RESEARCH

Multi-feature shape regression for face alignment

Wei-Jong Yang, Yi-Chen Chen, Pau-Choo Chung and Jar-Ferr Yang[*]

Abstract

For smart living applications, personal identification as well as behavior and emotion detection becomes more and more important in our daily life. For identity classification and facial expression detection, facial features extracted from face images are the most popular and low-cost information. The face shape in terms of landmarks estimated by a face alignment method can be used for many applications including virtual face animation and real face classification. In this paper, we propose a robust face alignment method based on the multi-feature shape regression (MSR), which is evolved from the explicit shape regression (ESR) proposed in Cao et al. (Int, Vis, 2014, 107:177–190, Comput). The proposed MSR face alignment method successfully utilizes color, gradient, and regional information to increase accuracy of landmark estimation. For face recognition algorithms, we further suggest a face warping algorithm, which can cooperate with any face alignment algorithm to adjust facial pose variations to improve their recognition performances. For performance evaluations, the proposed and the existing face alignment methods are compared on the face alignment database. Based on alignment-based face recognition concept, the face alignment methods with the proposed face warping method are tested on the face database. Simulation results verify that the proposed MSR face alignment method achieves better performances than the other existing face alignment methods.

Keywords: Face alignment, Face warping, Face recognition, Pose variation, Shape regression

1 Introduction

For smart living applications, the identification and behavior and emotion detection of a person become more and more important in our daily modern life. For identity verification and facial expression detection, the facial features extracted from the captured images are the most popular and low-cost information. The face shape in terms of the positions of landmarks is one of the important features. Once the face shape is extracted, the landmarks can be used for many applications including face animation for argument reality (AR) and virtual reality (VR) and emotion detection and face recognition for smart living services. Face recognition has been widely investigated in academic and industrial communities due to the extraordinary demands of security controls in sensitive areas, device and machine accesses, internet secure usages, etc. In practical face recognition systems, for example, a low-computation and accurate system could be operated under various challenges, such as pose variations,

illumination changes, and partial occlusions. To overcome the problem of facial pose variations, a suitable face alignment algorithm figured with an appropriate warping method becomes essential for face recognition.

Face alignment, which could locate the semantic key facial landmarks, such as facial contour, eye and mouth shapes, and nose and chin positions, is a necessary tool to estimate face contour and key facial characters in face images. From a captured facial image, the goal of face alignment is to minimize the difference between the estimated and ground true shapes defined by a set of facial landmarks. Over past decades, the shape estimation along the outer facial contour of a given facial image has been widely investigated for face alignment. The alignment algorithms can be generally categorized into optimization-based and regression-based approaches. The optimization-based algorithms depend on the design of error functions and optimization iterations. The most popular optimization-based algorithms include the active shape models (ASMs) [1, 2] and their extensions, called active appearance models (AAMs) [3–6]. For both ASM and AAM, the generative landmark positions from rough

* Correspondence: jfyang@ee.ncku.edu.tw
Department of Electrical Engineering, Institute of Computer and
Communication Engineering, National Cheng Kung University, Tainan,
Taiwan

Yang *et al. EURASIP Journal on Advances in Signal Processing* (2018) 2018:51

Page 2 of 13
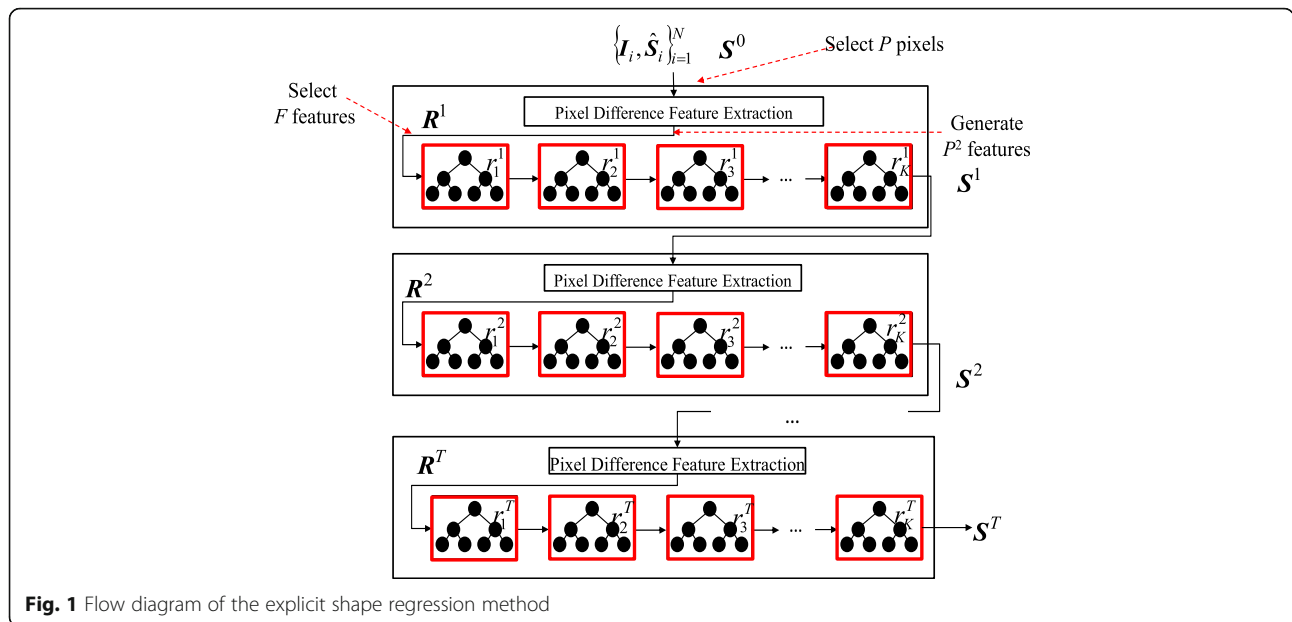
Fig. 1 Flow diagram of the explicit shape regression method

initial estimations are trained by the point distribution to iteratively refine the results. The parametric shape models utilized to keep shape constraints are not flexible enough to fit the faces with large variations and partial occlusions. The regression-based algorithms [7, 8] utilize regression functions to directly map an image appearance to the target output. Because the complex variation is trained from a large dataset, the testing process becomes efficient in general. In 2012, Cao et al. proposed the explicit shape regression (ESR) method [9] and realized the shape constraint to attain a good face alignment in non-parametric manners.

As to face recognition, numerous successful algorithms were proposed [10–14]. Over the past years, the subspace projection optimizations (SPO) with linear and non-linear approaches are the main research trends. The principal component analysis (PCA) [10–13] and linear discriminant analysis (LDA) [14] with linear approaches attempt to seek a low-dimensional subspace for computation

reduction and performance improvement. The kernel PCA (KPCA) [15–17] and kernel LDA (KLDA) [18–21] with non-linear projection approaches can uncover the underlying structure when the samples lie on a nonlinear manifold structure. The linear regression classification (LRC) proposed in [22] is simple in nature and effective in performance while the modular linear regression classification (MLRC) can deal with the occlusion problems. Simple computation in both training and testing procedures is the advantage of the above methods. Without re-training the existing candidates, the SPO face recognition methods can add the hyperplane of any new identity in the system directly. However, without any assistance, the SPO face recognition methods cannot achieve successful recognition in uncontrollable variations. Currently, some researches focused on contextual information and learning-based algorithms [23–25]. In [23], the context-aware local binary feature achieves better robustness than the local feature descriptor such as LDA. The convolutional
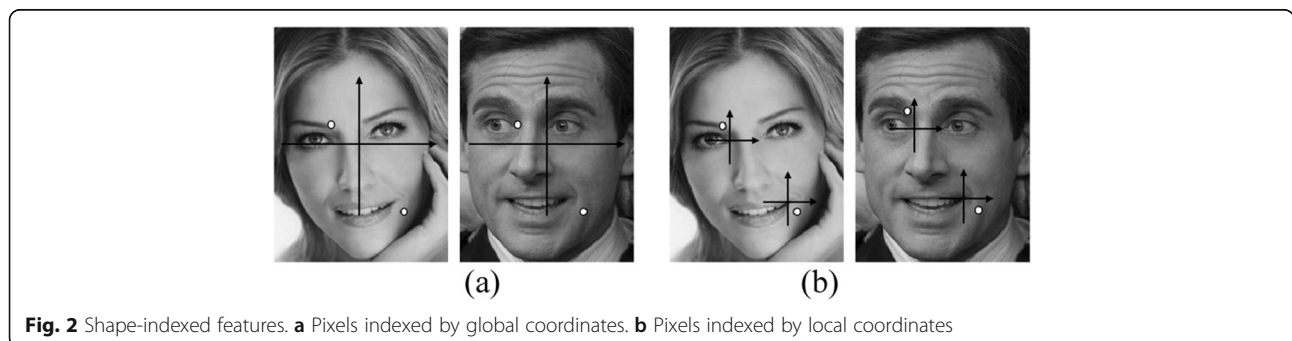
Fig. 2 Shape-indexed features. **a** Pixels indexed by global coordinates. **b** Pixels indexed by local coordinates

Yang *et al. EURASIP Journal on Advances in Signal Processing* (2018) 2018:51

Page 3 of 13

neural networks [24, 25] were introduced for face recognition to show better performance than the SPO approaches if a suitable deep network with a large tagged database is learnt. However, the learning approaches, which need intensive computation for training and testing computation, may not be suitable for real-time applications in current handheld devices.

The rest of the paper is organized as follows. In Section 2, the proposed methods in this paper are described. In Section 2.1, the explicit shape regression (ESR) face alignment method is first reviewed. Section 2.2 introduces the proposed multi-feature shape regression (MSR) face alignment method in details. In Section 2.3, the alignment-based face recognition with cross face warping is suggested to improve the performances of SPO face recognition methods. The detailed procedure of the cross warping method is described. To demonstrate the effectiveness of the methods, the performances of the proposed and existing face alignment methods are first evaluated on the famous face alignment database in Section 3. The face recognition performances with pose variations are then demonstrated on the face recognition database by using different SPO face recognition methods and different face alignment algorithms. In Section 4, the conclusions about this paper are finally addressed.

2 Methods

In this paper, we propose a robust face alignment method, which can estimate the positions of facial parameters and a cross face warping method to adjust the position of facial parameters. Thus, we can apply all SPO face recognition methods to the adjusted face image to achieve better recognition performance. The robust face alignment method is based on multi-feature shape regression (MSR) to achieve robust landmark estimation. With the estimated landmarks, a face cross warping method is proposed to reduce the pose variation of facial images such that the SPO face recognition methods can be improved to obtain better recognition performances.

2.1 Face alignment with shape regression

The face shape is generally defined by the positions of M selected landmarks as
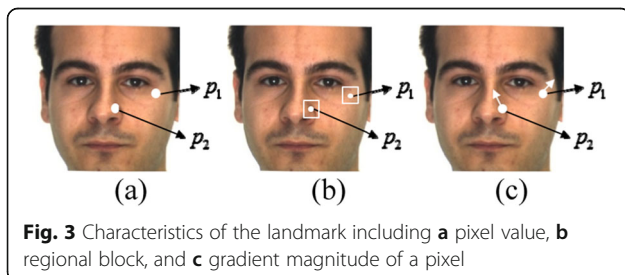
Fig. 3 Characteristics of the landmark including **a** pixel value, **b** regional block, and **c** gradient magnitude of a pixel
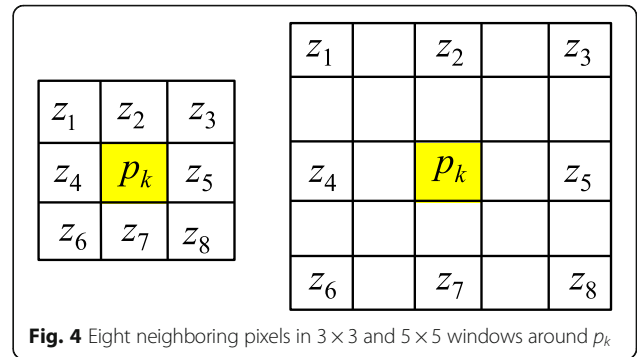
Fig. 4 Eight neighboring pixels in 3×3 and 5×5 windows around p_k

$$S = [x_1, y_1, x_2, y_2, ..., x_M, y_M] = [\boldsymbol{p}_1, \boldsymbol{p}_2, ..., \boldsymbol{p}_M], \quad (1)$$

where $\boldsymbol{p}_m = (x_m, y_m)$ denotes the position of the mth landmark in the facial image. To estimate M landmarks from a given facial image, we should design an effective face alignment method to estimate (x_m, y_m), $m = 1, 2, ..., M$. The explicit shape regression (ESR) algorithm [9] is a famous learning-based regression method. Figure 1 shows the basic framework of the ESR algorithm with a boosted regression process [26, 27], which combines T weak regressors, $\boldsymbol{R}^1, \boldsymbol{R}^2, ..., \boldsymbol{R}^T$, in an additive manner. Each regressor computes a shape increment δS from image features and updates the face shape as

$$S^t = S^{t-1} + \boldsymbol{R}^t(\boldsymbol{I}, S^{t-1}), t = 1, 2, ..., T. \quad (2)$$

Given N training data $(\boldsymbol{I}_i, \hat{S}_i)$ for $i = 1, 2, ..., N$, the regressors, $\boldsymbol{R}^1, \boldsymbol{R}^2, ..., \boldsymbol{R}^T$, are sequentially learnt until the training error no longer decreases. The tth regressor R^t is learnt by minimizing the regression error as

$$\boldsymbol{R}^t = \arg \min_{\boldsymbol{R}} \sum_{i=1}^{N} \|\hat{S}_i - (S_i^{t-1} + \boldsymbol{R}(\boldsymbol{I}_i, S_i^{t-1})\|, \quad (3)$$

where $\hat{S}_i$ denotes the ground truth shape and S_i^{t-1} is the estimated shape obtained from the previous $(t-1)$th regressor.

Table 1 Comparisons with different weights for pixel, region, and gradient differences

Weights (w_p, w_r, w_g)	Landmark error (%)
0.8, 0.1, 0.1	3.90
0.7, 0.2, 0.1	3.59
0.7, 0.1, 0.2	4.50
0.6, 0.3, 0.1	*3.30*
0.6, 0.2, 0.2	3.95
0.6, 0.1, 0.3	4.80
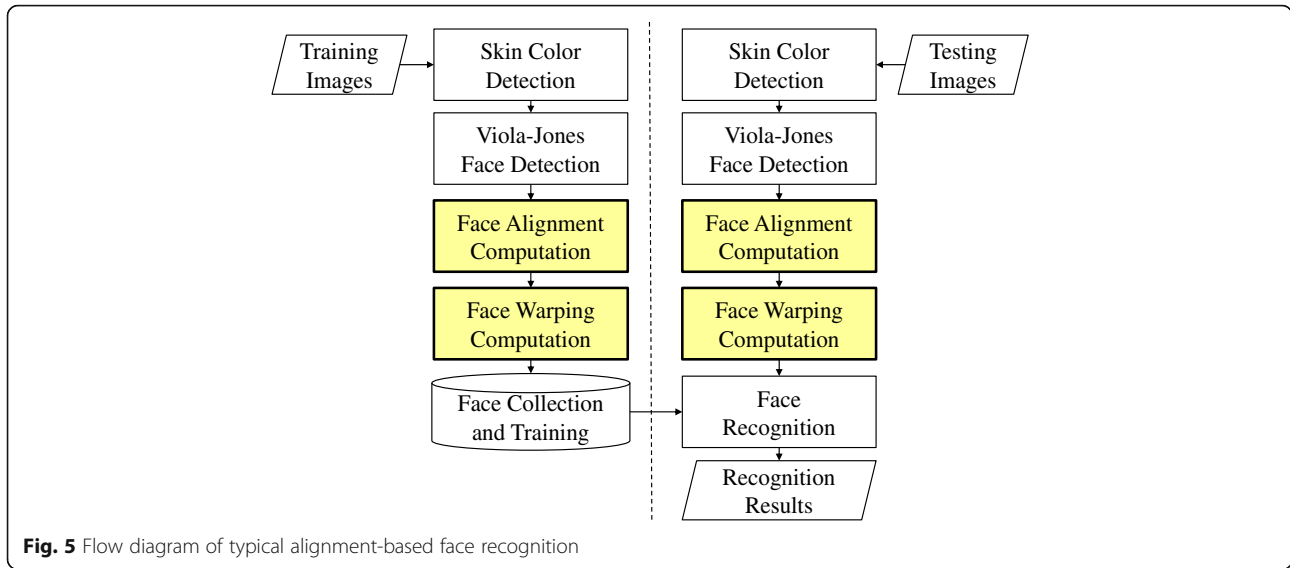
Fig. 5 Flow diagram of typical alignment-based face recognition

However, a simple weak regressor has the limited performance to reduce the error. For this reason, the two-level cascaded regression with the selected feature extraction as shown in Fig. 1 is proposed, and each weak regressor $\boldsymbol{R}^t$ is learnt by the second-level boosted regression, i.e., $\boldsymbol{R}^t = [r^t_1, r^t_2, ..., r^t_k, ..., r^t_K]$. Thus, the selected features are extracted by shape-indexed methods at each outer stage. Afterwards, each fern selects F of these features to infer an offset based on the correlation-based feature selection method. The fern-based regressor, shape-indexed feature, and correlation-based feature selection will be described in details as follows.

The fern is firstly applied for classification [26] and later used for regression [27]. In the ESR, each fern is composed of F features and thresholds. And the threshold is used to divide all the training samples into 2^F bins. After classification of all training samples, the regression output δS_b in each bin b minimizes the alignment error of Ω_b. The training samples falling into the bin as:

$$\delta\boldsymbol{S}_b = \ \arg\ \min_{\delta\boldsymbol{S}}\sum_{i\in\Omega_b}\left\|\hat{\boldsymbol{S}}_i - (\boldsymbol{S}_i + \delta\boldsymbol{S})\right\|, \qquad (4)$$

where $\boldsymbol{S}_i$ denotes the estimated shape in the previous step. According to (4), δS_b can be estimated by:

$$\delta\boldsymbol{S}_b = \frac{1}{\Omega_b}\sum_{i\in\Omega_b}\left(\hat{\boldsymbol{S}}_i - \boldsymbol{S}_i\right)^2. \qquad (5)$$

The training samples falling to the same bin own the same regression output, δS_b. Each outer stage regressor generates P pixels, $I(q_k)$, $k = 1, 2, ..., P$ randomly which are indexed relative to the nearest landmark of mean shape, as shown in Fig. 2. Total P^2 pixel-difference features,
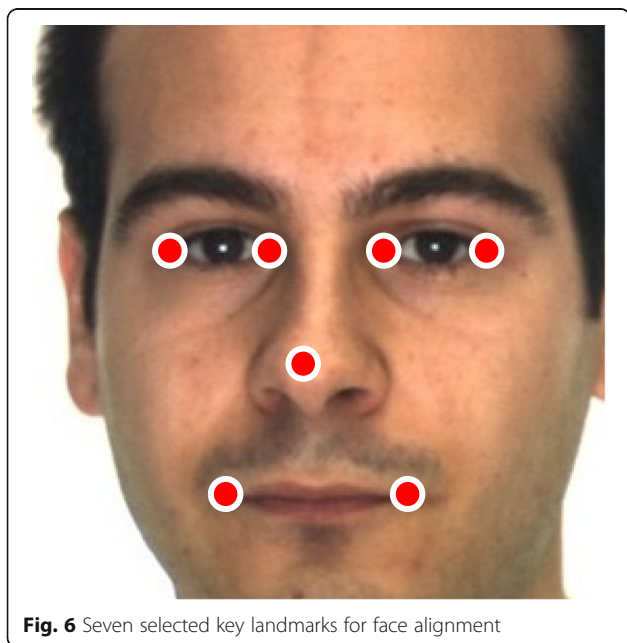
Fig. 6 Seven selected key landmarks for face alignment

(a)   (b)
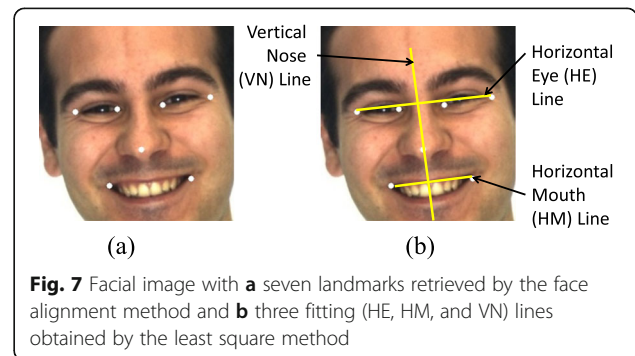
Fig. 7 Facial image with **a** seven landmarks retrieved by the face alignment method and **b** three fitting (HE, HM, and VN) lines obtained by the least square method

Yang *et al. EURASIP Journal on Advances in Signal Processing* (2018) 2018:51

Page 5 of 13
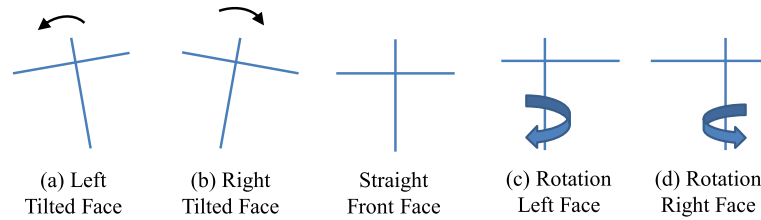
Fig. 8 Four major deformations based on detected cross in face images. **a** Left-tilted face. **b** Right-tilted face. **c** Rotation left face. **d** Rotation right face

$$f_{k,j} = I(q_k) - I(q_j), \qquad (6)$$

between all possible two pixels are generated. The local features are more discriminative than the global ones. Pixels indexed by the same local coordinates have the same semantic meaning, but pixels indexed by the same global coordinates have different semantic meanings due to face pose variation. Most of the useful features are distributed around salient landmarks such as eyes, nose, and mouth. To form a fern, F out of P^2 features are selected by calculating the correlation between the features and the regression target, which is the difference between the ground truth shape and the current estimated shape. The optimization can be achieved while we generate a random unit vector, then project each regression target onto it. We finally estimate the correlation coefficient between feature values and the lengths of projections to find the optimal shape.
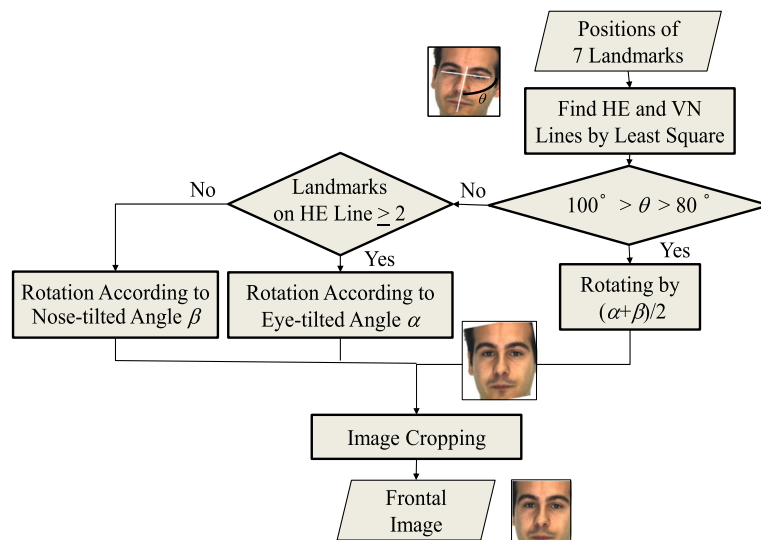
2.2 Multi-feature shape regression

The ESR algorithm detects the similarity of landmarks by the intensity difference of pixels as stated in (6);
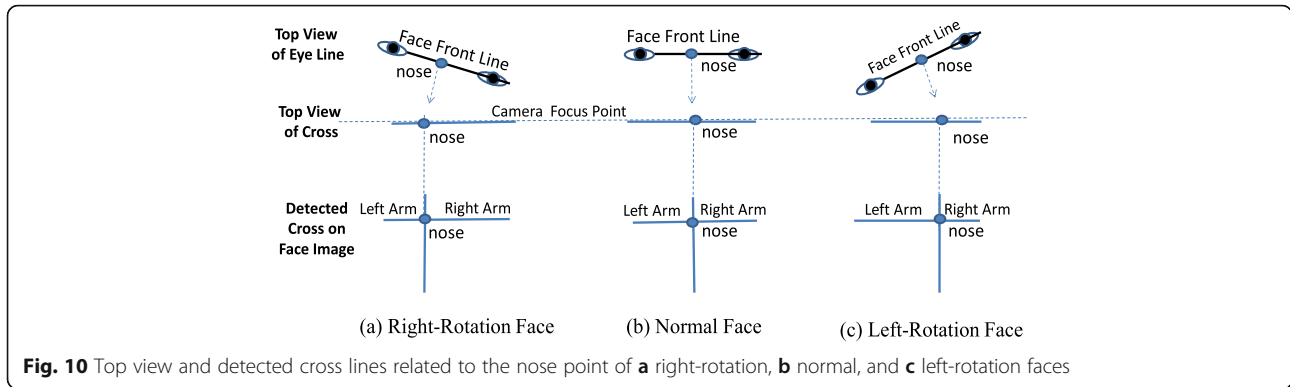
however, the characteristics of the landmarks are different not only with its pixel value. As shown in Fig. 3, we should further check the similarity of their surroundings to improve the detection performance. The multi-feature shape regression (MSR) method replaces the pixel difference feature with the multiple features to achieve more robust landmark detection than the ESR method. In the first feature set, as shown in Fig. 3a, the color values of the pixel at p_k is defined as

$$\boldsymbol{v}_k^p = [r(p_k), g(p_k), b(p_k)], \qquad (7)$$

where $r(p_k)$, $g(p_k)$, and $b(p_k)$ denote the red, green, and blue values of the pixel at p_k for the kth selected landmark of the image, respectively. To achieve reliable results, the intensities of eight neighboring pixels in 3×3 and 5×5 windows as shown in Fig. 4 can be used for detecting the similarity of the landmarks. Thus, for the second feature set as shown in Fig. 3b, we use the regional values, $\boldsymbol{v}_k^r$, at p_k as

Fig. 9 Flow diagram of conditional cross warping for tilted faces

Fig. 10 Top view and detected cross lines related to the nose point of **a** right-rotation, **b** normal, and **c** left-rotation faces

$$v_k^r = [I(z_1), I(z_2), ..., I(z_8)]. \tag{8}$$

For the last feature set, as shown in Fig. 3c, we choose the gradient magnitudes at p_k defined as

$$v_k^g = \left[\nabla_x(I(p_k)), \nabla_y(I(p_k)) \right], \tag{9}$$

where $\nabla_x(I(p_k))$ and $\nabla_y(I(p_k))$ denote the gradients along x and y directions are computed by horizontal and vertical Sobel filters, respectively. With pixel, region, and gradient features, the total difference between the jth pixel and the kth landmark at p_k is expressed by
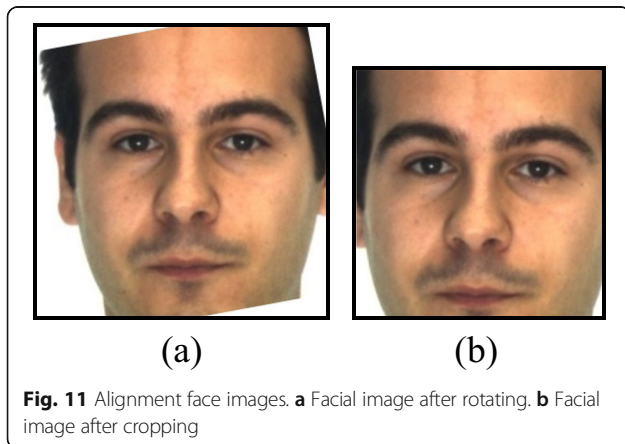
$$d_{k,j}^T = w_p d_{k,j}^p + w_r d_{k,j}^r + w_g d_{k,j}^g \tag{10}$$

where w_p, w_r, and w_g are the selected weights for the pixel, region, and gradient differences, respectively. In (10), $d_{k,j}^p$, d_{kj}^r, and d_{kj}^g are given as

$$d_{k,j}^p = \left| v_k^p - v_j^p \right|, \tag{11}$$

$$d_{kj}^r = \left| v_k^r - v_j^r \right|, \tag{12}$$

and

$$d_{kj}^g = \left| v_k^g - v_j^g \right|. \tag{13}$$

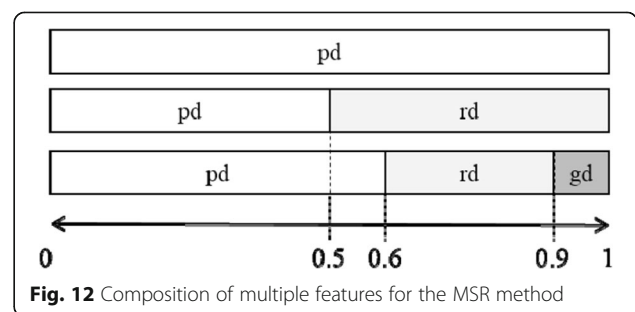which denote the pixel, region, and gradient differences between the kth and jth pixels, respectively. Thus, the feature $f_{k,j}$ stated in (6) suggested in EST method is changed to the total difference as

$$f_{k,j} = d_{k,j}^T, \tag{14}$$

in the proposed MSR method. To determine the weights depicted in (10), Table 1 shows the experimental results that exhibit landmark errors with different sets of weights. The pixel difference, which plays the main role in shape regression, is with the largest weight while the region and gradient differences, which are used for the feature refinements, are with slightly smaller weights. By experiments, we found that the weights with 0.6, 0.3, and 0.1 for pixel, region, and gradient differences, respectively, achieve the best performance for shape regression. It is noted that the above MSR concept can be extended to more features and can be applied for any landmark estimation of target objects.

2.3 Face warping method for alignment-based face recognition

Once the positions of the key landmarks are extracted by a face alignment method, they can be used for many

Fig. 11 Alignment face images. **a** Facial image after rotating. **b** Facial image after cropping

Fig. 12 Composition of multiple features for the MSR method

Yang *et al. EURASIP Journal on Advances in Signal Processing* (2018) 2018:51

Page 7 of 13

applications such as face animation for argument reality (AR) and virtual reality (VR) and emotion detection and face recognition for smart living services. In this section, we can use the face alignment to improve the performance of face recognition. Figure 5 shows the flow diagram of a typical alignment-based face recognition, which includes four major functions of face detection, face alignment, face warping, and face recognition. The face detection includes skin color detection, morphological operations, and Viola-Jones face detector [28]. The skin color is a simple and distinct feature for face detection to reduce the computation [29]. The morphological erosion and dilation are used to remove the noises of the detected skin areas. After morphological operation, the final face detection could be performed in large connected skin areas by Viola-Jones face detector. To improve SPO face recognition methods [10–21], the alignment-based face recognition approach needs a good selection of landmarks and acquires a good warping algorithm to adjust the pose variation of face images.

Figure 6 shows seven selected landmarks, including four eye canthi, one nose tip, and two mouth corners, which are used in the face warping method. After seven key landmarks are extracted by a face alignment method, we suggest a cross warping method to adjust the facial image with possible pose variation. First, three fitting lines are obtained by the least square method as shown in Fig. 7b. The horizontal eye (HE) line is detected by fitting four positions of landmarks on the canthi of eyes. The horizontal mouth (HM) line is obtained by two positions of landmarks on two mouth corners. The vertical nose (VN) line is found by fitting the position of the landmark at the nose and orthogonal to the HE and HM lines in the least square sense. Figure 8 shows the typical cross shapes, which are composed of VN and HE lines, of straight front, left-tilted, right-tilted, left-rotation, and right-rotation faces, will be used for adjusting the face alignment. The proposed cross warping method is described as follows.

For general facial images, the deformations could be mixed with tilted and rotated faces. The flow diagram of the cross warping method for correcting the face alignment is shown in Fig. 9. Since the estimated landmarks could not be always correct, we need to detect the reliability of all the landmarks at the same time. It is rational that the two cross lines should be nearly

Table 2 Landmark errors and failure rates compared with different features of the MSR method on LFPW database

Methods	MSR (pd)	MSR (pd + rd)	MSR (pd + rd + gd)
Landmark error (%)	4.06	3.78	3.30
Failure rate (%)	7.27	3.63	3.60

Table 3 Landmark errors and failure rates compared with different features of the MSR method on HELEN database

Methods	MSR (pd)	MSR (pd + rd)	MSR (pd + rd + gd)
Landmark error (%)	4.31	4.17	3.83
Failure rate (%)	3.63	1.52	0.91

orthogonal if the estimated landmarks are correct. Thus, the cross angle θ between the cross HE and VN lines is computed as

$$\begin{aligned} \theta &= \cos^{-1}\left(\frac{\boldsymbol{m}_1 \cdot \boldsymbol{m}_2}{|\boldsymbol{m}_1|\,|\boldsymbol{m}_2|}\right) \\ &= \cos^{-1}\left(\frac{1 + m_1 m_2}{[(1 + m_1^2)(1 + m_2^2)]^{1/2}}\right), \end{aligned} \tag{15}$$

where $\boldsymbol{m}_1 = [1, m_1]$ and $\boldsymbol{m}_2 = [1, m_2]$ are the slope vectors, which can characterize the HE and VN lines with slopes of m_1 and m_2, respectively. The dot operator in (15) denotes the inner product. Before the warping process, we first compute the eye-tilted and nose-tilted angles. The eye-tilted angle α between the horizontal and the HE line is expressed as

$$\alpha = \tan^{-1}(m_1), \tag{16}$$

while the nose-tilted angle β between the vertical and the VN lines is depicted by

$$\beta = \tan^{-1}(m_2) - 90. \tag{17}$$

If the angle of the cross is in range of $80^\circ \geq \theta \geq 100^\circ$, the rotation angle for face alignment is the average of eye-tilted and nose-tilted angles as

$$\theta_{rot} = (\alpha + \beta)/2. \tag{18}$$

By setting the nose position as the center, the face image is rotated by affine transform with θ_{rot} degrees and cropped. If the cross angle is out of $80^\circ \geq \theta \geq 100^\circ$, the rotation angle is determined either by eye-tilted angle or nose-tilted angle. If there more than two landmarks on the HE line, the rotation angle will be determined by eye-tilted angle, α, if not, the rotation angle becomes β, the nose-tilted angle.

As to rotation left and right deformations as depicted in Fig. 8c, d, the image faces slightly rotate toward the left and right directions, respectively. Figure 10 exhibits three top views of rotation faces. For the normal face, the VN line will evenly divide the HE line into two equal arms as shown in Fig. 10b. However, the right-rotation face will produce a longer right arm and a shorter left arm as shown in Fig. 10a while the left-rotation face will produce a shorter

Yang *et al. EURASIP Journal on Advances in Signal Processing* (2018) 2018:51

Page 8 of 13
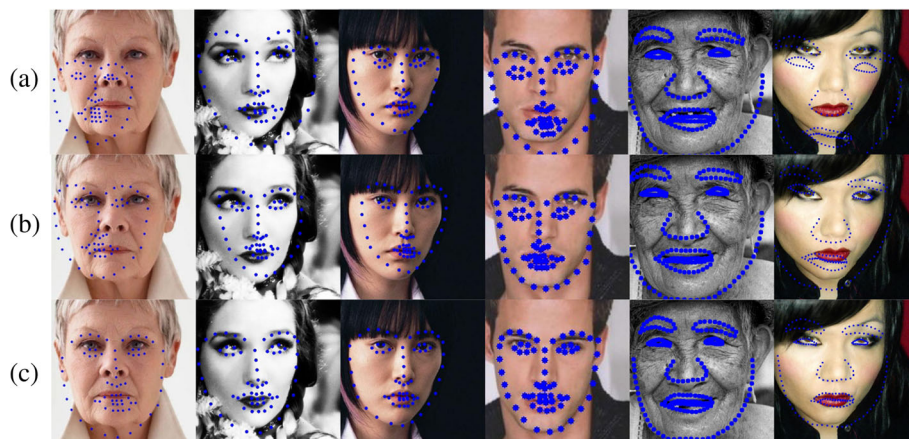
Fig. 13 Selected face alignment results by using **a** the ESR method (top row), **b** the MSR method with pixel and region difference features (bottom row), and **c** the MSR with pixel, region, and gradient difference features (final row)

right arm and a longer left arm as shown in Fig. 10c. For simplicity, we only allow the pointing face in + 6° and − 6°, The true angle warping angles, which is actually related to camera distance and focus length, with respect to the VN line, could be detected as − 6, − 3, 0, + 3 and + 6 by the ratio of segmented HE lengths separated by the VN line. If the face image is mixed with tilted and rotated variations, we should perform the adjustment of the tilt rotation first and then conduct the adjustment of rotation warping.

After warping transform of the face image, the facial image is adjusted to become a straight frontal face as Fig. 11a. Since the white (unknown) regions after affine transform are possibly yielded, the images are further cropped to 80% of the face image. Finally, the finally adjusted face image as shown in Fig. 11b will be used for face recognition.

3 Experimental results and discussion

For performance assessments of the proposed MSR face alignment, the experiments are divided into two main parts. For face alignment, the first part of simulations is performed to verify the alignment performance of the proposed MSR face alignment method while the second part is conducted to evaluate the recognition performance of alignment-based face recognition in use of the proposed MSR face alignment and cross warping methods.

3.1 Experiments for face alignment

In face alignment experiments, the proposed multi-feature shape regression (MSR), the explicit shape regression (ESR) [8], and the other face alignment methods are compared on the LFPW [29] and HELEN [30] face alignment databases. The LFPW database contains 792 facial images for the training phase and 220 facial images for the testing phase. These facial images were taken at different poses, facial expression, and head rotation. Each facial image has 68 landmarks which were annotated manually. The HELEN face database contains 1000 facial images for the training phase and 330 facial images for the testing phase. Each facial image contains 194 landmarks which were also annotated manually.

In order to evaluate the performances, the average landmark error and failure rate are the two important criteria to assess the face alignment algorithms. The average landmark error for all N testing images is defined as

$$\text{error} = \frac{1}{N}\sum_{n=1}^{N}\varepsilon^n, \text{with } \varepsilon^n$$
$$= \frac{1}{M}\sum_{m=1}^{M}\left(\frac{\left(x_m^n-\tilde{x}_m^n\right)^2}{w^n}+\frac{\left(y_m^n-\tilde{y}_m^n\right)^2}{h^n}\right)^{1/2}, \quad (19)$$

where (x_m^n, y_m^n) and $(\tilde{x}_m^n, \tilde{y}_m^n)$ respectively represent positions of the mth estimated landmark and the mth ground truth landmark, (w^n, h^n) is the image size of the nth image, and M denotes the number of landmarks. If the average of K

Table 4 Comparisons of different face alignment methods on FRGC database

Methods	LPCM	ERT	RCPR	SDM	ESR	MSR
Error	0.040	0.038	0.035	0.035	0.040	*0.033*

Table 5 Comparisons of different face alignment methods on HELEN database

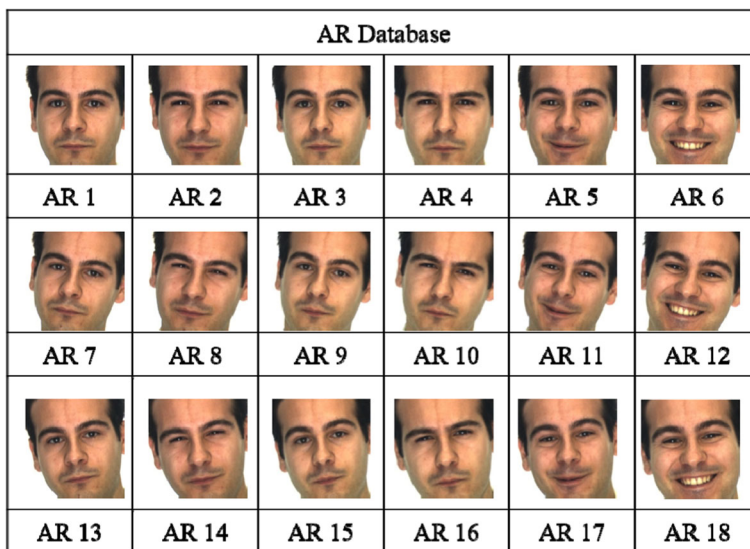| Methods	ERT	RCPR	SDM	ESR	MSR
Error	0.049	0.065	0.059	0.043	*0.038*

Fig. 14 Face images in AR database (AR1–6) and the synthesized images (AR7–18) for a sampled identify

landmarks of the testing image is more than 0.1, it will be treated as a fail case, and the number of the fail cases, f, is denoted as

$$f = \sum_{n=1}^{N} \delta(fail^n), \quad \delta(fail^n) = \begin{cases} 1, & \text{if } \varepsilon^n > 0.1, \\ 0, & \text{if } \varepsilon^n \le 0.1. \end{cases} \quad (20)$$

Thus, the failure rate is defined as

$$\text{failure rate } (\%) = \frac{f}{N} \times 100\%. \quad (21)$$

In addition, the experimental results for face alignment with AR and FRGC databases will also be presented. Since the two databases do not provide the ground truth shapes, we just can show some selected samples of facial images and their estimated shape.

The proposed multi-feature shape regression (MSR) method considers total differences of pixel difference (pd), region difference (rd), and gradient difference (gd). The three compositions of the multiple features for MSR are shown in Fig. 12. As shown in Tables 2 and 3, the landmark errors and failure rates by using different combinations of multiple features for the proposed MSR are tested on LFPW and HELEN databases, respectively. Some selected facial images with the detected landmarks by the proposed MSR methods and the ESR method are also shown in Fig. 13. Thus, the MSR face alignment method will use pixel difference (pd), region difference (rd), and gradient difference (gd) with 0.6, 0.3, and 0.1 weights for reminding simulations. For the comparisons of different face alignment methods, the LPCM (Localizing Parts of faces using a Consensus of Exemplars) [29], ERT (Ensemble of Regression Trees) [31], RCPR (Robust

Cascaded Pose Regression) [32], and SDM (Supervised Descent Method) [33] are shown in Tables 4 and 5. The results show that the proposed MSR is better than other methods.

3.2 Experiments for face recognition

For face recognition experiments on AR database [34], we select 100 subjects as shown in Fig. 14, which are used for performance evaluation. Each subject contains
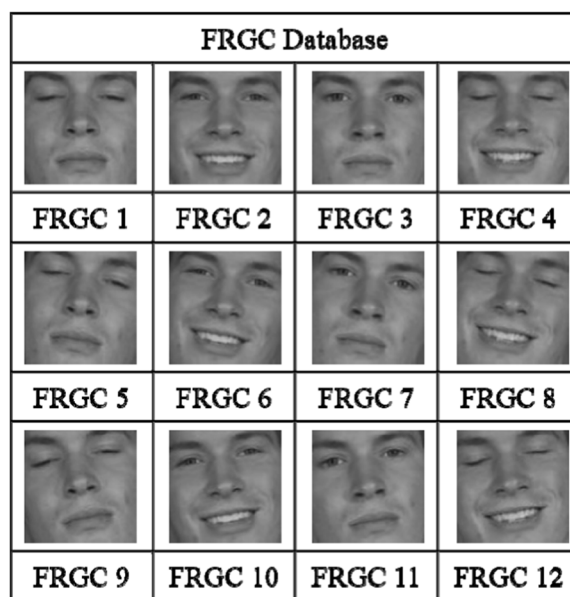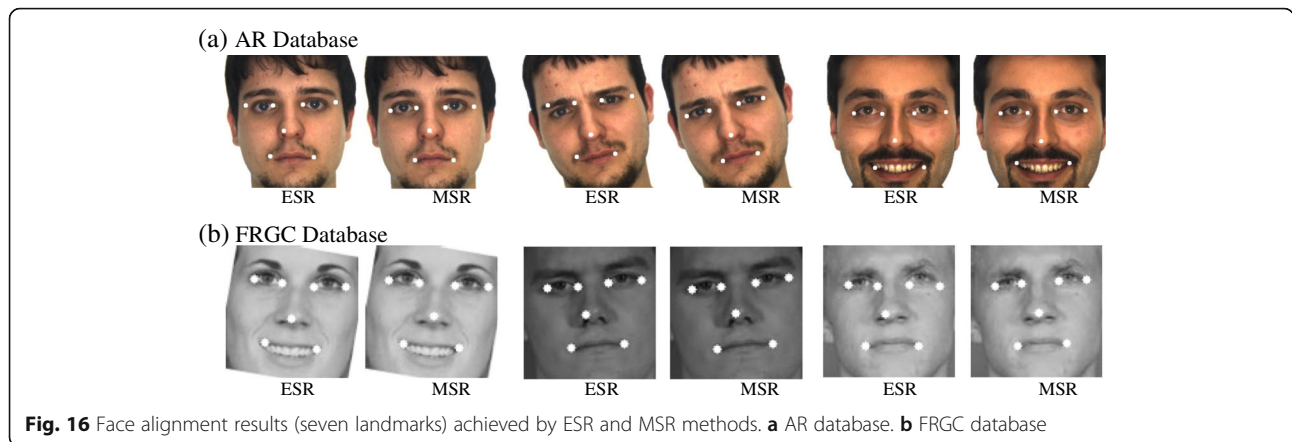

Fig. 15 Face images in FRGC database (FRGC 1–4) and the synthesized images (FRGC 5–12) for a sampled identify

Yang *et al. EURASIP Journal on Advances in Signal Processing* (2018) 2018:51

Page 10 of 13

Fig. 16 Face alignment results (seven landmarks) achieved by ESR and MSR methods. **a** AR database. **b** FRGC database

18 images, where AR1–AR6 face images are the original images, while AR7–AR18 are the synthesized ones. In face recognition experiments on FRGC database [35], as shown in Fig. 15, we also pick 100 subjects, which are used for performance evaluation. Each subject contains 12 images, where FRGC1–FRGC4 face images are the original images, while FRGC5–FRGC12 images are the synthesized ones. Each facial image is downsampled to 20 × 20 pixels.

To validate the proposed alignment-based face recognition system, the recognition performances achieved by the different algorithms will be simulated. The other face recognition algorithms used in the experiments include principal component analysis (PCA) [10, 11], linear discriminant analysis (LDA) [15], linear regression classification (LRC) [22], modular linear regression-based classification (MLRC) [22], sparse representation classification (SRC) [36, 37], locality preserving projection (LPP) [38], neighboring preserving embedding (NPE) [39], improved principal component regression (IPCR) [40], unitary regression classification (URC) [41], linear discriminant regression classification (LDRC) [42], and kernel linear regression classification (KLRC) [43] methods. From Fig. 14, six original face images, AR1, AR3, AR4, and AR5 for each identity are used for training while two original images AR2 and AR6 and four synthesized images are randomly selected for testing. From Fig. 15, three original face images, FRGC2, FRGC3, and FRGC4, for each identity are used for training while the original image FRGC1 and two synthesized images are randomly selected for testing.

In face recognition experiments, the abovementioned face recognition algorithms are compared in three categories: (1) without alignment, (2) with ESR alignment, and (3) with MSR alignment. After face alignment by using the ESR and MSR methods, the face images both are adjusted by using the proposed conditional cross warping method for fair comparisons. Figure 16 shows

the detected seven landmarks of some tested (normal and synthesized) images achieved by the ESR and MSR methods. The results also show that the MSR method has higher precision than the ESR method in landmark estimation on AR and FRGC databases.

If the testing face images are the normal face images, Tables 6 and 7 show the recognition performances on AR (AR2, AR6) and FRGC (FRGC1) databases, respectively. For the normal face images, it is noted that the ESR and the proposed MSR methods without any prior knowledges will still perform face alignment and face warping processes. The recognized results show that the proposed alignment-based face recognition systems are quite reliable while the proposed MSR shows better than the ESR method. For posed face images (synthesized face images), Tables 8 and 9 show the recognized rates on AR and FGGC databases, respectively. The simulation results show that the proposed MSR face alignment and

Table 6 Recognition performances (%) on AR database (normal faces)

Method	Alignment		
	Without alignment	Alignment by ESR	Alignment by MSR
PCA	85.00	86.67	91.00
LDA	98.75	95.00	97.00
LRC	97.00	97.50	98.00
MLRC	95.00	93.00	95.00
SRC	98.50	99.00	98.50
LPP	82.00	81.50	83.00
NPE	90.50	90.50	92.50
IPCR	97.00	96.50	96.50
URC	99.00	98.00	99.00
LDRC	97.00	97.00	96.00
KLRC	97.00	95.00	96.00
Average	94.25	93.61	94.77

Yang *et al. EURASIP Journal on Advances in Signal Processing* (2018) 2018:51

Page 11 of 13

Table 7 Recognition performances (%) on FRGC database (normal faces)

Method	Alignment		
	Without alignment	Alignment by ESR	Alignment by MSR
PCA	*98.00*	97.00	*98.00*
LDA	*99.00*	*99.00*	*99.00*
LRC	*98.00*	96.00	*98.00*
MLRC	*98.00*	95.00	96.00
SRC	*98.00*	96.00	*98.00*
LPP	*94.00*	90.00	88.00
NPE	*98.00*	96.00	97.00
IPCR	*98.00*	94.00	96.00
URC	*98.00*	*100.00*	98.00
LDRC	*97.00*	93.00	91.00
KLRC	*98.00*	97.00	*98.00*
Average	*97.63*	95.73	96.09

Table 9 Recognition rates on FRGC with different face recognition algorithms (synthesized faces)

Method	Alignment		
	Without alignment	Alignment by ESR	Alignment by MSR
PCA	69.00	*91.00*	*91.00*
LDA	54.50	76.50	*83.00*
LRC	43.50	80.50	*85.50*
MLRC	50.00	60.50	*62.00*
SRC	61.00	*85.50*	84.00
LPP	22.00	47.00	*50.50*
NPE	46.50	79.00	*80.00*
IPCR	41.00	*76.00*	75.50
URC	49.00	81.50	*85.00*
LDRC	6.50	37.00	*37.50*
KLRC	46.50	*76.50*	76.00
Average	44.50	71.91	*73.64*

conditional cross warping processes can effectively overcome the problems of pose variations. The proposed MSR method achieves better performances than the ESR method not only in face alignment but also in face recognition. Among all face recognition algorithms, the SRC and URC methods in conjunction with the proposed alignment-based face recognition system perform better than the other face recognition methods.

4 Conclusions

In this paper, the multi-feature shape regression (MSR) method, which considers pixel difference, region difference, and gradient difference together, is first proposed. For face

Table 8 Recognition performances (%) on FRGC database (Synthesized Faces)

Method	Alignment		
	Without alignment	Alignment by ESR	Alignment by MSR
PCA	46.25	78.29	80.00
LDA	55.00	89.08	*90.75*
LRC	48.00	88.75	*89.50*
MLRC	28.75	74.25	*76.50*
SRC	50.25	91.75	*92.75*
LPP	30.50	62.25	*64.25*
NPE	40.75	*80.00*	78.00
IPCR	43.00	84.75	*85.75*
URC	64.50	91.75	*93.50*
LDRC	27.25	76.00	*76.25*
KLRC	41.50	*88.00*	86.25
Average	43.25	82.26	*83.05*

recognition applications, a cross warping method is suggested to achieve alignment-based face recognition. The proposed MSR face alignment method can help to precisely estimate seven key landmarks of face images. Simulation results show that the multi-feature shape regression (MSR) method, which utilizes more features computed from surrounding pixels, shows better alignment performance than the explicit shape regression (ESR) algorithm, which only uses pixel difference. With seven selected face key landmarks, including four eye canthi, one nose tip, and two mouth corners, we can use the positions of seven landmarks to find a cross shape, which is defined by the estimated horizontal-eye (HE) and vertical-nose (VN) lines. By the cross warping process, we can adjust the tilted face image back to normal face image to overcome the problem of pose variations for face recognition. The experimental results show that the MSR method performs better than the ESR and other face alignment algorithms on face alignment database. For alignment-based face recognition, the MSR face alignment algorithm with the cross warping method can help the SPO face recognition methods to achieve better recognition performances. Simulation results show that the proposed multi-feature shape regression (MSR) face alignment method achieves better performances in both face alignment and face recognition than the existing face alignment methods.

Abbreviations
AAM: Active appearance model; AR: Argument reality; ASM: Active shape model; ESR: Explicit shape regression; IPCR: Improved principal component regression; KLDA: Kernel LDA; LDA: Linear discriminant analysis; LPP: Locality preserving projection; LRC: Linear regression classification; MLRC: Modular linear regression classification; MSR: Multi-feature shape regression; NPE: Neighboring preserving embedding; PCA: Principal component analysis; SPO: Subspace projection optimizations; URC: Unitary regression classification; VR: Virtual reality

Availability of data and materials
The face alignment data is obtained from the LFPW and HELEN face alignment databases provided in [29, 30], respectively. The face recognition data is retrieved from the AR and FRGC databases delivered in [34, 35], respectively. As to the augment face images, the datasets generated for the current study are available from the corresponding author on reasonable request.

Authors' contributions
W-JY carried out image processing studies, participated in the proposed system, assembled formulations, and drafted the manuscript. Y-CC carried out software simulations and face data augmentation by warping parameters. P-CC and J-FY conceived of the study, participated in its design and coordination, and helped to draft the manuscript. All authors read and approved the final manuscript.

Authors' information
W-J Yang received a B.S. degree in Computer Science from Tunghai University, Taiwan, in 2012 and an M.S. degree in Computer Science and Information Engineering from National University of Tainan, Taiwan, in 2015. Currently, he is a Ph.D. student with the Graduate Institute of Computer and Communication Engineering in National Cheng Kung University, Taiwan. His current research interests include pattern recognition, machine learning, and deep learning for designs of smart systems.
Y-C Chen received a B.S. degree in Electrical Engineering and an M.S. degree in Computer and Communication Engineering from the National Cheng Kung University, Tainan, Taiwan, in 2014 and 2016, respectively. Her current research interests include face recognition and machine learning.
P-C Chung received a Ph.D. degree in Electrical Engineering from Texas Tech University, Lubbock, TX, USA, in 1991. She was with the Department of Electrical Engineering, National Cheng Kung University (NCKU), Tainan, Taiwan, in 1991 and became a Full Professor in 1996. She applies most of her research results to healthcare and medical applications. Dr. Chung is a member of the Phi Tau Phi Honor Society, was a member of the Board of Governors of CAS Society from 2007 to 2009 and from 2010 to 2012, and is currently an ADCOM Member of the IEEE CIS and the Chair of CIS Distinguished Lecturer Program. She also is an Associate Editor of IEEE Transaction on Neural Networks and the Editor of Journal of Information Science and Engineering, the Guest Editor of Journal of High Speed Network, the Guest Editor of IEEE Transaction on Circuits and Systems-I, and the Secretary General of Biomedical Engineering Society of China. She is one of the Co–Founders of Medical Image Standard Association (MISA) in Taiwan and is currently on the Board of Directors of MISA. Her research interests include image/video analysis and pattern recognition, bio signal analysis, computer vision, and computational intelligence. She is an IEEE fellow.
J-F Yang received a Ph.D. degree in Electrical Engineering from the University of Minnesota, Minneapolis, MN, USA, in 1988. He joined the National Cheng Kung University (NCKU), Taiwan, in 1988 and was promoted to Distinguished Professor in 2004. Dr. Yang was the Distinguished Lecturer in the Program by the IEEE Circuits and Systems Society (CAS) from 2004 to 2005. He was the Chair of the IEEE CAS Multimedia Systems and Applications Technical Committee from 2008 to 2009. He was an Associate Editor of IEEE Transaction on Circuits and Systems for Video Technology and EURASIP Journal of Advances in Signal Processing. He is an IEEE Fellow. Currently, he is an Associate Editor of IET Signal Processing. He was a recipient of the NSC Excellent Research Award in Taiwan in 2008. He has published over 135 journals and 216 conference papers. Currently, his research interests include multimedia processing, coding, and recognition.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References
1. TF Cootes, CJ Taylor, in *Proc of the British Machine Vision Conference*. Active shape models—'smart snakes' (1992), pp. 266–275
2. D Cristinacce, TF Cootes, in *Proc of the British Machine Vision Conference*. Boosted regression active shape models (2007)
3. TF Cootes, GJ Edwards, CJ Taylor, in *European Conference on Computer Vision*. Active appearance models (1998)
4. I Matthews, S Baker, Active appearance models revisited. Int. J. Comput. Vis. **60**(2), 135–164 (2004)
5. P Sauer, TF Cootes, CJ Taylor, in *Proc of the British Machine Vision Conference*. Accurate regression procedures for active appearance models (2011)
6. J Saragih, R Goecke, in *Proc. of IEEE 11th International Conference on Computer Vision*. A nonlinear discriminative approach to AAM fitting (2007)
7. P Dollár, P Welinder, P Perona, in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*. Cascaded pose regression (2010)
8. M Valstar, B Martinez, X Binefa, in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*. Facial point detection using boosted regression and graph models (2010)
9. X Cao, Y Wei, F Wen, J Sun, Face alignment by explicit shape regression. Int. J. Comput. Vis. **107**(2), 177–190 (2014)
10. M Turk, A Pentland, Eigenfaces for recognition. J. Cogn. Neurosci. **3**(1), 71–86 (1991)
11. P. N. Belhumeur, ,J. P. Hespanha, and. D. J. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection IEEE Trans. Pattern Anal. Mach. Intell., vol. 19, no. 7, pp. 711–720, 1997.
12. B Moghaddam, A Pentland, Probabilistic visual learning for object representation. IEEE Trans. Pattern Anal. Mach. Intell. **19**(7), 696–710 (1997)
13. J Yang, D Zhang, AF Frangi, J-Y Yang, Two-dimensional PCA: a new approach to appearance-based face representation and recognition. IEEE Trans. Pattern Anal. Mach. Intell. **26**(1), 131–137 (2004)
14. AM Martínez, AC Kak, PCA versus LDA. IEEE Trans. Pattern Anal. Mach. Intell. **23**(2), 228–233 (2001)
15. J Shawe-Taylor, N Cristianini, *Kernel methods for pattern analysis* (Cambridge University Press, Oxford, 2004)
16. B Schölkopf, A Smola, K-R Müller, Nonlinear component analysis as a kernel eigenvalue problem. Neural Comput. **10**(5), 1299–1319 (1998)
17. M-H Yang, in *Proc. of the Fifth International Conference on Automatic Face and Gesture Recognition*. Kernel eigenfaces vs. kernel fisherfaces: face recognition using kernel methods (2002)
18. B. Scholkopft and K.-R. Mullert, Fisher discriminant analysis with kernels Neural networks for signal processing IX, 1 1 1999.
19. G Baudat, F Anouar, Generalized discriminant analysis using a kernel approach. Neural Comput. **12**(10), 2385–2404 (2000)
20. J Lu, KN Plataniotis, AN Venetsanopoulos, Face recognition using kernel direct discriminant analysis algorithms. IEEE Trans. Neural Netw. **14**(1), 117–126 (2003)
21. J Huang, PC Yuen, WS Chen, JH Lai, Choosing parameters of kernel subspace LDA for recognition of face images under pose and illumination variations. IEEE Trans. Syst. Man Cybern. B Cybern. **37**(4), 847–862 (2007)
22. I Naseem, R Togneri, M Bennamoun, Linear regression for face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **32**(11), 2106–2112 (2010)
23. Y Duan, J Lu, J Feng, J Zhou, Context-aware local binary feature learning for face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **40**(5), (2018)
24. W-Y Liu, Y-D Wen, Z-D Yu, M Li, B Raj, L Song, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. SphereFace: deep hypersphere embedding for face recognition (2017)
25. W Wu, M Kan, X Liu, Y Yang, S Shan, X Chen, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Recursive spatial transformer (rest) for alignment-free face recognition (2017)
26. N Duffy, D Helmbold, Boosting methods for regression. Mach. Learn. **47**(2), 153–200 (2002)
27. JH Friedman, Greedy function approximation: a gradient boosting machine. Ann. Stat. **29**(5), (2001)
28. P Viola, MJ Jones, Robust real-time face detection. Int. J. Comput. Vis. **57**(2), 137–154 (2004)

Yang *et al. EURASIP Journal on Advances in Signal Processing* (2018) 2018:51

Page 13 of 13

29. PN Belhumeur, DW Jacobs, DJ Kregman, N Kumar, Localizing parts of faces using a consensus of exemplars. IEEE Trans. Pattern Anal. Mach. Intell. **35**(12), 2930–2940 (2013)
30. V Le, J Brandt, Z Lin, L Bourdev, JS Huang, in *Proc. of European Conference on Computer Vision*. Interactive facial feature localization (2012)
31. V Kazemi, J Sullivan, in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*. One millisecond face alignment with an Ensemble of Regression Trees (2014)
32. XP Burgos-Artizzu, P Perona, P Dollár, in *Proc. of the IEEE International Conference on Computer Vision*. Robust face landmark estimation under occlusion (2013)
33. X Xiong, F De la Torre, in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*. Supervised descent method and its applications to face alignment (2013)
34. AM Martinez, in *CVC Technical Report*. The AR face database, vol 24 (1998)
35. PJ Phillips, FJ Flynn, T Scruggs, KW Bowyer, J Chang, K Hoffman, J Marques, J Ming, W Worek, in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. Overview of the face recognition grand challenge (2005)
36. J Wright, A-Y Yang, A Ganesh, Robust face recognition via sparse representation. IEEE Trans. Pattern Anal. Mach. Intell. **31**(2), 210–227 (2009)
37. X Jiang, J Lai, Sparse and dense hybrid representation via dictionary decomposition for face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **37**(5), 1067–1079 (2015)
38. X He, S Yan, Y Hu, P Niyogi, J-J Zhang, Face recognition using Laplacianfaces. IEEE Trans. Pattern Anal. Mach. Intell. **27**(3), 328–340 (2005)
39. X He, D Cai, Y Yang, H-J Zhang, in *Proc. of Tenth IEEE International Conference on Computer Vision (ICCV'05)*. Neighborhood preserving embedding, vol 1 (2005)
40. S-M Huang, J-F Yang, Improved principal component regression for face recognition under illumination variations. IEEE Sig. Process. Lett. **19**(4), 179–182 (2012)
41. S-M Huang, J-F Yang, Unitary regression classification with total minimum projection error for face recognition. IEEE Sig. Process. Lett. **20**(5), 443–446 (2013)
42. S-M Huang, J-F Yang, Linear discriminant regression classification for face recognition. IEEE Sig. Process. Lett. **20**(1), 91–94 (2013)
43. Y-T Chou, S-M Huang, J-F Yang, Class-specific kernel linear regression classification for face recognition under low-resolution and illumination variation conditions. EURASIP J. Adv. Sig. Process. https://doi.org/10.1186/s13634-016-0328-0