Hybrid Fusion for Biometrics: Combining Score-level and Decision-level Fusion

Qian Tao Raymond Veldhuis
Signals and Systems Group, University of Twente
Postbus 217, 7500AE Enschede, the Netherlands
{q.tao,r.n.j.veldhuis}@ewi.utwente.nl

Abstract

A general framework of fusion at decision level, which works on ROCs instead of matching scores, is investigated. Under this framework, we further propose a hybrid fusion method, which combines the score-level and decision-level fusions, taking advantage of both fusion modes. The hybrid fusion adaptively tunes itself between the two levels of fusion, and improves the final performance over the original two levels. The proposed hybrid fusion is simple and effective for combining different biometrics.

1. Introduction

Biometrics, which uses a variety of physical or behavioral characteristics to verify a person's identity, is widely used in a lot of security applications. To overcome the limitation of a single biometrics, information from multiple biometrics can be integrated to achieve more reliable and robust performance. For this purpose, fusion of diverse biometrics has been extensively studied in recent years. For a detailed review, see [10].

According to the different stages of a biometric system, fusion can be done at four distinct levels, namely: sensor (raw biometric data) level, feature level, matching score level, and decision level. Along these levels the biometric information is gradually extracted and reduced. On the first two stages, the information content is rich, but in most cases noisy and redundant. On the matching score level, the information is reduced into a single quantity, indicating the likelihood that the biometric data belongs to a certain class. On the final decision level, the information is further reduced to the discrete class labels. In this paper, we will concentrate on the last two levels of fusion, not only because of the simplicity, but also because of the possibility to build up a general fusion framework, without taking into account the specific type of biometric data processing and classification methods, which would closely influence the first two levels of fusion.

Fusion at matching score level is the most popular way of fusion, offering the best tradeoff between information content and fusion complexity [6, 15, 11, 9, 10]. Fusion at decision level, in comparison, is less studied, as it is often considered inferior to matching score level fusion, on the basis that decisions have less information content than the matching scores. Actually, the combination of the two decisions using AND and OR rule often has the risk of degrading the overall performance when the performance of component classifiers are significantly different [3].

A optimal decision fusion method by the AND and OR rule has been proposed in literature [12]. In this method, the fusion at decision level is done 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 the Neyman-Pearson sense [14]: at a given FAR (false accept rate) α, the decision-fused classifier has a FRR (false reject rate) β that is minimal, and never larger than the β of the component classifiers at the same α. Besides, the method has the advantage that in presence of outliers (i.e. the biometric data which belongs to the genuine user but deviate from the modeled distributions, possibly caused by the variability of collection conditions), the OR rule decision fusion can achieve a low FAR with little risk of increasing FRR [12]. In this paper, we will extend this work, constructing a more general framework of decision fusion oriented on performance, and propose a hybrid fusion scheme which combines the score-level and decision-level fusion.

Instead of dealing with the matching scores, the fusion framework works directly on the ROC (receiver operation characteristic). Although the ROC is derived from the matching scores, the problem is still made different: the matching scores are converted into a compact set of operations points, which convey the distribution information of matching scores in an indirect way. The optimization in the framework only involves those operation points, without reference to the matching scores.

Under this framework, any two (or more) ROCs can be fused together for improved performance. Those ROCs could characterize any biometric system, either of a single biometric, or of a already fused multi-biometrics. This en-

978-1-4244-2340-8/08/$25.00 ©2008 IEEE

ables us to do fusion in an hybrid manner, combining score-level and decision-level fusion and taking advantage of both fusion modes.

The paper is organized as follows. Section 2 reviews the decision-level fusion framework. Section 3 introduces the hybrid fusion. Section 4 shows the experimental results. Finally, Section 5 gives conclusions.

2. A Decision-level Fusion Framework

Each biometric system can be characterized by a ROC, i.e., the detection rate $p_{\rm d}$ ($p_{\rm d} = 1 - \beta$) as a function of false accept rate α, denoted by $p_{\rm d}(\alpha)$. The ROC is obtained by varying the threshold that discriminates the genuine and impostor matching scores, thus producing different detection rate p_d and false accept rate α. Each point on the ROC, a specific pair of $(\alpha, p_{\rm d})$, is called an operation point, corresponding to a particular threshold t of the matching scores. In this section we will show how multiple ROCs can be fused together simply by AND and OR rule for improved performance. When the optimal operation points on ROC are obtained, the thresholds of matching scores are obtained as well.

Suppose we have N independent biometric systems, each characterized by its ROC, $p_{{\rm d},i}(\alpha_i)$, $i = 1, ..., N$. The independency assumption is realistic in practice as fusion is often done between different biometric modalities. Besides, the independency assumption in this section makes the formulations much simpler and clearer. The dependent cases, however, will be discussed in Section 3.2.

If the AND rule is used for fusion, the final performance can be estimated, under the independent assumption, as

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

with α the false-accept rate and $p_{\rm d}$ the detection rate of the AND rule fused decision, respectively. In search of the optimal operation points, the fusion framework by AND rule can be formulated as

$$\hat{p}_{\rm d}(\alpha) = \max_{\alpha_i | \prod_{i=1}^{N} \alpha_i = \alpha} \left\{ \prod_{i=1}^{N} p_{{\rm d},i}(\alpha_i) \right\} \tag{2}$$

which means that the resulting detection rate $\hat{p}_{\rm d}$ at α is the maximal value of the product of the detection rates at a certain optimal combination of α_i, $i = 1, ..., N$, which satisfy $\prod_{i=1}^{N} \alpha_i = \alpha$. In other words, at a prefixed α, the optimal operation points of the component ROCs are obtained by optimizing (2). Consequently, the thresholds of component biometric systems can be readily obtained as the ones corresponding to the optimized operation points.

Likewise, if we define the reject rate for the impostors $p_{{\rm r},i} = 1 - \alpha_i$, the fusion framework by OR rule can be similarly formulated

$$\hat{p}_{\rm r}(\beta) = \max_{\beta_i | \prod_{i=1}^{N} \beta_i = \beta} \left\{ \prod_{i=1}^{N} p_{{\rm r},i}(\beta_i) \right\} \tag{3}$$

It can be easily proved that the optimized detection rate $\hat{p}_{\rm d}(\alpha)$ in (2) is never smaller than any of the component $p_{{\rm d},i}$, $i = 1, ..., N$, at the same α, and $\hat{p}_{\rm r}(\beta)$ in (3) is never smaller than any of the component $p_{{\rm r},i}$, $i = 1, ..., N$, at the same β [12]. If a certain classifier cannot help or possibly degrades the overall performance, the optimization will switch it off by tuning its operation points to $\alpha = 1$, $p_{\rm d} = 1$ in case of fusion by AND rule, or $\beta = 1$, $p_{\rm r} = 1$ in case of fusion by OR rule.

In practice, it is in most cases not possible to have the ROC $\hat{p}_{\rm d}(\alpha)$ in analytical form, instead, the ROC has to be estimated from the evaluation data. As a result, $\hat{p}_{\rm d}(\alpha)$ are characterized by a set of discrete operation points rather than a continuous function. The optimization problem formulated in (2) and (3), therefore, has to be solved numerically. In a brute-force way, the optimization could be done by first calculating the pool of operation points, i.e, estimating all the possible combinations by (1), and then select the ones optimal in the Neyman-Pearson sense. The fusion of three or more ROCs, as proved in Appendix A, can be reduced to iteratively fusing two ROCs. Therefore, the number of possible combinations does not explode rapidly with the number of ROCs, and the complexity of the optimization is kept low. An example is given to illustrate the optimization procedure, as shown in Fig. 1. The first ROC is obtained by generating genuine scores as the random variables of Gaussian distribution $N(1.5, 1)$, and impostor scores of $N(-1.5, 1)$, while the second ROC is obtained by generating genuine scores of $N(2, 1)$ and impostor scores of $N(-2, 1)$. The possible operation points after fusion are indicated by dots, while the final optimized points are marked by small squares. It can be observed that both the AND and OR fused ROCs are improved, in the Neyman-Pearson sense, over the two original ROCs.

3. Hybrid Fusion

The motivation for the hybrid fusion is twofold. Firstly, we show that the decision fusion framework using ROCs is very general and can be extended easily. Secondly, by hybrid fusion we hope to take advantage of the score-level and decision-level fusion, and eventually achieve an even more reliable and robust biometric system. In this section, we will first discuss the pros and cons of the score-level and decision-level fusion, and then present the hybrid fusion method.

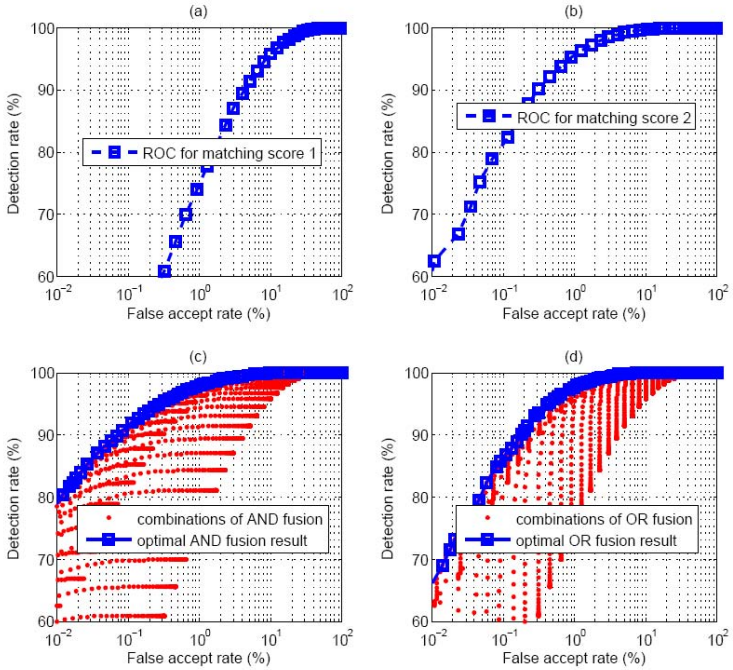


Figure 1. (a) the first component ROC; (b) the second component ROC; (c) all the possible AND fused points and the optimal ROC selected; (d) all the possible OR fused points and the optimal ROC selected.

3.1. Score-level and Decision-level Fusion: Pros and Cons

Score-level fusion is the most popular way of fusion. The advantage of it is obvious. As a quantitative similarity measure it contains rich information about the biometric input, and yet it is still easy to process compared to sensor-level or feature-level data. In many cases, score-level fusion is able to achieve theoretically optimal performance. For example, taking product of the matching scores, which are independent and proportional to the likelihood ratio (in the feature space), is an ideal estimation of the joint likelihood ratio. Also, in the density-based score-level fusion [2], the ROC corresponding to the likelihood ratio statistic (in the matching score space), is optimal in the Neyman-Pearson sense.

A disadvantage of score-level fusion is that, because it works in the matching score space, it is subject to considerable flexibilities. For example, different normalization methods of the matching scores lead to different decision boundaries. Also, a too small training set of scores might easily overfits the data, especially in methods with flexible boundaries.

There are also advantages and disadvantages of the decision-level fusion described in Section 2. First of all, the framework is simple and clear from a mathematical point of view. Only a compact set of operation points is involved, and the Neyman-Pearson criterion is very beneficial for any biometric system. Besides, the optimization is not influenced by any score normalization, to which the ROCs are strictly invariant. Furthermore, the OR rule fusion is very suitable for many real world biometric applications, with outliers existent in the genuine class [12]. Basically, when the distributions of the genuine and impostor class are not symmetric, as is often true, the AND or OR decision fusion is very likely to fit because they have unsymmetrical support for the two classes.

The common criticism on decision-level fusion is that it has small and rigid information content. In the framework described in Section 2, however, the decision-level fusion has been adapted in such a way that the operation points are not fixed anymore, instead they are tunable and can be optimized with respect to performance. The disadvantage of decision-level fusion, nevertheless, is still the limited possibility of decision boundaries, because the operations are restricted to thresholding, AND, and OR.

This paper presents a new fusion scheme, combining the score-level and decision-level fusion under the general fusion framework described in Section 2. As the fusion framework is orientated on performance, we expect the final classifier to automatically alternate between the two levels of fusion in different situations, and achieve improved performance.

3.2. Hybrid Fusion Method

Under the general decision fusion framework, any two or more ROCs can be fused together. A biometric system, which has already been fused, can be easily put into this framework. This enables us to design a new hybrid biometric fusion scheme, combining score-level and decision-level fusion. Suppose the decision-level fusion can be expressed by

$$r_{\text{decision}} = D(r_1, ..., r_N) \tag{4}$$

where $r_1, ...r_N$ are the component ROCs to be fused, D is the decision fusion function, and r_{decision} is the resulting ROC. Similarly, suppose the score-level fusion is expressed by

$$r_{\text{score}} = S(r_1, ..., r_N) \tag{5}$$

where S is the score fusion function, and r_{score} is the resulting ROC. The hybrid fusion function H is defined as

$$H(r_1, ..., r_N) = D\left(r_1, ..., r_N, S_1, ..., S_M\right) \tag{6}$$

where $S_1, ..., S_M$ denotes the ROCs of different score-level fusion methods.

In Section 2, we have assumed independency between the component ROCs. In hybrid fusion, however, the assumption is not satisfied, as the inputs in (6), $r_1, ..., r_N$ and $S(r_1, ..., r_N)$, are dependent. Strictly speaking, we have to go back to the matching score space, and take into account the joint probabilities of the component matching scores. For example, suppose we are fusing two classifiers with matching scores s_1 and s_2, with the genuine score distribution $p(s_1, s_2|\omega_1)$, and the impostor score distribution $p(s_1, s_2|\omega_0)$. The optimization at decision level, in the Neyman-Pearson sense, is

$$\hat{p}_{\text{d}}(\alpha) = \max_{t_1, t_2} \left\{ \int_{t_1}^{\infty} \int_{t_2}^{\infty} p(s_1, s_2|\omega_1) \text{d}s_1 \text{d}s_2) \right\} \tag{7}$$
$$\text{subject to} \quad \int_{t_1}^{\infty} \int_{t_2}^{\infty} p(s_1, s_2|\omega_0) \text{d}s_1 \text{d}s_2 = \alpha$$

There are methods to solve (7), however, in practice we found that the independency assumption, i.e., solving (2) to obtain the thresholds corresponding to the optimal α_i's, is just adequate. The independency assumption might change the estimation of $\hat{p}_{\text{d}}(\alpha)$, but the thresholds t_1 and t_2 corresponding to its maximal value is often unchanged, or close enough to the real t_1 and t_2 under the dependent assumption. This is similar to the Naive Bayes problem [5], which also assumes independency between features, but whose optimality in dependency cases has been acknowledged in a wide range of applications [16][4]. Actually, we have observed that in many cases, the results from independency assumption is even better than the results from the dependency solutions. This can be explained by that fact that the optimization problem in (7) has much larger complexity than (2) and therefore more prone to overfit the solutions to the specific training set of matching scores.

Solving the hybrid fusion using the ROCs, instead of the matching scores, not only preserves the simplicity of the method, but also makes the solution more robust to the deviations between the training and testing scores. We summarize the hybrid fusion method as follows:

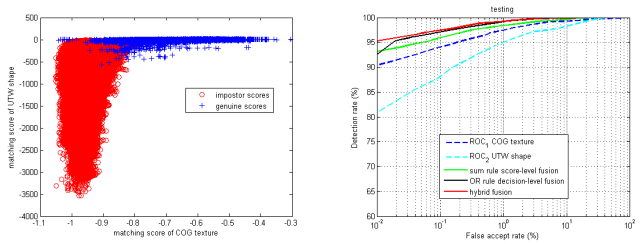
1. Given a set of component matching scores, and a set of score-level fusion methods.
2. (Training) Derive individual ROCs from the component matching scores and the score-level fused matching scores. Fuse all the ROCs under the fusion framework by the AND rule (2) or OR rule (3), and obtain the optimal combination of operation points.
3. Obtain the thresholds corresponding to those optimized operation points.
4. (Testing) Apply the trained thresholds on the component matching scores the score-level fused matching scores, and fuse the decisions by the AND rule or OR rule as the final decision.

4. Experiments and Results

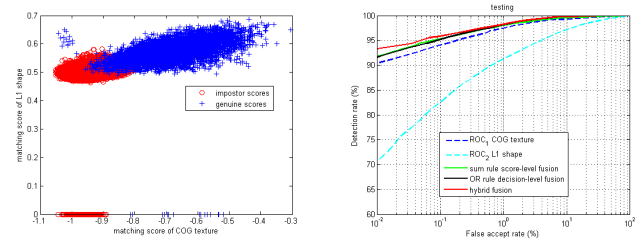
In this section, we present some experimental results of the proposed hybrid fusion. For the score-level fusion, we use the sum-rule, and preprocess the matchings by Z-normalization [10], which normalize the genuine scores to zero mean and unite variance. Many other score-level fusion methods could be inserted into the hybrid fusion, but in the preliminary experiments we only illustrate with Z-norm sum-rule score-level fusion, which is simple and robust. For the decision-level fusion, we use the OR rule, as in practice it is more suitable because of the outliers in the genuine class[1].

The first example is to combine the two-dimensional face texture and three-dimensional face shape information. The context of this work is EU FP6 3D-face project [1] which aims to combine two face modalities as a secure biometric for EU passports. The database that the algorithms are developed on is the FRGC database [13] which contains both 2D texture and 3D shape data. For either modality,

[1]There could also be outliers in the impostor class, but the outlier proportion in the genuine class is usually much higher. Generally the two opposite class are not balanced, either in size, or in distribution.

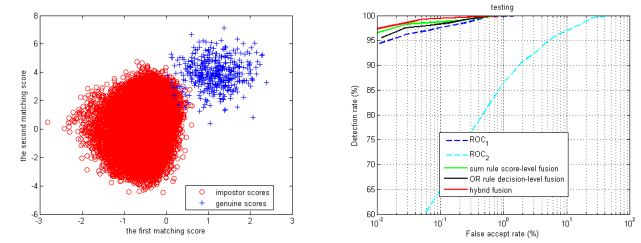


(a)

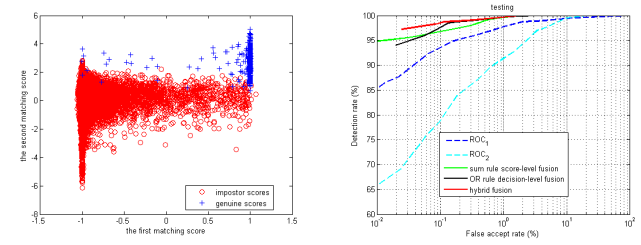


(b)

Figure 2. Example testing results of fusion between two face modalities, with matching scores from different institutes.



(a)



(b)

Figure 3. Example testing results of BA-fusion score database, with two typical type of score distributions.

the matching scores are derived by three algorithms, developed by the Cognitec Systems GmbH (COG), L-1 Identity Solutions (L1), and University of Twente (UTW), respectively.

The database contains data of 465 subjects and has in total 4,007 samples, with 2D texture data and 3D shape data collected simultaneously. The classifiers which 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 (fusion training data). The remaining 56 subjects are used for evaluation, resulting in 12,270 genuine scores and 700,910 impostor scores (fusion testing data). In the following experiments, we optimize the thresholds on the fusion training data, while evaluate the performance on the fusion testing data.

In Fig. 2, we give two examples of fusion between the 2D texture and 3D shape data. Both the scatterplot of the testing data and the fusion results on those data are shown. For comparison, we list the original ROCs, the sum rule fusion results, OR rule fusion results, and the hybrid fusion results. It can be observed that the hybrid fusion method outperforms both sum rule score-level fusion and OR rule decision-level fusion in both cases.

The second example is on the public database BA-fusion (Biometric Authentication Fusion Benchmark Database) [8] developed from the XM2VTS database [7], which contains the matching scores from face video and speech data. The matching scores are derived from various baseline systems (for details, see [8]). We show two examples with typical score distributions from the dataset, as in Fig. 3. Again we observed that the hybrid fusion method tunes the performance in such a way that it is always better than the score-level method or decision-level fusion methods.

The score-level fusion and decision-level fusion both have their advantages and fit different situations. For example, it can be observed that in Fig. 2 (a) the decision-level fusion is more beneficial, while in Fig. 2 (b) the decision-level fusion and score-level fusion have comparable performance. In Fig. 3 (a) sum rule fusion is more suitable, while in Fig. 3 (b), decision fusion and sum rule fusion fit different requirement of FARs. The hybrid fusion, which combines the two levels of fusion, adaptively tunes itself according to the different matching score distributions and specific performance requirements (i.e. prefix FAR or FRR). As can be observed, the final performance of hybrid fusion is improved over the better one, although sometimes with small margins due to the dependency.

Note that in both Fig. 2 and Fig. 3, the scatterplots are of the testing scores, different from the training scores on which the fusion is trained. In some cases, the improvement of the performance might also be accounted by the relative insensitivity of the ROC to overtraining, when a simple set of operation points are used to represent the original set of genuine and impostor training scores.

The hybrid fusion, therefore, is favorable in three senses, namely, adaptivity to different situations (alternating between the two levels of fusion), robustness to outliers, and relative insensitivity to deviations between the training and testing scores.

5. Conclusions

In this paper, we investigated a general fusion framework at decision level, by optimizing the operation points on the ROCs in the Neyman-Pearson sense. Under this framework, a hybrid fusion method is proposed, which combines the score-level fusion and the decision-level fusion, and takes advantage of both. Experiments show that in different cases, with different matching score distributions, the hybrid fusion method is able to adapt itself for improved performance over the two levels of fusion. More generally speaking, any fusion method could be integrated into this framework and optimized with respect to ROC, with improvements expected in the Neyman-Pearson sense.

A. Proof of Iterative Fusion

We show that the iterative fusion of two ROCs is optimal for the AND rule. The proof for the OR rule is similar.

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

$$\begin{aligned} p_{\mathrm{d}}^{\mathcal{I}}(\alpha) &= \max_{\alpha_i | \prod \alpha_i = \alpha} \prod_{i \in \mathcal{I}} p_{\mathrm{d},i}(\alpha_i), \\ p_{\mathrm{d}}^{\mathcal{J}}(\alpha) &= \max_{\alpha_j | \prod \alpha_j = \alpha} \prod_{j \in \mathcal{J}} p_{\mathrm{d},j}(\alpha_j) \end{aligned} \tag{8}$$

and

$$p_{\mathrm{d}}^{\mathcal{I}\mathcal{J}}(\alpha) = \max_{\alpha^{\mathcal{I}}\alpha^{\mathcal{J}} = \alpha} p_{\mathrm{d}}^{\mathcal{I}}(\alpha^{\mathcal{I}}) p_{\mathrm{d}}^{\mathcal{J}}(\alpha^{\mathcal{J}}). \tag{9}$$

First, expanding $p_{\mathrm{d}}^{\mathcal{I}\mathcal{J}}(\alpha)$ results in a product $\prod_{k=1}^{N} p_{\mathrm{d},k}(\alpha_k)$ for some α_k, $k = 1, \ldots, N$, satisfying $\prod_{k=1}^{N} \alpha_k = \alpha$. Therefore, we have

$$p_{\mathrm{d}}^{\mathcal{I}\mathcal{J}}(\alpha) \leq \max_{\prod_{k=1}^{N} \alpha_k = \alpha} \prod_{k=1}^{N} p_{\mathrm{d},k}(\alpha_k). \tag{10}$$

Second,

$$\begin{aligned} p_{\mathrm{d}}^{\mathcal{I}\mathcal{J}}(\alpha) &\geq p_{\mathrm{d}}^{\mathcal{I}}(\alpha^{\mathcal{I}}) p_{\mathrm{d}}^{\mathcal{J}}(\alpha^{\mathcal{J}}) \Big|_{\alpha^{\mathcal{I}}\alpha^{\mathcal{J}} = \alpha} \\ &\geq \prod_{i \in \mathcal{I}} p_{\mathrm{d},i}(\alpha_i) \Bigg|_{\prod \alpha_i = \alpha^{\mathcal{I}}} \prod_{j \in \mathcal{J}} p_{\mathrm{d},j}(\alpha_j) \Bigg|_{\prod \alpha_j = \alpha^{\mathcal{J}}} \\ &= \prod_{k=1}^{N} p_{\mathrm{d},k}(\alpha_k) \Bigg|_{\prod \alpha_k = \alpha} \\ &\geq \max_{\prod_{k=1}^{N} \alpha_k = \alpha} \prod_{k=1}^{N} p_{\mathrm{d},k}(\alpha_k). \end{aligned} \tag{11}$$

On combining (10) and (11) we have,

$$p_{\mathrm{d}}^{\mathcal{I}\mathcal{J}}(\alpha) = \max_{\prod_{k=1}^{N} \alpha_k = \alpha} \prod_{k=1}^{N} p_{\mathrm{d},k}(\alpha_k). \tag{12}$$

This means that *if the optimal ROCs are known for disjoint subsets, the overall optimal ROC can be found by optimally fusing the subsets*.

References

[1] 3D Face. 3D Face biometric research. `http://www.3dface.org/`, 2006.

[2] S. C. Dass, K. Nandakumar, and A. K. Jain. A principled approach to score level fusion in multimodal biometric systems. In *Audio- and Video-Based Biometric Person Authentication*, pages 1049–1058, 2005.

[3] J. Daugman. Combining multiple biometrics. `http://www.cl.cam.ac.uk/users/jgd1000/combine/combine.html`, 2000.

[4] P. Domingos and M. Pazzani. Beyond independence: Conditions for the optimality of the simple bayesian classifier. In *13th Internat. Conf. on Machine Learning*, 1996.

[5] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification (2nd ed.)*. John Wiley and Sons, New York, 2001.

[6] 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.

[7] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre;. Xm2vtsbd: The extended m2vts database. In *2nd Conference on Audio and Video-base Biometric Personal Verification*, 1999.

[8] N. Poh and S. Bengio. Database, protocols and tools for evaluating score-level fusion algorithms in biometric authentication. *Pattern Recognition*, 39(2):223–233, 2006.

[9] A. Ross and A. Jain. Information fusion in biometrics. *Pattern Recognition Letters*, 24(13), 2003.

[10] A. Ross, K. Nandakumar, and A. Jain. *Handbook of Multibiomtrics*. Springer Publishers, 2006.

[11] C. Sanderson and K. Paliwal. Information fusion and person verification using speech and face information. Technical report, IDIAP, Switzerland, September 2002.

[12] Q. Tao and R. Veldhuis. Optimal decision fusion for a face verification system. In *the 2nd International Conference on Biometrics*, pages 958–967, Seoul, Korea, 2007.

[13] FRGC. Frgc face database. `http://face.nist.gov/frgc/`.

[14] H. van Trees. *Detectioin, Estimation, and Modulation Theory*. John Wiley and Sons, New York, 1969.

[15] Y. Wang, T. Tan, and A. K. Jain. Combining face and iris biometrics for identity verification. In *Fourth International Conference on AVBPA*, pages 805–813, 2003.

[16] H. Zhang. The optimality of naive bayes. In *17th Internat. FLAIRS Conf.*, 2004.