Target Dependent Score Normalization Techniques and Their Application to Signature Verification

J. Fierrez-Aguilar, J. Ortega-Garcia, and J. Gonzalez-Rodriguez

Biometrics Research Lab., ATVS
Universidad Politecnica de Madrid, Spain
{jfierrez,jortega,jgonzalez}@diac.upm.es

Abstract. Score normalization methods in biometric verification, which encompass the more traditional user-dependent decision thresholding techniques, are reviewed from a test hypotheses point of view. These are classified into test dependent and target dependent methods. The focus of the paper is on target dependent methods, which are further classified into impostor-centric, target-centric and target-impostor. These are applied to an on-line signature verification system on signature data from SVC 2004. In particular, a target-centric technique based on a variant of the cross-validation procedure provides the best relative performance improvement both for skilled (19%) and random forgeries (53%) as compared to the raw verification performance without score normalization (7.14% EER and 1.06% EER for skilled and random forgeries respectively).

1 Introduction

Previous studies have shown that the performance of a number of biometric verification systems, specially those based on behavioral traits such as written signature [1] or voice [2], can be improved by using user-dependent decision thresholds. Even greater verification performance improvement can be expected through the use of score normalization techniques [3]. These methods (which include the user-dependent thresholding as a particular case) account not only for user specificities but also for intersession and environment changes [4]. The system model of biometric authentication with score normalization is depicted in Fig. 1.

The objectives of this work are: *i*) to provide a framework for score normalization collecting previous work in related areas, *ii*) to provide some guidelines for the application of these techniques in real world scenarios, and *iii*) to provide an example of a successful application of the proposed normalization methods regarding the first international Signature Verification Competition (SVC 2004).

D. Zhang and A.K. Jain (Eds.): ICBA 2004, LNCS 3072, pp. 498–504, 2004.

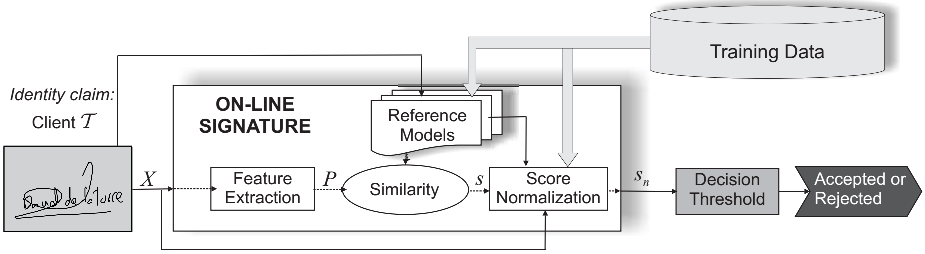


Fig. 1. System model of biometric authentication with score normalization

2 Theoretical Background

2.1 Score Normalization

Given a test sample X the problem of biometric authentication can be stated as a basic hypotheses test between two hypotheses:

$H0$**:** X is from hypothesized client $\mathcal{T}$.
$H1$**:** X is *not* from hypothesized client $\mathcal{T}$.

The optimum test to decide between these two hypotheses is a likelihood ratio test given by

$$\frac{p(X|H0)}{p(X|H1)} \begin{cases} > \theta \text{ Accept } H0 \\ < \theta \text{ Accept } H1 \end{cases} \tag{1}$$

where $p(X|H0)$ and $p(X|H1)$ are respectively the probability density functions for the hypotheses $H0$ and $H1$ evaluated for the observed biometric sample X. The decision threshold for accepting or rejecting $H0$ is θ. An equivalent log-likelihood ratio test is obtained transforming (1) into the log domain

$$\log p(X|H0) - \log p(X|H1) \begin{cases} > \log\theta \text{ Accept } H0 \\ < \log\theta \text{ Accept } H1 \end{cases} \tag{2}$$

A common practice in biometric verification (e.g., GMM in case of speaker recognition [5], HMM in case of signature recognition [6], etc.) consists in characterizing each client $\mathcal{T}$ by a statistical model $\lambda^{\mathcal{T}}$ (i.e., the reference model in Fig. 1). In this case, the similarity s is computed as

$$s = \log p(X|\lambda^{\mathcal{T}}) \tag{3}$$

which is an estimation of $\log p(X|H0)$. As a result, the optimal score normalization method for an authentication system based on statistical modeling is given by

$$s_n = s - \log p(X|H1) \tag{4}$$

Worth noting, the normalizing term is affected, in general, by:

Input Information: the input biometric sample X.

Information from Clients: example scores $s_1^{\mathcal{T}}, \ldots, s_{N_{\mathcal{T}}}^{\mathcal{T}}$ from the hypothesized target $\mathcal{T}$.

Information from Impostors: both the models $\left\{\lambda_1^{\overline{\mathcal{T}}}, \ldots, \lambda_{N_I}^{\overline{\mathcal{T}}}\right\}$ and example scores $\left\{s_1^{\overline{\mathcal{T}}}, \ldots, s_{N_{\overline{\mathcal{T}}}}^{\overline{\mathcal{T}}}\right\}$ from N_I possible impostors pretending the hypothesized client $\mathcal{T}$.

Estimation of $\log p(X|H1)$ based on the different information involved is nevertheless not a straightforward task. Thus, operational procedures are usually employed. Much effort has been done in order to derive such operational procedures for *score normalization* (also *likelihood normalization*) based on the statistical formalism described above, mainly in the speaker recognition community [3]. These operational procedures aim at designing a function

$$s_n = f(s, X, \left\{s_1^{\mathcal{T}}, \ldots, s_{N_{\mathcal{T}}}^{\mathcal{T}}\right\}, \left\{\lambda_1^{\overline{\mathcal{T}}}, \ldots, \lambda_{N_I}^{\overline{\mathcal{T}}}\right\}, \left\{s_1^{\overline{\mathcal{T}}}, \ldots, s_{N_{\overline{\mathcal{T}}}}^{\overline{\mathcal{T}}}\right\}) \tag{5}$$

so as to minimize the error rate of the verification task. Linear functions of various statistics of the information involved in Eq. (5) is the prevailing strategy. This is the case of the celebrated: *i*) z-norm, which considers only scores samples from impostors, *ii*) t-norm, based on the test sample and models from impostors, and *iii*) UBM-norm, which considers the test sample and a universal background model characterizing the average target. Other examples can also be found regarding face [7] or signature recognition [8].

In order to simplify the discussion yet providing a powerful framework for score alignment, the main focus of this paper is on considering neither input test information nor models from impostors, i.e.

$$s_n = f(s, \left\{s_1^{\mathcal{T}}, \ldots, s_{N_{\mathcal{T}}}^{\mathcal{T}}\right\}, \left\{s_1^{\overline{\mathcal{T}}}, \ldots, s_{N_{\overline{\mathcal{T}}}}^{\overline{\mathcal{T}}}\right\}) \tag{6}$$

This family of score normalization methods will be referred to as *target dependent score normalization techniques*. Other normalization methods using the test sample and models from impostors will be referred to as *test dependent normalization techniques*.

3 Target Dependent Score Normalization Techniques

3.1 Impostor-Centric Methods

In impostor-centric methods (*IC* for short) no information about client score intra-variability is used. Therefore

$$s_{IC} = f(s, \mathcal{I} = \left\{s_1^{\overline{\mathcal{T}}}, \ldots, s_{N_{\overline{\mathcal{T}}}}^{\overline{\mathcal{T}}}\right\}) \tag{7}$$

The following *IC* methods are considered in this work:

IC**-1:** $s_{IC\text{-}1} = s - \mu_{\mathcal{I}}$
IC**-2:** $s_{IC\text{-}2} = s - (\mu_{\mathcal{I}} + \sigma_{\mathcal{I}})$
IC**-3:** $s_{IC\text{-}3} = (s - \mu_{\mathcal{I}})/\sigma_{\mathcal{I}}$

where $\mu_{\mathcal{I}}$ and $\sigma_{\mathcal{I}}$ are respectively the mean and standard deviation of the impostor scores $\mathcal{I}$. Note that the impostor samples scores $\mathcal{I}$ can be, in general, either from casual impostors (cIC) or from real impostors (rIC).

3.2 Target-Centric Methods

In target-centric methods (TC for short) no information about impostor score variability is used. Therefore

$$s_{TC} = f(s, \mathcal{C} = \left\{s_1^{\mathcal{T}}, \ldots, s_{N_{\mathcal{T}}}^{\mathcal{T}}\right\}) \tag{8}$$

Similarly to the impostor-centric case, the following methods are obtained

TC**-1:** $s_{TC\text{-}1} = s - \mu_{\mathcal{C}}$
TC**-2:** $s_{TC\text{-}2} = s - (\mu_{\mathcal{C}} - \sigma_{\mathcal{C}})$
TC**-3:** $s_{TC\text{-}3} = (s - \mu_{\mathcal{C}})/\sigma_{\mathcal{C}}$

Client scores $\mathcal{C}$ should be obtained from the available training set. In this work, we propose to generate $\mathcal{C}$ by using either the resubstitution or the rotation sampling methods of error estimation [9].

3.3 Target-Impostor Methods

In target-impostor methods (TI for short) information from both client score intra-variability and impostor score variability is used. Therefore

$$s_{TI} = f(s, \mathcal{C} = \left\{s_1^{\mathcal{T}}, \ldots, s_{N_{\mathcal{T}}}^{\mathcal{T}}\right\}, \mathcal{I} = \left\{s_1^{\overline{\mathcal{T}}}, \ldots, s_{N_{\overline{\mathcal{T}}}}^{\overline{\mathcal{T}}}\right\}) \tag{9}$$

The two following methods are considered

TI**-1:** $s_{TI\text{-}1} = s - s_{\mathrm{EER}}(\mathcal{I}, \mathcal{C})$
TI**-2:** $s_{TI\text{-}2} = s - (\mu_{\mathcal{I}}\sigma_{\mathcal{C}} + \mu_{\mathcal{C}}\sigma_{\mathcal{I}})/(\sigma_{\mathcal{I}} + \sigma_{\mathcal{C}})$

where $s_{\mathrm{EER}}(\mathcal{I}, \mathcal{C})$ is the decision threshold at the empirical Equal Error Rate obtained from $\mathcal{I}$ and $\mathcal{C}$.

4 Experiments

For the experiments reported in this paper, the HMM-based on-line signature verification system from Universidad Politecnica de Madrid [6, 8] competing in the First Intl. Signature Verification Competition (SVC 2004)[1] has been used.

[1] http://www.cs.ust.hk/svc2004/

Development corpus of the extended task (including coordinate and timing information, pen orientation and pressure) from SVC 2004 has been used. It consists of 40 sets of signatures. Each set contains 20 genuine signatures from one contributor (acquired in two separate sessions) and 20 skilled forgeries from five other contributors.

Signature data from both sessions is used both for training and testing. Training data consists of 5 genuine signatures for each target. For a specific target user, casual impostor information is extracted from all the remaining targets. Results in which real impostor information is used for computing the normalization functions are also provided. Impostor data for the estimation of the normalization parameters, either casual or real, is used in a leave-one-out fashion, i.e., testing one impostor with a normalization scheme estimated with information from the remaining impostors and rotating the scheme.

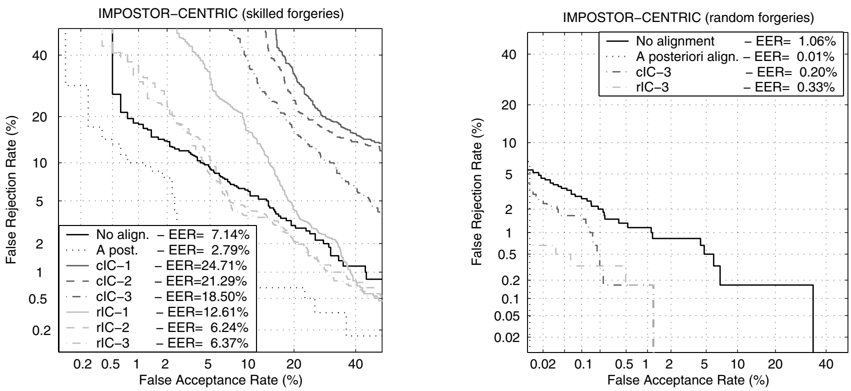
A priori score normalization methods are compared in the following. This means that only the information from the training set is used both for the enrollment of the targets and for the estimation of the parameters of the normalization scheme. In order to have an indication of the level of performance with an ideal score alignment between targets, the *a posteriori* target dependent score normalization TI-1 is also given. Only in this case test information is also used for the computation of the normalization functions.

4.1 Results

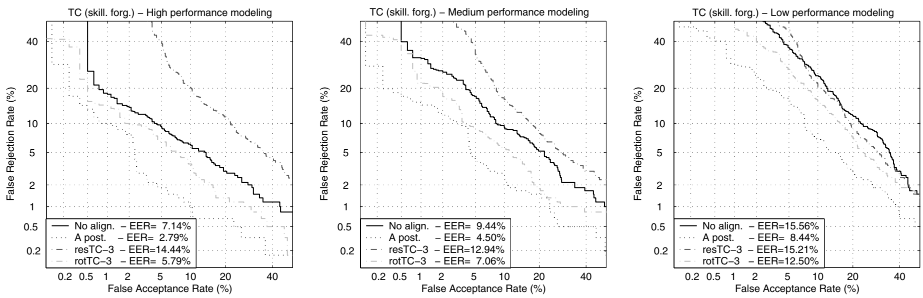
In Fig. 2 (a) the different impostor-centric methods described in Sect. 3.1 are compared either for skilled (left) or random forgeries (right). Raw verification performance with no normalization (7.14% and 1.06 EER for skilled and random forgeries respectively) is significantly improved by the *a posteriori* normalization scheme (2.79% and 0.01% respectively). Regarding the skilled forgeries test, *a priori* method IC-3 outperforms IC-1 and IC-2. Raw performance is only improved in this case by considering statistics from real impostors. Regarding the random forgeries test, significant improvements are obtained considering statistics either from casual or from real impostors.

Results of different resampling techniques for the estimation of target variability are summarized in Fig. 2 (b) for three different verification systems of decreasing verification performance (from left to right). As it can be observed, the rotation scheme always leads to verification improvements whereas the resubstitution strategy only leads to improvements in the low performance system. This result penalizes the biased estimation provided by the resubstitution scheme in favor of the unbiased rotation procedure.

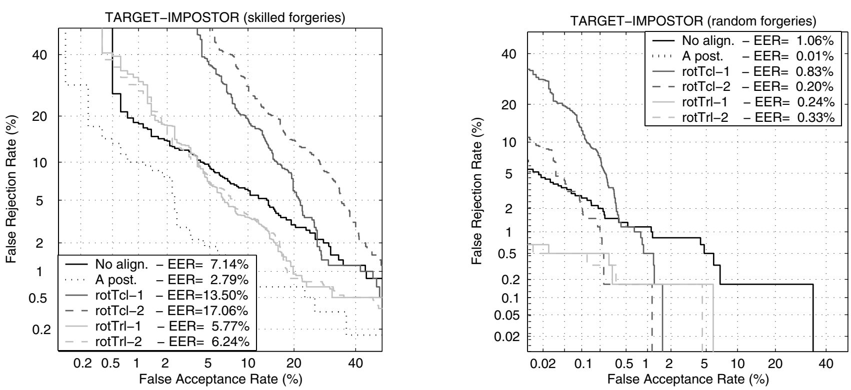
Verification performance for the target-impostor methods is shown in Fig. 2 (c). As in the impostor-centric experiment, only target-impostor normalization schemes based on real impostor statistics improve verification performance with respect to no score normalization in case of tests with skilled forgeries. With regard to the test considering random forgeries, verification performance improvements are obtained considering either casual impostor or real impostor statistics.



(a) Impostor-Centric: Different impostor-variability estimation methods.



(b) Target-Centric: Different client-variability estimation methods.



(c) Target-Impostor: casual/real information for impostor-variability estimation.

Fig. 2. Comparison of target dependent score normalization techniques

5 Conclusions

Target dependent score normalization techniques have been reviewed and applied to a HMM on-line signature verification system (7.14% EER and 1.06% EER for skilled and random forgeries respectively) on SVC 2004 signature data and various experimental finding have been obtained. Most remarkably, target-centric techniques based on a variation of the cross-validation procedure provided the best performance improvement both for skilled (5.79% EER) and random forgeries (0.50% EER). Other worth noting experimental findings are: *i*) the use of casual impostor statistics in either impostor-centric or target-impostor methods leads to the highest performance improvement when testing with random forgeries but lowers verification performance in case of testing against skilled forgeries, *ii*) the use of real impostor statistics in either impostor-centric or target-impostor methods leads to verification performance improvements when testing either with random or skilled forgeries, and *iii*) statistics for the estimation of target score intra-variability should be unbiased.

Acknowledgements

This work has been supported by the Spanish Ministry for Science and Technology under projects TIC2003-09068-C02-01 and TIC2003-08382-C05-01. J. F.-A. also thanks Consejeria de Educacion de la Comunidad de Madrid and Fondo Social Europeo for supporting his doctoral research.

References

[1] Jain, A., Griess, F., Connell, S.: On-line signature verification. Pattern Recognition 35 (2002) 2963–2972 498

[2] Matsui, T., Nishitani, T., Furui, S.: Robust methods of updating model and a priori threshold in speaker verification. In: Proc. of ICASSP. (1996) 97–100 498

[3] Auckenthaler, R., Carey, M., Lloyd-Tomas, H.: Score normalization for text-independent speaker verification systems. Digital Signal Processing **10** (2000) 42–54 498, 500

[4] Navratil, J., Ramaswamy, G.: The awe and mystery of t-norm. In: Proc. of ESCA Eur. Conf. on Speech Comm. and Tech., EuroSpeech. (2003) 2009–2012 498

[5] Reynolds, D.A., Quatieri, T.F., Dunn, R.B.: Speaker verification using adapted gaussian mixture models. Digital Signal Processing **10** (2000) 19–41 499

[6] Fierrez-Aguilar, J., Ortega-Garcia, J., Gonzalez-Rodriguez, J.: A function-based on-line signature verification system exploiting statistical signal modeling. Intl. Journal on Image and Graphics (2004) (accepted) 499, 501

[7] Sanderson, C., Paliwal, K.K.: Likelihood normalization for face authentication in variable recording conditions. In: Proc. of ICIP. Volume 1. (2002) 301–304 500

[8] Ortega-Garcia, J., Fierrez-Aguilar, J., Martin-Rello, J., Gonzalez-Rodriguez, J.: Complete signal modeling and score normalization for function-based dynamic signature verification. In: Proc. of IAPR Intl. Conf. on Audio- and Video-based Person Authentication, AVBPA, Springer (2003) 658–667 500, 501

[9] Jain, A., Duin, R., Mao, J.: Statistical pattern recognition: A review. IEEE Trans. on Pattern Anal. and Machine Intell. 22 (2000) 4–37 501