Cross-pollination of normalisation techniques from speaker to face authentication using Gaussian mixture models

Roy Wallace, *Member, IEEE,* Mitchell McLaren, *Member, IEEE,* Christopher McCool, *Member, IEEE,* and Sébastien Marcel, *Member, IEEE*

***Abstract*—This paper applies score and feature normalisation techniques to parts-based Gaussian mixture model (GMM) face authentication. In particular, we propose to utilise techniques that are well established in state-of-the-art speaker authentication, and apply them to the face authentication task. For score normalisation, T-, Z- and ZT-norm techniques are evaluated. For feature normalisation, we propose a generalisation of feature warping to 2D images, which is applied to discrete cosine transform (DCT) features prior to modelling. Evaluation is performed on a range of challenging databases relevant to forensics and security, including surveillance and access control scenarios. The normalisation techniques are shown to generalise well to the face authentication task, resulting in relative improvements in half total error rate (HTER) of between 17% and 62%.**

***Index Terms*—biometrics, face authentication, face recognition, feature warping, score normalisation, Gaussian mixture modelling**

I. Introduction

FACE authentication remains a challenging problem because of the high degree of variability across images, which can be influenced by lighting conditions, facial expression and pose. Recently, there has been a focus on removing this variability by normalising the images during preprocessing, for example by filtering the image to reduce the effects of illumination variation [1]. In contrast, this article focuses on robust techniques for normalisation in two stages that have so far received less attention, that is, normalisation of the features and normalisation of the output scores.

For face authentication, this work uses the parts-based approach proposed in [2], whereby the distribution of features extracted from images of a person's (subject's) face is described by a Gaussian mixture model (GMM) . In [3], [4], this approach was found to offer the best trade-off in terms of complexity, robustness and discrimination. Notably, a GMM modelling framework similar to that described above for face authentication has also been used with much success for speaker authentication [5], [6]. One of the reasons for the success of this framework in speaker authentication is the use of effective feature and score normalisation techniques, which have been shown to improve performance substantially [7]. The goal of this paper is to apply these well-established techniques to face authentication, and evaluate the results on a range of face authentication databases that are relevant to forensics and security, including surveillance and access control scenarios.

For score normalisation, we evaluate the techniques of T-norm, Z-norm [8] and ZT-norm [9]. These techniques aim to scale output scores to a global distribution in a subject-centric (Z-norm) or probe-centric (T-norm) manner in order to facilitate the application of a global score threshold across varying enrolment and probe image conditions. While Z-norm has received some limited attention in the field of face authentication [10]–[13], this paper is the first to present a comprehensive comparison of T-norm, Z-norm and ZT-norm for face authentication, on a variety of challenging databases.

For feature normalisation, we investigate mean and variance normalisation (MVN) [14] and the more advanced technique of feature warping (FW) [15]. While MVN assumes that the effects to be normalised are stationary throughout a given image, FW utilises local information under the assumption that these effects vary within the image. This is achieved by transforming each value to a normalised value representative of its rank with respect to neighbouring feature vectors. Feature warping was originally applied to speaker authentication [15], where normalisation was performed on a sequence of feature vectors in time. In this work, we develop a novel generalised feature warping algorithm that can be applied to a two-dimensional lattice of feature vectors for face authentication.

The application of feature and score normalisation results in substantial reductions in face authentication error rate when evaluated on the challenging and publicly-available BANCA, SCface and MOBIO databases, with complementary improvements from both feature and score normalisation techniques.

Section II first provides a background of GMM-based face authentication. Section III describes the score normalisation techniques pursued here, while Section IV focuses on feature normalisation including a description of the proposed method of applying feature warping to 2D face images. In Sections V and VI, experimental results are reported, followed by conclusions in Section VII.

Roy Wallace (roy.wallace@idiap.ch), Christopher McCool (christopher.mccool@idiap.ch) and Sébastien Marcel (sebastien.marcel@idiap.ch) are with Idiap Research Institute, Martigny, Switzerland. Mitchell McLaren is with Radboud University Nijmegen, The Netherlands (m.mclaren@let.ru.nl).

Manuscript received XXX XXX, 2011; revised XXX XXX, XXX.

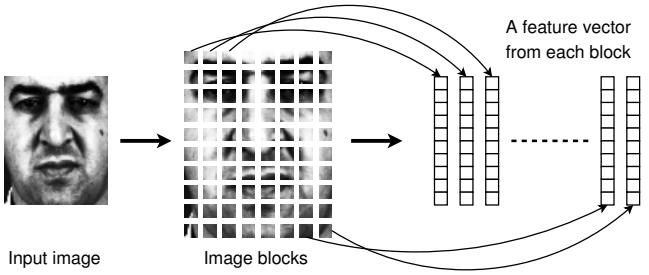


Fig. 1. This figure presents the concept of dividing the face into blocks and obtaining feature vectors from each block (a parts-based topology).

II. Gaussian mixture modelling for face authentication

The parts-based topology using Gaussian mixture modelling (GMM) was first applied to face authentication in [2] and has since been used by several researchers [4], [16], [17]. As shown in Figure 1, the method decomposes the face into an overlapping set of blocks, each of which is then considered to be a separate observation of the same signal (the face). Since features are extracted from each part of the face independently, the approach is naturally robust to occlusion, local transformation and face mis-localisation, and has been found to offer the best trade-off in terms of complexity, robustness and discrimination [3], [4]. This approach relies on estimating the distribution of features using a GMM for each subject, then performing authentication by calculating a likelihood ratio between the subject model and a universal background model (UBM). Another strength of this GMM-based framework is that the resulting likelihood ratio can theoretically be used in forensics cases where, in a Bayesian interpretation, such a measure represents the strength of forensic evidence. The rest of this section describes the main processing stages of the framework, including image registration, pre-processing, feature extraction and classification.

A. Image Registration and Pre-processing

The image is converted to grayscale, cropped and registered using manually or automatically-localised eye positions. In this paper, experiments use manually-annotated eye positions. The resulting image is 64×80 pixels with a distance between the eyes of 33 pixels, where the two eyes are aligned on the horizontal axis and the center of each eye is located 16 pixels down and 16 pixels in from the border of the image. Each cropped image is then processed using Tan & Triggs normalisation [1], which consists of gamma correction, difference of Gaussian (DoG) filtering then contrast equalisation. In section VI, images were not pre-processed using Tan & Triggs normalisation for experiments on the SCface database, as it did not improve performance in that case only.

B. Feature Extraction

Feature extraction consists of segmenting a pre-processed image into a set of overlapping blocks and extracting a feature vector of 2D-Discrete Cosine Transform (2D-DCT) coefficients from each block. Blocks are exhaustively sampled from the image, that is, sampled from the image with a step size of 1 pixel. This was consistently found to provide the best performance compared to sampling blocks with a lesser degree of overlap. The pixel values in each block are then mean and variance-normalised. From each of the K blocks in an image, we retain only the subset of D 2D-DCT coefficients that correspond to the low frequency range, since they are less susceptible to noise. The subset of low frequency coefficients is selected using the zig-zag pattern described in [18]. Each image is thus represented by a set of K feature vectors, $\boldsymbol{O} = \left\{\boldsymbol{o}^1, \boldsymbol{o}^2, \ldots, \boldsymbol{o}^K\right\}$.

C. GMM Classifier

The classifier works by modelling the distribution of feature vectors for a subject with a GMM, estimated using background model adaptation [4], [5], [16]. Background model adaptation utilises a universal background model (UBM), $\boldsymbol{m}$, as a prior for deriving subject models using maximum *a posteriori* (MAP) adaptation [5]. The UBM is trained using maximum likelihood training from face images of a large number of individuals [19]. The i^{th} subject model $\boldsymbol{s}_i$ is then formed by adapting the UBM to better match the observations of the subject's enrolment data. In this work we only adapt the means of the GMM components and use diagonal covariance matrices, as this requires fewer observations to perform adaptation [5] and has already been shown to be effective for face authentication [4], [16].

Once a subject model is trained, a probe image, $\boldsymbol{O}_t$, can be authenticated against the model by calculating a log-likelihood ratio (LLR),

$$h\left(\boldsymbol{O}_t, \boldsymbol{s}_i\right) = \log\left(\prod_{k=1}^{K} \frac{p(\boldsymbol{o}_t^k \mid \boldsymbol{s}_i)}{p(\boldsymbol{o}_t^k \mid \boldsymbol{m})}\right) \tag{1}$$

$$= \sum_{k=1}^{K} \log(p(\boldsymbol{o}_t^k \mid \boldsymbol{s}_i)) - \log(p(\boldsymbol{o}_t^k \mid \boldsymbol{m})). \tag{2}$$

By applying a threshold value, τ, the LLR (or score) can then be used in a decision rule where $\boldsymbol{O}_t$ is said to match to subject model $\boldsymbol{s}_i$ if and only if $h\left(\boldsymbol{O}_t, \boldsymbol{s}_i\right) \geq \tau$. Recently, [20] proposed an alternative classification scheme for pairs of images using an L_1 distance between histograms of zeroth order UBM statistics. In this paper, we choose to use standard LLR classification and focus on normalisation techniques within this framework.

In recent years, a simplified approximation of (2) termed *linear scoring* has been widely adopted in the speaker authentication literature and is also adopted in this work. For a full explanation readers are referred to [21]. Briefly, linear scoring uses the following first order approximation to (2):

$$h_{\text{linear}}\left(\boldsymbol{O}_t, \boldsymbol{s}_i\right) = \bar{\boldsymbol{s}}_i^{\top} \boldsymbol{\Sigma}^{-1} \bar{\boldsymbol{F}} \tag{3}$$

$$\bar{\boldsymbol{s}}_i = \boldsymbol{s}_i - \boldsymbol{m} \tag{4}$$

$$\bar{\boldsymbol{F}} = \boldsymbol{F} - \boldsymbol{N}\boldsymbol{m}. \tag{5}$$

In this notation, $\boldsymbol{s}_i$ and $\boldsymbol{m}$ are $CD \times 1$ *supervectors*, formed

by concatenating the C GMM component means. Similarly,

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & & \\ & \ddots & \\ & & \boldsymbol{\Sigma}_C \end{bmatrix}, \quad \boldsymbol{N} = \begin{bmatrix} N_1\boldsymbol{I} & & \\ & \ddots & \\ & & N_C\boldsymbol{I} \end{bmatrix} \tag{6}$$

where $\boldsymbol{\Sigma}_c$ is the covariance matrix of the c'th UBM component and $\boldsymbol{I}$ is the $D \times D$ identity matrix (D is the feature dimensionality). Finally, N_c and $\boldsymbol{F_c}$ are the zeroth and first order UBM statistics of the probe image,

$$N_c = \sum_{k=1}^{K} P\left(c|\boldsymbol{o}^k\right), \qquad \boldsymbol{F_c} = \sum_{k=1}^{K} \boldsymbol{o}^k P\left(c|\boldsymbol{o}^k\right) \tag{7}$$

and $\boldsymbol{F}$ is the $CD \times 1$ supervector obtained by concatenating $\boldsymbol{F_c}$ for $c = 1, \ldots, C$. This approximation leads to much faster scoring and no degradation in face authentication accuracy, as found in preliminary experiments.

III. Score normalisation

Score normalisation has long been an integral part of speaker authentication technology [6]. Score normalisation aims to counteract statistical variations in output scores due to changes in the conditions across different enrolment and probe samples. This is achieved by scaling distributions of system output scores to better facilitate the application of a single, global threshold for authentication.

The most widely adopted methods in speaker authentication literature are zero-normalisation (Z-norm) and probe- or test-normalisation (T-norm) [8]. Both techniques use mean and variance normalisation,

$$\overline{h} = \frac{h - \mu}{\sigma}, \tag{8}$$

where h and $\overline{h}$ are the raw and normalised scores, respectively, while the scaling parameters μ and σ are the mean and standard deviation of an impostor score distribution (assumed to be Gaussian) estimated on a held-out or *cohort* data set. It is the manner in which these scaling parameters are estimated that distinguishes Z-norm from T-norm.

Z-norm operates in a subject-centric manner such that μ_i and σ_i are determined from the impostor score distribution found by comparing all images in the cohort data set to the subject model $\boldsymbol{s}_i$. The Z-norm parameters can thus be pre-computed during the enrolment stage, which ensures that Z-norm adds negligible computational load to scoring. T-norm, on the other hand, works in a probe-centric manner. Firstly, impostor models are trained for each subject in the cohort data set, in the same way as the subject models. The probe image being authenticated is then compared to each of these models to generate an impostor score distribution. From this score distribution, μ_t and σ_t are derived and used to normalise the scores from probe image $\boldsymbol{O}_t$ via (8). T-norm thus introduces extra computation during scoring, as the probe image needs to be compared to each cohort model in order to estimate the T-norm parameters. This computational cost is somewhat ameliorated by using fast linear scoring (Section II-C), and can be controlled by tuning the size of the cohort set to meet the speed requirements of the application. Common practice in speaker authentication is to employ ZT-norm, in which Z-norm is applied prior to T-norm [6].

Score normalisation for face authentication has received limited attention in the literature. In [10], Z-norm was applied to face authentication, however, T-norm was not considered nor was the use of a GMM parts-based framework. The work of [12], [13] applied Z-norm for images that were captured either in one of several discrete conditions, or with an estimated quality. In [22], [23], a form of probe-centric score normalisation was applied whereby the development/test sets were used directly as the cohort data set for score normalisation. In this way, score normalisation was essentially applied in a closed-set authentication task, since scores were normalised with respect to the scores from the complete set of potential impostors. In contrast to [22], [23], this work aims to simulate performance in real world applications where anyone can attempt to access the system, that is, the considerably more difficult open-set authentication task. In practice, this requires that the subjects in the cohort data set are disjoint from those in the development/test sets.[1] This work is the first to evaluate T-norm in this context. We present an analysis of face authentication using Z- and T-, and ZT-norm in Section VI.

IV. Feature Normalisation

The previous section detailed the normalisation of classification scores. In this section, we instead focus on normalisation of the features prior to modelling. Two of the most successful approaches to feature normalisation in the field of speaker authentication are mean and variance normalisation (MVN) [14] and the more advanced technique of feature warping (FW) [15]. Both techniques aim to remove within-class variation that is realised as differences in the distributions of feature vectors from the same subject.

A. Mean and Variance Normalisation

Mean and variance normalisation (MVN) [14] is a straightforward technique in which feature values from a particular image are normalised to have zero mean and unit variance. This is applied independently to each feature dimension.

MVN was originally developed to remove stationary channel offsets from features extracted from a speech recording (or utterance), in an attempt to make features acquired over different channels more comparable. In terms of 2D-DCT features extracted from a facial image, this can viewed as a means of normalising for noise introduced to an image in different frequency bands, which may occur due to different image qualities, artifacts, noise, or lighting variations between enrolment and probe images, for example.

B. Feature Warping

Feature warping [15] performs local normalisation of feature vectors using a sliding-window. In speaker authentication,

[1] In an operational system, it may or may not be trivial to ensure that the set of cohort subjects is disjoint from the set of real users, however, this is only an issue for users in the intersection of the sets, which even in the worst case should be of negligible size relative to the number of users.

feature warping applies the sliding window in the time dimension, in order to remove time-varying noise from speech features. That is, in contrast to MVN, FW was developed to additionally counteract slow-varying channel or acoustic conditions in the speech signal. To apply the idea to face authentication, we propose to apply the sliding window in the spatial domain, to compensate for noise that varies across the face region. It is hypothesised that FW will assist face authentication performance particularly when light and other noise sources do not fall across the face in a uniform manner.

Feature warping-like techniques have been applied to some problems in image processing, however, to the best of our knowledge feature warping has thus far never been applied to face authentication. In [24], Struc *et al* apply a similar warping function for palmprint recognition. However, the function is applied across the dimensions of each feature vector, rather than within each dimension across the feature vectors within a neighbouring region. In [25], Struc *et al* propose global feature warping but apply it directly to pixel values rather than feature vectors, again for palmprint recognition. The same paper proposes the gaussianisation of image patches, which is similar to sliding window-based feature warping but again it is applied directly to pixel values and additionally requires a cumbersome weighting window to smooth discontinuities. In [26], Struc *et al* propose a histogram remapping technique for face recognition, however, this differs from feature warping as it is applied at the preprocessing stage to raw pixel intensities and does not perform any local normalisation, for example with a sliding window.

In the remainder of this section, we first describe the feature warping technique in general, following the description of Pelecanos and Sridharan in [15], before adapting the technique to the task of face authentication.

Feature warping is applied independently to each dimension of a sequence of feature vectors. Values in a given dimension are warped in the context of a neighbouring region or window. Specifically, a sliding window is exhaustively applied across values such that the value central to each instance of the window is transformed (or *warped*) to a new value. The central point is warped to a value m according to its rank R within the window of N values (in descending order) according to the following equation (Equation 5 of [15]),

$$\frac{N+\frac{1}{2}-R}{N} = \int_{-\infty}^{m} h(z)dz. \tag{9}$$

In the specific case of warping the features to match a standard normal distribution, as in [15],

$$h(z) = \frac{1}{\sqrt{2\pi}} exp\left(-\frac{z^2}{2}\right). \tag{10}$$

We would like to solve for the output value m, as a function of the rank of the input value, R. From (9) and (10),

$$\frac{N+\frac{1}{2}-R}{N} = \frac{1}{2}\operatorname{erf}\left(\sqrt{\frac{1}{2}}m\right) + \frac{1}{2} \tag{11}$$

$$\therefore m\left(R\right) = \sqrt{2}\operatorname{erf}^{-1}\left(\frac{N+1-2R}{N}\right). \tag{12}$$

Given that the size of the window is fixed, $m\left(R\right)$ can be pre-calculated for all of the possible ranks $R = 1, 2, \ldots, N$. This helps reduce the computational load of feature warping. For each feature value one need simply determine the rank, R, with respect to the surrounding values and look up the output value $m\left(R\right)$.

Although initially developed for speech features that vary in a single dimension (time), FW can readily be applied to features extracted from a facial image by retaining the two-dimensional contextual relationship between feature vectors. Figure 2 presents an overview of the procedure. We first reconstruct a 2D grid of values corresponding to a particular DCT coefficient, according to the position of the block of pixels from which each feature vector was extracted. A square sliding window of $N = M \times M$ points, with M an odd integer, is then used to apply the feature warping function to the value in the centre of the window, as described previously. It should be noted that an edge of $\left(M-1\right)/2$ feature vectors is dropped from the grid of values, since a complete window of neighbouring values does not exist near the edges. Hence, feature warping slightly reduces the number of feature vectors for each image. We also propose an alternative formulation, referred to as *global* feature warping. In this case, each point is warped according to its rank R with respect to the K values across the entire image, rather than its rank within a window of N neighbouring values. Both proposed techniques are evaluated in Section VI.

V. Databases and experimental protocols

For evaluation of the proposed techniques, we chose to restrict ourselves to publicly-available databases with separate training, development and test sets to allow for unbiased evaluation. Unfortunately, some popular databases such as FRGC [27] and LFW [28] were thus not applicable, as they do not include separate development and test sets[2]. We therefore chose to evaluate the proposed normalisation techniques on the challenging BANCA, SCface and MOBIO databases. These databases were selected due to their challenging conditions, relevance to forensics and security applications, and well-defined protocols that include training, development and test sets in which subjects are disjoint.

Performance is reported in terms of equal error rate (EER) on the development set for which a decision threshold is found. This threshold is then applied to scores from the unseen test set to obtain a half total error rate (HTER), so as to measure the expected performance in a real world situation. For evaluating the statistical significance of improvements in HTER, we use the methodology proposed by equation (15) and Figure 2 of [29], with a one-tailed test. In this work we use $C = 512$ GMM components. In preliminary experiments, using more components led to only marginal gains at a prohibitive computational cost. Tuning of other system hyper-parameters (block size, DCT coefficients, feature

[2]In the FRGC database, 153 subjects occur in both the training set as well as the test set, and there is no publicly-available development set. In the LFW database, 758 image pairs in the training/development set (*View 1*) are exactly repeated in the test set (*View 2*).

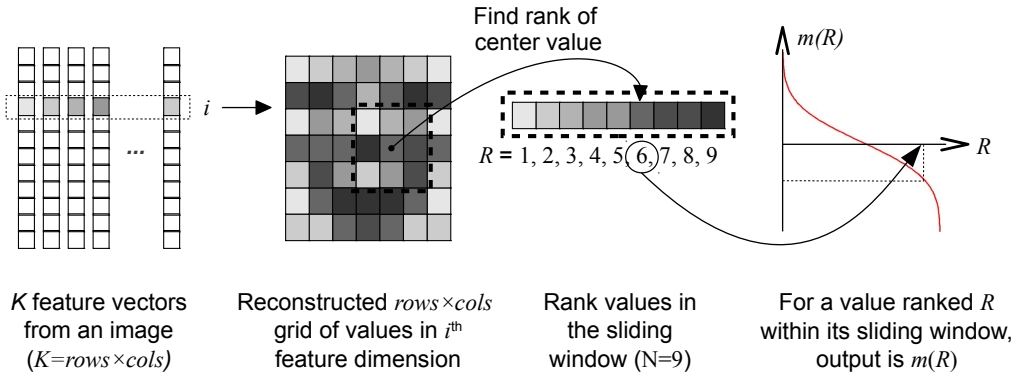


Fig. 2. 2D feature warping for face authentication

warping window size) was consistently based on minimising the EER on the development set only.

A. BANCA English

We report results on the Pooled (P) protocol of the BANCA (English) database [30]. According to this protocol, subjects were enrolled using 5 images acquired under controlled lighting conditions and probe images were taken from *controlled*, *degraded* and *adverse* lighting conditions. Figure 3a provides an example of the variation that exists between images of the same subject. The *g1* and *g2* groups of subjects (26 each) were used as the development and test sets, respectively, providing in each case a total of 2,730 scores (1,170 target trials, 1,560 impostor trials). The UBM was trained from 200 images of 20 subjects in the separate *world data* set. As score normalisation was empirically found to be more effective when using a set of subjects disjoint from those in the UBM training set, *g1* was used to normalise the scores for *g2*, and vice-versa. For T-norm, the enrolment data of the appropriate set was used to create the set of cohort models, while for Z-norm, the cohort set was formed from the appropriate set of probe images.

B. SCface

The Surveillance Cameras face database (SCface) [31] was acquired using commercially available surveillance equipment, in a range of challenging and realistic conditions. A face authentication protocol for SCface, based on the *DayTime tests* scenario [31], has recently been proposed [32] and made available online[3]. According to this protocol, facial images taken by five surveillance cameras at three specified distances (*close*, *medium*, *far*) are compared to a model trained using a single high-resolution mugshot image (See Figure 3b).

The world and test sets include 43 subjects while 44 subjects are included in the development set. There are 15 surveillance images from each subject to use as probe images, which for the test set results in 645 target trials and 27,090 impostor trials. Results are reported for a *combined* protocol, in which each probe image was assumed to originate from an unknown camera at an unknown distance. Two-thirds of the world data was used for UBM training (29 subjects), while the other third (14 subjects) was used for score normalisation. To match the enrolment and probe procedure, the T-norm cohort models were enrolled using the mugshot images, while the Z-norm cohort was formed from the surveillance probe images.

During pre-processing, low resolution images were upsampled where necessary. For experiments on SCface only, images were not pre-processed using Tan & Triggs normalisation, as it did not improve performance in this case.

C. MOBIO

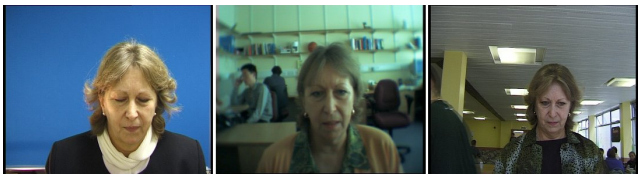
The MOBIO database contains videos of 150 participants captured in challenging real-world conditions using mobile phone cameras over a one and a half year period [33] (see Figure 3c for example images). The MOBIO protocol is supplied with the database[4] and defines three non-overlapping partitions: world (training), development and testing. The development and testing partitions are defined in a gender-dependent manner, such that subjects' models are only probed by images from subjects of the same gender. We chose to use the training data in a gender-independent manner to be consistent with the other databases, though future work could investigate gender-dependent training.

The distributors of the MOBIO database recently released a still-image protocol [32] which includes one image extracted from each video with manually annotated eye locations. Subjects were enrolled using 5 images each, as defined in the protocol, with a total of 42 and 58 subjects in the development and test sets, respectively. Across the male and female protocols, the development sets contain 4410 target trials and 90090 impostor trials while the test sets contain 6090 target trials and 187530 impostor trials. We use *MOBIO.mal* and *MOBIO.fem* to refer to the male and female protocols respectively.

From the full training data set containing 9,579 images of 50 subjects, a subset of 1,224 images of 34 subjects (36 images each) was used for UBM training, while the remaining 16 subjects were used for score normalisation. For Z-norm, we use all of the images from the training set of the 16 cohort subjects. For T-norm, in order to best replicate the enrolment conditions, we enrol cohort models from the first 5 images of each session for each cohort subject.

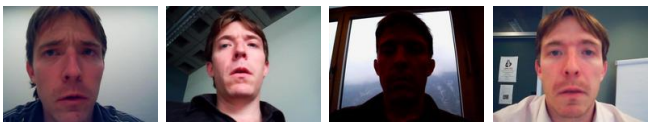
[3]http://scface.org/

[4]http://www.idiap.ch/dataset/mobio



(a) BANCA database: controlled, degraded, and adverse scenarios.

(b) SCface database: enrolment/mugshot image; probe images from close, medium and far distances.



(c) MOBIO database.

Fig. 3. Example images showing a wide range of within-subject variation.

B	D	BANCA		SCface		MOBIO.mal		MOBIO.fem	
		dev	**test**	**dev**	**test**	**dev**	**test**	**dev**	**test**
4	16	28.8	22.6	45.8	46.4	18.8	17.1	15.2	26.7
8	28	15.6	14.8	35.5	36.2	10.7	**11.9**	**9.8**	19.6
12	45	**11.9**	**12.8**	32.6	30.0	**9.1**	13.0	11.6	**18.1**
16	66	13.9	13.4	32.6	31.7	9.4	14.7	14.0	21.3
20	66	13.9	13.5	**30.3**	**29.8**	10.2	16.8	17.1	24.3
24	91	15.5	14.5	32.3	31.5	10.6	17.6	18.1	25.6

TABLE I
THE EFFECT OF BLOCK SIZE, $B \times B$, AND NUMBER OF 2D-DCT COEFFICIENTS, D (% EER ON DEV SET, % HTER ON TEST SET).

VI. RESULTS

A. Baseline system tuning

The system was first tuned to optimise the block size, $B \times B$, and number of 2D-DCT coefficients, D, retained during feature extraction. Table I shows that $B = 12$ provided the best results on BANCA. Considering MOBIO male and females jointly, the best results were obtained using a block size of $B = 8$ or $B = 12$. A larger block size of $B = 20$ was optimal on the SCface database, which may be due to the low resolution of the images. The optimised block sizes were also found to generalise quite well to the test sets. The remainder of this study uses $B = 12$ for BANCA and MOBIO and $B = 20$ for SCface experiments.

B. Feature Normalisation

The feature normalisation techniques of MVN and feature warping were evaluated on each database. Figure 4 illustrates the effect of changing the feature warping window size, in terms of performance on the development set of each database. Interestingly, a small window of 9×9 was optimal for BANCA evaluations, 7×7 for MOBIO, while a larger window of 15×15 was best for the lower resolution images in SCface.

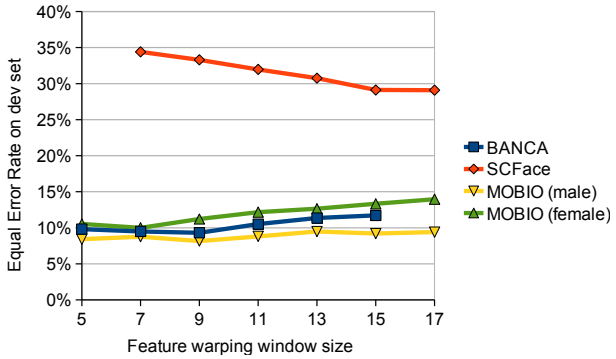


Fig. 4. Effect of feature warping sliding window size (% EER on dev set).

Table II compares MVN, sliding window-based feature warping (with tuned window size), and global FW to the baseline configuration. Several interesting trends can be observed in the table. While no single technique provided the best performance across all databases, feature warping consistently reduced the error rate when compared to the baseline. Furthermore, for BANCA and MOBIO, there was a consistent improvement from sliding window-based FW when compared to global FW. However, the same was not the case for SCface, which has much lower resolution images than the other databases. This may have been due to the relatively low amount of information contained within each region of a lower resolution image. We explored this hypothesis by artificially reducing the resolution of the BANCA images as follows. After image registration, eye centres are placed 33 pixels apart (see Section II-A). In the SCface database, the average distance for probe images before registration is 19.6 pixels. Therefore, to simulate images of a similar original resolution, cropped BANCA images were downsampled by a factor of 1.7, then upsampled by the same factor, except for enrolment images, which were kept at full resolution as in SCface. Then, we repeated the feature warping experiments on this artificially downsampled version of the BANCA database. As shown in Figure 5, the evidence supports the hypothesis that a larger window should be used for lower resolution images, and that the gain from sliding window-based FW when compared to global FW is indeed lessened in this case. For BANCA, using a window rather than global FW provides a 16% relative reduction in EER for the full resolution images, versus 5% for the downsampled images.

Finally, using global feature warping was not found to significantly improve performance compared to the simpler technique of (global) mean and variance normalisation. This suggests that the rank-based distribution mapping function of feature warping, i.e. (12) was not critical for feature normalisation in this case, and that normalising by the mean and variance in each feature dimension was sufficient.

Overall, substantial relative improvements were achieved by applying the best feature normalisation procedure for each database, with a 16% improvement on the SCface test set using MVN, and 30% on BANCA, 15% for MOBIO (male) and 9% for MOBIO (female) using the proposed feature warping technique. Each improvement is statistically significant at a level of at least 99%.

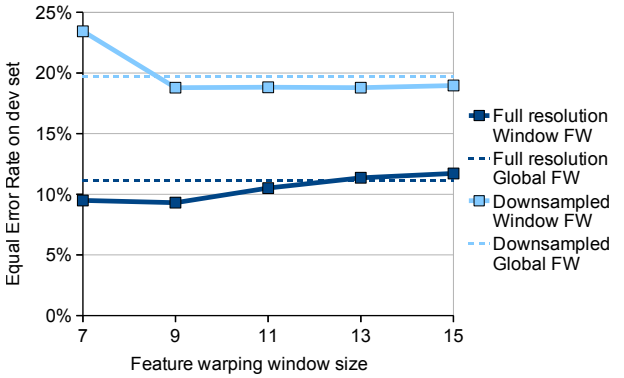


Fig. 5. The effect of feature warping sliding window size for the BANCA database, using either the standard full resolution images or a set of probe images that have been artificially reduced in resolution (% EER on dev set).

System	BANCA dev	BANCA test	SCface dev	SCface test	MOBIO.mal dev	MOBIO.mal test	MOBIO.fem dev	MOBIO.fem test
Baseline	11.9	12.8	30.3	29.8	9.1	13.0	11.6	18.1
MVN	11.0	11.1	**23.9**	**25.1**	8.9	11.9	10.3	18.2
FW (window)	**9.3**	**8.9**	29.1	28.8	8.8	**11.0**	**10.0**	**16.5**
FW (global)	11.1	10.8	25.2	25.7	**8.7**	11.8	10.9	18.4

TABLE II
COMPARISON OF FEATURE NORMALISATION TECHNIQUES (% EER ON DEV SET, % HTER ON TEST SET).

C. Score Normalisation

Table III details results on the evaluated databases when applying Z-, T- and ZT-norm to the baseline configuration. First we consider the results on the BANCA and SCface databases, followed by MOBIO results.

Results on BANCA and SCface: In these cases, Z-norm provided little benefit over the baseline configuration, while T-norm offered relative reductions in HTER of between 25% and 39%, which are statistically significant at a level exceeding 99.99%. This suggests that variations in score distributions are mostly due to variation between probe images, rather than between subject models. This is reasonable, especially considering the drastic differences between probe images taken in controlled, degraded and adverse scenarios in the BANCA database, and between probe images from different surveillance cameras in the SCface database. In contrast, all subject models are enrolled with images from relatively consistent conditions (i.e. the controlled scenario for BANCA, and a mugshot image for SCface). The use of Z-norm in combination with T-norm (ZT-norm) did, however, provide some further improvements in BANCA trials.

It was hypothesised that the reason for the limited gain observed from Z-norm may be due to the large variation in conditions between images in the development, test and cohort data sets. To test this hypothesis, an analysis was performed to demonstrate the effect of selecting a Z-norm cohort that better matches the probe image. For this purpose, more refined Z-norm cohort data sets were created by selecting subsets of images taken in similar conditions. For BANCA, the subsets are the pre-defined *controlled*, *adverse*, and *degraded* conditions. For SCface, the subsets correspond to the three pre-defined camera distances *close*, *medium*, and *far*. To perform Z-normalisation for each trial, the subset of the Z-norm cohort is used that matches the corresponding conditions of the probe image. The results of this approach, referred to as condition-specific normalisation, are listed in Table IV. On the BANCA database, selecting a Z-norm cohort to better match the specific probe image conditions provided considerable improvements, while this was not observed for SCface.

System	BANCA dev	BANCA test	SCface dev	SCface test	MOBIO.mal dev	MOBIO.mal test	MOBIO.fem dev	MOBIO.fem test
Baseline	11.9	12.8	30.3	29.8	**9.1**	13.0	11.6	18.1
Z-norm	12.6	11.8	30.8	29.6	**9.1**	12.1	**10.5**	18.6
T-norm	8.8	7.8	**22.7**	**22.0**	9.8	11.7	12.7	**16.6**
ZT-norm	**8.3**	**7.0**	23.3	22.7	10.1	**11.2**	10.7	19.0

TABLE III
COMPARISON OF SCORE NORMALISATION TECHNIQUES (% EER ON DEV SET, % HTER ON TEST SET).

System	BANCA dev	BANCA test	SCface dev	SCface test
Baseline	11.9	12.8	30.3	29.8
Z-norm	12.6	11.8	30.8	29.6
ZT-norm	8.3	7.0	23.3	**22.7**
Z-norm (condition-specific)	7.5	7.2	30.0	30.0
ZT-norm (condition-specific)	**7.4**	**5.8**	**23.2**	**22.7**

TABLE IV
RESULTS OF USING A CONDITION-SPECIFIC COHORT SUBSET FOR Z-NORM (% EER ON DEV SET, % HTER ON TEST SET).

Results on MOBIO: In this case, results showed limited gains from Z-norm, consistent with the results on the other databases. This is also consistent with the findings of [10], which applied Z-norm to the FERET and CAS-PEAL databases with limited gains (particularly on CAS-PEAL, when using a Fisherface classifier rather than the GMM classifier used in this work). With respect to T-norm, gains were observed on the MOBIO test set but not on the development set, for both male and female subjects.

Analysis of T-norm results: The key aspect of T-norm is the estimation of the impostor score distribution for each probe image, from which the scaling parameters μ and σ are applied to normalise the scores. If this estimated impostor score distribution is not well-matched to the actual impostor score distribution, it stands to reason that T-norm will be less effective. Therefore, an analysis was performed to evaluate the accuracy of impostor score distribution estimation for each data set. The metric chosen to represent score distribution mismatch is the per-image mean Kullback-Leibler divergence,

$$\mathrm{MKL} = \frac{1}{N}\sum_{j=1}^{N} \mathrm{KL}\left(p_j || q_j\right) \tag{13}$$

where N is the number of images, p_j is the score distribution for the image estimated using the models of actual impostors (i.e. subjects in the same data set as the image), while q_j is the score distribution estimated from the T-norm cohort models. As can be seen from Table V, in the cases where T-norm sub-

	BANCA		SCface		MOBIO.mal		MOBIO.fem	
	dev	**test**	**dev**	**test**	**dev**	**test**	**dev**	**test**
Error rate change (%)	-26	-39	-25	-26	8	-10	9	-8
Mean KL-divergence	0.09	0.11	0.18	0.17	0.37	0.25	1.40	0.52

TABLE V
CHANGE IN ERROR RATE CAUSED BY T-NORM, COMPARED TO MEAN KL-DIVERGENCE BETWEEN ESTIMATED AND ACTUAL IMPOSTOR SCORE DISTRIBUTIONS.

stantially reduced the error rate (e.g. on the BANCA and SCface databases), the estimated impostor score distributions were particularly well-matched to the scores observed for the real impostors (i.e. the mean KL-divergence was low). In contrast, on the MOBIO database, the mismatch was consistently worse on the development set, compared to the test set (particularly for female subjects), which corresponds to the increase in error rate observed on the development sets. This evidence supports the original hypothesis that the T-norm technique is less effective when there is a mismatch between the estimated impostor score distribution and the actual impostor score distribution. While this is a difficult problem to solve, it is reasonable to expect that increasing the size of the cohort should improve the estimation of score distributions. In comparison, for speaker authentication it has been shown that performance improves with an increase in T-norm cohort size up to 50 speaker models [8] and in some cases hundreds or thousands of models have been used [34].

Further analysis was performed to investigate the effect of T-norm cohort size. In Figure 6, the plots on the left-hand side show the effect of reducing the T-norm cohort set size on the mean KL-divergence. There is clear evidence that the use of a larger cohort improves the estimation of the impostor score distributions, with a lower divergence observed for larger cohort sets. The plots on the right-hand side show the corresponding error rate achieved after applying T-norm. For the databases tested in this work, considerable gains are achieved even with small cohort of 10–20 people, with some evidence that larger cohorts may further improve accuracy. These results suggest that in future work, in addition to increasing the size of the cohort, intelligently selecting cohort members may prove important, as has been investigated previously for speaker authentication [34].

D. Combining Normalisation Techniques

Table VI presents the results of applying score normalisation in conjunction with each of the feature normalisation techniques described in Section IV. It can be observed that score normalisation is highly complementary to each feature normalisation technique. As described earlier, for feature normalisation MVN provided the best performance on SCface while BANCA and MOBIO results were further improved by sliding window-based FW. Table VI shows that when T-norm was applied in addition to this feature normalisation, further relative improvements of 45%, 37%, 7% and 9% were achieved on the BANCA, SCface, MOBIO (male) and MOBIO (female) test sets, respectively. This combination of feature and score normalisation techniques thus resulted in overall

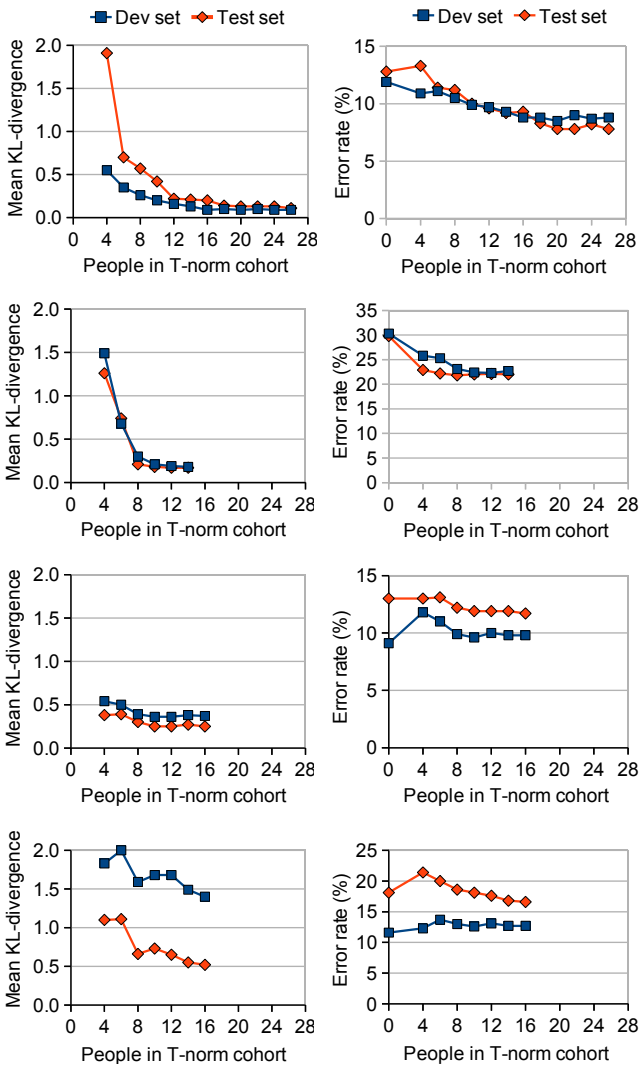


Fig. 6. Results for (top to bottom:) BANCA, SCface, MOBIO (male) and MOBIO (female) databases, showing the effect of reducing the T-norm cohort set size on (left-hand side:) the mean KL-divergence between estimated and actual impostor score distributions and (right-hand side:) the corresponding error rate (% EER on dev set, % HTER on test set) achieved after T-norm is applied.

improvements of 62%, 47%, 22% and 17% relative to the respective baseline results. Each improvement is statistically significant at a level exceeding 99.99%.

Table VII compares our results on the BANCA database to recently published work. For comparison, we report the HTER on the test set (*g2*), development set (*g1*)[5], and the average. Note that while Rua *et al* [13] used automatic face extraction from the BANCA videos, the other studies used the pre-selected images from each video as in this work. While the primary goal of this article was to provide an analysis of the effects of score and feature normalisation, Table VII further shows that our results are competitive with the state-of-the-art.

VII. CONCLUSIONS

This paper presented a novel study of score normalisation and feature normalisation for GMM-based face authentica-

[5]This is obtained by applying the EER threshold from *g2*.

System	BANCA dev	BANCA test	SCface dev	SCface test	MOBIO.mal dev	MOBIO.mal test	MOBIO.fem dev	MOBIO.fem test
Baseline	11.9	12.8	30.3	29.8	**9.1**	13.0	11.6	18.1
+ Z-norm	12.6	11.8	30.8	29.6	**9.1**	12.1	**10.5**	18.6
+ T-norm	8.8	7.8	**22.7**	**22.0**	9.8	11.7	12.7	**16.6**
+ ZT-norm	**8.3**	**7.0**	23.3	22.7	10.1	**11.2**	10.7	19.0
MVN	11.0	11.1	23.9	25.1	8.9	11.9	10.3	18.2
+ Z-norm	12.2	10.6	25.0	25.7	**8.6**	11.2	**9.3**	18.6
+ T-norm	8.0	6.2	**16.7**	**15.7**	9.3	10.9	12.2	**16.6**
+ ZT-norm	**7.8**	**6.1**	**16.7**	16.4	9.2	**10.5**	10.7	20.4
FW (window)	9.3	8.9	29.1	28.8	**8.8**	11.0	10.0	16.5
+ Z-norm	9.2	8.9	29.6	29.6	9.1	10.8	**9.2**	17.2
+ T-norm	6.9	**4.9**	**23.2**	**22.3**	9.4	**10.2**	11.8	**15.0**
+ ZT-norm	**6.3**	5.5	**23.2**	22.5	9.7	10.4	10.4	17.1
FW (global)	11.1	10.8	25.2	25.7	8.7	11.8	10.9	18.4
+ Z-norm	11.5	10.6	25.2	26.2	**8.5**	11.3	**9.6**	18.7
+ T-norm	8.5	**6.1**	**17.0**	16.5	8.9	10.9	12.3	**16.5**
+ ZT-norm	**7.5**	6.3	17.3	**16.0**	9.0	**10.5**	10.7	20.8

TABLE VI
SCORE NORMALISATION IN COMBINATION WITH FEATURE NORMALISATION (% EER ON DEV SET, % HTER ON TEST SET).

System	Dev	Test	Average
Rúa *et al* [13]	10.6	9.8	10.2
Ahonen *et al* [35]	-	-	9.1
Chan *et al* [23]	-	-	5.4
FW (window) + T-norm	7.0	4.9	5.9

TABLE VII
A COMPARISON TO PREVIOUSLY PUBLISHED RESULTS (% HTER) FOR THE P PROTOCOL OF THE BANCA ENGLISH DATABASE.

tion. For score normalisation, the probe-centric method of T-norm was particularly useful, with results suggesting that T-norm is an effective way to reduce the effects of probe image variability for open-set face authentication. Analysis further showed that Z-norm can improve performance if the cohort images are well-matched to the condition of the probe images. For feature normalisation, the proposed feature warping technique consistently reduced the face authentication error rate across all databases, when compared to the baseline. Together, these experimental results demonstrate that these techniques can significantly improve authentication accuracy, and confirm that these techniques generalise well from speaker to face authentication.

In future work, we aim to apply feature warping both in the space and time dimensions to face videos. This could improve the robustness of visual features to noise that is time-varying, for example, variations in pose, illumination and expression throughout a video. It would also be interesting to see if the results in this work generalise to approaches other than GMM-based systems. The use of additional training data from a combination of databases will also be explored.

With this work, we have successfully generalised normalisation methods across two very different biometric modalities. This should provide encouragement to those pursuing the consolidation of biometric research and fostering cross-pollination of ideas across modalities.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7) under grant agreements 238803 (BBfor2) and 257289 (TABULA RASA). Portions of the research in this paper use the SCface database of facial images. Credit is hereby given to the University of Zagreb, Faculty of Electrical Engineering and Computing for providing this database of facial images. We would also like to thank Niklas Johansson for completing the laborious task of manually annotating the images that now form the MOBIO still-image database.

REFERENCES

[1] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1635–1650, 2010.

[2] C. Sanderson and K. Paliwal, "Fast features for face authentication under illumination direction changes," *Pattern Recognition Letters*, vol. 24, pp. 2409–2419, 2003.

[3] F. Cardinaux, C. Sanderson, and S. Marcel, "Comparison of MLP and GMM classifiers for face verification on XM2VTS," in *International Conference on Audio- and Video-based Biometric Person Authentication*, 2003, pp. 911–920.

[4] F. Cardinaux, C. Sanderson, and S. Bengio, "User Authentication via adapted Statistical Models of Face images," *IEEE Transactions on Signal Processing*, vol. 54, pp. 361–373, 2006.

[5] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital Signal Processing*, vol. 10, pp. 19–41, 2000.

[6] T. Kinnunen and H. Li, "An overview of text-independent speaker recognition: from features to supervectors," *Speech Communication*, vol. 52, pp. 12–40, 2010.

[7] C. Barras and J.-L. Gauvain, "Feature and score normalization for speaker verification of cellular data," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2, 2003, pp. 49–52.

[8] R. Auckenthaler, M. Carey, and H. Lloyd-Thomas, "Score normalization for text-independent speaker verification systems," *Digital Signal Processing*, vol. 10, pp. 42–54, 2000.

[9] R. Vogt, B. Baker, and S. Sridharan, "Modelling session variability in text-independent speaker verification," in *Proc. Interspeech*, 2005, pp. 3117–3120.

[10] F. Yang, S. Shan, B. Ma, X. Chen, and W. Gao, "Using score normalization to solve the score variation problem in face authentication," in *Advances in Biometric Person Authentication*, ser. Lecture Notes in Computer Science, 2005, vol. 3781, pp. 31–38.

[11] N. Poh and S. Bengio, "F-ratio client-dependent normalisation for biometric authentication tasks," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2005, pp. 721–724.

[12] N. Poh, J. Kittler, S. Marcel, D. Matrouf, and J.-F. Bonastre, "Model and score adaptation for biometric systems: Coping with device interoperability and changing acquisition conditions," in *International Conference on Pattern Recognition*, 2010, pp. 1229–1232.

[13] E. Rúa, J. Castro, and C. Mateo, "Quality-based score normalization for audiovisual person authentication," in *Image Analysis and Recognition*, ser. Lecture Notes in Computer Science, 2008, vol. 5112, pp. 1003–1012.

[14] O. Viikki and K. Laurila, "Cepstral domain segmental feature vector normalization for noise robust speech recognition," *Speech Communication*, vol. 25, pp. 133–147, 1998.

[15] J. Pelecanos and S. Sridharan, "Feature warping for robust speaker verification," in *2001: A Speaker Odyssey - The Speaker Recognition Workshop*, 2001.

[16] S. Lucey and T. Chen, "A GMM parts based face representation for improved verification through relevance adaptation," in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, 2004, pp. 855–861.

[17] C. McCool, V. Chandran, S. Sridharan, and C. Fookes, "3D face verification using a free-parts approach," *Pattern Recognition Letters*, vol. 29, pp. 1190–1196, 2008.

[18] W. B. Pennebaker and J. L. Mitchell, *JPEG still image data compression standard*. New York: Van Nostrand Reinhold, 1993.

[19] D. Reynolds and R. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models," *IEEE Transactions on Speech and Audio Processing*, vol. 3, pp. 72–83, 1995.
[20] C. Sanderson and B. Lovell, "Multi-region probabilistic histograms for robust and scalable identity inference," *Lecture Notes in Computer Science*, vol. 5558, pp. 199–208, 2009.
[21] O. Glembek, L. Burget, N. Dehak, N. Brummer, and P. Kenny, "Comparison of scoring methods used in speaker recognition with joint factor analysis," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009, pp. 4057–4060.
[22] Y. M. Lui, J. Beveridge, B. Draper, and M. Kirby, "Image-set matching using a geodesic distance and cohort normalization," in *IEEE International Conference on Automatic Face and Gesture Recognition*, 2008, pp. 1–6.
[23] C. H. Chan, J. Kittler, N. Poh, T. Ahonen, and M. Pietikainen, "(Multiscale) local phase quantisation histogram discriminant analysis with score normalisation for robust face recognition," in *IEEE International Conference on Computer Vision*, 2009, pp. 633–640.
[24] V. Struc and N. Pavesic, "A comparison of feature normalization techniques for PCA-based palmprint recognition," in *MATHMOD*, 2009, pp. 2450–2453.
[25] ——, "Gaussianization of image patches for efficient palmprint recognition," *Electrotechnical review*, vol. 76, no. 5, pp. 245–250, 2009.
[26] V. Struc, J. Zibert, and N. Pavesic, "Histogram remapping as a preprocessing step for robust face recognition," *WSEAS Transactions on Information Science and Applications*, vol. 6, no. 3, pp. 520–529, 2009.
[27] 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 *IEEE Conference on Computer Vision and Pattern Recognition*, 2005, pp. 947–954.
[28] G. B. Huang, M. Ramesh, T. Berg, , and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments." University of Massachusetts, Amherst, Tech. Rep. 07-49, October 2007.
[29] S. Bengio and J. Mariéthoz, "A statistical significance test for person authentication," in *Proceedings of Odyssey 2004: The Speaker and Language Recognition Workshop*, 2004.
[30] E. Bailly-Baillière, S. Bengio, F. Bimbot, M. Hamouz, J. Kittler, J. Mariéthoz, J. Matas, K. Messer, V. Popovici, F. Porée, B. Ruiz, and J.-P. Thiran, "The BANCA database and evaluation protocol," in *International Conference on Audio- and Video-Based Biometric Person Authentication*, 2003, pp. 625–638.
[31] M. Grgic, K. Delac, and S. Grgic, "SCface-surveillance cameras face database," *Multimedia tools and applications*, vol. 51, pp. 863–879, 2011.
[32] R. Wallace, M. McLaren, C. McCool, and S. Marcel, "Inter-session variability modelling and joint factor analysis for face authentication," in *International Joint Conference on Biometrics*, 2011.
[33] C. McCool and S. Marcel, "MOBIO database for the ICPR 2010 face and speech competition," Idiap Research Institute, Tech. Rep. Idiap-Com-02-2009, November 2009.
[34] M. McLaren, R. Vogt, B. Baker, and S. Sridharan, "Improved GMM-based speaker verification using SVM-driven impostor dataset selection," in *Proc. Interspeech*, 2009, pp. 1267–1270.
[35] T. Ahonen and M. Pietikäinen, "Pixelwise local binary pattern models of faces using kernel density estimation," in *Advances in Biometrics*, ser. Lecture Notes in Computer Science, 2009, vol. 5558, pp. 52–61.

Roy Wallace received his BEng(Hon) in 2006 and PhD in Engineering in 2010 with the Speech, Audio, Image and Video Technologies (SAIVT) group, Queensland University of Technology (QUT), Australia. He has been involved in patent applications at QUT as well as Microsoft Research Asia, Beijing, and has performed commercial evaluations of his research. He is now a postdoctoral researcher in biometrics and machine learning at Idiap Research Institute, Switzerland, with a particular interest in the use of biometrics for forensics.

Mitchell McLaren received his PhD with the Speech, Audio, Image and Video Technologies (SAIVT) at the Queensland University of Technology (QUT), Brisbane, Australia, in 2010. He received his BCompSysEng also from QUT in 2006. Mitchell has been with the Centre for Language and Speech Technology (CLST) at Radboud University Nijmegen, The Netherlands, since 2010 where he is currently in a post-doctoral role. In 2007, he was a visiting intern within the Laboratoire Informatique D'Avignon in Avignon, France. His PhD research has concentrated on speaker verification using support vector machine techniques. Mitchell was awarded the 'Best Student Paper Award' at Interspeech 2008 and the 'IEEE 2009 Spoken Language Processing Student Grant' at ICASSP 2009.

Christopher McCool received his PhD with the Speech, Audio, Image and Video Technologies (SAIVT) group, Queensland University of Technology (QUT), Australia in 2007. He received his Bachelor of Information Technology and Bachelor of Engineering (Electronics) also from QUT in 2001. He is currently a Postdoctoral Researcher in biometrics and machine learning at the Idiap Research Institute, he has a particular interest in 2D and 3D face authentication, face detection and computer vision.

Sébastien Marcel received the Ph.D. degree in signal processing from Université de Rennes I in France (2000) at CNET, the research center of France Telecom (now Orange Labs). He is currently interested in pattern recognition and machine learning with a focus on multimodal biometric person recognition. He is a senior research scientist at the Idiap Research Institute (CH), where he leads a research team and conducts research on face recognition, speaker recognition and spoofing attacks detection. In 2010, he was appointed Visiting Professor at the University of Cagliari (IT) where he taught a series of lectures in face recognition. He serves on the Program Committee of several scientific journals and international conferences in pattern recognition and computer vision. Sébastien Marcel is the principal investigator of international research projects including MOBIO (EU FP7 Mobile Biometry) and TABULA RASA (EU FP7 Trusted Biometry under Spoofing Attacks).