Similar to the biological structures in the human brain, a set of specific filter structures is assigned to each node of the graph and analyzes the local facial characteristics [47,48]. With HGM, approximately 2000 characteristics are used to represent a face and an individual identity. For the analysis of a face, the shape ("landmarks") and the structure ("features") of the face are separated, making HGM a very robust facial recognition method providing a basis for both 2D and 3D face recognition. In the COG method, for 2D faces, the feature components are retrieved by applying local image Gabor transforms at facial feature locations. These components are then concatenated to form the raw 2D face

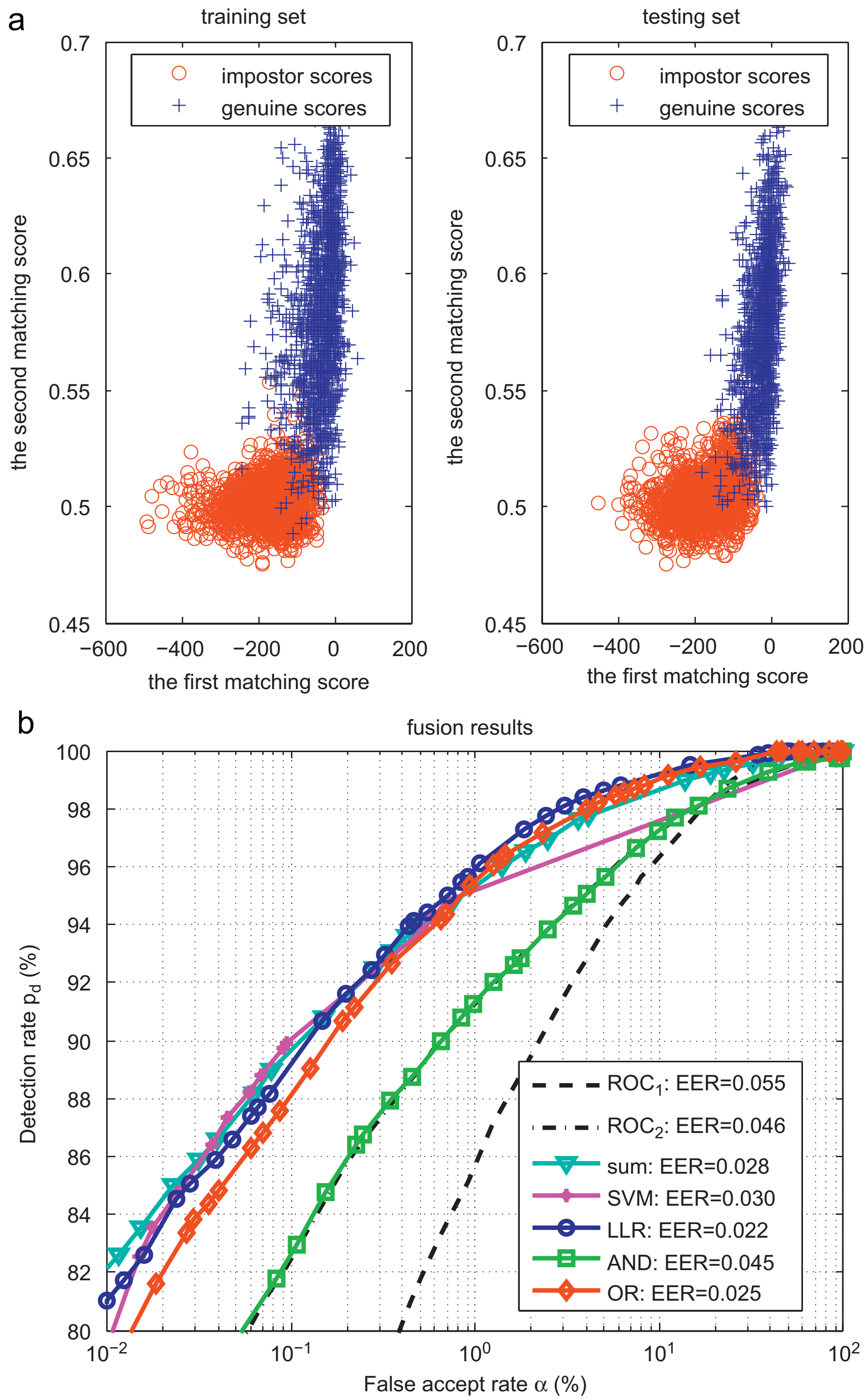

Fig. 7. Fusion between the UTW texture and L1 shape data: scatter plot and the fusion results.

feature vector. For 3D faces, the face shape is firstly registered and smoothed to form the raw 3D face feature vector. Global transformations are applied on the raw feature vectors in both cases, in order to maximize the ratio of inter-personal variance to intra-personal variance [49]. The final scores are obtained by simple similarity measures of the transformed feature vectors. In the UTW methods, holistic approach is taken, and the feature vectors are derived by the conventional PCA and LDA transformation, and the scores are computed as the likelihood ratio of the feature vector in the feature space. More details of the mathematics can be found in Ref. [50].

For comparison, we also implemented three other typical score-level fusion methods, namely, sum rule (transformation-based), likelihood ratio (density-based), SVM (classifier-based), which are explained in more detail in the following:

(1) Sum rule: In this transformation-based method, we used the simple and effective Z-normalization [1], which normal-

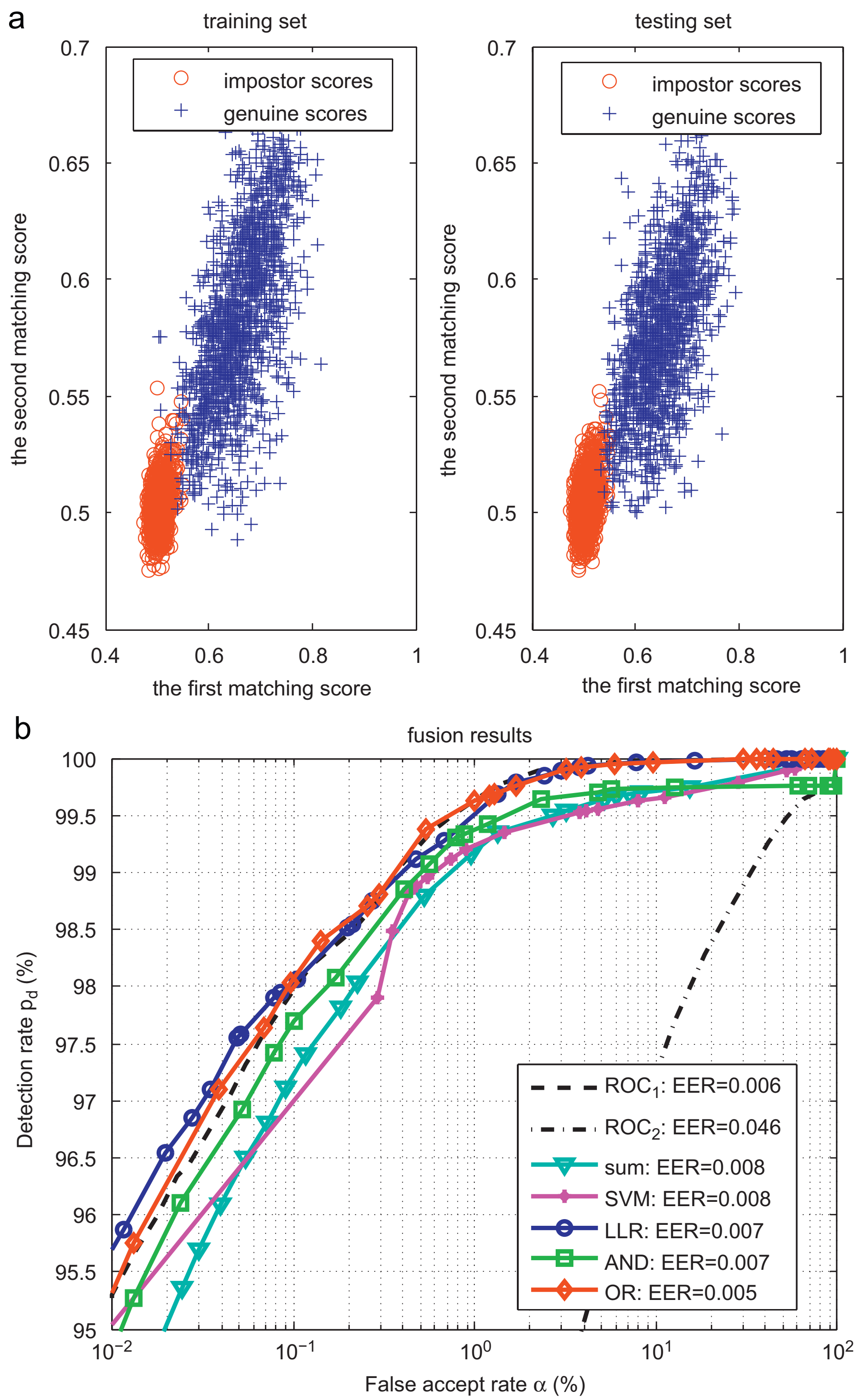

Fig. 8. Fusion between the L1 texture data and L1 shape data: scatter plot and the fusion results.

izes the genuine or impostor scores to unit variance. In this paper, we use the Z-normalization based on the genuine scores.[2]

(2) Likelihood ratio: In this density-based method, the score density is first estimated using Gaussian mixture models (GMM) [51], as in the work of Prabhakar and Jain [26]. Then the likelihood ratio is calculated based on the estimation of both genuine and impostor score distributions.
(3) SVM: In this classification-based methods, we used SVM as the classifier. The decision boundary is trained using the radius basis function (RBF) kernels [52]. The scores are firstly Z-normalized with a variance of 1, and the RBF radius is chosen as 1 as the

[2] We only present this one for readability of the figures. Z-normalization based on the impostor scores and other normalization techniques like min–max-normalization and tanh-normalization [1] have also been tried and yielded similar results.

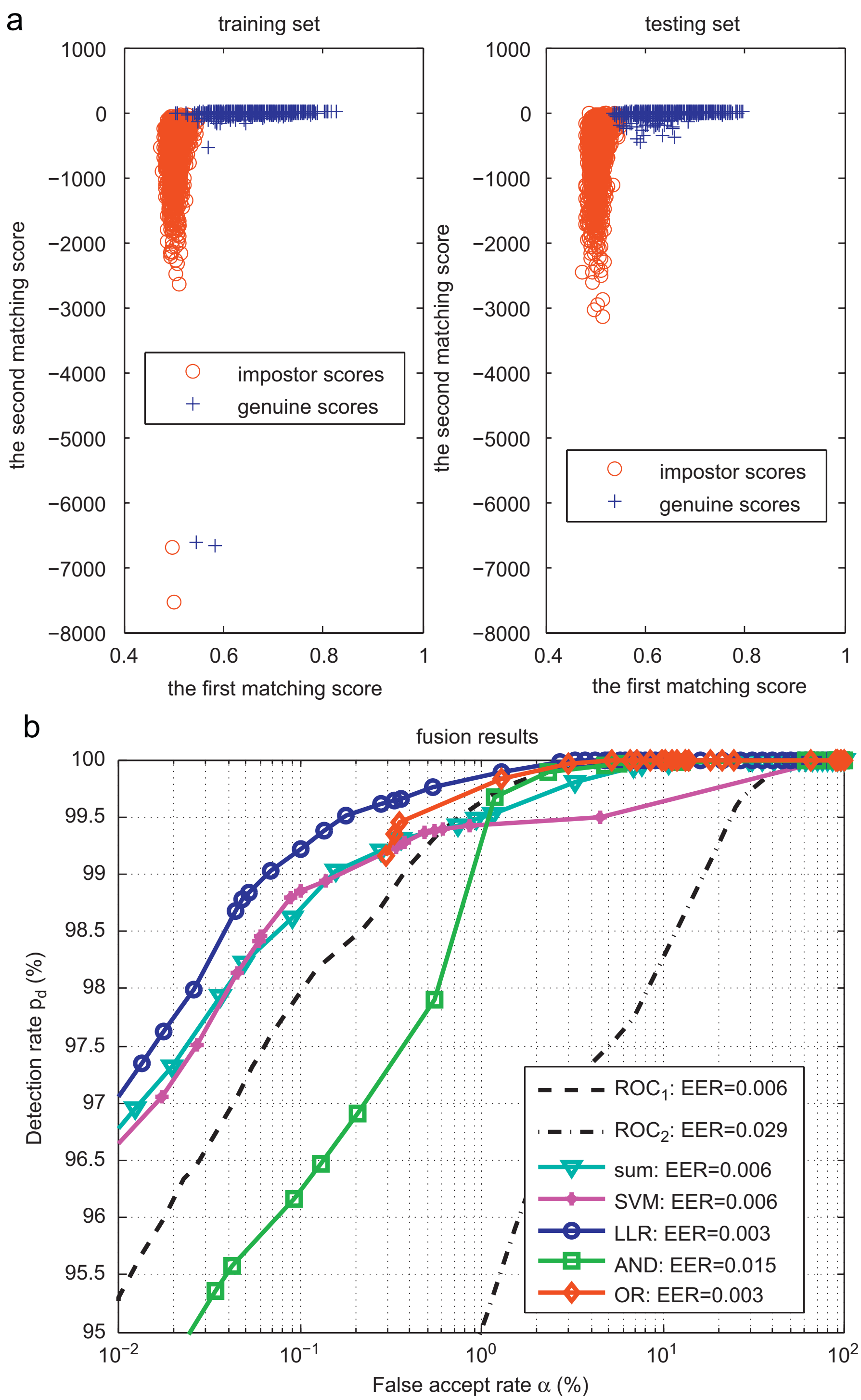

Fig. 9. Fusion between the L1 texture data and UTW shape data: scatter plot and the fusion results.

empirical value yielding robust performance. More implementation details can be found in Ref. [53].

Fusion is done between the two face modalities with scores derived from different algorithms. In each experiment, a training set is used first to find the parameters for fusion. In the decision-level fusion, the parameters refer to the optimized thresholds; while in the score-level fusion, the parameters refer to the normalization factors in the method (1), distributional parameters in the method (2), and SVM coefficients in the method (3). Then the evaluation of the methods are conducted on the testing data. For each fusion method, the resulting ROC are calculated and compared. Note that here we do not compare only a single operation point, instead we give an overview of the performance by plotting ROC, i.e., all the possible operation points.

Figs. 6–9 show 4 different combinations of the fusion between the face texture modality and shape modality. The ROC and the EER are shown, as well as the training score and testing score scatter plot. As observed, in Figs. 6 and 8 the OR rule fusion outperforms all other score-level fusion methods with respect to EER. The method even works better than the theoretically optimal LLR method. This can be explained by the discernible difference between the training and testing scores, which means that the probability density function might be over-tuned during the training. For the same reason, the support vectors are also different in the training and test set, thus accounting for the unsatisfactory performance of the SVM fusion method. Compared to the score-level fusion, the proposed ROC-based decision-level fusion are less sensitive to the training-testing data deviation, as indicated by Fig. 6. Another factor that makes the OR rule fusion favorable is its robustness against the outliers, as explained in Section 3. The AND rule fusion, however, does not work well, yielding performance sometimes even worse than the component ROC.[3]

In Fig. 7 the OR rule fusion works well, outperforming the score-level fusion methods except the likelihood ratio one, on the FAR range from 0.5% to 100%, but not as well on the lower FAR range (note the logarithm scale exaggerates this part). The likelihood ratio fusion method, as in Nandakumar et al. [24] and Prabhakar and Jain [26], remains the best. In Fig. 9 the OR rule fusion also performs worse than the LLR method in the lower FAR region, but equally well as far as the EER is concerned.

We have further combined all the 6 classifiers: 3 texture classifiers and 3 shape classifiers. The fusion is done in a recursive way as introduced in Appendix A. We show the fusion results on both the training ROCs and the testing ROCs in Fig. 10. It can be seen that the proposed decision-fusion outperforms other score-level methods, with considerable margin in the testing case. Such good performance is accounted by the outlier phenomenon existing in some component scores, as well as the obvious discrepancies between the training and testing data, which can be observed by comparing the 6 component ROCs in Fig. 10(a) and (b).

The LLR fusion is able to achieve the statistically optimal results and does outperform all the other methods in some experiments presented above, because it has the strongest theoretic support. Nevertheless, in this paper we still emphasize three properties of the proposed decision fusion: (1) Good performance at lower complexity. For example, the LLR fusion method needs to learn the joint probability distribution of the training scores, either in a parametric or a non-parametric way, and the SVM fusion methods needs to learn the support vectors and their corresponding weights, all of which has high computational complexity. (2) Tolerance to overtraining. This is mainly due to the simplicity of the decision boundary, as well as the fact that we only work on the ROC operation points, which is already a reduced representation of the scores. (3) Insensitivity to outliers. This has been elaborated on in Section 3, and illustrated in Fig. 6, in which the outlier phenomenon is most pronounced.

For the decision-level fusion, we have implemented the optimization methods derived in Section 2.2 and Appendix A, which is simple, but assumed independencies between the scores. Despite the certain degree of dependency between the two component scores, however, the final fused ROC on the testing data still demonstrate satisfactory performance. This can be explained by the fact that the main purpose of the proposed solutions is to find the optimal combination of thresholds which have the highest estimation of performance, instead of estimating the performance itself. In many dependent cases, the optimized thresholds are still plausible solutions, although the fused ROC is over-estimated. This is similar to the Naive Bayes classifier Duda et al. [28], which uses the independency assumption to estimate the class-conditional probabilities and then compare them. The estimated probabilities may very well be inaccurate, but the rank of them remains correct in many cases. The optimality of Naive Bayes classifier has been studied in literature [39,40].

[3] This may happen when the training set and testing set have different statistics.

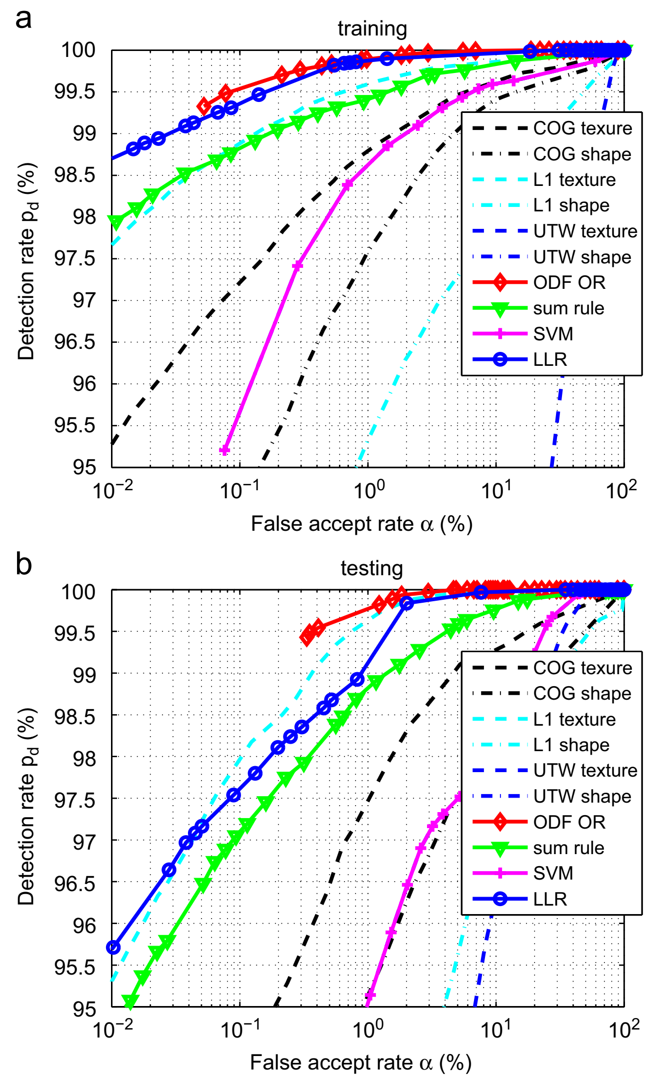

Fig. 10. Fusion of all the six classifiers: (a) training ROCs and (b) testing ROCs.

5. Conclusions

In this paper, a new fusion method called threshold-optimized decision-level fusion is proposed. Both the theoretical analysis and the experimental results have been presented. In theory, the proposed decision fusion will always bring improvements over the original classifiers that are fused, and in practice, it also improves the system performance effectively, in a way comparable or even better than the conventional matching score fusion.

Fusion at decision level by AND and OR rule is not a popular practice, but in this paper we have shown that it can be done in an optimal manner, by optimizing the thresholds of component classifiers, such that it can be very beneficial. By threshold-optimized decision fusion, matching score normalization is not needed, and the component classifiers are automatically balanced through the optimization process in training, thus reducing the risk of performance degradation, when the component classifiers have significantly different performances. In this way certain drawbacks related with AND/OR decision-level fusion [32] can be avoided. It is also noteworthy that the optimization is only based on the limited number of operation points on the ROC instead of directly on the matching scores. Furthermore, we have shown that the OR rule decision fusion is especially useful in presence of outliers. Compared to other score-level fusion, however, the disadvantage of decision-level fusion is the limited possibility of decision boundaries, because the operations are restricted to thresholding, AND, and OR. Therefore, a study of the matching score distribution characteristics and the classifier diversity [54,55] is strongly recommended before selecting the appropriate fusion method.

Threshold-optimized decision-level fusion based on optimizing the ROC is an interesting fusion method both in theory and in practice. From a Neyman–Pearson point of view, the improvements brought by the proposed decision fusion on FAR (FRR) with respect to a fixed FRR (FAR) is always very desirable for any biometric system.

Acknowledgments

This work is funded by the PNP2008 project of Freeband Netherlands, and the 3D-face project of the European Commission. The authors would also like to thank L-1 Identity Solutions, Bochum, Germany, and Cognitec GmbH, Dresden, Germany, for providing the matching scores used in this paper.

Appendix A. Optimality of recursive fusion

This procedure leads to an optimal solution, which is shown below for the OR rule. The proof for the AND rule is similar. As in Section 2.2, in the following derivation, the matching scores of different classifiers are assumed to be independent.

Let $\mathscr{I}$ and $\mathscr{J}$ denote the index sets, such that $\mathscr{I} \cap \mathscr{J} = \emptyset$ and $\mathscr{I} \cup \mathscr{J} = \{1, \ldots, N\}$. Define

$$p_{\mathrm{r}}^{\mathscr{I}}(\beta) = \max_{\beta_i | \prod \beta_i = \beta} \prod_{i \in \mathscr{I}} p_{\mathrm{r,i}}(\beta_i) \tag{A.1}$$

$$p_{\mathrm{r}}^{\mathscr{J}}(\beta) = \max_{\beta_j | \prod \beta_j = \beta} \prod_{j \in \mathscr{J}} p_{\mathrm{r,j}}(\beta_j) \tag{A.2}$$

and

$$p_{\mathrm{r}}^{\mathscr{I}\mathscr{J}}(\beta) = \max_{\beta^{\mathscr{I}}\beta^{\mathscr{J}} = \beta} p_{\mathrm{r}}^{\mathscr{I}}(\beta^{\mathscr{I}}) p_{\mathrm{r}}^{\mathscr{J}}(\beta^{\mathscr{J}}) \tag{A.3}$$

First, expanding $p_{\mathrm{r}}^{\mathscr{I}\mathscr{J}}(\beta)$ results in a product $\prod_{k=1}^{N} p_{\mathrm{r},k}(\beta_k)$ for some β_k, $k = 1, \ldots, N$, satisfying $\prod_{k=1}^{N} \beta_k = \beta$. Therefore, we have

$$p_{\mathrm{r}}^{\mathscr{I}\mathscr{J}}(\beta) \leqslant \max_{\beta_k | \prod \beta_k = \beta} \prod_{k=1}^{N} p_{\mathrm{r},k}(\beta_k) \tag{A.4}$$

Second,

$$\begin{aligned} p_{\mathrm{r}}^{\mathscr{I}\mathscr{J}}(\beta) &\geqslant p_{\mathrm{r}}^{\mathscr{I}}(\beta^{\mathscr{I}}) p_{\mathrm{r}}^{\mathscr{J}}(\beta^{\mathscr{J}})\big|_{\forall \{\beta^{\mathscr{I}}, \beta^{\mathscr{J}}\}:\ \beta^{\mathscr{I}}\beta^{\mathscr{J}} = \beta} \\ &\geqslant \prod_{i \in \mathscr{I}} p_{\mathrm{r,i}}(\beta_i)\Big|_{\forall \{\beta_i\}_{i \in \mathscr{I}}:\ \prod \beta_i = \beta^{\mathscr{I}}} \prod_{j \in \mathscr{J}} p_{\mathrm{r,j}}(\beta_j)\Big|_{\forall \{\beta_j\}_{j \in \mathscr{J}}:\ \prod \beta_j = \beta^{\mathscr{J}}} \\ &= \prod_{k=1}^{N} p_{\mathrm{r},k}(\beta_k)\Big|_{\forall \{\beta_k\}_{k=1}^{N}:\ \prod \beta_k = \beta} \\ &\geqslant \max_{\beta_k | \prod \beta_k = \beta} \prod_{k=1}^{N} p_{\mathrm{r},k}(\beta_k) \end{aligned} \tag{A.5}$$

The latter inequality follows by choosing the β_k such that they maximize $p_{\mathrm{r}}(\beta)$.

On combining Eqs. (A.4) and (A.5) we have

$$p_{\mathrm{r}}^{\mathscr{I}\mathscr{J}}(\beta) = \max_{\beta_k | \prod \beta_k = \beta} \prod_{k=1}^{N} p_{\mathrm{r},k}(\beta_k) \tag{A.6}$$

This means that *if the optimal ROCs are known for arbitrary disjoint index subsets $\mathscr{I}$ and $\mathscr{J}$, the overall optimal ROC can be found by optimally fusing the subsets*. Note that this statement is strictly true in ideal conditions, i.e., when the ROC is complete, with every point present on the ROC. In practice, however, the ROC cannot be complete, but represented by a limited number of operation points. The order of fusion, in this case, has some influences, but to an extent only as small as any other common numerical problems. As long as there are enough operation points from the ROC, the influences of the fusion order can well be neglected.

Appendix B. Threshold-optimized decision-level fusion on dependent decisions

It has been shown that to solve the proposed decision fusion problem under independency assumptions, we work directly on the ROCs and skip matching scores. In the dependent case, however, the fusion performance cannot be estimated as in Eq. (1). Instead, we return to the matching score space, and estimate the fusion performance in a nonparametric manner.

To illustrate the fusion process, we simulate two matching scores with dependency. The genuine matching scores have a multivariate Gaussian distribution of $N\left((2.5, 2.5), \begin{pmatrix} 1 & 0.25 \\ 0.25 & 1 \end{pmatrix}\right)$, while the impostor matching scores have a multivariate Gaussian distribution of $N\left((0, 0), \begin{pmatrix} 1 & 0.25 \\ 0.25 & 1 \end{pmatrix}\right)$. The matching scores are depicted by a scatter plot in a 2D space, as shown in Fig. B1 (a).

To estimate the performance of fusion, we created a threshold grid covering the matching score space, as shown in Fig. B1(a) by the cross points. The FAR α and detection rate p_{d} at each operation point can be estimated simply by applying the AND or OR rule, and then counting the number of false acceptances or false rejections. Suppose we have N_{gen} genuine samples and N_{imp} impostor samples, then from two classifiers, we have N_{gen} pair of genuine scores $(s_1^{\mathrm{gen}}, s_2^{\mathrm{gen}})$and N_{imp} pair of impostor scores $(s_1^{\mathrm{imp}}, s_2^{\mathrm{imp}})$. At any threshold (T_1, T_2), the ROC points by the OR and AND fusion can be easily calculated:

$$\alpha_{\mathrm{OR}}(T_1, T_2) = \frac{\|\{(s_1^{\mathrm{imp}}, s_2^{\mathrm{imp}}) | (s_1^{\mathrm{imp}} \geqslant T_1) \vee (s_2^{\mathrm{imp}} \geqslant T_2)\}\|}{N_{\mathrm{imp}}}$$

$$p_{\mathrm{d\ OR}}(T_1, T_2) = \frac{\|\{(s_1^{\mathrm{gen}}, s_2^{\mathrm{gen}}) | (s_1^{\mathrm{gen}} \geqslant T_1) \vee (s_2^{\mathrm{gen}} \geqslant T_2)\}\|}{N_{\mathrm{gen}}}$$

$$\alpha_{\mathrm{AND}}(T_1, T_2) = \frac{\|\{(s_1^{\mathrm{imp}}, s_2^{\mathrm{imp}}) | (s_1^{\mathrm{imp}} \geqslant T_1) \wedge (s_2^{\mathrm{imp}} \geqslant T_2)\}\|}{N_{\mathrm{imp}}}$$

$$p_{\mathrm{d\ AND}}(T_1, T_2) = \frac{\|\{(s_1^{\mathrm{gen}}, s_2^{\mathrm{gen}}) | (s_1^{\mathrm{gen}} \geqslant T_1) \wedge (s_2^{\mathrm{gen}} \geqslant T_2)\}\|}{N_{\mathrm{gen}}}$$

where $\|\cdot\|$ denotes size of the set. Consequently, for every threshold on the grid, ROC points can be shown in Fig. B1(c) and (d) by dots. Like the independent case, we again select those operation points which form a concave hull of the candidate points, as shown in Fig. B1(c) and (d). The optimized thresholds for decision-level fusion are therefore obtained as the thresholds corresponding to the selected points of operation.

Without the independency assumption, the presented decision fusion still has the good property that, similar to Eq. (4), the resulting ROC outperforms either component ROCs. This can be proved by the fact that in fusion, the original points of (α, p_{d})'s on ROC_1 and ROC_2 are still existent in the pool of candidate points to be

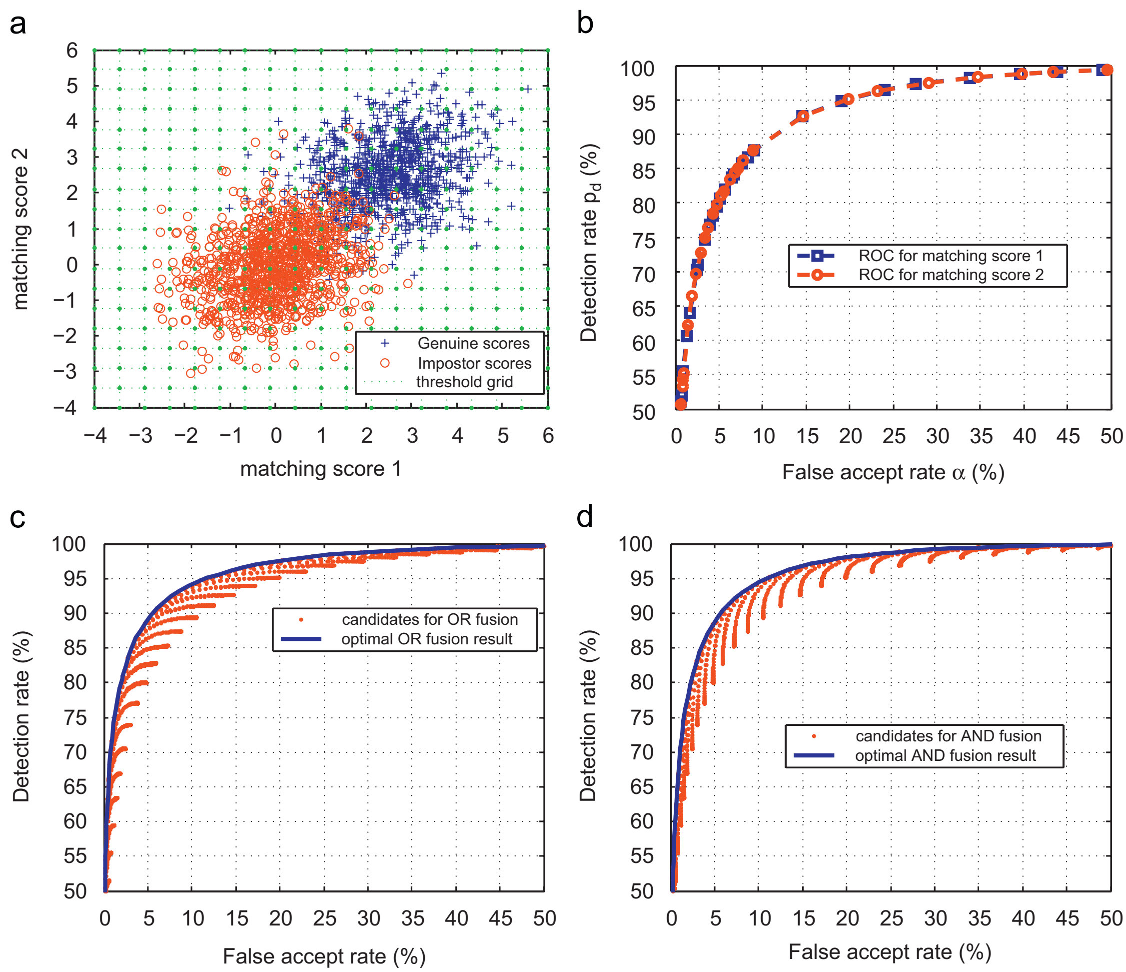

Fig. B1. Threshold-optimized decision fusion in the dependent case: (a) the scatter plot of two matching scores, and the threshold grid, (b) two ROCs of the two matching scores, respectively, (c) all the possible OR fused points and the optimal ROC selected, and (d) all the possible AND fused points and the optimal ROC selected.

selected.[4] Therefore, the resulting ROC, after the optimization of the concave hull, is again more favorable over the original ROCs in the Neyman–Pearson sense. It can be noticed, however, the margin of improvement becomes smaller compared to the independent case, as dependency of the two classifiers implies less added information.

References

[1] A. Ross, K. Nandakumar, A. Jain, Handbook of Multibiometrics (International Series on Biometrics), Springer-Verlag New York Inc., Secaucus, NJ, USA, 2006.
[2] B. Ulery, A. Hicklin, C. Watson, W. Fellner, P. Hallinan, Studies of biometric fusion, NIST Technical Report IR 7346, 2006.
[3] M. Faundez-Zanuy, Data fusion in biometrics, IEEE Aerosp. Electron. Syst. Mag. 20 (1) (2005) 34–38.
[4] H. Wang, W. Yau, A. Suwandy, Person recognition by fusing palmprint and palm vein images based on Laplacian palm representation, Pattern Recognition 41 (5) (2008) 1514–1527.
[5] F. Alsaade, A. Ariyaeeinia, A. Malegaonkar, Enhancement of multimodal biometric segregation using unconstrained cohort normalisation, Pattern Recognition 41 (5) (2008) 814–820.
[6] T. Theoharis, G. Passalis, G. Toderici, Unified 3d face and ear recognition using wavelets on geometry images, Pattern Recognition 41 (5) (2008) 796–804.
[7] T. Zhang, X. Li, D. Tao, Multimodal biometrics using geometry preserving projections, Pattern Recognition 41 (5) (2008) 805–813.
[8] D. Bouchaffra, A. Amira, Structural hidden Markov models for biometrics: fusion of face and fingerprint, Pattern Recognition 41 (5) (2008) 852–867.

[4] For ROC_1, the original operation points are obtained when the operation points of ROC_2 are tuned to extremes: for AND rule, $T_2 \to -\infty$; for OR rule, $T_2 \to \infty$. The same is true for ROC2.

[9] X. Zhou, B. Bhanu, Feature fusion of side face and gait for video-based human identification, Pattern Recognition 41 (5) (2008) 778–795.
[10] X. Jing, Y. Yao, D. Zhang, Face and palmprint pixel level fusion and kernel DCV–RBF classifier for small sample biometric recognition, Pattern Recognition 40 (11) (2007) 3209–3224.
[11] K. Wu, K. Yap, Fuzzy svm for content-based image retrieval—a pseudo-label support vector machine framework, IEEE Comput. Intell. Mag. 1 (5) (2006) 10–16.
[12] K. Wu, K. Yap, A soft relevance framework in content-based image retrieval systems, IEEE Trans. Circuits Syst. Video Technol. 15 (12) (2005) 1557–1568.
[13] J. Kittler, M. Hatef, R. Duin, L. Matas, On combining classifiers, IEEE Trans. Pattern Anal. Mach. Intell. 20 (3) (1998) 226–239.
[14] Y. Wang, T. Tan, A.K. Jain, Combining face and iris biometrics for identity verification, in: Fourth International Conference on AVBPA, 2003, pp. 805–813.
[15] C. Sanderson, K. Paliwal, Information fusion and person verification using speech and face information, Technical Report, IDIAP, Switzerland, September 2002.
[16] A. Ross, A. Jain, Information fusion in biometrics, Pattern Recognition Lett. 24 (13) (2003).
[17] B. Gokberk, L. Akarun, Comparative analysis of decision-level fusion algorithms for 3d face recognition, in: The 18th International Conference on Pattern Recognition, Washington, DC, USA, 2006, pp. 1018–1021.
[18] T. Saviv, N. Pavesic, Personal recognition based on an image of the palmar surface of the hand, Pattern Recognition 40 (11) (2007) 3152–3163.
[19] G. Marcialis, F. Roli, Fusion of appearance-based face recognition algorithms, Pattern Anal. Appl. 7 (2) (2004) 151–163.
[20] A. Lumini, L. Nanni, Combining classifiers to obtain a reliable method for face recognition, Multimedia Cyberscape J. 3 (3) (2005) 47–53.
[21] H. Van Trees, Detection, Estimation, and Modulation Theory, Wiley, New York, 1969.
[22] S. Dass, K. Nandakumar, A. Jain, A principled approach to score level fusion in multimodal biometric systems, in: Audio- and Video-Based Biometric Person Authentication, 2005, pp. 1049–1058.
[23] A. Jain, K. Nandakumar, A. Ross, Score normalization in multimodal biometric systems, Pattern Recognition 38 (12) (2005) 2270–2285.

[24] K. Nandakumar, Y. Chen, S. Dass, A. Jain, Likelihood ratio-based biometric score fusion, IEEE Trans. Pattern Anal. Mach. Intell. 30 (2) (2008) 342–347.
[25] D. Maurer, P. Baker, Fusing multimodal biometrics with quality estimates via a Bayesian belief network, Pattern Recognition 41 (5) (2008) 821–832.
[26] S. Prabhakar, A. Jain, Decision-level fusion in fingerprint verification, Pattern Recognition 35 (2002) 861–874.
[27] K. Toh, L. Kim, S. Lee, Biometric scores fusion based on total error rate minimization, Pattern Recognition 41 (5) (2008) 1066–1082.
[28] R. Duda, P. Hart, D. Stork, Pattern Classification, second ed., Wiley, New York, 2001.
[29] N. Littlestone, M. Warmuth, The weighted majority algorithm, Inf. Comput. 108 (2) (1994) 212–261.
[30] S. Kay, Fundamentals of Statistical Signal Processing: Detection Theory, Prentice-Hall, Englewood Cliffs, NJ, 1998.
[31] L. Kuncheva, J. Bezdek, R. Duin, Decision templates for multiple classifier fusion: an experimental comparison, Pattern Recognition 32 (2) (2001) 299–314.
[32] J. Daugman, Combining multiple biometrics ⟨http://www.cl.cam.ac.uk/users/jgd1000/combine/combine.html⟩, 2000.
[33] P. Tuyls, A. Akkermans, T. Kevenaar, G. Schrijen, A. Bazen, R. Veldhuis, Practical biometric authentication with template protection, in: 5th International Conference on Audio- and Video-Based Perosnal Authentication, Rye Brook, NY, USA, 2005, pp. 436–446.
[34] A. Juels, M. Wattenberg, A fuzzy commitment scheme, in: ACM Conference on Computer and Communications Security, 1999, pp. 28–36.
[35] K. Nandakumar, A. Jain, S. Pankanti, Fingerprint-based fuzzy vault: implementation and performance, IEEE Trans. Inf. Forensics Secur. 2 (12) (2007) 744–757.
[36] A. Martin, G. Doddington, T. Kamm, M. Ordowski, M. Przybocki, The DET curve in assessment of detection task performance, in: Proceedings of the Eurospeech '97, Rhodes, Greece, 1997.
[37] T. Fawcett, An introduction to ROC analysis, Pattern Recognition Lett. 27 (8) (2006) 861–874.
[38] L. Kuncheva, C. Whitaker, A. Shipp, R. Duin, Is independence good for combining classifiers?, in: 15th International Conference on Pattern Recognition, 2000, pp. 168–171.
[39] H. Zhang, The optimality of naive Bayes, in: 17th International FLAIRS Conference, 2004.
[40] P. Domingos, M. Pazzani, Beyond independence: conditions for the optimality of the simple Bayesian classifier, in: 13th International Conference on Machine Learning, 1996.
[41] W. Zhang, Y. Chang, T. Chen, Optimal thresholding for key generation based on biometrics, in: International Conference on Image Processing, 2004.
[42] A. Ortega, K. Ramachandran, Rate-distortion methods for image and video compression, IEEE Signal Process. Mag. 15 (6) (1998) 23–50.
[43] D. Tax, One class classification, Ph.D. Thesis, Delft University of Technology, 2001.
[44] 3D Face, 3D face biometric research ⟨http://www.3dface.org/⟩, 2006.
[45] P. Phillips, P. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, Overview of the face recognition grand challenge, in: Computer Vision and Pattern Recognition, 2005, pp. 947–954.
[46] M. Huesken, M. Brauckmann, S. Gehlen, K. Okada, C. von der Malsburg, Evaluation of implicit 3D modeling for pose-invariant face recognition, Biom. Technol. Hum. Identification 5404 (2004) 328–338.
[47] M. Husken, M. Brauckmann, S. Gehlen, C. Von der Malsburg, Strategies and benefits of fusion of 2d and 3d face recognition, in: (CVPR '05): proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)—Workshops, Washington, DC, USA, 2005, p. 174.
[48] L. Wiskott, J.-M. Fellous, N. Krüger, C. von der Malsburg, Face recognition by elastic bunch graph matching, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997).
[49] B. Moghaddam, T. Jebra, A. Pentland, Bayesian face recognition, Pattern Recognition 33 (2000) 1771–1782.
[50] A. Bazen, R. Veldhuis, Likelihood-ratio-based biometric verification, IEEE Trans. Circuits Syst. Video Technol. 14 (1) (2004) 86–94.
[51] M. Figueiredo, A. Jain, Unsupervised learning of finite mixture models, IEEE Trans. Pattern Anal. Mach. Intell. 24 (3) (2002) 381–396.
[52] N. Cristianini, J. Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernel-based Learning Methods, Cambridge University Press, Cambridge, 2000.
[53] T. Joachims, Making Large-Scale Support Vector Machine Learning Practical, MIT Press, Cambridge, MA, USA, 1999.
[54] E. Tang, P. Suganthan, X. Yao, An analysis of diversity measures, Mach. Learn. 65 (1) (2006) 247–271.
[55] L. Kuncheva, C. Whitaker, Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy, Mach. Learn. 51 (2003) 181C207.

About the Author—QIAN TAO received her B.Sc. and M.Sc. degree both from the Electrical Engineering Department, Fudan University, China, in 2001 and 2004, respectively. From 2004 Qian Tao joined the Signals and Systems Group, University of Twente as a Ph.D. student. Qian Tao currently works on face detection, recognition, and classifier combination. Her interest includes image processing and pattern recognition.

About the Author—RAYMOND VELDHUIS received his engineer degree in Electrical Engineering in 1981 from the University of Twente, The Netherlands. From 1982 until 1992 he worked as a researcher at Philips Research Laboratories in Eindhoven in various areas of digital signal processing, such as audio and video signal restoration and audio source coding. In 1988 he received the Ph.D. degree from Nijmegen University on a thesis entitled Adaptive Restoration of Lost Samples in Discrete-Time Signals and Digital Images. From 1992 until 2001 he worked at the IPO (Institute of Perception Research) Eindhoven in speech signal processing and speech synthesis. From 1998 until 2001 he was programme manager of the Spoken Language Interfaces research programme. He is now an associate professor at Twente University, working in the fields of biometrics and signal processing. He has published over 60 papers in international conferences and journals and has 20 patents in the field of sound, image and speech processing. He is also co-author of the book An Introduction to Source Coding, Prentice-Hall, and author of the book Restoration of Lost Samples in Digital Signals, Prentice-Hall. His expertise involves digital signal processing for audio, images and speech; statistical pattern recognition and biometrics. He has been active in the development of MPEG standards for audio source coding.