

Optimal Decision Fusion and Its Application on 3D Face Recognition

Qian Tao, Robin van Rootseler, Raymond Veldhuis,* Stefan Gehlen,† Frank Weber‡

Abstract: Fusion is a popular practice to combine multiple classifiers or multiple modalities in biometrics. In this paper, optimal decision fusion (ODF) by AND rule and OR rule is presented. We show that the decision fusion can be done in an optimal way such that it always gives an improvement in terms of error rates over the classifiers that are fused. Both the optimal decision fusion theory and the experimental results on the FRGC 2D and 3D face data are given. Experiments show that the optimal decision fusion effectively combines the 2D texture and 3D shape information, and boosts the performance of the system.

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 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.

In literature fusion at matching score level is more frequently discussed, see [KHDM98] [KLMS97] [RNJ06] [RJ03]. 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 [vT69]: 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 .

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

*Q. Tao, R. van Rootseler and R. Veldhuis are with the Signals and Systems Group, University Twente, the Netherlands.

†S. Gehlen is with the L-1 Identity Solutions.

‡F. Weber is with the Cognitec Systems GmbH.

2 Optimal Decision Fusion

2.1 Optimal AND Fusion

Suppose we have two (or more) classifiers which output binary decisions. Assume that the decisions are statistically independent. 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

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

where $\hat{p}_d(\alpha)$ is the optimized ROC by AND rule. Equation (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 $\alpha = \prod_{i=1}^N \alpha_i$.

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

$$\hat{p}_d(\alpha) \geq p_{d,i}(\alpha) \quad i = 1, \dots, N \quad (4)$$

Because, by definition

$$\begin{aligned} \hat{p}_d(\alpha) &= \max_{\alpha = \prod_{i=1}^N \alpha_i} \prod_{i=1}^N p_{d,i}(\alpha_i) \\ &\geq \prod_{j=1}^N p_{d,j}(\alpha_j) \Bigg|_{\prod_{i=1}^N \alpha_i = \alpha} \end{aligned} \quad (5)$$

As it holds for any classifier that, $p_{d,i}(1) = 1$, (4) readily follows by setting $\alpha_j = \alpha$ and $\alpha_i = 1, i \neq j$. This equation means that under the independency assumption, the optimal decision fusion can always bring improvement in terms of the FRR at a certain FAR, compared to any of the component classifiers.

2.2 Optimal OR Fusion

Likewise, if we define the correct reject rate for the impostors $p_{r,i} = 1 - \alpha_i$, under the assumption that all decisions are statistically independent, the optimal decision fusion by OR rule can be similarly formulated

$$\beta = \prod_{i=1}^N \beta_i \quad (6)$$

$$p_r(\beta) = \prod_{i=1}^N p_{r,i}(\beta_i) \quad (7)$$

Optimal OR rule fusion can be formally defined by finding

$$\hat{p}_r(\beta) = \max_{\prod_{i=1}^N \beta_i = \beta} \prod_{i=1}^N p_{r,i}(\beta_i) \quad (8)$$

where $\hat{p}_r(\beta)$ is the optimized ROC by OR rule. Equation (8) basically means that if we maximize the right hand term by tuning the component β_i 's, the optimal correct rejection rate $\hat{p}_r(\beta)$ can be found at a certain false reject rate β . It can be proved in the same way as for (4) that the fused classifier will outperform any of the component classifier in terms of the FAR at a certain FRR

$$\hat{p}_r(\beta) \geq p_{r,i}(\beta) \quad i = 1, \dots, N \quad (9)$$

which means that the improvement of performance can always be expected as in the AND fusion.

2.3 Solution to the Optimization Problem

Each pair of α and β on the ROC corresponds to a certain threshold t on the matching score level. When the threshold is gradually lowered, FRR will diminish, but FAR will grow. The reverse is true when the threshold is gradually increased. Therefore, in practice, the optimal decision fusion can be converted to the optimal tuning of the thresholds of component classifiers. This is done by solving the optimization problem in (3) and (8), which

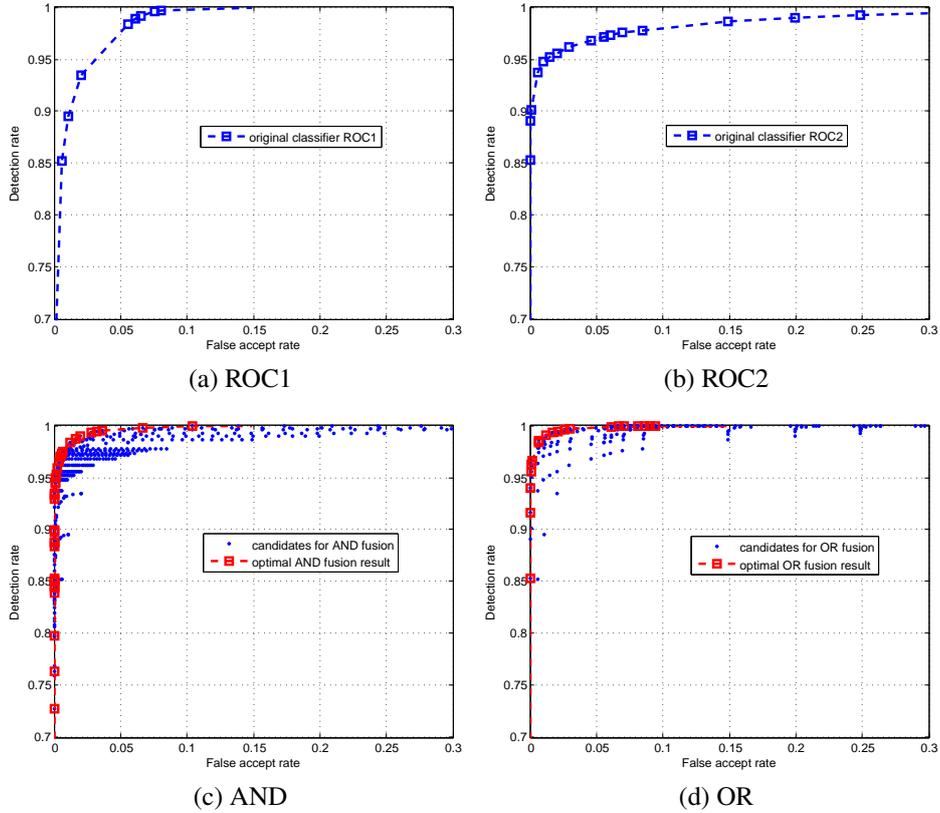


Abbildung 1: Optimal decision fusion: (a) operation points on the original ROC1; (b) operation points on the original ROC2; (c) all the possible AND fused points and the optimal ROC selected; (d) all the possible OR fused points and the optimal ROC selected.

result in the fused classifier with the optimal performance in the Neyman-Pearson sense [vT69]. In [ZCC04] the problem is reformulated in a logarithmic domain as an unconstrained Lagrange optimization problem. 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 (8) must be solved in a numerical way. We present our solution based on the set of discrete operation points, each points including the corresponding FAR α , detection rate p_d , and threshold t .

Take the fusion of two classifiers as an example. Suppose we have the ROC1 and ROC2, indicating two independent classifiers. If ROC1 has n_1 discrete points, where ROC2 has n_2 discrete points. The fusion of these two classifiers can have in total $n_1 \times n_2$ possible combinations $(\alpha_i^1 \alpha_j^2, p_{d_i}^1 p_{d_j}^2)$, where $i = 1, \dots, n_1, j = 1, \dots, n_2$. To get the optimal decision fusion, we select those points which form a concave hull of the candidate points. Taking two example ROCs from the 2D face recognition system developed by us [TV06], Fig. 1 illustrates the solution of the optimization problem. In Fig. 1 (c) and (d), the dots denote

the all the possible combinations from ROC1 and ROC2, and the dashed line runs across the selected operation points with the best performance. It can be seen that both the AND rule and OR rule optimal decision fusion result in a better ROC than ROC1 or ROC2. The corresponding points on the resulting ROC are the solution to the optimization problem. Fusion with more classifiers can be done in the same manner.

3 Application of Optimal Decision Fusion on 3D Face

In the following experiments the theory of fusion will be applied to real classifiers developed in the FP6 project on 3D face recognition. A classifier based on the texture of a face (2D) and a classifier that is based on shape (3D) will be fused at the decision level. In the first experiment the texture classifier from L-1 Identity Solutions and their shape classifier will be fused on the FRGC 2D+3D database. In a second experiment the texture classifier from L-1 Identity Solutions and the shape classifier from Cognitec Systems GmbH are fused. Both L-1 and Cognitec are partners in the 3D Face project.

The only information that we have about a classifier is a matrix of matching scores on the FRGC 2D+3D database. The dataset contains matching scores (similarity measure) for 466 individuals and has in total 4,007 samples. All 4,007 samples have been crossmatched yielding 16,056,049 matching scores (50,919 positive, 16,005,130 negative).

We will use the matching scores of 33% of the individuals for training and the other 67% will be used for testing. The individuals that will be used for training will be randomly selected from the total set. The optimal thresholds will be determined using the data for training and the performance will then be evaluated on the data for testing. For comparison the ROCs of the individual classifiers will be given. These ROCs are based on the test data that is used for fusion to make a fair comparison between the performance of the individual classifiers and their fused combination.

In Fig. 2 the results of AND-rule fusion on the L-1 texture and shape classifier have been depicted. The solid line gives the performance of the 2D texture classifier and the dashed line gives the performance of the 3D shape classifier. The dots indicate the performance of optimal AND-fusion applied on training data. Each dot corresponds to two thresholds (one for the texture classifier and one for the shape classifier). The squares give the performance on test data. From Fig. 2 it can be concluded that there is almost no gain by fusing the two classifiers using the AND-rule.

For the experiment using the OR-rule on the L-1 systems the results are given in Fig. 3. The results are similar to the AND-rule experiment. The output of the shape classifier does not provide additional information over the texture classifier that can be used for classification.

In the following two experiments two classifiers from different companies have been fused. The results are based on the 2D texture classifier from L-1 (the same as was used in the previous experiment) and the 3D shape classifier from Cognitec. In Fig. 4 the results of AND-rule fusion to these two classifiers have been depicted. The solid line gives the performance of the 2D texture classifier on the test dataset and the dashed line gives the performance of the 3D shape classifier. The dots indicate the performance of optimal

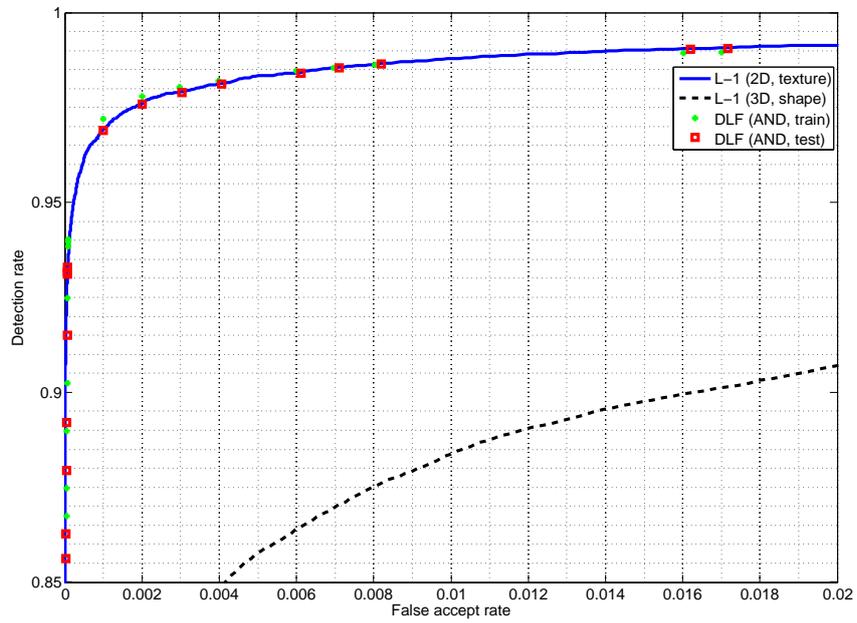


Abbildung 2: AND-fusion applied to 2D texture classifier (L-1) and 3D shape classifier (L-1)

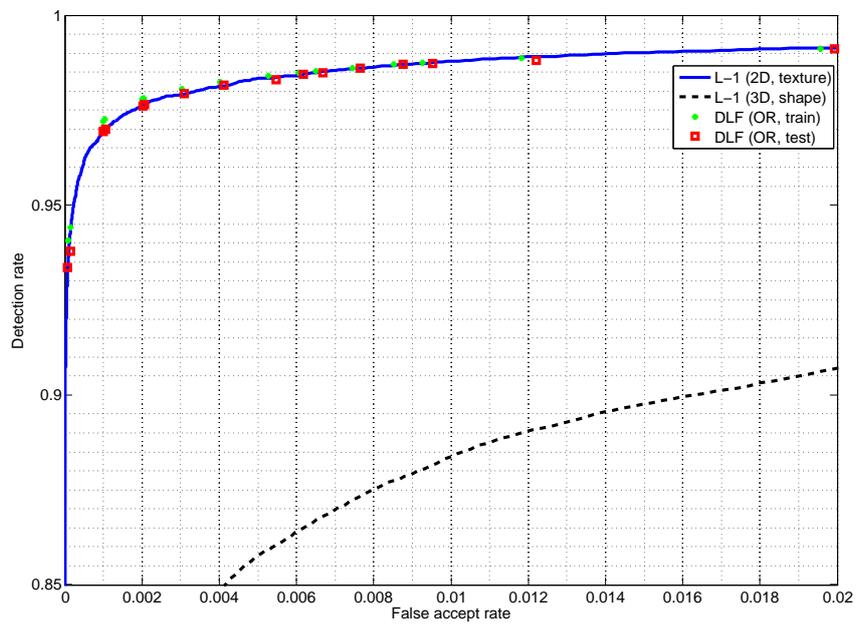


Abbildung 3: OR-fusion applied to 2D texture classifier (L-1) and 3D shape classifier (L-1)

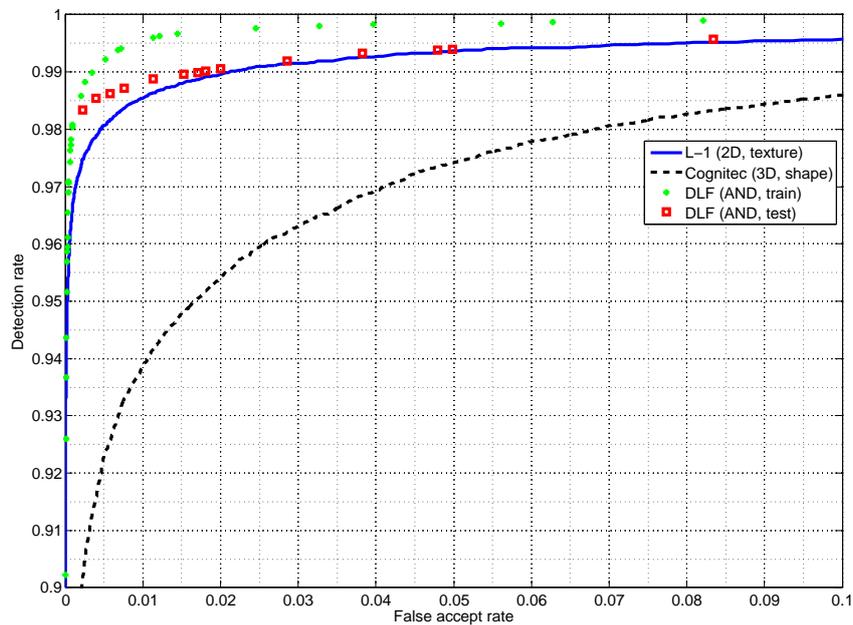


Abbildung 4: AND-fusion applied to 2D texture classifier (L-1) and 3D shape classifier (Cognitec)

AND-fusion applied on training data. These results look quite promising.

An evaluation of the performance of the fused classifier can be made using the thresholds corresponding to the dots. The performance of the fused classifier is indicated by squares in Fig. 4. The fused classifier outperforms the individual classifiers on all points, although the gain in performance for a FAR > 0.02 is hardly noticeable.

The same experiment has been repeated, but now using the OR-rule to fuse the two classifiers. Since both classifiers have more problems with their detection rates than with their false accept rates it can be expected that the OR-rule has more benefits than the AND-rule.

The results are depicted in Fig. 5. The gain in performance by fusing the two classifiers using the OR-rule is now clearly visible and significant. At a low FAR (FAR ≤ 0.001) the performance of the fused classifier is identical to the performance of the 2D texture classifier. This means that at a very low FAR the output of the 3D shape classifier outputs no useful information and its threshold will therefore be set to a very high value. A very high threshold will make the output of the shape classifier always 0. This means that at that point the output of the shape classifier does not influence the result using OR-rule fusion.

In Fig. 6 a scatter plot of the matching scores from the texture and the shape classifier is given. From this scatter plot it can be seen why the performance of the fused classifier using the OR-rule has improved. The stars denote the genuine scores and the circles denote the impostor scores. For both classifiers a threshold level has been chosen (indicated by the dotted line). These threshold values divide the plane in four sections. The outliers of the texture classifier can be found in section 1 and 3. The outliers of the shape classifier

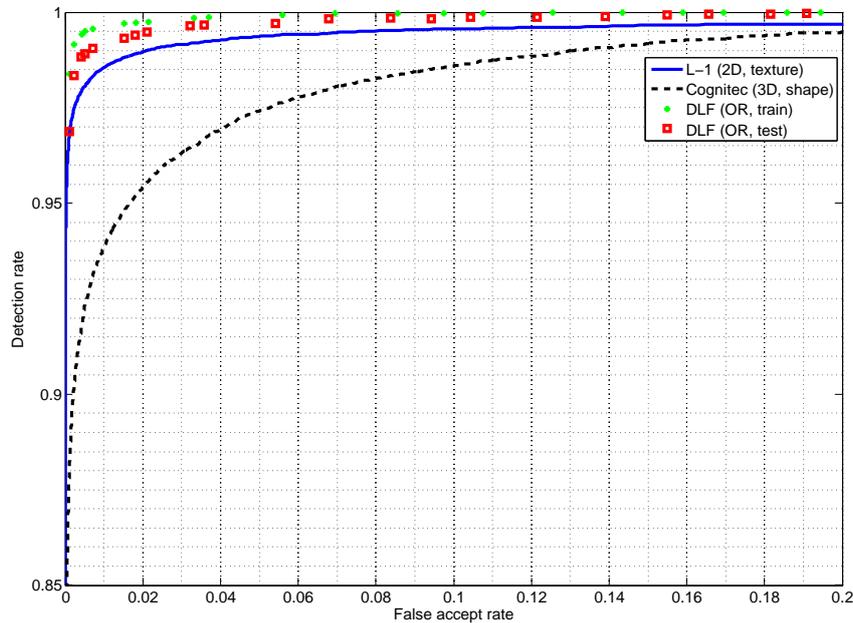


Abbildung 5: OR-fusion applied to 2D texture classifier (L-1) and 3D shape classifier (Cognitec)

can be found in section 3 and 4. By using the OR-rule, the outliers in section 1 and 4 can be correctly classified.

4 Conclusions

In this paper, optimal fusion at decision level by AND rule and OR rule is presented. Both the theoretical analysis and the experimental results are given. In theory optimal decision fusion can always improve the performance of the original classifiers, and in the experiments of 2D plus 3D face data optimal decision fusion proves to bring improvement over the original classifiers. A great advantage of the decision fusion is that its performance is invariant to any normalization method applied to the matching scores. Therefore unlike fusion on matching score level, optimal decision fusion is not influenced by the way of normalization and the scope of matching scores from different classifiers. To conclude, optimal decision fusion is an effective way to combine different classifiers or modalities in biometrics, and 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 any biometric systems.

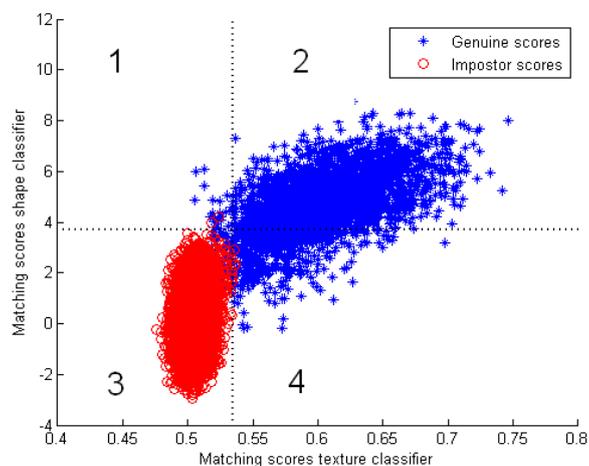


Abbildung 6: Scatter plot of matching scores

5 Acknowledgemet

The authors would like to thank the L-1 Identity Solutions and the Cognitec Systems GmbH for providing the matching scores used in this paper.

Literatur

- [KHDM98] J. Kittler, M. Hatef, R. Duin und J. Matas. On Combining Classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):226–239, 1998.
- [KLMS97] J. Kittler, Y Li, J. Matas und M. Sanchez. Combining Evidence in Multimodal Personal Identity Recognition Systems. In *Proc. 1st Int. Conf. on AVBPA*, Crans-Montana, Switzerland, 1997.
- [RJ03] A. Ross und A. Jain. Information Fusion in Biometrics. 24(13), 2003.
- [RNJ06] A. Ross, K. Nandakumar und A. Jain. *Handbook of Multibiometrics*. Springer Publishers, 2006.
- [TV06] Q. Tao und R. Veldhuis. Verifying a user in a personal face space. In *9th Int. Conf. Control, Automation, Robotics, and Vision*, Singapore, 2006.
- [vT69] H.L. van Trees. *Detectioin, Estimation, and Modulation Theory*. John Wiley and Sons, New York, 1969.
- [ZCC04] W. Zhang, Y. Chang und T. Chen. Optimal thresholding for key generation based on biometrics. In *International Conference on Image Processing*, 2004.