IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 11, 2019, accepted July 15, 2019, date of publication July 18, 2019, date of current version August 14, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2929753

Automatically Query Active Features Based on Pixel-Level for Facial Expression Recognition

ZHE SUN, ZHENGPING HU, AND MENGYAO ZHAO
Department of Information Science and Engineering, Yanshan University, Qinhuangdao 066000, China

Corresponding author: Zhengping Hu (hzp_ysu@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61071199 and Grant 61771420, in part by the China Postdoctoral Science Foundation under Grant 2018M641674, and in part by the Doctoral Foundation of Yanshan University under Grant 8190071.

ABSTRACT Feature extraction-based subspace learning methods normally learn a projection that can convert the high-dimensional data to the low-dimensional representation. However, they may not be suitable for better classification since features obtained by these methods ignore discriminability of the data-pixel itself. Given this, we propose a novel approach that automatically queries active features combing sparse representation classification for the facial expression recognition. The proposed approach aims to automatically query discriminative features from raw pixels, thereby fully considering the underlying characteristics existed in the source data. Especially, the proposed approach based on pixel-level adaptively selects the most active and discriminative feature for representation and classification. The intraclass low-rank decomposition and principal feature analysis are simultaneously used to guarantee that the extracted features can capture the most active energy of the raw data, and thus, the proposed approach can be also applied for other feature extraction and selection tasks. We conduct comprehensive experiments on four public datasets, and the results show superior performance than some state-of-the-art methods.

INDEX TERMS Facial expression recognition, automatically query active features, pixel-level, intraclass low-rank matrix, principal feature analysis.

I. INTRODUCTION

Feature extraction and selection play important roles in pattern recognition and machine learning and have attracted lots of attention in recent years [1], [2]. Especially for facial expression recognition (FER), the raw data usually contain redundant information and have high dimensions. Therefore, how to extract and select the most active features for FER is a challenging task [3].

Various feature extraction methods have been devised to improve efficiency and the ability of classification [4]–[8]. Some work aims to select the most important or active features from raw pixels to efficiently represent the original data [7], and some try to learn a mapping matrix that can transform the high dimensional data to low dimensional subspace [8]. Due to the difficulty of illumination to feature extraction, Ahmad *et al.* [9] proposed an ICA-based method for separating the illumination and reflectance components of a single illuminated image and their method also can be used as pre-processing methods for other recognition problems.

The associate editor coordinating the review of this manuscript and approving it for publication was Habib Ullah.

Besides, according to the physiological research methods for the study of the human brain, Ullah *et al.* [10] proposed ensemble learning algorithm using an electroencephalography (EEG) channel for internal emotion recognition, which is effective in improving computational efficiency and classification accuracy. Additionally, a lot of deep learning based supervised feature extraction methods have been presented and attracted wide attention [11]–[15]. For instance, Shao and Qian [12] proposed three novel convolutional neural network models with different architectures to address the problem caused by the complex architecture and over-fitting. From the view of 3D geometry, Liu *et al.* [14] designed an action unit synthesis framework for deep learning-based AU intensity estimation and extensive experiments demonstrated the effectiveness of their method. Besides, Fernandez *et al.* [15] proposed a FER with attention network architecture to generate synthetic data that improved the system classification performance. The deep learning-based algorithms above achieved good performance, while they need large-scale dataset and had a high requirement to train the feature extraction model.

Compared to the deep feature extraction methods, conventional methods showed more advantages in the tasks with

small size data and thus we mainly research on the conventional feature extraction methods in this paper. Some conventional methods performed superiority in FER application, while some of them extracted redundant features for classification. To solve this issue, Sun *et al.* [16] selected a certain percentage of features to reduce useless information for better classification, while it still needed manually selection of parameters. Given the factors above, in this paper, we propose an automatically query active features combining sparse representation classification approach to obtain the discriminative feature subspace. The proposed approach simultaneously preserves the global and local features and adaptively gives up some redundant information. In addition, the proposed approach fully explores the basic and representative features hidden in pixel intensity, and automatically selects the active features for classification. In brief, the proposed approach has the following contributions:

(1) Our approach is a simple yet effective approach to extract and select the active features for FER classification.

(2) The features achieved by our approach capture the main energy and thus hold the active information. Besides, the proposed approach can be regarded as the process of raising dimension and then lowering dimension, which guarantees the minimum loss of features and considers the feasibility of computing time. Experimental results also verify that our approach is superiority to some state-of-the-art methods.

(3) Compared to some conventional feature extraction methods, our approach is more easy to implement and can be applied for other classification tasks.

The remainder of the paper is organized as follows. Section II introduces the related works. In Section III, we present the proposed approach in detail. In Section IV, the experimental performance of the proposed approach is evaluated by using several public datasets: the Japanese Female Facial Expression (JAFFE) dataset [17], Karolinska Directed Emotional Faces (KDEF) [18], the Extended Cohn-Kanade (CK+) dataset [19], and the CMU Multi-PIE face database [20]. Section IV also provides an analysis of the proposed approach. Section V concludes the paper.

II. RELATED WORKS

In this section, we briefly introduce some related algorithms. For convenience, we first roughly divide the existed feature extraction methods into two categories: unsupervised learning methods [21]–[24] and supervised learning methods [25]–[27].

A. UNSUPERVISED LEARNING ALGORITHMS

Representatives based on unsupervised learning is the principal component analysis (PCA) [22] algorithm that tries to find the principal energy of raw data. From this point, PCA based on method has been widely used in feature extraction [28]. Considering the deep network structure can capture the abstract features of data, Chen *et al.* [29] proposed the principal component analysis network (PCANet) model that extended PCA to deep subspace learning and further preserved the high-level features during the unsupervised feature learning process. Though PCANet showed the excellent feature extraction ability, it ignores the non-linear relationship and the high dimensionality existed in features. Given this, Sun *et al.* [30] attempted to project the abstract features into kernel space to fully consider the use of features with non-linear matrics, and their experiments showed promising results. Besides, Sun's group [31] also presented an extended dictionary representation dictionary with deep subspace features based on PCANet method (EDR-PCANet) to remit the problem limited by high-dimension data. Although the methods above succeed in feature extraction, they do not take the discriminative features into consideration since they do not use the label information of data.

B. SUPERVISED LEARNING ALGORITHMS

Linear discriminant analysis (LDA) can be regarded as one of the classical supervised subspace learning, since it considered the label information when trying to project the raw data into new subspace for better classification [25]. Traditional LDA attempted to find an optimal projection matrix making the ratio of between-class distance larger to the within-class distance largest, so as to improve the classification performance [32]. However, LDA had the limitation in small size sample and feature extraction of the matrix data. To address this issue, two-dimension linear discriminant analysis (2DLDA) [27] making full use of structure information had been proposed to address the problem caused by above. Besides, Sun *et al.* [33] presented a discriminative feature learning method based on vertical 2DLDA that fully considered the matrix format of data and time complexity. Besides, some subspace learning methods are also proposed to perform FER classification tasks [34]–[37]. For example, paper [35] adopted a dictionary learning feature space via sparse representation classification (DLFS) for FER and achieved satisfying performance.

III. THE PROPOSED APPROACH

In this section, we present the details of the proposed approach. **Fig. 1** gives an illustration of the proposed approach. As is seen in this figure, the frame consists of three sequential steps: (1) generation of intraclass low-rank dictionary (2) extraction of active features based pixel-level, and (3) classification of sparse representation (SRC). Each step will be explained in detail in the following subsections.

A. GENERATION OF ICLR DICTIONARY

Let $D = [D_1, \cdots, D_i, \cdots, D_C] \in R^{mxN}$ be the training dictionary with C expression classes, where D_i is subset of the *ith* class. N and m mean the number and the dimension of training subset, respectively. Despired by the precious work [38] that used low-rank (LR) decomposition method, the training dictionary D can be decomposed into $N + E$, where N is the LR common dictionary and E is the sparse error dictionary. Towards this end, LR minimizes the rank of dictionary N while decreasing the ℓ_0-norm of E. As a

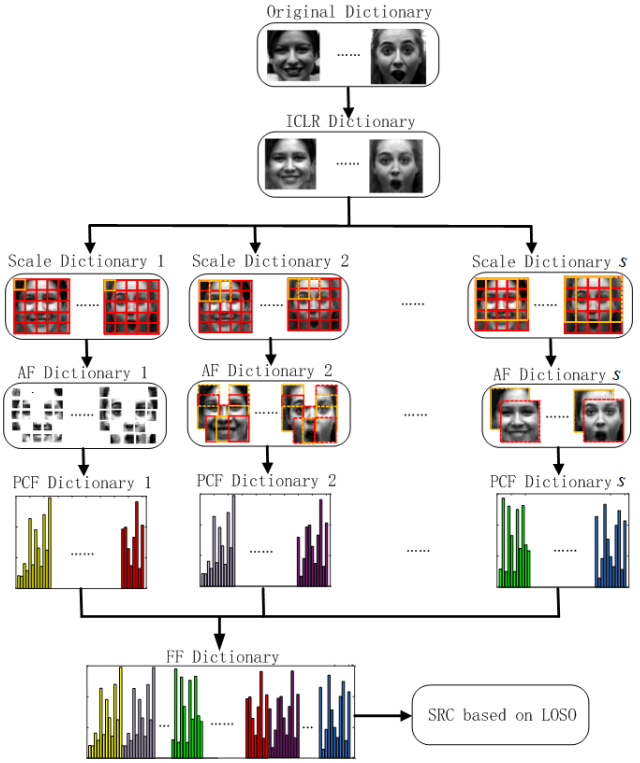

FIGURE 1. The illustration of the proposed approach. It roughly includes the following steps: Generating intraclass low-rank (ICLR) dictionary, extracting active features to form the feature fusion (FF) dictionary, and the final classification based sparse representation (SRC).

consequence, we need to solve the following minimization problem:

$$\min_{N,E} \ rank\,(N) + \lambda\|E\|_0 \quad s.t.\ D = N + E. \tag{1}$$

However, Eq. (1) is a NP-hard problem. Thus, ℓ_0-norm problem in Eq. (1) is converted to the ℓ_1-norm problem according to the compressed sensing:

$$\min_{N,E} \ rank\,(N) + \lambda\|E\|_1 \quad s.t.\ D = N + E. \tag{2}$$

To further solve the optimization problem in Eq. (2), Augmented Lagrange Multipliers (ALM) [39] was exploited owning to its computational efficiency.

It is worth noting that we do not apply the LR decomposition method to training dictionary directly, but apply to training subset of each intraclass since the training subset of each intraclass share the similar class information. Thus, the proposed intraclass low-rank (ICLR) decomposition approach can capture the similar intraclass structure.

Based on the Eq. (2), our ICLR decomposition approach can be represented as Eq. (3) and the solution can be iteratively addressed by the following minimization problem:

$$\min_{N_i,E_i} \ \|B_i\|_* + \lambda\|E_i\|_1 \quad s.t.\ D_i = B_i + E_i. \tag{3}$$

where D_i is the *ith* intraclass training subset, B_i is the *ith* ICLR dictionary, and E_i is the sparse error dictionary. The minimization problem in Eq. (3) can be solved by the structurally

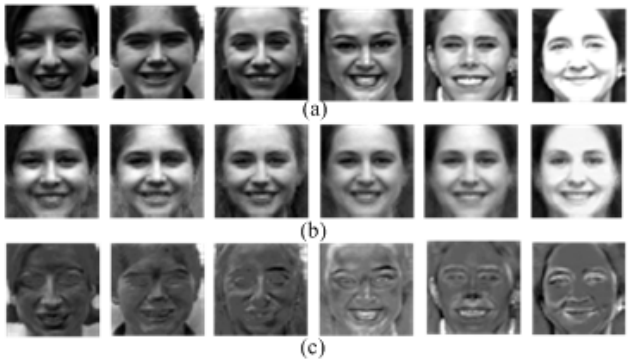

FIGURE 2. Six "Happy" expression images from CK+ dataset based on the proposed ICLR decomposing approach. (a) Six original facial images from training subset, (b) six ICLR images corresponding to (a), and (c) six sparse error images corresponding to (a).

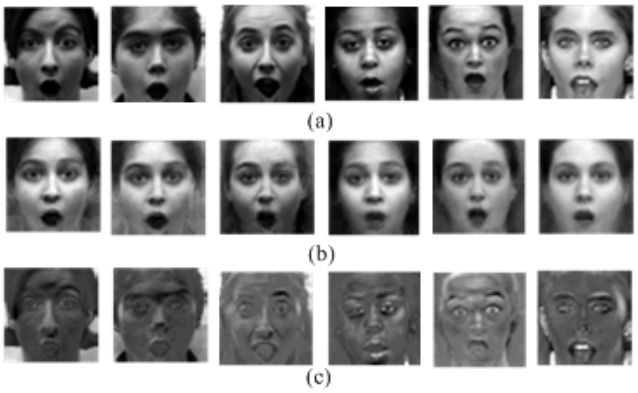

FIGURE 3. Six "Surprise" expression images from CK+ dataset based on the proposed ICLR decomposing approach. (a) Six original facial images from training subset, (b) six ICLR images corresponding to (a), and (c) six sparse error images corresponding to (a).

incoherent LR matrix decomposition algorithm [40] since it enforces the sparse error bases of different classes to be as independent as possible.

To make ICLR decomposing process more clear, we use two expressions ("Happy" and "Surprise" expression, respectively) from CK+ dataset as examples that are shown in **Fig. 2** and **Fig. 3**. From these two figures, we can see that the ICLR approach does well in decomposing the training subset D_i (the 1st row of **Fig. 2** and **Fig. 3**) into an ICLR subset B_i (the 2nd row of **Fig. 2** and **Fig. 3**) and sparse error subset E_i (the 3rd row of **Fig. 2** and **Fig. 3**). Obviously, the ICLR dictionary indeed has a better discriminative ability than the original training dictionary in expression features. Finally all ICLR subsets B_i form the ICLR dictionary $B = [B_1 \cdots, B_i, \cdots, B_C] = \left[b_1, \cdots, b_j, \cdots, b_N\right] \in R^{mxN}$ $(j = 1, \cdots, N)$, where b_j is the *jth* ICLR sample of B.

Although the ICLR decomposion approach succeeded in projecting the original space to the expression subspace, it inevitably ignores the local information hidden in the ICLR subspace that attributes the importance differently for representation and classification. Thus, we propose to dig the

locally discriminative features on the basis of ICLR subspace and the details will be given in the following subsections.

B. ACTIVE FEATURES EXTRACTION BASED ON PIXEL-LEVEL

Suppose that each ICLR sample b_j is divided into $k \times k$ patches and we set the scale as s ($s = 1, 2 \cdots k-1$). It's worth mention that we define $s|\bullet$ be the variable under the condition of scale s. We traverse all patches under different condition of scale s and obtain the scale dictionary $s|\varphi_p$ that can be expressed as:

$$s|\varphi_p = \left[s|f_{11}, \cdots s|f_{j,p}, \cdots, s|f_{N,n}\right] \quad (p = 1, \cdots, n). \tag{4}$$

where $s|f_{j,p}$ is the vector that is obtained by cascading of pixels corresponding to the *pth* patch in b_j under the condition of scale s, and n is the total number of patches that can be computed by

$$n = (k - s + 1)^2. \tag{5}$$

Thus, all $s|\varphi_p$ under different scales can from the scale dictionary set F:

$$F = \left\{\begin{array}{c} 1|\varphi_p \\ \cdots \\ s|\varphi_p \end{array}\right\} = \left\{\begin{array}{ccc} 1|f_{1,1}, & \cdots, & 1|f_{N,n} \\ \cdots & \cdots & \cdots \\ s|f_{1,1}, & \cdots, & s|f_{N,n} \end{array}\right\}. \tag{6}$$

Considering the fact that not all patches contribute equally for the representation, we propose to use sparse representation based classification (SRC) to select the active patches since SRC shows superior performing recognition tasks [41]. Suppose that *Function* $(\bullet)$ be the function of SRC. For a random vector of sub-patch $s|f_{j,p}$, we can obtain n accuracies corresponding to different scale dictionary based on leave-one-subject-out (LOSO) cross validation:

$$acc_p\left(s|f_{j,p}\right) = Function\left(s|\varphi_p\right). \tag{7}$$

where $acc_p\left(s|f_{j,p}\right)$ is the *pth* accuracy corresponding to the specific scale s. Subsequently, we use the threshold θ_s to effectively select the active patches and θ_s can be obtained by computing the average value of all accuracies:

$$\theta_s = \frac{1}{n}\sum_{p=1}^{n} acc_p\left(s|f_{j,p}\right). \tag{8}$$

When $acc_p\left(s|f_{j,p}\right)$ is greater than the threshold θ_s, we regard the corresponding $s|f_{j,p}$ as active feature (AF); otherwise, we regard $s|f_{j,p}$ as useless feature. Based on the criterion above, we can select all AFs under each scale s and then use these AFs to form the AF set $s|\widehat{\varphi}_{\widehat{p}}$ of each ICLR sample that can be expressed as follows:

$$s|\widehat{\varphi}_{\widehat{p}} = \left[s|\widehat{f}_{j,a}, \cdots, s|\widehat{f}_{j,b}\right]. \tag{9}$$

Here $s|\widehat{\varphi}_{\widehat{p}}$ is a subset of the scale dictionary $s|\varphi_p$, that is, $\left[s|\widehat{f}_{j,a}, \cdots, s|\widehat{f}_{j,b}\right]$ should be contained in $\left[s|f_{11}, \cdots s|f_{j,p}, \cdots, s|f_{N,n}\right]$ and AFs' patch-label set $[a, \cdots, b]$ should be contained in the total patch-label set $[1, \cdots, n]$ for each ICLR sample. Then we cascade the AF set for each scale as:

$$s|v_j = \left[s|\widehat{f}_{1,a}; \cdots; s|\widehat{f}_{N,b}, \cdots, s|\widehat{f}_{j,a}; \cdots; s|\widehat{f}_{j,b}, \cdots, s|\widehat{f}_{N,a}; \cdots; s|\widehat{f}_{N,b}\right] \tag{10}$$

where $s|v_j$ means the proposed descriptor for each ICLR sample. Subsequently all $s|v_j$ $(j = 1, 2, \cdots, N)$ further form the corresponding AF dictionary $s|V$ that can be expressed as:

$$s|V = \left[s|v_1, \cdots, s|v_j, \cdots, s|v_N\right]. \tag{11}$$

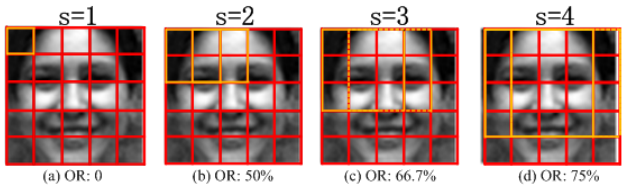

FIGURE 4. The schematic diagram and overlapping rate (OR) using a "happy" expression sample under different scale conditions. (a) $s = 1$, (b) $s = 2$, (c) $s = 3$, and (d) $s = 4$.

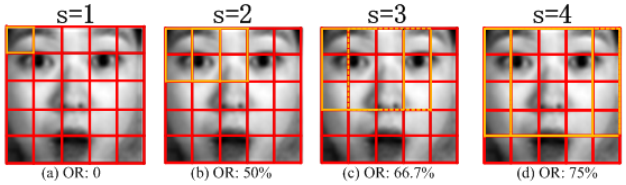

FIGURE 5. The schematic diagram and overlapping rate (OR) using a "surprise" sample under different scale values. (a) $s = 1$, (b) $s = 2$, (c) $s = 3$, and (d) $s = 4$.

Fig. 4 and **Fig. 5** give the schematic diagrams under different scales by using two expression samples. Among these two figures, the red box indicates all the divided patches and the yellow box indicates the selected scales. From **Fig. 4**, we observe that the scale s represents different region information variously and there also exists redundancy to some extent. For instance, when s is set to 3, the overlapping rate reaches up to 66.7% and when s is set to 4, the overlapping rate even reaches up to 75%. The same conclusion can be drawn from **Fig. 5**.

Although AF dictionary considers the contribution of local features existed in different patches, it causes the dimensionality of data increasing since it contains redundant information by stacking features under different scales. Given this, this paper proposes to further extract principal component features from the AF dictionary and Part *C* will describe the details.

C. PRINCIPAL COMPONENT FEATURE EXTRACTION

In this subsection, we use PCA method [22] to extract the principal component feature (PCF) of AF dictionary

since PCA succeeds in capturing the principal data structure and reducing the data dimension simultaneously. Suppose that AF dictionary also can be represented as $s|\,V = [s|\,V_1, \cdots, s|\,V_i, \cdots, s|\,V_C] \in R^{m_a \times N}$ with C classes, where $s|\,V_i$ is the subset of the *ith* class and m_a is the dimension of AF. For a given subset $s|\,V_i = \left[s|\,v_{i,1}, \cdots, s|\,v_{i,j}, \cdots, s|\,v_{i,n_i}\right] \in R^{m_a \times n_i}$, where $s|\,v_{i,j}$ is the *jth* sample of *ith* class in $s|\,V_i$ and n_i is the number of the *ith* class. First all samples in $s|\,v_{i,j}$ are centralized as:

$$s|\,\alpha_{i,j} = s|\,v_{i,j} - \frac{1}{n_i}\sum_{j=1}^{n_i} s|\,v_{i,j}. \tag{12}$$

Here $s|\,\alpha_{i,j}$ is the average value for $s|\,v_{i,j}$. Let the average matrix of each class be $s|\,A_i$ that can be represented as:

$$s|\,A_i = \left[s|\,\alpha_{i,1}, \cdots, s|\,\alpha_{i,j}, \cdots, s|\,\alpha_{i,n_i}\right] \quad (i = 1, 2, \cdots, C). \tag{13}$$

Then the covariance matrix corresponding to $s|\,V_i$ is represented as $(s|\,A_i)\,(s|\,A_i)^T$ and subsequently we do the eigenvalue decomposition:

$$s|\,\Lambda_i = s|\,U_i{}^T \left((s|\,A_i)\,(s|\,A_i)^T\right) s|\,U_i. \tag{14}$$

where $s|\,\Lambda_i$ is the diagonal matrix composed of eigenvalues and $s|\,U_i$ is the orthogonal matrix composed of eigenvectors. Define the eigenvalues be $s|\,\lambda_e\,(e = 1, 2, \cdots, m_a)$, where $s|\,\lambda_{i,1} \geq s|\,\lambda_{i,2} \geq \cdots \geq s|\,\lambda_{i,m_a}$ and define the eigenvector be $s|\,u_e\,(e = 1, 2, \cdots, m_a)$. The $s|\,\Lambda_i$ and $s|\,U_i$ can be represented as follows:

$$s|\,\Lambda_i = \begin{bmatrix} s|\,\lambda_1 & & \\ & \ddots & \\ & & s|\,\lambda_{m_a} \end{bmatrix}. \tag{15}$$

$$s|\,U_i = \left[s|\,u_1 \quad s|\,u_2 \quad \cdots \quad s|\,u_{m_a}\right]. \tag{16}$$

Then we select the matrix composed of the eigenvectors corresponding to the first d largest eigenvalues as the projection matrix $s|\,U_i{}^*$ that can be represented as:

$$s|\,U_i{}^* = [s|\,u_1 \quad s|\,u_2 \quad \cdots \quad s|\,u_d]. \tag{17}$$

After obtaining the projection matrix, we project the $s|\,V_i$ to the eigen subspace and achieve the d-dim principal component subset for each class as:

$$s|\,V\_P_i = \left(s|\,U_i{}^*\right)^T s|\,V_i. \tag{18}$$

where $s|\,V\_P_i$ is the PCF subset of the *ith* class. All $s|\,V\_P_i$ can form the PCF dictionary as $s|\,V\_P = [s|\,V\_P_1 \cdots s|\,V\_P_i \cdots s|\,V\_P_C]$ and PCF dictionaries from all classes under different scales can be expanded as the following matrix form:

$$\begin{bmatrix} 1|\,V\_P_1 & 1|\,V\_P_2, & \cdots & 1|\,V\_P_C \\ 2|\,V\_P_1, & 1|\,V\_P_2, & \cdots & 2|\,V\_P_C \\ \cdots & \cdots & \cdots & \cdots \\ (k-1)|\,V\_P_1 & (k-1)|\,V\_P_2 & \cdots & (k-1)|\,V\_P_C \end{bmatrix}. \tag{19}$$

TABLE 1. The accuracies under different values of $k * k$ on four datasets.

Datasets	$k*k$						
	3*3	4*4	5*5	6*6	7*7	8*8	9*9
JAFFE	68.34	69.23	**75.77**	71.41	71.54	71.59	73.97
CK+	88.80	90.17	**90.55**	89.57	88.88	90.08	90.02
KDEF	78.88	79.90	80.82	79.59	80.41	**81.12**	80.71
CMU Multi-PIE	71.85	72.92	74.34	73.18	74.34	**74.96**	74.82

Finally, we respectively fuse PCF dictionaries of each class to form the feature fusion (FF) dictionary V_P that can be represented as:

$$V\_P = \left[\sum_{s=1}^{k-1} s|\,V\_P_1, \sum_{s=1}^{k-1} s|\,V\_P_2, \cdots, \sum_{s=1}^{k-1} s|\,V\_P_C\right]. \tag{20}$$

D. THE PROPOSED APPROACH COMBING WITH SRC

After obtaining the proposed FF dictionary, we use the simple and effective SRC to classify the testing samples. The classification procedure is summarized as follows. For the sake of convenience, let $V\_PCA_i = \sum_{s=1}^{k-1} s|\,V\_P_i$, then the FF dictionary can be represented as:

$$V\_P = [V\_PCA_1, V\_PCA_2, \cdots, V\_PCA_C] \in R^{m_p \times N}. \tag{21}$$

Here m_p denotes the dimension of PCF. Given a test sample $y \in R^m$, first we obtain its AF vector y_a in a similar way as mentioned in Part *B* of Section III, and subsequently the PCF vector y_p of the test sample can be also obtained in a similar way as mentioned in Part *C* of Section III. Then we represent y_p over V_P as:

$$y_p \approx (V\_P)\,x. \tag{22}$$

where $x = [x_1, \cdots, x_i, \cdots, x_C]$ and x_i is the coefficient vector associated with the *ith* class. Generally, $y_p \approx V\_P_i x_i$ performs sparsely if y_p really comes from the *ith* class. In one word, most coefficients in x are close to zero while the coefficients in x_i has the significant entries. Then we solve the Eq. (23) via ℓ_1 minimization:

$$x = \arg\min_x \|x\|_1 \quad s.t.\ \left\|y_p - (V\_P)\,x\right\| < \varepsilon. \tag{23}$$

Then the residuals of the *ith* class is computed by:

$$r_i = \left\|y_p - (V\_P_i)\,\hat{x}_i\right\|. \tag{24}$$

where $\hat{x}_i$ is the representation coefficients corresponding to the *ith* class. Finally, y_p is classified the minimute class *i*:

$$identity\,(y) = \arg\min_i (r_i). \tag{25}$$

FIGURE 6. Some image samples from (a) JAFFE, (b) CK+, and (c) KDEF datasets.

FIGURE 7. Some image samples from CMU Multi-PIE dataset.

TABLE 2. AFs' patch-label under different scales on JAFFE dataset.

Scale	AFs' patch-label
s=1	1 2 4 5 6 7 9 10 18 19 22 23 24 25
s=2	1 3 4 8 14 15 16
s=3	2 3 6 8 9
s=4	2 3

TABLE 3. AFs' patch-label under different scales on CK+ dataset.

Scale	AFs' patch-label
s=1	1 5 6 7 9 10 11 12 14 15 16 18 19 22 23 24 25
s=2	1 4 5 8 9 11 12 14 15 16
s=3	1 2 3 4 5 6 7 8 9
s=4	1 2 3 4

TABLE 4. AFs' patch-label under different scales on KDEF dataset.

Scale	AFs' patch-label
s=1	2 3 6 7 8 9 10 11 13 14 15 16 18 19 22 23 36 37 44 45 49 51 52 53 54 56 57 64
s=2	1 2 3 5 6 7 8 9 10 12 13 14 16 31 32 33 38 39 40 41 43 44 45 46 47 48 49
s=3	1 2 3 4 5 6 7 8 9 10 11 12 27 28 29 30 31 32 33 34 35 36
s=4	1 2 3 4 5 6 7 9 10 18 21 22 23 24 25
s=5	1 2 3 4 6 7 10 11 16
s=6	1 2 4 5 6 9
s=7	1 2

IV. EXPERIMENTS AND ANALYSIS

In this section, we report on the experiments carried out to validate the performance of our approach using four publicly available datasets and adopted the leave-one-subject-out (LOSO) cross-validation method for all experiments. With LOSO, we picked one subject at a time for testing, and all images of other subjects are used for training.

TABLE 5. AFs' patch-label under different scales on CMU Multi-PIE dataset.

Scale	AFs' patch-label
s=1	1 2 8 9 10 11 14 15 16 17 18 19 22 23 24 36 37 44 45 51 52 53 54 56 57 58 63 64
s=2	1 6 7 8 9 10 12 13 14 15 16 17 19 20 21 38 39 40 43 45 46 47 48 49
s=3	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 27 28 33 34
s=4	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
s=5	1 2 3 6 7 9 10 11 12
s=6	4 5 6 7 8 9
s=7	3 4

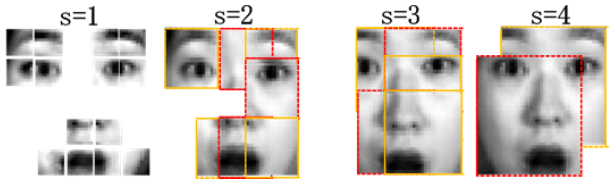

FIGURE 8. A "Su" sample example of active patches visualization from JAFFE dataset under different scale conditions. (a) $s = 1$, (b) $s = 2$, (c) $s = 3$, and (d) $s = 4$.

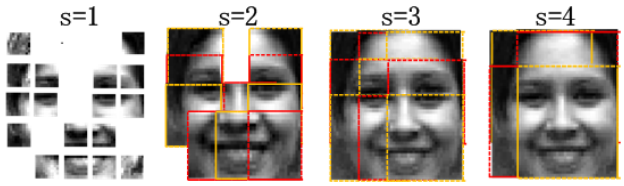

FIGURE 9. A "Ha" sample example of active patches visualization from CK+ dataset under different scale conditions. (a) $s = 1$, (b) $s = 2$, (c) $s = 3$, and (d) $s = 4$.

The input images were cropped to a size of 64*64 based on their two-eye locations [42]. These cropped images were then down-sampled to 48*48 pixels. We used abbreviations "An", "Di", "Fe", "Ha", "Sa", "Su", "Ne", "Sm", "Sq" and "Sc" to represent the expressions anger, disgust, fear, happiness, sadness, surprise, neutral, smile, squint and scream, respectively. All experiments were coded in MATLAB R2016a on a PC of Win 10 environment.

A. DATASETS

As mentioned above, JAFFE [17], KDEF [18], CK+ [19], and CMU Multi-PIE [20] datasets were used in all experiments. Amongst, the first three datasets include basic seven expression types and **Fig. 6** shows some expression images from them. CMU Multi-PIE has six expression categories that is different from the other three datasets and some sample examples are shown in **Fig. 7**.

B. AFS' PATCH SELECTION

In this subsection, we reported and analyzed the experimental results to verify the effectiveness of our approach. The results under different patch values $k*k$ on four datasets are shown in **Table 1**. From **Table 1**, we see that the results on JAFFE and

TABLE 6. Performance comparison with some the state-of-the-art methods on JAFFE dataset.

Methods	Recognition Rate (%)							
	An	Di	Fe	Ha	Sa	Su	Ne	Overall
Gray	60.00	44.83	43.75	58.06	51.61	60.00	40.00	50.92
LBP [39]	76.67	55.17	31.25	61.29	54.84	80.00	46.67	57.46
DLFS [23]	72.75	58.62	50.00	59.61	61.29	63.33	60.00	60.80
PCANet [17]	70.00	68.97	28.13	70.97	41.94	60.00	73.33	58.35
K-PCANet [18]	73.33	62.07	53.13	87.10	54.84	80.00	73.33	68.80
EDR-PCANet [19]	80.00	62.07	50.00	83.87	58.06	86.67	66.67	69.40
Proposed	73.33	75.86	78.13	77.42	67.74	80.00	80.00	**75.77**

TABLE 7. Performance comparison with some the state-of-the-art methods on CK+ dataset.

Methods	Recognition Rate (%)							
	An	Di	Fe	Ha	Sa	Su	Ne	Overall
Gray	42.22	77.97	36.00	94.20	14.29	95.12	88.57	75.92
LBP [39]	53.33	88.14	40.00	98.55	14.29	95.12	94.29	81.02
DLFS [23]	53.33	91.53	56.00	100.00	21.43	97.56	91.43	83.72
PCANet [17]	42.22	94.53	36.00	100.00	17.86	91.46	89.52	78.74
K-PCANet [18]	40.00	94.92	32.00	97.10	17.86	96.34	97.14	81.98
EDR-PCANet [19]	51.11	96.61	44.00	100.00	35.71	96.34	95.24	85.66
Proposed	84.44	86.44	48.00	100.00	50.00	97.56	100.00	**90.55**

TABLE 8. Performance comparison with some state-of-the-art methods on KDEF dataset.

Methods	Recognition Rate (%)							
	An	Di	Fe	Ha	Sa	Su	Ne	Overall
Gray	68.57	74.29	42.86	100.00	55.71	84.29	58.57	69.18
LBP [39]	64.29	78.57	45.71	95.71	51.43	90.00	82.86	72.65
DLFS [23]	74.29	80.00	62.12	97.14	59.71	91.29	85.71	78.60
PCANet [17]	54.29	75.71	41.43	97.14	50.00	82.86	85.71	69.59
K-PCANet [18]	74.29	85.71	48.57	100.00	70.00	92.86	90.00	80.20
EDR-PCANet [19]	74.29	84.29	50.00	100.00	71.43	92.86	91.43	80.61
Proposed	79.29	80.00	70.00	97.14	60.71	92.14	88.57	**81.12**

TABLE 9. Performance comparison with some state-of-the-art methods on CMU Multi-PIE dataset.

Methods	Recognition Rate (%)						
	Ne	Sm	Su	Sq	Di	Sc	Overall
Gray	88.03	60.64	67.82	54.26	54.26	97.87	70.48
LBP [39]	93.09	72.61	75.53	65.43	48.67	99.47	75.80
DLFS [23]	96.01	75.27	71.81	61.70	60.11	98.40	77.22
PCANet [17]	65.89	63.57	68.22	50.39	40.31	75.19	60.59
K-PCANet [18]	86.05	83.72	84.50	60.47	62.02	90.70	77.91
EDR-PCANet [19]	86.05	84.50	85.27	63.57	62.02	91.47	78.81
Proposed	94.88	85.09	85.81	57.45	60.90	97.61	**80.29**

CK+ reach the best accuracies of 75.77% and 90.55% when k is set to 5, and the results on KDEF and CMU Multi-PIE respectively reach the best accuracies of 81.12% and 74.96% when k is set to 8.

Also, **Tables 2-5** respectively show AFs' patch-label under different scale s on four datasets. It's worth mentioning that all results are achieved in **Table 2-5** are achieved under the best accuracies in bold in **Table 1**. For the results in **Table 2**, there should be 25 ($p = (k-s+1)^2$) patches in total for each sample in JAFFE dataset under the conditions of $k = 5, s = 1$, while we just select 14 AFs' patch-label that contribute higher than average of all patches in this case and cascade the pixels corresponding to these 14 active patches as the proposed AFs. Similar results can be seen on other datasets that is shown in **Table 3-5**. Thus, we conclude that not all AF patches contribute equally and are beneficial for representation and classification. To make the results more intuitive, we also use **Fig. 8-9** to show the visual results corresponding to the AF's patch-label in **Table 2-3**, respectively. For instance, the images under different scales in **Fig. 9** show

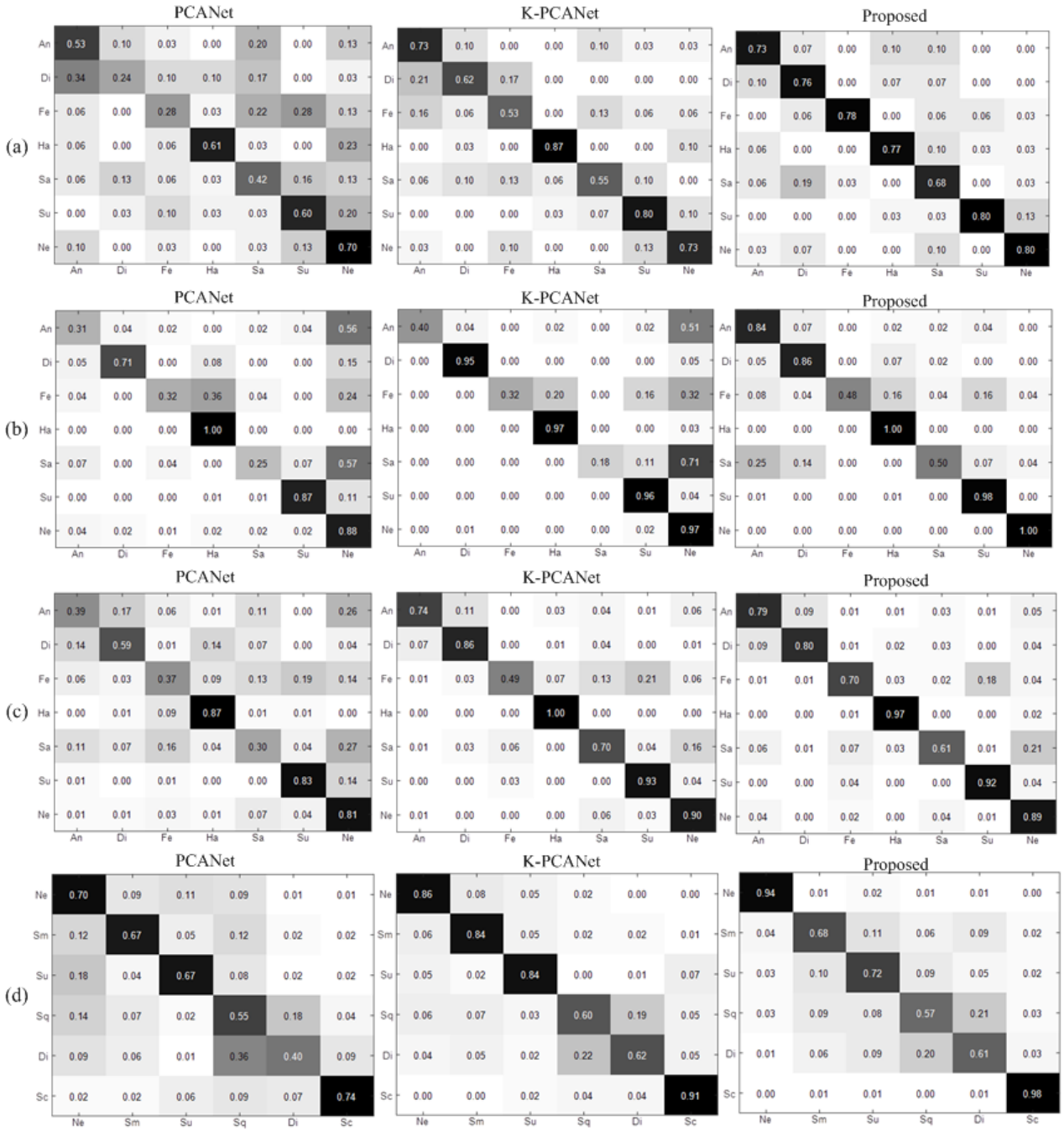

FIGURE 10. Confusion matrix for the proposed approach and comparison methods (PCANet and K-PCANet) on (a) JAFFE, (b) CK+, (c) KDEF, and (d) CMU Multi-PIE datasets, respectively.

that not all patches are selected as the AF patches. Similar results can be also found in **Fig. 8**. The visualization in these two figures also verifies that not all the areas play a positive role for representation.

C. CONFUSION MATRIX

Confusion matrix for the proposed approach and comparison methods (PCANet [29] and K-PCANet [30]) on four datasets are depicted in **Fig. 13**. We used the same parameters set in [29] and [30]. In **Fig. 10**, abscissa asix represents the true class and the vertical axis represents the predicted class, and the values on the diagonal represent the accuracies of expressions that are classified correctly. From these figures, we see that our approach performs superior to comparison methods in most cases though some expressions are wrongly classified. For example, **Fig. 10b** shows that "Ha" and "Ne" for the CK+ dataset are all classified correctly using our approach and the accuracies of other expressions are also higher than comparison methods. Accuracies of the proposed approach on other datasets are also higher than that of

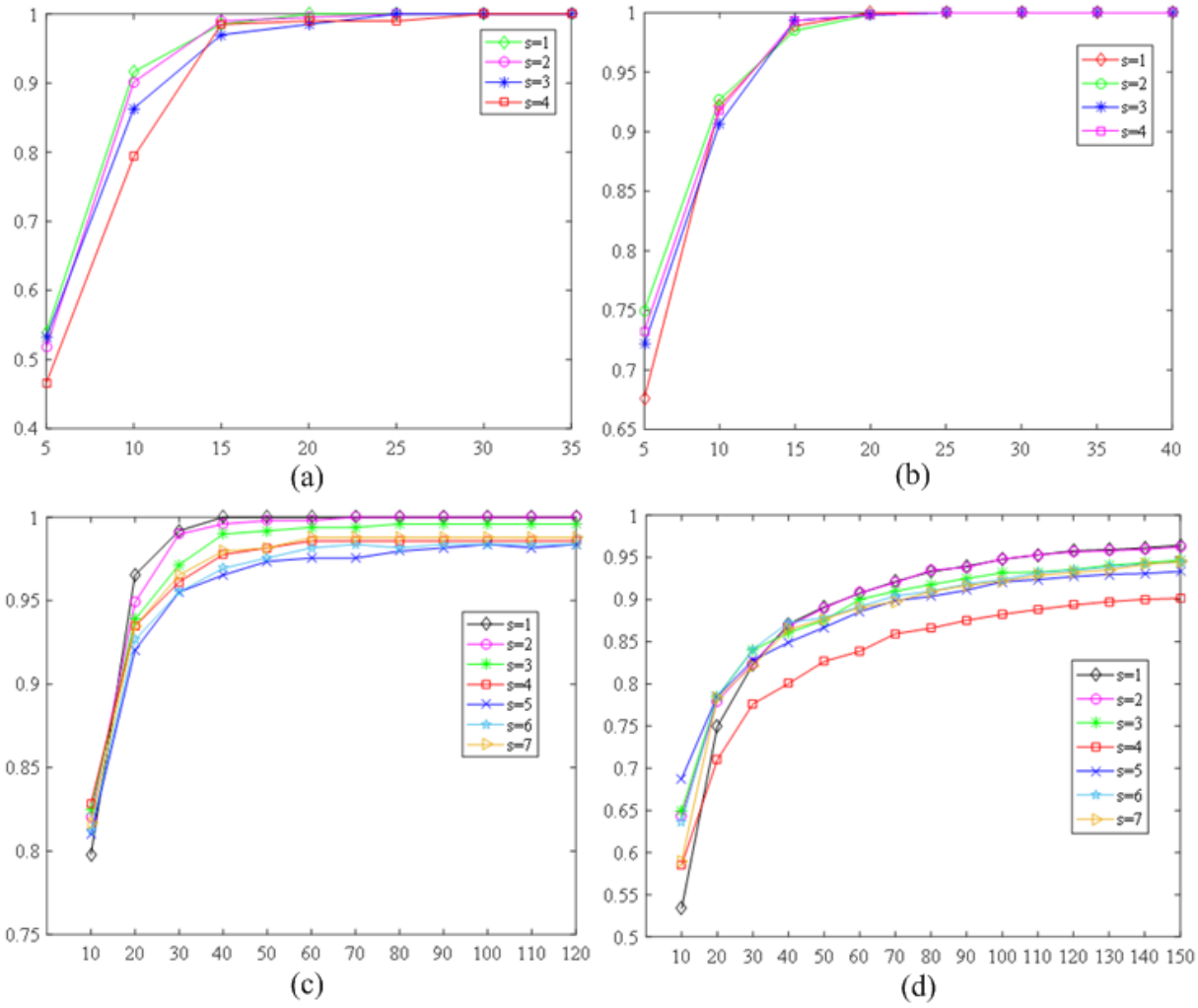

FIGURE 11. The accuracies of different dimensions under optimal patches on (a) JAFFE, (b) CK+, (c) KDEF, and (d) CMU Multi-PIE datasets, respectively. The curves indicate accuracies corresponding different scales under optimal patches.

comparable methods in most expression classes. As can be seen in **Fig. 10d**, some "Sm" images of the proposed approach are wrongly classified as "Su" or vice versa. This small portion of incorrect recognition results is mainly because these two expressions are more confused for the CMU Multi-PIE dataset. Also, the number of expression images for different classes is different. Despite the small number of incorrect classification, values that go diagonally across each matrix in **Fig. 10** indicate that most expression classes can be classified correctly, confirming that our approach is more stable and indeed performs well in expression classification.

D. COMPARISONS TO STATE-OF-THE-ART METHODS

To further verify the effectiveness of our approach, we compared the performance of the proposed approach to the state-of-the-art deep subspace learning methods including EDR-PCANet [31], K-PCANet [30], PCANet [29], and the traditional methods including DLFS [35], LBP [43] as well as the baseline Gray (raw pixel) methods. **Tables 6-9** show the recognition rates achieved by the different methods for each expression class and the overall based on the four datasets tested. From these tables, we see that the recognition rates of the proposed are much better in most cases compared to the other methods. For the JAFFE dataset, the proposed approach achieved an accuracy of 75.77% that performs superior to the second best EDR-PCANet method by 6%. Similarly, for the CK+ dataset, ours has the highest average accuracy rate of 90.55%. For the CMU Multi-PIE dataset, the proposed approach produces an average recognition rates above 80%, while the Gray can only get around 60% of accuracy. For the KDEF dataset, our proposed approach has overall accuracy rate of 81.12%. Based on the results, we can conclude that the performance of our proposed approach is superior to the other methods in comparison.

E. ANALYSIZE OF OUR APPROACH

We gave the analysis of our approach in this subsection. Our approach has the following superiorities: (1) ICLR dictionary that is projected to expression subspace mitigates

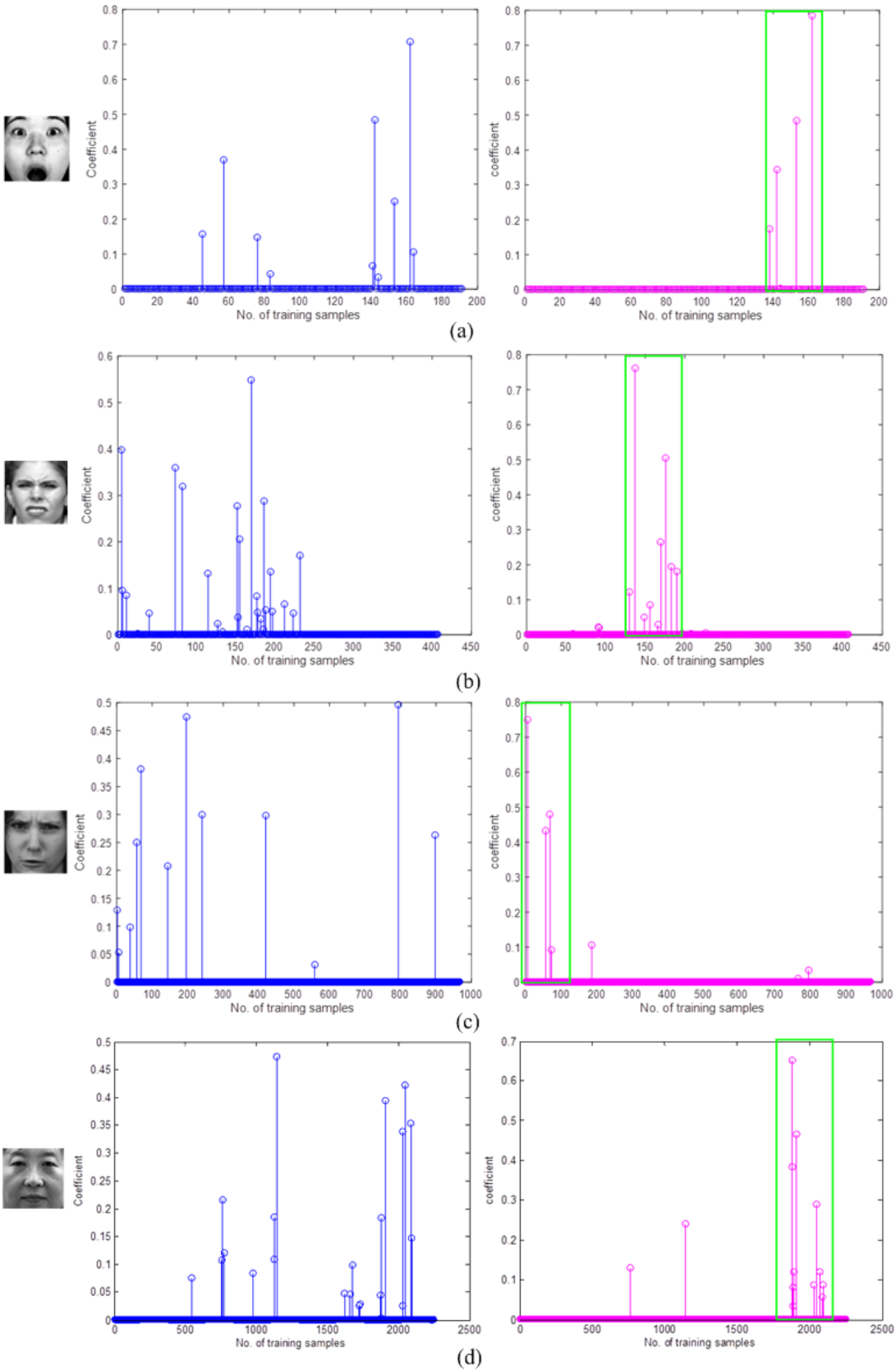

FIGURE 12. The sparseness of our approach and comparison method using four expression examples from (a) JAFFE, (b) CK+, (c) KDEF, and (d) CMU Multi-PIE datasets, respectively. Blue lines indicate the sparseness of our approach and rose red ones indicate the spareness of comparison method.

the impact of individuals and AF dictionary simultaneously takes the local importance into account; (2) PCF dictionary remits the influence of redundancy existed in AF dictionary and reduces the dimensionality, which makes the proposed approach easily performed; (3) the proposed approach mines the implicit abstract features of data based on pixel-level and avoid the high requirement for hardware as well; and (4) our approach automatically queries AF without manually setting parameters that is easy to perform for feature extraction and classification tasks.

We also provide **Fig. 11** to intuitively show the accuracies corresponding to different dimension under the optimal patch selection as demonstrated in **Tables 2-5**. As can be seen in **Fig. 11**, different curves in different colors represent different accuracies under different scales and each curve means the accuracies with the dimensions increasing. The dimension under the condition of best accuracy is selected as the parameter d. From **Fig. 11**, we also observe that curves under different scales are consistent, and thus fusing PCFs are beneficial for final representation and classification.

Besides, **Fig. 12** intuitively shows the sparseness of the proposed approach compared to the comparison method (raw pixel directly combining SRC) by respectively using four expression examples from four datasets (''Su'' from JAFFE; ''Di'' from CK+; ''An'' from KDEF, and ''Ne'' from CMU Multi-PIE, respectively). The abscissa axis means the No. of training samples and the vertical axis means the representation coefficients achieved by the proposed approach and the comparison method. From **Fig. 12**, we observe that the coefficients achieved by our approach are more discriminative than the comparison method. For instance, the coefficients obtained by comparison method on JAFFE dataset (the blue lines in **Fig. 12a**) are less clustered than that obtained by our approach. In contrast, the results by our approach (the rose red lines in **Fig. 12a**) are more focused on the samples corresponding to the true class (the 6^{th} class) and thus the sample is classified to the correct expression. Obviously, the representation coefficients (**Fig. 12b** to **Fig. 12d**) obtained using other three samples also keep consistency with the results in **Fig. 12a**. Therefore, we conclude that our approach enhances the discriminative power to a greater extent than the comparison method.

V. CONCLUSION

In this paper, we proposed a novel supervised feature extraction approach that automatically queries active features based on pixel-level for facial expression recognition. Generally speaking, our approach first succeeds in projecting the original space to ICLR subspace that removes the individual information to some extent. Then, by automatically querying active and principal component features, our approach simultaneously extracts and selects the most active features for classification. Third, our approach mines the abstract information implicated in raw data and there is no need to manually set the parameters that make our approach fast to converge. Substantial experiments on four public datasets also proved that our approach obtained promising performance compared to some state-of-the-art methods.

REFERENCES

[1] D. Mo and Z. Lai, ''Robust jointly sparse regression with generalized orthogonal learning for image feature selection,'' *Pattern Recognit.*, vol. 93, pp. 164–178, Sep. 2019.
[2] M. M. Hassan, M. G. R. Alam, M. Z. Uddin, S. Huda, A. Almogren, and G. Fortino, ''Human emotion recognition using deep belief network architecture,'' *Inf. Fusion*, vol. 51, pp. 10–18, Nov. 2019.
[3] O. Ekundayo and S. Viriri, ''Facial expression recognition: A review of methods, performances and limitations,'' in *Proc. Conf. Inf. Commun. Technol. Soc. (ICTAS)*, Mar. 2019, pp. 1–6.
[4] B. A. El-Rahiem, M. A. O. Ahmed, O. Reyad, H. A. El-Rahaman, M. Amin, and F. A. El-Samie, ''An efficient deep convolutional neural network for visual image classification,'' in *Proc. Int. Conf. Adv. Mach. Learn. Technol. Appl.*, Mar. 2019, pp. 23–31.
[5] S. Bhattacharya, G. S. Nainala, S. Rooj, and A. Routray, ''Local force pattern (LFP): Descriptor for heterogeneous face recognition,'' *Pattern Recognit. Lett.*, vol. 125, pp. 63–70, Jul. 2019.
[6] F. Z. Salmam, A. Madani, and M. Kissi, ''Fusing multi-stream deep neural networks for facial expression recognition,'' *Signal, Image Video Process.*, vol. 13, no. 3, pp. 609–616, Apr. 2019.
[7] J.-B. Yang and C.-J. Ong, ''An effective feature selection method via mutual information estimation,'' *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 6, pp. 1550–1559, Dec. 2012.
[8] Y. Lu, Z. Lai, Y. Xu, X. Li, D. Zhang, and C. Yuan, ''Low-rank preserving projections,'' *IEEE Trans. Cybern.*, vol. 46, no. 8, pp. 1900–1913, Aug. 2016.
[9] F. Ahmad, A. Khan, I. U. Islam, M. Uzair, and H. Ullah, ''Illumination normalization using independent component analysis and filtering,'' *Imag. Sci. J.*, vol. 65, no. 5, pp. 308–313, Jun. 2017.
[10] H. Ullah, M. Uzair, A. Mahmood, M. Ullah, S. D. Khan, and F. A. Cheikh, ''Internal emotion classification using EEG signal with sparse discriminative ensemble,'' *IEEE Access*, vol. 7, pp. 40144–40153, 2019.
[11] S. Xie, H. Hu, and Y. Wu, ''Deep multi-path convolutional neural network joint with salient region attention for facial expression recognition,'' *Pattern Recognit.*, vol. 92, pp. 177–191, Aug. 2019.
[12] J. Shao and Y. Qian, ''Three convolutional neural network models for facial expression recognition in the wild,'' *Neurocompting*, vol. 355, pp. 82–92, Aug. 2019.
[13] Y. Ye, X. Zhang, Y. Lin, and H. Wang, ''Facial expression recognition via region-based convolutional fusion network,'' *J. Vis. Commun. Image Represent.*, vol. 62, pp. 1–11, Jul. 2019.
[14] Z. Liu, G. Song, J. Cai, T.-J. Cham, and J. Zhang, ''Conditional adversarial synthesis of 3D facial action units,'' *Neurocomputing*, vol. 355, pp. 200–208, Aug. 2019.
[15] P. D. M. Fernandez, F. A. G. Peña, A. Cunha, and T. I. Ren, ''FERAtt: Facial expression recognition with attention net,'' 2019, *arXiv:1902.03284*. [Online]. Available: https://arxiv.org/abs/1902.03284
[16] Z. Sun, Z.-P. Hu, and M. Wang, ''Influenced factors reduction for robust facial expression recognition,'' *Multimedia Tools Appl.*, vol. 77, no. 13, pp. 16947–16963, Jul. 2018.
[17] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, ''Coding facial expressions with Gabor wavelets,'' in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit.*, Apr. 1998, pp. 200–205.
[18] E. Goeleven, R. De Raedt, L. Leyman, and B. Verschuere, ''The Karolinska directed emotional faces: A validation study,'' *Cognition Emotion*, vol. 22, no. 6, pp. 1094–1118, Aug. 2008.
[19] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, ''The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression,'' in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.-Workshops*, Jun. 2010, pp. 94–101.
[20] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, ''Multi-PIE,'' *Image Vis. Comput.*, vol. 28, no. 5, pp. 807–813, May 2010.
[21] S.-J. Wang, W.-J. Yan, G. Zhao, X. Fu, and C.-G. Zhou, ''Micro-expression recognition using robust principal component analysis and local spatiotemporal directional features,'' in *Computer Vision-ECCV* (Lecture Notes in Computer Science), vol. 8925. Cham, Switzerland: Springer, 2015, pp. 325–338.
[22] M. Turk and A. Pentland, ''Eigenfaces for recognition,'' *J. Cognit. Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.

[23] J. Yang, D. Chu, L. Zhang, Y. Xu, and J. Yang, "Sparse representation classifier steered discriminative projection with applications to face recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 7, pp. 1023–1035, Jul. 2013.
[24] Z. Zhang, Y. Xu, L. Shao, and J. Yang, "Discriminative block-diagonal representation learning for image recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 7, pp. 3111–3125, Jul. 2018. doi: 10.1109/TNNLS.2017.2712801.
[25] P. N. Belhumeur, J. P. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.
[26] G. Shikkenawis and S. K. Mitra, "On some variants of locality preserving projection," *Neurocomputing*, vol. 173, pp. 196–211, Jan. 2016.
[27] M. Li and B. Yuan, "2D-LDA: A statistical linear discriminant analysis for image matrix," *Pattern Recogn. Lett.*, vol. 26, no. 5, pp. 527–532, Apr. 2005.
[28] M. R. Mohammadi, E. Fatemizadeh, and M. H. Mahoor, "PCA-based dictionary building for accurate facial expression recognition via sparse representation," *J. Vis. Commun. Image Represent.*, vol. 25, no. 5, pp. 1082–1092, Jul. 2014.
[29] T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, "PCANet: A simple deep learning baseline for image classification," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5017–5032, Dec. 2015.
[30] Z. Sun, Z.-P. Hu, R. Chiong, M. Wang, and W. He, "Combining the kernel collaboration representation and deep subspace learning for facial expression recognition," *J. Circuits Syst. Comput.*, vol. 27, no. 8, Jul. 2018, Art. no. 1850121.
[31] Z. Sun, R. Chiong, and Z.-P. Hu, "An extended dictionary representation approach with deep subspace learning for facial expression recognition," *Neurocomputing*, vol. 316, pp. 1–9, Nov. 2018.
[32] M. H. Siddiqi, R. Ali, A. M. Khan, Y. T. Park, and S. Lee, "Human facial expression recognition using stepwise linear discriminant analysis and hidden conditional random fields," *IEEE Trans. Image Process.*, vol. 24, no. 4, pp. 1386–1398, Apr. 2015.
[33] Z. Sun, Z.-P. Hu, M. Wang, and S.-H. Zhao, "Discriminative feature learning-based pixel difference representation for facial expression recognition," *IET Comput. Vis.*, vol. 11, no. 8, pp. 675–682, Dec. 2017.
[34] S. H. Lee, K. N. K. Plataniotis, and Y. M. Ro, "Intra-class variation reduction using training expression images for sparse representation based facial expression recognition," *IEEE Trans. Affect. Comput.*, vol. 5, no. 3, pp. 340–351, Jul./Sep. 2014.
[35] Z. Sun, Z.-P. Hu, M. Wang, and S.-H. Zhao, "Dictionary learning feature space via sparse representation classification for facial expression recognition," *Artif. Intell. Rev.*, vol. 51, no. 1, pp. 1–18, Jan. 2019.
[36] P. Zarbakhsh and H. Demirel, "Low-rank sparse coding and region of interest pooling for dynamic 3D facial expression recognition," *Signal, Image Video Process.*, vol. 12, no. 8, pp. 1611–1618, Nov. 2018.
[37] A. Maronidis, D. Bolis, A. Tefas, and I. Pitas, "Improving subspace learning for facial expression recognition using person dependent and geometrically enriched training sets," *Neural Netw.*, vol. 24, no. 8, pp. 814–823, Oct. 2011.
[38] Z. Sun, Z.-P. Hu, M. Wang, F. Bai, and B. Sun, "Robust facial expression recognition with low-rank sparse error dictionary based probabilistic collaborative representation classification," *Int. J. Artif. Intell. Tools*, vol. 26, no. 4, Aug. 2017, Art. no. 1750017.
[39] L. Li, S. Li, and Y. Fu, "Learning low-rank and discriminative dictionary for image classification," *Image Vis. Comput.*, vol. 32, no. 10, pp. 814–823, 2014.
[40] C.-P. Wei, C.-F. Chen, and Y.-C. F. Wang, "Robust face recognition with structurally incoherent low-rank matrix decomposition," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3294–3307, Aug. 2014.
[41] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.
[42] Y.-L. Tian, "Evaluation of face resolution for expression analysis," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Workshops*, Jun./Jul. 2004, p. 82.
[43] Y. Ouyang, N. Sang, and R. Huang, "Robust automatic facial expression detection method based on sparse representation plus LBP map," *Optik*, vol. 124, no. 24, pp. 6827–6833, Dec. 2013.

ZHE SUN received the B.S. and Ph.D. degrees from Yanshan University, China, in 2013 and 2018, respectively. She was a Visiting Ph.D. Student with The University of Newcastle, Australia, from September 2017 to March 2018. She is currently a Lecturer with Yanshan University. Her main research interests include image processing and facial expression recognition. She has published more than 30 papers in these areas.

ZHENGPING HU received the M.Sc. degree from Yanshan University, China, and the Ph.D. degree from the Harbin Institute of Technology, China. He is currently a Professor and the Dean of the School of Electronic and Communication, Yanshan University. His main research interests include the theories and algorithms related to pattern recognition, one-class classification, and image processing.

MENGYAO ZHAO received the B.S. degree in electronic information engineering from Yanshan University, Qinhuangdao, China, in 2017, where she is currently pursuing the master's degree in electronics science and technology. Her current research interest includes video anomaly detection.