

RESEARCH

Open Access

Non-frontal facial expression recognition based on salient facial patches



Bin Jiang^{1*} , Qiuwen Zhang¹, Zuhe Li¹, Qinggang Wu¹ and Huanlong Zhang²

* Correspondence: jiangbin@zzuli.edu.cn

¹College of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, People's Republic of China
Full list of author information is available at the end of the article

Abstract

Methods using salient facial patches (SFPs) play a significant role in research on facial expression recognition. However, most SFP methods use only frontal face images or videos for recognition, and they do not consider head position variations. We contend that SFP can be an effective approach for recognizing facial expressions under different head rotations. Accordingly, we propose an algorithm, called profile salient facial patches (PSFP), to achieve this objective. First, to detect facial landmarks and estimate head poses from profile face images, a tree-structured part model is used for pose-free landmark localization. Second, to obtain the salient facial patches from profile face images, the facial patches are selected using the detected facial landmarks while avoiding their overlap or the transcending of the actual face range. To analyze the PSFP recognition performance, three classical approaches for local feature extraction, specifically the histogram of oriented gradients (HOG), local binary pattern, and Gabor, were applied to extract profile facial expression features. Experimental results on the Radboud Faces Database show that PSFP with HOG features can achieve higher accuracies under most head rotations.

Keywords: Facial expression recognition, Salient facial patch, Head rotation

1 Introduction

The problem of determining how to use face information in human–computer interaction has been investigated for several years. An increasing number of applications that employ facial recognition technology have emerged. However, current studies on facial expression recognition have yet to be fully and practically applied. Variations in head pose constitute one of the main challenges in the automatic recognition of facial expressions [1]. This problem arises when inadvertent or deliberate occlusions occur, which can obstruct nearly half of the face under large head pose changes. Automatically analyzing facial expressions from the pose-free human face is required to establish a technological framework for further research.

Recognition of profile facial expressions was first achieved by Pantic et al. [2]. They used particle filtering to track 15 facial landmarks in a sequence of face profiles, and an 87% recognition rate was achieved. Although only -90° face image sequences were used as experimental data, their work inspired further research. Hu et al. [3] are credited to be first to have researched the recognition of multi-view facial expressions.

Their experimental data included an increased number of subjects (100), six emotions with four intensity levels, and five viewing angles (0°, 30°, 45°, 60°, and 90°). The authors first calculated the geometric features of the facial components and then exploited five classifiers to recognize emotion features. Experimental results demonstrated that good recognition can be achieved on profile face images.

Moreover, Dapogny et al. [4] used spatio-temporal features to recognize facial expressions under head pose variations in videos. Zheng et al. [5] used additional head variations for face images and proposed a discriminant analysis algorithm to recognize facial expressions from pose-free face images. They chose 100 subjects from the BU-3DFE database [6]. The experimental results demonstrated that their algorithm could achieve satisfactory performance on subjects with a head pose under yaw or pitch. However, the face images with large pose variations yielded the lowest average recognition rate. Wu et al. [7] proposed a model called the locality-constrained linear coding-based bi-layer model. The head poses are estimated in the first layer. Then, the facial expression features are extracted using the corresponding view-dependent model in the second layer. This model improved recognition on face images with large pose variations. Lai et al. [8] presented a multi-task generative adversarial network to solve the problem of emotion recognition under large head pose variations. Mao et al. [9] considered the relationships between head poses and proposed a pose-based hierarchical Bayesian-themed model. Jampour et al. [10] found that linear or nonlinear local mapping methods provide more reasonable results for multi-pose facial expression recognition than global mapping methods.

Despite the above advancements in constructing models or functions for mapping the relationship between frontal and non-frontal face images, the feature point movements and texture variations are considerably more complex under head pose variations and identity biases. An effective feature extraction method is thus necessary for recognizing non-frontal facial expressions. Recently, a method based on salient facial patches, which seeks salient facial patches from the human face and extracts facial expression features from these patches, has played a significant role in emotion recognition [11–19]. In this method, select facial patches (e.g., eyebrows, eyes, cheeks, and mouth) are considered the key regions of face images, and the discriminative features are extracted from salient regions. The extracted features are instrumental in distinguishing one expression from another. Furthermore, the salient facial patches foster favorable conditions for non-frontal facial expression recognition. We therefore propose an algorithm based on salient facial patches that recognizes facial expressions from non-frontal face images. This method, called profile salient facial patches (PSFP), detects salient facial patches from non-frontal face images and recognizes facial expressions from these patches.

The remainder of this paper is organized as follows. Related work is described in the second section, and the details of PSFP are presented in the Method section. The design and analysis of experiments that validate the proposed approach are described in the Results and discussion section. Finally, conclusions are provided in the last section.

2 Related work

Sabu and Mathai [11] were the first to investigate the importance of algorithms based on salient facial patches for facial expression recognition. They found that,

to date, the most accurate, efficient, and reproducible system for facial expression recognition using salient facial patches was designed by Happy and Routray [12]. However, the salient regions can vary in different facial expressions and result in face deformation. Chitta and Sajjan [13] found that the most effective salient facial patches are located mainly in the lower half of the face. Thus, they reduced the salient region and extracted the emotion features from the lower face. However, their algorithm did not achieve high recognition rates in experiments. Zhang et al. [14] used a sparse group lasso scheme to explore the most salient patches for each facial expression, and they combined these patches into the final features for emotion recognition. They achieved an average recognition rate of 95.33% on the CK+ database. Wen et al. [15] used a convolutional neural network (CNN) [20] to train the salient facial patches on face images. A secondary voting mechanism trains the CNN to determine the final categories of test images. Sun et al. [16] presented a CNN that uses a visual attention mechanism and can be applied for facial expression recognition. This mechanism focuses on local areas of face images and determines the importance of each region. In particular, whole face images with different poses are used for CNN training. Yi et al. [17] expanded the salient facial patches from static images to video sequences. They used 24 feature points to show the deformation in facial geometry throughout the entire face. Yao et al. [18] presented a deep neural network classifier that can capture pose-variant expression features from depth patches and recognize non-frontal expressions. Barman and Dutta [19] used an active appearance model [21] to detect the salient facial landmarks, whose connections form triangles that can be deemed salient facial regions. The geometric features are extracted from the face for emotion recognition.

Given the above background, the following commonalities in facial expression recognition are identified:

1. Most existing methods are used on frontal face images.
2. There are three main components of salient facial regions: eyes, nose, and lips.
3. The appearances or texture features are crucial for recognizing facial expressions.

We contend that the salient facial patches method should be applied for both frontal and non-frontal facial expression recognition. Inspired by the method of Happy et al. [12], we designed PSFP for non-frontal facial expression recognition. Unlike previous non-frontal facial expression recognition methods, this method employs salient facial patches, which are composed mainly of the facial components that provide ample facial expression information under head pose variations. Thus, it can extract many appearance or texture features under these variations and identity biases. Furthermore, PSFP does not require the construction of a complex model for multi-pose facial expression classification. The PSFP details are presented in the following sections.

3 Method

There are three main steps in the non-frontal facial expression recognition system: face detection, feature extraction, and feature classification. The accurate detection of facial landmarks can improve the localization of salient facial patches on the non-frontal face images. Therefore, localization of fiducial facial points and estimation of the head pose are essential steps for identifying the salient facial patches. The head pose may be a

combination of different directions in a three-dimensional space. If the face detection method cannot obtain adequate information regarding the head rotations, the facial expression recognition rate will be low. In the methods of Jin and Tan [22], the tree-structured part model employs a unified framework to detect the human face and estimate head variations. This approach is highly suitable for non-frontal facial expression recognition. Thus, we adopt Yu et al.’s method [23] in our system for face detection and head pose estimation. Because this algorithm can estimate the head poses in pitch, yaw, and roll directions, it is adequate to detect the head poses and positions of human faces.

3.1 Face detection

To simultaneously detect the human face and track facial feature points, Yu et al. [23] presented a united framework. They define a “part” at each facial landmark and use global mixtures to model topological changes due to viewpoint variations. The different mixtures of the tree-structured model employ a shared pool of part templates, V . For each viewpoint i , $i \in (1, 2, \dots, M)$, they define N-node tree $T_i = (V_i, E_i)$, $V_i \subseteq V$. The connection between the two parts forms an edge in E_i . There are two main steps in their framework:

- (1) Initialization. For each viewpoint i , the measuring of landmark configuration $s = (s_1, s_2, \dots, s_N)$ is defined by scoring function f :

$$\begin{aligned}
 f_i(I, s) &= \sum_{j \in V_i} q_i(I, s_j) + \sum_{(j,k) \in E_i} g_i(s_j, s_k) \\
 s^* &= \arg \max_{i \in (1, 2, \dots, M)} f_i(I, s)
 \end{aligned}
 \tag{1}$$

where the first term uses local patch appearance evaluation function $q_i(I, s_j) = \langle w_j^{iq}, \Phi_j^{iq}(I, s_j) \rangle$, which indicates whether a facial landmark $s_j = (x_j, y_j)$, $j \in (1, 2, \dots, N)$ may occur at the aligned position in face image I . The second term uses shape deformation cost $g_i(s_j, s_k) = \langle w_{jk}^{ig}, \Phi_{jk}^{ig}(s_j, s_k) \rangle$, which maintains the balance of the relative locations of neighboring facial landmarks s_j and s_k . w_j^{iq} denotes the weight vector convolving the feature descriptor of patch j , $\Phi_j^{iq}(I, s_j)$. w_{jk}^{ig} are the weights controlling the shape displacement function, which is defined as $\Phi_{jk}^{ig}(s_j, s_k) = (dx, dy, dx^2, dy^2)$, $(dx, dy) = s_k - s_j$. The largest score may provide the most likely localization of the landmarks. Thus, the landmark positions can be obtained by maximizing scoring function f in Eq. 1. A group sparse learning algorithm [24] can be used to select the most salient weights, thereby forming a new tree.

- (2) Localization. Once the initial facial landmarks, s , have been detected, Procrustes analysis is employed to project the 3D reference shape model onto a 2D face image. $s = \bar{s} + Q \times u$ represents face shapes by mean shape \bar{s} and a linear combination of selected shape basis Q , and u is the coefficient vector.

Hence, the relationship is established between any two points in 3D space in Eq. 2.

$$s_j = a \times R \times s + T \tag{2}$$

where s_j is one of the defined landmarks, a denotes a scaling factor, R represents a rotation matrix, and T is the shift vector. The problem is to find such parameter, $\mathcal{P} = \{a, R, u, T\}$, to map the 3D reference shape to a fitted shape that best depicts the faces in an image.

Based on this probabilistic model, a two-step cascaded deformable shape model [23] is proposed to refine the facial landmark locations.

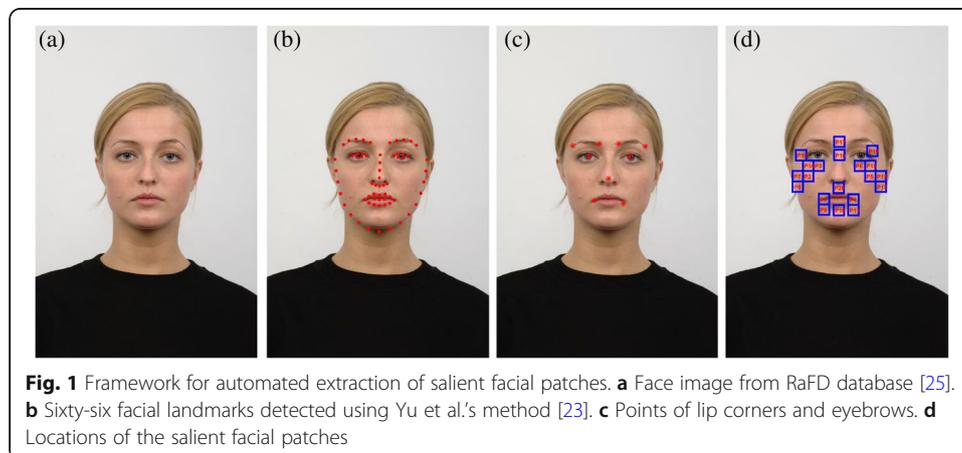
$$s^* = \arg \max_s p(s|\{v_i = 1\}_1^N, I) \tag{3}$$

$$\propto \arg \max_s p(s)p(\{v_i = 1\}_{i=1}^n | s, I) \tag{4}$$

$$= \arg \max_{\mathcal{P}} p(\mathcal{P}) \prod_{i=1}^n p(v_i = 1 | s_i, I) \tag{5}$$

In Eq. 3, vector $v = \{v_1, \dots, v_N\}$ indicates the likelihood of alignment in face image I . Here, $v = 1$ indicates that the facial landmarks are well aligned, and $v = 0$ indicates the opposite. Thus, Eq. 3 aims to maximize the likelihood of an alignment. Then, the Bayesian rule is used to derive Eq. 4. Hence, in Eq. 5, we know that parameter \mathcal{P} can determine 3D shape model s , $p(\mathcal{P}) = p(s)$. We suppose that $p(\mathcal{P})$ obeys the Gaussian distribution. In addition, logistic regression is used to represent the likelihood, $p(v_i = 1 | s_i, I) = \frac{1}{\exp(\vartheta\phi + b)}$, where ϕ is the local binary pattern (LBP) feature of facial landmark patch i , and parameters ϑ and b represent two regression weights that are trained from collected positive and negative samples.

Finally, the landmarks can be tracked and presented as $s_i = (x_i, y_i)$, $i = 1, 2, \dots, 66$. The locations of the landmarks for an image, such as that shown in Fig. 1a can be depicted as in Fig. 1b



3.2 Extraction of pose-free salient facial patches

The special salient facial patches are obtained from the face images according to the head pose. From an analysis of related work, we find that the eyes, nose, and lips are important facial components of the salient facial patches. The locations of these facial components for an image such as Fig. 1a can be shown as in Fig. 1c. The salient facial patches A_i can be extracted around the facial parts and the areas of the eyebrow, eye, nose, and lips:

$$A_i = \begin{bmatrix} \left(x_i - \frac{M}{2} + 1, y_i - \frac{N}{2} + 1\right) & \cdots & \left(x_i - \frac{M}{2} + 1, y_i + \frac{N}{2}\right) \\ \vdots & \ddots & \vdots \\ \left(x_i + \frac{M}{2}, y_i - \frac{N}{2} + 1\right) & \cdots & \left(x_i + \frac{M}{2}, y_i + \frac{N}{2}\right) \end{bmatrix} \quad (6)$$

where point $s_i = (x_i, y_i)$ is the center of A_i , and $M \times N$ is the size of A_i . If L salient facial patches have been selected from image R , the facial expression features will be extracted from L salient facial patches:

$$R_i = (A_1, A_2, \dots, A_L), i = 1, 2, \dots, k \quad (7)$$

where k is the number of images. The locations of 19 salient facial patches on a frontal face image are shown in Fig. 1d.

The rationale behind choosing the 20 given patches is based on the following facial action coding system. P_1 and P_4 are located at the lip corners, and P_9 and P_{11} are just below them. P_{10} is at the midpoint of P_9 and P_{11} , and P_{20} is at the upper lip. P_{16} is situated at the center of the two eyes, and P_{17} is at the center of inner brow. P_{15} and P_{14} are below the left and right eyes, respectively. P_3 and P_6 are respectively located at the middle of the nose and between the eyes. P_5 , P_{13} , and P_{12} were extracted from the left side of the nose and are stacked together; P_2 , P_7 , and P_8 are at the right side of the nose; and P_{18} and P_{19} are located on the respective outer eye corners.

The method of selecting facial patches in PSFP is similar to that in Happy et al., with two exceptions. The first difference is that the salient facial patches (SFP) method in Happy et al., which extracts facial expression features from salient facial patches, can only be used for frontal facial expression recognition; the face detection method is not applied for large head pose variations. As our method aims to recognize non-frontal facial expressions, the 66 facial landmarks are determined using Yu et al.'s method from face images with different head poses.

The second difference is the positions of P_{18} and P_{19} . When the face image is a frontal view, the Happy et al. method assigns the positions of these facial patches to the inner eyebrows, as shown in Fig. 2a (ours is shown in Fig. 2b). Two patches already exist at the inner eyebrows. Thus, if the patches are larger, they would likely overlap with those at the inner eyebrows. Moreover, Happy et al. do not consider the outer eye corner region.

When the image is a non-frontal facial view, the face will be partially occluded. Some patches may disappear under head pose variations. In such cases, the salient facial patches can be selected as shown in Fig. 3, and they are listed in Table 1.

As shown in Table 1, when the viewing angles increase from 0° to 90° , the number of patches decreases from 20 to 12. Thus, the feature dimensions of patches in the Happy et al. method are $19 \times M \times N$, whereas those in the PSFP algorithm are only $12 \times M \times$

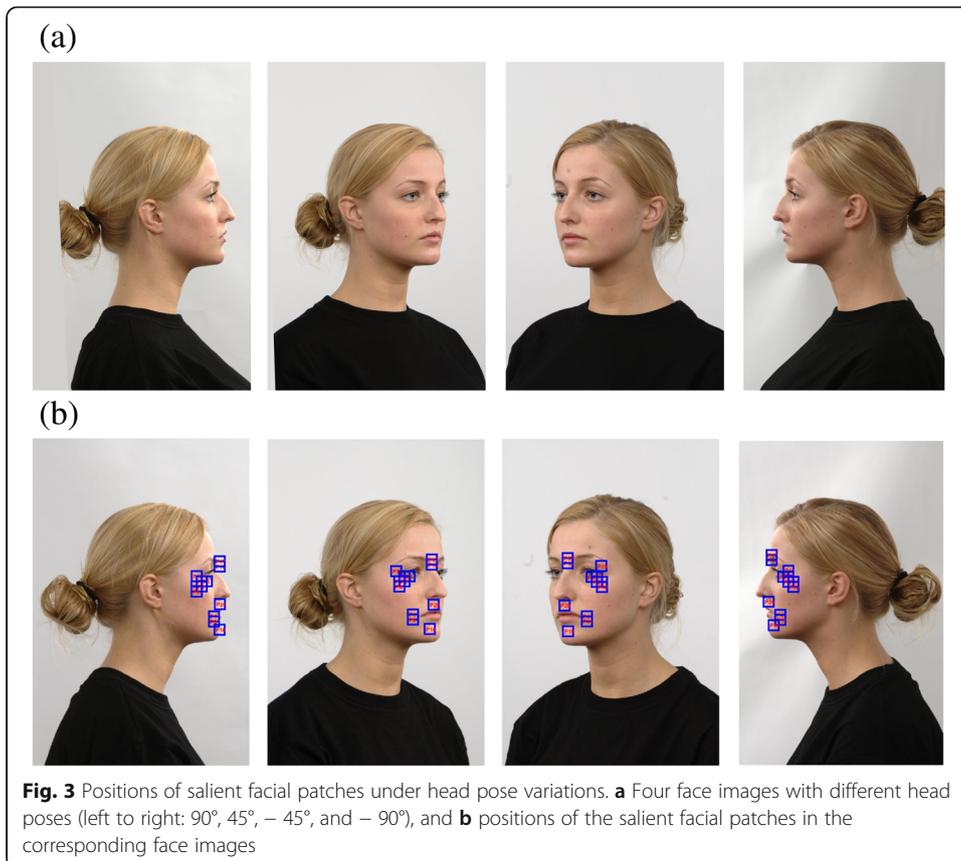
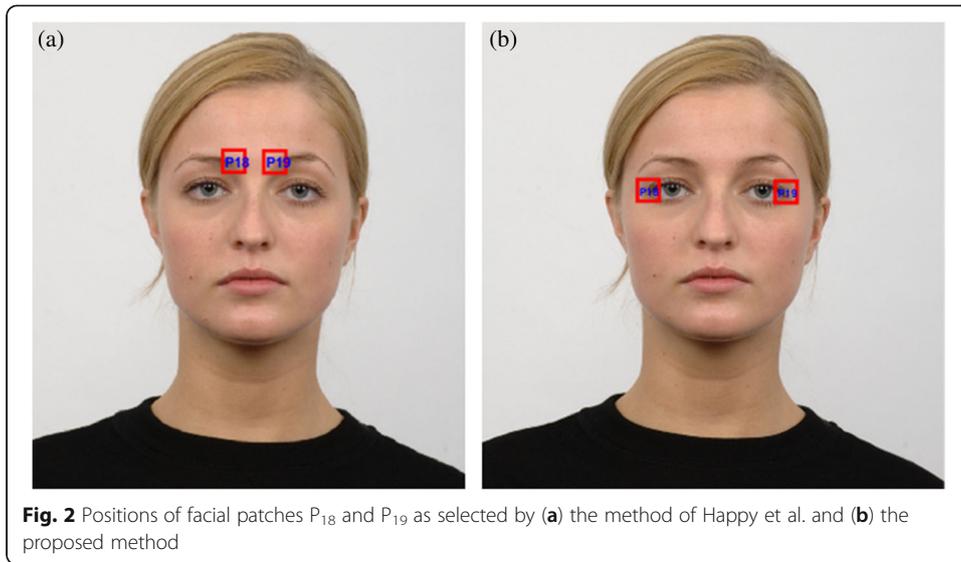


Table 1 Salient facial patches under different head poses

Head pose	Salient facial patches	Number of patches
90°	P ₁ , P ₂ , P ₃ , P ₇ , P ₈ , P ₉ , P ₁₀ , P ₁₄ , P ₁₆ , P ₁₇ , P ₁₈ , P ₂₀	12
45°	P ₁ , P ₂ , P ₃ , P ₇ , P ₈ , P ₉ , P ₁₀ , P ₁₄ , P ₁₆ , P ₁₇ , P ₁₈ , P ₂₀	12
0°	P ₁ –P ₂₀	20
– 45°	P ₄ , P ₅ , P ₆ , P ₁₀ , P ₁₁ , P ₁₂ , P ₁₃ , P ₁₅ , P ₁₆ , P ₁₇ , P ₁₉ , P ₂₀	12
– 90°	P ₄ , P ₅ , P ₆ , P ₁₀ , P ₁₁ , P ₁₂ , P ₁₃ , P ₁₅ , P ₁₆ , P ₁₇ , P ₁₉ , P ₂₀	12

N for non-frontal face images. Furthermore, we determined that the PSFP algorithm incurs a lower computational cost and has a time complexity of $O(2n \log 2n)$.

3.3 Feature extraction and classification

After the salient facial patches are obtained from the face images, the facial patch features must be extracted for classification. After these features are obtained, a representative classifier is applied for facial expression classification.

3.3.1 Feature extraction

Three classical feature extraction methods have been applied for extracting the facial expression information: the histogram of oriented gradients (HOG), LBP, and Gabor filters. They have been used in many important studies [3, 26] of non-frontal facial expression recognition. These methods can extract local facial expression features from face images. Therefore, in our experiment, we extracted features from salient facial patches in each image using the three methods separately to compare their recognition performances.

3.3.1.1 HOG First, we divided the whole-face image into parts; second, we obtained a histogram from each cell; and finally, we normalized the computed results and returned a descriptor.

3.3.1.2 LBP The $N \times N$ LBP operator was used to obtain the facial expression features. The operator weights were multiplied by the corresponding pixels of the face image, and $N \times N - 1$ pixels were used for the LBP features of the neighborhood. There are many variations of the LBP algorithm. In Happy et al.’s study, the highest recognition rate was attained using a uniform LBP. The $N \times N$ uniform LBP operator computes LBP features from a circular neighborhood. It has two important parameters: P, which is the number of corresponding pixels, and R, which is the circular neighborhood radius.

3.3.1.3 Gabor A two-dimensional Gabor filter can be formulated as [27]

$$G(x, y) = \frac{f^2}{\pi\gamma\eta} e^{-\frac{x'^2+y'^2}{2\sigma^2}} e^{i2\pi f x' + \phi}$$

$$x' = x \cos\theta + y \sin\theta, y' = -x \sin\theta + y \cos\theta \tag{8}$$

$$f = \frac{1/4}{\sqrt{2}^{u-1}}, u = 1, 2, \dots, 5. \theta = \frac{\pi}{8} \times (v-1), v = 1, 2, \dots, 8$$

where f is the frequency of the sinusoidal factor, and θ represents the orientation of the normal to the parallel stripes of the Gabor function. Further, ϕ is the phase offset, σ is

the standard deviation of the Gaussian envelope, and γ denotes the spatial aspect ratio that specifies the ellipticity of the support of the Gabor function. If image $I(x, y)$ is convolved with a Gabor filter, the Gabor features will be extracted by the particular f and θ values. In our experiments, we chose the largest value of f , and u was set to 1.

The above examples show feature extraction that was performed on only a single patch; thus, feature fusion was necessary for feature extraction of the salient facial patches.

3.3.2 Classification

After the facial expression features were extracted, the final task was feature classification. Non-frontal face images are hampered by a lack of emotion information. Thus, if the classifier is weak, the recognition rate may be very low. To address this problem, the adaptive boosting (AdaBoost) [28] algorithm was applied for the classification because it effectively combines many learning algorithms to improve the recognition performance and is thus suitable for classification.

4 Results and discussion

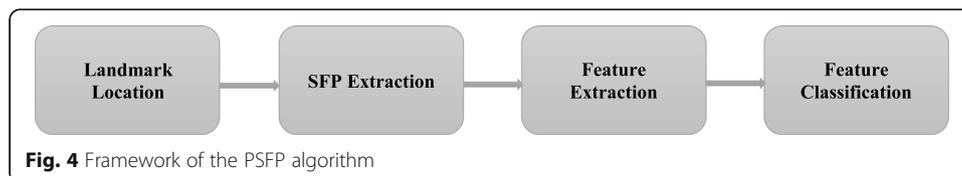
4.1 Experimental setting

This simulation environment of our experiment used MATLAB R2015b on a Dell personal computer. We evaluated the PSFP algorithm on the Radboud Faces Database (RaFD) [25]. RaFD is a free publicly available dataset that contains eight facial expressions: anger, contempt, disgust, fear, happiness, neutrality, sadness, and surprise. Each facial expression is shown with three different gaze directions: frontal, left, and right. The photographer captured photographs of 67 models with five different head poses. In this study, 1200 face images were used for the experiments, consisting of ten people, eight expressions, three gaze directions, and five head poses.

The framework of the PSFP algorithm is shown in Fig. 4.

For determining the facial landmark locations, the Yu et al. method was used, and salient facial patches were extracted from the face images under five different head poses. This method can estimate the head poses along pitch, yaw, and roll directions. However, in our experiments, the method was only needed to estimate the head poses along the yaw direction.

The size of the facial patches was typically set to 16×16 . HOG, LBP ($P = 8, R = 1$), and Gabor filters ($u = 1, v = 1, 2, \dots, 8$) were respectively applied for the feature extraction. Principal component analysis (PCA) was used for feature dimensionality reduction; the feature dimensionality was typically set to ten. We used the M1-type AdaBoost method (AdaBoost.M1) for the classification and applied the nearest-neighbor method (NN) for the AdaBoost.M1 basic classifier. The maximum number of iterations was 100.



4.2 Purposes

Experiments were conducted to validate the PSFP recognition performance with respect to the four different perspectives.

4.2.1 Testing PSFP performance under different training–testing strategies

There are two commonly used experimental approaches to performing non-frontal facial expression recognition: pose-invariant and pose-variant. In the former, training images and test images are obtained under the same head pose; thus, head pose estimation can be avoided. In the latter, the training and test images may have different head poses. This approach is thus more realistic. To analyze the recognition performance of the PSFP algorithm, two simulation experiments were performed, as described in the Pose-invariant non-frontal facial expression recognition section and Pose-variant non-frontal facial expression recognition section.

4.2.2 Testing PSFP performance under different parameter values

Generally, the selection of parameters depends on empirical values, and it is difficult to support them with rigorous proof. Therefore, it was necessary to use different parameter values for PSFP and to observe the recognition performance on a test set. As described in the Testing PSFP performance under different training–testing strategies section, the size of the facial patches was typically set to 16×16 , and the feature dimensionality was typically set to 10. Both of these key parameters could affect the expression recognition performance. The Comparison by facial patch size and Comparison by feature dimensionality sections describe the experiments that were conducted for this performance comparison.

4.2.3 Comparing PSFP with SFP for frontal facial expression recognition

In the Extraction of pose-free salient facial patches section, we discussed the two differences between the SFP method of Happy et al. and PSFP. Even if we replace the SFP face detection method with the Yu et al. method, this modified SFP method would still not be suitable for application to non-frontal-view face images. However, if we use PSFP to recognize the frontal-view face images, PSFP and SFP may be similar in the positions they select for facial salient patches. As PSFP and SFP should be compared, it is necessary to perform the experiments for frontal facial expression recognition. The experiment described in the Comparison with the Happy et al. SFP method section was designed for this purpose.

4.2.4 Comparing PSFP with non-SFP using whole-face images

A salient facial patch is in fact only part of a face image. According to common understanding, if the whole-face image is used for the recognition, the performance may be better. However, if the selection of salient facial patches is sufficiently good, PSFP could perform better than this non-SFP method. Therefore, we used the same feature extraction and classification method for the two methods and compared them, as described in the Comparison with non-SFP method using whole-face images section.

4.3 Pose-invariant non-frontal facial expression recognition

There are two training–testing strategies for facial expression recognition: person-dependent and person-independent. In our experiments on person-dependent facial expression recognition, the subjects appearing in the training set also appeared in the test set. Because every model had three different head poses, a three-fold cross-validation strategy was used for the person-dependent facial expression recognition.

The dataset could be divided into three segments according to head pose. Each time, two segments were used for training, and the remaining segment was used for testing. Thus, for each head rotation angle, the number of images in the training set was 160, and the number in the test set was 80. The same training–testing procedure was carried out three times and the average result of the three procedures was considered as the final recognition performance of the PSFP algorithm. The HOG, LBP, and Gabor methods were used for feature extraction, and the AdaBoost algorithm with the NN classifier was applied for classification. The recognition rates of these methods are shown in Table 2. Each row shows the recognition performance for five head rotation angles (90°, 45°, 0°, –45°, and –90°). The best recognition rates are highlighted in bold. For most angles, HOG has the best recognition performance, and at 0° and –45°, LBP has the best recognition performance. The best head rotation angle for recognition of non-frontal facial expressions is –45°.

In the experiments on person-independent facial expression recognition, the subjects appearing in the training set did not appear in the test set. For this reason, the leave-one-person-out strategy was used. That is, all photographs of one person were selected as the test set; the remaining photographs in the dataset were used for training. Thus, for each head rotation angle, the number of images in the training set was 216, and the number in the test set was 24. This procedure was repeated ten times, and the averaged result was taken as the final recognition rate. The results are shown in Table 3. For most angles, Gabor achieved the best recognition rate. For 0°, –45°, and 90°, Gabor and LBP achieved the best recognition rate. We found that the best head rotation angle for recognition of non-frontal facial expressions was 45°.

In summary, analyses of the pose-invariant non-frontal facial expression recognition experiments show the following: (1) When the head rotation angle is larger, the recognition rate may be lower. Because many facial patches are occluded by head rotation, the number of emotion features is not sufficient to achieve a high recognition rate. (2) Although identity bias and face occlusion interfere with facial expression recognition, the PSFP algorithm can achieve better recognition performance on non-frontal facial expression recognition.

Table 2 Recognition rates (%) for person-dependent facial expression recognition. The best recognition rates are highlighted in bold

Head pose	HOG	LBP	Gabor
90°	97.92	96.67	95.83
45°	99.17	98.75	98.33
0°	100	100	100
–45°	100	98.75	99.58
–90°	98.33	99.17	96.25

Table 3 Recognition rates (%) for person-independent facial expression recognition. The best recognition rates are highlighted in bold

Head pose	HOG	LBP	Gabor
90°	81.67	81.67	81.25
45°	95.00	91.25	92.50
0°	97.92	98.75	98.75
− 45°	88.33	89.17	92.08
− 90°	82.50	86.25	87.08

4.4 Pose-variant non-frontal facial expression recognition

In the experiments on person-dependent facial expression recognition, a three-fold cross-validation strategy was used for training and testing. The number of images in the training set was 800, and the number in the test set was 400. The same procedure was performed three times.

In the experiments on person-independent facial expression recognition, the leave-one-person-out strategy was used. The number of images in the training set was 1080, and the number in the test set was 120. This procedure was performed ten times for each dataset, and the average values are taken as the final recognition rate. The results are listed in Table 4.

As shown in the table, having different head pose rotations increases the difficulty of non-frontal facial expression recognition. However, the proposed method performed well. PSFP with the HOG algorithm again achieved the best recognition rates.

4.5 Performance comparisons

4.5.1 Comparison by facial patch size

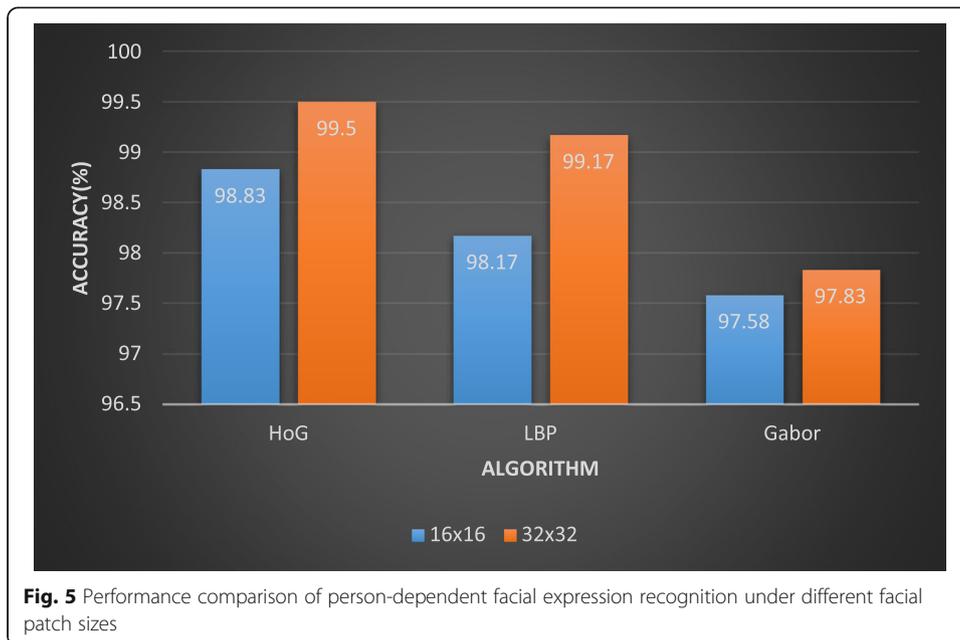
In the above experiments, the size of the facial patches was 16×16. We increased the size to 32×32, and the experiment results are shown in Figs. 5 and 6. When the results in Figs. 5 and 6 are compared, we can observe that the person-dependent results are better than the person-independent ones. Moreover, the 32×32 facial patches achieved higher recognition performance than the 16×16 facial patches in most cases. This is because the feature extraction methods can obtain much more information, which helps improve the recognition performance of non-frontal facial expression recognition.

4.5.2 Comparison by feature dimensionality

In the above experiments, the feature dimensionality was set to ten. We again conducted the experiments for pose-variant non-frontal facial expression recognition and the feature dimensionality was increased from ten to 100. AdaBoost with NN was used as the classifier, and the feature extraction methods were HOG, LBP, and Gabor. The results are shown in Figs. 7 and 8. It is observed that the recognition rates increase

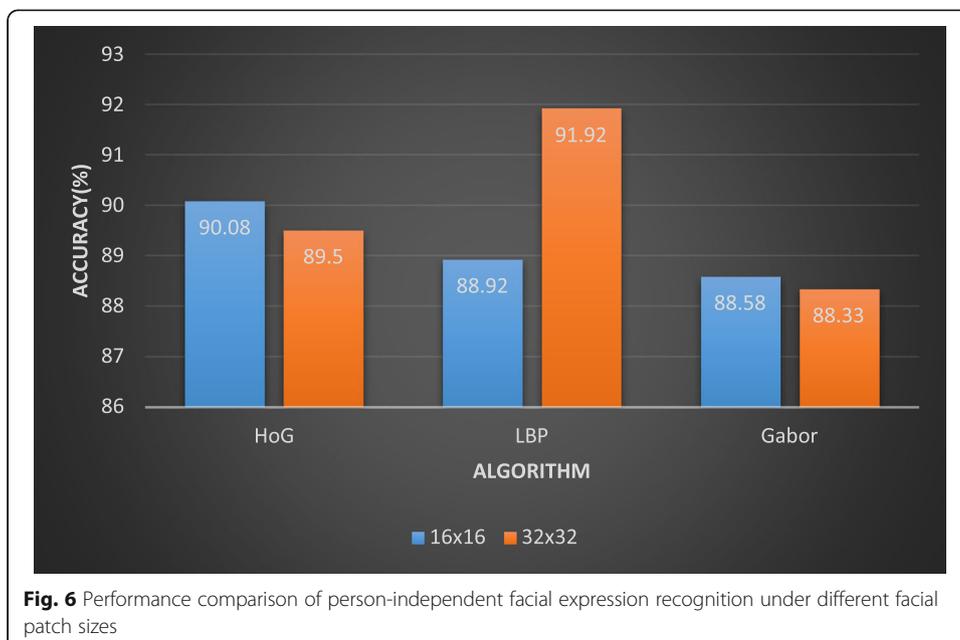
Table 4 Accuracy (%) for pose-variant non-frontal facial expression recognition

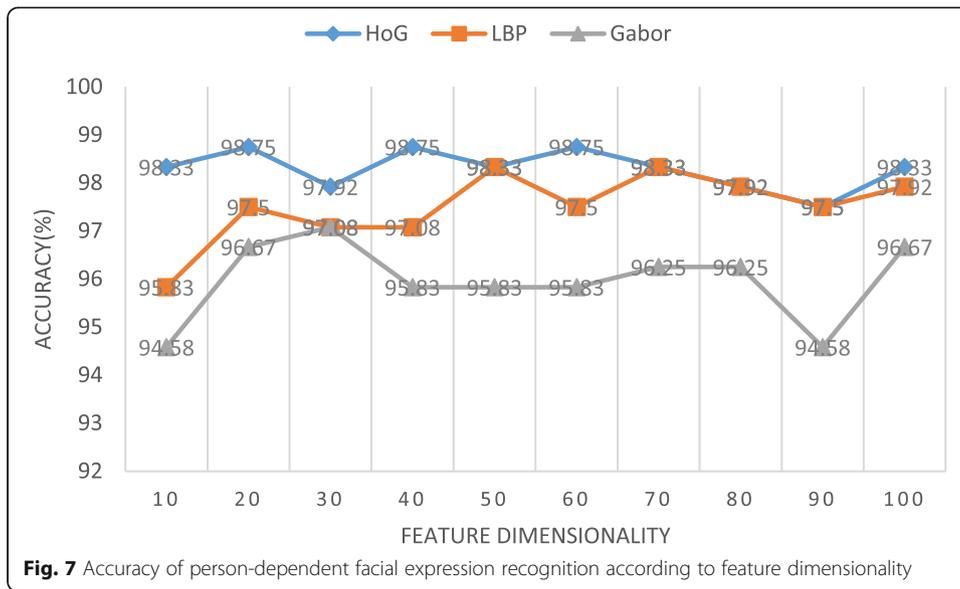
Strategy	HOG	LBP	Gabor
Person-dependent recognition	98.83	98.17	97.58
Person-independent recognition	90.08	88.92	88.58



from the initial allocation and eventually settle around a range of values. In the experiment on pose-variant non-frontal facial expression recognition, the magnitude of the range is from 2 to 7%. We find that the accuracy of person-independent facial expression recognition can increase with the increase in feature dimensionality. Because this model is trained and tested on different subjects, it leads to individual differences, which significantly hinders the recognition. When the feature dimension is increased, it improves the classification accuracy.

Although the recognition rate may increase with the increase in feature dimensionality, the computation cost of the algorithm is necessarily higher. We suggest that the

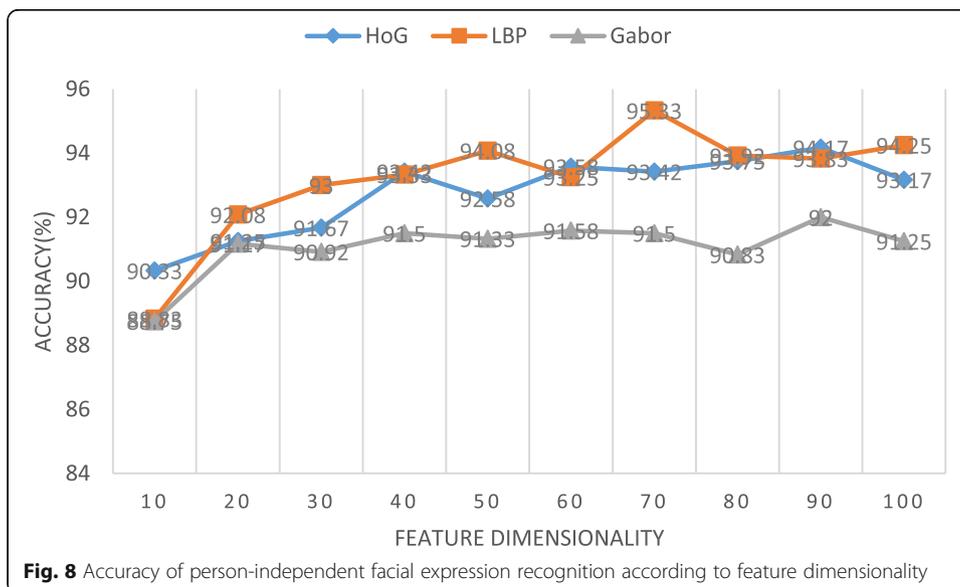


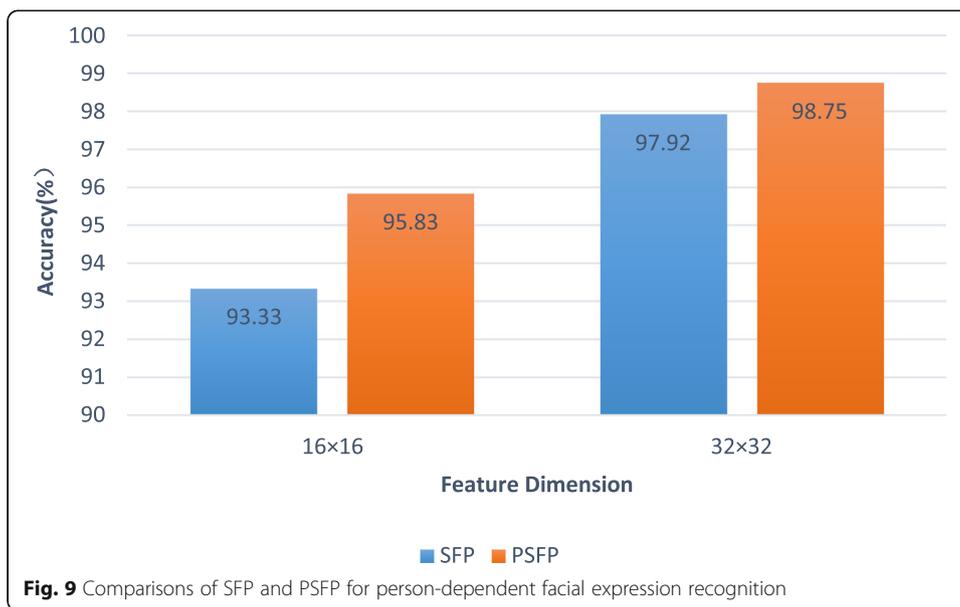


feature dimensionality should be set to a value that is as small as possible while maintaining good performance.

4.5.3 Comparison with the Happy et al. SFP method

To recreate the experimental conditions of Happy et al., the LBP and linear discriminant analysis (LDA) methods were used for feature extraction, and support vector machine (SVM) was used for classification. The results are shown in Figs. 9 and 10. When LBP parameters P and R are respectively equal to 8 and 1, the PSFP accuracy is higher than that of the Happy et al. SFP method. This finding demonstrates that the PSFP method can also outperform SFP for frontal facial expression recognition.





4.5.4 Comparison with non-SFP method using whole-face images

In this experiment, the LBP algorithm was used to extract the whole-face images, and the AdaBoost algorithm was applied for classification. The non-SFP method was compared with the PSFP method for pose-invariant non-frontal facial expression recognition. The recognition rates for person-dependent and person-independent strategies are shown in Figs. 11 and 12.

Even though the PSFP method does not use the whole-face image for recognition, its accuracy is not lower than that of the non-SFP method using whole-face images. The selection of salient facial patches enables the PSFP method to achieve a higher accuracy. Moreover, the size of the whole-face image is 128×128 , and the total areas of the salient facial patches are $16 \times 16 \times 20$, and $16 \times 16 \times 12$. Thus, the PSFP method substantially reduces the quantity of data.

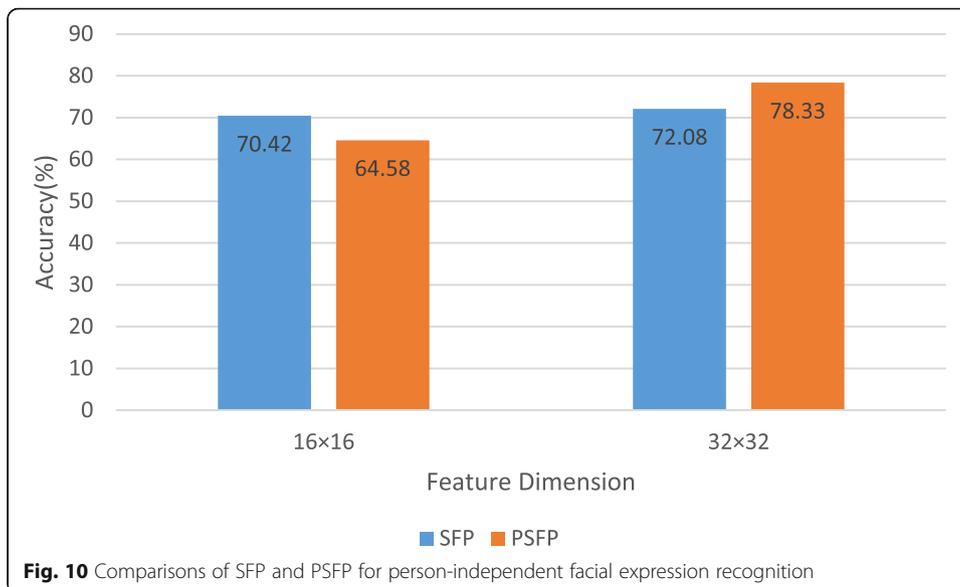




Fig. 11 Recognition rates for person-dependent facial expression recognition

4.5.5 CNN-based features perform for this non-frontal facial expression recognition task

As mentioned in the Related work section, several studies have employed salient patches with CNNs for face detection and classification. We thus used CNN for non-frontal facial expression recognition. The CNN model was 21-layer VGG [29] and AlexNet [30]. The number of images in the training set was 800, and the number in the test set was 400. The recognition rates are shown in Fig. 13, where we observe that the VGG recognition rate is lower than the AlexNet recognition rate.

We also used facial patches as images for training CNNs. However, this approach may be not suitable for recognition. CNNs typically require a whole face image for

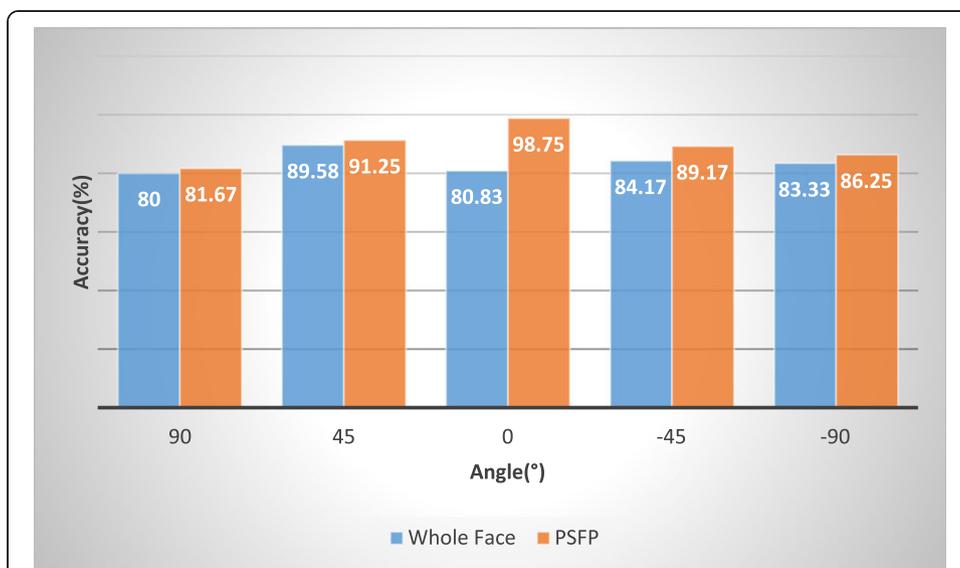


Fig. 12 Recognition rates for person-independent facial expression recognition

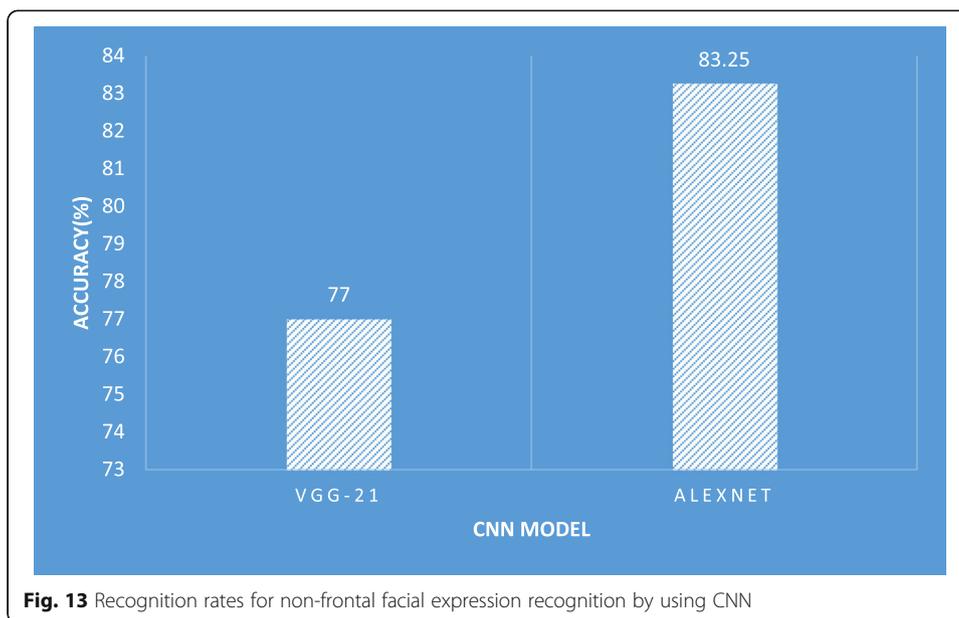


Fig. 13 Recognition rates for non-frontal facial expression recognition by using CNN

model training. The problem remains of how to use patches with CNNs and achieve good performance. This issue will be addressed in our future research.

4.6 Summary

From the above experiments, we find that the PSFP method has the following characteristics:

- (1) HOG features have better recognition performance than LBP features or Gabor features. We believe this is because the LBP features are based on the local image regions of the facial patch and the Gabor features are extracted from the whole-face patch, whereas HOG features are obtained from the small squared cells of the facial patch. Therefore, the HOG method can more effectively extract the emotion features under complex changes of light, scale, pose, and identity environments.
- (2) The PSFP method, an extension of the SFP method, can also be applied for frontal facial expression recognition.
- (3) PSFP can achieve high recognition rates while consuming fewer data.

5 Conclusion

This paper presented PSFP, an algorithm based on salient facial patches. PSFP employs the relevance of facial patches in non-frontal facial expression recognition and employs the facial landmark detection method to track key points from a pose-free human face. In addition, an algorithm for extracting the salient facial patches was proposed. This algorithm determines the facial patches under different head rotations. The facial expression features can be extracted from the facial patches and used for feature classification. The experiment results showed that PSFP can achieve high recognition rates while consuming fewer data.

Acknowledgments

The authors are very grateful to the editors and reviewers, to Dr. Xiang Yu for supplying the MATLAB code for face detection, and to Radboud University Nijmegen for providing the RaFD database.

Authors' contributions

BJ conceived the algorithm, designed the experiments, analyzed the results, and wrote the paper; QZ, ZL, and QW wrote the codes and performed the experiments; and HZ managed the overall research and contributed to the paper writing. The authors read and approved the final manuscript.

Authors' information

Bin Jiang received his M.S. degree from Henan University in 2009, and his Ph.D. from Beijing University of Technology, Beijing, China, in 2014. He joined the Zhengzhou University of Light Industry as a lecturer in 2014. His current research interests include image processing, pattern recognition, and machine learning.

Qiuwen Zhang received his Ph.D. degree in communication and information systems from Shanghai University, Shanghai, China, in 2012. Since 2012, he has been with the faculty of the College of Computer and Communication Engineering, Zhengzhou University of Light Industry, where he is an associate professor. He has published over 30 technical papers in the fields of pattern recognition and image processing. His major research interests include 3D signal processing, machine learning, pattern recognition, video codec optimization, and multimedia communication.

Zuhe Li received his M.S. degree in communication and information systems from Huazhong University of Science and Technology in 2008, and his Ph.D. degree in information and communication engineering from Northwestern Polytechnical University in 2017. He is currently an associate professor at Zhengzhou University of Light Industry. His current research interests include computer vision and machine learning.

Qinggang Wu received M.S. and Ph.D. degrees in computer science from Dalian Maritime University, Dalian, China, in 2008 and 2012, respectively. Since January 2013, he has been a lecturer at the School of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou, China. His research interests include remote sensing image processing, image segmentation, edge detection, pattern recognition, and computer vision.

Huanlong Zhang received his Ph.D. degree from the School of Aeronautics and Astronautics, Shanghai Jiao Tong University, China, in 2015. He is currently an associate professor at the College of Electric and Information Engineering, Zhengzhou University of Light Industry, Henan, Zhengzhou, China. He has published more than 40 technical articles in referred journals and conference proceedings. His research interests include pattern recognition, machine learning, image processing, computer vision, and intelligent human-machine systems.

Funding

This work was supported by the National Natural Science Foundation of China (Nos. 61702464, 61771432, 61873246, 61702462, and 61502435), the Scientific and Technological Project of Henan Province under Grant Nos. 16A520028, 182102210607, and 192102210108, and the Doctorate Research Funding of Zhengzhou University of Light Industry under Grant No. 2014BSJJ077.

Availability of data and materials

The raw/processed data required to reproduce these findings cannot be shared at this time as the data also form part of an ongoing study.

Declarations

Ethics approval and consent to participate

Authors have permissions on usage of photos of RaFD database as in strictly scientific publications RaFD images can be presented as stimulus examples.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Author details

¹College of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, People's Republic of China. ²College of Electric and Information Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, People's Republic of China.

Received: 19 September 2020 Accepted: 29 March 2021

Published online: 12 May 2021

References

1. E. Sariyanidi, H. Gunes, A. Cavallaro, Automatic analysis of facial affect: A survey of registration, representation, and recognition. *IEEE Transactions on Pattern Analysis & Machine Intelligence* **37**(6), 1113–1133 (2015)
2. M. Pantic, I. Patras, Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences. *IEEE Transactions on Systems Man & Cybernetics Part B* **36**(2), 433–449 (2006)
3. Y.X. Hu, Z.H. Zeng, L.J. Yin, X.Z. Wei, J.L. Tu, T.S. Huang, A study of non-frontal-view facial expressions recognition. *IEEE International Conference on Pattern Recognition*, 2008. ICPR, 1–4 (2008)
4. A. Dapogny, K. Bailly, S. Dubuisson, Dynamic pose-robust facial expression recognition by multi-view pairwise conditional random forests. *IEEE Transactions on Affective Computing* **10**(2), 167–181 (2019)

5. W.M. Zheng, H. Tang, Z.C. Lin, T.S. Huang, Emotion recognition from arbitrary view facial images. *Proceeding International Conference European Conference on Computer Vision* **2010**, 490–503 (2010)
6. L.J. Yin, X.Z. Wei, Y. Sun, J. Wang, M.J. Rosato, A 3D facial expression database for facial behavior research. *IEEE International Conference on Automatic Face and Gesture Recognition* **2006**, 211–216 (2006)
7. J.L. Wu, Z.C. Lin, W.M. Zheng, H.B. Zha, Locality-constrained linear coding based bi-layer model for multi-view facial expression recognition. *Neurocomputing* **239**, 143–152 (2017)
8. Y.H. Lai, S.H. Lai, Emotion-preserving representation learning via generative adversarial network for multi-view facial expression recognition. *IEEE International Conference on Automatic Face and Gesture Recognition* **2018**, 263–270 (2018)
9. Q.R. Mao, Q.Y. Rao, Y.B. Yu, M. Dong, Hierarchical Bayesian theme models for multipose facial expression recognition. *IEEE Transactions on Multimedia* **19**(4), 861–873 (2017)
10. M. Jampour, V. Lepetit, T. Mauthner, H. Bischof, Pose-specific non-linear mappings in feature space towards multiview facial expression recognition. *Image & Vision Computing* **58**, 38–46 (2017)
11. E. Sabu, P.P. Mathai, An extensive review of facial expression recognition using salient facial patches. *Proceeding International Conference Applied and Theoretical Computing and Communication Technology* **2015**, 847–851 (2015)
12. S.L. Happy, A. Routray, Automatic facial expression recognition using features of salient facial patches. *IEEE Transactions on Affective Computing* **6**(1), 1–12 (2015)
13. K.K. Chitta, N.N. Sajjan, A reduced region of interest based approach for facial expression recognition from static images. *IEEE Region 10 Conference* **2016**, 2806–2809 (2016)
14. R. Zhang, J. Li, Z.Z. Xiang, J.B. Su, Facial expression recognition based on salient patch selection. *IEEE International Conference on Machine Learning and Cybernetics* **2016**, 502–507 (2016)
15. Y.M. Wen, W. Ouyang, Y.Q. Ling, Expression-oriented ROI region secondary voting mechanism. *Application Research of Computers* **36**(9), 2861–2865 (2019)
16. W.Y. Sun, H.T. Zhao, Z. Jin, A visual attention based ROI detection method for facial expression recognition. *Neurocomputing* **296**, 12–22 (2018)
17. J.Z. Yi, A.B. Chen, Z.X. Cai, Y. Sima, X.Y. Wu, Facial expression recognition of intercepted video sequences based on feature point movement trend and feature block texture variation. *Applied Soft Computing* **82**, 105540 (2019)
18. N.M. Yao, H. Chen, Q.P. Guo, H.A. Wang, Non-frontal facial expression recognition using a depth-patch based deep neural network. *Journal of computer science and technology* **32**(6), 1172–1185 (2017)
19. A. Barman, P. Dutta, Facial expression recognition using distance and shape signature features. *Pattern Recognition Letters* **145**, 254–261 (2021)
20. Y. Sun, X.G. Wang, X.O. Tang, Deep convolutional network cascade for facial point detection. *IEEE International Conference on Computer Vision and Pattern Recognition*, 3476–3483 (2013, 2013)
21. T.F. Cootes, G.J. Edwards, C.J. Taylor, Active appearance models. *IEEE Transaction on Pattern Analysis and Machine Intelligence* **23**(6), 681–685 (2001)
22. X. Jin, X.Y. Tan, Face alignment in-the-wild: A survey. *Computer Vision and Image Understanding* **162**, 1–22 (2017)
23. X. Yu, J.Z. Huang, S.T. Zhang, W. Yan, D.N. Metaxas, Pose-free facial landmark fitting via optimized part mixtures and cascaded deformable shape model. *IEEE International Conference on Computer Vision* **2013**, 1944–1951 (2013)
24. J. Liu, S.W. Ji, J.P. Ye, *SLEP: Sparse Learning with Efficient Projections* (Arizona State University, Arizona, 2009)
25. O. Langner, R. Dotsch, G. Bijlstra, D.H.J. Wigboldus, S.T. Hawk, A.V. Knippenberg, Presentation and validation of the Radboud faces database. *Cognition & Emotion* **24**(8), 1377–1388 (2010)
26. S. Moore, R. Bowden, Local binary patterns for multi-view facial expression recognition. *Computer Vision Image Understand* **115**(4), 541–558 (2011)
27. M. Haghghat, S. Zonouz, M. Abdel-Mottaleb, Identification using encrypted biometrics. *Computer Analysis of Images and Patterns*, 440–448 (2013) York, United Kingdom
28. R.E. Schapire, A brief introduction to boosting. *IEEE International Joint Conference on Artificial Intelligence* **1999**, 1401–1406 (1999)
29. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations*, 1–4 (2015)
30. A. Krizhevsky, I. Sutskever, G.E. Hinton, *ImageNet classification with deep convolutional neural networks. Neural Information Processing Systems* (Curran Associates Inc, Red Hook, 2012), pp. 1097–1105

Publisher's Note

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

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
