Combination of Multiple Samples Utilizing Identification Model 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

In some cases, the test person might be asked to provide another authentication attempt besides the first one so that combination of the two input templates might give the system more confidence if the person is genuine or impostor. Instead of simply combining the matching scores which are associated with a single person compared to the two input templates, we investigate the use of matching scores corresponding to all enrolled persons. The dependencies between scores generated by the same input templates are accounted for the proposed combination algorithm. Such combination methods can be extended to large number of classes and input templates. Since matching scores are used, the proposed methods can also be applied on arbitrary biometric modalities. The experiments are conducted on NIST BSSR1 face and FVC2002 fingerprint datasets by using both likelihood ratio and multilayer perceptron combination methods.

1. Introduction

Generally, multimodal biometric system refers to utilization of more than one biometric measurements to improve the performance of biometric systems, for example, multiple independent or weakly correlated matchers (multibiometric system), multiple representations of faces, different matching algorithms of fingerprints, multiple faces from one video clip and so on. Multibiometric systems which are extensively studied show that superior biometric system performance can be achieved according to the combinations of multiple sensors or matchers [7]. Combination of multibiometric systems is due to the great diversity of different sensors or matching algorithms which lead to better recognition results compared to unimodal systems [10, 13].

What we want to investigate is the combination of matching results generated from repeated samples of the same biometric modality. This is actually corresponding to real scenarios, for example, video-based face matching requires fusion of matching results of multiple sequential frames. Another example is that for fingerprint verification system, if the first scan of the fingerprint is not good enough, the system might ask the person to provide another scan of the same fingerprint. Instead of using just one template to make acceptance or rejection decision, it is possible to increase the verification or identification accuracy of the biometric systems to combine two correlated templates.

Matching score is the measurement of distance between an enrolled template T_e and a test template T_t. The quality of the templates might influence the matching scores. For example, if the quality of a fingerprint image is good, the matching scores between this template and all enrolled ones might be high and the genuine score is more probably the best score in the score set which contains one genuine score and all other impostor scores. On the other side, if the image quality is bad, the matching scores in the score set will be low and the genuine score might be somewhere among the impostor scores. In [9], *user quality* is used to measure the confidence of user biometric samples in order to improve the performance for hand-based biometric systems during matching process. Instead of using genuine matching scores to estimate the quality of the biometric template, we use some statistic information from matching scores according to one test template and all stored templates so that it can be applied to large datasets.

The combination solutions can be some predetermined methods, *e.g.* summation rule. The problem can be derived to be a two-class classification, genuine and impostor. If the probability densities of genuine and impostor are not normal distributions and even more complex, the linear classification might not optimal and other trainable algorithms should be used. As a matter of fact, we use likelihood ratio and multilayer perceptron in our experiments to combine the matching results of multiple samples.

978-1-4577-1359-0/11/$26.00

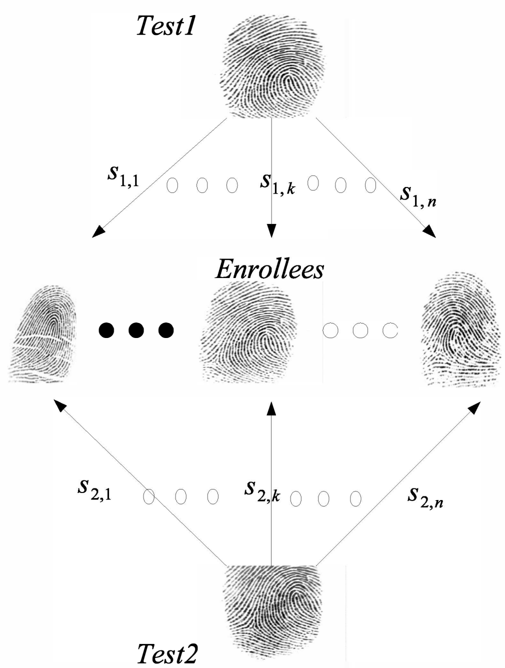


Figure 1. The sets of matching scores available for combinations for two test templates. In addition to using two matching scores, $s_{1,k}$ and $s_{2,k}$, between test templates and enrolled template for enrollee k, we want to utilize the entire set of matching scores $\{s_{i,j}\}$ between two test templates and all enrolled templates.

2. Previous Work

Multi-sample fusion can be applied at feature-, score-, and decision-level. Feature level fusion seems to be more effective than fusion at later stages, score and decision levels because features contain more information about the specific biometric data. The distribution of one feature can be estimated accurately by fusing the same features in multiple related samples. Such information can be used to generate a composite- or super-template which can successfully perform better than a single template since it's more reliable [4, 6]. On the other hand, fusion at the feature level requires that we need to access the raw data or features. And feature fusion algorithms are specific for different biometric modalities, for instance, face and fingerprint [4, 8].

Ryu *et al.* [6] generate a *super-template* from multiple templates in fingerprint biometric systems by using successive Bayesian estimation. The super-template is improved in the template fusion process in which the credibility of every minutia is updated based on the minutiae in the current input image. The results show that better accuracy is obtained when more impressions are used for fusion. They also claim that the algorithm can handle unlimited number of templates.

Jain *et al.* [4] use a modified iterative closest point algorithm to compute a transformation matrix in which the spatial relationship between the two input templates is defined. After input templates are mosaicked, augmented minutiae sets are extracted resulting in a better performance compared to utilizing only one template in fingerprint matching systems.

Compared to feature level fusion, score level fusion provides us a way to apply same algorithms on different biometric modalities since biometric matchers are treated as black boxes. As long as we have required matching scores, we don't have to know the details of feature extraction and matching algorithms. Especially in some cases where commercial biometric matchers hide features from public users, scores are the only available resources to use.

A sequence of frames are extracted from a clip of video and each frame is matched with an enrolled template of the person in [3]. The resulting matching scores are combined by updating the conditional entropy which captures the evolving uncertainty of the identity variable given observations. Similarly, Zhang and Martnez [16] use a weighted probabilistic method for score fusion in face recognition systems.

In this paper, we will consider matching scores as well as the dependencies between related scores. In verification system, the score between the test template and one stored template is used to authenticate the person. But as Fig. 1 shows, each input template can be compared to all enrolled templates to generate a set of scores, $s_{i,1}, \ldots, s_{i,k}, \ldots, s_{i,n}$ for test template i where n is the number of enrolled templates. Since the set of scores is achieved during one identification trial, we call the score set as *identification trial score set* and the algorithm utilizing the statistical information of this set to be *identification model.*

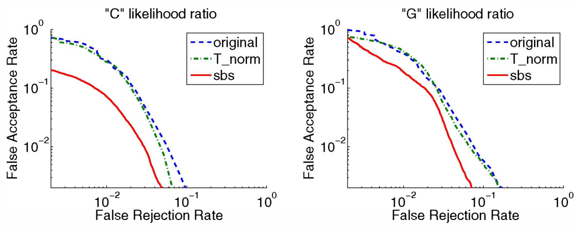
3. Identification Model

Identification model utilizes the identification trial scores which have dependencies between them because the same test template is used for their derivation. One example of using score sets is to convert raw scores to ranks. Ho *et al.* [14] claim that utilization of ranks is more powerful than raw scores in character recognization system.

One kind of identification model we want to use is T-normalization [11] which uses the mean μ and standard deviation σ from the set of scores. Each score will be normalized using such formula:

$$s \rightarrow \frac{s - \mu}{\sigma} \tag{1}$$

After T-normalization, each set of score will have mean to



(a) C (b) G

Figure 2. ROC curves for utilizing likelihood ratio method in NIST BSSR1 C G datasets

be 0 and variance to be 1. If the matching scores in different identification trials have only those variations, the T-normalization successfully accounts for score dependencies. Another identification model is called second best score model which has such formula:

$$s \to \{s, second\_best\_score(s)\} \tag{2}$$

where $second\_best\_score(s)$ is the best score besides the current score s in the identification trial. This model can be understood like this: if the second best score (best score besides current one) is higher than the current score s, we have more confidence that the current score is probably an impostor score and the second best sore is the genuine one. Reverse is true that if the second best score is lower than the current one, the current score is more probably the genuine score. Tulyakov *et al.* [15] applied identification models in multimodal biometric systems in which both T-normalization and second best score model provides better performance than utilization of original matching scores. Similar to identification model, the background model also utilizes the scores among templates. The difference is that the background model uses selective enrolled templates according to one enrolled template instead of all enrolled ones. One implementation of background model, the cohort based method which is to find a subset of enrolled templates close to a particular enrolled one under some considerations has been used for fingerprint verification [1].

4. Combination Rules

The multi-sample fusion algorithms used are likelihood ratio method and multilayer perceptron. Likelihood ratio is theoretically optimal combination method for verification systems [12]. For T-normalization, the equation is:

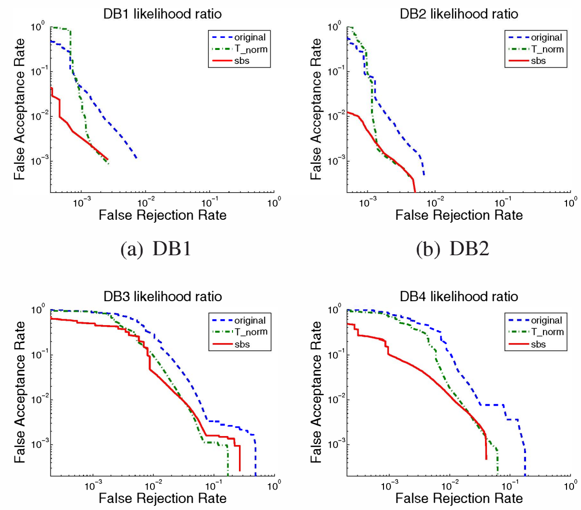
$$S = \frac{p_{gen}(s_1, s_2)}{p_{imp}(s_1, s_2)} \tag{3}$$

where s_i is the matching score from test template i and normalized by using formula 1, $p_{gen}(s_1, s_2)$ is the probability density of genuine scores s_1 and s_2 and $p_{imp}(s_1, s_2)$ is the probability density of impostor scores. Likelihood ratio assigns the combined score a value of ratio between genuine and impostor score densities. In order to use the second best score, four dimensional densities for genuine and impostor scores are constructed:

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

where s_i is the same as before, $sbs(s_i)$ is the second best score according to s_i in the same identification trial.

Multilayer perceptron is used to compare the identification models since direct approximation of score densities might be problematic in high dimensional space. The perceptron has two hidden layers with eight nodes in the first hidden layer and nine nodes in the second hidden layer. The input layer for traditional method and T-normalization contains two nodes for each matching score form each input template. The difference is that the scores are normalized in T-normalization. The input layer for second best score model has four nodes for two original scores and two second best scores $s_1, sbs(s_1), s_2, sbs(s_2)$ from the two input template. The output layer has one node with 0 standing for impostor matching and 1 for genuine matching.

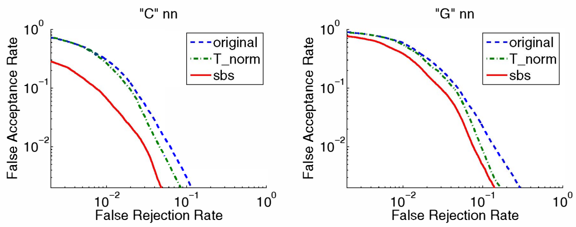


(a) DB1 (b) DB2 (c) DB3 (d) DB4

Figure 3. ROC curves for utilizing likelihood ratio method in FVC2002 DB1 DB2 DB3 and DB4

5. Experiment

We use both NIST BSSR1 dataset and FVC2002 four fingerprint datasets. For NIST BSSR1 dataset, we only use the set3 data from two face matchers 'C' and 'G' which run on images from 3000 individuals. The set contains one score from the comparison of face A with a later face, B, and a score from face A and another later face, C. Due to



(a) C (b) G

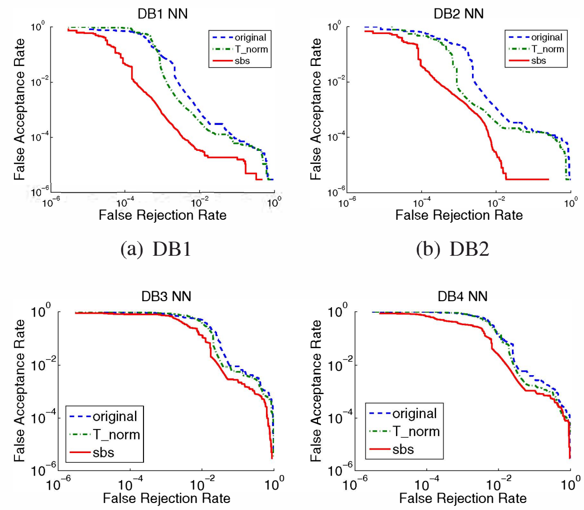
Figure 4. ROC curves for utilizing multilayer perceptron in NIST BSSR1 C G datasets

data collection errors, for instance, all match scores for one user or enrolled template being 0, some scores have to be discarded to get 2991 different enrollees and 5982 input templates where two faces are from the same person out of 2991 different persons. So each input-wise score set of the score matrix is an identification trial score set and there is only one genuine score and others are impostor scores.

FVC2002 has four datasets DB1, DB2, DB3 and DB4. Each dataset has 110 different persons with 8 different images for the same fingerprint of the person. The genuine matching case is generated by assuming one of eight images for the same fingerprint to be enrolled and other two to be tested. For each person, there are 56 such variations. Other scores used in the one identification trial score set are achieved by matching the test template to the first fingerprint image of other 109 persons. The impostor matching case is generated by matching one image from one person as the enrolled template and two test templates from another person. So that the score between the two test templates is genuine and the scores between the enrollee and two test templates are impostor scores. Other scores in the identification score set are generated as in genuine matching case.

Since FVC2002 datasets have fingerprint images, we use the fingerprint matching system proposed in [5] where 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.

For both likelihood ratio and multilayer perceptron, bootstrap sample testing technique is used [2]. In each bootstrap step of utilizing each FVC2002 dataset, twenty-five persons are selected as training set and another twenty-five ones for validation set. So the remaining sixty persons are used for testing. For each step of using NIST BSSR1 datasets, 800 persons are selected randomly for training and 800 for validation. The left 1391 persons are used for testing. We perform 100 bootstrap tests in each experiment. In likelihood ratio method, we use Parzen window with Gaussian kernels whose width is estimated by the maximum likelihood method. The activation function in multilayer perceptron is the Sigmoid function.



(a) DB1 (b) DB2 (c) DB3 (d) DB4

Figure 5. ROC curves for utilizing multilayer perceptron in FVC2002 DB1 DB2 DB3 and DB4

Fig. 2 and Fig. 3 show the ROC curves for NIST BSSR1 two face datasets and FVC2002 four fingerprint datasets using likelihood ratio method. The results of multilayer perceptron for both databases are shown in Fig. 4 and Fig. 5. As you can see, using identification models - T-normalization and second best score model perform better than using the original scores in both likelihood ratio and multilayer perceptron in biometric systems.

Tables 1 and 2 show the mean and standard deviation of equal error rate(EER) for NIST BSSR1 and FVC2002 using likelihood ratio in the bootstrap steps and Tables 3 and 4 have the EER for both databases utilizing multilayer perceptron. Both T-normalization and second best score model have smaller equal error rate than the original system. And second best score model performs better than T-normalization in almost all cases.

6. Conclusions

Utilizing the dependencies between related matching scores can improve the performance of multi-sample biometric systems. In this paper, we have shown two identification models - T-normalization and second best score model utilizing identification trial score sets. Two trainable methods are used to combine multiple samples of the same biometric trait - likelihood ratio and multilayer perceptron. The experimental results using both identification methods in both trainable functions show better performance than utilizing the original scores in the multi-sample face and fingerprint biometric systems.

EER(%)	Original	T-norm	SBS
C	3.70 ± 0.05	3.41 ± 0.07	2.04 ± 0.19
G	4.19 ± 0.26	3.65 ± 0.28	3.01 ± 0.34

Table 1. Error equal rate(mean$\pm$ standard deviation) for NIST BSSR1 datasets using likelihood ratio method

EER(%)	Original	T-norm	SBS
db1	0.35 ± 0.06	0.25 ± 0.08	0.17 ± 0.01
db2	0.31 ± 0.07	0.15 ± 0.08	0.17 ± 0.04
db3	3.28 ± 0.38	2.05 ± 0.30	1.78 ± 0.26
db4	2.04 ± 0.32	1.21 ± 0.16	1.26 ± 0.18

Table 2. Error equal rate(mean$\pm$ standard deviation) for FVC2002 datasets using likelihood ratio method

EER(%)	Original	T-norm	SBS
C	3.41 ± 0.30	2.89 ± 0.31	1.98 ± 0.20
G	5.75 ± 0.42	5.38 ± 0.60	4.38 ± 0.43

Table 3. Error equal rate(mean$\pm$ standard deviation) for NIST BSSR1 datasets using multilayer perceptron

EER(%)	Original	T-norm	SBS
db1	0.25 ± 0.08	0.21 ± 0.05	0.05 ± 0.01
db2	0.23 ± 0.10	0.12 ± 0.04	0.08 ± 0.04
db3	3.03 ± 0.58	2.20 ± 0.39	1.38 ± 0.46
db4	1.99 ± 0.31	1.52 ± 0.27	0.93 ± 0.19

Table 4. Error equal rate(mean$\pm$ standard deviation) for FVC2002 datasets using multilayer perceptron

In the NIST BSSR1 'C' and 'G' face datasets, the faces are captured at sequential time so the experiment is absolutely suitable for this application. But we don't have specific information on how the FVC2002 fingerprint datasets are generated and whether the scans are obtained at the same time. The real application may different, for instance, two fingerprint scans are captured at one time and the enrolled scan is obtained at another time, so some adjustments might be needed.

It is obvious that the proposed algorithms can be extended to more than two input samples or enrolled multiple samples. We can also see that because only scores are used, the identification models and combination methods can be used for any biometric systems generating matching scores.

References

[1] G. Aggarwal, N. K. Ratha, and R. M. Bolle. Biometric verification: Looking beyond raw similarity scores. *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW '06. Conference on*, pages 31–31, 2006. 3

[2] R. M. Bolle, N. K. Ratha, and S. Pankanti. Error analysis of pattern recognition systems–the subsets bootstrap. *Computer Vision and Image Understanding*, 93(1):1–33, 2004. doi: DOI: 10.1016/j.cviu.2003.08.002. 4

[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. 2

[4] 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. 2

[5] T.-Y. Jea and V. Govindaraju. A minutia-based partial fingerprint recognition system. *Pattern Recognition*, 38(10):1672–1684, 2005. 4

[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. 2

[7] 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. 1

[8] 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. 2

[9] A. Kumar and D. Zhang. Improving biometric authentication performance from the user quality. *Instrumentation and Measurement, IEEE Transactions on*, 59(3):730–735, 2010. doi: 10.1109/TIM.2009.2028773. 1

[10] 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. 1

[11] J. Mariethoz and S. Bengio. A unified framework for score normalization techniques applied to text-independent speaker verification. *Signal Processing Letters, IEEE*, 12(7):532–535, 2005. 2

[12] S. Prabhakar and A. K. Jain. Decision-level fusion in fingerprint verification. *Pattern Recognition*, 35(4):861–874, 2002. doi: DOI: 10.1016/S0031-3203(01)00103-0. 3

[13] A. A. Ross, K. Nandakumar, and A. K. Jain. *Handbook of Multibiometrics*. International Series on Biometrics. Springer-Verlag New York, Inc., 2006. 1

[14] H. Tin Kam, J. J. Hull, and S. N. Srihari. Decision combination in multiple classifier systems. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 16(1):66–75, 1994. 2

[15] S. Tulyakov, Z. Zhi, and V. Govindaraju. Comparison of combination methods utilizing t-normalization and second best score model. *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW '08. IEEE Computer Society Conference on*, pages 1–5, 2008. 3

[16] 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. 2