

Optimal Decision Fusion for a Face Verification System

Qian Tao, Raymond Veldhuis

Signals and Systems Group, Faculty of EEMCS
University of Twente, the Netherlands

Abstract. Fusion is a popular practice to increase the reliability of the biometric verification. In this paper, optimal fusion at decision level by AND rule and OR rule is investigated. Both a theoretical analysis and the experimental results are given. Comparisons are presented between fusion at decision level and fusion at matching score level. For our face verification system, decision fusion proves to be a simple, practical, and effective approach, which significantly improves the performance of the original classifier.

1 Introduction

Fusion is a popular practice to increase the reliability of the biometric verification by combining the outputs of multiple classifiers. Often, fusion is done based on these matching scores, because this combines a good performance with a simple implementation. In decision fusion, each classifier outputs an accept or reject decision and the fusion is done based on these decisions. The diagram of decision fusion can be drawn as in Fig. 1.

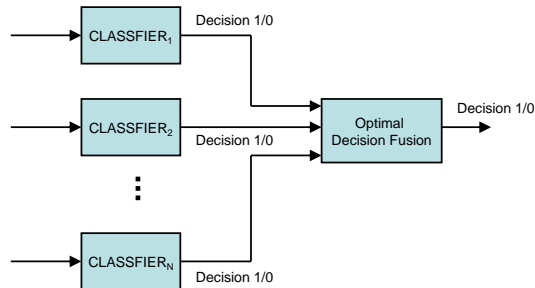


Fig. 1. Diagram of optimal decision fusion

In literature fusion at matching score level is more frequently discussed [2] [3] [6] [5]. In this paper, however, we will show that fusion at decision level by AND rule and OR rule can be applied 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 taken in Neyman-Pearson sense [9]: at a given false-reject rate α , the decision-fused classifier has a false-reject rate β that is minimal and never larger than the false-reject rates of the classifiers that are fused at the same α .

In this paper we apply the optimal decision fusion to a likelihood-ratio-based face verification system. At decision level the classifier outputs binary values: 0 for reject and 1 for accept. At matching score level, the classifier outputs the log likelihood ratio. Optimal decision fusions by AND rule and OR rule are compared to matching score fusion by sum rule.

This paper is organized as follows. In Section 2 theoretical analysis on optimal decision fusion is given. In Section 3 the application on face verification is described, and the results of optimal decision fusion on this system are shown. Section 4 gives the conclusions.

2 Optimal Decision Fusion

2.1 Optimal Decision Fusion Theory

Suppose we have two (or more) classifiers which output binary decisions. Assume that the decisions are statistically independent. (Note that this independency may arise from independent classifiers, or independent samples.)

Each decision D_i is characterized by two error probabilities: the first is the probability of a false accept, the false-accept rate (FAR), α_i , and the second is the probability of a false reject, the false-reject rate (FRR), β_i . To analyze the AND rule it is more convenient to work with the detection probability or detection rate $p_{d,i} = 1 - \beta_i$. It is assumed that $p_{d,i}$ is a known function of α_i , $p_{d,i}(\alpha_i)$, known as the ROC (Receiver Operating Characteristic). In practice, the ROC has to be derived empirically. After application of the AND rule to decisions $D_i, i = 1, \dots, N$, we have, under the important assumption that all decisions are statistically independent, that

$$\alpha = \prod_{i=1}^N \alpha_i \quad (1)$$

$$p_d(\alpha) = \prod_{i=1}^N p_{d,i}(\alpha_i) \quad (2)$$

with α the false-accept rate and p_d the detection rate of the fused decision, respectively. Optimal AND rule fusion can be formally defined by finding

$$\hat{p}_d(\alpha) = \max_{\alpha = \prod_{i=1}^N \alpha_i} \prod_{i=1}^N p_{d,i}(\alpha_i) \quad (3)$$

(3) means that the resulting detection rate p_d at a certain α is the maximal value of the product of the detection rates at some combination of α_i 's under the condition that $\alpha = \prod_{i=1}^N \alpha_i$. In other words, the α_i 's of component classifiers are tuned so that the

fused classifier can give maximal detection rate at a fixed $\alpha = \prod_{i=1}^N \alpha_i$. Likewise, if we define the reject rate for the impostors $p_{r,i} = 1 - \alpha_i$, the optimal decision fusion by OR rule can be similarly formulated

$$\hat{p}_r(\beta) = \max_{\beta = \prod_{i=1}^N \beta_i} \prod_{i=1}^N p_{r,i}(\beta_i) \quad (4)$$

where $\hat{p}_d(\alpha)$ and $\hat{p}_r(\beta)$ are the optimized ROCs by AND rule and OR rule, respectively.

For AND rule, it is easily proved that the optimized detection rate $\hat{p}_d(\alpha)$ is never smaller than any of the $p_{d,i}$'s at the same FAR α

$$\hat{p}_d(\alpha) \geq p_{d,i}(\alpha) \quad i = 1, \dots, N \quad (5)$$

Because, by definition

$$\begin{aligned} \hat{p}_d(\alpha) &= \max_{\alpha = \prod_{i=1}^N \alpha_i} \prod_{i=1}^N p_{d,i}(\alpha_i) \\ &\geq \prod_{j=1}^N p_{d,j}(\alpha_j) \Bigg|_{\prod_{i=1}^N \alpha_i = \alpha} \end{aligned} \quad (6)$$

As it holds for any classifier that, $p_{d,i}(1) = 1$, (5) readily follows by setting $\alpha_j = \alpha$ and $\alpha_i = 1, i \neq j$. For OR rule, it can be similarly proved that the optimized reject rate $\hat{p}_r(\beta)$ is never smaller than any of the $p_{r,i}$'s at the same FRR β .

By solving the optimization problem in (3) and (4), the operation points for every component classifiers are obtained, hence the fused classifier which yields the optimal performance in the Neyman-Pearson sense. Because in real situations, the ROCs, i.e. $\hat{p}_d(\alpha)$ or $\hat{p}_r(\beta)$, are characterized by a set of discrete operation points rather than analytically, the optimization in (3) and (4) must be solved in a numerical way. In [11] the problem is reformulated in a logarithmic domain as an unconstrained Lagrange optimization problem.

2.2 Optimal Decision Fusion on Identical Classifiers

In this section we will discuss, in particular, the optimal decision fusion on identical classifiers. This is a very useful setting in real applications, as will be shown in Section 3. Fusion on identical classifiers, in practice, means that given one classifier and multiple independent input samples, we make optimal fusion on the multiple output decisions.

In this paper, for simplicity, we analyze optimal fusion on two decisions. Fusion on three or more decisions can be done in a similar manner. Because the classifiers are identical, we have that $p_{d,1} = p_{d,2}$ and the optimization problem can be formulated to

$$p_{\text{fusion}}(x; \alpha) = p_d(x) \cdot p_d\left(\frac{\alpha}{x}\right) \quad (7)$$

$$\hat{p}_{\text{fusion}}(\alpha) = \max_{\alpha \leq x \leq 1} \{p_{\text{fusion}}(x; \alpha)\} \quad (8)$$

where x is a changing variable in the search process, and $\hat{p}_{\text{fusion}}(\alpha)$ is the detection rate at α under optimal AND fusion.

The optimum can be found by looking for the stationary point where the derivative of (7) w.r.t x is zero. As this derivative can be written

$$p'_{\text{fusion}}(x; \alpha) = p'_d(x)p_d\left(\frac{\alpha}{x}\right) - \frac{\alpha}{x^2}p_d(x)p'_d\left(\frac{\alpha}{x}\right) \quad (9)$$

Obviously when $x = \sqrt{\alpha}$, i.e. $\alpha_1 = \alpha_2 = \sqrt{\alpha}$, the derivative reaches zero. However, for some ROCs and for some α , this stationary point corresponds to a minimum, then the optimum is found at the border, either $\alpha_1 = 1$ or $\alpha_2 = 1$, which means only one of the two ROCs is taken.

In practice, therefore, under the optimal situation, either the two component classifiers work on identical operation points, or one of them does not effect at all. Although the former situation happens more often in practice, the later one does occur in certain cases.

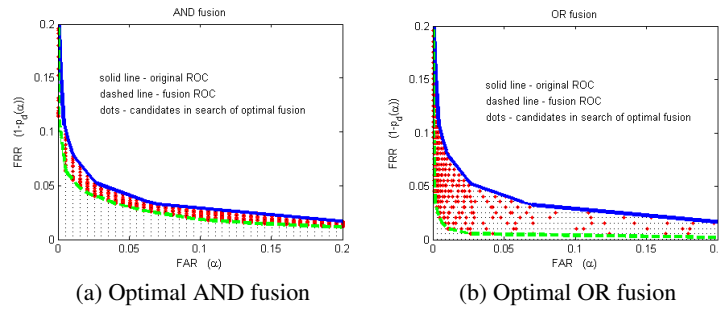


Fig. 2. Optimal decision fusion on ROC, example 1

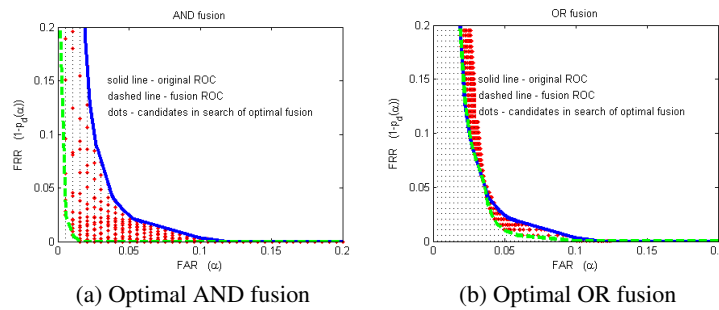


Fig. 3. Optimal decision fusion on ROC, example 2

Examples are shown to illustrate the optimal decision fusion. Fig. 2 and Fig. 3 show two examples of optimal decision fusion. In both figures, the solid line represents the original ROC, the dots represent the candidates in search of optimal point by (7) with different x , and the dashed line represents the resulting optimal ROC. Improvements of performance can be clearly seen in both cases. Furthermore, it can be observed that OR rule is more suitable for the ROC in Fig. 2, and AND rule is more suitable for the ROC in Fig. 3. In Fig. 3 (b) we can see that for a certain range of α , the two component classifiers are not working on the same operation point, but one of the two is taken.

To better explain the improvement brought by fusion, Fig. 4 visualizes different decision boundaries of the original classifier, AND fusion, OR fusion, and the sum rule. The crosses represent the scattering of two independent matching scores (in this case the logarithm likelihood ratio) for the user, and the circles represent the scattering of two independent matching scores for the impostors. In Fig. 4 (b) (c) (d), fusion facilitates decision boundaries spanning across a two dimensional space, where the two "clouds" are better separated compared to the case of a one dimensional space in Fig. 4 (a) (in which only one dimension is valid). Even better separation can be expected in higher dimensional spaces.

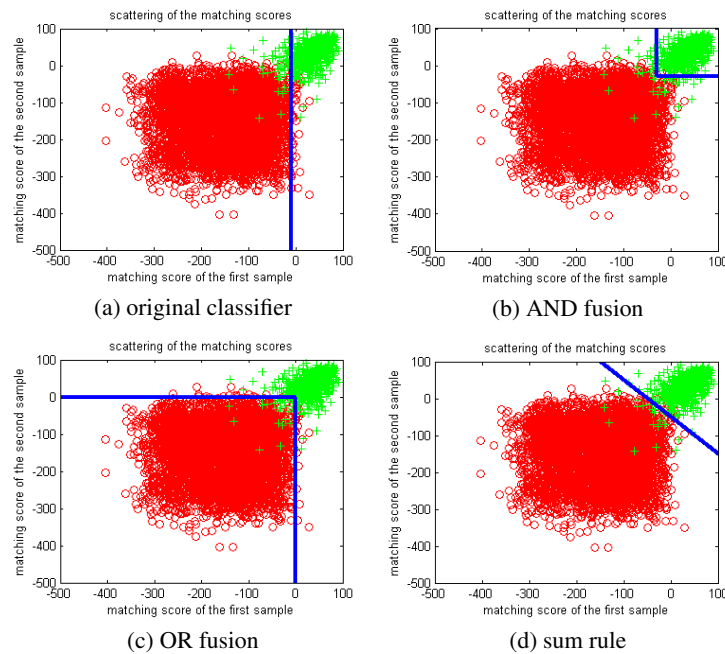


Fig. 4. Boundaries of different classifiers based on the original classifier

3 Application of Optimal Decision Fusion on a Face Verification System

3.1 The Face Verification System on a Mobile Device

In Section 2 the optimal decision fusion theory has been presented. In this section we describe a real application of a biometric verification system, on which the optimal decision fusion will be applied.

In a larger context, our biometric verification system acts as a link between a user and a private PN (personal network), via an intermediate MPD (mobile personal device) [7]. To achieve high security for the PN, it is specially demanded, among other requirements, that the authentication should be done not only at logon time, but also ongoing, in order to prevent the scenario that a MPD is taken away by the impostors after logged in by the user.

We use the face as the biometrics, and a camera on the MPD as the biometric sensor. In our standard system, features are extracted from each frame of face image, and a decision of YES or NO is made. In our original face recognition system, face detection is done by Viola-Jones method [10], face registration is done by aligning prominent facial landmarks detected also by Viola-Jones method. Illumination normalization is done by apply local binary patterns (LBP) [4] [1] as an preprocessing method. A likelihood ratio classifier is used which is based on the relative distribution between the user and the background data [8]. As the user-specific distribution has to be learned from extensive user training data which is beyond most public face databases, we collected our dataset under laboratory conditions. More than 500 frames of face images are collected per subject. (The database is still under construction, but the data used in this paper is available on request.)

In our new system with decision fusion, multiple frames with certain intervals are taken as the input, and the decision is made based on optimal fusion. It can be argued that the independency assumption is rendered less true when the intervals are chosen too small, but we will show that even in case of partial dependency, the decision fusion brings improvements to the performance of the system.

3.2 Experiments Setup

In the experiments, the face images are collected with a frequency of 5 frames per second, and stored as a function of time. For each subject, the data are independently collected in different sessions under different illuminations. Examples of the cross session data are shown by Fig. 5.

We use the data of two independent sessions for training and testing. Firstly, the classifier is trained on the first session. Secondly, the classifier is tested on the second session, and a ROC is obtained. The ROC represents the component classifier in the decision fusion. Then optimal decision fusion is then made on the ROC according to (3) or (4). Finally, the optimal decision fusion scheme is tested on multiple inputs from the second session, with each component classifier working on its optimal operation points.

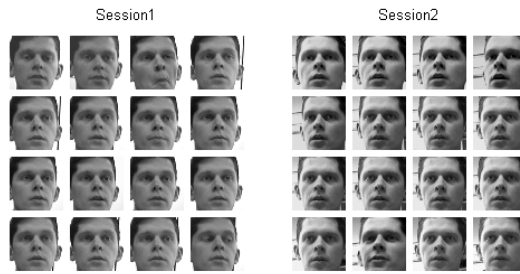


Fig. 5. The face data of a user collected in different sessions

3.3 Results on Optimal Decision Fusion

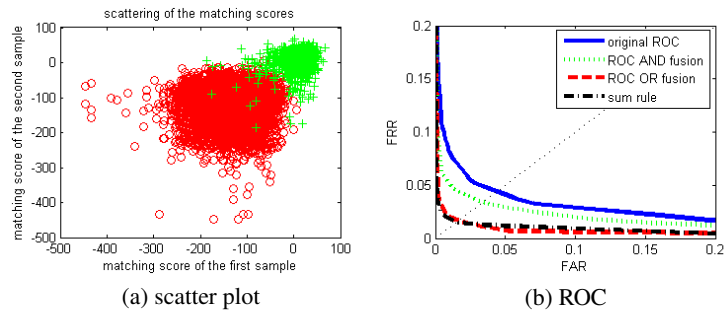


Fig. 6. Experiment results with randomly chosen samples

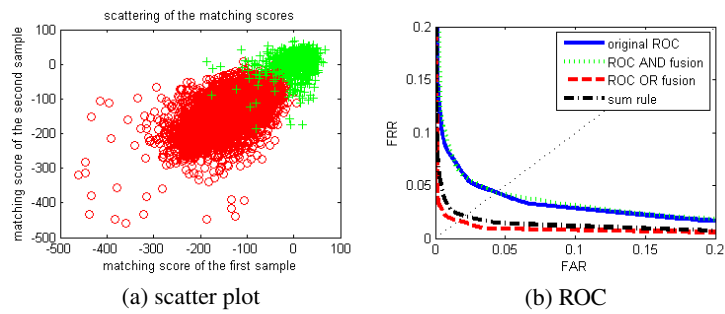


Fig. 7. Experiment results with samples chosen at a time interval of 0.5 second

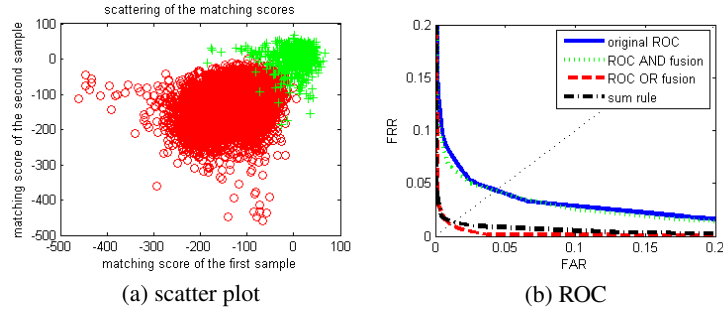


Fig. 8. Experiment results with samples chosen at a time interval of 15 second

In the following experiments, optimal decision fusion is done on two samples. The samples are taken in three ways. (1) The two samples are randomly taken; (2) The two samples are taken on a short interval of 0.5 second; (3) The two samples are taken on a longer interval of 15 second. For comparison, we also do sum rule matching score fusion, which is the theoretically optimal scheme for logarithm likelihood ratio matching scores. Fig. 6, Fig. 7, and Fig. 8 shows the results of these three sampling ways, respectively.

Improvements in performance can be clearly seen from Fig. 6, Fig. 7, and Fig. 8, with the EER (equal error rate) reduced to less than half of the original value. In Fig. 7 and Fig. 8 there exists certain correlation between the two samples, but despite this partial independency, OR rule still works very well and yields a performance comparable to or even better than the sum rule matching score fusion. The scatter plot indicates that in certain cases, a corner-shaped OR rule boundary is favored over a straight-line sum rule boundary.

3.4 Outliers and OR-Rule Optimal Decision Fusion

Outliers, in face verification, means the face images which belong to the user, but deviate from the user distribution because of extraordinary expressions or poses. Outliers occur in biometric verification, and cause rejections of the user. In our ongoing face verification system on a MPD, this harms the convenience aspect of the system [7]. Fig. 9 illustrates the outlier faces rejected by the original classifier.

The optimal decision fusion by OR rule, fortunately, can effectively reduce the FRR cause by the outliers at almost no expenses of FAR. Suppose the outlier distribution of the genuine user sample x is denoted by $\Psi_{G_o}(x)$, with a prior probability of a small quantity p_o , and suppose the distribution of the genuine user sample in normal cases is $\Psi_G(x)$, with a prior probability of $1 - p_o$. Taking into account the outlier distribution, the probability $\Psi(x)$ of a genuine user sample x can be expressed by

$$\Psi(x) = (1 - p_o) \cdot \Psi_G(x) + p_o \cdot \Psi_{G_o}(x) \quad (10)$$

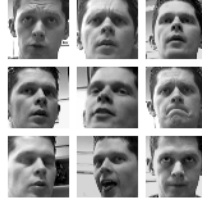


Fig. 9. Outliers in user data which are rejected by the classifier

For two samples x_1 and x_2 , assuming independency, their joint probability is

$$\begin{aligned} \Psi(x_1, x_2) = & (1 - p_o)^2 \cdot \Psi_G(x_1)\Psi_G(x_2) + p_o^2 \cdot \Psi_{Go}(x_1)\Psi_{Go}(x_2) \\ & + p_o n(1 - p_o) \cdot \Psi_{Go}(x_1)\Psi_G(x_2) + p_o(1 - p_o) \cdot \Psi_G(x_1)\Psi_{Go}(x_2) \end{aligned} \quad (11)$$

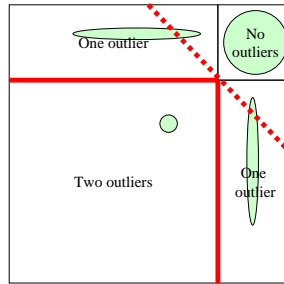


Fig. 10. The distribution of the two samples in fusion, taking into account the outlier distribution. The solid lines are the OR rule boundary, and the dotted line is the sum rule boundary.

The four terms in (11) describe the probability of the four different joint occurrences of the two samples, corresponding to Fig. 10. Note the second term, which describes the simultaneous occurrences of two outliers, is extremely small due to p_o^2 . In this case, OR rule boundary denoted by the solid line works better than the sum rule boundary denoted by the dotted line, with fewer false rejections. Real examples in our experiments also confirms the advantage of OR rule, as shown in Fig. 11. In this experiment, the cross session data are more extensive, therefore the outlier effects are more prominent.

It can be seen that in realistic situations, in presence of outliers, the OR rule works best. Comparing the OR rule performance with the sum rule performance in Fig. 11 (b), it can be seen that the FRR is effectively reduced at the same FAR.

4 Conclusions

In this paper, optimal fusion at decision level by AND rule and OR rule is proposed and investigated. Both the theoretical analysis and the experimental results are given,

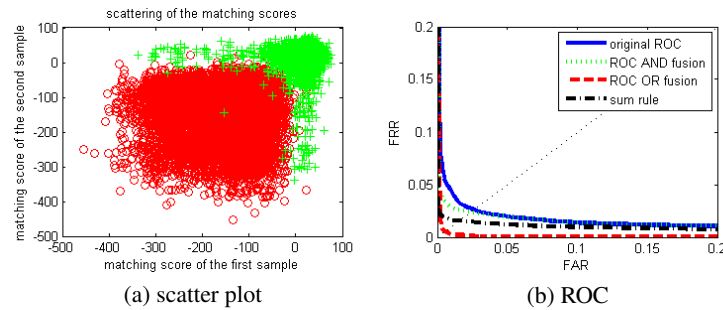


Fig. 11. Experiment results with samples with outliers

showing optimal decision fusion can always give an improvement to the performance of the original classifier. For our face verification system, decision fusion proves to be a simple, practical, and effective approach, which significantly improves the performance of the system. The improvements brought by optimal decision fusion on FAR with respect to a fixed FRR (or FRR with respect to FAR) is very desirable for biometric systems.

References

1. G. Heusch, Y. Rodriguez, and S. Marcel. Local binary patterns as image preprocessing for face authentication. In *IEEE International Conference on Automatic Face and Gesture Recognition*, 2006.
2. J. Kittler, M. Hatef, R. Duin, and J. Matas. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):226–239, 1998.
3. J. Kittler, Y. Li, J. Matas, and M. Sanchez. Combining evidence in multimodal personal identity recognition systems. In *Proc. 1st Int. Conf. on AVBPA*, Crans-Montana, Switzerland, 1997.
4. T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2004.
5. A. Ross and A. Jain. Information fusion in biometrics. 24(13), 2003.
6. A. Ross, K. Nandakumar, and A. Jain. *Handbook of Multibiometrics*. Springer Publishers, 2006.
7. Q. Tao and R. Veldhuis. Biometric authentication for mobile personal device. In *First International Workshop on Personalized Networks*, San Jose, USA, 2006.
8. Q. Tao and R. Veldhuis. Verifying a user in a personal face space. In *9th Int. Conf. Control, Automation, Robotics, and Vision*, Singapore, 2006.
9. H.L. van Trees. *Detection, Estimation, and Modulation Theory*. John Wiley and Sons, New York, 1969.
10. P. Viola and M. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.
11. W. Zhang, Y. Chang, and T. Chen. Optimal thresholding for key generation based on biometrics. In *International Conference on Image Processing*, 2004.