Journal of Information & Communication Technology
Vol. 3, No. 1, (Spring 2009) 29-35

Effectiveness of Score Normalisation in Multimodal Biometric Fusion

F.Alsaade*
King Faisal University, Al-Ahssa, Saudi Arabia
Noor Zaman*
Hamdard University, Karachi Pakistan
Mansoor Z. Dawood*
Institute of Business & Technology, Biztek, Karachi, Pakistan
Sayed Hyder Abbas Musavi*
Hamdard University, Karachi Pakistan

ABSTRACT

An important problem which can adversely influence the overall effectiveness of biometric recognition is the undesired variations in the biometric data. This is because such variations are reflected in the corresponding biometric scores. This paper investigates the effectiveness of different score normalisation techniques for tackling this issue as well as enhancing the accuracy of multimodal biometrics. The fusion process is accomplished at the score level.

INSPEC Classification : B2360, B6430, C1250, C4240

Keywords : Multimodal biometrics, score-level fusion, biometric verification, score normalisation.

1. INTRODUCTION

Automatic personal recognition is becoming a basic need now a days. A number of different applications are in use for the same, for example physical access control, teleshopping , and telebanking are the few famous one out of those. Conventional recognition methods have few lacking , while biometric recognition is becoming more safe, secure and powerful. In recent years, an area of considerable interest in biometric recognition has been the use of multiple modalities, to make it more robust. Keeping in view the features, safety and importance of multimodal, this paper provides the opportunity for enhancing the recognition accuracy beyond that achievable with unimodal.

In general, multimodal biometrics is based on notion that the sets of data obtained from different modalities are complementary to each other [Jain, et al., 2004]. It's really more meaningful to use the different datasets in this fashion rather than use them individual. For this purpose, there are various data combination levels that can be considered, examples are the feature level, score level and decision level [Jain, et al., 2004]. It has, however, been reported that the most appropriate and effective approach to multimodal biometrics is through the fusion of data at the score level [Indovina, et al., 2003].

* The material presented by the authors does not necessarily portray the viewpoint of the editors and the management of the Institute of Business and Technology (BIZTEK) or King Faisal University, Al-Ahssa, Saudi Arabia, Hamdard University, Karachi Pakistan and Institute of Business & Technology Biztek.

*F.Alsaade :falsaade@kfu.edu.sa
*Noor Zaman :noorzaman650@hotmail.com
*Mansoor Z. Dawood :dr.mzdawood@biztek.edu.pk
*Sayed Hyder Abbas Musavi :sayed.abbas@hamdard.edu.pk

© JICT is published by the Institute of Business and Technology (BIZTEK).
Ibrahim Hydri Road, Korangi Creek, Karachi-75190, Pakistan.

A major issue with unimodal and multimodal biometric technique is the undesired variation of biometric data. Such variations are reflected in the corresponding biometrics scores, and that's why they have adversely influence the overall effectiveness of biometric recognition. The same undesired variation occurs due to different reasons ,for example, the effects of data capturing apparatus and various non-ideal data taking conditions such as light effects, background noise, etc.

In recent work, [Alsaade, et al., 2008] has presented an investigation into the effects, on the accuracy of multimodal biometrics, of introducing unconstrained cohort normalisation (UCN) into the score-level fusion process. The study has demonstrated that the capabilities provided by UCN can significantly improve the accuracy of fused biometrics. This paper, on the other hand, experimentally compares the effectiveness of two different score normalisation techniques with the UCN for enhancing the accuracy of multimodal biometrics. The focus of the study is on the score-level fusion of face and voice biometrics using SVM (support vector machine). The use of SVM in this work is based on earlier studies reporting it as one of the most effective methods for multimodal biometric fusion [Jain, et al., 2004; Gutschoven and Verlinde, 2000]. However, because of the generality of the approach proposed in this paper, the outcomes should be applicable to other fusion methods as well.

The rest of the paper is organised as follows. Section 2 introduces the proposed approaches and discusses the motivation behind their use. The experimental investigations and an analysis of the results are presented in Section 3, and the overall conclusions are given in Section 4.

2. SCORE NORMALISATION

The main purpose of using score normalisation in the field of multimodal biometrics is to facilitate the suppression of the individual biometric scores for impostors in relation to those for the clients. In practice, this is particularly important in order to minimise the effects of variations in the data from the individual modalities deployed. This would then lead to the maximisation of the recognition accuracy in the presence of variation (e.g. due to contamination) in some or all types of biometric data involved. There are two main categories of score normalisation (i.e. Bayesian and standardisation). Different methods under these two categories of score normalisation have already been subjected to thorough comparative evaluations in the context of speaker recognition [Ariyaeeinia and Sivakumaran, 1997; Ariyaeeinia, et al., 2006]. These are Cohort Normalisation (CN), Unconstrained Cohort Normalisation (UCN), Universal Background Model (UBM) Normalisation, T-norm and Z-norm. The results have clearly confirmed the importance of score normalisation in speaker verification. The aim of this paper is to explore the potential usefulness of score normalisation in enhancing accuracy in multimodal biometrics. The following describes three different methods in the two main categories of score normalisation mentioned earlier.

2.1. Bayesian solution

Under the Bayesian framework, the normalised matching score can be expressed as follows [Ariyaeeinia and Sivakumaran, 1997; Ariyaeeinia, et al., 2006]

$$p(\lambda * x) = \frac{p(x * \lambda) p(\lambda)}{p(x)}$$

where $p(.)$ is the probability function. In this equation, the speaker model probability, $p(\lambda)$, can be assumed equal for all speakers, and therefore ignored. $p(x)$, on the other hand, will need to be approximated. Two different approaches are presented in this paper for such approximation.

2.1.1. Cohort Normalisation (CN)

As described in [Ariyaeeinia and Sivakumaran, 1997; Ariyaeeinia, et al., 2006], given a test token of certain biometrics type, the normalised matching score provide through CN can be expressed as

$$S^i=\log\rho_T^i - \frac{1}{N}\sum_{n=1}^{N}\log\rho_n^i \qquad (2)$$

where i denotes the biometrics type, S^i is the normalised score of biometric *i*, ρ_T^i is the score for the target model, ρ_n^i are the scores obtained for a set of competing models, and N is the number of competing models considered. Here, the competitiveness of any two models is determined based on their closeness in biometric space. The entire cohort selection is carried out prior to the test phase. More information about CN can be found in [Ariyaeeinia and Sivakumaran, 1997; Ariyaeeinia, et al., 2006].

2.1.2. Unconstrained Cohort Normalisation (UCN)

In this technique, the normalised matching score provide through UCN can be calculated as in equation (2). However, the competing models, in this case, are selected dynamically from a group of background models, based on their closeness to the test token [Ariyaeeinia and Sivakumaran, 1997; Ariyaeeinia, et al., 2006].

2.2. Standardisation of score distributions

In speaker recognition, a method based on the standardisation of score distributions is a slightly different approach for score normalisation. Such approach aims to facilitate the use of a single threshold for all registered speakers [Ariyaeeinia, et al., 2006]. A major difficulty in setting a global threshold in speaker verification (SV) is that both impostor score distribution and true speaker score distribution have different characteristics for different registered speakers. Fixing the characteristics of one of the score distribution types for all registered speakers can tackle this issue. Usually, the common practice is to focus on standardising the impostor score distributions. The main reason for operating on the impostor score distributions, rather than on the true speaker score distributions, is the unavailability of sufficient data (in the existing databases) for a reliable estimation of the standardisation parameters in the latter approach. The following presents the descriptions of one approach in this category.

2.2.1. Test Normalisation (Tnorm)

This technique is based on using the mean and standard deviation of the impostor distribution. Such parameters are determined dynamically in the test phase using a set of example impostor models. The score normalisation is obtained based on the following equation [Ariyaeeinia, et al., 2006]:

$$S^i=(\log\rho_T^i - \mu_n^i)/\sigma_n^i \qquad (3)$$

where i denotes the biometrics type, S^i is the normalised score of biometric i, ρ_T^i is the score for the target model, and μ_n^i and σ_n^i are the mean and standard deviation for a set of competing models respectively.

3. EXPERIMENTAL INVESTIGATIONS AND RESULTS

The experimental studies are concerned with the score-level fusion of face and voice biometrics in the recognition mode of verification. The modeling and pattern matching approaches used with each modality is not discussed here, as these are outside the scope of this paper. The investigations involve three different data conditions. The first two are formed by using scores for clean face images together with scores for either clean or degraded utterances. The third one is based on the use of scores for degraded face images and degraded utterances.

In each experiment, the individual biometric score types involved are subjected to the range equalisation process using the Z-score normalisation [Indovina, et al., 2003]. In this work, the process of score-level fusion is based on the use of support vector machine [Burges, 1998]. The fusion scheme is applied to the biometric scores with and without subjecting them to the selected score normalisation schemes. The procedures for speech feature extraction and speaker classification are as detailed in [Ariyaeeinia, et al., 2006; Fortuna, et al., 2004]. The face recognition scores are based on the approach detailed in [Zafeiriou, et al., 2006; Bengio, et al., 2002].

3.1. Fusion under Clean Data Conditions

The aim of the experiments in this part of the study is to investigate the effectiveness of the selected score normalisation schemes in enhancing the reliability of multimodal fusion when the biometric datasets are free from degradation. The datasets considered for the face and voice modalities in this investigation are extracted from the XM2VTS and TIMIT databases respectively [Zafeiriou, et al., 2006; Alsaade, et al., 2005]. Using these biometric datasets, a total number of 140 client tests and 19460 (i.e. 140×[140-1]) non-client tests is used from the development data. While the total number of client and non-client tests used in investigating the performance for the proposed schemes is 140 and 19460 respectively. The verification results are presented as equal error rates (EERs) in Table 1.

Table 1
Effectiveness of score normalisation in Multimodal verification based on clean biometric data.

Modality	EER% (Without score normalisation)	EER% (With Tnorm)	EER% (With CN)	EER% (With UCN)
Voice (TIMIT)	2.30	0.90	. 0.00	. 0.00
Face (XM2VTS)	3.57	2.87	1.64	1.58
Fused: voice and face	0.12	0.17	. 0.00	. 0.00

Table 1 shows that the use of CN and UCN reduces the EER to zero for the individual modalities and for the fused biometrics with linear SVM. Tnorm, however, reduces the verification error rate, particularly, in the individual modalities. On the other hand, it is seen that the verification accuracy offered by fused biometrics is decreased slightly (by about 2 %) through the use of Tnorm prior to fusion. The effectiveness of CN, UCN and Tnorm under clean data condition is due to their ability to suppress the scores for impostors in relation to those for true users. It is noted that the usefulness of such score normalisation techniques in fused biometrics is mostly due to its performance with the voice modality. However, the corrective effect that CN, UCN and Tnorm have on the face modality is seen to be also considerable. This has in turn helped further enhance the accuracy of classification based on the fused data.

3.2. Fusion under Varied Data Quality Conditions

The purpose of the experiments presented in this section is to investigate the usefulness of the three score normalisation techniques in multimodal fusion when the qualities of the biometric data types are considerably different. The datasets considered for the face and voice modalities in this case are extracted from the XM2VTS (clean images) [Zafeiriou, et al., 2006] and from the 1-speaker detection task of the NIST Speaker Recognition Evaluation 2003 (degraded speech) databases respectively [Fortuna, et al., 2004]. Using these datasets, again a total number of 140 client tests and 19460 (i.e. 140×[140-1]) non-client tests is used from the development data. While the total number of client and non-client tests used in investigating the performance for the proposed schemes is 140 and 19460 respectively. The results of verification in this case are presented in Table 2.

Table 2
Performance of score normalisation in biometric verification based on mixed-quality data.

Modality	EER% (Without score normalisation)	EER% (With Tnorm)	EER% (With CN)	EER% (With UCN)
Voice (TIMIT)	30.71	10	25.71	11.37
Face (XM2VTS)	3.57	2.87	1.64	1.58
Fused: voice and face	2.93	1.95	1.95	0.71

It is noted (Table 2) that whilst the error rates for the face modality are exactly the same as those in the previous investigation, due to the use of a degraded speech database, the accuracy rates for the voice modality in this case are lower than the corresponding ones in Section 3.1. The results in Table 2 demonstrate the capability of the selected score normalisation schemes in reducing the verification error rate, particularly, in fused biometrics. This is achieved by a combination of enhancing the client scores when these are affected by data degradation, and suppressing the impostor scores in relation to the client ones. It is noted that without subjecting the scores to a normalisation process, the fusion process results in improving the EER associated with the better modality by about 18%. According to the results, this reduced EER (2.93%) is further decreased by about 51%, 33% and 76% through the use of Tnorm, CN and UCN respectively.

3.3. Fusion under Degraded Data Conditions

The experiments in this section investigate the effectiveness of CN, UCN and Tnorm in enhancing the reliability of multimodal fusion when the biometric datasets are contaminated. The datasets considered for the face and voice modalities in this investigation are extracted from the BANCA [Bengi, et al., 2002] and NIST Speaker Recognition Evaluation 2003 [Fortuna, et al., 2004] databases respectively. Using these biometric datasets, a total of 26 subjects have been used for the experiments. The face recognition scores are obtained based on images captured in four sessions, and affected by two different forms of distortion [Bengi, et al., 2002]. Based on these and the corresponding score data for NIST, a development score dataset is formed for the experiments. This consists of 104 client tests and 2600 (i.e. 4×{26×(26-1)}) non-client tests. While the total number of client and non-client tests used in investigating the performance for the proposed schemes is 104 and 2600 respectively. The results for the verification experiments in this part of the study are presented as equal error rates (EERs) in Table 3.

It is observed from the experimental results in Table 3 that the fusion process results in an EER which is slightly better than the EER offered by the best unimodal biometrics. It is worth noting that the accuracy of fused biometrics with CN (Table 3) is away below the accuracy obtained for the two single modalities involved as well as for the fused biometrics. The reason for such a phenomenon is that the two databases involved in this part of the study are both heavily degraded. Another reason is that the entire cohort selection in CN is carried out prior to the test phase. This would then lead to the minimisation of the recognition accuracy in the presence of variation in some or all types of biometric data involved. However, it should be emphasised that using either Tnorm or UCN together with the fusion process, successfully reduces the EER.

Table 3
Effectiveness of score normalisation in multimodal verification based on degraded data.

Modality	EER% (Without score normalisation)	EER% (With Tnorm)	EER% (With CN)	EER% (With UCN)
Voice (TIMIT)	35.69	16.35	29.80	15.38
Face (XM2VTS)	18.27	14.42	18.26	13.46
Fused: voice and face	16.55	8.65	51.21	6.35

Some important outcomes of the experimental investigations can be observed by considering the results in all the tables shown above. From these results, it is clearly seen that in all three data conditions, subjecting the scores to UCN prior to the fusion process consistently lead to the best performance. This is shown to be due to the twofold characteristic of this score normalisation method. Firstly it provides a means for enhancing the scores when the test data is degraded, and secondly, it aims to suppress the scores from impostors in relation to those for clients.

4. CONCLUSION

This paper has presented a comparison investigation into the use of score normalisation with score-level fusion for multimodal biometrics. Three different score normalisation methods have been used in this study. These are Cohort Normalisation (CN), Unconstrained Cohort Normalisation (UCN) and Test Normalisation (TN). Amongst the three score normalisation methods considered (CN, UCN and Tnorm), UCN has appeared to provide better performance in terms of reducing error rates in both degraded and clean data conditions. This is shown to be due to the twofold characteristic of this score normalisation method. Firstly it provides a means for enhancing the scores when the test data is degraded, and secondly, it aims to suppress the scores from impostors in relation to those for clients.

5. REFERENCES

Alsaade, F., Malegaonkar, A. and Ariyaeeinia, A., 2005. Fusion of Cross Stream Information in Speaker Verification. Proc. COST 275 Workshop on Biometrics on the Internet, 63-66.

Alsaade, F., Ariyaeeinia, A., Malegaonkar, A., Pawlewski, M and Pillay, S., 2008. Enhancement of Multimodal Biometric Segregation Using Unconstrained Cohort Normalisation. Pattern Recognition, special issue on multimodal biometrics, 41, 814-820.

Ariyaeeinia, A. and Sivakumaran, P., 1997. Analysis and comparison of score normalisation methods for text-dependent speaker verification. Proc. Eurospeech'97, 3, 1379-1382.

Ariyaeeinia, A., Fortuna, J., Sivakumaran, P. and Malegaonkar, A. 2006. Verification Effectiveness in Open-Set Speaker Identification. IEE Proc. Vision, Image and Signal Processing, 153, 618-624.

Bengio, S., Bimbot, F., Mariethoz, J., Popovici, V., Poree, F., Bailly-Bailliere, E., Mate, G. and Ruiz, B., 2002. Experimental protocol on the BANCA database. Tech. Report, IDIAP-RR 02-05.

Burges, C. J. C., 1998. tutorial on support vector machines for pattern recognition, Data Mining and Knowledge Discovery, Volume. 2, 955-974.

Fortuna, J., Sivakumaran, P., Ariyaeeinia, A. and Malegaonkar, A., 2004. Relative effectiveness of score normalization methods in open-set speaker identification. Proc. IEEE Speaker and language Recognition Workshop (Odyssey'04) 369-376.

Gutschoven, B. and Verlinde, P., 2000. Multi-modal identity verification using support vector machines (SVM). Proc. Int. Conf. Information Fusion, 3-8.

Indovina, M., Uludag, U., Snelick, R., Mink, A. and Jain, A.K., 2003. Multimodal Biometric

Authentication Methods: A COTS Approach. Proc. Multi-Modal User Authentication (MMUA), 99-106.
Jain, A.K., Ross, A. and Prabhakar, S., 2004. An Introduction to Biometric Recognition. IEEE Trans. Circuits and Systems for Video Technology, 14, 4-19.
Zafeiriou, S., Tefas, A., Buciu, L. and Pitas, I., 2006. Exploiting Discriminant Information in Nonnegative Matrix Factorization With Application to Frontal Face Verification. IEEE Trans. Neural Networks, 17, 683-695.