Multiple-sample Fusion of Matching Scores in Biometric Systems

Xi Cheng, Sergey Tulyakov, and Venu Govindaraju
Center for Unified Biometrics and Sensors
University at Buffalo, NY, USA
xicheng,tulyakov,govind@buffalo.edu

Abstract

If a biometric matching attempt does not succeed, the person might be asked to repeat authentication attempt second time. In such situations, the biometric system has acquired two test templates, and could construct a combined matching score, for example, by averaging two scores from the matches between these templates and the enrolled template. This paper investigates the enhancement of combination algorithm performance by utilizing the third score - the matching score between the two test templates. Two kinds of combination strategies are considered, linear combination and likelihood ratio. The results show, that using the scores between test templates can improve the performance of both linear and likelihood ratio combination methods. We illustrate the theory by experiments performed on four FVC2002 fingerprint databases.

1. Introduction

It is well known that the combination of multiple biometric matching results can lead to superior biometric system performance [9]. Typically, multibiometric systems consider the combinations of the results produced by the different sensors or by the different matching algorithms. The construction of such systems follows the general principle that the greater diversity of available information leads to better recognition results [11]. For example, multimodal biometric systems, which incorporate sensors for different biometric modalities, show definite increase in matching performance compared to unimodal systems, and such systems were the main focus of the research in biometric systems fusion methods [16].

The lesser researched topic, which we will investigate in this paper, is the combination of matching results obtained from repeated scans of the same biometric modality, multi-sample or multi-instance biometric fusion. Suppose, our biometric system consists of a single sensor and a single matching algorithm. If a person needs to be authenticated, we perform the first biometric scan and match the scanned test (or probe) template to the enrolled template in the biometric database. If the matching score is good enough, we accept the person, and if not, we reject. In case of rejection, the person might try to authenticate the second time by allowing the second sensor scan and presenting second test template to the matching algorithm. The second matching score could be used for acceptance decision this time. Instead of using single second matching score for acceptance decision, it is possible to combine both matching scores and possibly increase the performance of biometric systems.

The intuitive solution for the fusion of multi-sample scores is to apply some fixed combination rule, e.g. simple sum rule. Indeed, since the pairs of scores in both types of verification trials, genuine and impostor, are obtained by using prints of the same fingers, then it is reasonable to assume the normal distributions of genuine and impostor pairs of scores with the same correlation matrices. Under this assumption, the optimal decision (ratio between two normal densities) will be linear and coincide with simple sum rule, if we assume the symmetry among scores as well. If these assumptions do not hold, then might try to use an optimal trainable algorithm, e.g. likelihood ratio [14], and try to get better performance than simple sum rule. As a matter of fact, we tried to compare likelihood ratio and simple sum rule in our experiments, but the results were inconclusive.

We can notice that in the situations dealing with multi-sample trials an additional matching score, the matching score between two test templates, can be obtained (Fig. 1). This score can bear additional information, for example related to the quality of test templates or their diversity, and this score can be used during the combination of matching scores between enrolled and test templates. In this paper we try to improve the combination algorithms by utilizing this third score. The experiments, performed using fingerprint images from FVC2002 database, show consistent improvements resulting from utilizing the matching score between test templates.

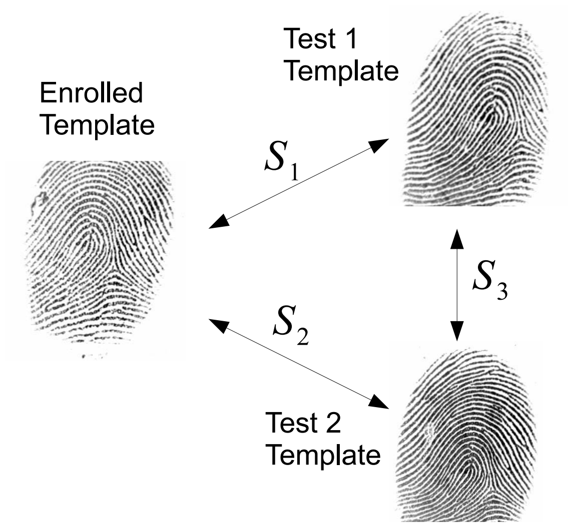


Figure 1. In addition to using two matching scores, s_1 and s_2, between test templates and enrolled template, we utilize the matching score s_3 between two test templates.

2. Previous Work

The usual approach to utilize the information contained in multiple instances of the same biometric is to construct a single more reliable template which incorporates the information of these instances. Such approaches include the *image-level* or *feature-level* fusion methods and the fused template can be called *super-template* [5, 6, 7].

Toh *et al*. [7] reduce false rejection in fingerprint matching systems by setting one template as background and updating minutiae points from other templates. Several transformation models are used and compared for fingerprint image alignment. They claim that the algorithm not only reduces false rejection but also has low computational cost and low memory storage compared to existing mosaicking techniques.

In order to define the spatial relationship between two templates, a transformation matrix is calculated by Jain *et al*. [5] to use a modified iterative closest point algorithm. Augmented minutiae sets are extracted after a composite image of two fingerprint images is constructed. It shows that mosaicking two templates together can improve the performance of matching systems compared to utilizing only one template.

Composition of multiple templates is generated by using successive Bayesian estimation in Ryu *et al*. [6].The composed image is called super-template which is updated sequentially. It shows that by utilizing the successive Bayesian estimation method, the algorithm can handle unlimited number of templates so that more inputs are considered, better error reduction will be achieved.

The feature level fusion could be important step if the extraction of the reliable template is required. For example, fuzzy fingerprint vault method requires the exact minutia positions as input. Such positions can be calculated as the averages of minutia positions in few instances of the fingerprint, and the performance of fuzzy vault matching can be improved [8].

Other biometric modalities could also employ some algorithms for feature-level fusion and the creation of the super-template. The more precise model of the face can obtained from the sequence of video images [10, 12] than from a single image. The handwritten signature verification methods typically construct an enrolled template, e.g. as a trained HMM, from the set of few signatures [15].

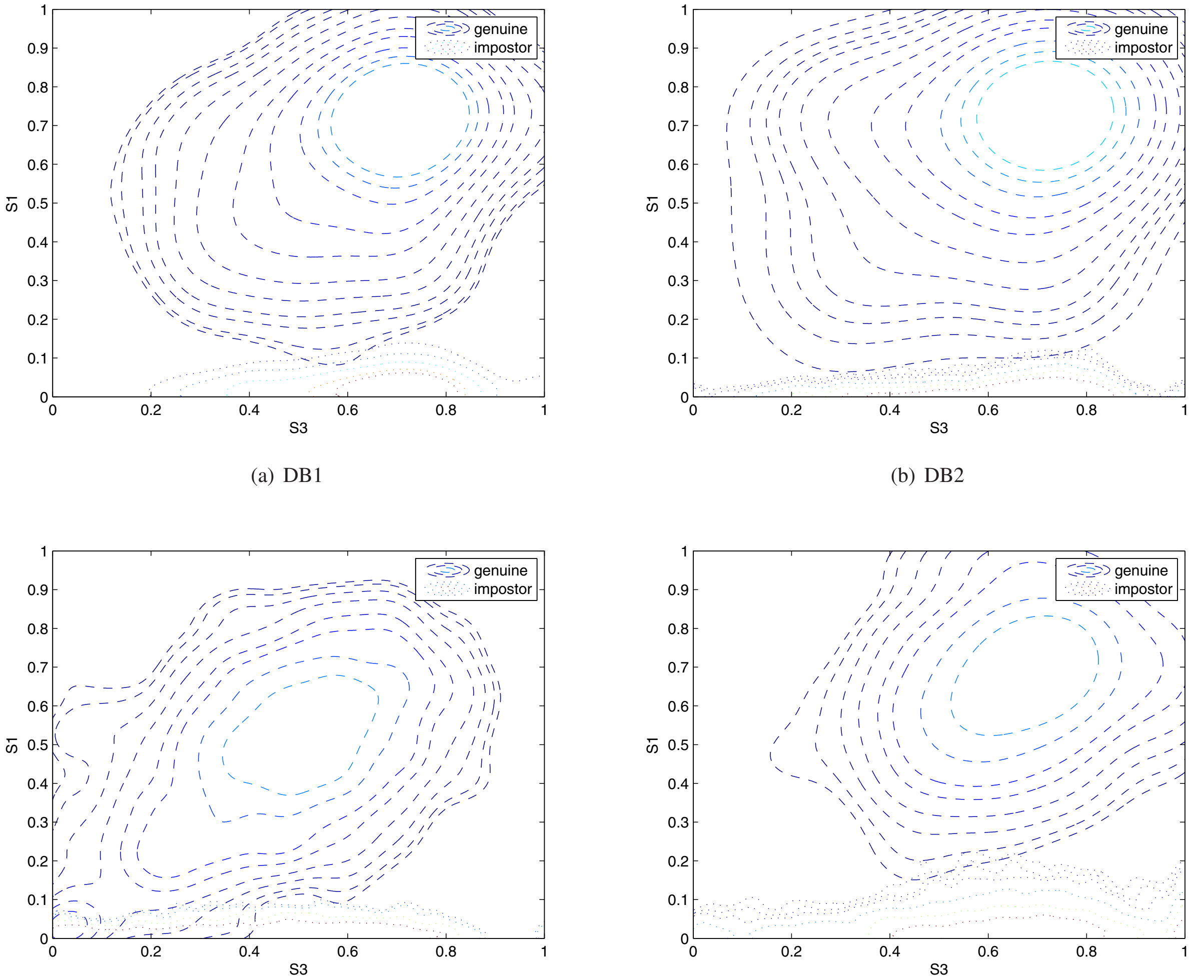
Some methods attempt to use the matching scores from multiple instances directly instead of performing feature-level fusion of constructing super-templates. In a video base face recognition, Chellappa et al. [3] extract a sequence of frames from a clip of video and each frame is matched with a stored template of the person, resulting in a sequence of matching scores, and the scores are combined by updating the conditional entropy. Similar approach is explored by Zhang and Martnez [18] with simple averaging used for score fusion. Cheung et al. [13] explored some methods in combining matching scores for speaker verification including weighted sum of the scores.

Fusion at the feature level seems to be more effective than the fusion at a later stage, matching score level. Indeed, features contain more information about the biometric data than matching scores. For example, multiple samples of a particular feature provide better estimation of this feature's distribution than a single sample, and such information could be successfully used during matching of a constructed super-template to some other template. But fusion at feature level is more difficult to achieve compared to matching score fusion - we need access to raw data or features, and fused features should correspond to each other. Besides, the feature fusion algorithms are specific for each biometric modality.

Even if the commercial biometric matchers do not provide features for public users [4], the fusion at the score level will still be feasible for such matchers. Additionally, same combination methods will be applicable to the matchers of different modalities. In our research, we propose to use matching scores for fusion. Although not as powerful as feature level fusion, our method also takes into consideration the consistency of the test templates by exploiting the statistical dependence between matching scores.

3. Experimental Setup

In this paper, we only consider the combination scenario shown in Fig. 1 - one enrolled and two input templates. One verification attempt consists of three matching scores. Two scores are matching scores between the enrolled template and the two input templates - s_1 and s_2. The third score



(a) DB1 (b) DB2

(c) DB3 (d) DB4

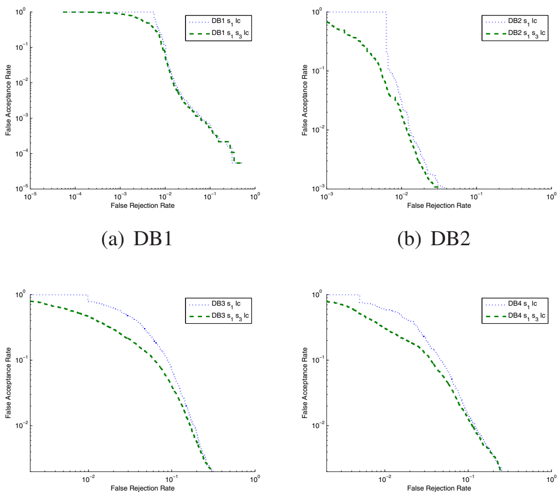
Figure 2. probability densities for genuine and impostor matchings in FVC2002 DB1 DB2 DB3 and DB4

is from the matching between the two input templates s_3. There are two types of verification attempts, genuine and impostor. The genuine verification attempt is when the enrolled and the two input templates come from the same fingerprint. So s_1,s_2 and s_3 are all genuine matching scores. The impostor verification attempt is when the enrolled template comes from one fingerprint and the two input templates are obtained from another fingerprint. So s_3 is a genuine score, s_1 and s_2 are both impostor scores.

The fingerprint matching system which we use in this paper is proposed in [2]. Localized distances and angles between nearby minutiae are calculated and used to overcome the conventional methods for partial fingerprint matching. Minutia matching problem between two fingerprints is converted to minimum cost flow problem which gives an efficient way to achieve optimal matching.

We use the FVC2002 [1]DB1, DB2, DB3 and DB4 fingerprint datasets. Each database contains 110 different persons with eight different images for the same fingerprint of one person. We generate the genuine matching scores by assuming one of eight images for the same fingerprint to be enrolled and two others to be test templates. For each person, there are 168 such variations, and each variation represents one verification attempt of Fig. 1 with all of the scores s_1,s_2 and s_3 being genuine. The impostor scores are generated by using one image from one person and the two test templates from another one. The scores between the enrolled image and two test templates are impostor scores but the score between the two test templates is genuine.

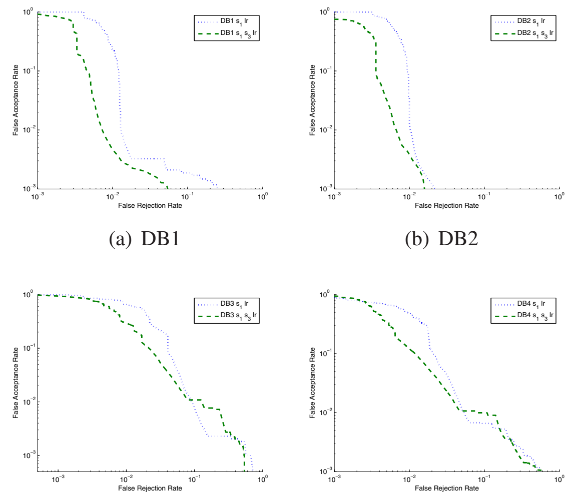
For the likelihood ratio combination method, bootstrap sample testing technique is used. For each bootstrap step, twenty-five persons are selected as training set and another twenty-five ones for validation set. The remaining sixty persons are used for testing. In the training process, the kernel



(a) DB1 (b) DB2

(c) DB3 (d) DB4

Figure 3. ROC curves for verification decisions of single score s_1 and two scores s_1 and s_3 utilizing linear combination in FVC2002 DB1 DB2 DB3 and DB4



(a) DB1 (b) DB2

(c) DB3 (d) DB4

Figure 4. ROC curves for verification decisions of single score s_1 and two scores s_1 and s_3 utilizing likelihood ratio in FVC2002 DB1 DB2 DB3 and DB4

sizes of the likelihood ratio methods are estimated by using the validation sets.

4. Matching Score Dependencies

In order to fully understand the role of additional matching score s_3 in our system, we performed a series of experiments on determining the existing matching score dependencies. Specifically, we explored the situation where scores s_1 and s_3 are available for combination, but not score s_2. For genuine verification attempts, both scores are genuine, and for impostor identification attempts, s_1 is impostor and s_3 is genuine. The question, which we want to answer, is whether using both scores, s_1 and s_3, results in better separation of genuine and impostor trials than using only single score s_1.

Fig. 2 contains the contour curves of the approximated densities of the score pair samples $\{s_1, s_3\}$ of two classes - genuine and impostor. Note, that for all datasets we have a slight positive correlation between s_1 and s_3 for both genuine and impostor classes, and, for example, the densities of genuine pairs are somewhat elongated along the line $s_1 = s_3$. Thus, the decision surfaces for the optimal algorithm separating genuines from impostors will probably be different from the horizontal lines, $s_1 = c$, which are decision surfaces for the classification algorithm utilizing single score s_1.

Our first experiment on measuring the effect of additional score s_3 on classifying genuine and impostor matching attempts was linear combination. We compare the performance by using one matching score s_1 and two scores $s_1 - a * s_3$. We choose $s_1 - a * s_3$ because the separation boundary for genuine and impostor scores appears to be close to such linear function with parameter a somewhere between 0 and 1. The parameter a was chosen to minimize the equal error rate, and its value was slightly different for four datasets: 0.10 for DB1, 0.12 for DB2, 0.14 for DB3 and 0.12 for DB4.

In the second experiment we used the Bayesian classification with the score densities approximated by the Parzen window method with Gaussian kernels. When using a single score s_1, the following combination formula was used

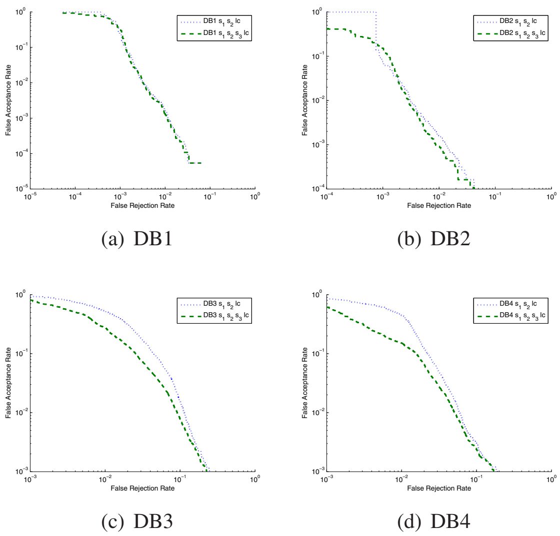
$$S = \frac{p_{gen}(s_1)}{p_{imp}(s_1)} \tag{1}$$

where $p_{gen}(s_1)$ is the probability density of genuine scores and $p_{imp}(s_1)$ is the probability density of impostor scores in s_1. The performance of utilizing only s_1 will be compared to using both s_1 and s_3 which has following combination formula

$$S = \frac{p_{gen}(s_1, s_3)}{p_{imp}(s_1, s_3)} \tag{2}$$

where $p_{gen}(s_1, s_3)$ is the joint probability density of genuine scores and $p_{imp}(s_1, s_3)$ is the joint probability density of impostor scores in s_1 and s_3. Two dimensional Parzen window approximation is performed in this case for score pair densities.

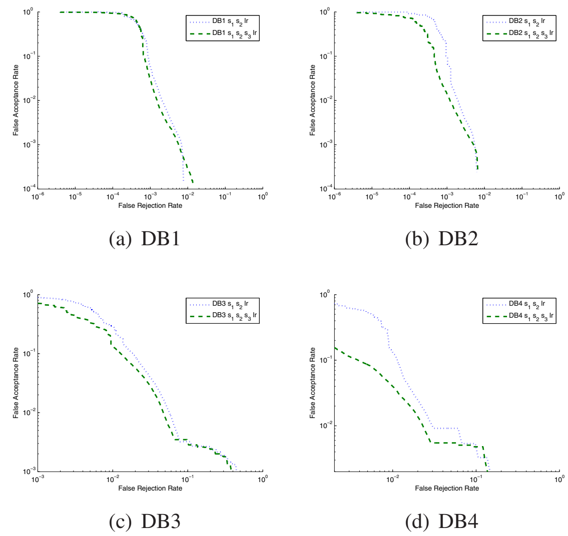
The results of the experiments are shown in Fig. 3 for linear combinations and in Fig. 4 for likelihood ratio combinations. Utilizing the score between test templates seems to produce better results than just using matching scores between enrolled and test templates in most cases.



(a) DB1 (b) DB2

(c) DB3 (d) DB4

Figure 5. ROC curves for verification decisions of two scores s_1 and s_2 and three scores s_1,s_2 and s_3 utilizing linear combination in FVC2002 DB1 DB2 DB3 and DB4



(a) DB1 (b) DB2

(c) DB3 (d) DB4

Figure 6. ROC curves for verification decisions of two scores s_1 and s_2 and three scores s_1,s_2 and s_3 utilizing likelihood ratio in FVC2002 DB1 DB2 DB3 and DB4

5. Combination Experiments

The goal of our research is to perform the combination of scores s_1 and s_2. As we saw in the previous section, the use of additional s_3 can improve the decision making based on score s_1. The question here is whether similar improvement will be observed in the combination algorithm.

The verification trial scenario of Fig. 1 seems to be symmetric with regards to scores s_1 and s_2. Thus, a simple sum $(s_1 + s_2)/2$ seems to be the optimal linear combination of these two scores. Similarly, due to symmetry conditions, we use the formula $(s_1 - a * s_3 + s_2 - a * s_3)/2$ for the linear combination involving the third score s_3. The coefficients a were chosen to be the same as in section 4.

But linear combination is not expected to deliver the best possible performance. For example, it does not take into account the possibility that some input templates have poor quality biometric measurement. For example, in the video scenario, the error of one corrupted frame will be repeated in other frames, and the fusion result will be erroneous as well.

Generally, the distributions of genuine and impostor score tuples will be quite arbitrary and linear functions might not provide best possible combinations. Therefore, we use the likelihood ratio combination method, which is known to be the optimal combination method in verification systems [14]. The likelihood ratio combination of s_1 and s_2 has the following form

$$S = \frac{p_{gen}(s_1, s_2)}{p_{imp}(s_1, s_2)} \tag{3}$$

To use the matching score between the two input templates, s_3, we adjust the likelihood ratio to use three dimensional probability densities for the score triplets $\{s_1, s_2, s_3\}$:

$$S = \frac{p_{gen}(s_1, s_2, s_3)}{p_{imp}(s_1, s_2, s_3)} \tag{4}$$

The results of experiments are given in Fig. 5 for linear combinations and Fig. 6 for likelihood ratio combinations. Similar to results of the previous section, utilizing the score between test templates seems to improve the combination results in most cases.

6. Conclusions

Utilizing the scores between test templates can improve the accuracy of biometric systems. In this paper, we showed two methods to use the score between test templates - linear combination and likelihood ratio. In both methods, results show that using the scores between test templates can improve the performance.

The observed improvement appears to be the result of the dependence between matching scores. The dependence is present in both genuine and impostor tuples of scores $\{s_1, s_2, s_3\}$. The existence of the matching score dependencies can be explained by the fact that same templates serve for score derivation. For example, first test template is used to derive scores s_1 and s_3, which results in the positive correlations of s_1 and s_3. The existence of similar dependencies between matching scores in NIST BSSR1 database has been noted before [17], and therefore we can expect our algorithm to work for other biometric matchers also.

We can think of the additional score s_3 as a sample of a typical genuine score for particular person being identified. If the matching score score s_1 between test and enrolled templates is comparable to s_3, then we are more certain that this is genuine verification attempt, and impostor otherwise. Whereas considering the correspondences of features in feature-level fusion or super-template construction can provide more useful information for the fusion, utilizing only a single matching score between test templates in presented approach delivers a performance boost as well. In contrast to feature-level fusion, the current method does not depend on specifics of biometric template construction, and can be applicable to many biometric modalities.

We do not have specific information on how the used fingerprint database (FVC 2002) was generated and whether the fingerprint scans were obtained at the same time. Thus, the real life application scenario (two test fingerprints are obtained at the same time and one enrolled fingerprint is obtained at another time) could have dependence properties different from our experiments and some adjustments or retraining of the algorithm might be needed.

It is possible to extend the proposed methods to more general scenarios, e.g. more than two test templates, two or more enrolled templates. In particular, in the task of matching a single enrolled face image to the video of the person, the number of video frames normally is more than two. It seems that our research could be beneficial in this area - the combination algorithm will derive additional matching scores between different video frames, and use such scores for combinations.

References

[1] Fingerprint verification competition. 2006. http://bias.csr.unibo.it/fvc2002/. 139

[2] Anonymous. Reference removed for review. 139

[3] R. Chellappa, V. Kruger, and Z. Shaohua. Probabilistic recognition of human faces from video. In *Image Processing. 2002. Proceedings. 2002 International Conference on*, volume 1, pages I–41–I–44 vol.1, 2002. 138

[4] A. Jain, K. Nandakumar, and A. Ross. Score normalization in multimodal biometric systems. *Pattern Recogn.*, 38(12):2270–2285, 2005. 1746649. 138

[5] A. Jain and A. Ross. Fingerprint mosaicking. In *Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., 1993 IEEE International Conference on*, volume 4, pages IV–IV, 2002. 138

[6] T. Kanade, A. Jain, N. Ratha, C. Ryu, Y. Han, and H. Kim. Super-template generation using successive bayesian estimation for fingerprint enrollment. In *Audio- and Video-Based Biometric Person Authentication*, volume 3546 of *Lecture Notes in Computer Science*, pages 261–277. Springer Berlin / Heidelberg, 2005. 138

[7] T. Kar-Ann, Y. Wei-Yun, J. Xudong, C. Tai-Pang, L. Juwei, and L. Eyung. Minutiae data synthesis for fingerprint identification applications. 3:262–265 vol.3, 2001. 138

[8] E. Kelkboom, J. Breebaart, R. Veldhuis, X. Zhou, and C. Busch. Multi-sample fusion with template protection. In *BIOSIG 2009: Proceedings of the Special Interest Group on Biometrics and Electronic Signatures*, Darmstadt, Germany, 2009. 138

[9] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas. On combining classifiers. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(3):226–239, 1998. 137

[10] L. Kuang-Chih, J. Ho, Y. Ming-Hsuan, and D. Kriegman. Video-based face recognition using probabilistic appearance manifolds. 1:I–313–I–320 vol.1, 2003. 138

[11] L. I. Kuncheva, C. J. Whitaker, C. A. Shipp, and R. P. W. Duin. Is independence good for combining classifiers? In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, volume 2, pages 168–171 vol.2, 2000. 137

[12] Y. Li, S. Gong, and H. Liddell. Constructing facial identity surfaces for recognition. *International Journal of Computer Vision*, 53(1):71–92, 2003. 138

[13] C. Ming-Cheung, M. Man-Wai, and K. Sun-Yuan. Multi-sample data-dependent fusion of sorted score sequences for biometric verification. In *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on*, volume 5, pages V–681–4 vol.5, 2004. 138

[14] K. Nandakumar, C. Yi, S. C. Dass, and A. K. Jain. Likelihood ratio-based biometric score fusion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(2):342–347, 2008. 137, 141

[15] G. Rigoll and A. Kosmala. A systematic comparison between on-line and off-line methods for signature verification with hidden markov models. In *Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on*, volume 2, pages 1755–1757 vol.2, 1998. 138

[16] A. A. Ross, K. Nandakumar, and A. K. Jain. *Handbook of Multibiometrics*. International Series on Biometrics. Springer-Verlag New York, Inc., 2006. 137

[17] S. Tulyakov and V. Govindaraju. Use of identification trial statistics for the combination of biometric matchers. *Information Forensics and Security, IEEE Transactions on*, 3(4):719–733, 2008. 141

[18] Y. Zhang and A. M. Martnez. A weighted probabilistic approach to face recognition from multiple images and video sequences. *Image and Vision Computing*, 24(6):626–638, 2006. doi: DOI: 10.1016/j.imavis.2005.08.004. 138