IET Biometrics

Research Article

Low-resolution face alignment and recognition using mixed-resolution classifiers

ISSN 2047-4938
Received on 29th February 2016
Revised 10th March 2017
Accepted on 19th March 2017
E-First on 24th April 2017
doi: 10.1049/iet-bmt.2016.0026
www.ietdl.org

Yuxi Peng[1] ✉, Luuk Spreeuwers[1], Raymond Veldhuis[1]
[1]Faculty of EEMCS, University of Twente, Enschede, The Netherlands
✉ *E-mail: y.peng@utwente.nl*

Abstract: A very common case for law enforcement is recognition of suspects from a long distance or in a crowd. This is an important application for low-resolution face recognition (in the authors' case, face region below 40 × 40 pixels in size). Normally, high-resolution images of the suspects are used as references, which will lead to a resolution mismatch of the target and reference images since the target images are usually taken at a long distance and are of low resolution. Most existing methods that are designed to match high-resolution images cannot handle low-resolution probes well. In this study, they propose a novel method especially designed to compare low-resolution images with high-resolution ones, which is based on the log-likelihood ratio (LLR). In addition, they demonstrate the difference in recognition performance between real low-resolution images and images down-sampled from high-resolution ones. Misalignment is one of the most important issues in low-resolution face recognition. Two approaches – matching-score-based registration and extended training of images with various alignments – are introduced to handle the alignment problem. Their experiments on real low-resolution face databases show that their methods outperform the state-of-the-art.

1 Introduction

Biometric face recognition for high-resolution images has been highly successful. However, low-resolution face recognition, which refers to the case where at least the probe images are of low resolution (in our case, a face region below 40 × 40 pixels), is a challenging task because low-resolution face images contain less discriminative information than higher-resolution face images.

In low-resolution face recognition, a common surveillance task is, given a list of suspects, to try to verify whether a person in the surveillance scene is on this list (in face recognition known as gallery). Usually, the gallery images are high-resolution frontal images of high quality. The probe images are taken at a distance and without user cooperation. We call this sort of images ‘real low-resolution images’ as opposed to low-resolution images obtained by down-sampling. The images are not only of low resolution, but also have a higher-noise level and deviate in other ways from the high-resolution gallery images. However, most classifiers are designed to work properly for images of the same high resolution and cannot handle the resolution and quality mismatch. There are three approaches to deal with the resolution mismatch in low-resolution face recognition. The first one is to reconstruct higher-resolution probes using super-resolution techniques, and perform comparison with the gallery images in the high-resolution space [1, 2]. The second approach is to down-sample the high-resolution galleries and compare them with the probes in the low-resolution space. Since the low-resolution images are smaller than the high-resolution ones, this approach involves lower computational costs. The third approach is to compare the low-resolution probe images to the high-resolution gallery images directly. Most methods following this approach find mappings to project both low-resolution and high-resolution images to a common space in which a direct comparison is performed.

It is hard to compare the existing methods directly due to the lack of a common protocol: everyone uses a different experimental setting. Another issue is that most researchers only present closed-set identification results such as the rank-1 recognition rate. These results depend highly on the size of the test sets, the number of subjects and on whether the classifier was trained on the galleries. Verification results are better suited for direct comparison and they directly address a relevant biometric question: ‘Are the two images of the same person?’.

The evaluation of most existing methods for low-resolution face recognition is based on images that were aligned at high resolution and then smoothed and down-sampled (ds) to low resolution. However, we observe that in [3–6] the face recognition performance on ds images is much better than on real low-resolution images. In this paper, we will confirm these observations and conclude that, in order to produce realistic results, experiments should be based on real low-resolution images.

In this paper, instead of proposing a general face recognition method, we deal with problems in the specific scenario as mentioned above: comparison of low-resolution probes with high-resolution galleries. We extend our work in [7] concerning a method especially designed for comparing images captured at different distances, called mixed-resolution biometric comparison. This method not only works for different resolutions, but also captures other differences between images recorded at various distances. Since proper alignment proves to be crucial for a good recognition performance, and precise alignment (e.g. via accurately detected facial landmarks) is difficult for low-resolution images, we provide two methods to deal with the alignment problem: matching-score-based registration and extended training with images with slightly varying alignment. We demonstrate that the combination of the three methods outperforms the state-of-the-art for real low-resolution face recognition. In addition, we address the differences in face recognition performance between ds and real low-resolution images and pay special attention to proper evaluation protocols.

We conduct experiments of different settings using images of varying resolutions. First, we duplicate the experimental protocol of a state-of-the-art method so that we can directly compare our methods to it. Then, we set up a more realistic experiment to demonstrate that our methods are effective in realistic situations.

In the remainder of this paper, we use the terms *HiRes*, *LoRes* and *SupRes* for *high-resolution*, *low-resolution* and *super-resolution*.

This paper is organised as follows: Section 2 is a literature review of existing low-resolution face recognition methods. In Section 3, we explain why researchers should present results from real low-resolution images by demonstrating the difference

Table 1 Papers using ds data as probe

Method	Database	N_C	N_G	N_P	IMsize	Rank-1, %	Approach
CKE [5]	multi-PIE	229	7	13	6×6	88	LoRes–HiRes
CDA [4]	multi-PIE	100	10	10+	6×6	79	LoRes–HiRes
MDS [3]	multi-PIE	237	1	1	12×10	81	LoRes–HiRes
DTCWT [8]	multi-PIE	202	1	20	20×20	99	SupRes–HiRes
SDA [9]	multi-PIE	149	1	10+	12×12	70	LoRes–HiRes
S2R2 [10]	multi-PIE	224	1	13	6×6	73	SupRes–HiRes
MFF [11]	FERET	200	1	1	12×12	84	SupRes–HiRes
NMCF [12]	FERET	1195	1	1	12×12	84	LoRes–HiRes
CLPM [13]	FERET	1195	1	1	12×12	90	LoRes–HiRes
graph DA on multi-manifold [14]	AR	126	7	7	8×7	72	SupRes–HiRes
MM [1]	CMUvideo	68	1	16	23×23	81	SupRes–HiRes
EigenSR [15]	CMUvideo	68	1	16	10×10	74	SupRes–HiRes
EigenTr [16]	XM2VTS	295	1	1	10×10	59	SupRes–HiRes
DSR [6]	FRGC v2.0	311	8	2	7×6	78	SupRes–HiRes

N_C is number of subjects for testing, N_G is number of images per subject in the gallery set, N_P is number of images per subject in the probe set and IMsize is the probe image size (in pixels). CKE: Coupled kernel embedding, CDA: Coupled discriminant analysis, MDS: Multidimensional scaling, SDA: simultaneous discriminant analysis, MFF: Multi-resolution feature fusion, NMCF: nonlinear mappings on coherent features, DTCWT: Dual-Tree Complex Wavelet Transform , CLPM: coupled locality preserving mappings, DSR: discriminative super-resolution, CBD: Semi-coupled basis and distance metric learning, MM: Morphable model, EigenTr: eigen transformation.

Table 2 Papers using real LoRes data as probe

Method	Database	Gallery	Probe	N_C	N_G	N_P	Rank-1, %	Approach
CKE [5]	SCface	mug-shot	dist1	130	1	5	8	LoRes–HiRes
DSR [6]	SCface	dist2	dist1	130	5	5	22	SupRes–HiRes
CBD [17]	SCface	dist3	dist2	100	4	1	53	LoRes–HiRes
CDA [4]	localvideo	Photograph	video	161	5	5	53	LoRes–HiRes
DTCWT [8]	localvideo	Photograph	video	34	1?	1	56	SupRes–HiRes

N_C is number of subjects for testing, N_G is number of images per subject in the gallery set and N_P is number of images per subject in the probe set. As for the SCface database, the images from dist1, dist2 and dist3 are captured at a distance of 4.2, 2.6 and 1.0 m.

between using ds and real low-resolution probes. Our proposed methods – mixed-resolution biometric comparison, matching-score-based registration and extended training are introduced in Section 4. In Section 5 we report results and Section 6 presents the conclusions.

2 Approaches to LoRes–HiRes comparison

Face recognition for low resolution is different from face recognition for high resolution. First, it is much harder to detect landmarks reliably and accurately in LoRes images. In addition, LoRes images contain far less discriminative information than HiRes images. There is a small, but growing body of literature that specifically addresses the problem of LoRes face recognition. Important references of LoRes face recognition are listed in Tables 1 and 2. Table 1 lists papers that present experiments on ds probe images and Table 2 lists papers using real LoRes probe images. We separate the papers in two tables because the evaluation of the two types of data is different. We will discuss the differences in Section 3. In the tables, we include the methods, the experimental settings and rank-1 recognition rates. As we can see, each paper presents experiments in a different setting even when they use the same database. For example, there are three papers which conducted experiments on the surveillance camera face (SCface) database in Table 2, but they have different numbers of subjects for testing, different numbers of gallery and probe images per subject. This makes it impossible to compare the methods objectively.

As mentioned in Section 1, existing methods solve the resolution mismatch problem mainly by following the three approaches: applying SupRes to LoRes probes, down-sampling HiRes gallery images and direct LoRes–HiRes comparison. We will discuss the existing methods based on the three approaches.

2.1 SupRes versus HiRes comparison

Since most face recognition systems are designed for HiRes images, many researchers reconstruct HiRes versions from the LoRes probes using SupRes techniques to make use of the information contained in the HiRes gallery images and conduct comparison in the HiRes space.

The simplest SupRes approach is interpolation. It up-samples the LoRes images, but does not use additional information about the images, e.g. that they are faces. Thus, the resulting images usually have poor recognition performance because there is still a large difference between the gallery and probe images.

Face hallucination refers to SupRes techniques that were specially designed to improve face image quality. Baker and Kanade [18] proposed a method that learns a priori on the spatial distribution of image gradients for frontal face images. Then, this prior is incorporated in the maximum a posteriori (MAP) framework. Wang and Tang [16] proposed a face hallucination method using eigen transformation. The input LoRes image is represented as a linear combination of the LoRes images in the training set by principal component analysis (PCA). The SupRes image is reconstructed using the corresponding HiRes training images with the same coefficients.

Though the face hallucination methods can enhance visual image quality, the restored information may not contribute to better face recognition. Therefore, researchers started to work on SupRes methods that aim to improve face recognition performance. Gunturk *et al.* [15] proposed to apply SupRes in an eigen domain that reconstructs only the necessary information for recognition. Hennings-Yeomans *et al.* [19] built a model for SupRes based on Tikhonov regularisation and a linear feature extraction stage. This model can be applied when images from training, gallery and probe sets have varying resolutions. This approach is extended in [10] by adding a face prior to the model and using relative residuals as measures of fit. Zou and Yuen [2] developed a data constraint to minimise both the distances between the constructed SupRes images and the corresponding HiRes images as well as the distances between SupRes images from the same class. This method is extended in [6] with a linearity cluster. Bilgazyev *et al.* [8] proposed a method that uses dual-tree complex wavelet transform to extract high-frequency components of training images. Then, the SupRes features are represented as a weighted combination of the HiRes training images and the weights are the same as the ones that represent the input LoRes probe using corresponding LoRes training images. Zhang *et al.* [1] proposed a SupRes method in morphable model space, which provides HiRes information required by both reconstruction and recognition.

Though applying SupRes to LoRes probes makes the comparison with HiRes galleries possible in the HiRes space, the process of SupRes usually brings in artefacts or noise which might influence the face recognition performance.

2.2 LoRes versus ds HiRes comparison

Down-sampling HiRes gallery images and conducting comparison in the LoRes space is a very simple way for LoRes to HiRes comparison. It requires lower computational costs than using HiRes images. Although some information in the HiRes images will be lost in the down-sampling process, it has been reported by some researchers that this approach has similar recognition performance as SupRes methods for LoRes face recognition. For instance, Hu *et al.* [20] conducted experiments using a video database of moving faces and people. Their experimental results show that applying SupRes methods and then comparing with HiRes galleries have similar performance as LoRes to LoRes comparison at a far range (5–10 pixel eye-to-eye distance). Xu *et al.* [21] showed that when image resolution is low enough, LoRes to LoRes comparison is superior to using SupRes methods. In their experiments conducted on Yale B and acceptance rate (AR) databases with ds images, SupRes methods perform much poorer than LoRes to LoRes comparison when the image size is 8 × 8 pixels. These results suggest that down-sampling gallery images and comparing with LoRes probes has at least as good face recognition performance as applying SupRes on LoRes probes and comparing with galleries in the HiRes domain.

2.3 LoRes versus HiRes comparison

Direct comparison of LoRes probes and HiRes galleries is a new area that has drawn researchers' attention in recent years. Most methods of this approach find transformations for both LoRes and HiRes images and compare their features in a common space. This approach avoids losing information as a result of down-sampling HiRes or adding artefacts by SupRes. The mappings between HiRes gallery and LoRes probe data can also be learnt in such a way that different variations are modelled.

Li *et al.* [13] proposed a method that projects both HiRes galleries and LoRes probes to a unified feature space for classification using coupled mappings. The mappings are learnt by optimising the objective function that minimises the difference between corresponding HiRes and LoRes images. Huang and He [12] proposed a method that uses canonical correlation analysis to project the PCA features of HiRes and LoRes image pairs to a coherent feature space. Radial-based functions are then applied to find the mapping between the HiRes and LoRes pairs. A multidimensional scaling-based method is proposed by Biswas *et al.* [3]. Both HiRes and LoRes images are transformed to a common space where the distance between them approximates the distance when they are both HiRes. The transformations are learnt using an iterative majorisation algorithm. Ren *et al.* [5] proposed a method called coupled kernel embedding. It projects the original HiRes and LoRes images onto reproducible kernel space using coupled non-linear functions. The dissimilarities captured by their kernel Gram matrices are minimised in this space. Lei *et al.* [4] proposed a coupled discriminant analysis (DA) method. They find coupled transformations to project HiRes and LoRes images to a common space in which the low-dimensional embedding is well classified. The locality information in kernel space is also used as a constraint for the DA process. This method is also suitable for images of different modalities, for example, visible and infrared faces. Moutafis and Kakadiaris [17] proposed a method that learns semi-coupled mappings for HiRes and LoRes images for optimised representations. The mappings aim at increasing class-separation for HiRes images and projecting LoRes images to their corresponding class-separated HiRes data.

3 Evaluation using ds or real LoRes images

3.1 Problem statement

As we stated in the previous section, most papers test the recognition performance of their proposed methods on ds probe images. Some researchers conducted experiments on real LoRes databases as well as on ds probe images. However, we observe that face recognition methods perform much worst on real LoRes images than ds images. In [6] for instance, the proposed method achieves 78% rank-1 recognition rate on images ds to 7 × 6 pixels from the Face Recognition Grand Challenge (FRGC) v2.0 database, whereas the result is only 22% on dist1 images (about 30 × 30 pixels) from the SCface database. Also in [5], the rank-1 recognition rate on ds images of 6 × 6 pixels from the multi-PIE databases is 88%; however, it is only 6% on dist1 images from the SCface database. In [4], the rank-1 recognition rate of the proposed method is 97% on images ds to 16 × 16 pixels from the multi-PIE database, whereas it is only 52% on a self-collected database, where the image size is 35 × 35 pixels.

3.2 Analysis experiment

Our hypothesis is that there are differences between ds images and real LoRes images that result in poorer face recognition results for the latter. To investigate this and exclude other influences, we set up a face recognition experiment where the ds images and the real LoRes images are from the same source and pose, and there are no pose and illumination variations. Since there is no database available that meets this requirement, we collected a database ourselves.

We used a commercial camera CASIO EX-FC100. The images were recorded in the following way: the subject sat still and faced the camera. The first photograph of each subject was taken at a distance of 2 m, and then we moved the camera 1 m away and took a photograph each time until we had nine photographs of this subject. Thus, the face images (probes) were taken at nine distances in total from 2 to 10 m. Two weeks later, gallery images of each person were captured at a distance of 1 m with the same setup. The faces were always frontal and with the same illumination. About 25 subjects are included in our database. The original images are around 1 MB and of resolution 1600 × 1200. The image file format is JPEG. During pre-processing, all images were aligned using manually annotated eye-coordinates. The size of cropped face regions is 243 × 243 pixels for the galleries, and 131 × 131, 87 × 87, 64 × 64, 51 × 51, 44 × 44, 36 × 36, 33 × 33, 28 × 28 and 23 × 23 pixels for distance 2–10 m, respectively. An elliptic mask is applied to select the region of interest. Histogram equalisation is used to normalise the illumination. Unfortunately, we are not able to make this database available, but we are willing to evaluate algorithms on request. Sample images are shown in Fig. 1.

To test face recognition performance on our database, we use three different probe sets: the first one contains images taken at distances from 2 to 10 m (Real); the second one contains images taken at 2 m ds to the same resolution as the first probe set and then aligned using the eye-coordinates annotated at the highest resolution (Downsample); the third one also contains ds images, but they are aligned using eye-coordinates annotated after the down-sampling process (Align after ds). The type of images in the second set, which are pre-aligned before down-sampling, is commonly used for evaluation of LoRes face recognition methods. From now on, we shall refer to pre-aligned ds images when we mention ds images. The gallery images are ds to the same resolution as the probe images for comparison. To train the face classifiers, 4471 high-quality images of 275 subjects from the FRGC database [22] are used. We use four face recognition methods. One of them is the state-of-the-art LoRes face recognition

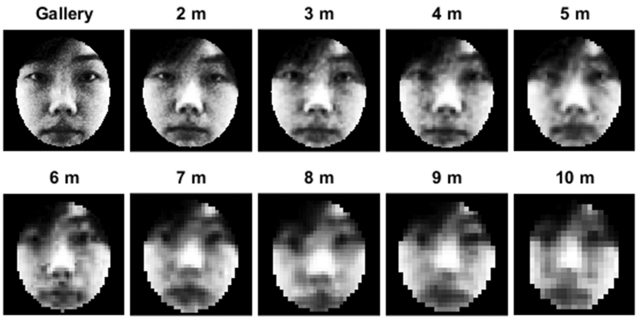

Fig. 1 *Sample images from our own database*

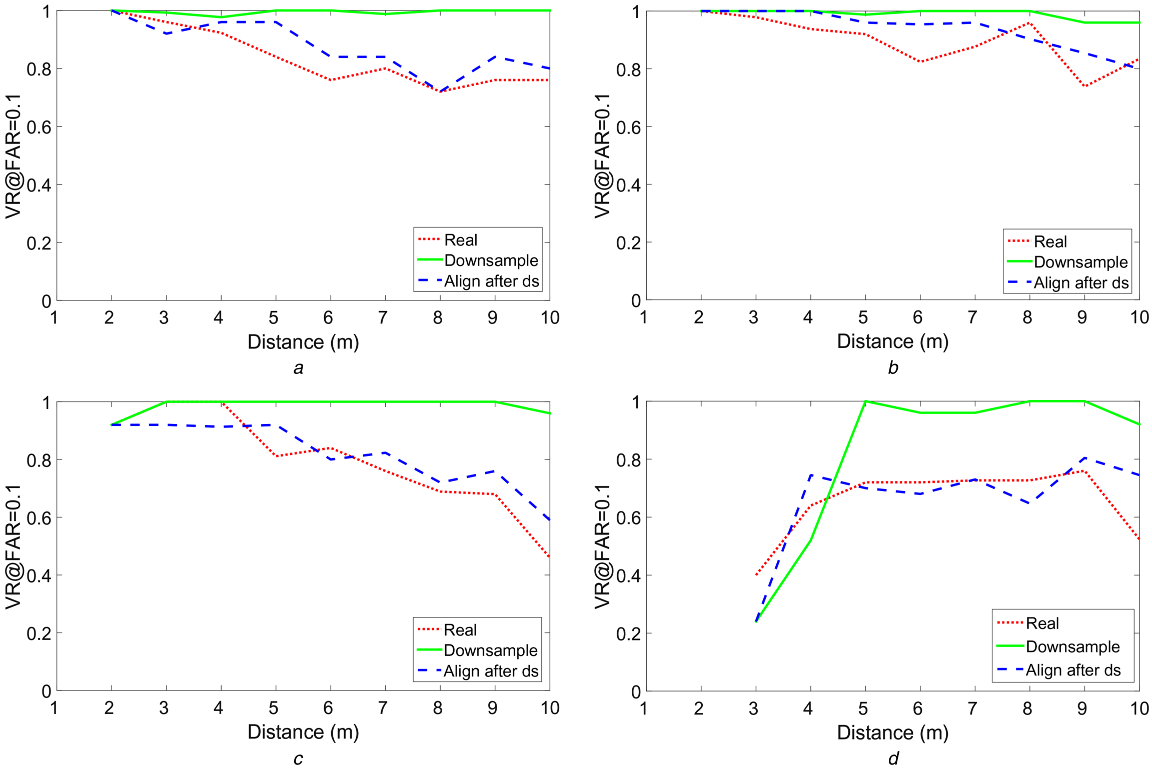

Fig. 2 *VR at FAR 10% on images from various distances using different classifiers*
(a) PCA, ***(b)*** LDALLR, ***(c)*** LBP, ***(d)*** CLPM

method CLPM [13] (detailed explanation of CLPM is in Section 5). The other three methods are classical face classifiers, namely PCA [23], linear DA LLR (LDALLR) [24] and local binary patterns (LBPs) [25]. The LDALLR is based on the LDA classifier [26], but computes LR similarity scores instead of distance measurements. Distance measures employed for PCA and LBP are $L1$-norm and chi square, respectively. To obtain overall good performance, 80 PCA and 50 LDA vectors are chosen. The images are divided into 6×6 regions for LBP. The feature dimension selected for the CLPM method is 170 (Li *et al.* [13] chose 80, but 170 gives better performance).

We present verification results because, as argued in Section 1, they allow better comparison and address a more realistic biometric question than rank-1 recognition rates. It must be remarked in advance that the verification performance for low-resolution images is poorer than for high-resolution images. This is also illustrated in [7], where the performance of the state-of-the-art low-resolution face recognition is presented. Besides, LoRes face recognition is commonly used for surveillance, by which good verification performance is more important than limiting false AR (FAR). For the above reasons, verification rates (VRs) (also known as genuine AR) are presented at FAR 10% rather than the more common FAR equals to 1 or 0.1% (see Fig. 2).

As we can see in this figure, the performance of all the four classifiers has a similar trend. When the images are pre-aligned and then ds, the face recognition performance remains stable for every resolution. The performance on both the real LoRes images and the post-aligned ds images become worst when the distances increase from 2 to 10 m. Although the post-aligned ds images generally have better results than the real LoRes images, the differences are much smaller than when compared with the results from the pre-aligned images. This trend demonstrates that face recognition performance on the pre-aligned ds images is different from the performance on real LoRes images and alignment plays a very important role.

There is also different behaviour of different face recognition methods on images from different distances. The CLPM method, as it was designed to operate optimally for LoRes images, could not handle images from 3 and 4 m and it performs stably for the rest of the resolutions. We have a memory problem for the 2 m images from the available code from the authors of this method, but it does not affect the trend. The other three methods all perform worst on lower-resolution images than higher-resolution ones, while LBP is more sensitive to resolution changes than PCA and LDALLR. In addition, LDALLR has the best performance among the four face recognition methods.

3.3 Conclusion

In this section, we have demonstrated that ds images are not fully representative of realistic low-resolution images and hence should not be used as probes when testing the effectiveness of LoRes face recognition methods. Face recognition methods perform much better on ds images than on real LoRes images. The fact that ds images are usually pre-aligned at HiRes while real LoRes images are aligned at LoRes plays an important role. Other factors such as viewing angle, Bayer mask, noise level and compression artefacts may influence the recognition performance as well. We identify alignment as one of the most important factors. Besides, in real LoRes face recognition applications, pose, illumination and facial expression may differ significantly between gallery and probe, though we do not deal with them in this paper. Thus, real LoRes data should be used for the evaluation of LoRes face recognition methods to ensure their feasibility in reality.

4 Proposed methods

4.1 Mixed-resolution biometrics comparison

Here, we present a method especially designed for comparing images captured at different distances, called mixed-resolution biometric comparison. This method not only works for different resolutions, but also learns variations in the image quality. It was first derived in [7], but is repeated here for completeness. A similar method for homogeneous cases was proposed in [27], but we consider heterogeneous cases where the gallery and probe are from different scenarios.

Given two biometric feature vectors $\boldsymbol{x} \in \mathbb{R}^M$ and $\boldsymbol{y} \in \mathbb{R}^N$ obtained from multi-resolution acquisition devices we look for support for the hypothesis H_s: *the samples originate from the same individual* versus H_d: *the samples originate from different individuals*, quantified by the LR

$$l(\boldsymbol{x}, \boldsymbol{y}) = \frac{p\left(\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix} | H_\mathrm{s}\right)}{p\left(\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix} | H_\mathrm{d}\right)} \tag{1}$$

It is well known that an optimal classifier in the Neyman–Pearson sense is obtained by thresholding the LR, compare for example [28]. This means the LR will give the highest VR at a given FAR.

Note that $\boldsymbol{x}$, with dimension M, is a realisation of a feature vector of a random individual, characterised by its feature mean, which is therefore also random. Similarly, $\boldsymbol{y}$ is also a realisation of a feature vector of a random individual, but with dimension N. We take $M \geq N$, i.e. $\boldsymbol{x}$ is of higher resolution than $\boldsymbol{y}$. Let ω and θ denote the identities of $\boldsymbol{x}$ and $\boldsymbol{y}$, we assume that $\boldsymbol{x} = \mu_\omega + \boldsymbol{w}_\omega$ and $\boldsymbol{y} = \mu'_\theta + \boldsymbol{w}'_\theta$, with $\mu_\omega = E\{\boldsymbol{x} | \omega\} \in \mathbb{R}^M$ and $\mu'_\theta = E\{\boldsymbol{y} | \theta\} \in \mathbb{R}^N$ the subject-specific mean, modelling the between-subject variations, and with $\boldsymbol{w}_\omega$ and $\boldsymbol{w}'_\theta$ the statistically independent, zero-mean within-subject variations. Furthermore, we assume normal, zero-mean probability densities for μ_ω and μ'_θ for unknown ω and θ, and for $\boldsymbol{w}_\omega$ and $\boldsymbol{w}'_\theta$. If $\boldsymbol{x}$ and $\boldsymbol{y}$ are not zero mean, estimated means have to be subtracted prior to comparison. Such a simple model cannot be expected to work well for HiRes face recognition, but when LoRes faces with fewer details are involved it can still be applied successfully. The covariance matrices of $\boldsymbol{x}$ and $\boldsymbol{y}$ are

$$\boldsymbol{\Sigma}_{xx} = \mathrm{E}\{\boldsymbol{x}\boldsymbol{x}^\mathrm{T}\} \in \mathbb{R}^{M \times M} \quad \text{and} \quad \boldsymbol{\Sigma}_{yy} = \mathrm{E}\{\boldsymbol{y}\boldsymbol{y}^\mathrm{T}\} \in \mathbb{R}^{N \times N} \tag{2}$$

respectively, and the cross-covariance matrices are

$$\boldsymbol{\Sigma}_{xy} = \mathrm{E}\{\boldsymbol{x}\boldsymbol{y}^\mathrm{T}\} \in \mathbb{R}^{M \times N} \quad \text{and} \quad \boldsymbol{\Sigma}_{yx} = \boldsymbol{\Sigma}_{xy} \tag{3}$$

respectively. Then $\boldsymbol{\Sigma}_{xy} = \mathrm{E}\{\mu_\omega\mu'\mathrm{T}_\theta | \omega = \theta\}$. If $\omega \neq \theta$, $\boldsymbol{\Sigma}_{xy} = 0$. For the probability densities of the pairs of feature vectors we then have, respectively

$$\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix} | H_\mathrm{s} \sim \mathcal{N}\left(\boldsymbol{0}, \begin{pmatrix}\boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy}\\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy}\end{pmatrix}\right) \tag{4}$$

$$\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix} | H_\mathrm{d} \sim \mathcal{N}\left(\boldsymbol{0}, \begin{pmatrix}\boldsymbol{\Sigma}_{xx} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{\Sigma}_{yy}\end{pmatrix}\right) \tag{5}$$

Covariance and cross-covariance matrices need to be estimated in a training process. The cross-covariance matrix $\boldsymbol{\Sigma}_{xy}$ is estimated as

$$\hat{\boldsymbol{\Sigma}}_{xy} = \frac{1}{K}\sum_{i=1}^{K} \hat{\mu}_i\hat{\mu}'\mathrm{T}_i \tag{6}$$

with K as the number of individuals involved in training and $\hat{\mu}_i$ and $\hat{\mu}'_i$ as the estimated sample means of subject i. It can be shown that estimating $\boldsymbol{\Sigma}_{xy}$ in this way is equivalent to estimating $\boldsymbol{\Sigma}_{xy}$ from all the possible combinations of pairs of feature vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ in a training set. The rank of $\hat{\boldsymbol{\Sigma}}_{xy}$ can be at most min $(N, K - 1)$. The −1 is included because the sample means are zero mean.

By substituting the normal probability density functions (PDFs) corresponding to (4) and (5) into (1), taking the log and ignoring some constants, we arrive at the following similarity score:

$$s(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x}^\mathrm{T}\boldsymbol{y}^\mathrm{T})\left(\begin{pmatrix}\boldsymbol{\Sigma}_{xx} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{\Sigma}_{yy}\end{pmatrix}^{-1} - \begin{pmatrix}\boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy}\\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy}\end{pmatrix}^{-1}\right)\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix} \tag{7}$$

This score is optimal because it monotonically increases with the LLR. To simplify (7) and to assure that the estimated covariance matrices have full rank and can be inverted, we simultaneously reduce the dimensionality and apply whitening transforms to $\boldsymbol{x}$ and $\boldsymbol{y}$, resulting in

$$\boldsymbol{x}_\mathrm{w} = \boldsymbol{W}_\mathrm{H}\boldsymbol{x} \in \mathbb{R}^{M_\mathrm{w}} \quad \text{and} \quad \boldsymbol{y}_\mathrm{w} = \boldsymbol{W}_\mathrm{L}\boldsymbol{y} \in \mathbb{R}^{N_\mathrm{w}} \tag{8}$$

with dimensionalities M_w and N_w, respectively. Usually $M_\mathrm{w} < M$, $N_\mathrm{w} < N$ and $M_\mathrm{w} \geq N_\mathrm{w}$. As a result we have that

$$\begin{gathered}\boldsymbol{\Sigma}_{xx}^\mathrm{w} = E\{\boldsymbol{x}_\mathrm{w}\boldsymbol{x}_\mathrm{w}^\mathrm{T}\} = \boldsymbol{I}, \quad \boldsymbol{\Sigma}_{yy}^\mathrm{w} = \mathrm{E}\{\boldsymbol{y}_\mathrm{w}\boldsymbol{y}_\mathrm{w}^\mathrm{T}\} = \boldsymbol{I} \quad \text{and} \quad \boldsymbol{\Sigma}_{xy}^\mathrm{w}\\ = \boldsymbol{W}_H\boldsymbol{\Sigma}_{xy}\boldsymbol{W}_L^\mathrm{T}\end{gathered} \tag{9}$$

with $\boldsymbol{I}$ as an identity matrix of appropriate size. The similarity score then becomes

$$s(\boldsymbol{x}_\mathrm{w}, \boldsymbol{y}_\mathrm{w}) = (\boldsymbol{x}_\mathrm{w}^\mathrm{T}\boldsymbol{y}_\mathrm{w}^\mathrm{T})\left(\begin{pmatrix}\boldsymbol{I} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}\end{pmatrix}^{-1} - \begin{pmatrix}\boldsymbol{I} & \boldsymbol{\Sigma}_{xy}^\mathrm{w}\\ \boldsymbol{\Sigma}_{yx}^\mathrm{w} & \boldsymbol{I}\end{pmatrix}^{-1}\right)\begin{pmatrix}\boldsymbol{x}_\mathrm{w}\\ \boldsymbol{y}_\mathrm{w}\end{pmatrix} \tag{10}$$

We will further simplify (10). First, we apply a singular value decomposition to $\boldsymbol{\Sigma}_{xy}^\mathrm{w}$, such that

$$\boldsymbol{\Sigma}_{xy}^\mathrm{w} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\mathrm{T} \tag{11}$$

with $\boldsymbol{U} \in \mathbb{R}^{M_\mathrm{w} \times M_\mathrm{w}}$, and orthonormal, $\boldsymbol{V} \in \mathbb{R}^{N_\mathrm{w} \times N_\mathrm{w}}$, and orthonormal and $\boldsymbol{D} \in \mathbb{R}^{M_\mathrm{w} \times N_\mathrm{w}}$. The first N_w rows of $\boldsymbol{D}$ form a diagonal matrix consisting of singular values ν_i, $i = 1, \ldots, N_\mathrm{w}$ in decreasing order. The last $M_\mathrm{w} - N_\mathrm{w}$ rows of $\boldsymbol{D}$ are an all-0 matrix. In a trained classifier, the rank of $\boldsymbol{D}$ can be at most $D =$ min $(N_\mathrm{w}, K - 1)$, with K as the number of individuals in the training set. If a smaller feature vector is desired, D can be chosen to be less than min $(N_\mathrm{w}, K - 1)$. We now transform the feature vectors again, such that

$$\boldsymbol{x}_c = (\boldsymbol{U}_{*,1:D})^\mathrm{T}\boldsymbol{x}_\mathrm{w} \in \mathbb{R}^D \quad \text{and} \quad \boldsymbol{y}_c = (\boldsymbol{V}_{*,1:D})^\mathrm{T}\boldsymbol{y}_\mathrm{w} \in \mathbb{R}^D \tag{12}$$

where the subscript *, 1:D denotes that only the first D columns of matrix are taken. The subscript c indicates that these transformations map the feature vectors to a common subspace. It can be shown that these transformations, which reduce the feature dimensionality to D, will result in the same similarity score as transformations using the full matrices $\boldsymbol{U}$ and $\boldsymbol{V}$. For the similarity score we now have

$$s(\boldsymbol{x}_c, \boldsymbol{y}_c) = (\boldsymbol{x}_c^\mathrm{T}\boldsymbol{y}_c^\mathrm{T})\left(\begin{pmatrix}\boldsymbol{I} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}\end{pmatrix}^{-1} - \begin{pmatrix}\boldsymbol{I} & \boldsymbol{D}\\ \boldsymbol{D} & \boldsymbol{I}\end{pmatrix}^{-1}\right)\begin{pmatrix}\boldsymbol{x}_c\\ \boldsymbol{y}_c\end{pmatrix} \tag{13}$$

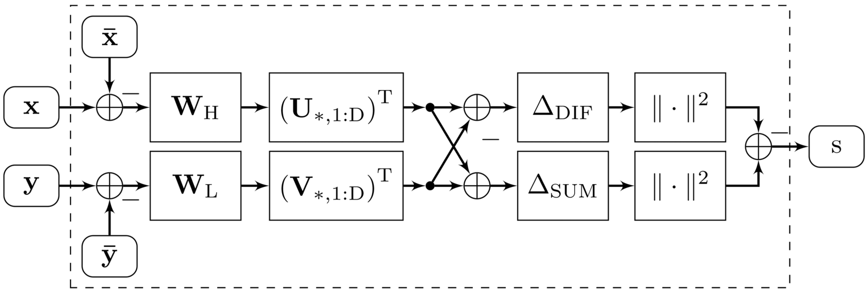
$\bar{\mathbf{x}}$
$\mathbf{x}$
−
$\mathbf{W}_{\mathrm{H}}$
$(\mathbf{U}_{*,1:\mathrm{D}})^{\mathrm{T}}$
−
Δ_{DIF}
$\|\cdot\|^2$
−
s
$\mathbf{y}$
−
$\mathbf{W}_{\mathrm{L}}$
$(\mathbf{V}_{*,1:\mathrm{D}})^{\mathrm{T}}$
Δ_{SUM}
$\|\cdot\|^2$
$\bar{\mathbf{y}}$

Fig. 3 *Block diagram of the mixed-resolution classifier according to (14)*

with $\boldsymbol{D} \in \mathbb{R}^{D\times D}$ redefined as a diagonal matrix with the D largest singular values ν_i of $\boldsymbol{\Sigma}_{xy}^{\mathrm{w}}$ on the diagonal. After some manipulations, we obtain

$$s(\boldsymbol{x}_c, \boldsymbol{y}_c) = -\sum_{i=1}^{D} \frac{\nu_i}{1-\nu_i}(x_{c,i} - y_{c,i})^2 + \sum_{i=1}^{D} \frac{\nu_i}{1+\nu_i}(x_{c,i} + y_{c,i})^2 \qquad (14)$$

In (14) a factor of 1/2 has been left out. A full expression for the LLR that includes all the constants that have been ignored is

$$\log(l(\boldsymbol{x}_c, \boldsymbol{y}_c)) = -\frac{1}{2}\sum_{i=1}^{D}\log\left(1-\nu_i^2\right) + \frac{1}{4}s(\boldsymbol{x}_c, \boldsymbol{y}_c) \qquad (15)$$

Since the ν_i depends on training data, the use of this full expression is recommended in n-fold cross-validation experiments, since then the first term may differ slightly per validation step. Fig. 3 shows a block diagram of the classifier according to (14). The blocks perform matrix multiplications, except the rightmost ones, which compute a squared vector norm. The vectors $\bar{\boldsymbol{x}}$ and $\bar{\boldsymbol{y}}$ are the average HiRes and LoRes facial images, respectively. The matrices Δ_{DIF} and Δ_{SUM} are diagonal matrices, defined by $\Delta_{\mathrm{DIF},\,ii} = \sqrt{\nu_i/(1-\nu_i)}$, $i = 1, \ldots, D$ and $\Delta_{\mathrm{SUM},\,ii} = \sqrt{\nu_i/(1+\nu_i)}$, $i = 1, \ldots, D$, respectively.

In a similar way, a likelihood-ratio-based classifier can be derived for other types of heterogeneous features, e.g. for visual light and near infrared facial images, and for the case that feature sets of possibly different numbers if multiple captures must be compared. In the following part of this paper, we use the term *MixRes* for this method. The MixRes method has been implemented in MATLAB. The code is available on request.

4.2 Matching-score-based registration

In the LoRes face recognition field, manually marked eye-coordinates are often still used, because landmarks cannot reliably be detected automatically. However, even the manual landmarks are usually not accurate enough because the eyes are not clear when the images are too small [29]. On the other hand, the variations between the true eye-coordinates and the manually marked ones are small in pixels, for example, the variations are within [−2, 2] pixels if the distance between the eyes is 10 pixels. Thus, we propose to use matching-score-based registration to benefit LoRes face recognition. This method was proposed for inaccurately aligned faces in [30], and its effectiveness on HiRes images has already been demonstrated in [31, 32].

In a normal face recognition system, the probe and reference images are registered using facial landmarks, and then the aligned images are compared. For a probe image x_{p} and a reference image x_{r}, given the eye-coordinates of the probe image ρ (reference images are assumed to be pre-aligned), the similarity score is written as $s(x_{\mathrm{p}}(\rho), x_{\mathrm{r}})$. For matching-score-based registration, the alignment of the probe image is varied resulting in several aligned images. All those images are compared with each gallery image using the chosen classifier. The best result for each gallery image is stored as the genuine or imposter score which is used for the subsequent verification or identification process. Matching-score-based registration tries to find the eye-coordinates ρ^* that maximises the similarity between $x_{\mathrm{p}}(\rho)$ and x_{r} (16)

$$\boldsymbol{\rho}^* = \underset{\rho}{\operatorname{argmax}}\, s(x_{\mathrm{p}}(\rho), x_{\mathrm{r}}) \qquad (16)$$

The output similarity score is then $s^*(x_{\mathrm{p}}(\rho^*), x_{\mathrm{r}})$. We call this method MSBR for short.

4.3 Extended training

An alternative way to compensate for misalignment is to include misaligned images in the training set [33]. Unlike in MSBR, different aligned images are generated in the training set instead of on the probe images. The training set contains HiRes and LoRes image pairs. HiRes images are assumed to be perfectly aligned. However, for each LoRes image, the eye-coordinates are varied to generate a set of (mis-)aligned images. The possible misalignment in the probe is thus modelled using this extended training set. The given eye-coordinates are noted as ρ which can be obtained by manual marking. We vary ρ to obtain m different coordinates $[\rho_1, \rho_2, \ldots, \rho_m]$. Thus, for one LoRes training image x_{L}, m images are generated with different alignment $[x_{\mathrm{L}}(\rho_1), x_{\mathrm{L}}(\rho_2), \ldots, x_{\mathrm{L}}(\rho_m)]$. This extended LoRes set together with the original HiRes set will form the new training set. This method will be denoted as ET.

5 Experiments

In this section, we demonstrate the effectiveness of our proposed methods using real LoRes data. The SCface database is chosen in our experiments because it contains surveillance quality face images. We did not choose other commonly used databases such as The Facial Recognition Technology (FERET) database, FRGC or Labeled Faces in the Wild (LFW) because their image resolutions are much higher.

As shown in Tables 1 and 2, many publications use different experimental settings and there is no straightforward way to select the best method. We chose to compare our method with three state-of-the-art LoRes face recognition methods CBD [17], CLPM [13] and DSR [2] and the state-of-the-art HiRes face recognition method FaceVACS [34]. The CBD method is proposed in a recent publication and we are able to duplicate their protocol of evaluation using surveillance quality images from the SCface database. The CLPM method was not evaluated using real LoRes data in [13], but the source code is available so we can test it in our experiments. The correctness of the CLPM code was evaluated using the same setting on FERET database as in [13]. The rank-1 recognition rate with 80 features is 90%, which is the same as reported in [13]. For the DSR method, we used our own implementation of the method described in [2]. We repeated the experiment described in [2] on SCface data to verify our implementation and found that our implementation had a 24% rank-1 recognition rate, which was slightly better than the result in [2]. FaceVACS is a commercial face recognition software developed by Cognitec Systems GmbH. We use FaceVACS for comparison because commercial systems are also used in real surveillance cases. Note that FaceVACS was not designed for very LoRes facial images and thus is used here outside its normal

Table 3 Parameters of MixRes and its combination with MSBR and ET. M_w and N_w are the number of feature vectors of HiRes and LoRes training after the first dimensionality reduction

Method	M_w	N_w	D
MixRes, +ET	70	60	40
+MSBR, +both	100	100	60

D is the number of feature vectors after the second dimensionality reduction.

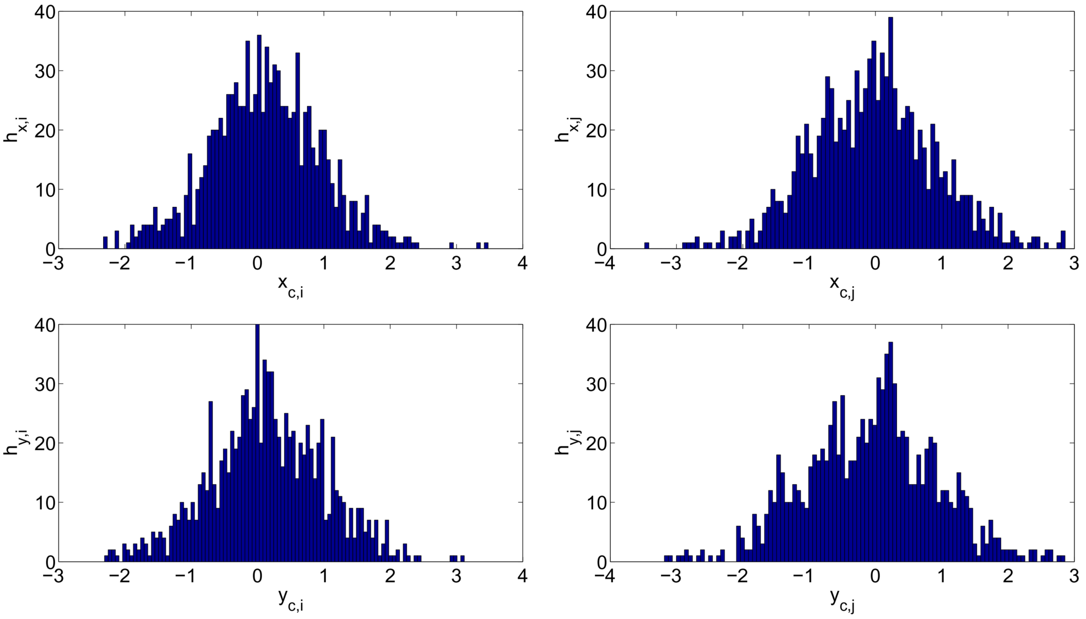

Fig. 4 *Histograms illustrating the normal distribution for facial features for MixRes. Top row: histograms of HiRes feature elements and bottom row: histograms of LoRes feature elements*

Table 4 Resolutions (pixels) of gallery and probe images in each section

Section	HiRes gallery		LoRes probe	
	Original	Actual	Original	Actual
5.1	68 × 55	30 × 24	47 × 38	15 × 12
5.2	800 × 800	80 × 80	32 × 32	32 × 32
5.3	243 × 243	80 × 80	33 × 33	32 × 32

specifications. For HiRes images, it outperforms our method by a large margin.

In our experiments, the parameters of each method are chosen to ensure a good performance. The feature dimension selected for CLPM is 170. The DSR method is employed with the LDALLR classifier, for which 50 PCA and 40 LDA feature vectors are chosen. The LR of the MixRes method in all experiments is calculated using (15). The parameters of the MixRes method are shown in Table 3. Larger numbers are chosen when MixRes is used in combination with MSBR. The experiments using FaceVACS are conducted with the eye-coordinates provided. Otherwise, eight images from dist3, 30 images from dist2 and 151 images from dist1 could not be processed successfully because FaceVACS could not detect the eyes in those images.

In Section 4.1, we assume a normal distribution for the facial features. We illustrate the appropriateness of this assumption using images from the FRGC database which were also used for training in Section 3. We choose two resolutions: 131 × 131 and 23 × 23 as HiRes and LoRes, respectively. We use 3464 images of the first 185 subjects to train the MixRes classifier. Then, we use the remaining 1007 images of 90 subjects for testing. We apply the transformation matrices $\boldsymbol{W}_{\rm H}$ and $\boldsymbol{U}_{*,1:D}$ on the HiRes testing images, and $\boldsymbol{W}_{\rm L}$ and $\boldsymbol{V}_{*,1:D}$ on the LoRes testing images. This results in a feature vector of dimensionality 40 for each image. We randomly select two HiRes feature elements $\boldsymbol{x}_{c,i}$ and $\boldsymbol{x}_{c,j}$ and corresponding LoRes feature elements $\boldsymbol{y}_{c,i}$ and $\boldsymbol{y}_{c,j}$ and plot the histograms $\boldsymbol{h}_{x,i}$, $\boldsymbol{h}_{x,j}$, $\boldsymbol{h}_{y,i}$ and $\boldsymbol{h}_{y,j}$ in Fig. 4. As we can see, all of the four seem to follow a normal distribution, which illustrates that our assumption for MixRes is reasonable.

In Section 5.1, we compare our methods with the four methods described above, following the protocol that was used to evaluate the CBD method in [17]. Thanks to the authors, who had provided us with the necessary details, we were able to duplicate their protocol, which will be used in this section. The images of the second longest distances from the SCface database are ds and used as probe in these experiments.

In Section 5.2, we further demonstrate the effectiveness of our methods on real LoRes data following a more realistic protocol. The experiments are more challenging because images from the SCface database captured at the greatest distance are used as probe and single mug-shot images are used as gallery. The CLPM method and FaceVACS are used for comparison.

In Section 5.3, we show results obtained on our own database to demonstrate that our methods are not optimised only for the SCface database.

In Table 4, we list the original resolutions of probe and gallery images and the resolutions actually used in each sections. In Section 5.1, the testing protocol defined in [17] describes that the input images are rescaled to a lower resolution. In Sections 5.2 and 5.3, we also ds the HiRes gallery images, but the LoRes probe images preserve their original resolution.

5.1 Comparison with the state-of-the-art

The experiments are conducted on the SCface database [35]. The SCface database contains images of 130 subjects taken by five SCs

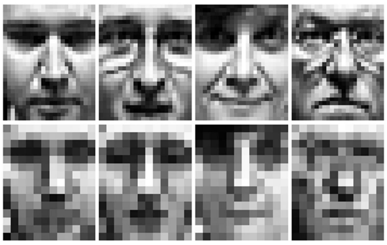

Fig. 5 *Sample images from the SCface database in the experiments in Section 5.1. First row: HiRes and second row: LoRes*

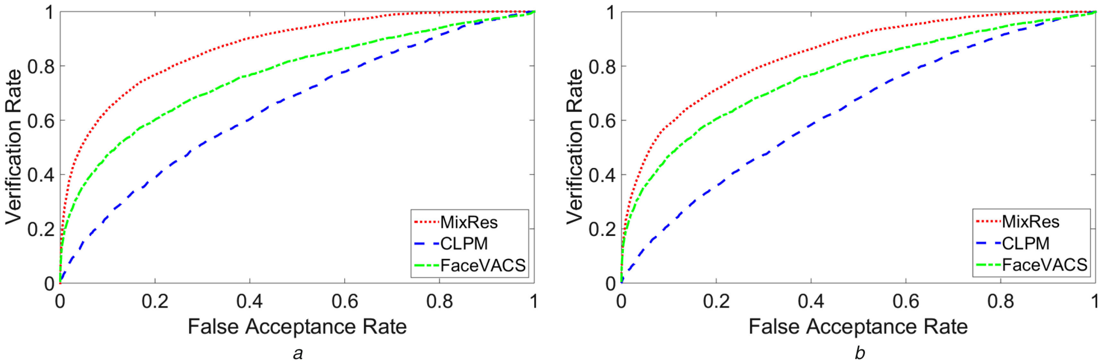

Fig. 6 *ROC curves for comparing MixRes to CLPM and FaceVACS **(a)** MG, **(b)** SG. Probe: dist2, 15×12 pixels. Gallery: dist3, 30×24 pixels*

Table 5 Comparison of MixRes to CBD, DSR, CLPM and FaceVACS

Setting	Method	AUC	Verification, %	Rank-1, %
MG	CBD	0.77 (0.03)	—	52.7 (9.9)
	DSR	0.69 (0.03)	29.3 (6.1)	31.7 (7.3)
	CLPM	0.64 (0.05)	23.9 (8.4)	9.0 (5.4)
	MixRes	**0.88 (0.03)**	**64.6 (8.1)**	**57.3 (9.5)**
	FaceVACS	0.76 (0.02)	47.0 (3.5)	19.3 (3.2)
SG	DSR	0.69 (0.03)	28.9 (6.0)	30.2 (8.8)
	CLPM	0.62 (0.05)	19.1 (7.8)	7.0 (5.1)
	MixRes	**0.84 (0.03)**	**57.4 (7.5)**	**47.9 (8.9)**
	FaceVACS	0.76 (0.02)	46.5 (1.6)	19.1 (1.5)

Values are in the format: average value (standard deviation). The VRs are obtained at FAR = 10%. Probe: dist2, 15 × 12 pixels. Gallery: dist3, 30 × 24 pixels. The bold values are the best results of MG or SG setting.

The bold values are the best results of MG or SG setting.

at three distances, namely 4.20 (dist1), 2.60 (dist2) and 1.00 m (dist3). There are 5 × 130 = 650 images for each distance. It also contains one frontal mug-shot image for each subject.

The protocol of [17] is duplicated so that we can compare the reported results of the CBD method directly. The region of interest is obtained by cropping the face region of the images based on the eye-coordinates provided in the database. Images from dist2 are used as LoRes and images from dist3 are used as (relatively) HiRes. The sizes of the HiRes and LoRes images after align and rescaling using bicubic interpolation are 30 × 24 pixels and 15 × 12 pixels, respectively. Histogram equalisation is used to normalise the illumination. Sample images are shown in Fig. 5. We randomly select 100 subjects and four images of these subjects for training. The remaining 30 subjects are used for testing, of which four images per subject from dist3 are randomly selected for the gallery and one image per subject from dist2 is used for the probe. The gallery and probe images of the same subject are taken by different cameras. Thus, we have 400 training images for both HiRes and LoRes, 120 gallery images and 30 probe images each time. This is the same setting as in [17]. In addition, we repeated the experiments using a single-gallery (SG) image per subject as this is a more realistic setting. Each experiment is repeated 100 times.

The area under the curve (AUC) and rank-1 identification rates using the CBD method are reported in [17]. We choose their best results and compare with the results from our experiments in Table 5. In addition, we provide VRs at FAR 10%. We also collect all the genuine and imposter scores from the 100 repetitions of each experiment to plot receiver operating characteristic (ROC) curves (Fig. 6).

As we can see, the MixRes method performs best of the five. The CBD method, which was also designed for real LoRes face recognition, performs better than the rest, but still worst than our MixRes method. DSR performs the second worst. Although CLPM was demonstrated to perform well on ds images (see Table 1), it gives the worst results in our real LoRes experiments. FaceVACS outperforms DSR and CLPM despite the fact that it was designed for HiRes face recognition. When a SG image per subject is used, all the methods decrease in performance, but the MixRes method remains the best.

To further improve the face recognition performance, MSBR and ET are employed. Rigid transformation and isotropic scaling are used for alignment in our experiments. The variation of each eye-coordinate for probe images is [−2, 2] pixels based on the manually marked eye-coordinates. The training procedure of ET uses 20 randomly selected (mis-)aligned images for each original image. The results are shown in Table 6 and Fig. 7.

As we can see, the verification results are significantly improved in all cases, especially when we combine MSBR and ET the improvement is more than 10%. The rank-1 recognition rate only improved slightly using ET and not in the other settings. It is due to the fact that closed-set recognition only picks the best score, whereas MSBR optimises all the similarity scores (imposter scores could benefit more from MSBR than genuine scores in some cases).

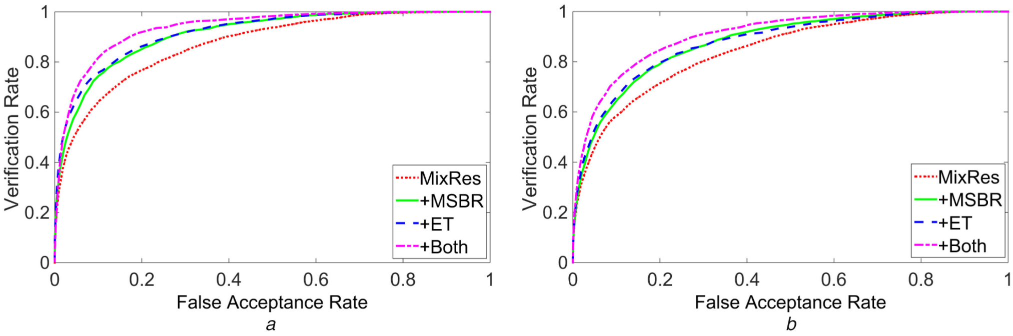

Fig. 7 *ROC curves of MixRes and its combination with MSBR and ET* ***(a)*** *MG,* ***(b)*** *SG. Probe: dist2, 15×12 pixels. Gallery: dist3, 30×24 pixels*

Table 6 MixRes combined with MSBR and ET

Settings	Verification, %	Rank-1, %
MG: MixRes	64.6 (8.1)	57.3 (9.5)
MG: + MSBR	74.5 (6.2)	55.6 (7.2)
MG: + ET	75.4 (6.5)	**58.2 (9.4)**
MG: + both	**81.4 (6.1)**	51.8 (7.6)
SG: MixRes	57.4 (7.5)	47.9 (8.9)
SG: + MSBR	65.2 (7.4)	46.7 (8.9)
SG: + ET	66.8 (7.4)	**50.3 (8.1)**
SG: + both	**72.5 (9.4)**	47.5 (9.1)

Values are in the format: average value (standard deviation). The VRs are obtained at FAR = 10%. MG: multi-gallery, SG: single-gallery. Probe: dist2, 15 × 12 pixels. Gallery: dist3, 30 × 24 pixels. The bold values are the best results of MG or SG setting.

The bold values are the best results of MG or SG setting.

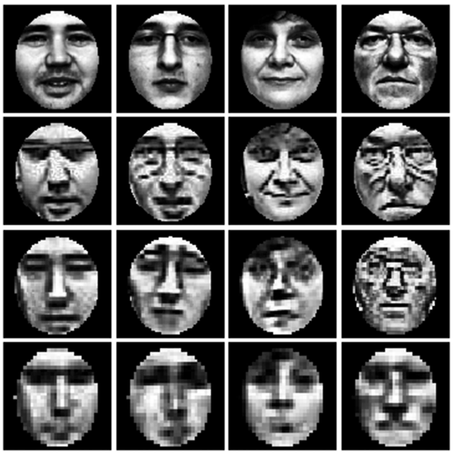

Fig. 8 *Sample images from the SCface database in the experiments in Section 5.2. First row: mug-shots, second row: dist3 and third row: dist2, last row: dist1*

5.2 Real LoRes experiments

In the experiment of the previous section, the gallery images were not separately captured mug-shots, but simply higher-resolution images from the sequence. A more realistic setting is to use separately captured mug-shots as a gallery.

We follow the guidelines below to make sure the experiments are representative of realistic applications:

i. Report verification results.
ii. Only use real LoRes images as probe.
iii. Gallery images should be HiRes, preferably mug-shots.
iv. Neither training on a gallery, nor on users represented in the gallery.
v. Use only one HiRes gallery image per subject.
vi. For small databases, use cross-validation in order to produce statistically more significant results. Use as many probe images as possible to reduce the standard deviation of the results.

The images are processed in such a way that they preserve their original resolution. The sizes of the cropped images for dist3, dist2 and dist1 are 80 × 80, 56 × 56 and 32 × 32, respectively (note that the cropping is different than in Section 5.1). The images are aligned using eye-coordinates provided in the database. An elliptic mask is applied to select the region of interest. Thus, the selected regions are smaller than the above resolutions. Histogram equalisation is used to normalise the illumination. Some samples of processed images are shown in Fig. 8. The original size of mug-shot face images is around 800 × 800 pixels. To make the comparison more efficient, mug-shot images are ds to the same resolution as dist3 images in our experiments. The probe images (dist1) are always of the same resolution as the original.

Dist1 (the largest distance) images are used as probes in all the following experiments. Different combinations of training and gallery sets are employed to test the performance of our methods in different situations. In all experiments, 100 subjects are randomly selected for training and the rest are for testing. The HiRes training images are of the same size as dist3 (80 × 80 pixels), and the LoRes training images are of the same size as dist1 (32 × 32 pixels). There are five images per subject from dist1, dist2 and dist3, and one image per subject from mug-shots. Our implementation of the MixRes method requires that the HiRes and LoRes training sets have the same number of images. Therefore, we replicate the mug-shot images so that each subject has five identical mug-shots.

Three different training settings are used. The first one is to have mug-shot images as HiRes training and dist1 images as LoRes training. Since there is only one mug-shot image per subject, we add images from dist3 or dist2 to the training set to obtain better training results in the other two settings. All the three training settings are used for MixRes and CLPM. Then, we select the best setting to conduct the combinations of MixRes, MSBR and ET.

All the mug-shots are used as gallery and all 650 images from dist1 are used as probe for FaceVACS, because it does not require training images. Thus, no standard deviation is presented for FaceVACS because no cross-validation was used. MSBR and ET are conducted in the same way as in Section 5.1. In addition, MSBR is also conducted with automatic initialisation (denoted as MSBRa). That is, instead of manually marked eye-coordinates, the starting points are determined by the region of interest detected by the Viola–Jones face detector [36]. The scale factor is set to 1.002 to detect the small face regions in the images. The false positives are then manually discarded. We compute VRs at FAR = 10% and rank-1 recognition rates, see Table 7.

MixRes significantly outperforms CLPM and FaceVACS in all three settings. FaceVACS performing the worst shows that this is a difficult experiment for state-of-the-art algorithms which are designed for HiRes images. The CLPM method still performs poorly as in the experiments in the previous section, while it had very good performance on ds data as shown in Table 1. This further supports our conclusion made in Section 3 that methods designed for LoRes face recognition should be evaluated using real LoRes data instead of images ds from HiRes.

A sufficient number of training images is required for MixRes to achieve good results. If we only use the mug-shot images as HiRes training and dist1 as LoRes training, the verification result at

Table 7 Results on the SCface database with different experiment settings

Principle	Train (HiRes)	Train (LoRes)	Verification. %	Rank-1, %
MixRes	mug-shot	dist1	47.8 (4.8)	34.9 (5.0)
	mug-shot + dist3	dist1 + dist2ds	63.1 (5.8)	43.4 (5.6)
	mug-shot + dist2	dist1 + dist2ds	67.2 (6.0)	48.3 (5.3)
CLPM	mug-shot	dist1	10.2 (2.7)	3.4 (1.7)
	mug-shot + dist3	dist1 + dist2ds	15.9 (3.5)	5.9 (2.0)
	mug-shot + dist2	dist1 + dist2ds	18.3 (3.6)	7.2 (2.2)
FaceVACS	—	—	21.7	2
Mix Res + MSBR	mug-shot + dist2	dist1 + dist2ds	**73.4 (5.8)**	**48.4 (5.2)**
Mix Res + ET	mug-shot + dist2	dist1 + dist2ds	61.0 (6.9)	39.0 (6.7)
Mix Res + both	mug-shot + dist2	dist1 + dist2ds	53.2 (7.2)	25.6 (4.5)
Mix Res + MSBRa	mug-shot + dist2	dist1 + dist2ds	72.0 (5.7)	48.0 (6.1)

Gallery: mug-shots, 80 × 80 pixels. Probe: dist1, 32 × 32 pixels. The values are in the format: average value (standard deviation). The VRs (%) are obtained at FAR = 10%. The abbreviation ds stands for down-sampled images.

Table 8 Results on our own database

	MixRes, %	+MSBRa, %	FaceVACS, %
verification	91	84	43
rank-1	80	64	20

Gallery: 1 m, 80 × 80 pixels. Probe: 8 m, 32 × 32 pixels. The VRs are obtained at FAR = 10%

FAR = 10% is <50%. When we enlarge the training sets by combining two sets of images, the VRs increase significantly up to 67.2%. The rank-1 recognition rates also show an improvement of around 10%. If we have more realistic data (mug-shot images and the corresponding LoRes images) to add in the training set, we would expect even more improvement. With the help of MSBR, MixRes can reach 73.4% VR. However, ET does not improve the results as it did in Section 5.1. A possible explanation is that the added training data may not be fully representative of the testing data, causing the results to become worst in this experiment. This could also be the reason that MixRes in combination with both MSBR and ET has worst performance. It is promising that the results from MSBR with automatic initialisation (MSBRa) show only marginal degradation as compared with the results from MSBR with initialisation based on manual eye-coordinates. This means that the LoRes face recognition system can be fully automatic if the detection is correct.

5.3 Evaluation using other data

To demonstrate that our methods are not optimised only for the SCface database, we conduct another experiment using SCface images for training, but using our own database, which is described in Section 3, for testing. The images taken at a distance of 8 m are selected as probe because their original resolution is similar to that of images at dist1 from the SCface database. The gallery images are the ones that were taken at a distance of 1 m in a separate session (same as in Section 3). The training setting that provided the best results from the previous SCface experiments is selected, that is, mug-shots and dist2 images are combined for HiRes training while dist1 and ds dist2 images are used for LoRes training. MSBR is conducted with automatic initialisation. FaceVACS is used with manually marked eye-coordinates. The results are shown in Table 8. Since this is an easier experiment on a small database with only 25 subjects and controlled environment, MixRes gives very good results. The fully automatic method + MSBRa performs close to MixRes with manually marked eye-coordinates. On the other hand, FaceVACS performs much poorer than MixRes (about a 50% difference in VR at FAR 10%).

6 Conclusion

In a common case of low-resolution face recognition, recognition of suspects from a long distance, there is usually a resolution mismatch between low-resolution probe and high-resolution gallery. Most existing methods, which are designed to match high-resolution images, cannot handle low-resolution probes well. In this paper, we present a classifier, mixed-resolution biometric comparison, specifically designed for heterogeneous cases, which allows direct comparison of different resolution images. We identify that proper alignment is one of the major challenges in low-resolution face recognition. We also show that there is a large difference in face recognition performance between ds and real low-resolution images and only the results from real low-resolution can be trusted. To cope with the alignment problem, we investigate two methods, matching-score-based registration and extended training. Extended training is less effective, but it helps when there is not enough training data. Matching-score-based registration is useful in combination with our mixed-resolution classifier. It can also be initialised using a region obtained from face detection, which results in a fully automatic low-resolution face recognition method. Our experiments using real low-resolution data show that our methods outperform the state-of-the-art. In our experiment on the SCface database, the combination of our mixed-resolution classifier with matching-score-based registration and extended training outperform the state-of-the-art by 20% of VR at FAR 10% in the multi-gallery (MG) setting.

7 Acknowledgments

We thank the Cognitec Systems GmbH for supporting our research by providing the FaceVACS software. Results obtained for FaceVACS were produced in experiments conducted by the University of Twente, and should therefore not be construed as a vendor's maximum effort full capability results. We would also thank Moutafis and Kakadiaris [17] for providing us with the necessary details to duplicate their protocol and Li *et al.* [13] for availing software.

8 References

[1] Zhang, D., He, J., Du, M.: 'Morphable model space based face super-resolution reconstruction and recognition', *Image Vis. Comput.*, 2012, **30**, pp. 100–108

[2] Zou, W., Yuen, P.: 'Very low resolution face recognition problem'. 2010 Fourth IEEE Int. Conf. on Biometrics: Theory Applications and Systems (BTAS), 2010, pp. 1–6

[3] Biswas, S., Bowyer, K., Flynn, P.: 'Multidimensional scaling for matching low-resolution face images', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2011, **34**, (10), pp. 2019–2030

[4] Lei, Z., Liao, S., Jain, A.*, et al.*: 'Coupled discriminant analysis for heterogeneous face recognition', *IEEE Trans. Inf. Forensics Sec.*, 2012, **7**, (6), pp. 1707–1716

[5] Ren, C., Dai, D., Yan, H.: 'Coupled kernel embedding for low-resolution face image recognition', *IEEE Trans. Image Process.*, 2012, **21**, (8), pp. 3770–3783

[6] Zou, W., Yuen, P.: 'Very low resolution face recognition problem', *IEEE Trans. Image Process.*, 2012, **21**, (1), pp. 327–340

[7] Peng, Y., Spreeuwers, L.J., Veldhuis, R.N.J.: 'Likelihood ratio based mixed resolution facial comparison'. Third Int. Workshop on Biometrics and Forensics (IWBF2015), Gjøvik, Norway, March 2015, pp. 1–5

[8] Bilgazyev, E., Efraty, B., Shah, S.*, et al.*: 'Improved face recognition using super-resolution'. 2011 Int. Joint Conf. on Biometrics (IJCB), 2011 Int. Joint Conf. on, October 2011, pp. 1–7

[9] Zhou, C., Zhang, Z., Yi, D.*, et al.*: 'Low-resolution face recognition via simultaneous discriminant analysis'. 2011 Int. Joint Conf. on Biometrics (IJCB), October 2011, pp. 1–6

[10] Hennings-Yeomans, P., Kumar, B., Baker, S.: 'Robust low-resolution face identification and verification using high resolution features'. 2009 16th IEEE Int. Conf. on Image Processing (ICIP), 2009, pp. 33–36

[11] Pong, K.-H., Lam, K.-M.: 'Multi-resolution feature fusion for face recognition', *Pattern Recognit.*, 2014, **47**, (2), pp. 556–567

[12] Huang, H., He, H.: 'Super-resolution method for face recognition using nonlinear mappings on coherent features', *IEEE Trans. Neural Netw.*, 2011, **22**, (1), pp. 121–130

[13] Li, B., Chang, H., Shan, S., *et al.*: 'Low-resolution face recognition via coupled locality preserving mappings', *IEEE Signal Process. Lett.*, 2010, **17**, (1), pp. 20–23

[14] Jiang, J., Hu, R., Han, Z., *et al.*: 'Graph discriminant analysis on multi-manifold (GDAMM): a novel super resolution method for face recognition'. 2012 19th IEEE Int. Conf. on Image Processing (ICIP), September 2012, pp. 1465–1468

[15] Gunturk, B., Batur, A., Altunbasak, Y., *et al.*: 'Eigenface-domain super-resolution for face recognition', *IEEE Trans. Image Process.*, 2003, **12**, (5), pp. 597–606

[16] Wang, X., Tang, X.: 'Hallucinating face by eigen transformation', *IEEE Trans. Syst. Man Cybern., Appl. Rev.*, 2005, **35**, (3), pp. 425–434

[17] Moutafis, P., Kakadiaris, I.A.: 'Semi-coupled basis and distance metric learning for cross-domain matching: application to low-resolution face recognition'. IEEE Int. Joint Conf. on Biometrics, Clearwater, FL, 2014, pp. 1–8

[18] Baker, S., Kanade, T.: 'Hallucinating faces'. . Proc. Fourth IEEE Int. Conf. on Automatic Face and Gesture Recognition, 2000, 2000, pp. 83–88

[19] Hennings-Yeomans, P., Baker, S., Kumar, B.: 'Recognition of low-resolution faces using multiple still images and multiple cameras'. Second IEEE Int. Conf. on Biometrics: Theory, Applications and Systems, 2008 BTAS 2008, 2008, pp. 1–6

[20] Hu, S., Maschal, R., Young, S.S., *et al.*: 'Face recognition performance with super resolution', *Appl. Opt.*, 2012, **51**, (18), pp. 4250–4259

[21] Xu, X., Liu, W., Li, L.: 'Face hallucination: how much it can improve face recognition'. 2013 3rd Australian Control Conf. (AUCC), November 2013, pp. 93–98

[22] Phillips, P., Flynn, P., Scruggs, T., *et al.*: 'Overview of the face recognition grand challenge'. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, 2005 CVPR 2005, June 2005, vol. **1**, pp. 947–954

[23] Turk, M., Pentland, A.: 'Face recognition using eigen faces'. Proc. 1991 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, June 1991, pp. 586–591

[24] Veldhuis, R.N.J., Bazen, A.M.: 'One-to-template and one-to-one verification in the single- and multi-user case'. 26th Symp. on Information Theory in the Benelux, Brussels, Belgium, Brussels, May 2005, pp. 39–46

[25] Ahonen, T., Hadid, A., Pietikainen, M.: 'Face description with local binary patterns: application to face recognition', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2006, **28**, (12), pp. 2037–2041

[26] Belhumeur, P., Hespanha, J., Kriegman, D.: 'Eigenfaces vs. Fisherfaces: recognition using class specific linear projection', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1997, **19**, (7), pp. 711–720

[27] Chen, D., Cao, X., Wang, L., *et al.*: 'Bayesian face revisited: a joint formulation'. Computer Vision – ECCV 2012, 2012 (LNCS, **7574**), pp. 566–579

[28] Bazen, A., Veldhuis, R.: 'Likelihood-ratio-based biometric verification', *IEEE Trans. Circuits Syst. Video Technol.*, 2004, **14**, (1), pp. 86–94

[29] Vera-Rodriguez, R., Tome, P., Fierrez, J., *et al.*: 'Comparative analysis of the variability of facial landmarks for forensics using CCTV images'. Proc. Image and Video Technology: Sixth Pacific-Rim Symp., PSIVT 2013, October 28–November 1 2013, Guanajuato, Mexico, 2014, pp. 409–418

[30] Min, J., Bowyer, K., Flynn, P.: 'Eye perturbation approach for robust recognition of inaccurately aligned faces'. Audio- and Video-Based Biometric Person Authentication, 2005 (LNCS, **3546**), pp. 41–50

[31] Boom, B.J., Spreeuwers, L.J., Veldhuis, R.N.J.: 'Automatic face alignment by maximizing similarity score'. Proc. of the Seventh Int. Workshop on Pattern Recognition in Information Systems, Biosignals, Madeira, Portugal, June 2007, pp. 221–230

[32] Spreeuwers, L.J., Boom, B.J., Veldhuis, R.N.J.: 'Better than best: matching score based face registration'. Proc. of the 28th Symp. on Information Theory in the Benelux, Enschede, May 2007, pp. 125–132

[33] Rowley, H.A., Baluja, S., Kanade, T.: 'Human face detection in visual scenes'. Advances in Neural Information Processing Systems, 1996, vol. **8**, pp. 875–881

[34] Cognitec Systems GmbH. FaceVACS-SDK Version 8.7.0.2, 2015

[35] Grgic, M., Delac, K., Grgic, S.: 'SCface – surveillance cameras face database', *Multimedia Tools Appl.*, 2011, **51**, pp. 863–879

[36] Viola, P., Jones, M.: 'Rapid object detection using a boosted cascade of simple features'. Proc. of the 2001 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, 2001 CVPR 2001, 2001, vol. **1**, pp. I–511–I–518