Non-parametric score normalization for biometric verification systems

Vitomir Štruc[1], Jerneja Žganec Gros[2], and Nikola Pavešić[1]
[1]*Faculty of Electrical Engineering, University of Ljubljana, Tržaška 25, 1000 Ljubljana, Slovenia*
[2]*Alpineon Ltd., Ulica Iga Grudna 15, SI-1000 Ljubljana, Slovenia*
vitomir.struc, nikola.pavesic@fe.uni-lj.si, jerneja.gros@alpineon.si

Abstract

In this paper we study the problem of score normalization in biometric verification systems. Specifically, we introduce a new class of normalization techniques, which unlike the commonly used parametric score normalization techniques, such as z- or t-norm, make no assumptions regarding the shape of the underlying score distribution. The proposed class of normalization techniques first estimates the relevant score distribution in an impostor-centric manner using kernel density estimation and then maps the estimated distribution to a common one. Our experimental results obtained on the FRGCv2 face database show that the proposed non-parametric score normalization techniques consistently outperform their parametric counterparts when the target distribution takes a log-normal form and that all assessed techniques, i.e., z-, t-, zt- and tz-norms, improve upon the setting where no score normalization is used.

1. Introduction

Score normalization techniques have received quite some attention in the fields of speaker or signature authentication [2, 3, 10, 6] and have found their way to other biometric modalities used in biometric verification systems as well [9, 8, 13, 14]. As stated in [13], the goal of these techniques is to counteract statistical variations in matching scores due to changes in conditions across different enrollment (target, gallery) and probe (test, query) samples. This is commonly achieved by assuming that the relevant score distribution takes a Gaussian form and then adjusting the first and second statistical moments of the presumed distribution to a common value. This process scales the score distribution and makes it possible to use a single global threshold for the task of identity verification.

In this paper we propose a different approach to score normalization, which makes no assumptions regarding the shape of the score distribution. Instead, it estimates the relevant distribution using kernel density estimation (KDE) and uses the obtained *non-parametric* estimate to remap the entire distribution to a common one. As no assumption is made regarding the type of the underlying score distribution and the entire distribution is adjusted instead of only the first two statistical moments, we naturally expect that the proposed class of non-parametric score normalization techniques will offer superior recognition performance when compared to their parametric counterparts.

In the remainder of the paper, we first briefly review the theory behind score normalization. Next, we introduce the novel class of non-parametric score normalization techniques and, finally, present an experimental assessment of the proposed techniques.

2. Theoretical background

Score normalization techniques can be categorized based on different criteria. One such criterion, which is also highly relevant to this paper, is the nature of the score distribution to be normalized. If the sample score population is generated based on client verification attempts, the score normalization is said to be *client-centric*, and similarly, if the sample score population is generated based on impostor verification attempts, the score normalization is said to be *impostor-centric* [2, 3, 10]. Since client scores are extremely scarce in practice, most normalization techniques fall into the latter category. This class of techniques also represents the main focus of this paper and among others comprises the popular *zero-normalization* or *z*-norm as well as the *test-normalization* or *t*-norm [2].

Consider a set of N users that are enrolled in a biometric verification system with corresponding class labels $\omega_1, \omega_2, ..., \omega_N$. Score normalization techniques try to define the mapping ψ of the following form[10]:

$$\psi : s \rightarrow s', \text{for } i \in \{1, 2, ..., N\}, \tag{1}$$

where s denotes a raw score representing the output of the matching module of a biometric verification system and s' stands for the normalized version of the score. Impostor-centric score normalization techniques typically define the mapping ψ based on the class-conditional impostor distribution $p(s|\overline{\omega_i})$ (where $^-$ implies that we are dealing with impostor scores s). The most popular impostor-centric score normalization techniques *z*- and *t*-norm, for example, both assume that the impostor distribution takes a Gaussian form, i.e., $p(s|\overline{\omega_i}) = \mathcal{N}(s; \mu, \sigma)$. Hence, they define the mapping ψ as follows:

$$\psi(s) = s' = \frac{s - \mu}{\sigma}. \tag{2}$$

Both the z- as well as the *t*-norm try to normalize the score distribution to $\mathcal{N}(s'; 0, 1)$, with the difference that the $z-$norm generates the required sample score population by comparing the enrolled template corresponding to the claimed identity ω_i to a number of impostor probe samples (or *z*-impostors), while the $t-$norm generates a similar sample score population by comparing the probe sample that was subjected to the verification procedure to a number of impostor templates (or *t*-models). Based on these sample score populations each type of score normalization computes its corresponding first and second statistical moment, μ and σ, respectively.

3. Non-parametric score normalization

Instead of presuming a certain shape for the class-conditional impostor distribution $p(s|\overline{\omega_i})$ and performing score normalization through parameter adjustment, we introduce in this section a class of non-parametric score normalization techniques which make no assumptions regarding the shape of the impostor distribution.

To be able to lift the gaussian assumption pertaining to the impostor score distribution we require an estimate of the probability density function (pdf) of the impostor distribution $p(s|\overline{\omega_i})$, which can be conveniently calculated using kernel density estimation (KDE) [1].Once the pdf is estimated, we can normalize the impostor score distribution by mapping it to a common predefined shape. This procedure can be formalized as follows: let ρ be a random variable with the property [4]

$$\rho = F(s) = \int_{q=-\infty}^{s} p_s(q)dq, \tag{3}$$

where q is a dummy variable for integration. Furthermore, let s' be another random variable with the property:

$$\rho = G(s') = \int_{x=-\infty}^{s'} p_{s'}(x)dx, \tag{4}$$

where x again denotes a dummy variable for integration. If we assume that the pdf $p_s(q)$ represents our impostor score pdf $p(s|\overline{\omega_i})$ and that $p_{s'}$ represents the pdf of a predefined target distribution, then we can describe the new class of non-parametric score normalization techniques with the following mapping:

$$\psi(s) = G^{-1}(\rho) = G^{-1}(F(s)) = s', \tag{5}$$

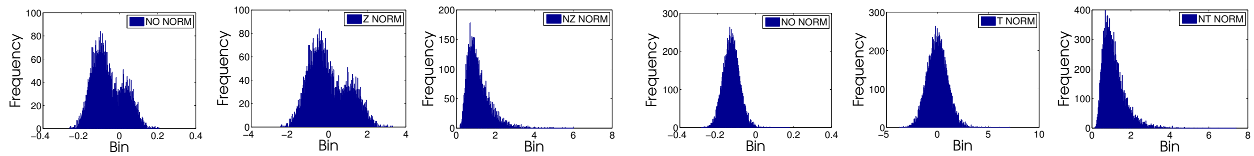
where $G(.)$ and $F(.)$ denote cumulative density functions (CDF) of their corresponding arguments and $G^{-1}(.)$ represents the inverse of the CDF.

Based on the above expressions, it is possible to implement most parametric normalization techniques in a non-parametric manner. Examples of possible normalization techniques that can be implemented using the presented formalism include among others non-parametric versions of the *z*-norm, *t*-norm, *zt*-norm (*t*-norm followed by *z*-norm) and the *tz*-norm (*z*-norm followed by *t*-norm).

4. Experiments

4.1. Database and experimental setup

For our experiments we use the FRGC database (version 2), which is a large scale database featuring more than 40000 facial images [7]. To assess the performance of our normalization techniques we select the hardest and most challenging experimental configuration defined for the FRGC, namely, FRGC experiment 4. In this configuration three data sets are available for experimentation: *a) the training set*, containing 12776 (controlled and uncontrolled) images belonging to 222 subjects, *b) the target set*, containing 16028 (controlled) images belonging to 466 subjects, and *c) the query set*, containing 8014 (uncontrolled) images belonging to 466 subjects. The training set is used to train potential background models (e.g., PCA or LDA transformation matrices, universal background models - UBMs, etc.) needed by the recognition system, while the target and query sets serve as the basis for matching score calculation. All images from the FRGC database were subjected to a preprocessing procedure that geometrically



(a) Raw z-scores (b) Z-norm (c) NZ-norm (d) Raw t-scores (e) T-norm (f) NT-norm

Figure 1. Comparison of different score normalization techniques (parametric and non-parametric) on the estimated histogram

aligns facial images using manually annotated eye positions, crops the facial region and scales the cropped region to a fixed size of 128×128 pixels.

For the experiments we implement a basic feature extraction technique and use it in conjunction with a simple similarity-based scoring procedure to generate the 8014×16028 similarity (or matching score) matrix that forms the foundation for assessing our score normalization techniques. Specifically, we implement a variant of Principal Component Analysis (PCA) [11] and use it in conjunction with the cosine Mahalanobis similarity measure and the nearest neighbor classifier [5]. We set the PCA feature vector length to 500. Note here that the feature extraction technique, the similarity measure used for the experiments as well as the achieved (baseline) recognition performance are only of minor importance. We are instead interested in the relative improvements gained by applying score normalization techniques to the computed similarity scores.

4.2. Experiments and results

To obtain an estimate of the performance gain achievable with parametric score normalization techniques, we first implement four popular normalization schemes, namely, the *t*-norm, the *z*-norm, the *tz*-norm and the less common *zt*-norm, and assess their performance on the FRGCv2 database. Next, we implement non-parametric versions of all of the above mentioned normalization techniques and compare their performance to their parametric counterparts. These non-parametric versions of the normalization techniques are referred to as *nt*-, *nz*- *ntz*-, and *nzt*-norm in the remainder of this section.

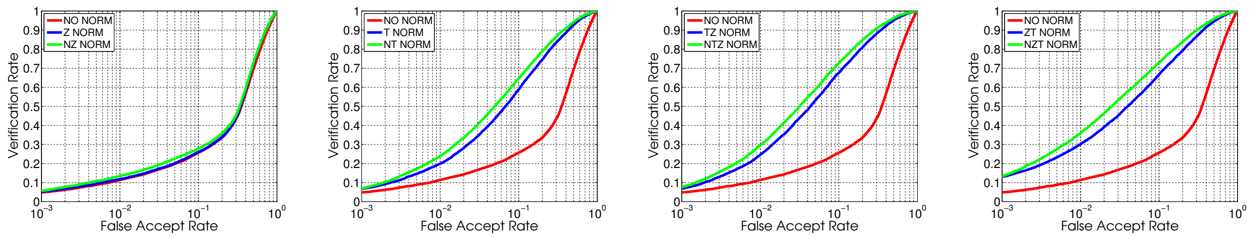
Note that before the non-parametric normalization techniques could be implemented, a target distribution has to be selected for Eq. 4. To this end, we assessed different types of target distributions, such as uniform, Gaussian and log-normal distributions, during our preliminary evaluations and concluded that the log-normal distribution ensures the best normalization performance. In fact, we did not observe any improvements at all with uniform or Gaussian target distributions. Hence, we select the following target probability density function (pdf) for our assessments:

$$p_{s'}(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right), \qquad (6)$$

where μ and σ denote the mean and standard deviation of the distribution. We select $\mu = 0$ and $\sigma = 0.5$ and make no effort to determine the parameters of the target distribution for optimal performance. An evaluation of the impact of the pdf's parameters on the final recognition performance is planned for our future work.

When estimating the statistics needed by the normalization techniques, we do not use any independent background data, but rely instead on the remaining gallery and query images to serve as our *t*-models and *z*-impostors. This setting corresponds (in most parts) to a closed-set experimentation scenario and is in our opinion indicative of the performance of the normalization techniques with independent background impostor data (i.e., open-set experiments).

Before we turn our attention to the actual recognition experiments, let us first look at Fig. 1. Here, the first three images from the left depict the histogram of the score distribution estimated on a sample target (gallery) image from the FRGC database and a set of *z*-norm impostors for: *a)* the raw non-normalized scores, *b)* the parametric *z*-norm normalized scores, and *c)* the proposed non-parametric *z*-norm normalized scores. Similarly, the three images on the right depict the histogram of the score distribution estimated on a sample query (test) image from the FRGC database and a set of *t*-norm models for: *d)* the raw non-normalized scores, *e)* the parametric *t*-norm normalized scores, and *f)* the proposed non-parametric *t*-norm normalized scores. The first thing to notice here is that the score distribution estimated for the *z*-norm is obviously non-Gaussian, which is a typical setting for FRGC database. Under the assumption of gaussianity, the parametric version of the *z*-norm tries to adjust the first and second statistical moment of the estimated distribution, which results in simple distribution scaling and consequently in sub-optimal recognition performance. The *nz*-norm, on the other hand, maps the entire distribution to a common



(a) Z-, NZ-norm (b) T-, NT-norm (c) TZ-, NTZ-norm (d) ZT-, NZT-norm

Figure 2. Effect of (parametric and non-parametric) score normalization techniques on the recognition performance

one and, hence, makes no assumptions regarding the nature of the initial score distribution. Clearly, this non-parametric procedure requires significantly more data than its parametric equivalent to be applicable, but since this data can be generated off-line this is of no major concern. The score distribution estimated for the *t*-norm, on the other hand, coincides better with the Gaussian assumption. Here, the simple scaling procedure performed by the *t*-norm seems to be sufficient for score normalization, but is, as will be shown in the remainder, still inferior to the non-parametric *nt*-norm when it comes to the final recognition performance.

We present the results of our experiments in the form of Receiver Operating Characteristic (ROC) curves, which plot the verification rate (VER) of the assessed recognition system against the false accept rate (FAR) on a semi-log scale. From Fig. 2, where the effect of each of the assessed parametric normalization techniques is compared to the effect of their non-parametric counterparts as well as the baseline performance of the system without any normalization techniques, we can see that all normalization techniques except for the *z*- and *nz*-norm result in noteworthy performance gains when compared to the baseline performance (denoted as NO NORM in the figures). Furthermore, we observe that all non-parametric normalization techniques consistently improve upon their parametric equivalents.

Fig. 3 shows all generated ROC curves in one graph. We can see that both types of *z*-norm perform the worst while the overall best performance is achieved by the *nzt*-normalization technique. Decent results are also achieved with the *ntz*-normalization technique, which in fact performs a little bit better than the *nzt*-technique at the higher values of the FAR. This is also evidenced by the characteristic error and verification rates tabulated in Table 1. Here, the first column corresponds to the minimal achievable half total error rate (minHTER), the second column corresponds to the equal error rate (EER), the third to the verification rate at the false accept rate of 0.1% (FAR_01) and the fourth column corresponds

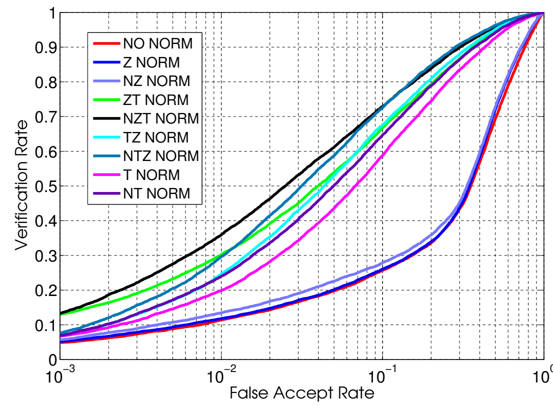


Figure 3. ROC curves for parametric and non-parametric normalization techniques

to the verification rate at the FAR of 1% (FAR_1). We can see that for all observed operating points, the non-parametric normalization techniques consistently outperform their parametric counterparts while all normalization techniques (parametric and non-parametric) improve upon the case where no normalization is used.

The results of our experiments show that non-parametric normalization techniques are capable of delivering additional performance gains when compared to the commonly used parametric normalization techniques, but at least for the schemes were the *t*-norm is involved induce a computational overhead as the data needed to remap the score distribution needs to be generated on-line.

5. Conclusion

We have introduced a new class of non-parametric score normalization techniques, which relax the Gaussian assumption commonly used with the classical parametric normalization techniques, such as the *z*- or *t*-score normalization procedures. Instead, the proposed

Table 1. Quantitative comparison of the normalization techniques

Norm.	minHTER	EER	FAR_01	FAR_1
No norm	0.402	0.413	0.049	0.114
Z norm	0.388	0.407	0.052	0.118
NZ norm	0.385	0.401	0.057	0.135
ZT norm	0.202	0.204	0.129	0.302
NZT norm	0.177	0.178	**0.133**	**0.359**
T norm	0.226	0.226	0.067	0.199
NT norm	0.205	0.206	0.069	0.240
TZ norm	0.194	0.196	0.069	0.247
NTZ norm	**0.173**	**0.174**	0.076	0.296

class of non-parametric score normalization techniques relies on a common target distribution to normalize the scores. We experimentally show that the proposed non-parametric normalization techniques consistently outperform their parametric counterparts when the target distribution takes a log-normal form, albeit at the expense of a higher computational load. As part of our future research we plan to further investigate the effect the shape of the target distribution has on the normalization/recognition performance as well as the possibilities of using composite (i.e., combinations of parametric and non-parametric) normalization schemes.

Acknowledgements

This work has been supported in parts by the national research program P2-0250(C) Metrology and Biometric Systems, the postdoctoral project BAMBI (ARRS ID Z2-4214) and by the European Union, European Regional Fund, in scope of the framework of the Operational Programme for Strengthening Regional Development Potentials for the Period 2007-2013, contract No. 3211-10-000467 (KC Class).

References

[1] T. Ahonen and M. Pietikäinen. Pixelwise local binary pattern models of faces using kernel density estimation. In *Proc. of ICB '09*, volume 1, pages 52–61, 2009.

[2] R. Auckenthaler, M. Careya, and H. Lloyd-Thomas. Score normalization for text-independent speaker verification systems. *Dig. Sig. Proc.*, 10(1–3):42–54, 2000.

[3] J. Fierrez-Aguilar, J. Ortega-Garcia, and J. Gonzalez-Rodriguez. Target dependent score normalization techniques and their application to signature verification. *IEEE TSMC - Part C*, 35(3):418–425, 2005.

[4] R. Gonzales and R. Woods. *Digital Image Processing, 3rd Ed.* Prentice Hall, New Yersey, 2008.

[5] C. Liu. The bayes decision rule induced similarity measures. *IEEE TPAMI*, 29(6):1086 – 1090, 2007.

[6] J. Mariethoz and S. Bengio. A unified framework for score normalization techniques applied to TI speaker verification. *IEEE Sig. Pr. Lett.*, 12(7):532–535, 2005.

[7] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *Proc. of CVPR*, volume 1, pages 947 – 954, 2005.

[8] N. Poh and S. Bengio. F-ratio client-dependent normalisation on biometric authentication tasks. In *Proc. of ICASSP*, volume 1, pages 721–724, 2005.

[9] N. Poh, C. Chan, J. Kittler, S. Marcel, C. McCool, E. Rua, J. Castro, M. Villegas, R. Paredes, V. Štruc, N. Pavešić, A. Salah, H. Fang, and N. Costen. An evaluation of video-to-video face verification. *IEEE Tran. on Inf. Forensics and Security*, 5(4):781–801, 2010.

[10] N. Poh and J. Kittler. Incorporating model-specific score distribution in speaker verification systems. *IEEE Tr. Aud., Speech and Lang. Proc.*, 16(3):594–606, 2008.

[11] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[12] V. Štruc, J. Žibert, and N. Pavešić. Histogram remapping as a preprocessing step for robust face recognition. *WSEAS Trans. ISCA*, 6(3):520–529, 2009.

[13] R. Wallace, M. McLaren, C. McCool, and S. Marcel. Cross-pollination of normalization techniques from speaker to face authentication using gaussian mixture models. *IEEE IFS*, 7(2):553–562, 2012.

[14] F. Yang, S. Shan, B. Ma, X. Chen, and W. Gao. Using score norm. tech. to solve the score variation problem in face authentication. *Springer LNCS*, 3781:31–38, 2005.