IET Biometrics

Research Article

IET Journals
The Institution of Engineering and Technology

Low-resolution face recognition and the importance of proper alignment

ISSN 2047-4938
Received on 24th January 2018
Revised 20th November 2018
Accepted on 29th January 2019
E-First on 20th February 2019
doi: 10.1049/iet-bmt.2018.5008
www.ietdl.org

Yuxi Peng[1] ✉, *Luuk J. Spreeuwers*[1], *Raymond N.J. Veldhuis*[1]

[1]*Faculty of Electrical Engineering, Mathematics and Computer Science, SCS Research Group, University of Twente, Enschede, The Netherlands*

✉ *E-mail: y.peng@utwente.nl*

Abstract: Face recognition methods for low resolution are often developed and tested on down-sampled images instead of on real low-resolution images. Although there is a growing awareness that down-sampled and real low-resolution images are different, few efforts have been made to analyse the differences in recognition performance. Here, the authors explore the differences and demonstrate that alignment is a major cause, especially in the absence of pose and illumination variations. The authors found that the recognition performances on down-sampled images are flattered mostly due to the fact that the images are perfectly aligned before down-sampling using high-resolution landmarks, while the real low-resolution images have much poorer alignment. To obtain better alignment for real low-resolution images, the authors apply matching score-based registration which does not rely on accurate landmarks. The authors propose to divide low resolution into three ranges to harmonise the terminology: upper low resolution (ULR), moderately low resolution (MLR), and very low resolution (VLR). Most face recognition methods perform well on ULR. MLR is a challenge for commercial systems, but a low-resolution deep-learning method can handle it very well. The performance of most methods degrades significantly for VLR, except for simple holistic methods which perform the best.

1 Introduction

Face recognition is one of the most popular biometric modalities [1, 2]. Comparing to fingerprint and iris, facial images can be captured at a distance and thus, they are natural and non-intrusive. Face recognition for high-resolution facial images has achieved great success in the past decades. However, common applications for face recognition such as surveillance cannot always capture facial images at a close range. The images captured at a distance suffer from various problems which make them more difficult to recognise than those captured at close range. One of the problems is that the images are of low resolution, so that they contain less information. Another problem is that the images are usually captured in uncontrolled situations, which results in illumination and pose variations. In addition, the images are often noisy because of low light and suffering from compression artefacts. Thus, low-resolution face recognition is still challenging and receiving substantial attention nowadays.

One approach to improve low-resolution face recognition is by using super-resolution. Hennings-Yeomans *et al.* [3] proposed an algorithm that combines the underlying assumptions of super-resolution methods with subspace distance metrics used for classification. Zhang *et al.* [4] proposed a super-resolution method in the morphable model space, which provides high-resolution information required for both reconstruction and recognition. Zou and Yuen [5] developed a data constraint for reconstructing super-resolution image features so that both the distances between the reconstructed images and the corresponding high-resolution images and the distances between super-resolution images from the same class are minimised.

Researchers also developed face recognition methods that perform face recognition directly on low-resolution images. Li *et al.* [6] proposed a method that projects both high-resolution gallery and low-resolution probe to a common feature space for classification using coupled mappings which minimise the difference between corresponding images. Moutafis and Kakadiaris [7] proposed a method that learns semi-coupled mappings for optimised representations. The mappings aim at increasing class-separation for high-resolution images and mapping low-resolution images to their corresponding class-separated high-resolution data. Peng *et al.* [8] proposed a likelihood ratio-based method for direct comparison between images of different resolutions.

Owing to the lack of appropriate data sets of real-life surveillance situations, most of the research is conducted on down-sampled images, for example [3, 4, 6]. Researches like [5, 9, 10] include experiments on real low-resolution images, but most of their experiments and analysis are conducted using down-sampled images. It is shown in [11, 12] that face recognition and super-resolution methods both perform much better on down-sampled images than on real low-resolution images. In [12], it was mentioned that alignment is important for the performance differences, but the experiments were conducted on a small data set and only simple basic face recognition methods are used.

In this paper, we analyse the performance differences between down-sampled and real low-resolution facial images in more depth using various face recognition methods. We demonstrate that in the absence of pose and illumination variations, the loss of recognition performance for real low-resolution images compared to down-sampled images can largely be attributed to poor alignment (or registration). When images are captured at uncontrolled situations (with illumination and pose variations), alignment still plays an important role in recognition performance. Proper alignment is a problem for low-resolution face recognition, because the landmarks, such as eye-coordinates, that are used for high-resolution face recognition are much less reliable or even unavailable. We particularly demonstrate that: (i) if down-sampled images are aligned based on landmarks estimated after down-sampling instead of before, which is common but unrealistic, then the recognition performance becomes similar/closer to that obtained with real low-resolution images. (ii) Conversely, if real low-resolution images are better registered by means of matching score-based registration, then the recognition performance increases towards that obtained with down-sampled images.

In the literature on low-resolution face recognition, what in some papers is considered as low resolution, is still considered as high resolution in other papers. For example, the interpupillary distances (IPD) of low-resolution images described in [9] are from 2 pixels to 8 pixels, and the IPD of high-resolution images is 16

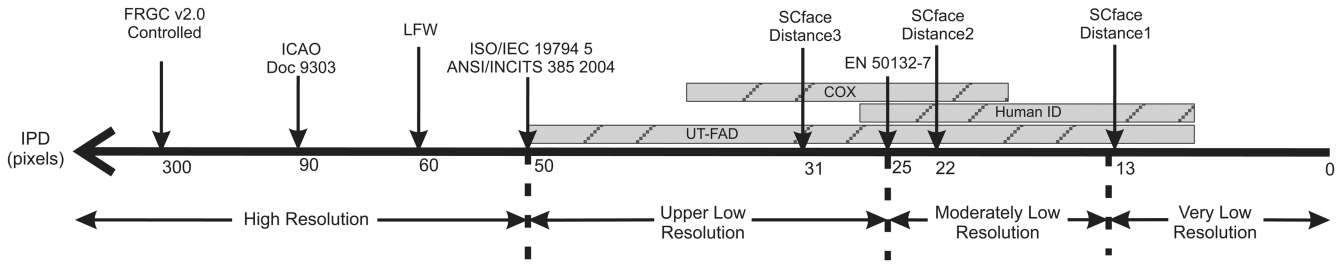

Fig. 1 *Resolution scale*

Fig. 2 *Down-sampled and real low-resolution image examples*
(a) High-resolution, IPD 96 pixels, ***(b)*** Down-sampled, IPD 10 pixels, ***(c)*** Real, IPD 10 pixels, ***(d)*** Real with pose, IPD 12 pixels (subject from the Human ID data set)

pixels. While in [13], the IPD of low-resolution images is 20 pixels. To harmonise the terminology in low-resolution face recognition, we propose the resolution scale as shown in Fig. 1. We use IPD as the measure of resolution.

There are four biometric standards in this graph. Two of them are ISO/IEC 19794-5:2005 [14] and ANSI/INCITS 385-2004 [15]. They describe an example of proper face position in an image [16], where the size of the face is $\sim 106 \times 96$ pixels and the corresponding IPD is ∼50 pixels. We use this point to separate high resolution and low resolution. Another standard, the European norm EN 50132-7 [17], describes the recommended minimum size of the object for the CCTV system. We calculate the IPD, which is ∼25 pixels according to the description. ICAO Doc 9303 [18] mentions that standardised size portrait results in a facial image with IPD ∼90 pixels. We divide low resolution into upper low resolution (ULR), moderately low resolution (MLR), and very low resolution (VLR), and we use IPD 25 pixels (EN 50132-7) and 13 pixels as the separating points. ULR is a relatively high resolution and most existing face recognition methods including commercial systems that are designed for high-resolution face recognition can handle it well. Images of MLR are harder to recognise than ULR. Methods that are designed for low-resolution face recognition start to show their advantages in this range. VLR is extremely difficult for face recognition and we expect poor results for most methods. In addition, we added in the graph the resolution of several popular face data sets: FRGC v2.0 [19], LFW [20], and SCface [21], and the data sets used in our experiments: UT-FAD [22], Human ID [23], and COX [24]. The latter three contain data over a range of resolutions and are presented as bars in Fig. 1. We focus on low resolution, so we divide it into three ranges, but high resolution can also be further divided into subranges.

The body of this paper is divided into two parts. The first part analyses and demonstrates that alignment is an important factor which causes the loss in recognition performance of real low-resolution images. The second part is about improving face recognition performance by compensating alignment problem on real low-resolution images using matching score-based registration.

2 Difference between down-sampled and real low-resolution images

2.1 Analysis

We have mentioned in the previous section that down-sampled images are commonly used as a substitute for real low-resolution images, but face recognition performance on down-sampled images is always much better than on real low-resolution images [11]. In this section, we will analyse the differences between down-sampled and real low-resolution facial images.

Firstly, we compare down-sampled images with real low-resolution images from the visual aspect. In Fig. 2, we show some example images. Fig. 2*a* is a high-resolution facial image, Fig. 2*b* is obtained by down-sampling Fig. 2*a*, and Fig. 2*c* is a real low-resolution image of the same subject captured at a distance. Although Figs. 2b and *c* are of the same resolution, the down-sampled image seems to have much better quality than the real low-resolution one. Firstly, the down-sampled image is much sharper. We can still see a lot of details from the down-sampled image, for example, the eyebrows are very clear, the white of eyes, and the colour of the lips are visible. The real low-resolution image is much more blurred: the eyes become two blurred dark dots, the lips become grey, and the nose and eyebrows are almost invisible. Secondly, the real low-resolution image contains more noise than the down-sampled image, which mostly can be seen on the cheeks. Despite the above differences we can see from the pictures, there are other factors that could play a role. For example, the perspective is different between images taken at different distances; there is more air between the subject and the camera in large distance which may affect the shape of the image [25], lens distortion [26] etc. Pose and illumination are well-known factors that have a large influence on face recognition for both high-resolution and low-resolution facial images. In low-resolution face recognition, for example, surveillance, the images are more likely to be captured under uncontrolled situations, thus pose and illumination can be a serious problem (for example, in Fig. 2*d*).

Secondly, we compare the differences in processing real and down-sampled low-resolution images for face recognition. The face recognition processing flow graphs are presented in Fig. 3. These flow graphs are derived from [2]. Fig. 3*a* is the recognition process for real low-resolution images. Fig. 3*b* is the procedure of how down-sampled images are normally processed in order to test low-resolution face recognition methods. We can see that a significant difference between the two is where the landmarks are detected. For real low-resolution images, the landmarks are detected on the low-resolution images which are very likely to be inaccurate. While for down-sampled images, the landmarks are detected on the high-resolution images which are usually very accurate. This could be essential because inaccurate landmarks will result in poor alignment and subsequently poor recognition results. In this case, Fig. 3*c* is a more proper way of processing down-sampled images because it is closer to real low-resolution face recognition.

We choose to focus on the recognition process aspect and propose the following hypothesis: *in the absence of pose and illumination variations, alignment is the most important cause of the performance differences in face recognition between down-sampled and real low-resolution images*.

2.2 Controlled situation

In this subsection, we set up experiments to test our hypothesis that alignment is the most important cause of the recognition performance differences between down-sampled and real low-resolution images.

To test the hypothesis, we first need both down-sampled images and real low-resolution images captured at the same moment of the corresponding high-resolution images: the former ones can be achieved by down-sampling the high-resolution images and the latter ones should be captured at a farther distance. It is preferred that images captured at different distances are available, so that we can see if the results are valid for all distances. Ideally, the images

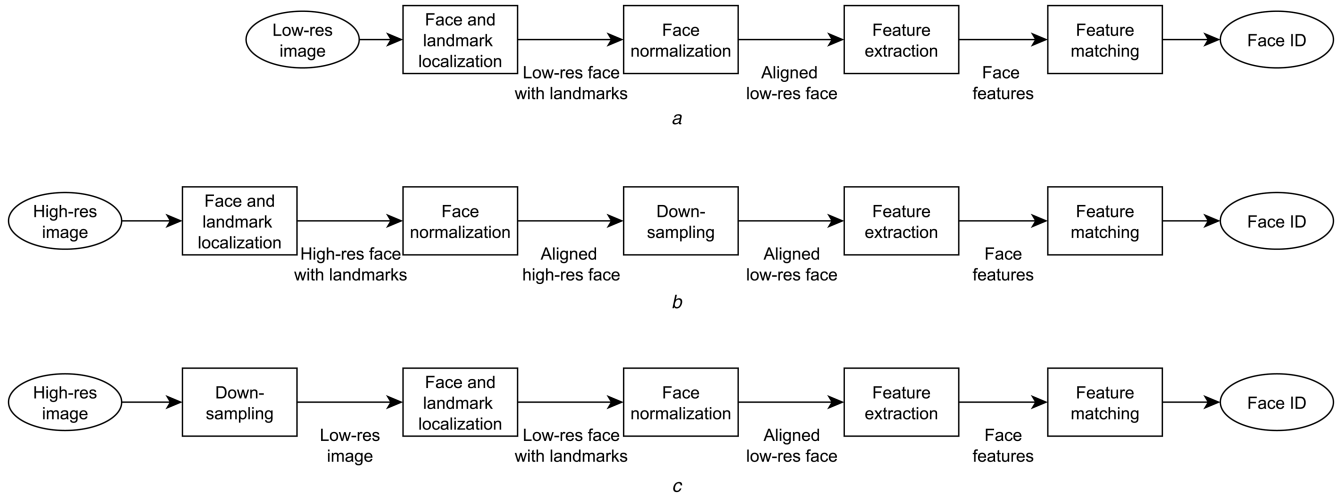

Fig. 3 *Face recognition processing flow of*
(a) Real low-resolution facial images, ***(b)*** Down-sampled facial images (commonly), ***(c)*** Down-sampled images (proposed)

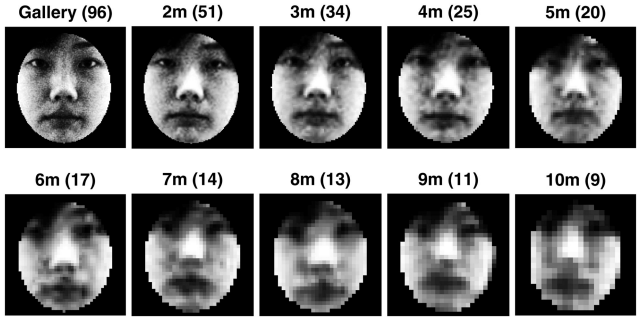

Fig. 4 *Sample images from the UT-FAD data set (with IPD in the brackets)*

Table 1 Four state-of-the-art low-resolution face recognition methods

Method	Testing data set	Rank-1 rate, %
RIDN [29]	SCface	74
MixRes [8]	SCface	48
CLPM [6]	FERET	90
DSR [30]	SCface	22

should be frontal with uniform illumination because we only want to test the influence of alignment.

The University of Twente-Faces At Distances (UT-FAD) data set [22] is used in our experiment. Although it is a small data set of only 22 subjects, it is the only data set available that meets the above requirements. Besides, noise is allowed in the experimental results as long as the trend is clear. The images of the UT-FAD data set were captured using a commercial camera CASIO EX-FC100. To minimise the changes of the condition on the subjects' face, images were recorded in the following way: subjects remained a fixed position and the photos were taken at different distances. The first photo of each subject was taken at a distance of 2 m, and then we moved the camera 1 m away and took a photo each time until we had nine photos of this subject. Thus, the photos were taken at nine distances from 2 to 10 m. Those images are used as the probe in the following experiments. The faces are always frontal and have the same illumination. In addition, a gallery image of each subject was captured at a distance of 1 m after 2 weeks using the same camera with the same illumination and frontal pose.

The size of the cropped face regions of gallery images is 243×243 pixels with 96 pixels between the eyes. The sizes of the probes from distance 2 to 10 m are 131×131, 87×87, 64×64, 51×51, 44×44, 36×36, 33×33, 28×28, and 23×23 pixels, respectively. The corresponding average IPDs are 51, 34, 25, 20, 17, 14, 13, 11, and 9 pixels, respectively. The first three resolutions are ULR, the middle four resolutions are MLR, and the last two resolutions are VLR according to our divisions of low resolutions. The facial images are aligned using manually annotated eye-coordinates. An ellipse mask is added to define the region of interest. Sample images are shown in Fig. 4.

We use eight different face recognition methods. Two basic face classifiers, principal component analysis (PCA) [27] and local binary patterns (LBP) [28], are used. We also use four state-of-the-art low-resolution face recognition methods, namely RIDN [29], MixRes [8], CLPM [6], and DSR [30]. In Table 1, we present the published rank-1 recognition rates of the four methods. Note that the methods are tested under different protocols (even when using the same testing data set), and therefore, the performance of the methods cannot be compared directly using these numbers. RIDN, MixRes, and CLPM are specially designed for comparing low-resolution probes with high-resolution galleries. RIDN is based on a deep neural network and developed for low-resolution face recognition. DSR is a super-resolution method which reconstructs high-resolution features of the probes. In our experiments, we used the original source code of the RIDN, MixRes, and CLPM methods and therefore, we are confident about the correctness of the implementations. We used our own implementation of the DSR method. We repeated the experiment described in [30] on SCface data to verify our implementation and our implementation had a 24% rank 1 recognition rate, which was slightly better than the result in [30]. We use DSR in combination with PCA for recognition, which was also used in the original paper [30]. PCA, MixRes, CLPM, and DSR + PCA are trained using 4064 high-quality images of 254 subjects from the FRGC v2.0 data set [19]. Those images are aligned at high resolution using the eye-coordinates provided by the data set. Distance measures employed for PCA and LBP are L1 norm and χ^2, respectively. In addition, two commercial face recognition methods are used. We call them system A and system B.

The eight methods are designed for different situations and thus have different requirements for the image resolutions. PCA and LBP require probe and gallery images have the same resolution, so the training and gallery images are down-sampled to the same resolution as the probe images. RIDN is a pre-trained system and all the testing images have to be resized to 55×60 pixels. MixRes, CLPM, and DSR are capable of comparing gallery and probe with different resolutions, so gallery images of their original size are used. The commercial systems should be able to cope with different resolutions, thus the images of original size are used.

Five of the methods, PCA, LBP, MixRes, CLPM, and DSR + PCA, have parameters that can be changed. We choose the following parameters as in Table 2 to ensure a good performance of all the face recognition methods. System A is provided with manual landmarks. Its own landmark detection could not function well for low-resolution images, for example, it could not generate features for 25% of the images at IPD 13 pixels if no eye-coordinates were provided. However, system B has to use its own

detecting function because it does not allow manual input of landmarks.

Three experimental settings are conducted. A brief description is in Table 3 and detailed explanations are in the following.

i. *REAL*: REAL low-resolution images captured at different distances are used as probe, those are, the images captured at 2–10 m in the UT-FAD data set. The images are aligned using manually marked eye-coordinates.
ii. *ADS*: Align before down-sample. The images of the highest resolution of 'REAL' are down-sampled to the same resolution as the ones captured at other distances. In the UT-FAD data set, the images from 2 m (IPD 51 pixels) are down-sampled to the same resolution as images from 3 m (IPD 34 pixels) to 10 m (IPD 9 pixels). The images are aligned before down-sampling using eye-coordinates manually marked on the images of the highest resolution.
iii. *DSA*: Down-sample then align. The images of the highest resolution of 'REAL' are down-sampled to the same resolution as the images captured at other distances. The images are aligned after down-sampling process using eye-coordinates marked on the down-sampled images.

'ADS' is not applicable for system B because system B does not allow manual input of eye-coordinates.

We provide verification results obtained from our experiments. Verification means to compare two images and verify if they are from the same subject. It must be remarked in advance that the verification performance for low-resolution images is poorer than on high-resolution images. This is also illustrated in [8] where the performance of the state-of-the-art low-resolution face recognition is presented. For that reason, verification rates (or genuine acceptance rate) are presented at false acceptance rate (FAR) 0.1 rather than the more common FAR equals to 0.01 or 0.001. Also note that the results of the first two resolutions for CLPM and DSR + PCA are not shown because, to be able to perform correctly, they require the number of training images to be larger than the dimensionality of the low-resolution images in the training set.

The verification results on the UT-FAD data set are shown in Fig. 5.

Although the UT-FAD data set is small, there is a clear trend over most of the methods shown in Fig. 5. Firstly, the recognition performance on pre-aligned down-sampled images ('ADS') is the best and there is almost no performance degradation when resolution changes for the basic recognition methods. Secondly, the results of 'REAL' and 'DSA' settings are strikingly alike for all the eight methods. The results of 'REAL' and 'DSA' settings both decrease when the resolution decreases, but they remain similar at the same resolution. The above two points support our hypothesis that the poorer alignment of low-resolution images is the key factor for the performance degradation. The CLPM method behaves 'strange' in that its performance does not drop when image resolution decrease and it even goes up for 'ADS', because it is only designed for low-resolution face recognition and could not perform well on high-resolution images. In [6], it was only tested using images of IPD 6 pixels. However, its results still follow the above-described trend.

While the above hypothesis holds for seven methods, system A is an outlier. Its results of 'ADS' are similar to those of 'REAL' and 'DSA'. Apparently, the resolution changes are more dominant than the alignment issue for system A. As we do not have internal knowledge about this commercial system, we cannot explain why it performs so differently than other methods. We guess that it might also include alignment compensation algorithms.

The results also show the different capabilities for low-resolution face recognition between the methods. PCA, LBP, and MixRes have similar verification rates and most results become worse when resolution decreases (MLR and VLR), but their verification rates are all >0.6 at FAR 0.1. RIDN significantly outperforms other methods at higher resolutions (ULR and MLR), but the results at the lowest resolution are not better than the three former mentioned methods. CLPM does not perform well for ULR and the results for VLR are also worse than most of the other methods. DSR + PCA have a reasonable performance over all the resolutions; however, DSR is actually making the performance worse when compared with the results of PCA alone. Systems A and B are much more sensitive to resolution changes. System A maintains good performance for images of ULR, but its performance drops significantly when the resolution decreases. For the images of VLR, it completely collapses. It is an even harder task for system B: without the aid of manual eye-coordinates, the performance of system B decreases much faster than that of system A.

2.3 Uncontrolled situation

In the previous subsection, we show support for our hypothesis that alignment is the most important cause for the face recognition performance difference between down-sampled and real low-resolution images, in the absence of pose and illumination variations. The previous experiments were conducted on a data set that was captured under strictly controlled conditions. Now, we are going to use images which are captured at uncontrolled situations to explore whether alignment is still important in combination with pose and illumination changes.

The Human ID data set [23] is chosen for our experiment. It has parallel gait videos that were captured when the subjects were walking towards the camera. We obtain facial images captured at different distances from those videos. The images are near-frontal. There were maximum four sessions for each subject and the intervals between sessions were >7 days. It is captured in less controlled conditions. The poses of the captured images were mostly frontal but could change when the subjects were getting close and turning away from the camera. The viewing angle was not consistent during the recording of the videos because the camera was placed lower than the subjects' face. The illumination was affected by sunlight, thus could vary within a video or between different videos.

We use the Viola–Jones face detector (from MATLAB function) [31] to detect the faces in the videos. From the detected faces, we choose images suitable for our experiments. The facial images with nine different resolution are selected: 70×70, 60×60, 50×50, 45×45, 40×40, 35×35, 30×30, 25×25, and 23×23 pixels. The corresponding average IPDs are 27, 23, 20, 18, 16, 14, 12, 10, and 9 pixels, respectively. For each resolution, two images are randomly selected from each video. The number of images for each resolution is different because some of the videos do not have images of all the nine resolutions. Images of the highest resolution (IPD 27 pixels) are used as gallery and images of other resolutions are used as the probe. Compared with the previous experiments on the UT-FAD data set, the probe images are of lower resolutions that are only in the range of MLR and VLR. Detailed information about the data are shown in Table 4.

All the images are aligned using manually marked eye-coordinates. The regions of interest are defined by an elliptic mask. Sample images are shown in Fig. 6.

The eight face recognition methods from the previous experiments are also used. The configurations of the methods are similar to the previous experiments, only that the number of regions the images are divided into for LBP is 6×6.

The three probe settings corresponding to previous experiments are also applied on this data set as follows:

i. *REAL*: the images of IPD 23 to 9 pixels in the Human ID data set. The images are aligned using manually marked eye-coordinates.
ii. *ADS*: the images of IPD 23 pixels are down-sampled to IPD 20 to 9 pixels. The images are aligned before down-sampling.
iii. *DSA*: the images of IPD 23 pixels are down-sampled to IPD 20 to 9 pixels, but they are aligned using eye-coordinates marked on the down-sampled images.

The verification results are shown in Fig. 7. Note that the images are captured at uncontrolled situation, so that illumination and pose variations are present in the experiments. The resolution of the

Table 2 Parameter settings

Method	Parameter	Number
PCA	number of eigenvectors	50
LBP	number of divided regions for IPD 51 to 14 pixels	7×7
	number of divided regions for IPD 13 to 9 pixels	5×5
MixRes	first dimension reduction high-res feature vectors	70
	first dimension reduction low-res feature vectors	60
	feature vectors after second dimension reduction	40
CLPM	number of feature vectors	100
DSR + PCA	number of eigenvectors	60

Table 3 Three probe settings

Setting	Description
REAL	real low-resolution images
ADS	first align then down-sample
DSA	first down-sample then align

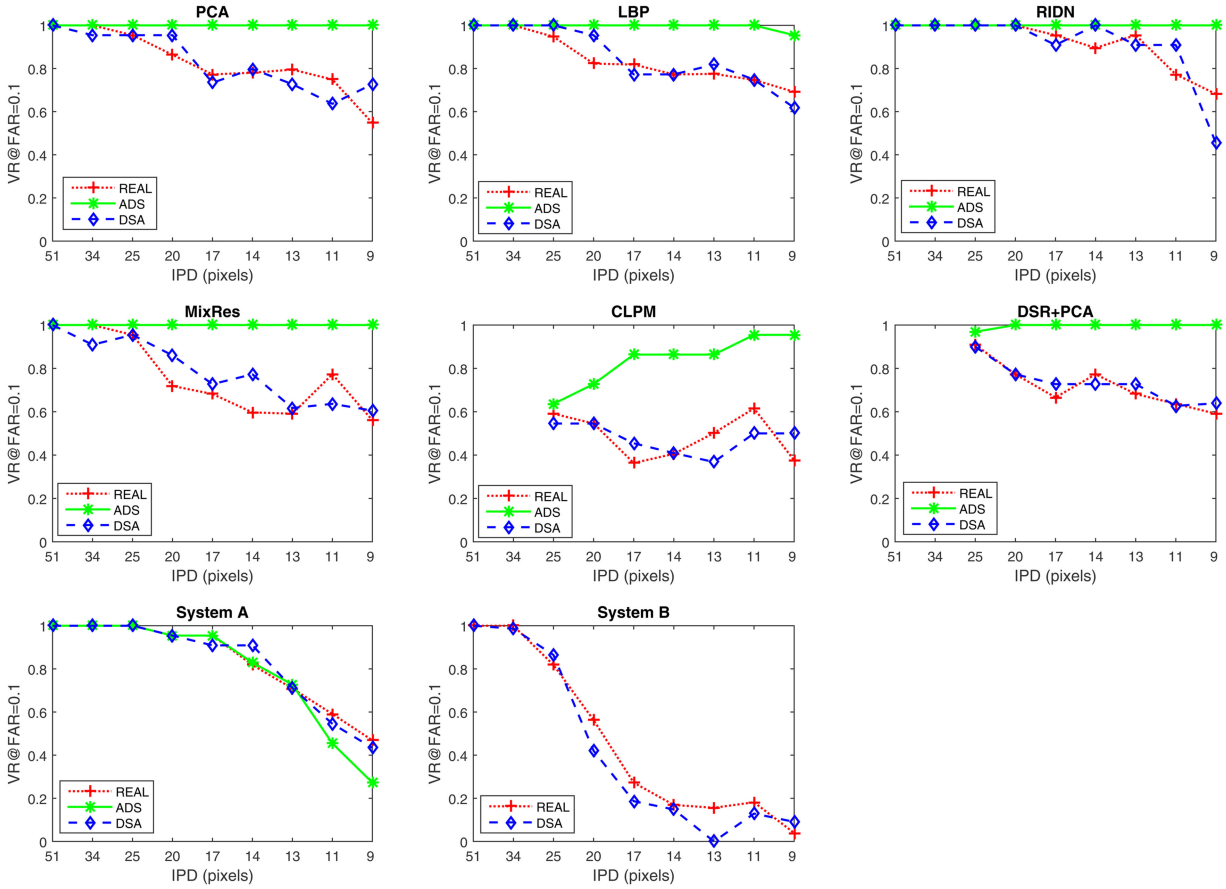

Fig. 5 *Comparing 'REAL', 'ADS', and 'DSA' settings on the UT-FAD data set. X-axis: IPD (pixels); Y-axis: verification rate (VR) at FAR 0.1*

Table 4 Details of our experimental data from the Human ID data set. Ni, total number of images. Ns, number of subjects

IPD	27	23	20	18	16	14	12	10	9
Ni	707	664	755	837	768	811	827	877	873
Ns	251	259	279	282	281	276	276	276	272

gallery images is also much lower than that from the UT-FAD data set.

We observe a similar trend as in the experiments on the UT-FAD data set. The resolution changes almost have no influence on the results of the basic methods for 'ADS' setting. The two settings for which the images are aligned at low resolution, 'REAL' and 'DSA', have worse results than the 'ADS' setting. This demonstrates that alignment still is an important factor for the face recognition performance in this uncontrolled situation. The difference to the experiments on the UT-FAD data set is that the results of 'DSA' are better than 'REAL', which is as expected because, besides alignment, other factors like pose and illumination also lead to performance loss for the 'REAL' setting. System A is still the outlier, but its behaviour is consistent with the previous subsection except for that the pose and poor illumination cause the performance to drop for the 'REAL' setting.

The capability of handling images under uncontrolled situations is also different between the eight methods. RIDN outperforms other methods by a large margin for MLR. However, its performance decreases significantly when the probe resolution

decreases to the range of VLR. Especially for 'REAL' images, RIDN performs worse than most of the other methods at the lowest resolution IPD 9 pixels. PCA, LBP, and MixRes have similar trend. They have a reasonable performance for MLR and perform better than other methods for VLR. LBP performs slightly better among the three. Although CLPM does not perform well on higher resolutions, its results on VLR images are at a similar level as PCA, LBP, and MixRes. DSR + PCA use PCA for recognition and thus has a very similar trend as PCA, but the fact, that it performs worse than PCA, suggests that this super-resolution method does not improve face recognition. System A is the second best method for the highest probe resolution; however, it is very sensitive to resolution changes and thus not suitable for low-resolution face recognition. System B completely collapses for the Human ID data set.

2.4 Uncontrolled situation with uncontrolled training

In the above experiments, PCA, MixRes, CLPM, and DSR + PCA are trained using images from the FRGC v2.0 data set. The training images are of much higher quality than the testing images from the Human ID data set. The results are not very promising since the best result on the lowest resolution real images is only ∼0.3

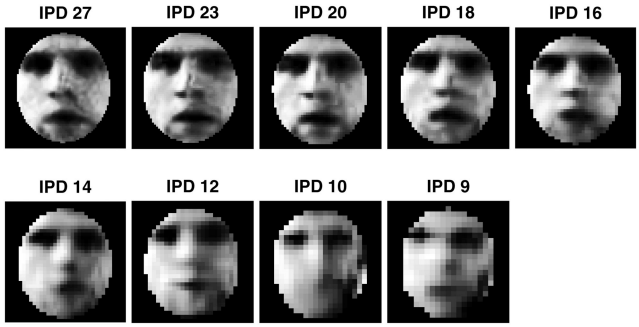

Fig. 6 *Sample images after pre-processing from the Human ID data set*

verification rate at FAR 0.1. To maximise the recognition capability of these methods, we use another data set for training in this subsection, the COX face data set [24]. The COX data set contains video sequences that simulate video surveillance with three different video-based face recognition scenarios. We crop facial images from these videos. We randomly select five images per subject for training in our experiments. We use images of 996 subjects. The parameters for each method are chosen to be the same as in the previous subsection. The results are shown in Fig. 8.

As we can see, the results of MixRes and CLPM are significantly improved using COX training, but PCA results remain similar as in previous experiments. At the lowest resolution, the verification rate of MixRes increased by 0.13 at FAR 0.1. The results at the highest resolution also increased by 0.07. This makes that MixRes significantly outperforms the other methods in the previous subsection for VLR images. CLPM performs overall worse than MixRes. It could not perform well on higher resolutions and is still slightly worse than MixRes on the lowest resolution.

2.5 Discussion

In this section, we demonstrate that poor alignment is a very important factor that causes the loss in the recognition performance for real low-resolution images compared to down-sampled images. Our experiments on the UT-FAD data set demonstrate this for under controlled situation in the absence of pose and illumination variations. The face recognition performance on pre-aligned down-sampled images remains similar to the high-resolution images. However, when we align the images after down-sampling, the performance becomes much worse and strikingly alike the performance on real low-resolution images. Furthermore, we demonstrate, using the Human ID data set, that under uncontrolled situations with pose and illumination variations, poor alignment is still one of the more important reasons for the performance loss.

Our experiments also show the capabilities of different face recognition methods on different ranges of low-resolution. ULR is not a difficult task for most of the recognition methods. System A

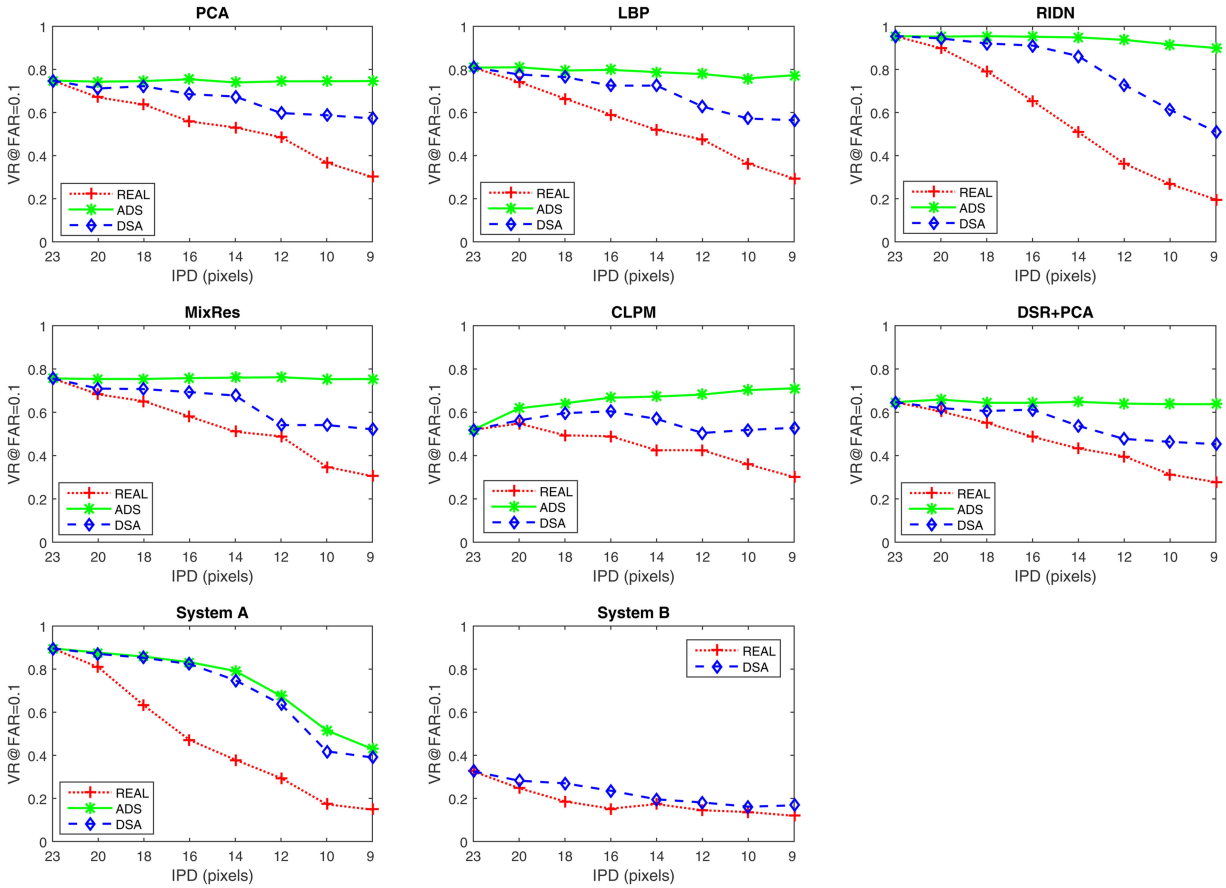

Fig. 7 *Comparing 'REAL', 'ADS', and 'DSA' settings on the Human ID data set. X-axis: IPD (pixels); Y-axis: verification rate (VR) at FAR 0.1*

IET Biom., 2019, Vol. 8 Iss. 4, pp. 267-276

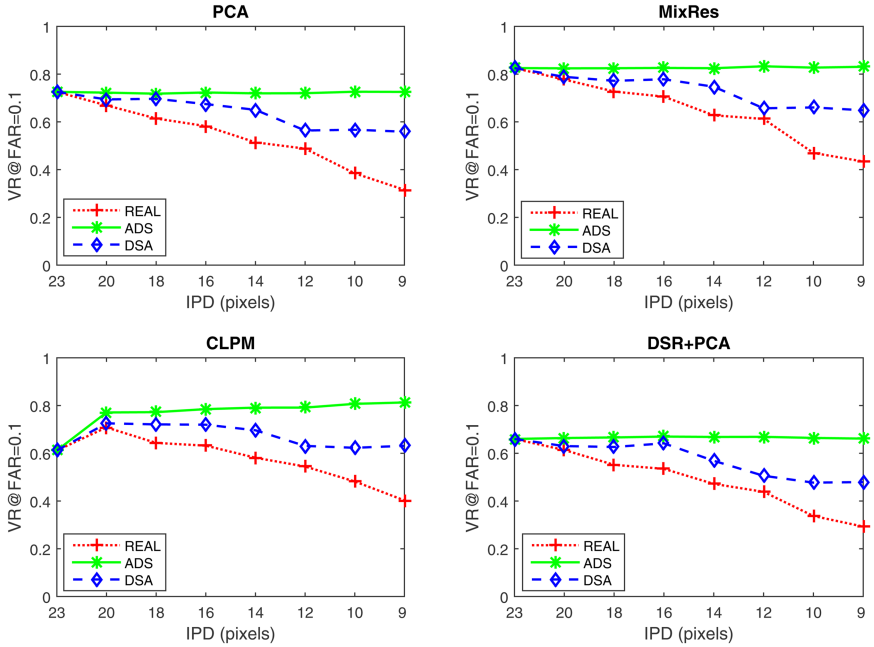

Fig. 8 *Comparing 'REAL', 'ADS', and 'DSA' settings on the Human ID data set using COX training. X-axis: IPD (pixels); Y-axis: verification rate (VR) at FAR 0.1*

and RIDN perform the best for this range. The performance of system B shows some drop at the lower range, but still reasonably well. Images of MLR are harder to recognise than ULR. The performance of most of the face recognition methods, especially the commercial systems, starts to become worse. The basic methods and low-resolution face recognition methods show an overall better performance than the commercial systems. The low-resolution deep-learning face recognition method RIDN still outperforms the other methods generally. VLR is extremely difficult for face recognition and all the methods give poor results. The commercial systems and the RIDN method perform a lot worse than other methods. Simple methods have an advantage in this range. When benefiting from real low-resolution training images, MixRes, which is a simple method based on holistic features, provides the best results for VLR.

3 Matching score-based registration

In the previous section, we demonstrated that poor alignment is a major factor that causes the loss in face recognition performance for real low-resolution images. Thus, if we improve the alignment, we would expect improvement in the recognition performance. We use matching score-based registration for this purpose.

3.1 Method

In the previous experiments, we show that aligning low-resolution images using manually marked landmarks are not good enough for face recognition. An intuitive way for a better alignment is to use better landmark detection methods. There are many good detection methods available for high-resolution facial images. However, because low-resolution images are much more blurry and lack of details, those landmark detection methods are very likely to fail on them. For example, in our experiments, the two commercial face recognition systems both have their own landmark detection function, but both of them could not perform well when manual eye-coordinates are not provided.

If we have a look at the low-resolution facial images, we notice that the size of the images is very small. This has two consequences. On the one hand, facial landmarks cannot be detected accurately. Automatic face detectors which are commonly used for registration usually fail to detect the landmarks on these images. Also, even manually marked landmarks can be inaccurate. On the other hand, because the images are so small, the possible position of the landmarks will be within a range of only a few pixels. Thus, if we have the rough locations of the landmarks, we could try to find the locations for best recognition performance by varying a few pixels around them. Therefore, matching score-based registration, which scales and aligns the images to reach maximum matching score based on crude initial estimates landmarks, can be effective way of improving face recognition performance on low-resolution images. Its effectiveness on high-resolution images has been demonstrated in [32–34].

The process of matching score-based registration is a variation of the general face recognition procedure. The method described here is based on the eye-coordinates because they are the most commonly used landmarks for alignment. It can be extended to more landmarks with higher computational cost. In a standard face recognition system, the probe and reference images are registered using facial landmarks, and then compared. For a probe image x_p and a reference image x_r, given the eye-coordinates ρ of the probe image (reference images are assumed to be pre-aligned), the similarity score is written as $\Delta(x(x_p, \rho), x_r)$. For matching score-based registration, the alignment of the probe image is varied so that there are several aligned images. All those images are compared with each gallery image using a classifier. The best result for each gallery image is stored as the genuine or imposter score which can be used for the subsequent verification or identification process. Matching score-based registration tries to find the eye-coordinates ρ^*, that maximise the similarity between $x(x_p, \rho)$ and x_r, as

$$\rho^* = \arg\max_{\rho} \Delta(x(x_p, \rho), x_r) \tag{1}$$

The output similarity score is then $\Delta^*(x(x_p, \rho^*), x_r)$.

As the identity of the probe image is unknown, when we apply matching score-based registration, we have to pick the best result when comparing with each gallery images. As a result, all the similarity scores will increase no matter they are from the genuine or imposter pairs. There is a risk that the overall face recognition performance will become worse when imposter scores benefit more from the process than the genuine scores. In the following experiments, we will explore the effectiveness of matching score-based registration.

3.2 Experimental set-up

The goal of our experiments is to show matching score-based registration improves the face recognition performance on the real low-resolution images. The results of 'REAL' and 'ADS' from

previous experiments are used here for comparison. In addition, one more setting, *REALM*, is included, where matching score-based registration is applied on the REAL low-resolution probe images (Table 5).

All the face recognition methods except system B are used with the same settings as in previous experiments. System B is not used in the following experiments because it does not allow input of eye-coordinates so that matching score-based registration is not applicable.

The manually marked eye-coordinates which we used to align the 'REAL' images are used as the initial points for matching score-based registration. The range of variation of the eye-coordinates is a 5×5 pixel-region around the manually marked eyes. Within this range, we search for the maximum similarity scores. This region is well balanced between face recognition performance and computational cost and it works for all the images in our experiments. However, a region of 3×3 pixels is chosen for system A and RIDN on the Human ID data set because otherwise the computing time would be too long.

3.3 Experimental results

First, we compare the results of 'REAL', 'ADS', and 'REALM' settings on the UT-FAD data set in Fig. 9.

Matching score-based registration has different impact on different recognition methods. It significantly improves the performance of PCA, DSR + PCA, and MixRes: results of 'REALM' are much better than that of 'REAL' and are getting close to the results of 'ADS'. This further confirms that alignment is the major difference between down-sampled and real low-resolution images. Matching score-based registration also improves the performance of LBP and CLPM, though the improvement is not as much. However, matching score-based registration could not benefit RIDN and system A. This is probably because the imposter scores benefit more as we have mentioned in the previous subsection.

In Fig. 10, we present the results of matching score-based registration compared with 'REAL' and 'ADS' settings on the Human ID data set. Matching score-based registration has a similar effect for each method as on the UT-FAD data set. The difference is that, for most of the methods, the results of 'REALM' are even better than the results of 'ADS' on the higher resolutions. This confirms that poor alignment is an important problem in low-resolution face recognition and also indicates that for the Human ID data set, even the highest resolution probes do not have perfect alignment.

Finally, we present the results on the Human ID data set with COX training in Fig. 11. The results of PCA and DSR + PCA are very similar to the results from FRGC2 training. MixRes and CLPM significantly improved by both matching score-based registration and COX training. When comparing those two, MixRes has a better performance, especially at the lowest resolution. MixRes obtains 0.55 verification rate at FAR 0.1 which is 0.1 more than the second best (PCA) at the lowest resolution.

Table 5 Three probe settings with matching score-based registration

Setting	Description
REAL	real low-resolution images
ADS	first align then down-sample
REALM	matching score-based registration on REAL low-resolution images

3.4 Discussion

From the above experiments on two data sets, we show that improving the alignment by means of matching score-based registration can indeed improve the face recognition performance on real low-resolution images no matter under controlled or uncontrolled situations. However, the improvement varies for different methods. For face recognition methods using global features, the improvement of matching score-based registration is significant. In the experiments using the Human ID data set, because even the alignment of the highest resolution images is not

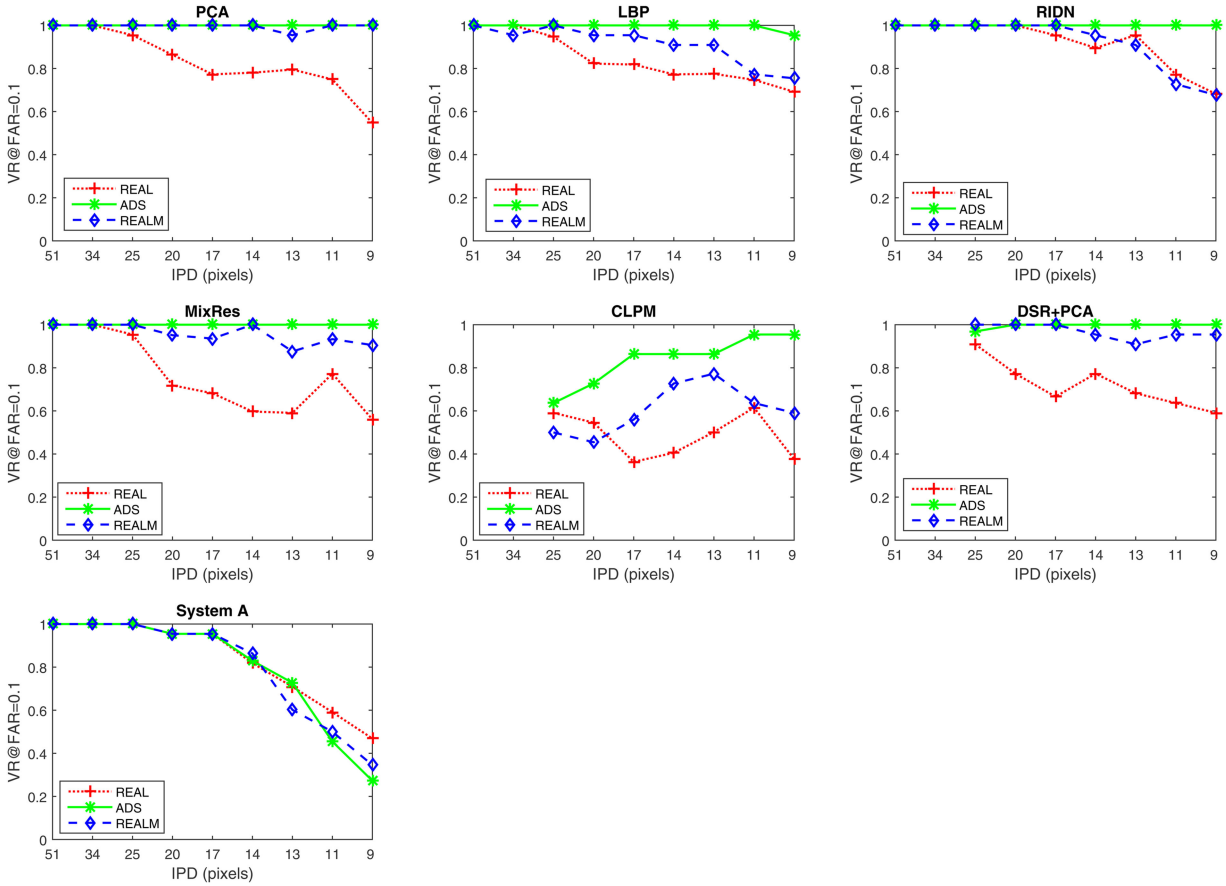

Fig. 9 *Comparing 'REAL', 'ADS', and 'REALM' settings on the UT-FAD data set. X-axis: IPD (pixels); Y-axis: verification rate (VR) at FAR 0.1*

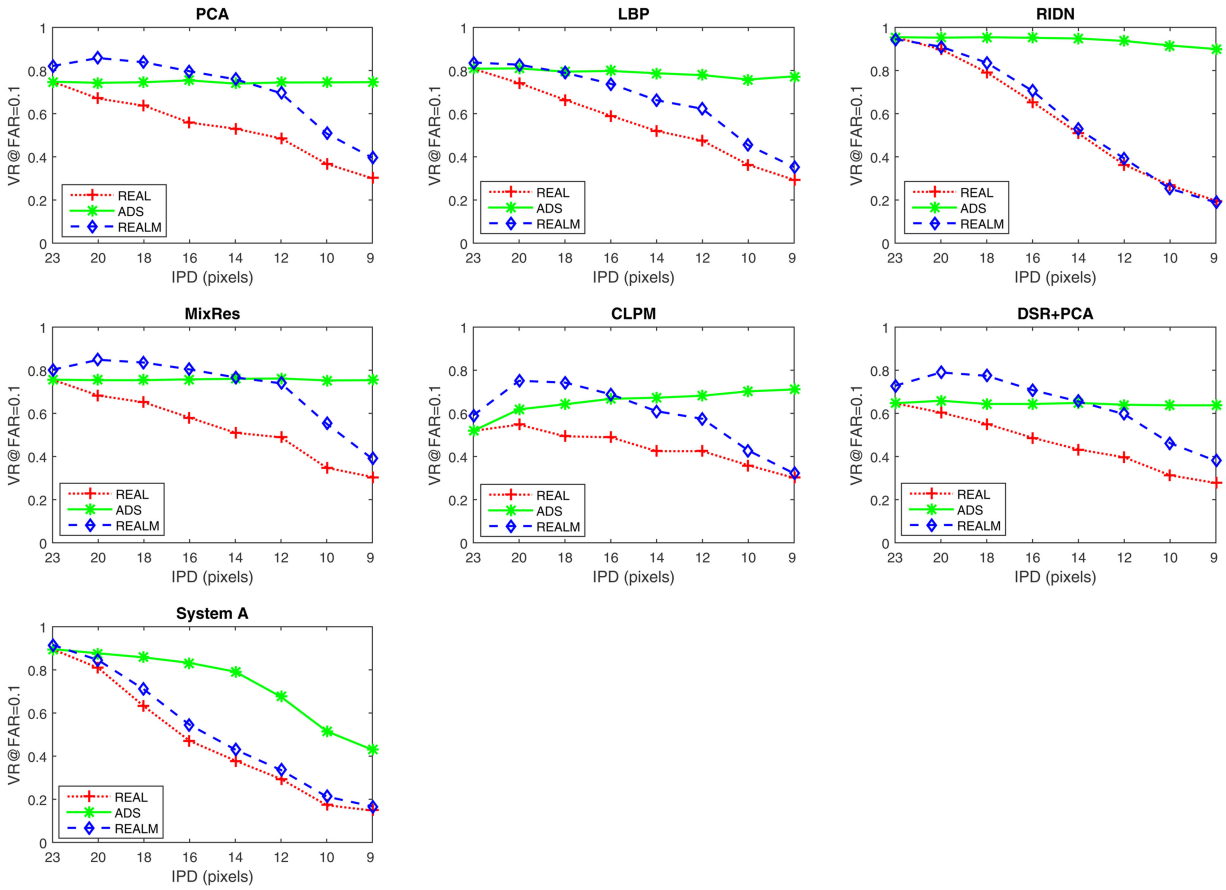

Fig. 10 *Comparing 'REAL', 'ADS', and 'REALM' settings on the Human ID data set. X-axis: IPD (pixels); Y-axis: verification rate (VR) at FAR 0.1*

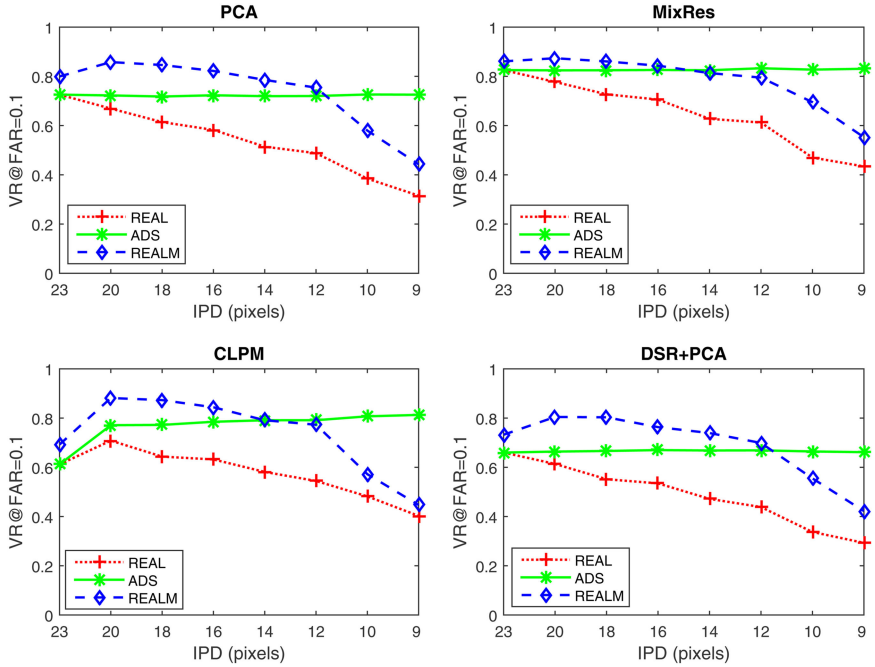

Fig. 11 *Comparing 'REAL', 'ADS', and 'REALM' settings on the Human ID data set with COX training. X-axis: IPD (pixels); Y-axis: verification rate (VR) at FAR 0.1*

good enough, some results of matching score-based registration are even better than the pre-aligned down-sampled images. On the other hand, RIDN and the commercial face recognition system could not benefit from matching score-based registration: there is hardly any improvement on the face recognition performance. Commercial systems are designed strictly for high-resolution images and behave unpredictably for low resolution. Deep-learning is less dependent on alignment accuracy and therefore does not profit from matching score-based registration. Furthermore, with the help from both real low-resolution training and matching score-based registration, MixRes significantly outperforms all the other methods for VLR.

4 Conclusion

In low-resolution face recognition, the effectiveness of a new method is usually tested on images that are down-sampled from high resolution. However, face recognition on down-sampled

images is much easier than on real low-resolution images. In this paper, we provide an insight into the difference between down-sampled and real low-resolution images. Surprisingly, it turns out that the degradation of face recognition performance is not caused in the first place because low-resolution image would contain less information. We demonstrate that, under controlled situations (in the absence of pose and illumination variations), poor alignment is the major problem that causes the poor recognition performance on real low-resolution images. When images are captured under uncontrolled situations, alignment still plays an important role, but performance also degrades due to pose and illumination variations. The results on down-sampled images are flattered because they are usually aligned using landmarks annotated on the high-resolution images while landmarks of real low-resolution images are much less accurate. If the down-sampled images are aligned based on landmarks estimated after down-sampling, the face recognition performance appears to be similar to that obtained with real low-resolution images. In addition, we propose to use matching score-based registration to achieve better alignment and hence better face recognition performance. Matching score-based registration allows for fine tuning the registration of the poorly aligned images, and does not need accurate landmarks. Our experiments show that matching score-based registration significantly improves the face recognition performance of most of the methods, though it did not improve the performance of a deep-learning method and commercial face recognition systems. Furthermore, we divide low resolution into three classes: ULR, MLR, and VLR. Most face recognition methods, including commercial systems that are developed for high-resolution face recognition methods, can perform well on ULR images. The performance of commercial systems starts to become worse for MLR images, while methods that are aimed at low-resolution face recognition (especially a deep-learning-based method) still perform well. VLR images are very difficult to recognise and the performance of the deep-learning method degrades significantly, while simple holistic methods perform the best. In our experiments, a simple method based on holistic features, which can benefit from real low-resolution training and matching score-based registration, outperforms other methods for the range of VLR and the lower range of MLR. Finally, we recommend that if not enough real low-resolution images are available, to use down-sampled images that are aligned based on landmarks estimated after down-sampling as substitutes.

5 References

[1] Introna, L.D., Nissenbaum, H.: 'Facial recognition technology: a survey of policy and implementation issues'. Technical report, Center for Catastrophe Preparedness and Response, New York University, 2009

[2] Li, S.Z., Jain, S.Z.: '*Handbook of face recognition*' (Springer, London, 2011, 2nd edn.)

[3] Hennings-Yeomans, P., Baker, S., Kumar, B.: 'Recognition of low-resolution faces using multiple still images and multiple cameras'. 2nd IEEE Int. Conf. on Biometrics: Theory, Applications and Systems, 2008. BTAS 2008, Arlington, VA, USA, 29th September–1st October 2008, pp. 1–6

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

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

[6] 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

[7] Moutafis, P., Kakadiaris, I.A.: 'Semi-coupled basis and distance metric learning for cross-domain matching: application to low-resolution face recognition'. Proc. Int. Joint Conf. on Biometrics, Clearwater, FL, 29 September–2 October 2014

[8] Peng, Y., Spreeuwers, L.J., Veldhuis, R.N.J.: 'Likelihood ratio based mixed resolution facial comparison'. 3rd Int. Workshop on Biometrics and Forensics (IWBF2015), Gjøvik, Norway, March 2015, pp. 1–5

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

[10] 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

[11] Peng, Y., Spreeuwers, L.J., Gökberk, B., *et al.*: 'Comparison of super-resolution benefits for downsampled images and real low-resolution data'. Proc. of the 34rd Symp. on Information Theory in the Benelux and the 3rd Joint WIC/IEEE Symp. on Information Theory and Signal Processing in the Benelux, Leuven, Belgium, WIC, May 2013, pp. 244–251

[12] Peng, Y., Spreeuwers, L., Veldhuis, R.: 'Low-resolution face alignment and recognition using mixed-resolution classifiers', *IET Biometrics*, 2017, **6**, (6), pp. 418–428

[13] 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

[14] ISO/IEC 19794-5: 'Information technology – biometric data interchange formats – Part 5: face image data', 2005

[15] ANSI/INCITS 385-2004[R2014]: Information technology – face recognition format for data interchange, 2014

[16] Marciniak, T., Chmielewska, A., Weychan, R., *et al.*: 'Influence of low resolution of images on reliability of face detection and recognition', *Multimedia Tools Appl.*, 2015, **74**, (12), pp. 4329–4349

[17] BS EN 50132-7: Alarm systems. CCTV surveillance systems for use in security applications. Application guidelines, 1996

[18] ICAO, Doc9303: Machine Readable Travel Documents, Seventh Edition, Part 9: Deployment of biometric Identification and Electronic Storage of Data in eMRTDs, 2015

[19] Phillips, P.J., Flynn, P.J., Scruggs, T., *et al.*: 'Overview of the face recognition grand challenge'. IEEE Computer Vision and Pattern Recognition, San Diego, CA, USA, 2005, pp. 947–954

[20] Huang, G.B., Ramesh, M., Berg, T., *et al.*: 'Labeled faces in the wild: a database for studying face recognition in unconstrained environments'. Technical Report 07-49, University of Massachusetts, Amherst, October 2007

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

[22] 'University of Twente – faces at distances database'. Available at http://scs.ewi.utwente.nl/downloads/show,University%20of%20Twente%20-%20Faces%20At%20Distances%20(UT-FAD), accessed 8 November 2017

[23] O'Toole, A.J., Harms, J., Snow, S.L., *et al.*: 'A video database of moving faces and people', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2005, **27**, (5), pp. 812–816

[24] Huang, Z., Shan, S., Wang, R., *et al.*: 'A benchmark and comparative study of video-based face recognition on Cox face database', *IEEE Trans. Image Process.*, **24**, (12), pp. 5967–5981, 2015

[25] Kovalevsky, J. 'Atmospheric effects on image formation', in '*Modern astrometry*' (Springer, Berlin Heidelberg, 2002), pp. 33–59

[26] Jähne, B.: '*Digital image processing*' (Springer-Verlag, Berlin Heidelberg, 2002, 5th edn.)

[27] Turk, M., Pentland, A.: 'Face recognition using eigenfaces'. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, 1991. Proc. CVPR '91, Maui, HI, USA, June 1991, pp. 586–591

[28] 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

[29] Zeng, D., Chen, H., Zhao, Q.: 'Towards resolution invariant face recognition in uncontrolled scenarios'. 2016 Int. Conf. on Biometrics (ICB), Halmstad, Sweden, 2016, pp. 1–8

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

[31] 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, Kauai, HI, USA, 2001, vol. 1, pp. I–511–I–518

[32] Boom, B.J., Spreeuwers, L.J., Veldhuis, R.N.J.: 'Automatic face alignment by maximizing similarity score'. Proc. of the 7th Int. Workshop on Pattern Recognition in Information Systems, Madeira, Portugal, Biosignals, June 2007, pp. 221–230

[33] 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

[34] 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