QUALITATIVE FUSION OF NORMALISED SCORES IN MULTIMODAL BIOMETRICS

Fawaz Alsaade, Aladdin Ariyaeeinia, Amit Malegaonkar, Surosh Pillay

University of Hertfordshire, Hatfield, UK

Abstract

A new approach to enhancing the accuracy of multimodal biometrics is investigated. The proposed approach, which involves combining score normalisation and qualitative-based fusion, is shown to considerably improve the accuracy of multimodal biometrics under different data conditions.

Keywords- Multimodal biometrics, score-level fusion, unconstrained cohort normalisation, qualitative-based fusion.

1. INTRODUCTION

The problem of automatic recognition of individuals using biometrics has received a great deal of interest in the last few decades [Jain, et al., 2004; Prabhakar, et al., 2003]. This is mainly due to the obvious advantages that biometrics has to offer over conventional means of identity verification. However, whilst considerable advances have been achieved in this area, the performance of unimodal biometric techniques is still being impeded by such issues as non-universality and impersonation. Recent studies have shown that a potentially viable way of addressing such problems is the use of multiple modalities in biometric recognition [Indovina, et al., 2003; Gutschoven and Verlinde, 2000]. Another major benefit offered by multimodal biometrics is the capability to outperform the recognition accuracy of unimodal systems. This paper is concerned with the latter characteristic of multimodal biometrics.

Multimodality is based on the concept that the information obtained from different modalities complement each other [Indovina, et al., 2003]. Consequently, an appropriate combination of such information can be more useful than using information from any of the single modalities involved alone. 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 shown that the score-level fusion is the most effective approach to multimodal biometrics [Indovina, et al., 2003]. This approach involves combining together the matching scores obtained through the individual modalities involved.

In general, one of the main problems associated with biometric systems, unimodal or multimodal, is the undesired variations in the biometric data. Such variations are reflected in the corresponding biometric scores, and thereby can adversely influence the overall effectiveness of biometric recognition. The said variations can arise due to a variety of factors such as the sensors used in capturing the biometric data and various non-ideal operating conditions such as background noise or poor lighting conditions.

In this paper, a combined approach is proposed to enhance the accuracy of multimodal biometrics. This is based on the normalisation of the matching scores obtained for face and voice biometrics. The quality of the normalised scores for each modality is then measured and using this prior knowledge, the score-level fusion is carried out 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 results can also be applied to other fusion methods.

The rest of the paper is structured as follows. Section 2 introduces the proposed approach for multimodal biometrics. Section 3 describes the experimental investigations, and the overall conclusions are presented in Section 4.

2. PROPOSED METHOD

As indicated earlier, data variations are considered as one of the main problems in multimodal fusion. Such variations are reflected in the corresponding biometric scores, and can thereby adversely influence the overall effectiveness of biometric recognition. As a result, there has recently been considerable research into ways of tackling the problem of data variations, through quality learning schemes [Poh and Bengio, 2005; Nandakumar, et al., 2006; Alsaade, et al., 2006] or score normalisation approaches [Alsaade, et al., 2008] in fusion-based biometrics. The quality learning schemes are, in general, concerned with adjusting the balance of weighting in fusion in favour of the modalities of better quality [Poh and Bengio, 2005; Nandakumar, et al., 2006; Alsaade, et al., 2006]. In other words, such techniques involve emphasising or deemphasising the scores for the individual biometric modalities during the fusion process, depending on an estimate of their relative degradation. On the other hand, it has been shown in [Alsaade, et al., 2008] that the use of Unconstrained Cohort Normalisation (UCN) helps improve the robustness of multimodal biometrics. This is because the approach provides a useful means for appropriately adjusting the individual biometric scores for a client, without any prior knowledge of the level of degradation of each biometric data type involved. Another motivation for using UCN in multimodal biometrics is that it facilitates the suppression of the individual biometric scores for impostors in relation to those for the clients. However, it is believed that the accuracy of multimodal biometrics can be further enhanced if the scores from the individual modalities involved are first subjected to UCN [Alsaade, et al., 2008] and then passed on to the relative quality learning mechanism [Alsaade, et al., 2006].This process is expected to enhance the overall accuracy of score-level fusion in multimodal biometrics due to the individual capabilities of each of the two techniques involved. The combined method should help enhance the multimodal biometrics reliability in clean, degraded and mixed-quality data conditions.

As described in [Alsaade, et al., 2008], given a test token of certain biometrics type, the normalised matching score provide through UCN can be expressed as

$$\Gamma_{(f/s)} = \log \rho_T^{(f/s)} - \frac{1}{N} \sum_{n=1}^{N} \log \rho_n^{(f/s)},$$ (1)

where f and s denote the biometrics type (face and speech), $\Gamma_{(f/s)}$ is the normalised score for face and speech, $\rho_T^{(f/s)}$ is the score for the target model, $\rho_n^{(f/s)}$ are the scores obtained for a set of competing models, and N is the number of competing models considered. These competing models are selected dynamically from a group of background models, based on their closeness to the test token.

The exploitation of information about the relative qualities of biometric data for the purpose of fusion is a multistage process [Alsaade, et al., 2006]. For each biometrics type, the relative quality is determined by the characteristics of the normalised scores obtained with the development and test samples of that modality. The quality of the face scores (Q_f) and speech scores (Q_s) are calculated as follows:

$$Q_{(f/s)} = Q(\Gamma_{d(f/s)}) \times Q(\Gamma_{ts(f/s)}),$$ (2)

where $Q_{(f/s)}$ is the quality for face and speech, $\Gamma_{d(f/s)}$ and $\Gamma_{ts(f/s)}$ are the normalised scores for the involved modality (face/speech) obtained using the development data and test data sample respectively.

Based on equation (2), the computation for the quality of samples is divided into two steps. These are described in the following sections.

2.1. Estimation of the quality aspects for the development data samples

$Q(\Gamma_{d(f/s)})$ in equation (2) denotes the quality of the normalised score obtained with face or speech development data. It is computed in the development phase as follows [Alsaade, et al., 2006].

$$Q(\Gamma_{d(f)}) = \frac{l_s}{l_s + l_f}, \tag{3}$$

$$Q(\Gamma_{d(s)}) = \frac{l_f}{l_s + l_f}, \tag{4}$$

where $l_{(f/s)}$ is obtained (in the development phase) as follows:

$$l_{(f/s)} = \sqrt{\frac{\left(\sigma_{(f/s)}^{C}\right)^2}{B_{(f/s)}^{C}} + \frac{\left(\sigma_{(f/s)}^{I}\right)^2}{B_{(f/s)}^{I}}}, \tag{5}$$

where $\sigma_{(f/s)}^{C}$ is the variance of the normalised client scores for the individual biometric data types (face/speech). Similarly, $\sigma_{(f/s)}^{I}$ is the variance of the normalised impostor scores for the individual biometric data types (face/speech). $B_{(f/s)}^{C}$ and $B_{(f/s)}^{I}$ are the total number of clients and impostors respectively for the individual face and speech biometric data in the development set.

2.2. Estimation of the quality aspects for the test data samples

$Q(\Gamma_{ts(f/s)})$ in equation (2) represents the relative qualities of the test data for face and speech. These quality aspects are calculated using a subset of the test data as follows

$$Q(\Gamma_{ts(f)}) = \frac{k_{ts(f)}}{k_{ts(s)} + k_{ts(f)}}, \tag{6}$$

$$Q(\Gamma_{ts(s)}) = \frac{k_{ts(s)}}{k_{ts(s)} + k_{ts(f)}}, \tag{7}$$

where $k_{ts(f/s)}$ is computed in the testing phase as follows:

$$k_{ts(f/s)} = \frac{\left| \dfrac{(\Gamma_{ts(f/s)} - \mu_{(f/s)}^C)^2}{\sigma_{(f/s)}^C} - \dfrac{(\Gamma_{ts(f/s)} - \mu_{(f/s)}^I)^2}{\sigma_{(f/s)}^I} \right|}{\mu_{(f/s)}^C},$$ (8)

where $\mu_{(f/s)}^C$ and $\sigma_{(f/s)}^C$ are the mean and variance of the normalised client scores for the individual types of biometrics (face/speech). Similarly, $\mu_{(f/s)}^I$ and $\sigma_{(f/s)}^I$ are the mean and variance of the normalised impostor scores for the individual types of biometrics. Such means and variances are calculated during the development phase.

2.3. Exploitation of the estimated data qualities

Once $Q_{(f/s)}$ for face and speech modalities are obtained, there are two approaches that can be used for passing this information to the classifier [Alsaade, et al., 2006]. The first approach is known as "relative quality aspect as an independent feature (RQ-IF)". In this approach, SVM is fed with four input vectors/ data values, two of these (vectors/ data values) present the actual individual biometric scores (face/ speech) based on the current stage (development/ test) whilst the other two present the relative quality of both the development and test data. During the development stage, the SVM uses these four input vectors (particularly the former two input vectors) to generate prior knowledge of the expected level of degradation of each biometric data type involved (in the test phase). This helps SVM to tune its parameters to fit the incoming test data [Alsaade, et al., 2006]. In the test stage, four input data values are passed on into the classifier (fusion stage), with two of them presenting the quality of the test data. As stated earlier, these are computed based on the parameters obtained from the development data. The other two data values present the test data itself (i.e. test scores). In the fusion stage, the four input data values are combined and then classified

based on the tuned SVM parameters obtained from the development stage [Alsaade, et al., 2006].

The second approach is known as "relative quality aspect as a modality specific weight (RQ-MSW)". In this approach, the quality of face and speech normalised scores is considered as weights [Alsaade, et al., 2006]. These weights for face or speech scores are then multiplied by their corresponding face or speech normalised scores, respectively. In the development phase, the results of these multiplications, two weighted input vectors, are used in order to optimise (tune) the parameters of SVM. This is because these parameters are believed to provide useful information about the relative degradation in the different types of biometric data. In the test phase, the weights for face or speech scores, which are computed based on their respective normalised sample test scores, are multiplied by their corresponding face or speech normalised scores. The resulted two weighted input data values are fused and then classified based on the tuned SVM parameters obtained in the previous phase.

Figures 1 and 2 illustrate the process involved in multimodal biometric fusion based on unconstrained score normalisation (UCN) and the use of the relative quality information (RQ-IF/MSW) respectively. Further details of the UCN approach and different quality learning mechanisms adopted in this study are presented in [Alsaade, et al., 2008] and [Alsaade, et al., 2006] respectively.

3. EXPERIMENTAL INVESTIGATIONS AND RESULTS

The experimental studies are concerned with the fusion of face and voice biometrics in the two recognition modes of verification and open-set identification. The modelling and pattern matching approaches used with each modality are not discussed here, as these are outside the scope of this study. The investigations in each mode involve three different data conditions. Two of them use scores for clean face images together with scores for either clean or degraded

utterances. The third uses scores for degraded face images together with scores for degraded utterances.

The individual biometric score types involved (in each experiment) are subjected to the range equalisation process using the ZS normalisation [Indovina, et al., 2003]. The fusion process is applied to the biometric scores with and without subjecting them to the UCN process. The fusion process, with UCN, is achieved via three different fusing configurations (Figure 2). In the first configuration, the normalised scores for face and voice are combined using linear SVM. This approach is termed Support Vector Machine with Unconstrained Cohort Normalisation (SVM-UCN). However, in the other two configurations, the normalised scores for face and voice are subjected to the relative quality analysis (discussed in the previous section) before being passed on to the SVM for fusion. As indicated earlier, the relative quality information can be used in two different ways in the fusion process. These are RQ-IF and RQ-MSW. Further descriptions of these can be found in [Alsaade, et al., 2006]. The fusion configurations involving the use of relative quality information together with UCN are referred to as RQ-IF-UCN and RQ-MSW-UCN in this paper.

The competing models required for UCN are selected from within the set of registered users during the test phase. The cohort size of the competing models is set to 1 and 3 for clean and degraded data respectively [Ariyaeeinia and Sivakumaran, 1997; Ariyaeeinia, et al. 2006]. The procedures for speech feature extraction and speaker classification are as detailed in [Alsaade, et al. 2005; Fortuna, et al. 2004]. The face recognition scores are based on the approaches detailed in [Zafeiriou, et al. 2006; Bengio, et al., 2002]. The experimental results (i.e. verification in both modes) are accompanied by a 95% confidence interval.

Test token , x_V

Target model – type V (λ_T)

λ_1 λ_2 λ_3 ------- λ_M

Background models (biometrics type V)

Scores for the best N models
$\{\lambda_n \ n = 1,2,..., N\}, N < M$

$$\Gamma\left(x_V\right) = \log \ \rho_T^V - \frac{1}{N}\sum_{n=1}^{N} \log \ \rho_n^V \qquad \text{Normalised Score (NS-V)} \longrightarrow$$

$$\Gamma\left(x_F\right) = \log \ \rho_T^F - \frac{1}{N}\sum_{n=1}^{N} \log \ \rho_n^F \qquad \text{Normalised Score (NS-F)} \longrightarrow$$

Scores for the best N models
$\{\psi_n \ n = 1,2,..., N\}, N < M$

Background models (biometrics type F)

Target model – type F (ψ_T)

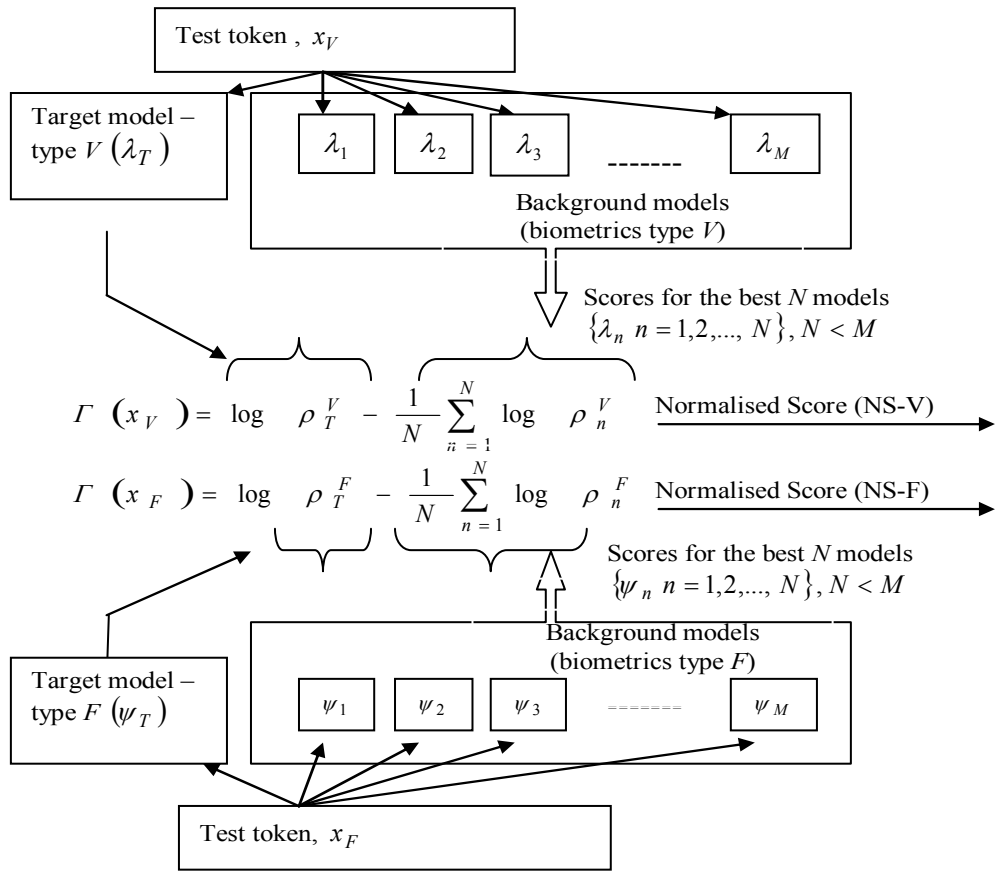ψ_1 ψ_2 ψ_3 ------- ψ_M

Test token, x_F

Fig. 1: Unconstrained cohort normalisation (UCN) of scores from the individual biometric modalities.
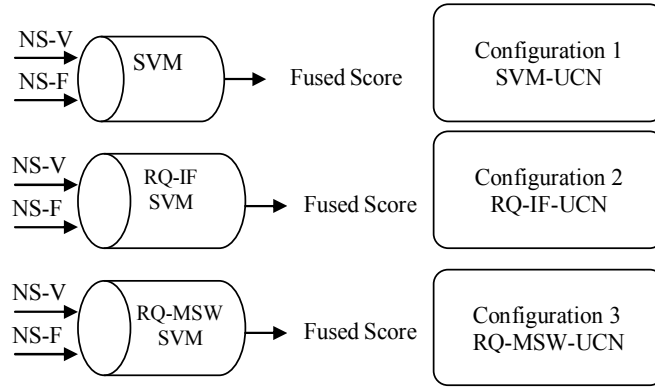
Fig. 2: Fusion of normalised scores of two different types of biometrics.
RQ-IF indicates the use of data relative qualities as independent features in the fusion process.
RQ-MSW indicates the use data relative qualities as modality specific weights in the fusion process.
SVM: support vector machine used for fusion.
UCN: unconstrained cohort normalisation.

3.1 Fusion under Clean Data Conditions

The purpose of the experiments in this part of the study is to investigate the effectiveness of the proposed method 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]. A total of 120 chimerical identities are formed for this purpose. These consist of 70 clients, 25 development impostors and 25 test impostors.

The development data comprises 70 and 6580 (i.e. $70 \times \{25+[70-1]\}$) tokens from the same-users and impostors (including cross-users) respectively whilst the total number of client and impostor tests used in estimating the quality of the test data is 70 and 6580 (i.e. $70 \times \{25+[70-1]\}$) respectively. In order to investigate the performance of the proposed scheme, 70 client tests and 6580 (i.e. $70 \times \{25+[70-1]\}$) impostor tests are used. The verification results for this

part of the study are presented as equal error rates (EERs) with a 95% confidence interval in Table 1.

Modality	EER ± CI 95(%) (Without UCN)	EER ± CI 95(%) (With UCN)
Voice (TIMIT)	2.86 ± 0.36	0.03 ± 0.04
Face (XM2VTS)	3.44 ± 0.39	1.56 ± 0.27
Fusing by SVM	0.16 ± 0.09	≈0.00 ± 0
RQ-MSW	0.08 ± 0.06	≈0.00 ± 0
RQ-IF	0.09 ± 0.07	≈0.00 ± 0

Table 1: Effectiveness of combining qualitative linear SVM with UCN based on clean biometric data.

The second column in Table 1 shows that the use of either of RQ-MSW and RQ-IF results in better performance than the individual modalities and the fused biometrics with linear SVM. Moreover, it is observed that the use of UCN results in a reduction of EERs for the individual modalities and for the fused biometrics. It is also noted that the use of UCN with fused biometrics reduces the EER to zero in all three configurations considered. The advantages of performing quality measurements on the normalised data prior to fusion are not clearly visible in this case because of the use of clean datasets for both modalities.

Table 2 presents the results of open-set identification (OSI) experiments with clean data. These are expressed in terms of IER (identification error rate) and OSI-EER that occur in the first and second stages of the process respectively.

Modality	Without UCN		With UCN	
	IER%	OSI-EER ± CI 95%	IER%	OSI-EER ± CI 95%
Voice (TIMIT)	≈0.00	18.57 ± 0.85	≈0.00	1.43 ± 0.26
Face (XM2VTS)	12.86	11.11 ± 0.69	12.86	3.57 ± 0.41
Fusion by SVM	≈0.00	4.28 ± 0.44	≈0.00	≈0.00 ± 0
RQ-MSW	≈0.00	2.86 ± 0.36	≈0.00	≈0.00 ± 0
RQ-IF	≈0.00	2.94 ± 0.37	≈0.00	≈0.00 ± 0

Table 2: Experimental results for open-set identification based on clean biometric data.

As before, the advantages of the proposed combined method are not clearly visible in the case of IER since the databases contain clean data. It is noted that, as in the verification scenario, the use of qualitative SVM results in better OSI-EER compared to standard linear SVM (or the individual modalities involved). It is also observed that subjecting the individual biometric scores to UCN prior to fusion in each of the three different configurations, reduces the error rates of the fused scores to zero.

3.2 Fusion under Varied Data Conditions

The purpose of the experiments presented in this section is to investigate the effectiveness of combining qualitative SVM with UCN when the biometric data types have different levels of quality. 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 [Fortuna, et al., 2004]. Using these datasets, again a total of 120 chimerical identities are formed which consist of the same number of clients, development impostors and test impostors as in the previous experiments (Section 3.1). The development and test datasets also consist of the same number of tokens from the same-users and impostors as those considered in the previous section. The results of verification and open-set identification for this part of the study are presented in tables 3 and 4 respectively.

Modality	EER $\pm$ CI 95(%) (Without UCN)	EER $\pm$ CI 95(%) (With UCN)
Voice (NIST)	30 ± 1.00	11.43 ± 0.70
Face (XM2VTS)	3.44 ± 0.39	1.56 ± 0.27
Fusion by SVM	3.69 ± 0.41	1.43 ± 0.26
RQ-MSW	3.32 ± 0.39	0.97 ± 0.21
RQ-IF	2.86 ± 0.36	0.86 ± 0.20

Table 3: Performance of UCN and quality learning in biometric verification based on mixed-quality data.

Modality	Without UCN		With UCN	
	IER%	OSI-EER $\pm$ CI 95%	IER%	OSI-EER $\pm$ CI 95%
Voice (NIST)	45.71	41.43 $\pm$ 1.08	45.71	15.56 $\pm$ 0.79
Face (XM2VTS)	12.86	11.11 $\pm$ 0.69	12.86	3.57 $\pm$ 0.41
Fusion by SVM	11.43	7.14 $\pm$ 0.56	5.71	2.66 $\pm$ 0.35
RQ-MSW	8.89	6.43 $\pm$ 0.54	4.33	2.11 $\pm$ 0.31
RQ-IF	8.57	6.67 $\pm$ 0.55	4.29	2.16 $\pm$ 0.32

Table 4: Experimental results for open-set identification based on mixed-quality data.

There are several observations to be made from these results. Firstly, it is noted that whilst the error rates for the face modality are exactly the same as those in the previous investigation, due to the use of degraded speech database, the accuracy rates for the voice modality in this case are lower than the corresponding ones in Section 3.1. It is observed from the results in Table 3, that the fusion process (SVM) on its own may not necessarily lead to the reduction of EER offered by the best single biometric modality involved. However, it is noted that the use of qualitative SVM (particularly SVM with RQ-IF) results in improvement of the EER associated with the better modality by about 17%. On the other hand, using linear SVM together with UCN reduces this EER by about 58%. It is interesting to note that the use of relative quality learning mechanisms (i.e. RQ-MSW and RQ-IF) together with UCN results in considerable improvement in the accuracy. In this case, it is seen that the use of RQ-IF with UCN reduces the EER for fused biometrics by about 76%.

It is observed from the results in Table 4 that the lowest IER and OSI-EER are obtained when the fusion is based on the use of UCN together with RQ-MSW/IF. It is noted that IER and OSI-EER obtained with SVM-based fusion are reduced by 62% and 70% respectively when UCN together with RQ-MSW are incorporated into the process.

3.3 Fusion under Degraded Data Conditions

The aim of the experiments in this part of the study is to investigate the effectiveness of combining qualitative SVM with UCN 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 [Bengio, et al., 2002] and NIST Speaker Recognition Evaluation 2003 [Fortuna, et al., 2004] databases respectively. Using these biometric datasets, a total of 52 chimerical identities consisting of 26 clients and 26 impostors are formed. The face recognition scores are obtained based on images captured in six sessions [Bengio, et al., 2002]. These sessions are separated as follows. Two sessions are used for the development data, while four sessions are used for the test data. Two of these four sessions are used to measure the relative quality of the test data whilst the other two are used to investigate the performance of the proposed scheme. Based on these and the corresponding score data for NIST, a development score dataset is formed for the experiments. This consists of 52 (i.e. 2×26) and 2652 (i.e. 2×{26×[26+(26-1)]}) score tokens from the same-users and impostors (including cross-users) respectively whilst the total number of score tokens from the same-users and impostors (including cross-users) used for finding the relative quality of the test data is 52 (i.e. 2×26) and 2652 (i.e. 2×{26×[26+(26-1)]}) respectively. In order to investigate the performance of the proposed scheme 52 (i.e. 2×26) client tests and 2652 (i.e. 2×{26×[26+(26-1)]}) impostor tests are used.

The experimental results for the verification and open-set identification scenarios are presented in tables 5 and 6 respectively. The performances of the two recognition modes are measured in terms of EERs for the former and IERs and OSI-EERs for the latter.

It is observed that the accuracy rates for the face modality are lower than the corresponding ones in the previous sections. This is due to the use of a degraded face database. On the other

hand, although the speech database is degraded as in the previous section, the accuracy rates for the speech modality in this section are observed to be lower. This is because of the fact that the subset of NIST used in this case has a different size from that used previously. Based on the experimental results, it is noted that the use of SVM on its own does not lead to better performance than that of the best individual modality involved. The results in Table 5 demonstrate the capability of reducing the verification error rates by combining UCN with the qualitative SVM. This is thought to result from the three-fold characteristics of this combination. The first is that UCN provides a means for enhancing the scores when the test data is degraded. Secondly, it aims to suppress the scores for impostors in relation to those for clients. Finally, the use of relative quality measurements further facilitates the reduction in error rates. This is achieved by either assigning higher weights (RQ-MSW) to the best biometric scores or by feeding the SVM with new features (RQ-IF).

Modality	EER $\pm$ CI 95(%) (Without UCN)	EER $\pm$ CI 95(%) (With UCN)
Voice (NIST)	40.09 $\pm$ 1.85	11.98 $\pm$ 1.22
Face (BANCA)	17.68 $\pm$ 1.44	13.46 $\pm$ 1.29
Fusing by SVM	20.93 $\pm$ 1.53	5.42 $\pm$ 0.85
RQ-MSW	12.65 $\pm$ 1.25	4.15 $\pm$ 0.75
RQ-IF	14.78 $\pm$ 1.34	4.94 $\pm$ 0.82

Table 5: Effectiveness of UCN and quality learning in verification based on degraded data.

Modality	Without UCN		With UCN	
	IER%	OSI-EER $\pm$ CI 95(%)	IER%	OSI-EER $\pm$ CI 95(%)
Voice (NIST)	28.85	59.62 $\pm$ 1.85	28.85	15.38 $\pm$ 1.36
Face (BANCA)	32.69	32.69 $\pm$ 1.77	32.69	25 $\pm$ 1.63
Fusing by SVM	46.15	34.62 $\pm$ 1.79	19.23	7.69 $\pm$ 1.00
RQ-MSW	21.15	32.69 $\pm$ 1.77	19.23	5.77 $\pm$ 0.88
RQ-IF	32.69	30.77 $\pm$ 1.74	23.08	3.85 $\pm$ 0.72

Table 6: Experimental results for open-set identification based on degraded data.

It can be seen from the results in Table 6 that the use of SVM alone results in an increase in both of the IER and OSI-EER obtained with the best single modality. The use of qualitative SVM (i.e. RQ-MSW and RQ-IF) on the other hand cannot reduce both types of error together. This is in conflict with the results in Sections 3.1 and 3.2. The reason for such a phenomenon is that the two databases involved in this part of the study are both heavily degraded. However, it should be emphasised that the combination of UCN with qualitative SVM, successfully reduces both the IER and OSI-EER.

A performance assessment of the results in terms of OSI-EERs is presented in Figure 3. The results for each of the two individual modalities used in this investigation are given as baselines. It can be observed from Figure 3 that combining UCN with the qualitative SVM appears to provide better performance in terms of reducing error rates (OSI-EERs).

These outcomes confirm the earlier suggestion (Section 2) that the reliability of multimodal biometrics can be further increased if the scores from the individual modalities involved are first subjected to UCN and then passed on to the relative quality learning mechanism.

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, combining UCN with relative quality learning mechanism consistently lead to the best performance whether it is in verification or open-set identification mode. It can also be seen that there is a difference in the effectiveness of the fusion processes using RQ-IF and RQ-MSW. This is thought to be due to the different processes involved in passing on the quality of the scores to the SVM. Another possible reason for such behaviour is the biometric data conditions involved.
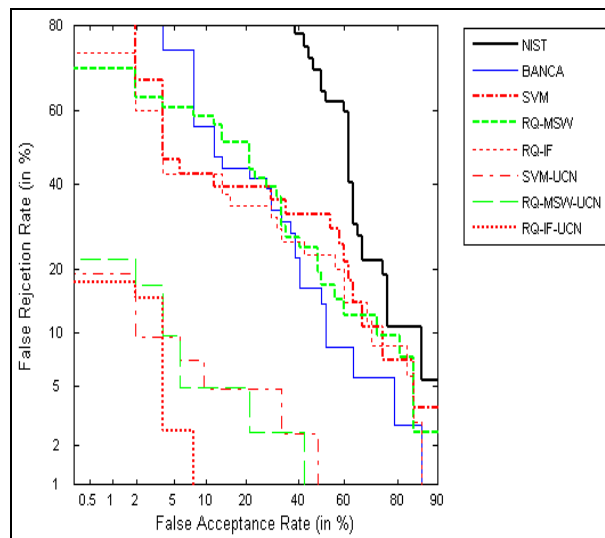

Fig. 3: DET plots showing the effects of qualitative SVM and UCN on the verification process in the second stage of open-set identification experiments with degraded data.

4. CONCLUSIONS

An investigation into the use of unconstrained cohort normalisation (UCN) combined with qualitative score-level fusion for multimodal biometrics has been presented. The experimental investigations have been carried out under three different data conditions. As expected in the case of clean data condition (for all modalities), the results have shown that the use of SVM with UCN leads to the highest accuracy. However, the effectiveness of combined UCN and data relative quality with SVM becomes evident when at least one of the biometric modalities is degraded. The experimental results have shown that, in such a

situation, the combination of UCN with relative quality learning measurements leads to a higher accuracy than either the best single modality performer or the use of only one of these techniques with SVM. This has been observed for both verification and open-set identification. The reason for this seems to relate to the individual characteristics of the two techniques: UCN aims to compensate for degraded scores and to suppress the impostor scores with respect to the client scores; whilst RQ-SVM makes use of the knowledge of the relative level of degradation of biometric data types involved.

REFERENCES

[1] 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.

[2] Alsaade, F., Ariyaeeinia, A., Meng, L. and A. Malegaonkar, 2006. Multimodal Authentication using Qualitative Support Vector Machines. INTERSPEECH'06, 2454-2457.

[3] 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.

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

[5] 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.

[6] 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.

[7] 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.

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

[9] 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.

[10] 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.

[11] Nandakumar, K., Chen, Y., Jain, A.K. and Dass, S.C., 2006. Quality-based Score Level Fusion in Multibiometric Systems. Proc. 18th Int. Conf. Pattern Recognition, 473-476.

[12] Poh, N. and Bengio, S., 2005. Improving Fusion with Margin-Derived Confidence in Biometric Authentication Tasks. Proc. fifth Int. Conf. Audio and Video-Based Biometric Person Authentication (AVBPA'05), 474–483.

[13] Prabhakar, S., Pankanti, S. and Jain, A.K., 2003. Biometric recognition: security and privacy concerns. IEEE Security & Privacy, 33-42.

[14] 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.