Beyond parametric score normalisation in biometric verification systems

Vitomir Štruc[1], Jerneja Žganec-Gros[2], Boštjan Vesnicer[2], Nikola Pavešić[1]

[1]Faculty of Electrical Engineering, University of Ljubljana, Tržaška 25, SI-1000 Ljubljana, Slovenia
[2]Alpineon d.o.o., Ulica lga Grudna 15, SI-1000, Ljubljana, Slovenia
E-mail: vitomir.struc@fe.uni-lj.si

Abstract: Similarity scores represent the basis for identity inference in biometric verification systems. However, because of the so-called miss-matched conditions across enrollment and probe samples and identity-dependent factors these scores typically exhibit statistical variations that affect the verification performance of biometric systems. To mitigate these variations, score-normalisation techniques, such as the *z*-norm, the *t*-norm or the *zt*-norm, are commonly adopted. In this study, the authors study the problem of score normalisation in the scope of biometric verification and introduce a new class of non-parametric normalisation techniques, which make no assumptions regarding the shape of the distribution from which the scores are drawn (as the parametric techniques do). Instead, they estimate the shape of the score distribution and use the estimate to map the initial distribution to a common (predefined) distribution. Based on the new class of normalisation techniques they also develop a hybrid normalisation scheme that combines non-parametric and parametric techniques into hybrid two-step procedures. They evaluate the performance of the non-parametric and hybrid techniques in face-verification experiments on the FRGCv2 and SCFace databases and show that the non-parametric techniques outperform their parametric counterparts and that the hybrid procedure is not only feasible, but also retains some desirable characteristics from both the non-parametric and the parametric techniques.

1 Introduction

Biometric verification systems typically rely on a matching module to measure the similarity (or dissimilarity) between the feature representation extracted from the 'live' biometric sample (the probe sample) and the template of a given identity stored in the systems database. Here, the template can be a feature representation, a classifier with a corresponding discriminant function or some other type of mathematical model that can be probed for its similarity with the input biometric sample. The output of the matching module is a matching (similarity/dissimilarity) score, which is compared to a decision threshold to make a decision regarding the identity of the 'live' biometric sample. However, because of changes in the conditions across different enrollment and probe samples, similarity scores typically exhibit statistical variations [1], which negatively affect the recognition performance and render approaches relying on a single, global decision threshold as suboptimal. These statistical variations are as well as being dependent on varying external conditions often also identity dependent, which is known in the literature as the Dodington Zoo effect [2].

To mitigate the problem of score variation the computed matching score is commonly subjected to a score-normalisation technique, which in a sense calibrates the score and makes it comparable to a global decision threshold. In the context of biometric verification, ('parametric') normalisation techniques, which assume that the matching score is drawn

from a Gaussian-shaped distribution and that adjusting this distribution to zero mean and unit variance successfully alleviates the score-variation problem, have emerged as the most popular. Since different strategies can be adopted to produce the required estimates of the first and second statistical moment of the Gaussian distribution, different normalisation techniques have been proposed by different researchers (e.g. [3–6]).

In this paper, we build on our work from [7, 8] and introduce a new class of score-normalisation techniques, which make no assumption regarding the shape of the distribution from which the matching scores are drawn. We present a simple procedure based on the rank transform for mapping the initial score distribution to a common (predefined) one. Although the basic idea of adjusting the score distribution to a common one is still the same as in the case of parametric techniques, this class of so-called 'non-parametric' score-normalisation techniques relaxes the assumptions pertaining to the shape of the score distribution and is, therefore expected to have an advantage when it comes to the verification performance after normalisation. Note that a parametric technique exploiting a similar idea was already proposed in [9], where it was shown explicitly that relaxing the Gaussian assumption is indeed beneficial for the recognition performance.

Although the assumption-free nature of the non-parametric normalisation techniques is expected to positively affect the verification performance of a given biometric system, it

comes at a price. As will be elaborated in the remainder of the paper, non-parametric techniques require a much higher number of score samples than their parametric counterparts for the normaliaation procedure and this then results in a significantly higher computational complexity. The reason for this setting is the fact that non-parametric normalisation techniques operate on the entire score distribution, which needs to be estimated reliably, while the parametric techniques rely only on the estimation of two parameters. These issues may limit the deployment of non-parametric normalisation techniques in verification systems in which the operational speed is crucial.

We will show that for certain types of score-normalisation techniques it is possible to retain the performance gains because of the non-parametric normalisation, while preserving the (run-time) computational complexity of the parametric normalisation techniques. This, however, applies only to score-normalisation techniques that combine two normalisation techniques into a single two-step procedure. An example of such a two-step normalisation technique can be found in the popular *zt*-norm, which combines the so-called *z*- and *t*-norms [Details on the *t*-, *z*- and *zt*-norms are given in the remainder.] [3, 10, 11]. The *zt*-norm is considered to be amongst the most effective score-normalisation techniques and is regularly used in the fields of speaker- and signature-verification [3, 4]. Instead of performing the *zt*-norm in either a pure parametric or a pure non-parametric manner, in this paper, we propose to use a 'hybrid' technique, where the *z*-norm step, which can be conducted off-line, is performed non-parametrically, and the *t*-norm step is performed parametrically.

Our experimental assessments, conducted on the Face Recognition Grand Challenge (FRGC) and SCFace databases, show that the non-parametric score-normalisation techniques indeed ensure a better verification performance than their parametric equivalents and that the proposed hybrid normalisation scheme still outperforms its parametric counterpart, while exhibiting the same computational complexity.

The rest of the paper is structured as follows. In Section 2, we first describe the theoretical background required for understanding the problem of score normalisation and review the existing work done in this field. In Section 3, we introduce the new class of non-parametric and hybrid score-normalisation techniques and elaborate on their characteristics in Section 4. We present experiments on the FRGCv2 and SCFace databases in Section 5 and conclude the paper with some final comments and directions for future work in Section 6.

2 Theoretical background

2.1 Problem formulation

Score normalisation represents a problem that arises in the context of biometric verification [Score normalisation is important in other areas (e.g. classifier fusion [12, 13]) as well, but in this paper, we focuson the problem of biometric verification.]. Before approaching score normalisation, it is, therefore necessary to define the problem of biometric verification in a more formal manner.

Let us assume that there are N identities enroled in the given biometric system and that these identities are labelled with $C_1, C_2, ..., C_N$. Furthermore, let us assume that we are given an input feature vector $\boldsymbol{x}$ and a claimed identity C_i from the pool of the N enrolled identities. The goal of

biometric verification is to assign the pair $(C_i, \boldsymbol{x})$ to class w_1 (a genuine/client claim) if the claim of identity is accepted, and to class w_2 (a false/impostor claim) otherwise. Typically, the acceptance of the claim is determined based on the return value [The matching or similarity scores] of the scoring function $s = \delta_i(\boldsymbol{x})$ associated with the identity C_i [14, 15], that is,

$$(C_i, \boldsymbol{x}) = \begin{cases} w_1, & \text{if} \quad \delta_i(\boldsymbol{x}) \leq \Delta \text{ for } i \in \{1, 2, \ldots, N\} \\ w_2, & \text{otherwise} \end{cases}$$
(1)

where Δ stands for the so-called decision threshold [Here, we assume that the scoring function $\delta_i(.)$ returns small values for large similarities and large values for small similarities.].

There are usually at least two possibilities for how to compute the matching score for a given probe vector $\boldsymbol{x}$ and a claimed identity C_i, which depend on the way the identities are modelled in the biometric system. In the case, the identities are modelled using a classifier, the matching score is produced based on a discriminant function associated with the classifier, where the return value of the discriminant function serves as an indicator of class-membership. When the identities are not represented with a classifier, but with some other type of template, the matching score is typically generated by measuring the similarity of the probe feature vector and the enrolled template using some similarity measure. In the remainder of the paper, we will refer to any function returning a matching score, albeit a discriminant function or a similarity measure, as a 'scoring function'.

From a generative point of view, verification can be seen as a two-class generating process, w_1 and w_2, that generate observations s according to the probability density functions (PDFs) $p(s/w_1)$ and $p(s/w_2)$ [16]. Based on this formulation and a specific value for the decision threshold Δ, it is possible to define the total Bayes' error of the verification process as

$$ER = \int_{-\infty}^{\Delta} p(s|w_2)p(w_2)\,\mathrm{d}s + \int_{\Delta}^{\infty} p(s|w_1)p(w_1)\,\mathrm{d}s \quad (2)$$

where $p(w_2)$ and $p(w_1)$ represent the prior probabilities of the two classes, w_2 and w_1, respectively.

To illustrate the importance of score normalisation let us have a look at the simple toy example presented in Fig. 1. Here, the classes w_1 and w_2 rely on two base classifiers (identities) C_1 and C_2 to generate the scores s [5, 17]. The green plot represents the client-score distribution $p(s|w_1)$, the black plot represents the impostor-score distribution $p(s|w_2)$, the dotted red and blue plots represent the identity-specific client-score distributions $p(s|w_1, C_1)$ and $p(s|w_1, C_2)$, and the dotted magenta and cyan plots represent the identity-specific impostor-score distributions $p(s|w_2, C_1)$ and $p(s|w_2, C_2)$, respectively. In this example, the identity-specific class-conditional distributions form the basis for computing the impostor- and client-score distributions (i.e. the green plot represents a weighted sum of the red and blue plots and the black plot represents a weighted sum of the cyan and magenta plots). Assume for the moment that the criterion for determining the decision threshold was an equal contribution of both terms of (2) to the Bayes' error. As we can see from Fig. 1, this criterion results in the threshold value of Δ if only the class-conditional distributions $p(s|w_1)$ and $p(s|w_2)$ are
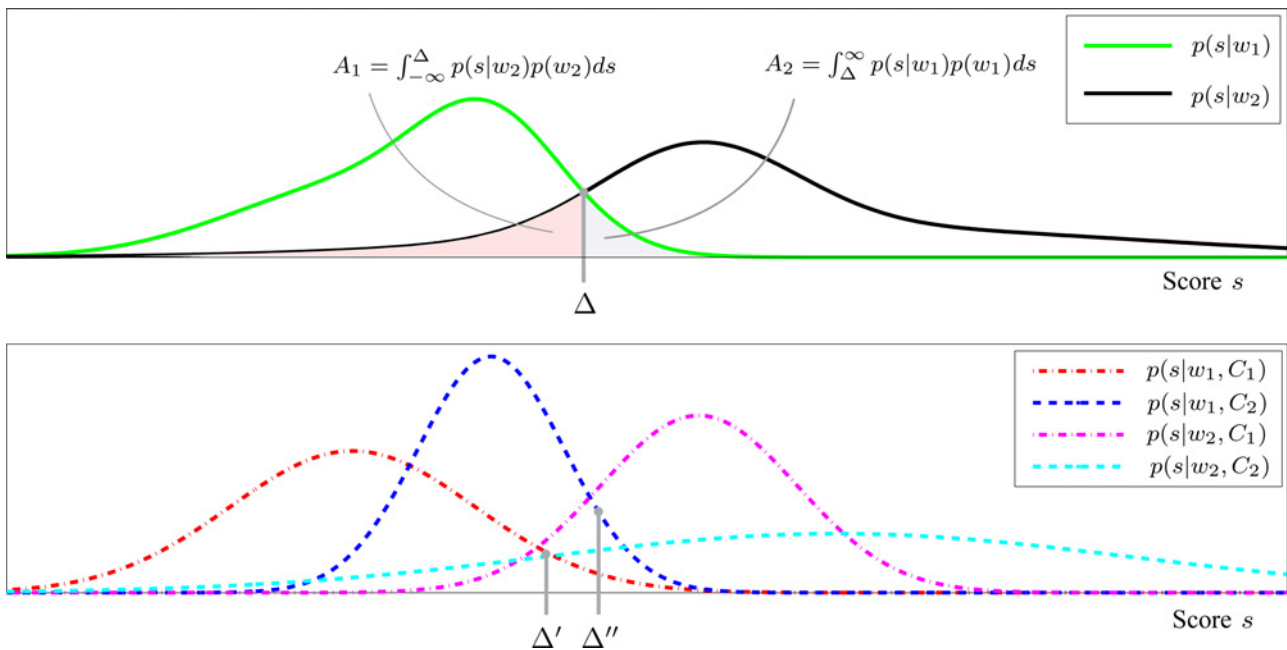
Fig. 1 *Illustration of the role of score normalisation*

The plots in the upper figure show the impostor and client-score distributions, while the plots in the lower figure show identity-specific impostor and client-score distributions. The distributions in the upper figure were created based on the distributions in the lower figure (best viewed in colour). **For all plots equal priors are assumed**. It should be noted that the decision thresholds for the EERs are set in a such a way that the cumulative density functions of the presented densities take the same value, for example, $A1 = A_2$

considered. However, if the identity-specific information is considered as well, the decision threshold for the identities C_1 and C_2 has to be set adaptively at Δ' and Δ'', respectively, to ensure the optimal operation of the system with respect to the selected criterion (assuming equal priors for the two classes) [18–20]. Another possibility is to apply score-normalisation techniques to the computed scores s with the goal of aligning the identity-specific class-conditional [Here, the term 'class' refers to either the impostor or the client class.] score distributions $p(s|w_1, C_i)$ and/or $p(s|w_2, C_i)$, for $i \in \{1, 2, …, N\}$ and making the scores comparable to a single global decision threshold [3, 5, 10]. Note that the presented example applies only to target-centric score-normalisation techniques (such as the z-norm). However, a similar example could also be presented for the probe-centric techniques (such as the t-norm).

2.2 Score normalisation – background and related work

Score-normalisation techniques can be divided into two groups based on the nature of the scores that are used for the normalisation. If the scores used by the normalisation technique are generated based on client-verification attempts, then the technique is said to be 'client-centric' and, similarly, if the scores used by the normalisation technique are generated based on impostor-verification attempts, then the normalisation technique is said to be 'impostor-centric'. As the data for producing client scores is usually not available (or extremely scarce), most of the existing techniques fall into the latter group, even though attempts have been made to develop client-centric score-normalisation techniques as well [20].

Among the group of impostor-centric score-normalisation techniques, the zero-normalisation or z-norm, the 'test-normalisation' or t-norm [3] and combinations of the two

(zt-norm) are among the most popular. Promising efforts on incorporating client scores into impostor-centric normalisation techniques (i.e. efforts towards 'client-impostor-centric' methods) have also been proposed in the literature in the form of the F-norm [11], the EER-norm [4], the MS-LLR-norm [5] and other recently proposed techniques, e.g. [21, 22].

Formally, score-normalisation techniques try to define a mapping ψ [5]

$$\psi: s \rightarrow s' \tag{3}$$

in such a way that the resulting normalised scores s' are well calibrated and, thus, comparable to a single global decision threshold. In the above equation, s denotes a raw score representing the output of the matching module of a biometric verification system and s' stands for the normalised version of the score. The total Bayes' error of the verification process after normalisation can then be written as

$$ER' = \int_{-\infty}^{\Delta_\psi} p(s'|w_2)p(w_1)\,\mathrm{d}s' + \int_{\Delta_\psi}^{\infty} p(s'|w_1)p(w_1)\,\mathrm{d}s' \tag{4}$$

where Δ_ψ again defines a decision threshold determined based on the selected performance criterion. Obviously, the goal of the score normalisation is to ensure that $ER' < ER$.

Impostor-centric score-normalisation techniques typically define the mapping ψ based on the impostor-score distributions $p(s|w_2)$ [7, 8]. The most popular techniques from this group are presented in the next section.

2.3 Parametric solutions

Parametric score-normalisation techniques commonly assume a certain shape for the class-conditional score-distributions and then apply normalisation steps in accordance with the

assumptions made. The most popular impostor-centric score-normalisation techniques, such as the z- or t-norm, for example, assume that the impostor-score distribution $p(s|w_2)$ (estimated based on the model of the claimed identity or the given probe vector) takes a Gaussian form, that is,

$$p(s|w_2) = \mathcal{N}(s; \mu, \sigma) \qquad (5)$$

where μ and σ denote the mean and standard deviation of the impostor-score distribution. Score normalisation is then conducted by shifting and scaling the assumed Gaussian distribution to zero mean and unit variance

$$\psi(s) = s' = \frac{s - \mu}{\sigma} \qquad (6)$$

As we can see, the goal of these techniques is to normalise the impostor-score distributions for all the identities (in the case of the z-norm) or for each probe vector $\boldsymbol{x}$ (in the case of the t-norm) to $\mathcal{N}(s'; 0, 1)$ and to ensure that the scores from all the verification attempts are drawn from the same distribution. This in turn suggests that the scores are well calibrated and can be compared to a single global threshold.

The z- and t-norm both use (6) to normalise the impostor-score distributions to $\mathcal{N}(s'; 0, 1)$, with the difference being that the z-norm generates the required scores by subjecting a number of probe samples (or z-impostors) to the scoring function $\delta_i(.)$ associated with the target/claimed identity (i.e. C_i), while the t-norm generates its scores by subjecting the given probe sample to a number of scoring functions $\delta_j(.)$ (or t-models) corresponding to some background identities $C_{1t}, C_{2t}, \ldots, C_{jt}, \ldots, C_{mt}$, where m denotes the total number of said identities (t-models). Based on the generated score sets each type of score normalisation computes its μ and σ, respectively [7, 8].

3 Beyond parametric techniques

Parametric score-normalisation techniques operate by making certain assumptions regarding the shape of the relevant score distributions and by adjusting the parameters of the assumed distributions to normalise the scores. In this section, we introduce a new class of non-parametric normalisation techniques that make no such assumptions and show how they can be combined with parametric techniques into hybrid normalisation approaches.

3.1 Non-parametric score normalisation

Conducting score normalisation in a non-parametric manner pursues the same goal as conducting the normalisation procedure in a parametric way, namely, making the normalised scores comparable to a single, global decision threshold. However, unlike parametric techniques, the non-parametric approaches do not make any assumptions regarding the shape of the impostor-score distribution $p(s|w_2)$.

To be able to relax the Gaussian assumption, the non-parametric normalisation techniques need to estimate the PDF of the impostor-score distribution $p(s|w_2)$, which can be achieved by using kernel density estimation (KDE) [23]. Once the PDF is estimated, the impostor-score distribution can be normalized by mapping it to a predefined shape. This procedure forms the basis for the work presented in [7, 8], but requires an estimate of the open hyper-parameters of the KDE and is computationally relatively intense. In this paper we introduce a slightly

different approach, which follows the same basic steps, but relies on a rank transform, has no open hyper-parameters and is computationally simpler.

Non-parametric score normalization can be formally described as follows: let ρ be a random variable with the property [8, 24]

$$\rho = F(s) = \int_{q=-\infty}^{s} p_s(q)\,\mathrm{d}q \qquad (7)$$

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

$$\rho = G(s') = \int_{x=-\infty}^{s'} p_{s'}(x)\,\mathrm{d}x \qquad (8)$$

where x again denotes a dummy variable for the integration. If we assume that the PDF $p_s(q)$ represents our impostor-score distribution $p(s|w_2)$ and that $p_{s'}(x)$ represents the PDF of a predefined target distribution, then the class of non-parametric score-normalisation techniques can be described as follows

$$\psi(s) = G^{-1}(\rho) = G^{-1}(F(s)) = s' \qquad (9)$$

where $G(.)$ and $F(.)$ denote the cumulative density functions (CDF) of their corresponding arguments and $G^{-1}(.)$ represents the inverse of the CDF [7]. Using the above expressions, it is possible to implement most of the existing parametric score-normalisation techniques in a non-parametric fashion. This includes the popular z- and t-norms [7, 8].

In practice, it is not necessary to estimate the PDF of (7) explicitly using the KDE. Instead, the left-hand side of the equation can be computed directly using the so-called rank transform [25, 26]. With the rank transform, the score samples that would otherwise serve to estimate the PDF of (7) are first put in ascending order. Each score in the ordered sequence is then assigned an index (or rank) that represents the position of the score in the ordered sequence [In other words, each score is assigned a number (i.e. the rank) that counts the number of scores smaller than the currently observed score in the available set of scores.]. Once the rank R of each score is determined, the left-hand side of (7) can be computed as

$$\rho = F(s) = \frac{n - R + 0.5}{n} \qquad (10)$$

where n denotes the number of elements in the set of scores and the rank R corresponds to the input score s. Note again that with the presented procedure we use the available score samples to estimate a CDF [i.e. $F(s)$] rather than a PDF. From this point of view, the proposed technique could also be interpreted so as to relate to the copula family of algorithms from the field of multivariate statistics [27].

To illustrate the idea of non-parametric score normalisation an example of an impostor-score distribution was generated through preliminary experimentation on the FRGCv2 database. The distribution is shown in Fig. 2 (left). It is clear that the non-parametric z-norm (denoted as nz-norm) maps the entire score distribution to a predefined shape, while the parametric z-norm only shifts and scales the score distribution, but leaves its shape unchanged.
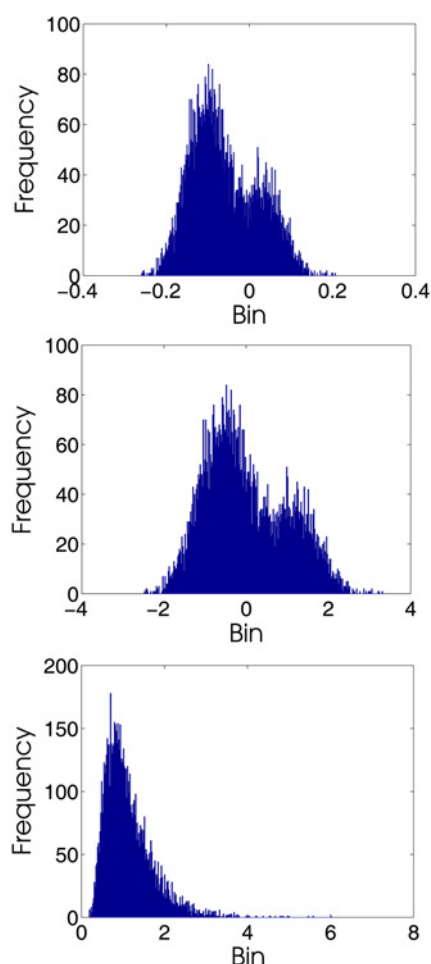
Fig. 2 *Illustration of the idea of non-parametric score normalisation*

The original score distribution (left) is only scaled and shifted with the parametric normalisation technique (middle)

With the non-parametric technique, the entire distribution is mapped to a predefined shape – in this case to a log-normal shape (right)

Let us emphasise again that the proposed non-parametric score-normalisation techniques allow us to conduct score normalisation without any assumptions regarding the shape of the impostor-score distribution. However, there are still a number of issues that need to be kept in mind:

• *Computational complexity:* non-parametric score-normalisation techniques require significantly more data for the normalisation procedure than their parametric counterparts and as such may result in an increased computational complexity of the verification system during the run-time, when the score distribution needs to be computed. This is especially true for normalisation techniques involving scores generated with the probe vector (such as the *t*-norm), as these cannot be pre-computed off-line.

• *Target distribution:* parametric score-normalisation techniques only shift and scale the score distributions in accordance with the assumptions they make; non-parametric techniques, on the other hand, require a target distribution for the mapping in (9). This target distribution is an open parameter of the non-parametric techniques that can affect their performance.

In the next section, we introduce hybrid score-normalisation techniques that try to mitigate the problem of

computational complexity by combining non-parametric and parametric score-normalisation techniques into hybrid procedures. The target distribution still remains an open parameter and needs to be selected empirically.

3.2 Hybrid score normalisation

We have pointed out before that popular score-normalisation techniques such as the *z*- and the *t*-norm rely on the same (Gaussian) assumption when normalising the scores, but differ in the way that the scores (which are used for estimating the parameters of the normalisation techniques) are created. Owing to this difference, score-normalisation techniques that operate on different scores are often combined into two-step normalisation techniques, where one technique is applied after the other. An example of such a two-step normalisation technique can be found in the *zt*-norm [3], where the *t*-norm technique is applied on *z*-norm normalised scores. However, there are also other examples, for example, [28].

Formally, two-step normalisation techniques can be defined based on the following mapping

$$s' = \psi(s'') \text{ and } s'' = \phi(s) \qquad (11)$$

where both ψ and ϕ denote the score-normalisation techniques, as defined in (3), s'' denotes the normalised version of the input score s after the first normalisation, and s' stands for the final normalised score. Although parametric techniques are often combined into two-step procedures, the authors of [7] suggested combining non-parametric techniques. As we emphasised in the previous section, such an approach is computationally demanding and is, therefore often not suitable for deployment in biometric verification systems.

To overcome this shortcoming, parametric and non-parametric normalisation techniques can be combined into a 'hybrid' (two-step) normalisation technique. In the case of the *zt*-norm this means that the first step of the normalisation technique, which relates to the score samples created based on the scoring function associated with the target/claimed identity, is conducted in a non-parametric manner, whereas the second step of the normalisation techniques, which relates to the sample-score population created with the probe biometric sample, is conducted in a parametric manner. Since the rank transform needed for the non-parametric *z*-norm step can be conducted off-line and the mapping in (9) can be implemented in the form of a look-up table, the first step of the hybrid normalisation does not induce any computational burden during the run-time. The second step of the hybrid technique is identical to the parametric version of the *zt*-norm.

Note that the proposed hybrid approach can be applied to any two-step normalisation technique, where the first step involves scores computed with scoring functions associated with the target/claimed identities, and is not limited only to the *zt*-norm studied in this paper.

4 Score-normalisation revisited

To provide a better insight into the characteristics of the different score-normalisation techniques and emphasise the differences between them conceptual diagrams of all the normalisation techniques are presented in Fig. 3. Here, 'no norm' represents the matching process without score
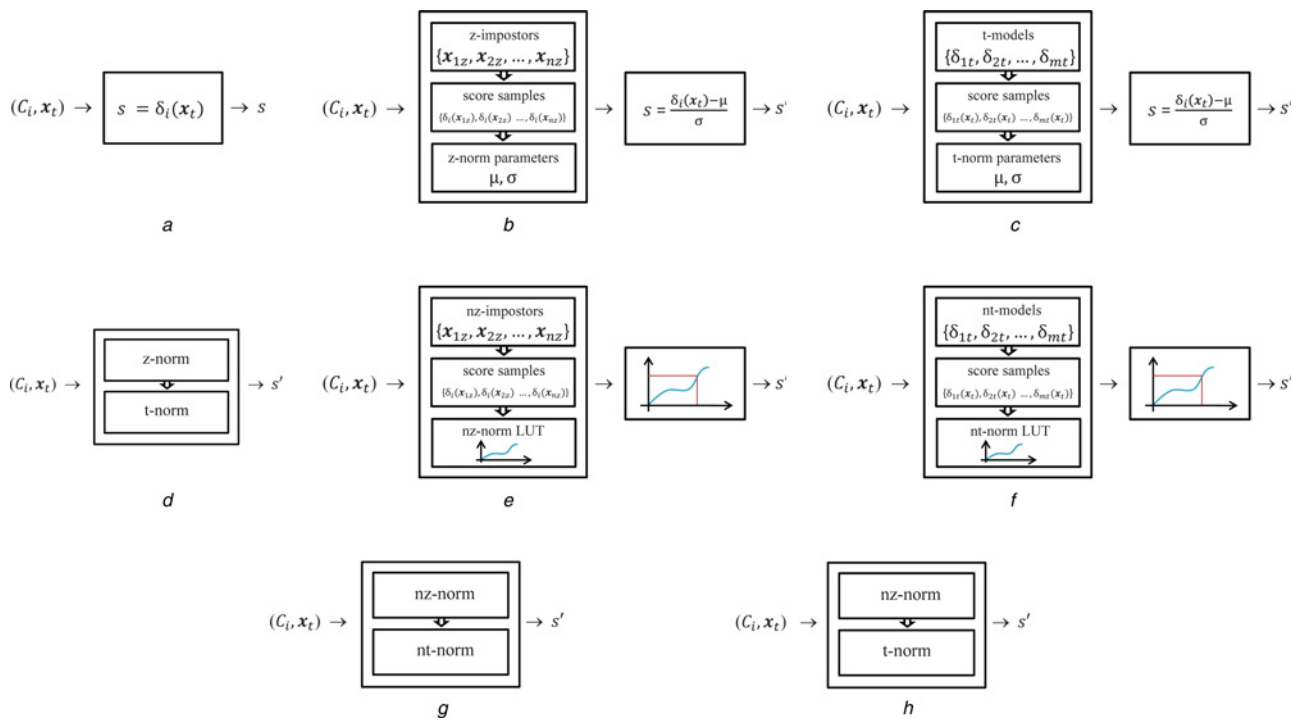
Fig. 3 *Conceptual diagrams of different score-normalisation techniques*

a No-norm
b z-norm
c t-norm
d zt-norm
e nz-norm
f nt-norm
g nzt-norm
h hzt-norm

Here, C_i denotes the claimed identity, x_t denotes the probe feature vector, δ_i stands for the scoring functions associated with the ith identity, $\delta_{\cdot t}$ stands for the scoring functions of the given t-model, n represents the number of z-impostor used to create the sample scores for the z-norm variants, m represents the number of t-models used to create the sample scores for the t-norm variants and the abbreviation LUT stands for 'look-up table'

normalisation, 'z-norm' (Fig. 3b), 't-norm' (Fig. 3c) and 'zt-norm' (Fig. 3d) represent the parametric techniques, 'nz-norm' (Fig. 3e), 'nt-norm' (Fig. 3f) and 'nzt-norm' (Fig. 3g) stand for the non-parametric versions of the techniques shown in Figs. 3b–d, and 'hzt-norm' (Fig. 3h) stands for the hybrid zt-norm. The same notation is also used in the remainder of this paper.

Note that the score-normalisation techniques, which operate on score distributions created by comparing the target tempate/model with a number of probe vectors (i.e. variants of the z-norm) [These techniques are often referred to as target-centric normalisation techniques], can typically pre-compute the parameters required for the normalisation and, therefore do not result in an additional computational burden. Normalisation techniques that operate on score distributions created by comparing the probe sample to a number of background identities (i.e. variants of the t-norm) [These techniques are often referred to as probe-centric normalisation techniques], on the other hand, need to compute the score samples on-line and, therefore require an additional computational effort. A summary of the storage requirements and the computational steps that need to be conducted on-line for different normalisation techniques is presented in Table 1.

Table 1 Comparison of different score-normalisation techniques

Normalisation technique	Steps needed				Need to store for each client	Need to compute on-line for each probe
	z-norm	*t*-norm	*nz*-norm	*nt*-norm		
z-norm	×	—	—	—	two parameters: μ, σ	n/a
t-norm	—	×	—	—	n/a	two parameters: μ, σ
nz-norm	—	—	×	—	look-up-table	n/a
nt-norm	—	—	—	×	n/a	look-up-table
zt-norm	×	×	—	—	two parameters: μ, σ	two parameters: μ, σ
nzt-norm	—	—	×	×	look-up-table	look-up-table
hzt-norm	—	×	×	—	look-up-table	two parameters: μ, σ

The table shows the steps needed for the given technique (not necessarily in the correct order) and the storage requirements per client/probe sample.
n/a stands for 'not applicable' since there is nothing to compute or store.

5 Experiments

5.1 Experimental database and setup

For the experiments presented in the remainder of this paper, we make use of two challenging databases, that is, the second version of the Face Recognition Grand Challenge database (FRGCv2) [29] and the SCFace database [30].

The FRGCv2 database features more than 40 000 facial images (of 466 distinct subjects) captured in diverse conditions (e.g. outdoors, indoors, under artificial lighting etc.) and exhibits characteristics that are known to affect the performance of existing face-recognition technology [31]. We select the most challenging experimental configuration defined for the FRGCv2 database for our experiments, that is, FRGCv2 Experiment 4. Here, 12 776 images are available for the training, 8024 images for the probe/query, and 16 028 images are at disposal for the gallery/target set.

The SCFace database contains 2080 images of 130 subjects (16 images per subject) taken in uncontrolled indoor environments using five video-surveillance cameras. The images are of varying quality and resolution, except for one image per subject, which represents a mug-shot image captured in controlled conditions. For the SCFace database we define a similar experimental protocol as for the FRGCv2 database. Thus, we divide the data into disjoint sets used for the training and testing. With the defined protocol, 480 images of 30 subjects are used for the training, whereas the rest is randomly partitioned into gallery/target and probe/query sets. The experimental protocols for both databases are summarised in Table 2.

Prior to the evaluation, we subject all of the images from both databases to a preprocessing procedure that geometrically normalises the facial images in accordance with manually marked eye-centre coordinates, crops the facial regions to a pre-defined size of 128×128 pixels and finally applies the histogram equalisation to the cropped images to improve the contrast and (partially) compensate for the lighting conditions present during the image acquisition. Some examples of the preprocessed facial images from both databases are shown in Fig. 4.

For the experiments on the FRGCv2 database, we implement a variant of principal component analysis (PCA) [32] and use it in conjunction with the whitened-cosine-based (or cosine Mahalanobis) similarity scoring procedure. Based on the implemented classification technique, we generate the $8014 \times 16\,028$ similarity (or matching score) matrix that forms the basis for assessing the score-normalisation techniques [Note that considering the entire $8014 \times 16\,028$ similarity matrix when producing

Table 2 Characteristics of FRGCv2 experiment 4

Database	Image set	#Images	Image quality
FRGCv2	training	12 776	controlled and uncontrolled
	target/gallery	16 028	controlled
	probe/query	8014	uncontrolled
SCFace	training	480	controlled and uncontrolled
	target/gallery	800	randomly partitioned
	probe/query	800	randomly partitioned

performance metrics on the FRGCv2 database corresponds to the so-called all vs. all experimental scenario.]. This setting more or less corresponds to a closed-set verification scenario.

For the experiments on the SCFace database, we use a more competent classification technique, that is, the technique proposed by Liu [33]. The technique relies on Gabor filters and Kernel Fisher Analysis for the classification. Like with the FRGCv2 database, we again generate a similarity matrix (of size 800×800) that forms the basis for the score normalisation.

5.2 Performance measures

We measure the effectiveness of the score-normalisation techniques based on the performance metrics derived from the false-acceptance error rate (FAR) and the false-rejection error rate (FRR). These error rates are defined as [34–37]

$$ \mathrm{FAR}(\Delta) = \frac{\left| \left\{ s_{\mathrm{imp}} \middle| s_{\mathrm{imp}} \leq \Delta \right\} \right|}{\left| \left\{ s_{\mathrm{imp}} \right\} \right|} \tag{12} $$

and

$$ \mathrm{FRR}(\Delta) = \frac{\left| \left\{ s_{\mathrm{cli}} \middle| s_{\mathrm{cli}} > \Delta \right\} \right|}{\left| \left\{ s_{\mathrm{cli}} \right\} \right|} \tag{13} $$

where $\{s_{\mathrm{cli}}\}$ and $\{s_{\mathrm{imp}}\}$ represent sets of client and impostor scores generated during the experiments, respectively, $|.|$ returns the cardinality of its argument, Δ denotes the decision threshold and that the inequalities are set in a way that assumes that dissimilarity measures were used to produce the matching scores. The FAR and FRR are typically used as estimates for the two terms of the Bayes error defined in (2).

Fig. 4 Samples from the FRGCv2 (left) and SCFace (right) databases

Note that the images are presented at different stages of the preprocessing chain – the histogram of the SCFace samples has not been equalised yet

a Geometric normalisation, cropping, scaling and histogram equalisation

b Geometric normalisation, cropping and scaling

Note that both the FAR and FRR are dependent on the value of the decision threshold Δ. Thus, selecting a specific value for this threshold results in different performance metrics. For our assessments, we selected three such metrics, that is, the equal error rate (EER), the verification rate at the false-acceptance error rate of 0.1% ($\mathrm{VER}_{0.1\mathrm{FAR}}$) and the verification rate at the false-acceptance rate of 1% ($\mathrm{VER}_{0.1\mathrm{FAR}}$). Here, the performance metrics can be computed as follows

$$\mathrm{EER} = \frac{1}{2}\big(\mathrm{FAR}(\Delta_{\mathrm{eer}}) + \mathrm{FRR}(\Delta_{\mathrm{eer}})\big) \qquad (14)$$

$$\mathrm{VER}_{0.1\mathrm{FAR}} = 1 - \mathrm{FRR}(\Delta_{\mathrm{ver01}}) \qquad (15)$$

and

$$\mathrm{VER}_{1\,\mathrm{FAR}} = 1 - \mathrm{FRR}(\Delta_{\mathrm{ver1}}) \qquad (16)$$

where the corresponding decision thresholds are defined as

$$\Delta_{\mathrm{eer}} = \underset{\Delta}{\mathrm{argmin}}\, |\mathrm{FAR}(\Delta) - \mathrm{FRR}(\Delta)| \qquad (17)$$

$$\Delta_{\mathrm{ver01}} = \underset{\Delta}{\mathrm{argmin}}\, |\mathrm{FAR}(\Delta) - 0.001| \qquad (18)$$

and

$$\Delta_{\mathrm{ver1}} = \underset{\Delta}{\mathrm{argmin}}\, |\mathrm{FAR}(\Delta) - 0.01| \qquad (19)$$

In addition to the presented performance metrics, we also compute the half total error (HTER) at each operating point. The HTER is defined as

$$\mathrm{HTER}_k = \frac{1}{2}\big(\mathrm{FAR}(\Delta_k) + \mathrm{FRR}(\Delta_k)\big) \qquad (20)$$

where $k \in \{_{\mathrm{eer}},\ _{\mathrm{ver01}},\ _{\mathrm{ver1}}\}$. Here, the HTER serves as an estimate of the Bayes error defined in (2) under the assumption of equal priors, that is, $p(w_1) = p(w_2) = 0.5$.

Next to the quantitative performance metrics presented above, we also use performance curves to present the results of our experiments. Specifically, we use receiver operating characteristic (ROC) curves, which plot the verification rate (VER) against the FAR for various values of the decision threshold Δ, and the detection error tradeoff (DET) curves, which plot the FAR against the FAR for various values of the decision threshold Δ. In contrast to the ROC curves, the DET curves use a non-linear scale for the x- and y-axes and, therefore better emphasise the performance differences between the assessed techniques [38].

5.3 Results and discussion

In our first series of verification experiments, we implement three popular (parametric) normalisation techniques and compare their performance to the performance of their non-parametric counterparts. Specifically, we implement the *t*-norm, *z*-norm and the two-step *zt*-norm as well as the non-parametric equivalents, which are denoted as *nt*-norm, *nz*-norm and *nzt*-norm in the remainder of this paper. Note that it would be possible to conduct a *tz*-norm procedure (by applying *z*-norm on *t* normalised scores); however, since this would involve enormous amounts of computation

for each test sample during the run-time, this options is not feasible in practice and is therefore omitted from our experiments. For this series of experiments, we use only the FRGCv2 database.

Before the non-parametric normalisation techniques can be implemented, a target distribution for (8) needs to be selected. To this end, we assess different types of target distributions when implementing the non-parametric normalisation techniques, that is, uniform, Gaussian and log-normal distributions. The three target distributions are defined as

$$p_{\mathrm{unif}}(s') = \frac{1}{b-a} \qquad (21)$$

where in our case $a = 0$ and $b = n$ [n denotes the number of scores used for the normalisation – see (10)]

$$p_{\mathrm{gaus}}(s') = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(s'-\mu)^2}{2\sigma^2}\right) \qquad (22)$$

where μ and σ represent the mean and the standard deviation of the Gaussian distribution, and

$$p_{\mathrm{logn}}(s') = \frac{1}{s'\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln s'-\mu)^2}{2\sigma^2}\right) \qquad (23)$$

where μ and σ represent the mean and the standard deviation of the log-normal distribution. For the experiments we select $\mu = 0$ and $\sigma = 1$ for the Gaussian target distribution and $\mu = 0$ and $\sigma = 0.5$ for the log-normal distribution.

The results from this series of experiments are shown in the form of DET and ROC curves in Fig. 5. Here, the pair of graphs on the upper-left (marked as Fig. 5*a*) presents a comparison of the baseline performance of the PCA technique (denoted as NO NORM) and the performance of the parametric and non-parametric versions of the *z*-norm. For the non-parametric *z*-norms the name in the brackets indicates the target distribution that was used to generate the performance curves. The pair of graphs on the upper-right (marked as Fig. 5*b*) shows the same comparison for the parametric *t*-norm and the different versions of the non-parametric *nt*-norm. The pair of graphs on the lower-left side of Fig. 5 (marked as Fig. 5*c*) shows the comparison of the baseline PCA technique and the *zt*- and *nzt*-norms, while the last graphs on the lower-right (marked as Fig. 5*d*) depict a comparison of all the assessed techniques in this series of experiments and their relative ranking – for the non-parametric techniques only the results for the log-normal target distributions are shown. The same comparison is also shown in Table 3, where the characteristic error rates are tabulated for the assessed techniques.

The first thing to notice from the presented results is the fact that except for the *z*-norms, where no significant improvement over the baseline performance was observed, all the other techniques resulted in significant performance gains. In general, the two-step normalisation techniques outperformed the single-step ones and the non-parametric normalisation techniques consistently outperformed their parametric equivalents along most operating points of the DET and ROC curves when the log-normal distribution was used as the target distribution. When the uniform distribution was used as the target distribution for the non-parametric score-normalisation techniques, the performance of the biometric verification systems increased
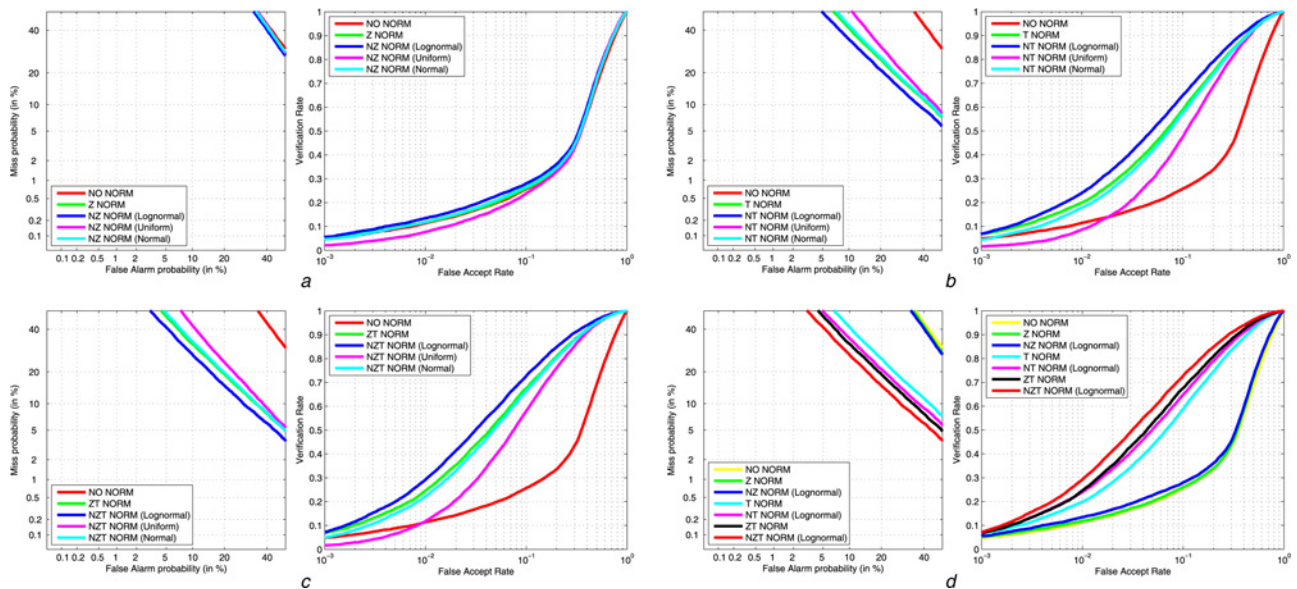
Fig. 5 *Effect of (parametric and non-parametric) score normalisation on the verification performance – DET (left) and ROC (right) curves generated during Experiment 4 on the FRGCv2 database (best viewed in colour)*

a Comparison no-, *z*-, *nz*-norm
b Comparison no-, *t*-, *nt*-norm
c Comparison no-, *zt*-, *nzt*-norm
d Comparison – all

for most operating points over the baseline (i.e. NO NORM), but was worse than that of the parametric equivalents. For the normal target distribution the performance of the non-parametric score-normalisation techniques was very similar to that of the parametric techniques, with very little difference over all the operating points for all the implemented methods.

Note that the improved performance of the non-parametric normalisation techniques (with a log-normal target distribution) comes at the expense of an increased computational complexity. The non-parametric techniques require significantly more data (i.e. a larger score population) to estimate the entire score-distribution reliably, while the parametric techniques only require enough data to estimate the mean and the variance. This fact results in the need for storing a larger background set of templates in a face-verification system to ensure a sufficient number of scores for the non-parametric normalisation techniques. If storage is no issue, the non-parametric score-normalisation

techniques can ensure a better verification performance than their parametric counterparts.

When looking at the characteristic error rates in Table 3, similar conclusions to the ones presented above can be made. Of particular interest here are the values of the HTER, which indicate that the Bayes error is indeed reduced with (most of) the score-normalisation techniques. Here, the HTER at the equal-error operating point equals the EER and is, therefore not tabulated separately in Table 3. We can again see that the non-parametric normalisation techniques with a log-normal target distribution ensure the best verification performance. We, therefore select this distribution for all of our subsequent experiments.

Another interesting issue relevant in the context of non-parametric score normalisation is the effect of the normalisation procedure on the unseen client-score distribution. With parametric score-normalisation techniques this distribution is simply shifted and scaled based on the mean and standard deviation computed from the

Table 3 Quantitative comparison of the score-normalisation techniques on the FRGCv2 database with experiment 4

| Perform. metric | No norm | Parametric techniques | | | Non-parametric techniques | | | | | | | | |
| | | | | | Uniform distribution | | | Log-normal distribution | | | Normal distribution | | |
		z	*t*	*zt*	*nz*	*nt*	*nzt*	*nz*	*nt*	*nzt*	*nz*	*nt*	*nzt*
EER	0.413	0.407	0.226	0.196	0.411	0.250	0.220	0.401	0.206	**0.173**	0.409	0.230	0.198
$VER_{0.1FAR}$	0.049	0.052	0.067	0.069	0.021	0.015	0.017	0.055	0.068	**0.072**	0.046	0.043	0.048
VER_{1FAR}	0.114	0.118	0.199	0.247	0.077	0.086	0.118	0.135	0.241	**0.263**	0.118	0.179	0.223
$HTER_{ver01}$	0.526	0.524	0.516	0.516	0.540	0.542	0.542	0.522	0.516	**0.514**	0.527	0.528	0.526
$HTER_{ver1}$	0.493	0.491	0.450	0.427	0.512	0.507	0.491	0.483	0.430	**0.401**	0.491	0.461	0.439

The results for the non-parametric techniques are presented for three different target distributions.
The best results for each presented performance metric are presented in bold.
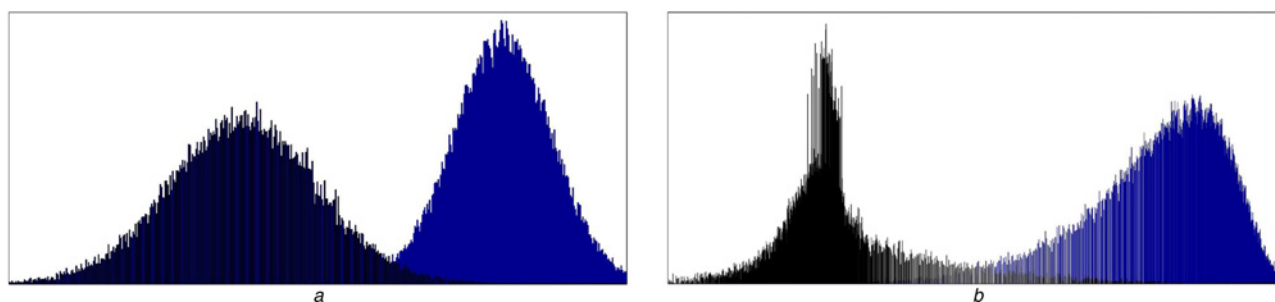
Fig. 6 *Effect of a non-parametric score-normalisation technique (with a log-normal target distribution) on synthetic client and impostor-score distributions*

a Sample histograms before normalisation
b Sample histograms after normalisation

impostor-score distribution. However, with the non-parametric techniques the client distribution obtains re-mapped in accordance with the CDF computed from the impostor-score distribution. To examine this effect more closely, we generate two synthetic score populations and estimate their distributions as shown in Fig. 6. Here, the light-blue distribution corresponds to the impostor distribution that forms the foundation for the non-parametric normalisation procedure and the dark-blue distribution corresponds to the client distribution, which is mapped along with the impostor distribution. Note that with a target log-normal distribution the impostor-score distribution takes an approximately log-normal shape after the normalisation, while for the client-score distribution most of the mass obtains concentrated around a specific value and heavy tails appear on both sides of the distribution. It is important to emphasise at this point that when implementing the non-parametric techniques, we need to ensure that the CDF from (7) (estimated from the impostor-score distribution) is defined over the domain of the client-score distribution as well, otherwise the part of

the client-score distribution over the undefined domain obtains mapped onto a single point.

So far, we have established the relative ranking of the score-normalisation techniques, examined some of the characteristics of the non-parametric score-normalisation procedures and demonstrated the importance of choosing an appropriate target distribution for the normalisation. The next issue we need to investigate is the feasibility of the hybrid two-step normalisation technique. In Section 3.2, where we have introduced the hybrid normalisation scheme, we proposed carrying out the first normalisation step in a non-parametric manner based on a pre-computed look-up table and conduct the second normalisation step parametrically. This procedure exhibits approximately the same computational complexity during the run-time as the parametric techniques, but hopefully results in an improved recognition performance. To assess the proposed hybrid two-step *zt*-norm (denoted as *hzt*-norm in the figures) we apply it to the FRGCv2 similarity matrix and compare the generated results to the remaining two two-step procedures, that is, the *zt*-norm and the *nzt*-norm, and the baseline
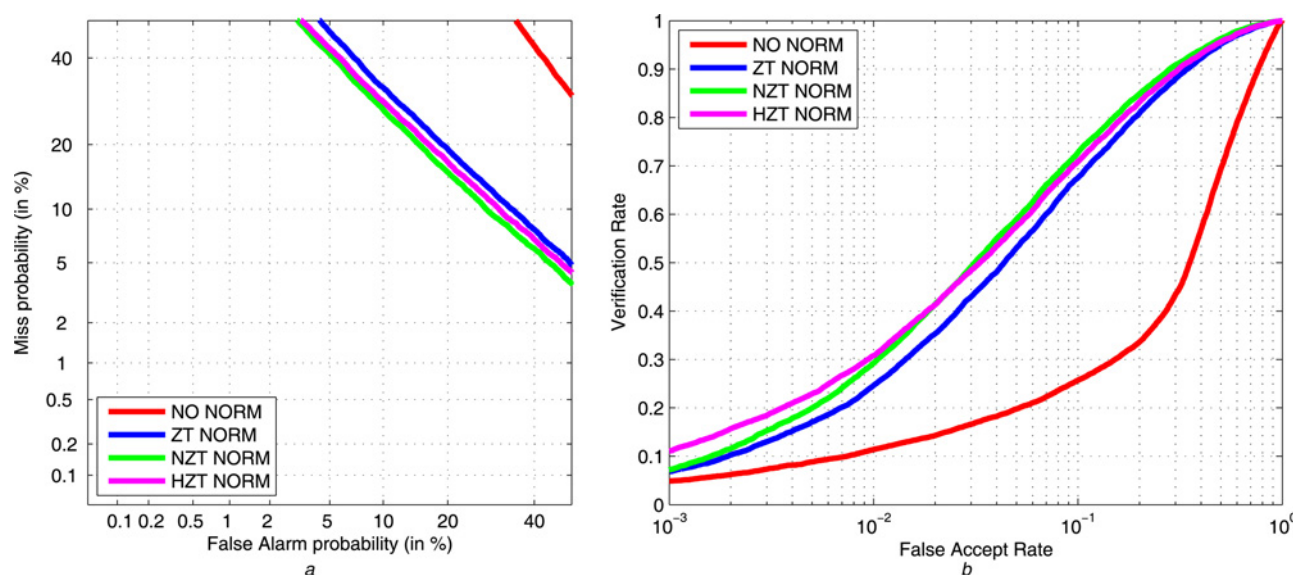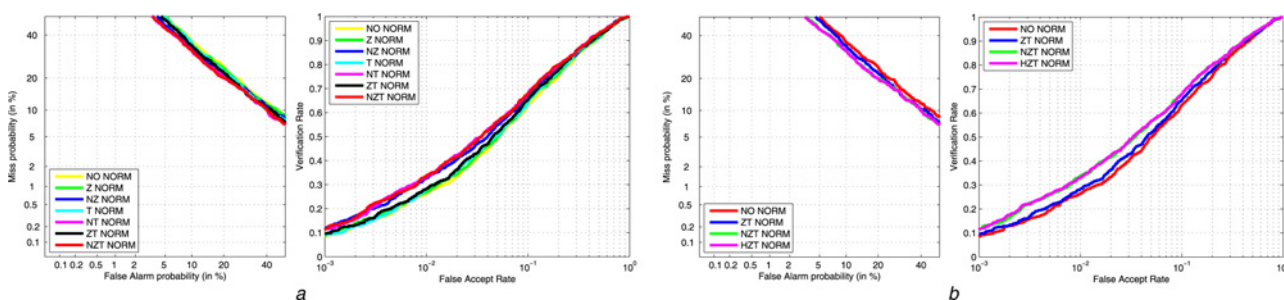

Fig. 7 *Comparison of the two-step techniques – DET (left) and ROC (right) curves generated during Experiment 4 on the FRGCv2 database (best viewed in colour)*

The target distribution for the non-parametric part of the HZT-normalisation technique was again chosen to be lognormal
The results for the *nzt* technique are shown here again for comparison purposes – they are the same as in Fig. 5, where the technique was labelled *nzt*-norm (lognormal)
a Comparison of parametric and non-parametric techniques
b Comparison of the two-step score-normalisation techniques

Fig. 8 *Effect of (parametric and non-parametric) score normalisation on the verification performance – DET (left) and ROC (right) curves generated during experiments on the SCFace database (best viewed in colour)*

a Comparison of parametric and non-parametric techniques
b Comparison of the two-step score normalization techniques

performance of the PCA. The results of the experiments are again shown in the form of DET (on the left) and ROC (on the right) curves in Fig. 7.

The hybrid *zt*-norm technique achieves the $VER_{0.1FAR}$ of 11.1%, the VER_{1FAR} of 30.8% and the EER of 18.3%. In comparison to the parametric two-step normalisation technique, the composite procedure produces better verification results along all the operating points of the DET and ROC curves. When compared to the non-parametric two-step normalisation techniques, the proposed hybrid procedure still results in a very competitive performance. At this point, we need to stress that the lower computational cost of the hybrid techniques comes at the expense of storage requirements, as a look-up table needs to stored in the verification system for each registered client. Furthermore, the lower computational load is ensured only for the run-time operation of the system (in verification mode), as the look-up table needs to be computed during the enrolment of a client.

To validate the findings made so far, we conduct additional experiments on the SCFace database using a state-of-the-art face-recognition technique (i.e. GaborKFA [33]). We again implement all the normalisation techniques and again use the log-normal target distribution for the non-parametric techniques. We apply all the normalisation approaches to our 800 × 800 similarity matrix and generate all the relevant performance metrics and curves. The results of this series of experiments are presented in Fig. 8 and Table 4.

As can be seen, the results of this series of experiments are much closer together than in the case of the FRGCv2 database. Although almost all of the score-normalisation techniques improve the verification performance over the baseline, this improvement is not as large as on the FRGCv2 database. The reasons for such a setting can undoubtedly be found in the characteristics of the database

and the better performing classifier, which leaves less room for an improvement through score normalisation. Overall, the non-parametric and hybrid normalisation procedures are again the best performing techniques considering all the operating points on the ROC and DET curves. The non-parametric techniques (with a log-normal target distribution) again outperform their parametric counterparts and the HTER is again reduced after the score normalisation for most of the assessed normalisation techniques. All in all, it seems that the findings made during the experiments on the FRGCv2 database, generalise to other databases as well.

In the last series of experiments, we tried to assess the significance of the obtained results. To this end, we observed the relative change in the EER with respect to the baseline performance without normalisation as suggested by Sun *et al.* [39]

$$\text{rel. change of EER} = \frac{EER_{norm} - EER_{baseline}}{EER_{baseline}} \quad (24)$$

Here, EER_{norm} stands for the EER after normalisation with the selected score-normaliaation technique and $EER_{baseline}$ for the EER achieved without score normalisation. A negative change of this metric implies an improvement over the baseline.

To establish the significance of the observed results on the FRGCv2 and SCFace databases, we randomly sub-sample the original similarity matrices 20 times and generate box plots for the relative change in the EERs. The results are shown in Fig. 9. Note that because of the smaller number of experiments on the SCFace database the dispersion of the results is much larger for this database than for the FRGCv2 database. In general, all the normalisation techniques improve significantly over the baseline on the

Table 4 Quantitative comparison of the score normalization techniques on the scface database

Perform. metric	No norm	Parametric techniques			Non-parametric techniques			Hybrid technique
		z	t	zt	nz	nt	nzt	hzt
EER	0.219	0.220	0.212	0.212	0.209	0.203	**0.200**	**0.200**
$VER_{0.1FAR}$	0.083	0.089	0.089	0.095	**0.120**	0.116	0.114	0.113
VER_{1FAR}	0.267	0.267	0.281	0.283	0.326	0.330	**0.338**	0.331
HTER $VER_{0.1FAR}$	0.506	0.506	0.506	0.503	**0.490**	0.492	0.493	0.494
HTER VER_{1FAR}	0.418	0.409	0.409	0.409	0.387	0.385	**0.381**	0.384

The results for the non-parametric techniques are presented for the log-normal target distributions.
The best results for each presented performance metric are presented in bold.
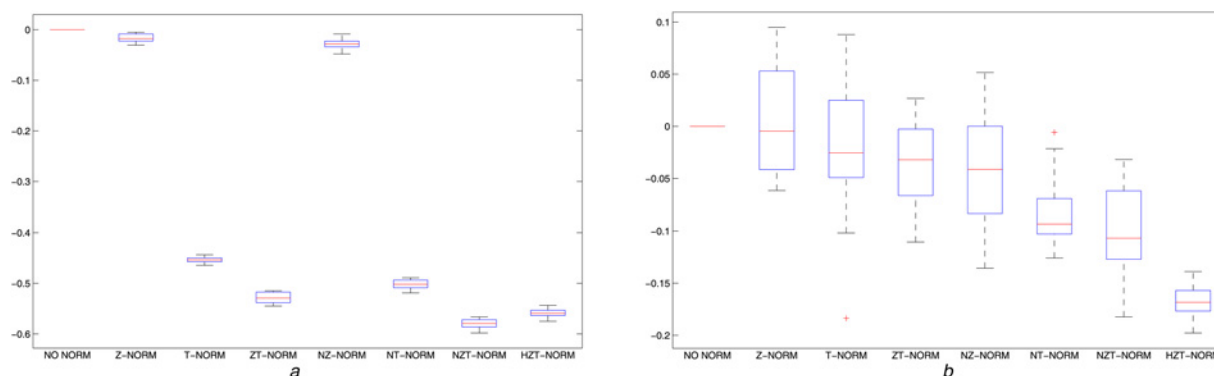
Fig. 9 *Box plots for some characteristic operating points on the FRGCv2 database (the plots were created by randomly sub-sampling the original ROC data 20 times)*

a Box plots for the FRGCv2 database
b Box plots for the SCFace database

FRGCv2 database with the non-parametric and hybrid two-step procedures significantly outperforming all the other normalisation techniques. On the SCFace database not all the normalisation techniques performed significantly better than the baseline. However, the non-parametric and hybrid two-step procedures significantly outperform the baseline and the hybrid procedure also performs significantly better than all the remaining normalisation techniques (except for the *nzt*-norm).

To sum up, the results of our experiments suggest that non-parametric score-normalisation techniques have the potential to improve upon the performance ensured by the parametric methods. Moreover, combining parametric and non-parametric methods into a hybrid normalisation procedure can reduce the run-time computational complexity of the normalisation procedure, while ensuring approximately the same recognition performance.

6 Conclusion

We have presented a new family of non-parametric techniques for the score normalisation in face-verification systems. We have shown that the techniques are capable of ensuring an improved verification performance when compared to their parametric counterparts, albeit at the expense of a higher computational complexity. Furthermore, we have demonstrated that parametric and non-parametric normalisation techniques can be combined into hybrid normalisation schemes to provide a trade-off between the computational complexity and the performance. As part of our future work, we plan to examine possibilities to incorporate the non-parametric normalisation approach into other possible client-impostor centric normalisation techniques and assess their performance in open-set verification experiments. Possible starting points for our future work are state-of-the-art normalisation schemes such as the one presented in [40].

7 Acknowledgments

8 References

1 Wallace, R., McLaren, M., McCool, C., Marcel, S.: 'Cross-pollination of normalization techniques from speaker to face authentication using Gaussian mixture models', *IEEE Inf. Forensics Sec.*, 2012, **7**, (2), pp. 553–562
2 Doddington, G., Liggett, W., Martin, A., Przybocki, M., Reynolds, D.: 'Sheep, goats, lambs and woves: a statistical analysis of speaker performance in the nist 1998 speaker recognition evaluation'. Proc. Int. Conf. on Spoken Language Processing (ICSLP), Sydney, Australia, 1998, vol. 1
3 Auckenthaler, R., Careya, M., Lloyd-Thomas, H.: 'Score normalization for text-independent speaker verification systems', *Digit. Signal Process.*, 2000, **10**, (1–3), pp. 42–54
4 Fierrez-Aguilar, J., Ortega-Garcia, J., Gonzalez-Rodriguez, J.: 'Target dependent score normalization techniques and their application to signature verification', *IEEE Trans. Syst. Man Cybern., C*, 2005, **35**, (3), pp. 418–425
5 Poh, N., Kittler, J.: 'Incorporating model-specific score distribution in speaker verification systems', *IEEE Trans. Audio Speech Lang. Process.*, 2008, **16**, (3), pp. 594–606
6 Mariethoz, J., Bengio, S.: 'A unified framework for score normalization techniques applied to text-independent speaker verification', *IEEE Signal Process. Lett.*, 2005, **12**, (7), pp. 532–535
7 Štruc, V., Žganec Gros, J., Pavešić, N.: 'Non-parametric score normalization for biometric verification systems'. Proc. ICPR, Tsukuba, Japan, 2012, pp. 2395–2399
8 Štruc, V., Pavešić, N., Žganec Gros, J.: 'Composite score normalization for face verification'. Proc. first Int. Workshop on Biometrics and Forensics (IWBF), Lisbon, Portugal, 2013, pp. 1–4
9 Merati, A., Poh, N., Kittler, J.: 'User-specific cohort selection and score normalization for biometric systems', *IEEE Trans. Inf. Forensics Sec.*, 2012, **7**, (4), pp. 1270–1277
10 Yang, F., Shan, S., Ma, B., Chen, X., Gao, W.: 'Using score normalization techniques to solve the score variation problem in face authentication', *LNCS*, 2005, **3781**, pp. 31–38
11 Poh, N., Bengio, S.: 'F-ratio client-dependent normalisation on biometric authentication tasks'. Proc. Int. Conf. on Acoustics, Speech and Signal Processing, 2005, vol. 1, pp. 721–724
12 Jain, A., Nandakumara, K., Ross, A.: 'Score normalization in multimodal biometric systems', *Pattern Recognit.*, 2005, **38**, pp. 2270–2285
13 Žibert, J., Mihelič, F.: 'Fusion of acoustic and prosodic features for speaker clustering'. Lecture Notes in Computer Science, Springer, 2013, pp. 210–217
14 Jain, A., Ross, A., Prabhakar, S.: 'An introduction to biometric recognition', *IEEE Trans. Curcuits Syst. Video Technol.*, 2004, **14**, (1), pp. 4–20

15 Štruc, V., Pavešić, N.: 'The corrected normalized correlation coefficient: a novel way of matching score calculation for lda-based face verification'. Proc. FSKD 2008, China, 2008, pp. 110–115

16 Kryszczuk, K., Drygajlo, A.: 'Improving biometric verification with class-independent quality information', *IET Signal Process.*, 2009, **3**, (4), pp. 310–321

17 Kryszczuk, K., Drygajlo, A.: 'Q-stack: uni- and multimodal classifier stacking with quality measures'. *Lecture Notes in Computer Science*, 2007, vol. 4472, pp. 367–376

18 Jonsson, K., Kittler, J., Li, Y., Matas, J.: 'Support vector machines for face authentication', *Image Vis. Comput.*, 2002, **20**, pp. 269–275

19 Chen, K.: 'Towards better making a decision in speaker verification', *Pattern Recognit.*, 2003, **36**, (2), pp. 329–346

20 Saeta, J., Hernando, J.: 'On the use of score pruning in speaker verification for speaker dependent threshold estimation'. The Speaker and Language Recognition Workshop (Odyssey), Toledo, Spain, 2004, pp. 215–218

21 Poh, N., Rattani, A., Tistarelli, M., Kittler, J.: 'Group-specific score normalization for biometric systems'. IEEE Computer Society Workshop on Biometrics (CVPR), 2010, pp. 38–45

22 Poh, N., Tistarelli, M.: 'Customizing biometric authentication systems via discriminative score calibration'. Proc. CVPR, 2012, pp. 2681–2686

23 Ahonen, T., Pietikäinen, M.: 'Pixelwise local binary pattern models of faces using kernel density estimation'. Proc. ICB '09, Alghero, Italy, 2009, vol. 1, pp. 52–61

24 Gonzales, R., Woods, R.: 'Digital image processing' (Prentice-Hall, New Jersey, 2008, 3rd edn.)

25 Pelecanos, J., Sridharam, S.: 'Feature warping for robust speaker verification'. Proc. Speaker Recognition Workshop (Odyssey), p. 213

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

27 Iyengar, S.G., Varshney, P.K., Damarla, T.: 'Biometric authentication: a copula based approach'. In Bhanu, B., Govindaraju, V. (Eds.), 'Multibiometrics for human identification' (Cambrdige University Press, 2011) p. 23

28 Poh, N., Merati, A., Kittler, J.: 'Adaptive client-impostor centric score normalization: a case study in fingerprint verification'. Proc. BTAS, Washington, USA, 2009, pp. 1–7

29 Phillips, P., Flynn, P., Scruggs, T., *et al.*: 'Overview of the face recognition grand challenge'. Proc. CVPR, 2005, vol. 1, pp. 947–954

30 Grgic, M., Delac, K., Grgic, S.: 'Scface – surveillance cameras face database', *Multimedia Tools Appl. J.*, 2011, **51**, (3), pp. 863–879

31 Vesnicer, B., Žganec Gros, J., Pavešić, N., Štruc, V.: 'Face recognition using simplified probabilistic linear discriminant analysis', *Int. J. Adv. Robot. Syst.*, 2012, **9**, pp. 1–10

32 Turk, M., Pentland, A.: 'Eigenfaces for recognition', *J. Cognitive Neurosci.*, 1991, **3**, (1), pp. 71–86

33 Liu, C.: 'Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2006, **28**, (5), pp. 725–737

34 Gajšek, R., Mihelič, F., Dobrišek, S.: 'Speaker state recognition using an hmm-based feature extraction method', *Comput. Speech Lang.*, 2013, **27**, (1), pp. 135–150

35 Gajšek, R., Mihelič, F., Dobrišek, S.: 'Analysis and assessment of state relevance in hmm-based feature extraction method'. Lecture Notes in Computer Science, Springer, 2012

36 Vesnicer, B., Mihelič, F.: 'The likelihood ratio decision criterion for nuisance attribute projection in gmm speaker verification', *EURASIP J. Adv. Signal Process.*, 2008, Article ID 786431, pp. 1–11

37 Günther, M., Costa-Pazo, A., Ding, C., *et al.*: 'The 2013 face recognition evaluation in mobile environment'. Proc. Sixth IAPR Int. Conf. on Biometrics, June 2013

38 Dobrišek, S., Žibert, J., Mihelič, F.: 'Towards the optimal minimization of a pronunciation dictionary model'. Lecture Notes in Computer Science, Springer, 2013, pp. 267–274

39 Sun, Y., Tistarelli, M., Poh, N.: 'Picture-specific cohort score normalization for face pair matching'. Proc. BTAS, 2013, vol. 1

40 Poh, N., Merati, A., Kittler, J.: 'Adaptive client-impostor centric score normalization: a case study in fingerprint verification'. IEEE Conf. on Biometrics: Theory, Applications and Systems (BTAS'09), 2009, pp. 1–7