*Research Article*

# Facial Expression Recognition with LBP and ORB Features

**Ben Niu [iD],[1] Zhenxing Gao,[2] and Bingbing Guo[3]**

[1]*School of Electronic and Information Engineering, Jinling Institute of Technology, Nanjing 211169, China*
[2]*College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China*
[3]*School of Psychology, South China Normal University, Guangzhou 510631, China*

Correspondence should be addressed to Ben Niu; niuben@jit.edu.cn

Emotion plays an important role in communication. For human–computer interaction, facial expression recognition has become an indispensable part. Recently, deep neural networks (DNNs) are widely used in this field and they overcome the limitations of conventional approaches. However, application of DNNs is very limited due to excessive hardware specifications requirement. Considering low hardware specifications used in real-life conditions, to gain better results without DNNs, in this paper, we propose an algorithm with the combination of the oriented FAST and rotated BRIEF (ORB) features and Local Binary Patterns (LBP) features extracted from facial expression. First of all, every image is passed through face detection algorithm to extract more effective features. Second, in order to increase computational speed, the ORB and LBP features are extracted from the face region; specifically, region division is innovatively employed in the traditional ORB to avoid the concentration of the features. The features are invariant to scale and grayscale as well as rotation changes. Finally, the combined features are classified by Support Vector Machine (SVM). The proposed method is evaluated on several challenging databases such as Cohn-Kanade database (CK+), Japanese Female Facial Expressions database (JAFFE), and MMI database; experimental results of seven emotion state (neutral, joy, sadness, surprise, anger, fear, and disgust) show that the proposed framework is effective and accurate.

## 1. Introduction

Emotion recognition plays an important role in communication and is still a challenging research field. Facial expression is an indicator of feelings, providing valuable emotional information of human [1, 2]. Therefore, people can immediately recognize the emotional state of another person based on his/her facial expression. Consequently, information on facial expressions is often used in automatic systems of emotion recognition. With the development of technology, it has become convenient for us to solve problems with automatic systems, such as recognizing a person's emotion from a facial image. A machine can accurately interact with humans if it has the ability to recognize human emotions. The real-world applications that involve such interactions include human–computer interaction (HCI) [3, 4], virtual reality (VR) [5], augmented reality (AR) [6], advanced driver assistant systems (ADASs) [7], and entertainment [8, 9]. For facial expression recognition

system, faces are detected and recorded as 2D images by using various devices, such as electromyographs (EMGs), electrocardiographs (ECGs), electroencephalographs (EEG), and cameras.

Based on these 2D images, there are three main research directions in facial expression recognition. The first approach defines the movement of facial muscles as an action unit (AU), and the facial expression is represented by the movements of several AUs. According to the movements of AUs from an input image, the corresponding facial expression can be determined by decoding the detected AUs. However, since muscle movements are often slight, detecting AUs has accurate problems [10], which have severely restricted the development of this approach. The second approach is feature-based methods, which include three steps. In the first stage, the face is detected from an input image, and landmarks are identified in the face area. In the second stage, features are extracted from the face region by using Histogram of Oriented Gradient (HOG), LBP, and

Gabor wavelet methods. In the third stage, facial expressions are finally recognized and classified by using classifiers such as SVMs, AdaBoost, and random forests. The third approach is DNN-based methods, which perform very well in facial expression recognition. Different from conventional approaches, a feature map of the input image is produced through filter collection in the convolution layers [11–13].

Although DNN-based methods generally perform better than conventional methods, they have problems related to the processing time and memory consumption. Considering low hardware specifications used in some real-life conditions, to gain better results without DNNs, we aim to recognize seven basic emotional states (neutral, happy, surprise, sadness, fear, anger, and disgust) based on facial expressions using the second (feature based) approach. As we knew, the LBP is a powerful feature for texture representation. When facial expression appears, different textures are generated by the actions of the related muscles on the face. For this reason, we chose LBP in our study for facial expression recognition. In addition, to achieve the requirements of real time that are usually overlooked in previous studies, we combine LBP and OBR, since the ORB has been shown to have high computing speed.

However, the traditional ORB is more concentrated on the distribution of the extracted feature points. As traditional ORB, abundant feature points are extracted to achieve a higher accuracy, but the excessively dense feature points are very inconvenient for feature description. To solve this problem, we innovatively employed region division in the traditional ORB. This improved ORB algorithm calculates the number of feature points extracted from each region according to the total number of feature points to be extracted and the number of regions to be divided. In the present study, we did not just combine LBP and ORB, but combined LBP with improved ORB which solved the problem of feature point overlap and redundancy in the extraction process. Moreover, to test our improved LBP and ORB algorithm, we conducted experiments on several facial expressions in CK+ database, JAFFE database, and MMI database.

This paper is organized as follows. Section 2 summarizes some of the previous works. Section 3 presents the algorithm framework of facial expression recognition. Faces are detected and extracted using the Dlib library because of its fast processing speed, and LBP and improved ORB features are extracted separately from the face region. Then, the Z-score method is used to fuse both types of features. The facial expression databases are given in Section 4, and the experimental results and comparison with state-of-the-art methods are presented in Section 5. Finally, conclusions are stated in Section 6.

## 2. Related Work

We describe related works of facial expression recognition systems that have been studied to date. These algorithms can largely be divided into three directions: the geometric feature extraction method, the appearance feature extraction method, and the deep learning-based method. The geometric feature extraction method extracts geometric elements of the facial structure and motion of the facial muscles, the appearance feature extraction method extracts the features of the entire facial criterion, and the deep learning-based method uses convolutional neural network (CNN) to achieve the automatic learning of the extracted facial features. Some of the recent algorithms involving the aforementioned algorithms are described in the following subsections.

The geometric features algorithm extracts the temporal or dynamic changes of the landmark of the face. Geometric relationship between landmark points which are detected from face regions are exploited for expression representation. Tang et al. designed geometric features based on psychology and physiology [14]. Liang et al. focused on fine-grained facial expression recognition in the wild and built a brand-new benchmark FG-Emotions, extended the original six classes to more elaborate thirty-three classes, and proposed a new end-to-end Multiscale Action Unit- (AU-) Based Network (MSAU-Net) for facial expression recognition [15]. Zhang et al. proposed an end-to-end deep learning model that allows to synthesize simultaneous facial images and pose-invariant facial expression recognition by exploiting shape geometry of the face image [16]. However, when the landmark points are detected incorrectly, the recognition accuracies will be decreased.

The appearance features are extracted from facial image intensities to represent a discriminative textural pattern. Mandal et al. extracted the adaptive positional thresholds for a facial image; threshold parameters in the local neighborhood can be adaptively adjusted for different images; then multidistance magnitude features are encoded. SVM is used as a classifier for the facial expression recognition [17]. Tsai combined the Haar-like features method with the self-quotient image (SQI) filter for facial expression recognition. The angular radial transform (ART), the discrete cosine transform (DCT), and the Gabor filter (GF) are simultaneously employed in the system. An SVM is used to achieve the best results [18]. Chen et al. used visual modalities (face images) and audio modalities (speech) for facial expression recognition. They proposed a descriptor named Histogram of Oriented Gradients from Three Orthogonal Planes (HOG-TOP) to extract dynamic textures from video sequences to characterize facial appearance changes [19]. Ryu et al. used a coarse grid for stable codes (highly related to nonexpression), and a finer one for active codes (highly related to expression) as the features for facial expression recognition [20]. Liu et al. extracted LBP and HOG features from the salient areas that were defined on the faces; then Principal Component Analysis (PCA) is used to reduce the dimensions of the features which fused with LBP and HOG [21]. Ghimire et al. divided the whole face region into domain-specific local regions; then region-specific appearance features and geometric features are extracted from the domain-specific regions for facial expression recognition [22]. Nigam et al. retrieved Histogram of Oriented Gradients (HOG) feature in DWT domain and an SVM was used for expression recognition [23]. Kamarol et al. proposed spatiotemporal texture map (STTM) to capture subtle spatial

and temporal variations of facial expressions with low computational complexity. And an SVM was used for classification [24]. Turan and Lam performed 27 local descriptors while the Local Phase Quantization (LPQ) and Local Gabor Binary Pattern Histogram Sequence (LGBPHS) achieved the best results [25]. Bougourzi et al. combined histogram of oriented gradients, local phase quantization, and binarized statistical image features to recognize the facial expressions from the static images [26]. Meena et al. used graph signal processing along with the curvelet transform for recognizing the facial expressions. Not only the dimension of the feature vectors has been reduced but also recognition of the facial expression has been significantly improved [27]. Ashir et al. used the compressive sensing technique with statistical analysis of the extracted compressed facial signal to represent a more robust feature representation for each individual facial expression class [28].

Recently, deep learning is widely used in many computer vision tasks including facial expression recognition and achieves remarkable success. The facial expression features are learned automatically by Convolutional Neural Networks (CNN). Dong et al. introduced dense connectivity across pooling to enforce feature sharing in a relatively shallow CNN structure for effective representation of facial expressions under limited training data [29]. Xie et al. embedded the feature sparseness into deep feature learning to boost the generalization ability of the convolutional neural network for facial expression recognition [30]. Yu et al. designed a multitask learning framework for global–local representation of facial expressions [31]. Wen et al. designed a new neural network with domain information loss and dynamic objectives learning; the network can avoid the gradient disappearance and obtain the higher semantic features [32]. Wang et al. proposed a simple yet efficient Self-Cure Network (SCN) which suppresses the uncertainties efficiently and prevented deep networks from overfitting uncertain facial images [33]. Although deep learning based methodology may achieve higher recognition rate, more computing resources and data are needed. For these reasons, our research uses the former framework.

## 3. Methodology

In this section, the proposed methodology is introduced in detail. As shown in Figure 1, before LBP and ORB features extraction, the input images should be preprocessed (face detection and face extraction). The LBP is a fine-scale descriptor that captures small textural details. Local spatial invariance is achieved by locally pooling (e.g., with a histogram). Given that this approach is very resistant to lighting changes, LBP is a good choice for coding the fine details of facial appearance information over a range of coarse scales. In addition, ORB methods are 2 orders of magnitude faster than SIFT methods, and they perform as well in many situations. Moreover, our improved ORB avoided feature point overlap and redundancy in extraction processing. Then, features fusion of LBP and improved ORB were performed. Finally, by using the SVM, the classification accuracy was calculated to estimate whether our method can achieve outstanding results in facial expression recognition.

*3.1. Face Detection and Face Extraction.* Many algorithms have been proposed for face detection, such as the Haar cascade algorithm, HOG method with an SVM, and CNN models. In this paper, we use the Dlib library to detect faces because of its fast processing speed as shown in Figure 2. The Dlib library contains a facial landmark detector with a pretrained model [34, 35]. This model marks 68 points on the face, and the points specify regions of the face. The jaw line is highlighted using points 1–17, the left eyebrow is highlighted using points 18–22, the right eyebrow is highlighted using points 23–27, the left eye region is highlighted using points 37–42, the right eye region is highlighted using points 43–48, the region of the nose is highlighted using points 28–36, the outer lip area is highlighted using points 49–60, and the inner lip structure is highlighted using points 61–68. In this method, a regression tree model is created to find these facial landmark points directly from the pixel intensities without feature extraction; thus, the detection process may be fast enough to overcome the accuracy and quality problems associated with real-time analysis.

Usually, emotions are mainly conveyed via the eyes, nose, eyebrows, and some facial regions, so other parts of the face, such as the ears and forehead, can be excluded in further analysis. This face detection algorithm is appropriate for obtaining exact facial regions, and points 1–27 are used to extract the face region from original images.

*3.2. Fusion with LBP and ORB Features*

*3.2.1. Local Binary Patterns (LBPs).* LBPs, which were introduced by Ojala, can effectively describe the texture information of an image [36–38]. The LBP operator is defined for $3 \times 3$ neighborhoods, takes each pixel as the central pixel, and assesses the 8 pixels around the selected pixel based on a given threshold. The resulting binary-valued image patch forms a local image descriptor [39]. The LBP operator takes the following form:

$$\text{LBP}(x_c, y_c) = \sum_{n=0}^{7} 2^n s(i_n - i_c). \tag{1}$$

For the 8 neighbors of the central pixel $c$, $i_c$ and $i_n$ are the gray values at $c$ and $n$, and $s(u)$ is 1 if $u \geq 0$ and 0 otherwise. An original LBP operator that is shown in Figures 3 and 4 depicts the original LBP features of facial expressions.

If an LBP operator contains at most one 0-1 and one 1-0 transition in a binary code, then a uniform pattern exists. The uniform pattern contains primitive structural information for edges and corners. This information can be used to reduce the length of the feature vector and implement a simple rotation-invariant descriptor. In our research, a uniform-pattern LBP descriptor is applied to obtain features from faces, and the length of the feature vector for a single cell can be reduced from 256 in the traditional method to 59. The size of the face region is $130 \times 130$, and the LBP face is
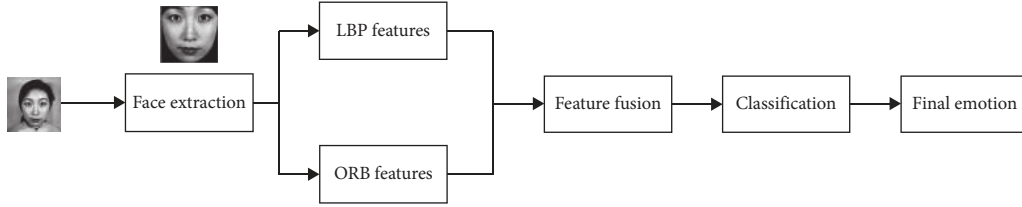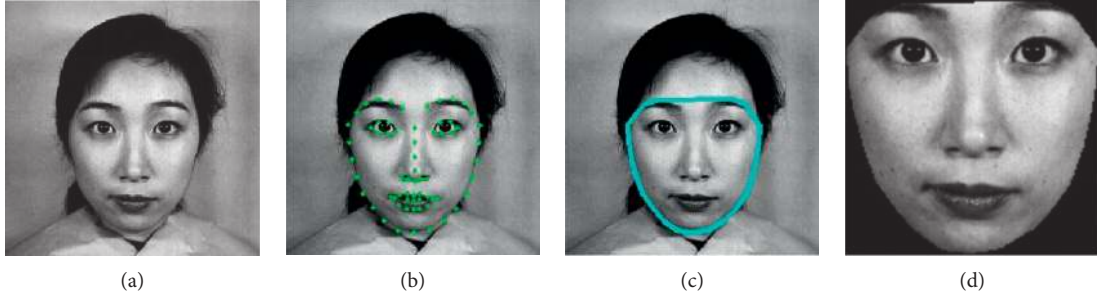
FIGURE 1: Flow diagram of emotion recognition.



FIGURE 2: The framework of face detection and face extraction. (a) The original image. (b) The detected face with 68 landmarks. (c) The face region using points 1–27. (d) The extracted face region.
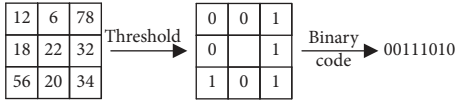


FIGURE 3: Original LBP operation.

$128 \times 128$, which is separated into small $16 \times 16$ patches with a resolution of $8 \times 8$. The uniform LBP features are extracted from each small patch and mapped to a 59-dimensional histogram.

*3.2.2. Oriented FAST and Rotated BRIEF (ORB) Features.* FAST features are widely used because of their computational properties. The FAST operation uses one parameter, the intensity threshold between the center pixel and each pixel in circular ring around the center, as a simple but effective measure of corner orientation. This parameter is called the intensity centroid. The intensity centroid assumes that the intensity of a corner is different from that of a pixel center, and this vector may be used to determine an orientation [40].

*(1) FAST Detector.* The FAST detector first proposed by Rosten is widely used in corner detection for computer vision because of its rapid operations and low computational complexity compared to other corner detectors. The segment test criterion is based on analyzing a circle of sixteen pixels around the candidate corner $p$. The original detector classifies $p$ as a corner if a set of $n$ contiguous pixels that are brighter than the intensity of the candidate corner $I_p$ plus a threshold $t$ or are darker than $I_p$ minus $t$ exists in the circle. Notably, $n$ is chosen as 12 because it yields a high-speed test that can be used to exclude a very large number of

noncorners. The expression of the FAST detector is presented as follows:

$$
S_{p \longrightarrow x} = \begin{cases} d, & I_{p \longrightarrow x} \leq I_p - t, \\ s, & I_p - t < I_{p \longrightarrow x} < I_p + t, \\ b, & I_p + t \leq I_{p \longrightarrow x}, \end{cases} \tag{2}
$$

where $I_p$ is the intensity of $p$, $I_{p \longrightarrow x}$ is the intensity of the sixteen pixels around the corner, and $t$ is a threshold. If $S_{p \longrightarrow x}$ is equal to $d$, the pixel belongs to the darker group; if $S_{p \longrightarrow x}$ is equal to $s$, the pixel belongs to the similar group. If $S_{p \longrightarrow x}$ is equal to $b$, the pixel belongs to the brighter group. If 12 continuous pixels that belong to the darker or brighter group exist, $p$ is regarded as a corner.

The corners are determined when all image pixels are tested using the above process. The corners will converge in some areas. To find the most robust corners, a non-maximal suppression method based on a score function is adopted. The score values of each detected corner are calculated, and the corners with the low score values are removed. The corners with high score values are kept using the nonmaximal suppression method. There are several intuitive definitions for the score value:

(1) The maximum value of $n$ for which $p$ is still a corner,

(2) The minimum value of $t$ for which $p$ is still a corner,

(3) The sum of the absolute difference between pixels in a contiguous arc and the center pixel.

Definitions (1) and (2) are highly quantifiable measures, and many pixels share these same values. To improve the computational speed, a slightly modified version of (3) is used. The score value is calculated as follows:

(a)          (b)          (c)          (d)          (e)          (f)          (g)
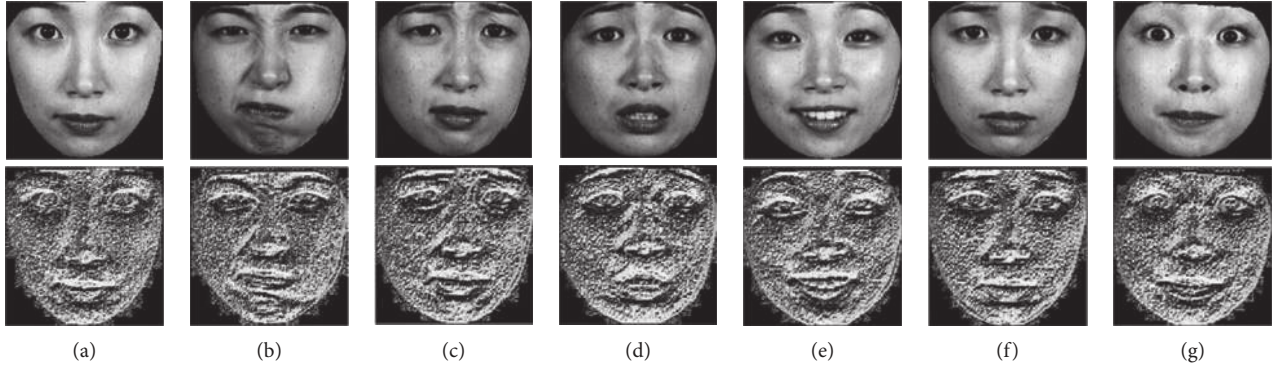
FIGURE 4: Original LBP features of facial expressions. (a) Neutral. (b) Anger. (c) Disgust. (d) Fear. (e) Happy. (f) Sad. (g) Surprise.

$$\text{score} = \max\left( \sum_{x \in S_{\text{bright}}} \left| I_{p \longrightarrow x} - I_p \right| - t, \sum_{x \in S_{\text{dark}}} \left| I_p - I_{p \longrightarrow x} \right| - t \right),$$

(3)

where $I_p$ is the intensity of $p$, $I_{p \longrightarrow x}$ is the intensity of the sixteen pixels around the corner, and $t$ is a threshold.

*(2) BRIEF Descriptor.* The BRIEF descriptor first proposed by M. Calonder is adopted to describe the detected corners. The form of the BRIEF descriptor consists of "1" and "0" values, and the length of the BRIEF descriptor is generally defined as 128 bits, 256 bits, or 512 bits. The following formula clearly shows the definition of the BRIEF descriptor:

$$\lambda\left(p; r_1, c_1, r_2, c_2\right) = \begin{cases} 1: I\left(r_1, c_1\right) < I\left(r_2, c_2\right), \\ 0: I\left(r_1, c_1\right) \geq I\left(r_2, c_2\right), \end{cases}$$

(4)

where $I\left(r_1, c_1\right)$ and $I\left(r_2, c_2\right)$ are the intensities of the pixels at $\left(r_1, c_1\right)$ and $\left(r_2, c_2\right)$, respectively. If $I\left(r_1, c_1\right)$ is less than $I\left(r_2, c_2\right)$, then $\lambda = 1$; otherwise, $\lambda = 0$. The length of $\lambda$ is designated as 256 bits in this paper.

In the description of one point, a subimage with a size of 35 columns × 35 rows (here, the definition in the following situation is the same) is used. Because the BRIEF descriptor is sensitive to noise, the intensity value of each patch pair is calculated using a smoothing filter with a $5 \times 5$ subwindow centered at $\left(r_i, c_i\right)$, where $(i = 1, 2, \Lambda, 512)$. To reduce the impact of the image boundary, the intensity values of the image boundary are removed from the computation; thus, the actual size of the subimage is reduced to $31 \times 31$. Next, $\{(r_1, c_1), (r_2, c_2)\}$ is defined as a patch pair instead of a point pair, and there is a total of 256 patch pairs in the subimage. The locations $(r_i, c_i)$ of the 256 point pairs are determined by a Gaussian distribution, where $(r_i, c_i) \sim i.i.d.$ Additionally, Gaussian $(0, S^2/25)$: $(r_i, c_i)$ values are determined from an isotropic Gaussian distribution, where $S$ is the size of a patch. Details regarding how to determine the locations of the 256 patch pairs can be found in [40]. Figure 5 depicts the ORB features of facial expressions.

*3.2.3. Region-Based ORB.* Traditional ORB is more concentrated on the distribution of the extracted feature points. In traditional ORB, abundant feature points are extracted to achieve a higher accuracy, but the excessively dense feature points are very inconvenient for feature description. To solve this problem, we innovatively employed region division in the traditional ORB. This improved ORB algorithm calculates the number of feature points to be extracted for each region according to the total number of feature points to be extracted and the number of regions to be divided. The steps are described as follows.

*Step 1.* Evenly divide the images into $M \times N$ regions of the same size. $M$ is the row and $N$ is the column of the division. Feature points are randomly distributed in the divided regions, and the regions are sequenced as $\{h_1, h_2, \ldots, h_{M \times N}\}$.

*Step 2.* Set a threshold $T$

$$T = \frac{n}{MN},$$

(5)

where $n$ is the number of feature points.

*Step 3.* The feature points are detected in each region; if the number of features is not less than $T$, we choose $T$ as the feature number. If the number of features is less than $T$, the threshold should be smaller and repeat this step.

*Step 4.* When the number of feature points is bigger than $n$, nonmaximal suppression method is used to select the best feature points.

*Step 5.* All regions are traversed until the number of feature points meets the conditions.

*3.2.4. Feature Fusion Scheme.* Feature normalization is used to increase the recognition rate before feature fusion. In this article, the LBP features are normalized to the range of 0-1, and the following formula is applied to normalize the LBP and ORB features:

$$L = \frac{l}{\max(l)},$$

(6)

where $l$ is the value of a feature.

Feature fusion is usually performed because one type of feature cannot describe every image characteristic. In this
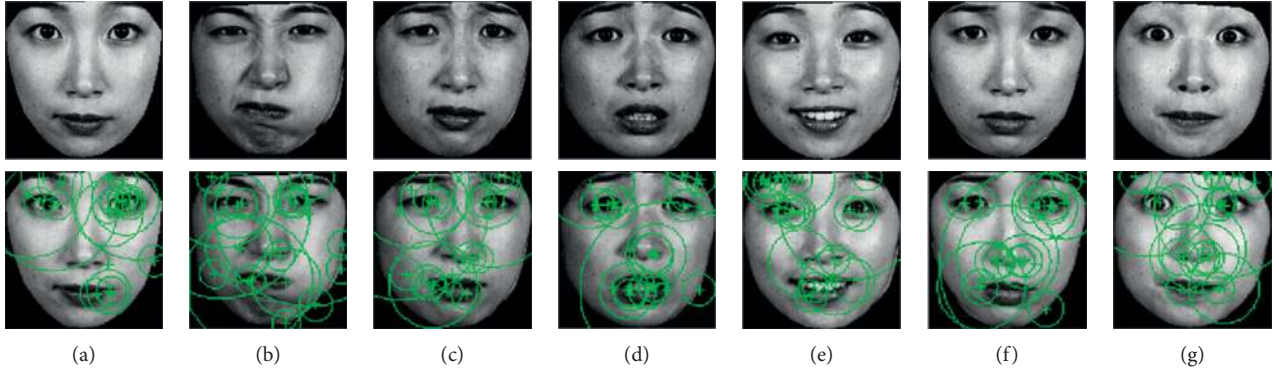
FIGURE 5: ORB features of facial expressions. (a) Neutral. (b) Anger. (c) Disgust. (d) Fear. (e) Happy. (f) Sad. (g) Surprise.

article, the LBP and ORB descriptors are fused by using the Z-score method.

$$\sigma = \sum_{j}^{J} \left( f_j - \mu \right)^2,$$

$$\mu = \frac{\sum_{j}^{J} f_j}{J},$$ (7)

$$\widehat{f}_j = K \frac{\left( x_j - \mu \right)}{\sigma + C},$$

where $f_j$ is an LBP feature or ORB feature and $\widehat{f}_j$ is the feature data after fusion. $K$ is a factor multiplied by $\widehat{f}_j$, and $K$ is 100 in the following experiments.

## 4. Database of Facial Expressions

*4.1. Japanese Female Facial Expressions (JAFFE) Database.* The JAFFE database contains 213 images from ten different Japanese female subjects, and each subject has 3 or 4 examples of seven facial emotions (six basic facial emotions and one neutral emotion). The images are gray with a resolution of $256 \times 256$ [41]. In our experiment, the total 213 images (anger: 30 images; disgust: 29 images; fear: 32 images; happiness: 31 images; neutral: 30 images; sad: 31 images; and surprise: 30 images) are used to evaluate the proposed algorithm. Some images from the JAFFE database are shown in Figure 6.

*4.2. The Extended Cohn-Kanade (CK+) Database.* The CK+ database contains 593 sequences from 123 subjects from 18 to 30 years of age. Overall, 327 sequences are labeled with anger (45), contempt (18), disgust (59), fear (25), happiness (69), sad (28), and surprise (83). Different facial expressions have different numbers in the corresponding sequences, and each expression contains images from neutral to the peak expression. The neutral and three peak frames sampled from each sequence are used for testing while contempt is not considered. In our experiments, we selected 309 sequences which are labeled as one of the six basic facial expressions excluding contempt. We chose the three last frames of each

sequence, so there were 927 images for anger, disgust, fear, happiness, sad, and surprise. In addition, we defined the first frame of each sequence as neutral facial expression (309 images). Taken together, 1236 images from CK+ database were used in our experiments. The images have resolution of $640 \times 490$ or $640 \times 480$ [42, 43].

*4.3. MMI Database.* The MMI facial expression database [44, 45] includes 208 videos of both genders aged from 19 to 62 years. Each sequence labels as one of the six basic emotions and begins with the neutral expression and ends with it, while the expression is in the middle. In our experiments, we approximated the three peak frames from the middle of each sequence. Thus, we obtained 624 color images in total with the resolution $720 \times 576$ pixels. The MMI database has more challenging conditions including illumination, gender, age, ethnicities, and insufficient number of subjects, of which many wear accessories (e.g., glasses, moustache). Figure 7 shows some images from the MMI database.

In the collection of the static image from the sequences, the most commonly used method is to select the first frame (as the neutral expression) and the three-peak frame in each sequence. Also, ten-fold person-independent cross-validation is conducted for the experiments.

## 5. Experiments

This section gives the details of the experiments performed. Two experiments involving the JAFFE, CK+, and MMI databases were conducted to verify the effectiveness of the proposed facial expression recognition method. In the first experiment, the subjects used for training were part of the testing set. In the second experiment, the subjects used for training were not used for testing. That means subject-dependent (SD) and subject-independent (SI), and cross-validation schemes are used in Experiment 1 and Experiment 2, respectively. The subject-independent experimental scheme has been widely investigated in the past years because it is more plausible for the real applications, which need to recognize the facial expressions from new persons. An SVM was the basic classifier used in the experiments, and
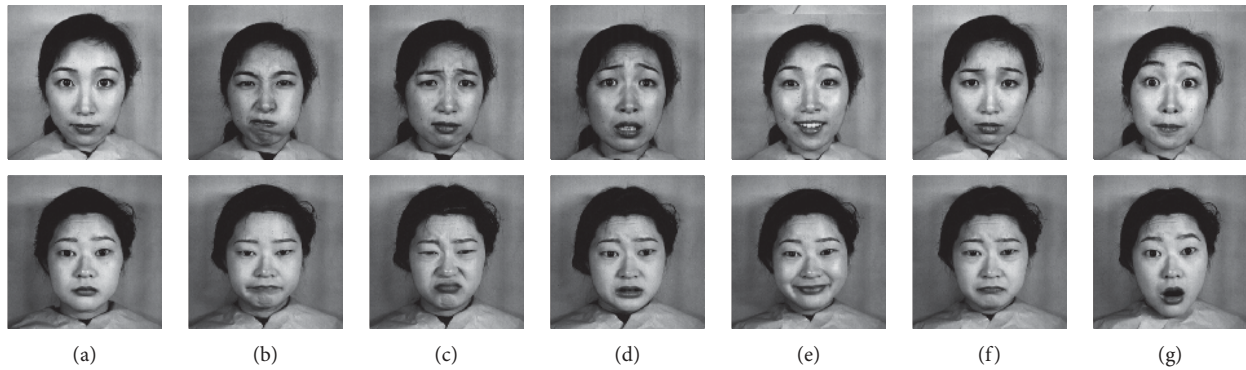
FIGURE 6: Some images from the JAFFE database. (a) Neutral. (b) Anger. (c) Disgust. (d) Fear. (e) Happy. (f) Sad. (g) Surprise.



FIGURE 7: Some images from the MMI database.

10-fold cross-validation procedure was used in all experiments.

### 5.1. Experiments Based on the JAFFE Database.

All the facial region images were resized to $128 \times 128$ pixels, and the classifiers were trained using LibSVM [46]. The parameters were defined through a grid search, and 10-fold cross-validation was used to improve algorithm performance. Each experiment was repeated 10 times, and the results were reported as the average of these 10 replications. Fusion features were used in the proposed algorithm, and the input data were fused with LBP and ORB features.

The accuracy of experiments with different features is shown in Table 1. Results of LBP + ORB were better than the results of LBP/ORB in Experiment 1 (LBP + ORB: 92.4%; LBP: 86.7%; and ORB: 89.2%) and Experiment 2 (LBP + ORB: 88.5%; LBP: 65.4%; and ORB: 73.1%).

A comparison of our results with those of other methods from the literature is shown in Table 2. Recognition rate based on JAFFE dataset suggested the proposed method obtained higher accuracy (88.5%) than recent methods (73.24%, 55.71%, and 87.6%).

### 5.2. Experiments Based on the CK+ Database.

The procedure for the experiments based on the CK+ database was the same as that for the JAFFE database. All facial expression images were extracted from video sequences. The accuracy of experiments with different features is shown in Table 3. Results of LBP + ORB were better than the results of LBP/ORB in Experiment 1 (LBP + ORB: 99.2%; LBP: 97.3%; and ORB: 98.5%) and Experiment 2 (LBP + ORB: 93.2%; LBP: 86.7%; and ORB: 88.4%).

A comparison of our results with those of other methods from the literature is shown in Table 4. Recognition rate based on CK+ dataset suggested the proposed method obtained higher accuracy (93.2%) than one recent method (84.6%), but lower than three recent methods (95.96%, 94.5%, and 95.1%).

### 5.3. Experiments Based on the MMI Database.

The procedure for the experiments based on the MMI database was the same as that for the JAFFE database and CK+ database. All facial expression images were extracted from video sequences. The accuracy of experiments with different features is shown in Table 5. Results of LBP + ORB were better than the results of LBP/ORB in Experiment 1 (LBP + ORB: 84.2%; LBP: 73.3%; and ORB: 78.5%) and Experiment 2 (LBP + ORB: 79.8%; LBP: 69.9%; and ORB: 73.3%).

A comparison of our results with those of other methods from the literature is shown in Table 6. The proposed method obtained higher accuracy (79.8%) than all four recent methods (73.57%, 70.63%, 76.3%, and 58.98%). Notably, the proposed method performed better than the other methods based on MMI dataset.

### 5.4. Discussion.

Based on the results of our experiments, it is obvious that our proposed method can deal with the facial expression recognition efficiently. In this paper, we proposed

TABLE 1: The accuracy of experiments with different features.

| Experiment 1 | | Experiment 2 | |
| --- | --- | --- | --- |
| Features | Accuracy (%) | Features | Accuracy (%) |
| LBP | 86.7 | LBP | 65.4 |
| ORB | 89.2 | ORB | 73.1 |
| LBP + ORB | 92.4 | LBP + ORB | 88.5 |

TABLE 2: Recognition rate based on JAFFE.

| Study | | Bougourzi et al. [26] | Mandal et al. [17] | Tong and Chen [47] | Proposed algorithm |
| --- | --- | --- | --- | --- | --- |
| Method | | PCA-fusion | RADAP | LDDSCP | LBP + ORB |
| Validation setting | | 10-fold | 10-fold | 10-fold | 10-fold |
| Accuracy (%) | SD | — | 90.5 | — | 92.4 |
| | SI | 73.24 | 55.71 | 87.6 | 88.5 |

TABLE 3: The accuracy of experiments with different features.

| Experiment 1 | | Experiment 2 | |
| --- | --- | --- | --- |
| Features | Accuracy (%) | Features | Accuracy (%) |
| LBP | 97.3 | LBP | 86.7 |
| ORB | 98.5 | ORB | 88.4 |
| LBP + ORB | 99.2 | LBP + ORB | 93.2 |

TABLE 4: Recognition rate based on CK+.

| Study | | Bougourzi et al. [26] | Makhmudkhujaev et al. [48] | Ji et al. [49] | Mandal et al. [17] | Proposed algorithm |
| --- | --- | --- | --- | --- | --- | --- |
| Method | | PCA-fusion | LDDP | ICID | RADAP | LBP + ORB |
| Validation setting | | 10-fold | — | — | 10-fold | 10-fold |
| Accuracy (%) | SD | — | — | — | 95.4 | 99.2 |
| | SI | 95.96 | 94.5 | 95.1 | 84.6 | 93.2 |

TABLE 5: The accuracy of experiments with different features.

| Experiment 1 | | Experiment 2 | |
| --- | --- | --- | --- |
| Features | Accuracy (%) | Features | Accuracy (%) |
| LBP | 73.3 | LBP | 69.9 |
| ORB | 78.5 | ORB | 73.3 |
| LBP + ORB | 84.2 | LBP + ORB | 79.8 |

a new framework of emotion recognition system from static images by using feature-based approach. We combined LBP and improved ORB features extracted by descriptors possessing different properties. Recently, numerous studies have reported that DNN-based approach performed well in facial expression recognition. However, DNN-based approach needs excessive hardware specifications requirement. Considering low hardware specifications used in real-life condition, to gain better results without DNNs, in this paper, we proposed an algorithm with the combination of the improved ORB features and LBP features extracted from facial expression.

Compared with other methods, LBP is more effective to extract facial expression. However, the speed is also important for effective facial expression recognition. Instead of using LBP only, ORB is also employed for facial expression recognition in our system since the ORB has been shown to have high computing speed. Notably, compared with LBP/ORB, experimental results based on JAFFE, CK+, and MMI databases showed that our method can enhance the accuracy of facial expression recognition, in other words, can improve the discriminative ability. Meanwhile, based on JAFFE, CK+, and MMI databases, we found that our method was in general better than recent studies on facial expression recognition. Our method improved the accuracy of facial expression recognition based on JAFFE database; results of recognition rate based on CK+ database indicated that our method was not as good as the previous methods but outperformed [17] which used similar experiment. Recognition rate based on MMI database suggested the proposed method achieved superior recognition accuracy than all four state-of-the-art approaches. Considering the characteristics of the facial expression and the requirements of real time, we combined LBP and improved OBR as the features of facial expression to

TABLE 6: Recognition rate based on MMI.

| Study | Bougourzi et al. [26] | Makhmudkhujaev et al. [48] | Ji et al. [49] | Mandal et al. [17] | Proposed algorithm |
|---|---|---|---|---|---|
| Method | PCA-fusion | LDDP | ICID | RADAP | LBP + ORB |
| Validation setting | 10-fold | — | — | 10-fold | 10-fold |
| Accuracy (%) SD | — | — | — | 86.6 | 84.2 |
| Accuracy (%) SI | 73.57 | 70.63 | 76.3 | 58.98 | 79.8 |

improve recognition performance, but there are still much more challenges. In the future research, we will focus on how to improve the accuracy and increase the speed at the same time.

## 6. Conclusions

This paper proposed a framework for facial expression recognition with fused features. LBP and improved ORB descriptors were used for feature extraction, and SVM classification was performed for facial expression recognition. The experimental results showed that the proposed framework performed better than some widely used methods based on the JAFFE database, CK+ database, and MMI database.

This article mainly focused on facial expressions from static images. In our future work, facial expressions from video sequences will be further considered, as will calculation times for embedded systems.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon reasonable request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

## Acknowledgments

## References

[1] A. Mehrabian, "Communication without words," *Psychology Today*, vol. 2, pp. 53–56, 1968.

[2] K. Kaulard, D. W. Cunningham, H. H. Bülthoff, and C. Wallraven, "The MPI facial expression database-avalidated database of emotional and conversational facial expression," *PLoS One*, vol. 7, Article ID e32321, 2012.

[3] F. Dornaika and B. Raducanu, "Efficient facial expression recognition for human robot interaction," in *Proceedings of the 9th International Work-Conference onArtificial Networks on Computational and Ambient Intelligence*, pp. 700–708, San Sebastián, Spain, June 2007.

[4] C. Bartneck and M. J. Lyons, "HCI and the face: towards an art of the soluble," in *Proceedings of the International Comference on Human-Computer Interaction: Interaction Design and Usability*, pp. 20–29, Beijing, China, July 2017.

[5] S. Hickson, N. Dufour, A. Sud, V. Kwatra, and I. A. Essa, "Eyemotion: classifying facial expressions in VR using eye-tracking cameras," 2017, https://arxiv.org/abs/1707.07204.

[6] C.-H. Chen, I.-J. Lee, and L.-Y. Lin, "Augmented reality-based self-facial modeling to promote the emotional expression and social skills of adolescents with autism spectrum disorders," *Research in Developmental Disabilities*, vol. 36, pp. 396–403, 2015.

[7] M. A. Assari and M. Rahmati, "Driver drowsiness detection using face expression recognition," in *Proceedings of the IEEE International Conference on Signal and Image Processing Applications*, pp. 337–341, Kuala Lumpur, Malaysia, November 2011.

[8] C. Zhang, W. Li, P. Ogunbona, and F. Safaei, "A real-time facial expression recognition system for online games," *International Journal of Computer Games Technology*, vol. 2008, Article ID 542918, 7 pages, 2008.

[9] A. Mouráo and J. Magalháes, "Competitive affective gaming: winning with a smile," in *Proceedings of the ACM International Conference on Multimedia*, pp. 83–92, Barcelona, Spain, October 2013.

[10] K. Zhao, W. S. Chu, and H. Zhang, "Deep region and multi-label learning for facial action unit detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3391–3339, Las Vegas, NV, USA, June 2016.

[11] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

[12] S. E. Kahou, V. Michalski, and K. Konda, "Recurrent neural networks for emotion recognition in video," in *Proceedings of the ACM on International Conference Multimodal Interaction*, pp. 467–474, Seattle, WA, USA, November 2015.

[13] R. Walecki, O. Rudovic, V. Pavlovic, and B. Schuller, "Deep structured learning for facial expression intensity estimation," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 259, pp. 143–154, Honolulu, HI, USA, July 2017.

[14] Y. Tang, X. M. Zhang, and H. Wang, "Geometric-convolutional feature fusion based on learning propagation for facial expression recognition," *IEEE Access*, vol. 6, pp. 42532–42540, 2018.

[15] L. Liang, C. Lang, Y. Li, S. Feng, and J. Zhao, "Fine-grained facial expression recognition in the wild," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 482–494, 2021.

[16] F. Zhang, T. Zhang, Q. Mao, and C. Xu, "Geometry guided pose-invariant facial expression recognition," *IEEE Transactions on Image Processing*, vol. 29, pp. 4445–4460, 2020.

[17] M. Mandal, M. Verma, S. Mathur, S. K. Vipparthi, S. Murala, and D. Kranthi Kumar, "Regional adaptive affinitive patterns (RADAP) with logical operators for facial expression recognition," *IET Image Processing*, vol. 13, no. 5, pp. 850–861, 2019.

[18] H.-H. Tsai and Y.-C. Chang, "Facial expression recognition using a combination of multiple facial features and support vector machine," *Soft Computing*, vol. 22, no. 13, pp. 4389–4405, 2018.

[19] J. Chen, Z. Chen, Z. Chi, and H. Fu, "Facial expression recognition in video with multiple feature fusion," *IEEE Transactions on Affective Computing*, vol. 9, no. 1, pp. 38–50, 2018.

[20] B. Ryu, A. R. Rivera, J. Kim, and O. Chae, "Local directional ternary pattern for facial expression recognition," *IEEE Transactions on Image Processing*, vol. 26, no. 12, pp. 6006–6018, 2017.

[21] Y. Liu, Y. Li, X. Ma, and R. Song, "Facial expression recognition with fusion features extracted from salient facial areas," *Sensors*, vol. 17, no. 4, p. 712, 2017.

[22] D. Ghimire, S. Jeong, J. Lee, and S. H. Park, "Facial expression recognition based on local region specific features and support vector machines," *Multimedia Tools and Applications*, vol. 76, no. 6, pp. 7803–7821, 2017.

[23] S. Nigam, R. Singh, and A. K. Misra, "Efficient facial expression recognition using histogram of oriented gradients in wavelet domain," *Multimedia Tools and Applications*, vol. 77, no. 21, pp. 28725–28747, 2018.

[24] S. K. A. Kamarol, J. Parkkinen, M. H. Jaward, and R. Parthiban, "Spatiotemporal feature extraction for facial expression recognition," *IET Image Processing*, vol. 10, no. 7, pp. 534–541, 2016.

[25] C. Turan and K.-M. Lam, "Histogram-based local descriptors for facial expression recognition (FER): a comprehensive study," *Journal of Visual Communication and Image Representation*, vol. 55, pp. 331–341, 2018.

[26] F. Bougourzi, K. Mokrani, Y. Ruichek, F. Dornaika, A. Ouafi, and A. Taleb-Ahmed, "Fusion of transformed shallow features for facial expression recognition," *IET Image Processing*, vol. 13, no. 9, pp. 1479–1489, 2019.

[27] H. K. Meena, K. K. Sharma, and S. D. Joshi, "Effective curvelet-based facial expression recognition using graph signal processing," *Signal, Image and Video Processing*, vol. 14, no. 2, pp. 241–247, 2020.

[28] A. M. Ashir, A. Eleyan, and B. Akdemir, "Facial expression recognition with dynamic cascaded classifier," *Neural Computing and Applications*, vol. 32, no. 10, pp. 6295–6309, 2020.

[29] J. Dong, H. Zheng, and L. Lian, "Dynamic facial expression recognition based on convolutional neural networks with dense connections," in *Proceedings of the International Conference on Pattern Recognition*, pp. 3433–3438, Stockholm, Sweden, 2018.

[30] W. Xie, X. Jia, L. Shen, and M. Yang, "Sparse deep feature learning for facial expression recognition," *Pattern Recognition*, vol. 96, Article ID 106966, 2019.

[31] M. Yu, H. Zheng, Z. Peng, J. Dong, and H. Du, "Facial expression recognition based on a multi-task global-local network," *Pattern Recognition Letters*, vol. 131, pp. 166–171, 2020.

[32] G. Wen, T. Chang, H. Li, and L. Jiang, "Dynamic objectives learning for facial expression recognition," *IEEE Transactions on Multimedia*, vol. 22, no. 11, pp. 2914–2925, 2020.

[33] K. Wang, X. Peng, J. Yang, S. Lu, and Y. Qiao, "Suppressing uncertainties for large-scale facial expression recognition," 2020, https://arxiv.org/abs/2002.10392.

[34] D. E. King, "Dlib-ml: a machine learning toolkit," *Journal of Machine Learning Research*, vol. 10, pp. 1755–1758, 2009.

[35] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1867–1874, Columbus, OH, USA, June 2014.

[36] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[37] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[38] M. Topi, O. Timo, P. Matti, and S. Maricor, "Robust texture classification by subsets of local binary pattern," *Pattern Recognition*, vol. 3, no. 7, pp. 935–338, 2000.

[39] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: a comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803–816, 2009.

[40] E. Rublee, V. Rabaud, and K. Konolige, "ORB: an efficient alternative to SIFT or SURF," in *Proceedings of the International Conference of Computer Vision*, pp. 2564–2571, Barcelona, Spain, November 2011.

[41] M. Lyons, S. Akemastu, and M. Kamachi, "Coding facial expressions with gabor wavelets," in *Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 200–205, Nara, Japan, April 1998.

[42] T. Kanade, J. F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 46–53, Grenoble, France, March 2000.

[43] P. Lucey, F. Cohn, and T. Kanade, "The Extended Cohn-Kanade Dataset (CK+): a complete expression dataset for action unit and emotion-specified expression," in *Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB 2010)*, pp. 94–101, San Francisco, CA, USA, 2010.

[44] M. Pantic, M. Valstar, and R. Rademaker, "Web-based database for facial expression analysis," in *Proceedings of the IEEE International Conference on Multimedia & Expo (ICME)*, Amsterdam, Netherlands, 2005.

[45] M. Valstar and M. Pantic, "Induced disgust, happiness and surprise: an addition to the mmi facial expression database," in *Proceedings of the Third International Workshop on EMOTION (Satellite of LREC): Corpora for Research on Emotion and Affect*, Valletta, Malta, May 2010.

[46] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 27, pp. 1–27, 2011.

[47] Y. Tong and R. Chen, "Local dominant directional symmetrical coding patterns for facial expression recognition," *Computational Intelligence and Neuroscience*, vol. 2019, Article ID 3587036, 13 pages, 2019.

[48] F. Makhmudkhujaev, M. Abdullah-Al-Wadud, M. T. B. Iqbal, B. Ryu, and O. Chae, "Facial expression recognition with local prominent directional pattern," *Signal Processing: Image Communication*, vol. 74, pp. 1–12, 2019.

[49] Y. Ji, Y. Hu, Y. Yang, F. Shen, and H. T. Shen, "Cross-domain facial expression recognition via an intra-category common feature and inter-category distinction feature fusion network," *Neurocomputing*, vol. 333, pp. 231–239, 2019.