Non-frontal facial expression recognition based on salient facial patches

Bin Jiang (jiangbin@zzuli.edu.cn)
Zhengzhou University of Light Industry https://orcid.org/0000-0002-6338-4051
Qiuwen Zhang
Zhengzhou University of Light Industry
Zuhe Li
Zhengzhou University of Light Industry
Qinggang Wu
Zhengzhou University of Light Industry
Huanlong Zhang
Zhengzhou University of Light Industry

Research

Keywords: facial expression recognition, salient facial patches, head rotations

Posted Date: January 27th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-80959/v2

Non-frontal facial expression recognition based on salient facial patches

Bin Jiang[1*], Qiuwen Zhang[1], Zuhe Li[1], Qinggang Wu[1], Huanlong Zhang[2]

***Correspondence: jiangbin@zzuli.edu.cn**
[1]College of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, People's Republic of China
[2]College of Electric and Information Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, People's Republic of China

Abstract. Methods using salient facial patches (SFP) play a significant role in research on facial expression recognition. However, most SFP methods use only frontal face images or videos for recognition, and do not consider variations of head position. In our view, SFP can also be a good choice to recognize facial expression under different head rotations, and thus we propose an algorithm for this purpose, called Profile Salient Facial Patches (PSFP). First, in order to detect the facial landmarks from profile face images, the tree-structured part model is used for pose-free landmark localization; this approach excels at detecting facial landmarks and estimating head poses. Second, to obtain the salient facial patches from profile face images, the facial patches are selected using the detected facial landmarks, while avoiding overlap with each other or going beyond the range of the actual face. For the purpose of analyzing the recognition performance of PSFP, three classical approaches for local feature extraction-histogram of oriented Gradients (HOG), local binary pattern (LBP), and Gabor were applied to extract profile facial expression features. Experimental results on radboud faces database show that PSFP with HOG features can achieve higher accuracies under the most head rotations.

Keywords: facial expression recognition, salient facial patches, head rotations.

1 Introduction

The problem of determining how to use face information in human computer interaction has been the subject of analysis for a number of years. An increasing number of applications using face recognition technology have appeared and are being used routinely. However, current studies on facial expression recognition are not yet being fully and realistically applied. Variations in head pose constitute one of the main challenges in the automatic recognition of facial expressions[1]; the problem arises when there are deliberate occlusions and because nearly half of the face can disappear under large changes in head pose. Achieving the automatic analysis of facial expressions from the pose-free human face will be necessary for the establishment of a technological framework for further research.

Recognition of profile facial expressions was first achieved by Pentic et al.[2] Particle filtering was used to track 15 facial landmarks in a sequence of face profiles, and a recognition rate of 87% was achieved. Although only -90° face image sequences were used as experimental data, their work inspired further research. Hu et al.[3] are considered the 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 around the facial components and then exploited five classifiers to recognize emotion features. Their extensive experiment results demonstrate that good recognition can be achieved on profile face images.

Dapogny et al.[4] used spatio-temporal features to recognize facial expressions under variations in head pose from videos; thus, the extracted features were not limited to spatial features. 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. These authors chose 100 subjects from the BU-3DFE database, [6] and their experiment results demonstrated that their proposed algorithm could obtain good 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, and then the facial expression features are extracted using the corresponding view-dependent model in the second layer. This model has 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 variations in head pose. Mao et al.[9] considered relationships between head poses and proposed a pose-based hierarchical Bayesian-

themed model. Jampour et al.[10] found that linear or non-linear local mapping methods provide more reasonable results for multi-pose facial expression recognition than global mapping methods. However, none of the above algorithms is sufficient to correctly recognize expressions on faces in non-frontal images. Although the researchers have sought to achieve higher recognition rates, 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 bias. An effective feature extraction method is necessary for the recognition of 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, a few prominent facial patches (e.g., eyebrows, eyes, cheeks, and mouth) are relied on as the key points in face images, and the discriminative features are extracted from salient regions. The extracted features are important for distinguishing one expression from another, and the salient facial patches create favorable conditions for non-frontal facial expression recognition. Thus, we propose an algorithm based on salient facial patches designed to recognize 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 salient facial patches. The remainder of this paper is organized as follows. Related work is described in Sec. 2, and the details of PSFP are presented in Sec. 3. We provide the design and analysis of experiments for facial expression recognition in Sec. 4. Finally, we conclude the paper in Sec. 5.

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 the most accurate and efficient system of the methods proposed to date was that by Happy and Routray,[12] who provided the system for facial expression recognition using salient facial patches. These salient regions can vary in different facial expressions and can be responsible for deformation of the face. This system is easy to reproduce and is efficient for recognizing frontal-view facial expressions. 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 their experiment. Zhang et al.[14] used a sparse group lasso scheme to explore the most salient patches for each facial expression and 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 CNN (convolutional neural network)[20] to train the salient facial patches on face images and then employed a secondary voting mechanism to help the trained convolutional neural network determine the final category of test images. Sun et al.[16] presented a convolutional neural network that uses a visual attention mechanism and can be used for facial expression recognition. This mechanism pays attention to local areas of face images and determines the importance of each region. In particular, whole face images with different poses were used for training the convolutional neural network. Yi et al.[17] expanded the salient facial patches from static images to video sequences, and 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, which can capture pose-variant expression features from the depth patches and recognize non-frontal expressions. Barman and Dutta.[19] used an active appearance

model (AAM) [21] to detect the salient facial landmarks, whose connections form triangles that can be regarded as salient facial regions. The geometric features are extracted for the recognition of emotion in the face.

Based on this overview of salient facial patches algorithms, we find the following three commonalities in facial expression recognition:

1. Most of the proposed methods are used on frontal face images.
2. There are three main components of salient facial regions: eyes, nose, and lips.
3. The appearance or texture feature is very important for recognizing facial expressions.

In our view, the salient facial patches method should be applied not only for frontal facial expression recognition, but also for non-frontal facial expression recognition. Inspired by the method in **Happy et al.'s method**, we propose the PSFP method for non-frontal facial expression recognition. In contrast with previous non-frontal facial expression recognition methods, this method employs salient facial patches, which are composed mainly of those facial components that carry much facial expression information under variations in head pose. Thus, it can extract many appearance or texture features under head pose variations and identity bias. Furthermore, the PSFP method does not require the construction of a complex model for multi-pose facial expression classification. The details of the PSFP method are introduced 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 localizations of salient facial patches on the non-frontal face images. Therefore, localization of fiducial facial points and estimation of the head pose of the faces are essential intermediate steps for identifying the salient facial patches. Especially, head pose may be the

combination of different directions in three dimensional space. **If the face detection method can't obtain adequate information about the rotations undergone, the recognition rates of facial expression will be low.** In the careful descriptions of prominent methods provided by Jin and Tan, [22] we found that the tree-structured part model can use a unified framework to detect the human face and estimate head variations; this is highly suitable for non-frontal facial expression recognition. Thus, we adopt Yu et al.'s method[23] in our system to accomplish the face detection and head pose estimation. **Because this algorithm can estimate the head poses along pitch, yaw and roll directions, and good enough to detect the head poses and face positions of human faces.**

3.1 Face Detection

Yu et al.[23] presented a united framework to detect the human face and track facial feature points simultaneously. There are two main steps in their framework:

(1) Initialization. The mixtures of the tree-structured part model are used to formulate the problem as follows:

$$s^* = \arg \max_{s \in S, i \in (1,M)} \sum_{j \in V_i} q_i(I, s_j) + \sum_{(j,k) \in E_i} g_i(s_j, s_k), \quad (1)$$

where the first term uses local patch appearance evaluation function q_i, which indicates whether a facial landmark may be at the aligned position; the second term uses shape deformation cost g_i, which will maintain the balance of the relative locations of neighboring facial landmarks; and the tree has been defined as $T_i = (V_i, E_i)$, $i \in (1, M)$, in which V represents the shared pool of parts and E represents an edge between two parts. For each viewpoint i, this scoring function is applied to measure the facial landmark configuration s. Eq. 1 assigns a larger score to more likely positions of facial landmarks. **In order to solve the Eq. 1, a group sparse learning algorithm[24] can be used to select the most salient weights, and form a new tree.**

(2) Localization. The initial facial landmarks having been detected, the authors use the Procrustes analysis method to project their 3D reference shape model onto a 2D face image. **They first transform this problem into parametric form, and then build the probabilistic model.** Based on this probabilistic model, **a two-step cascaded deformable shape model was proposed to refine the locations of the facial landmarks.**

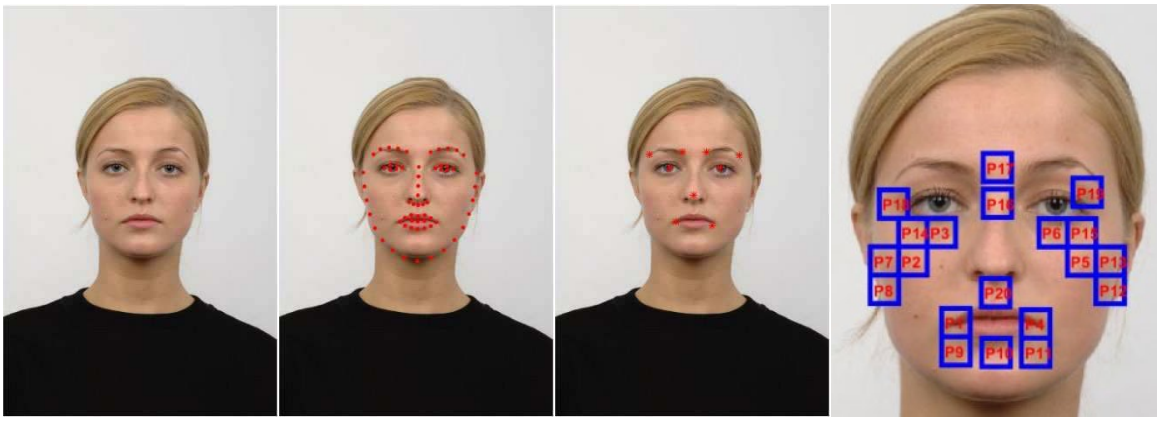
$$s^* = \arg\max_s p(s|\{v_i = 1\}_1^N, I) \tag{2}$$

$$\propto \arg\max_s p(s)p(\{v_i = 1\}_{i=1}^n|s, I) \tag{3}$$

$$= \arg\max_{\mathcal{P}} p(\mathcal{P}) \prod_{i=1}^n p(v_i = 1|s_i, I) \tag{4}$$

In Eq. 2, the vector $\mathrm{v} = \{v_1, \dots, v_N\}$ indicates the likelihood of alignment in the face image *I*. $\mathrm{v} = 1$ indicates that facial landmarks have been well aligned; $\mathrm{v} = 0$ indicates the opposite. Thus, Eq. 2 aims to maximize the likelihood of alignment. Then, we can use the Bayesian rule to derive Eq. 3. In Eq. 4, we know that the parameter $\mathcal{P}$ can determine 3D shape model *s*, $p(\mathcal{P}) = p(s)$. The authors suppose that $p(\mathcal{P})$ obeys the Gaussian distribution. In addition, the logistic regressor is used to interpret $p(v_i = 1|s_i, I) = \frac{1}{\exp(\vartheta\varphi + b)}$, where φ is the feature descriptor of facial landmark patch *i*, and the other parameters ϑ and *b* represent two regressor weights.

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 Fig. 1 (a) can be shown as in Fig. 1 (b).



(a) (b) (c) (d)

Fig. 1 Framework for automated extraction of salient facial patches. (a) Face image from RaFD database,[25] (b) the 66 facial landmarks detected using Yu et al.'s method,[23] (c) the points of lip corners and eyebrows, **and (d) locations of the salient facial patches.**

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 the analysis of related work, we find that 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. 1 (a) can be shown as in Fig. 1 (c). The salient facial patches A_i can be extracted around the facial parts and the areas among eyebrow, eye, nose, and lip areas:

$$A_i = \begin{bmatrix} (x_i - \frac{M}{2} + 1, y_i - \frac{N}{2} + 1) & \cdots & (x_i - \frac{M}{2} + 1, y_i + \frac{N}{2}) \\ \vdots & \ddots & \vdots \\ (x_i + \frac{M}{2}, y_i - \frac{N}{2} + 1) & \cdots & (x_i + \frac{M}{2}, y_i + \frac{N}{2}) \end{bmatrix} \quad (5)$$

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 the L salient facial patches:

$$R_i = (A_1, A_2, \cdots, A_L), i = 1,2, \cdots k \quad (6)$$

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

The rationale behind choosing the given 20 patches is following facial action coding system. P_1 and P_4 are at the lip corners; P_9 and P_{11} are just below them. P_{10} is at the midpoint of P_9 and P_{11}. **P_{20} is at the upper lip.** P_{16} is 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 located from the middle of the nose and the eyes. P_5, P_{13}, and P_{12} are stacked together, extracted from the left side of the nose,

and P_2, P_7, and P_8 are at the right side of the nose. **P_{18} and P_{19} are located on the respective outer eye corners**.

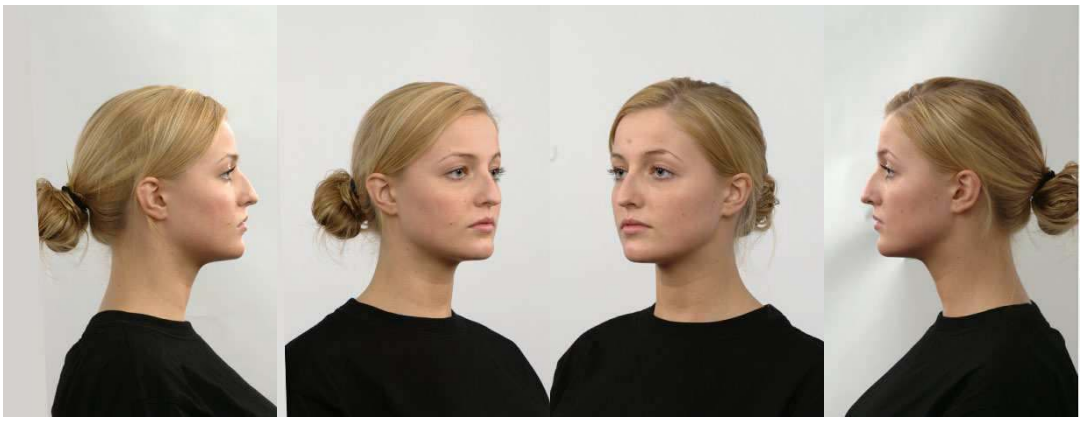
The method of selection of facial patches in PSFP is similar to that in **Happy et al.**, with two exceptions. The first difference is that the salient facial patches (SFP) method used in **Happy et al.**, which extracts facial expression features from salient facial patches, can only be used for frontal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

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

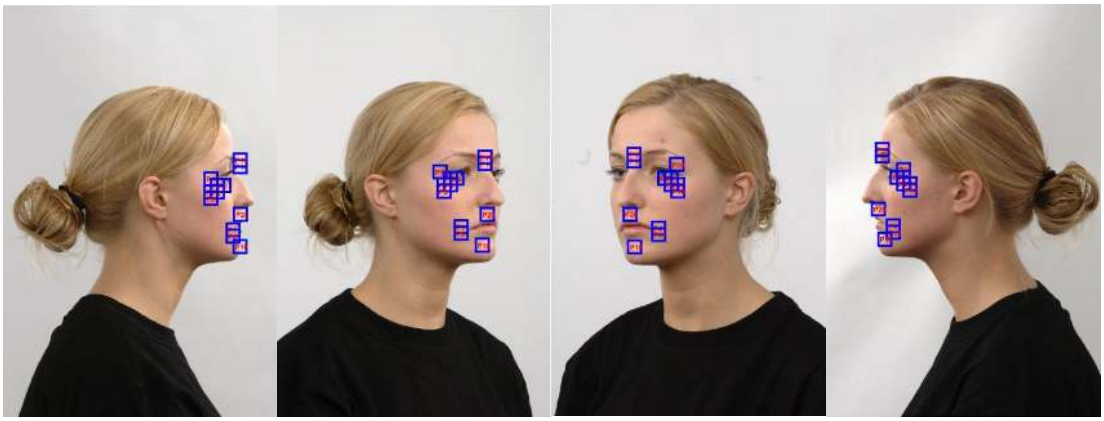
(a) (b)

Fig. 2 Positions of facial patches P_{18} and P_{19}. (a) As selected by the method of **Happy et al.**, **(b) as selected by the proposed method.**

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



(a)



(b)

Fig. 3 Positions of salient facial patches under variations in head pose. (a) Four face images with different head poses (left to right: 90°, 45°, −45°, and −90°), and **(b) positions of the salient facial patches in the corresponding face images.**

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

As shown in Table 1, when the viewing angles are increasing from 0° to 90°, the number of patches is decreasing from 20 to 12. So the feature dimension of patches in Happy et al.'s method is 19*M*N, and the feature dimension of patches in PSFP algorithm is only 12*M*N for non-frontal face images. The PSFP algorithm has the less computational cost. Additionally, we have analyzed the programming codes and found that PSFP algorithm has time complexity of $O(2n\log 2n)$.

3.3 Feature Extraction and Classification

After the salient facial patches have been obtained from the face images, the features of the facial patches need to be extracted for the classification. After the features have been 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), local binary pattern (LBP), and Gabor filters. These have appeared in many important studies[3,26] of non-frontal facial expression recognition. These methods are all adept in extracting local facial expression features from face images, so in our experiment we extracted features from salient facial patches in every image using each of the three methods separately to compare their recognition performance.

(1) HOG:

First, we break up the whole-face image into parts; second, we obtain a histogram from each cell; finally, we normalize the computed results and return a descriptor.

(2) LBP:

The $N \times N$ LBP operator is used to obtain the facial expression features. The operator weights of operator are multiplied by the corresponding pixels of the face image, and the summation of the $N \times N$ - 1 pixels are used for the LBP features of the neighborhood.

There are much variations of the LBP algorithm; in **Happy et al.** the highest recognition rate was attained using the 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 radius of the circular neighborhood.

(3) Gabor:

Gabor filters can be formulated as

$$\psi_{u,v}(z) = \frac{\|k_{u,v}\|^2}{\sigma^2} e^{-\frac{\|k_{u,v}\|^2 \|z\|^2}{2\sigma^2}} \left(e^{ik_{u,v}z} - e^{-\frac{\sigma^2}{2}} \right) \quad (7)$$

where u represents the orientation, and v represents the scale. If an image is convolved with a Gabor filter, the Gabor features will be extracted by the particular u and v values.

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

3.3.2 Classification

After the facial expression features have been extracted, the final task is feature classification. Non-frontal face images are hampered by a lack of emotion information, so if the classifier is weak, the recognition rate may be very low. For this problem, the adaptive boosting (AdaBoost)[27] algorithm is applied for the classification because it is good at combining many learning algorithms to improve recognition performance and is thus suitable method for the task of classification.

4 Results and discussion

4.1 Experimental Setting

This simulation environment used the MATLAB R2015b platform running on a Dell personal computer. We evaluated the PSFP algorithm on the radboud (RaFD) [25] database. RaFD is a publicly available and free 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 took 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 for the PSFP algorithm was implemented as shown in Fig. 4.

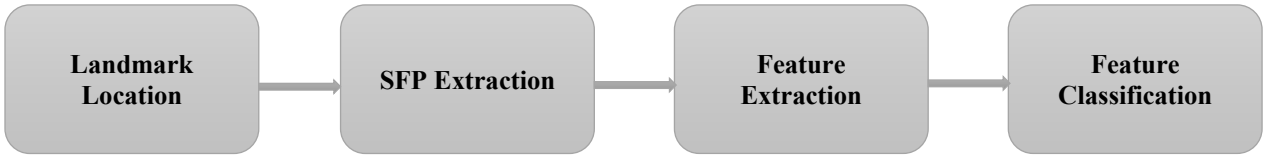


Fig. 4 Framework for the PSFP algorithm.

For the facial landmark location, Yu et al.'s method was used, and salient facial patches were extracted from the face images under five different head poses. Yu et al.'s method can estimate the head poses along pitch, yaw and roll directions. In our experiments, Yu et al.'s method was only need to estimate the head poses along yaw direction.

The size of the facial patches was typically set to 16×16. HOG, LBP ($P = 8, R = 1$) and Gabor filters ($u = 1$， $v = 1,2,\cdots,8$) were each 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 basic classifier of AdaBoost.M1. The maximum number of iterations was 100.

4.2 Purposes

In this study, experiments were used to validate the recognition performance of PSFP from four different perspectives.

4.2.1 Testing PSFP performance under different training–testing strategies

There are two commonly used experimental ways of performing non-frontal facial expression recognition: pose-invariant and pose-variant. In the former, training images and test images have under the same head pose, so head pose estimation can be avoided; in the latter, the training and test images may have different head poses, so it is more realistic. To analyze the recognition performance of the PSFP algorithm, two simulation experiments were performed, described in Sec. 4.3 and Sec. 4.4.

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 a rigorous proof. Therefore, it is necessary to use different parameter values for PSFP and observe the recognition performance on the test set. As described in Sec. 4.1, 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 can affect the expression recognition performance. Secs. 4.5.1 and 4.5.2 describe the experiments carried out for this performance comparison.

4.2.3 Comparing PSFP with SFP for frontal facial expression recognition

In Sec. 3.2, we discussed the two differences between the SFP method of **Happy et al.** and PSFP. Even if we were to replace the face detection method of SFP with Yu et al.'s method, this modified SFP method would still not 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 selected for facial salient patches. As PSFP and SFP should be compared with each other, it is necessary to perform the experiments for frontal facial expression recognition. The experiment described in Sec. 4.5.3 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 the 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 Sec. 4.5.4.

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 the experiments on person-dependent facial expression recognition, the subjects appearing in the training set also appear in the test set. Because every model has three different head poses, a three-fold cross-validation strategy was used for the person-dependent facial expression recognition.

The dataset can be divided into three segments according to head pose. Each time, two segments were used for training and the remaining segment for testing. Thus, the number of images in the training set was 160, and the number in the test set was 80, for each head rotation angle. The same

training–testing procedure can be repeatedly carried out three times and the average result of the three procedures is considered 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 recognition performance with 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. We also find that the best head rotation angle for recognition of non-frontal facial expressions is −45°.

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

In the experiments on person-independent facial expression recognition, the subjects appearing in the training set do not appear in the test set. For this reason, the leave-one-person-out strategy was used for the experiments: All photographs of one person are selected as the test set, and the remaining photographs in the dataset are used for training. Thus, the number of images in the training set was 216, and the number in the test set was 24, for each head rotation angle. This procedure was repeated 10 times, and the averaged result is taken as the final recognition rate.

The experiment results are shown in Table 3.

For most angles, **Gabor achieved the best recognition rate, and at 0°, −45° and 90°, Gabor and LBP achieved the best recognition rate. We find that the best head rotation angle for recognition of non-frontal facial expressions is 45°.**

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**

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 the facial expression recognition, the PSFP algorithm can achieve better recognition performance on non-frontal facial expression recognition.

4.4 Pose-Variant Non-frontal Facial Expression Recognition

Again, there are two training–testing strategies for facial expression recognition: person-dependent and person-independent.

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 10 times for each dataset, and the average values are taken as the final recognition rate. The experiment results are listed in Table 4.

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

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 size of facial patches

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 results in Figs. 5 and 6 are compared, we can see that 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.

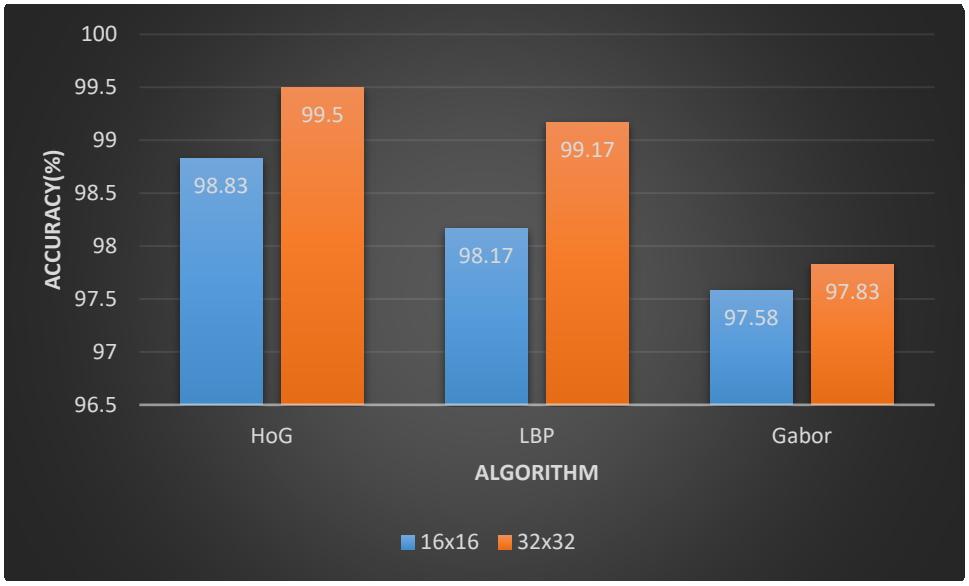


Fig. 5 Comparison of performance for person-dependent facial expression recognition under different facial patch sizes.

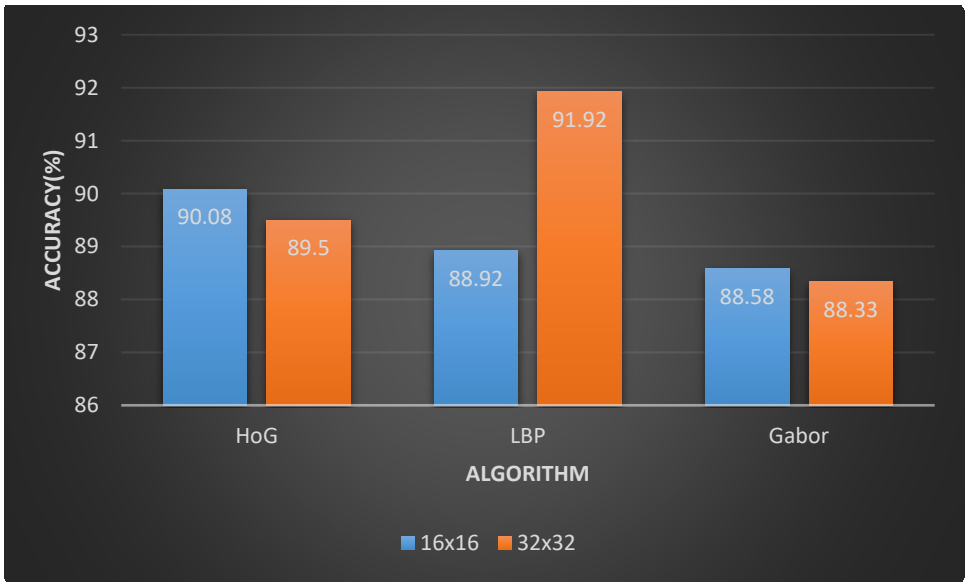


Fig. 6 Comparison of performance for person-independent facial expression recognition under different facial patch sizes.

4.5.2 Comparison by feature dimensionality

In the above experiments, the feature dimensionality was set to 10. We re-ran the experiments for pose-variant non-frontal facial expression recognition and allowed the feature dimensionality to

increase from 10 to 100. AdaBoost with NN was used as the classifier, and the feature extraction methods were HOG, LBP, and Gabor, shown separately.

The experiment results are shown in Figs. 7 and 8. As shown in the figures, the recognition rates grow 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 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, this leads individual difference which does great harm to recognition. When the feature dimension is increased, it is helpful to improve classified accuracy.

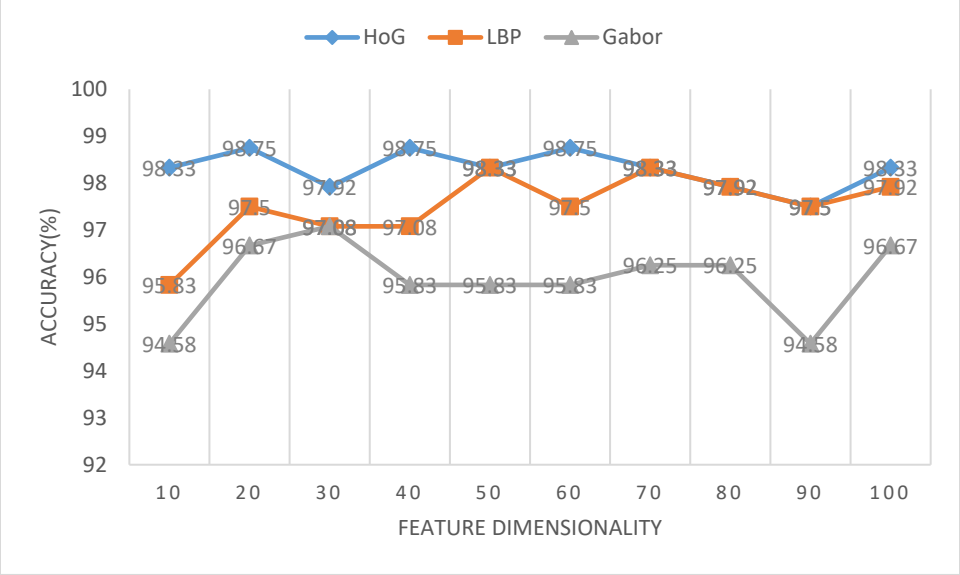


Fig. 7 Accuracy of person-dependent facial expression recognition by feature dimensionality.

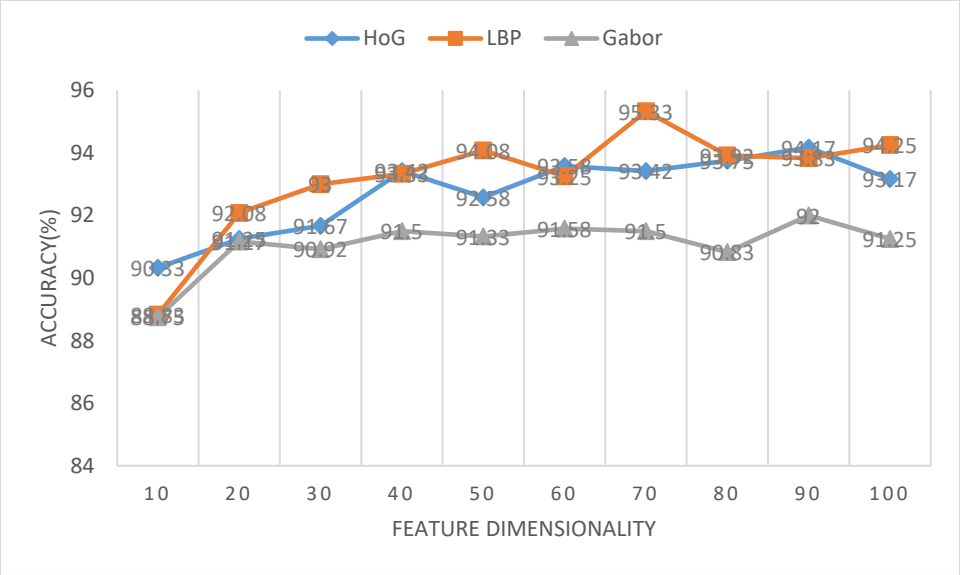


Fig. 8 Accuracy of person-independent facial expression recognition by feature dimensionality.

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 as small as possible while maintaining good performance.

4.5.3 Comparison with the SFP method of ***Happy et al.***

In order 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 the LBP parameters *P* and *R* were equal to 8 and 1,** respectively, the accuracy of PSFP was higher than that of the **SFP method of Happy et al**. This demonstrates that the PSFP method can also outperform SFP for frontal facial expression recognition.

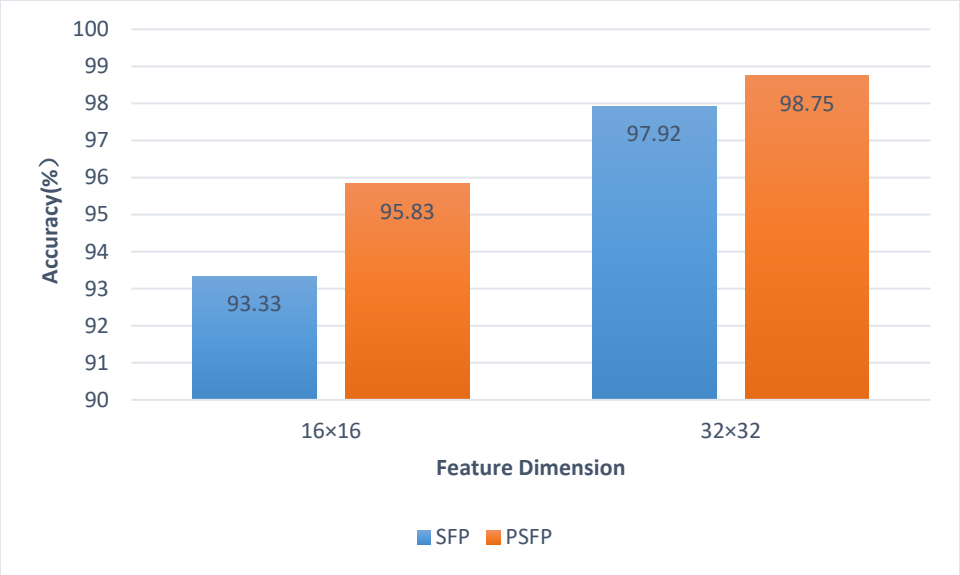


Fig. 9 Comparisons of SFP and PSFP for person-dependent facial expression recognition.

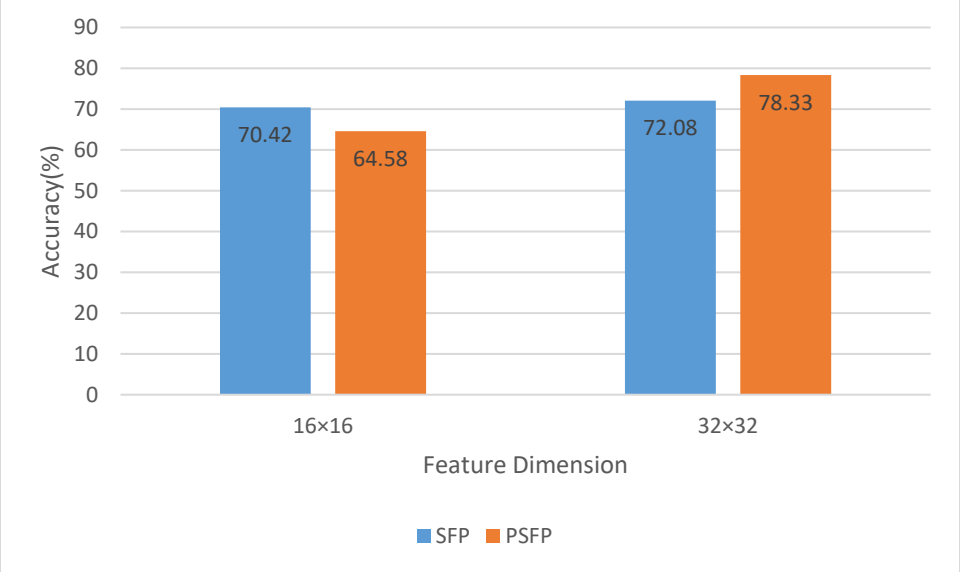


Fig. 10 Comparisons of SFP and PSFP for person-independent facial expression recognition.

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 -independent strategies are shown in Fig. 11 and Fig. 12.

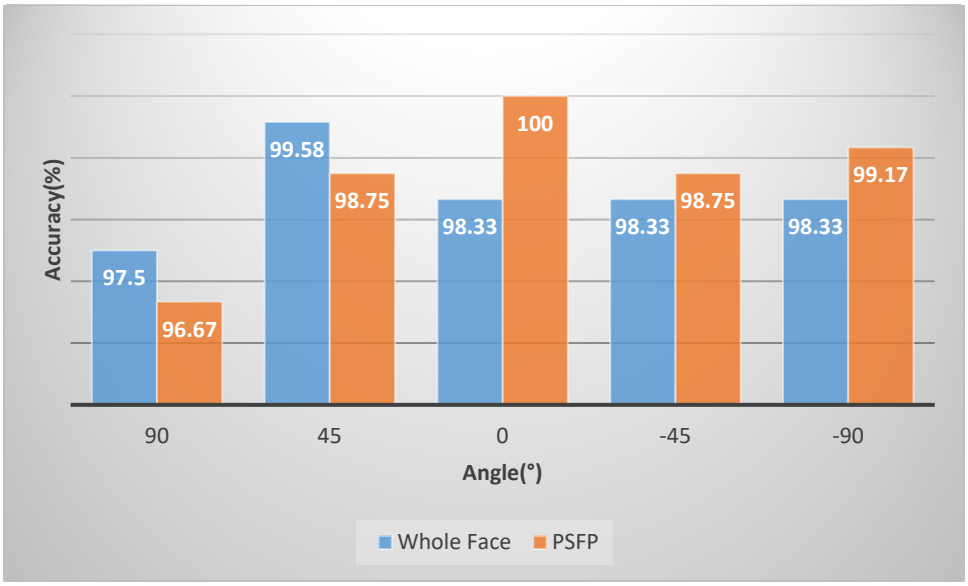


Fig. 11 Recognition rates for person-dependent facial expression recognition.

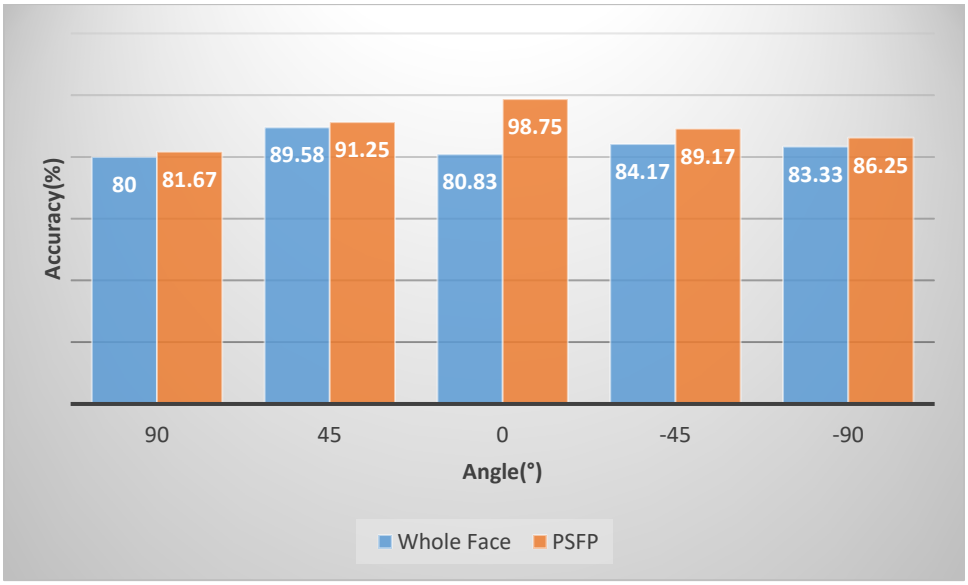


Fig. 12 Recognition rates for person-independent facial expression recognition.

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

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

As mentioned in the related work, several studies have employed salient patches with CNNs for face detection and classification. So we used CNN for non-frontal facial expression recognition. The CNN model was 21-layers VGG [28] and AlexNet [29]. 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 the Figure 13.

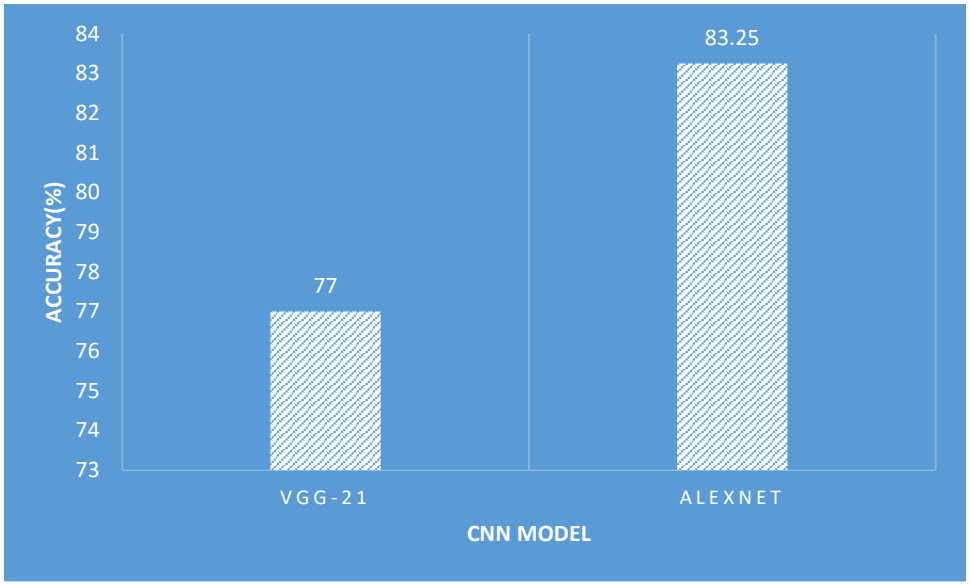


Fig. 13 Recognition rates for non-frontal facial expression recognition by using CNN.

In Fig 13, we can find that the recognition rate of VGG is lower than recognition rate of AlexNet. We also used facial patches as an image for training CNNs. However, this approach may be not suitable for recognition. CNNs usually need whole face image for training model. How to use patches with CNNs, and achieve good performance? It will be the next research plan.

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 the reason is that whereas LBP features are based on local image regions of the facial patch

and Gabor features are extracted from the whole-face patch, 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 can also be applied for frontal facial expression recognition. (It is an extension of the SFP method.);

(3) PSFP can achieve high recognition rates while consuming fewer data.

5 Conclusion

This paper has presented an algorithm based on salient facial patches, called PSFP. It employs the relevance of facial patches in non-frontal facial expression recognition, and uses the facial landmark detection method to track key points from the 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 could be extracted from the facial patches and finally used for feature classification. The experiment results show that PSFP can achieve high recognition rates while consuming fewer data.

6 Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

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

Competing interests

The authors declare that they have no competing interests

Funding

This work is 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.

Authors' contributions

Bin Jiang conceived the algorithm, designed the experiments, analyzed the results, and wrote the paper; Qiuwen Zhang, Zuhe Li and Qinggang Wu wrote the codes and performed the experiments; Huanlong Zhang was in charge of the overall research and contributed to the paper writing. The author(s) read and approved the final manuscript.

Acknowledgments

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

Author information

Bin Jiang received his MS degree from the Henan University in 2009, and his Ph.D. from the Beijing University of Technology 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 currently an Associate Professor. He has published over 30 technical papers in the field 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 is currently an Associate Professor at the Zhengzhou University of Light Industry. He received his MS degree in communication and information system from the Huazhong University of Science and Technology in 2008, and his Ph.D. degree in information and communication engineering from the Northwestern Polytechnical University in 2017.His current research interests include computer vision and machine learning.

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

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. 2008:1-4.
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. Proc. Int. Conf. European Conference on Computer Vision, 2010. ECCV. 2010:490-503.
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. FG. 2006:211-216.

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. FG. 2018:263-270.
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. Proc. Int. Conf. Applied and Theoretical Computing and Communication Technology, 2015. iCATccT.2015:847-851.
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. TECON. 2016:2806-2809.
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. ICMLC. 2016:502-507.
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, 1-8 (2017).
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, 2013. CVPR, 2013:3476-3483.
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. ICCV. 2013:1944-1951.
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. R. E. Schapire, A brief introduction to boosting. IEEE International Joint Conference on Artificial Intelligence, 1999. IJCAI. 1999:1401-1406.
28. K Simonyan, A Zisserman. Very deep convolutional networks for large-scale image recognition. International Conference on Learning Representations, May 2015, 1-4.
29. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks[C]// Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc. 2012:1097-1105.

Figures

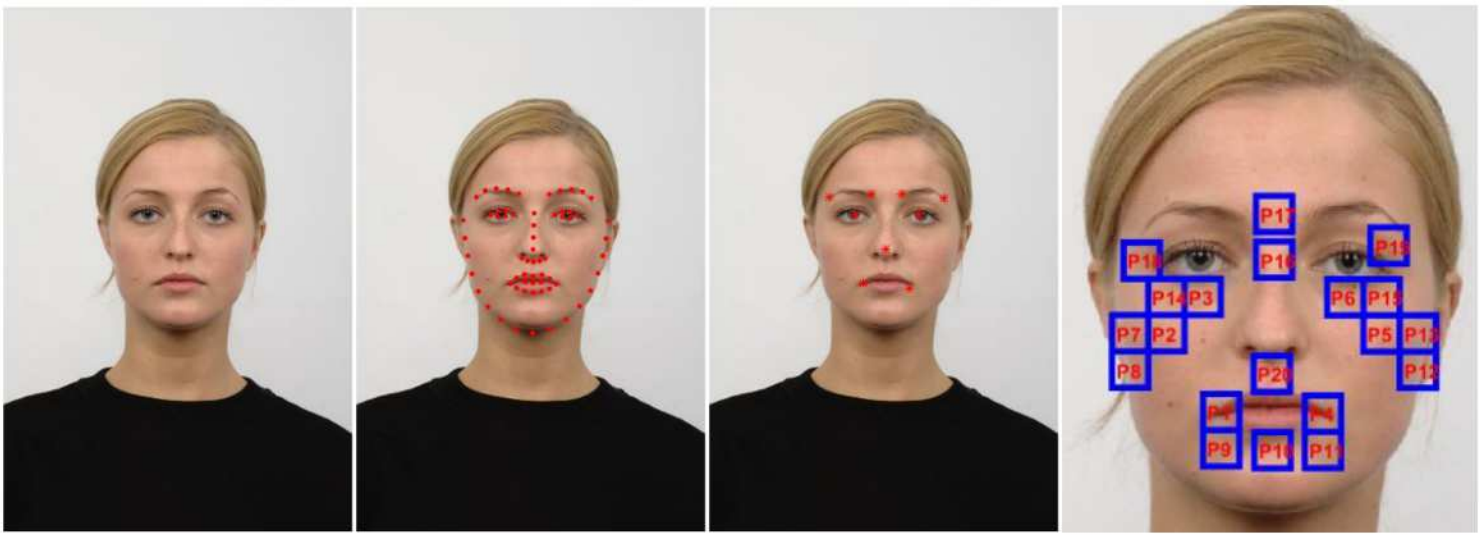


Figure 1

Framework for automated extraction of salient facial patches. (a) Face image from RaFD database,25 (b) the 66 facial landmarks detected using Yu et al.'s method,23 (c) the points of lip corners and eyebrows, and (d) locations of the salient facial patches.

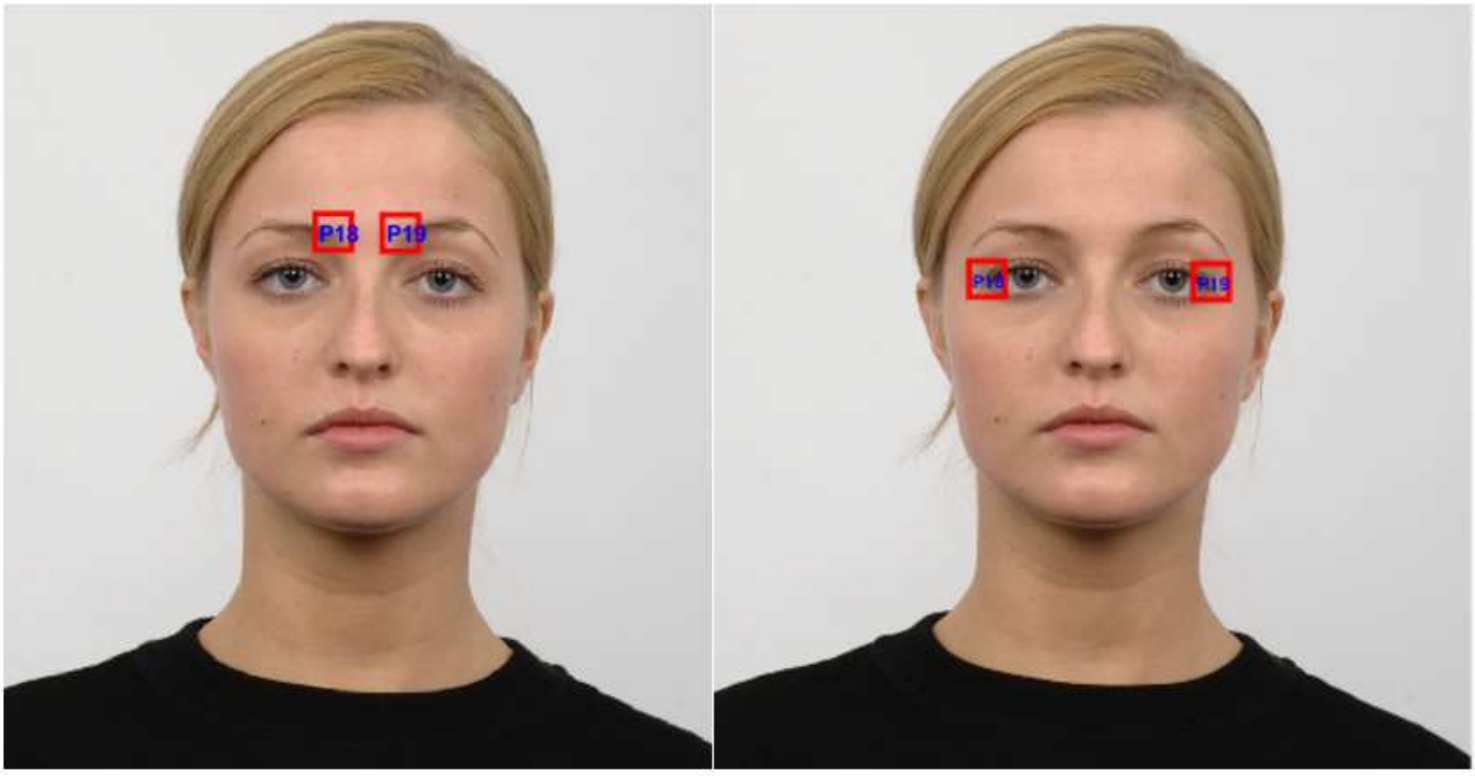
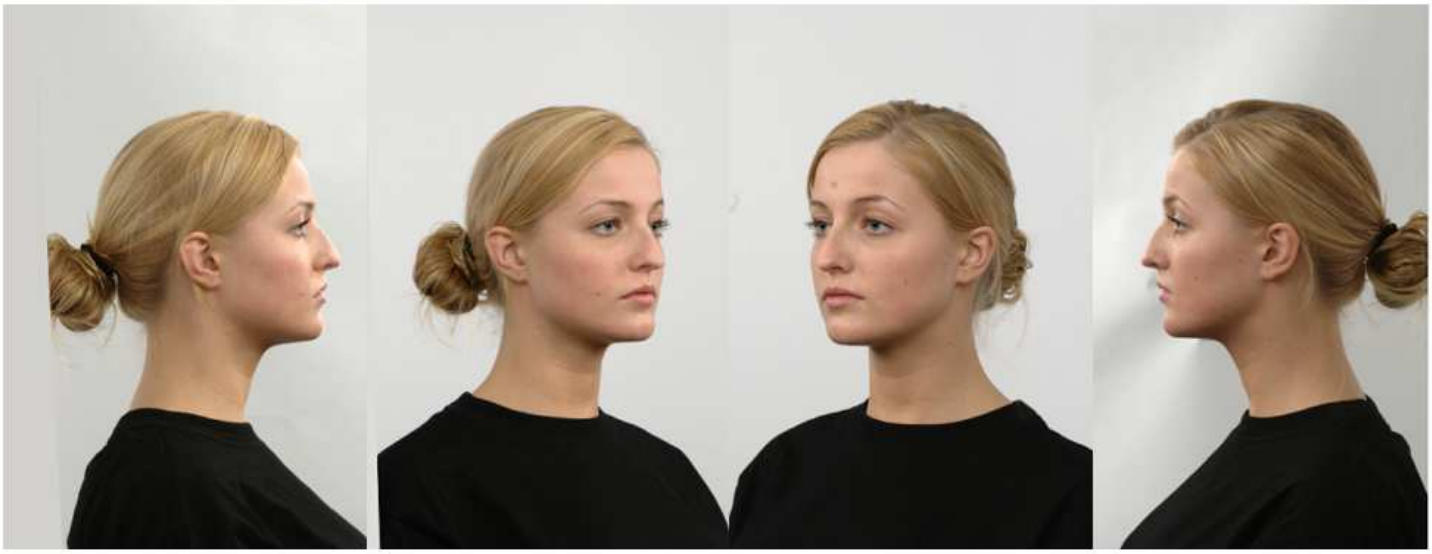
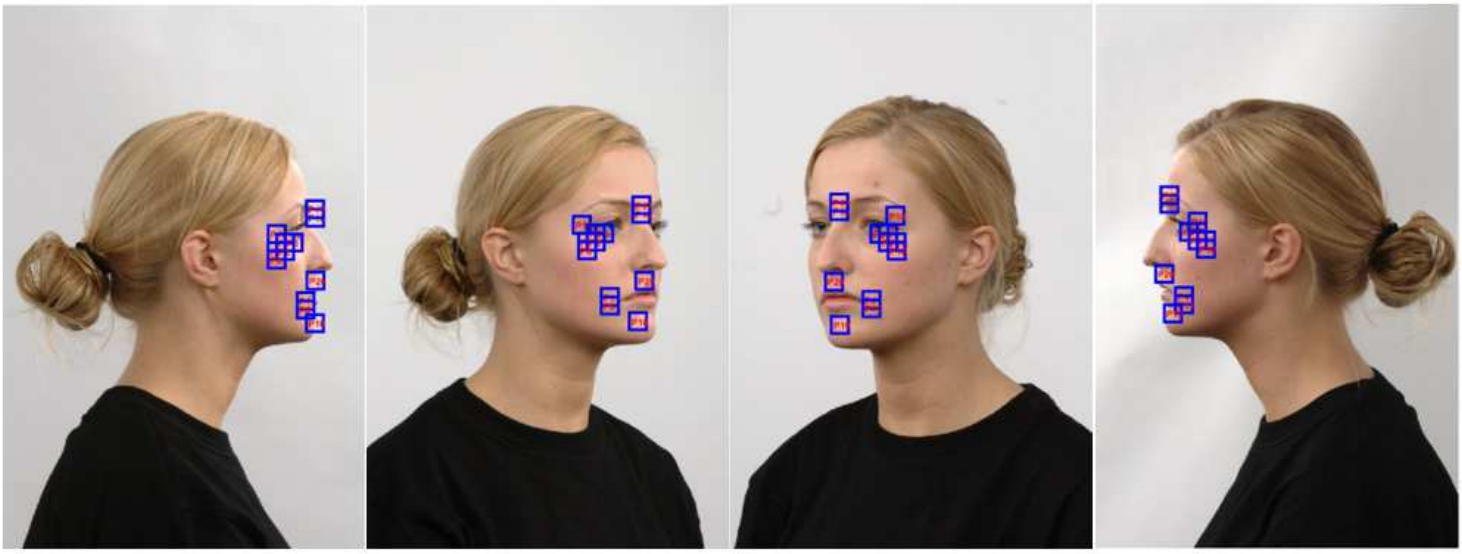


Figure 2

Positions of facial patches P18 and P19. (a) As selected by the method of Happy et al., (b) as selected by the proposed method.



(a)



(b)

Figure 3

Positions of salient facial patches under variations in head pose. (a) Four face images with different head poses (left to right: 90°, 45°, −45°, and −90°), and (b) positions of the salient facial patches in the corresponding face images.

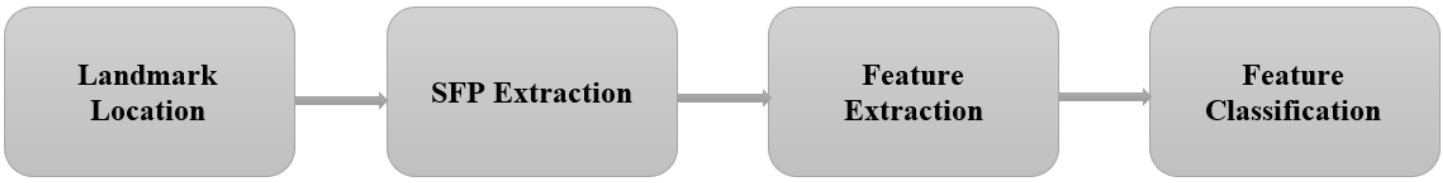


Figure 4

Framework for the PSFP algorithm.

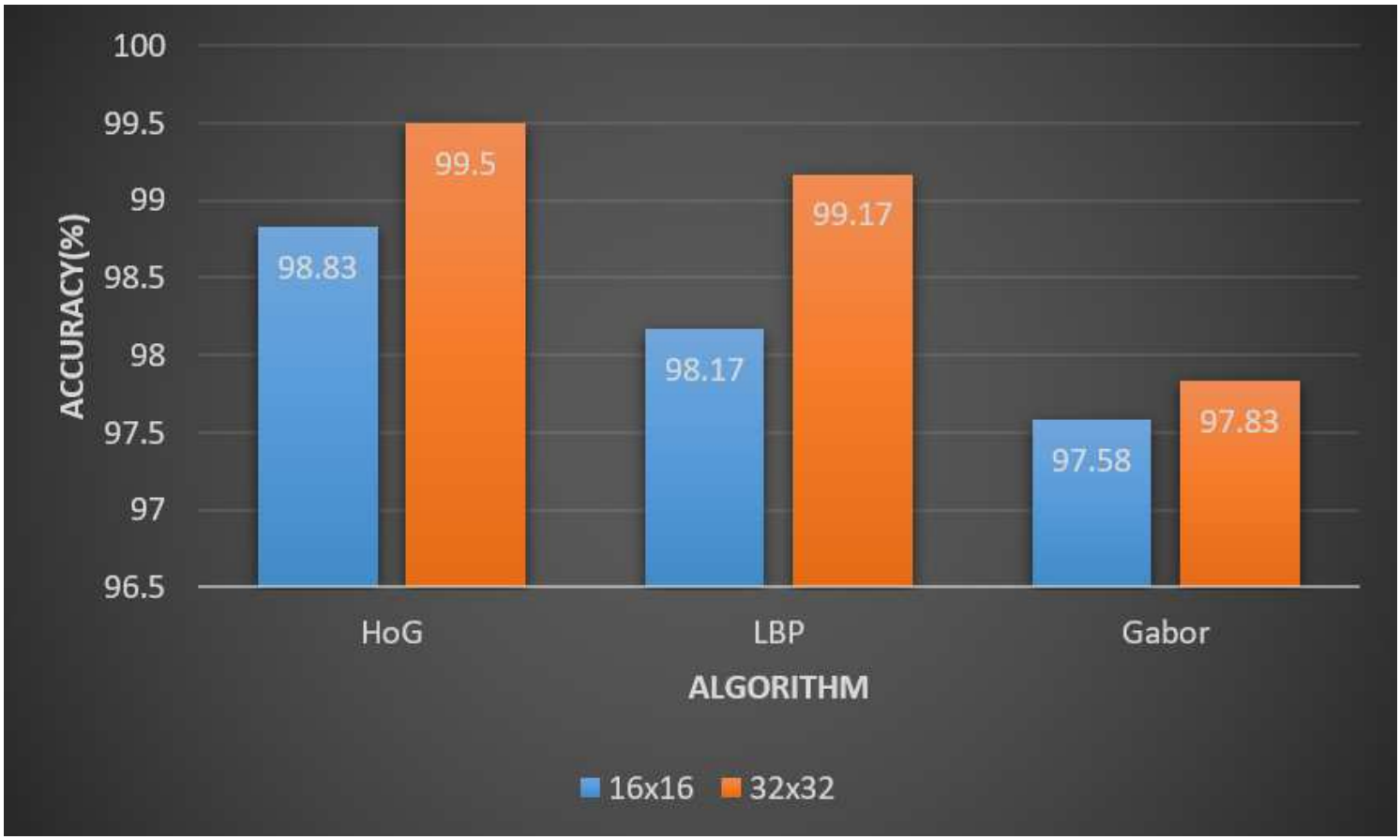


Figure 5

Comparison of performance for person-dependent facial expression recognition under different facial patch sizes.

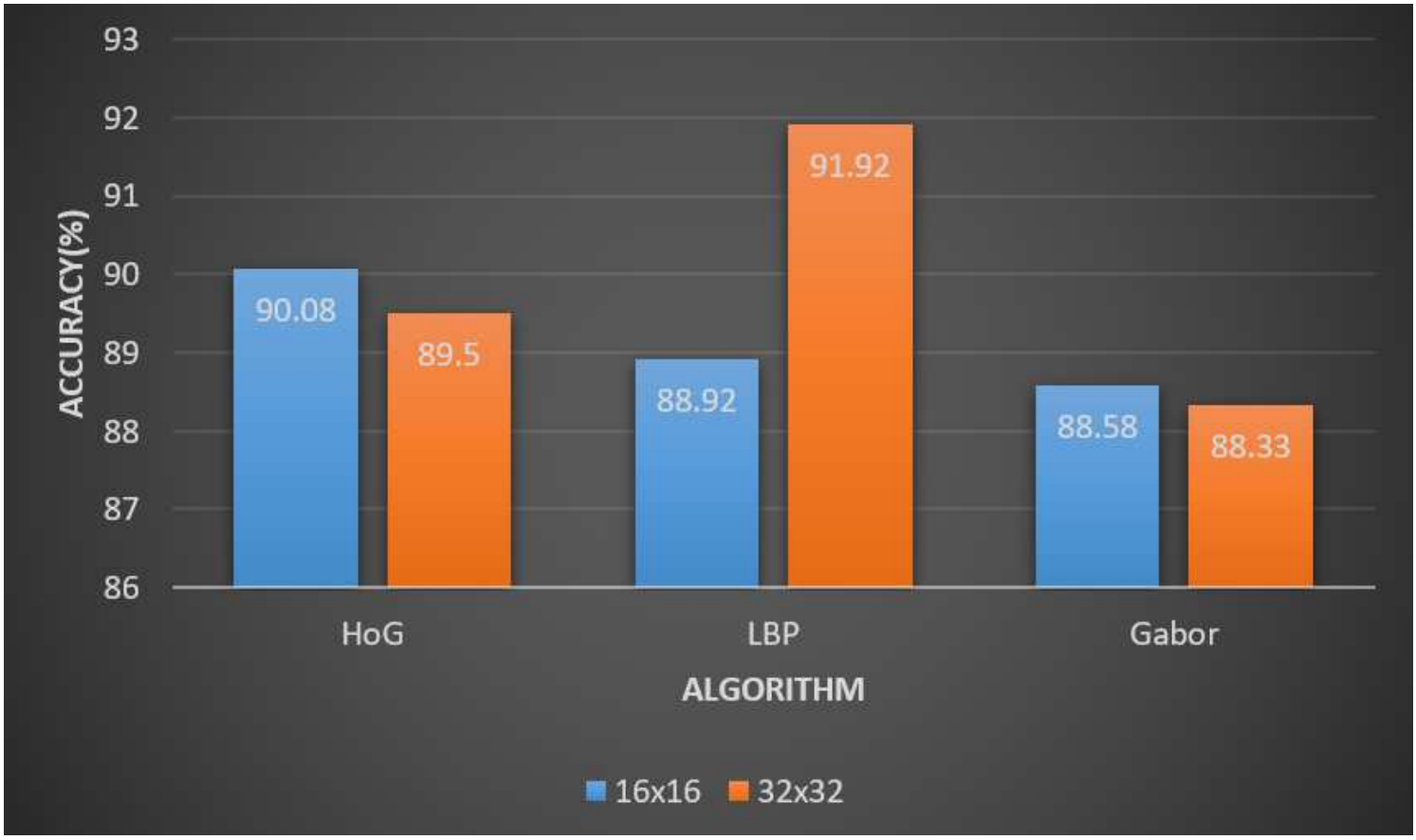


Figure 6

Comparison of performance for person-independent facial expression recognition under different facial patch sizes.

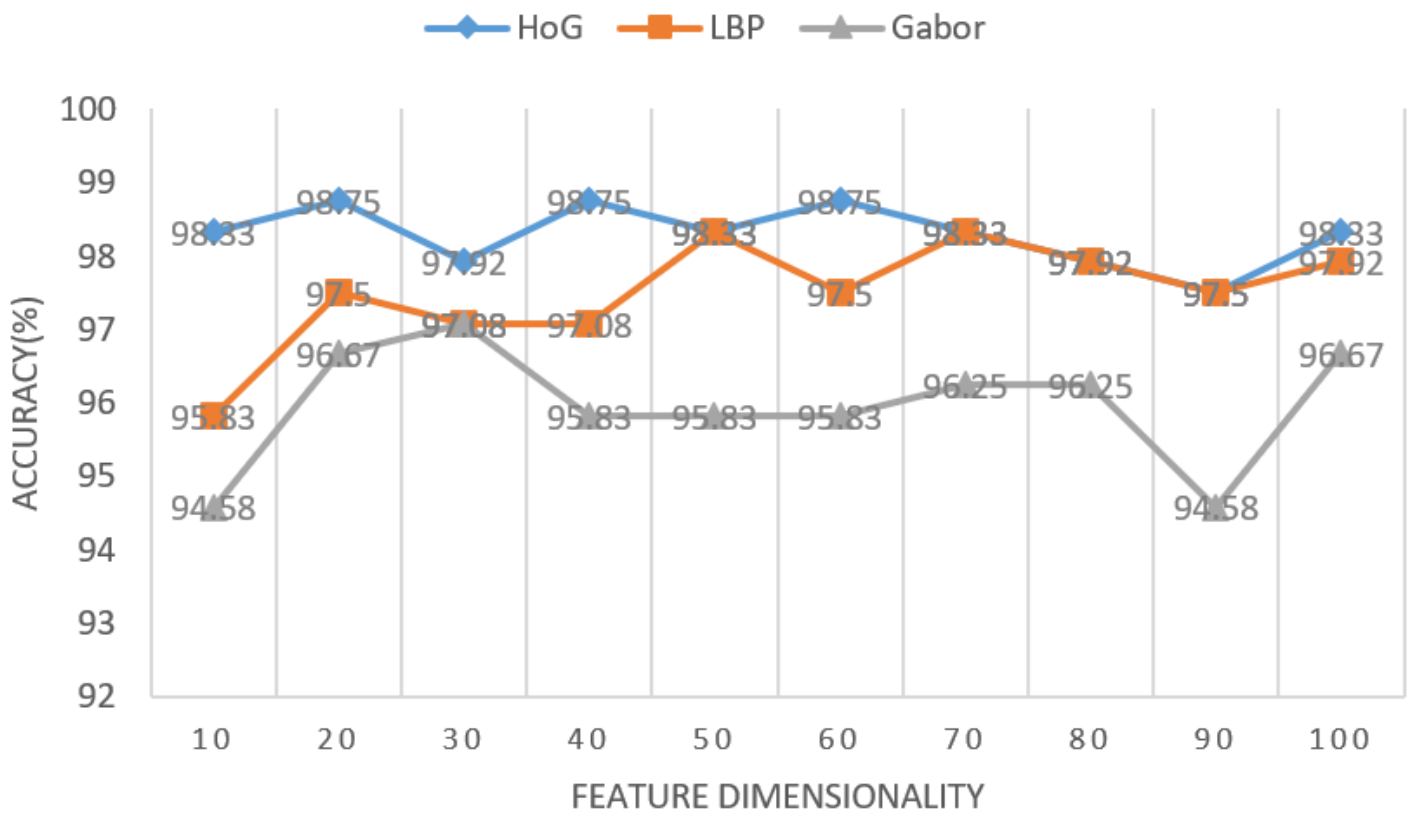


Figure 7

Accuracy of person-dependent facial expression recognition by feature dimensionality.

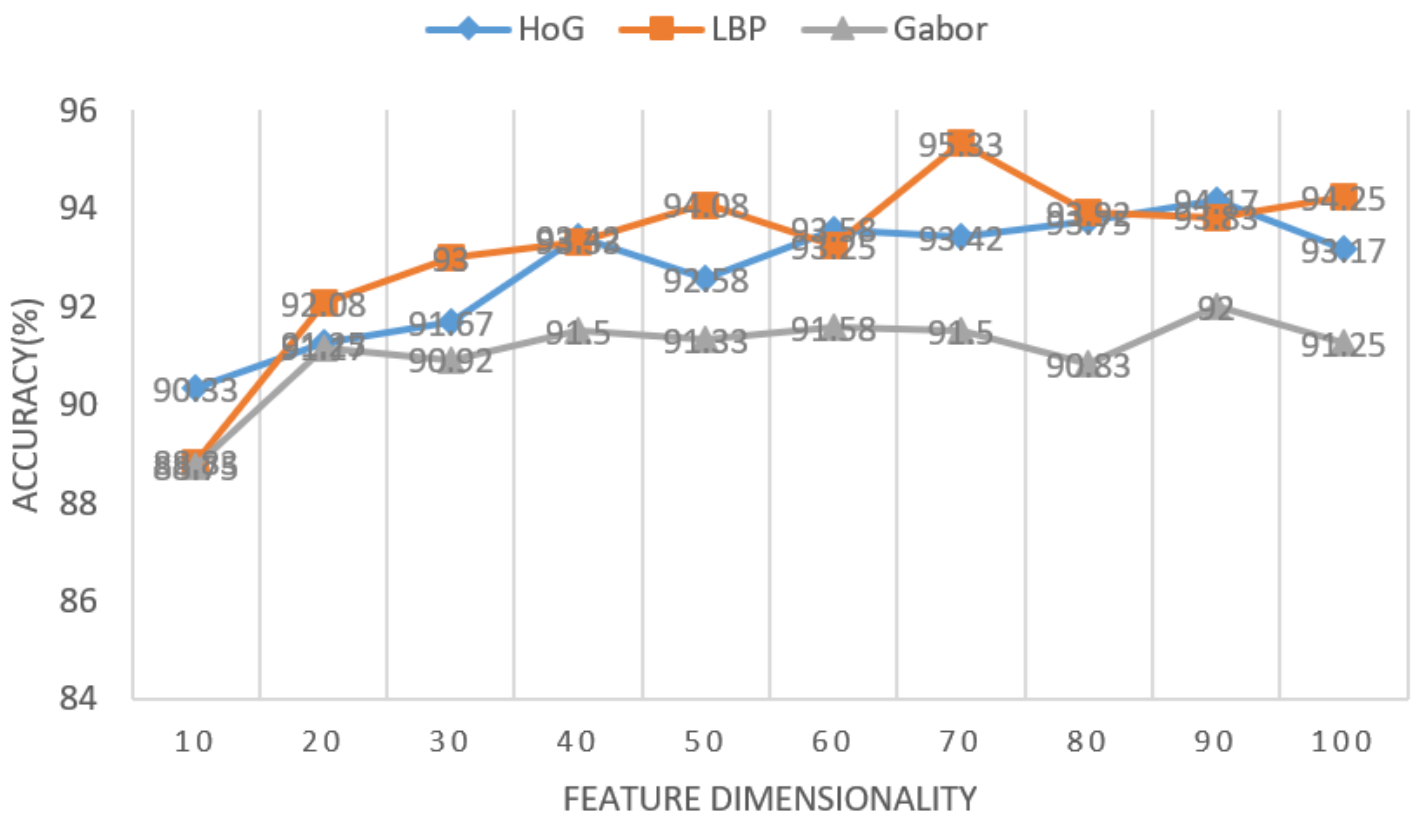


Figure 8

Accuracy of person-independent facial expression recognition by feature dimensionality.

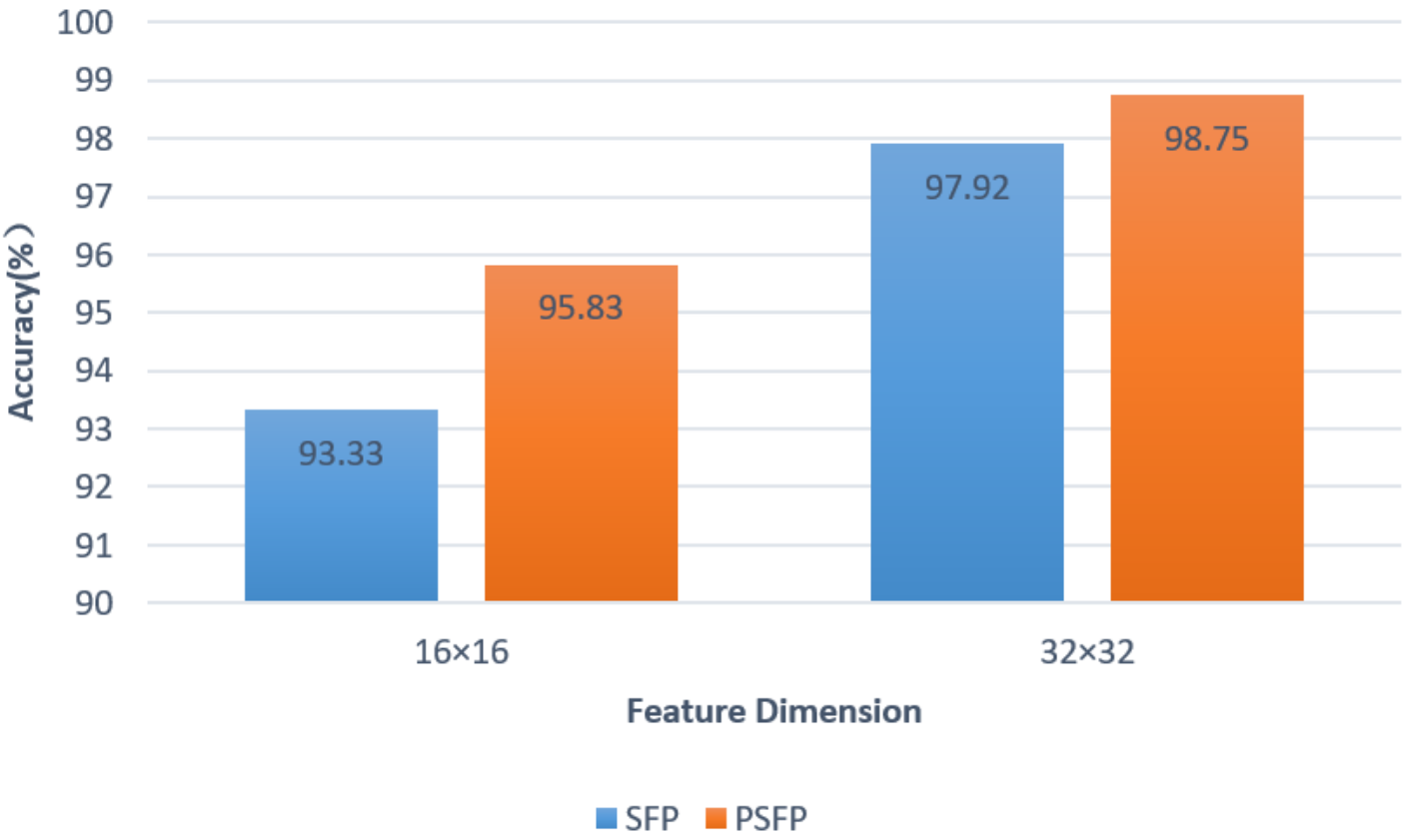


Figure 9

Comparisons of SFP and PSFP for person-dependent facial expression recognition.

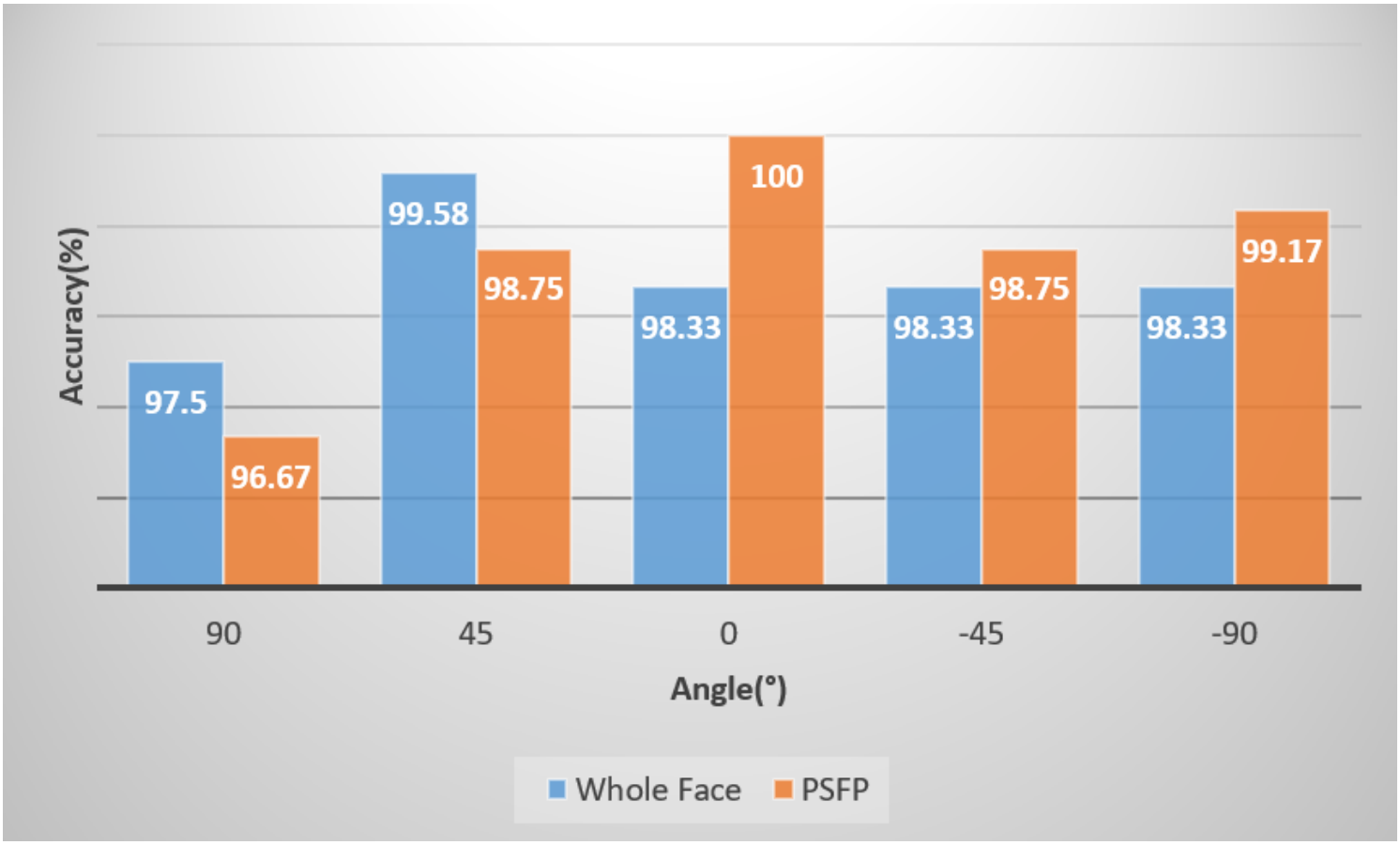


Figure 11

Recognition rates for person-dependent facial expression recognition.

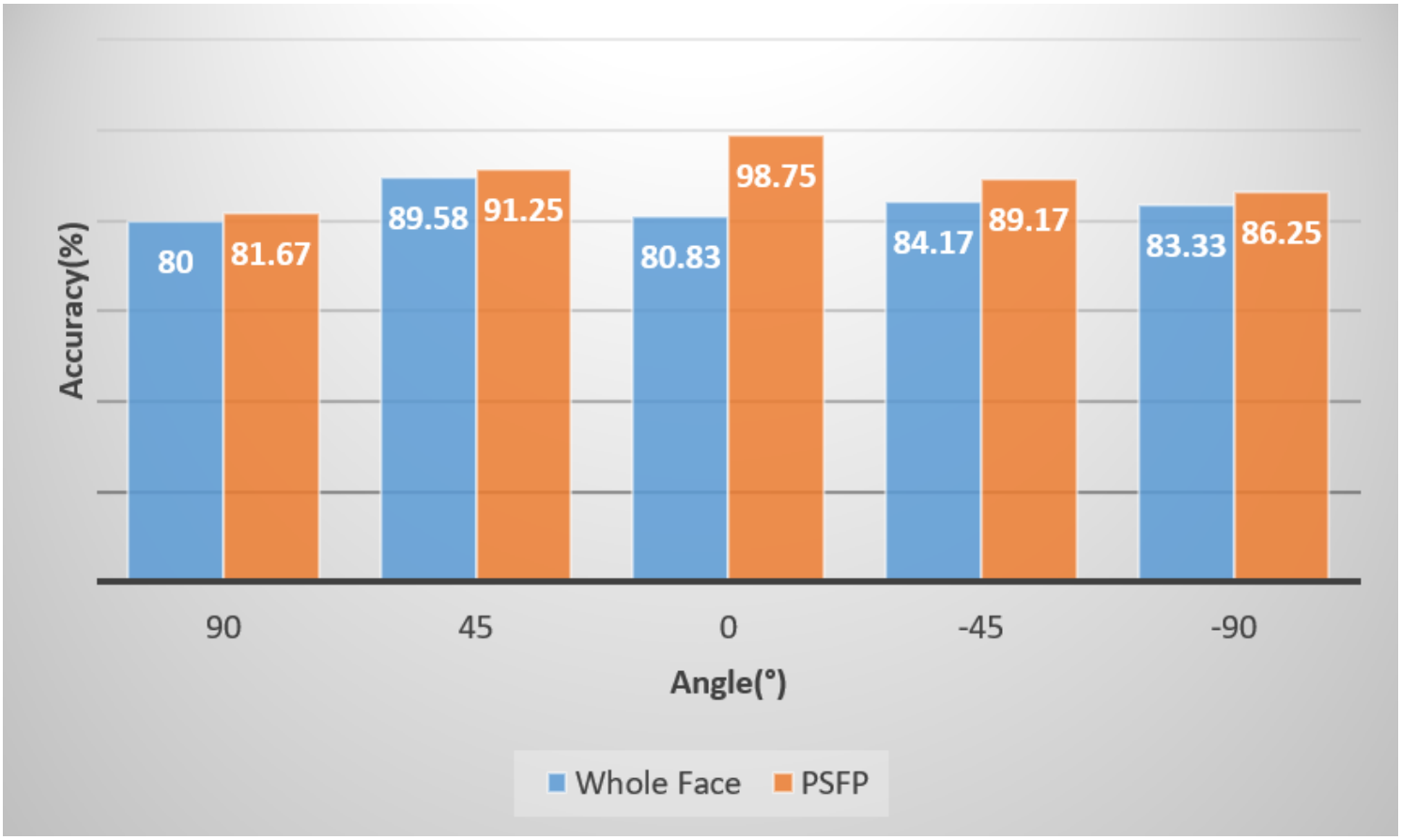


Figure 12

Recognition rates for person-independent facial expression recognition.

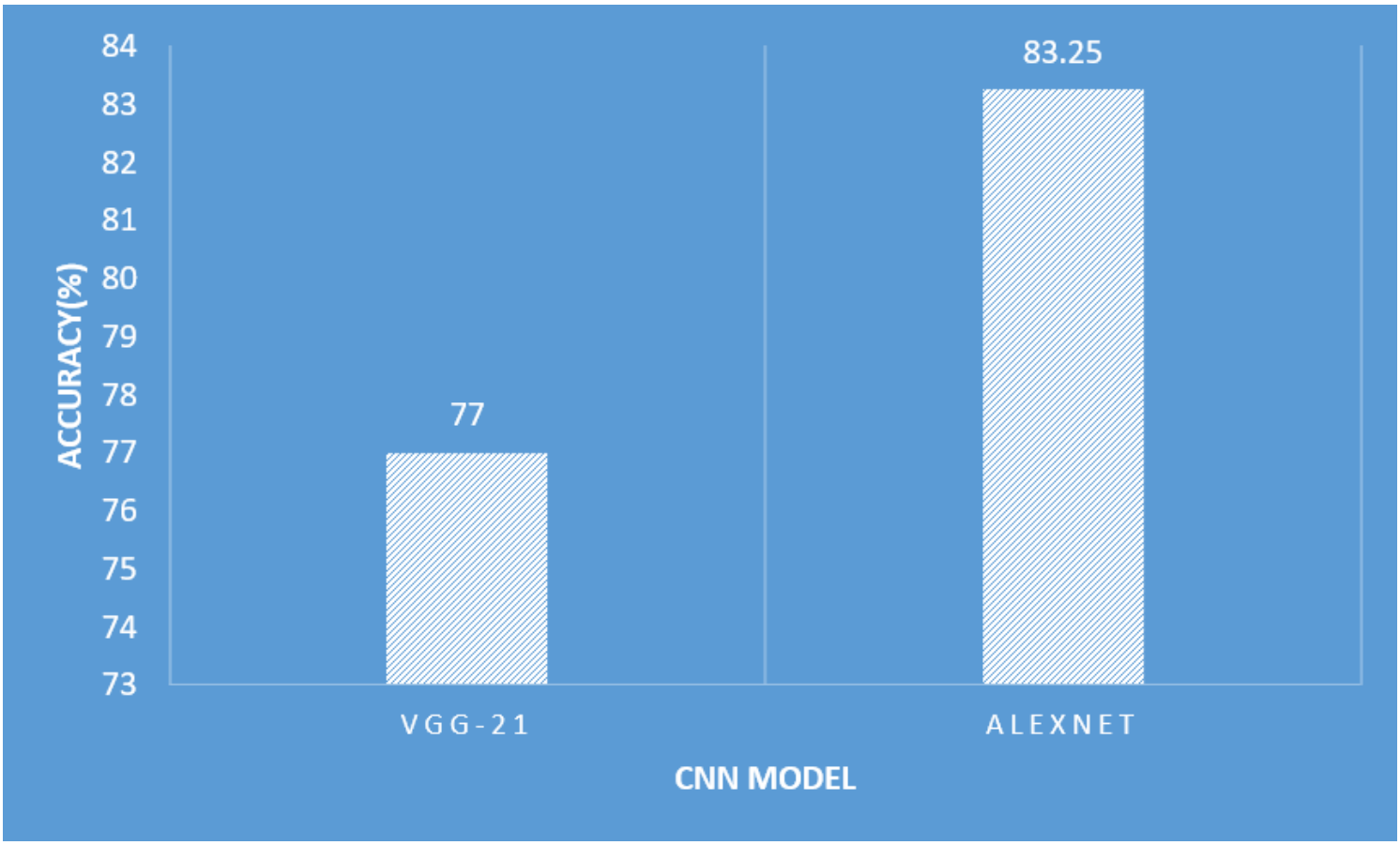


Figure 13

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