Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 9340147, 10 pages
https://doi.org/10.1155/2021/9340147

Research Article

Sparse Representation Classifier Embedding Subspace Mapping and Support Vector for Facial Expression Recognition

Shaoqin Lu,[1] Lei Xue,[2] and Xiaoqing Gu[2]

[1]*School of Digital Economics, Changzhou College of Information Technology, Changzhou, 213164 Jiangsu, China*
[2]*School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou, 213164 Jiangsu, China*

Correspondence should be addressed to Xiaoqing Gu; guxq@cczu.edu.cn

Received 17 November 2021; Revised 13 December 2021; Accepted 14 December 2021; Published 28 December 2021

Academic Editor: Xin Ning

With the development of integration and innovation of Internet and industry, facial expression recognition (FER) technology is widely applied in wireless communication and mobile edge computing. The sparse representation-based classification is a hot topic in computer vision and pattern recognition. It is one type of commonly used image classification algorithms for FER in recent years. To improve the accuracy of FER system, this study proposed a sparse representation classifier embedding subspace mapping and support vector (SRC-SM-SV). Based on the traditional sparse representation model, SRC-SM-SV maps the training samples into a subspace and extracts rich and discriminative features by using the structural information and label information of the training samples. SRC-SM-SV integrates the support vector machine to enhance the classification performance of sparse representation coding. The solution of SRC-SM-SV uses an alternate iteration method, which makes the optimization process of the algorithm simple and efficient. Experiments on JAFFE and CK+ datasets prove the effectiveness of SRC-SM-SV in FER.

1. Introduction

At present, many countries are actively developing intelligent technologies focusing on mobile edge computing [1]. The data architecture and forms in mobile edge computing are complex and diverse, resulting in great limitations of application. The main success of artificial intelligence application comes from image processing, natural language processing, social network, robots, and so on. Companies in many fields develop their applications combined with artificial intelligence technology. In the field of mobile edge computing, vision-based human-computer interaction is a very active research field. Among them, facial expression recognition (FER) technology is widely used [2]. The recognition of facial expression through artificial intelligence technology can promote the rapid development of human-computer interaction technology. The main function of FER is to recognize the expressions in the natural environment to judge people's emotions and inner activities and produce a series of FER systems [3, 4].

The applications of FER include the following: (1) security monitoring. At present, face recognition system has been widely used. It can realize the recognition of specific faces in complex crowd, predict the behavior and activities of the recognized people through the facial state of the recognized people, and analyze their action intention. In order to prevent dangerous situations in public places, these FER systems can be placed in specific public places, and facial expressions can be used to determine whether someone is engaging in illegal activities or entering illegal places. If an abnormality is found, the system will sound an alarm to avoid emergencies in public places. (2) Medical care assistance. Many hospitals have introduced robots that can detect the facial expressions of patients. These robots can determine whether there is a problem with the patient's body based on the various facial expressions of the patient. For example, when the patient's facial expression is very stable, then, the patient's physical condition is good; when the patient's face shows painful and uncomfortable expression characteristics, then, the patient's body may have problems. When a problem occurs, the ward monitor will immediately send an alarm to inform the medical staff that the patients in the ward need emergency treatment. Not only that, this type

of robot can also be used in the home of the elderly who live alone. (3) Safe driving. In order to prevent traffic accidents, the driver's FER technology application safety assistance system has begun to spread all over the world. Driving for a long time is prone to fatigue, which may lead to improper driving. Therefore, it is necessary to stop driving and have a period of rest. A camera for recognizing facial expression is installed in the vehicle, which can monitor the driver's facial expression in real time. When the expression characteristics of fatigue appear, it will remind the driver that he should rest and assist the driver to stop. The auxiliary system improves the driver's safety factor to a great extent and reduces the occurrence of unnecessary traffic accidents. (4) Entertainment. The development of interactive games has enriched people's lives. Such games mainly make corresponding judgments according to the changes of people's facial expressions. For example, when the expression shows the characteristics of panic, the expression monitoring camera of the game will immediately add relaxed elements after receiving expression feedback to relax ones nervous mood. When the expression has the characteristics of carelessness, it will give one some stimulating elements to increase the attraction of the game. (5) In the field of criminal investigation, the analysis of the subtle expression changes of the suspect can be used to determine whether the other party is lying and assist the police in solving the case.

Generally speaking, FER is to extract and analyze the features related to emotional expression of facial images and judge the emotion contained in facial images by using the prior knowledge of human emotional information. In the process of human-computer intelligent interaction, the development of good and reliable FER technology will enable the computer to well understand people's emotional state and obtain the ability to perceive and understand human social behavior, which will make the human-computer interaction system intelligent. In generally, the FER system mainly includes three stages: face detection and preprocessing, feature extraction, and FER [5]. In a complete FER process, firstly, the image is obtained through the external image acquisition equipment and detected to segment the corresponding facial expression image. Then, the facial expression-related features are extracted to obtain a good emotional feature representation of the facial expression image. Finally, the classification model is trained based on emotional feature recognition.

There are four commonly used kinds of facial expression feature extraction algorithms: geometry, texture, representation, and deep learning algorithms [6]. The geometry feature extraction algorithm is aimed at representing the structural changes of the face as a whole. It mainly uses the geometric relationship of facial points to extract the facial expression features. Lee [7] established shape model and texture model for training samples at the same time in the model establishment stage and then combined them to form active appearance models to obtain reliable expression features. It is a geometric feature extraction method. The texture feature extraction method uses the characteristics of face image pixels to represent the local subtle changes of the face expression image. Representative methods include local binary pattern (LBP) [8] and scale invariant feature transform (SIFT) [9].

The last type is to use the deep learning method to automatically learn and extract facial expression image features. For example, Kuo et al. [10] used a convolutional neural network (CNN) for FER and achieved good recognition performance. Li and Deng [11] used a bimanifold CNN (DBM-CNN) to learn the discriminative features of facial expression images. In the whole process of FER, the last stage is facial expression classification. At this stage, the obtained efficient feature representation and the labels of training data are used to train a good classifier for FER system. Sparse representation-based classification (SRC) was proposed in 2009 [12]. This algorithm has been successfully applied to face recognition, especially when the samples are damaged or occluded. However, because the SRC algorithm involves the optimization of the l_1 norm, when the data scale of the linear combination is large, the computation cost will be greatly increased, and it is not suitable for practical applications. Researches have proposed a large number of improvement algorithms to solve this problem. These algorithms can be simply divided into two types: one is to select representative training samples, and the other is to refine training sample information through dictionary learning. By selecting representative training samples, the data can be compressed to reduce the computation scale, thereby speeding up the efficiency of sparse decomposition. Li et al. [13] selected the nearest neighbor samples as the representation data. Hui et al. [14] combined SRC and local linear embedding strategy together into the speed up sparse decomposition. Ortiz and Becker [15] developed a sparse representation classification algorithm. This algorithm used linear regression to filter training samples before sparse optimization, thereby reducing computation time.

The sparse classification method based on dictionary learning can effectively accelerate the efficiency of sparse decomposition. Dictionary learning can obtain a dictionary with a small scale but a large amount of information [16, 17]. The most classical dictionary learning algorithm is the K-SVD algorithm proposed by Aharon et al. [18]. Zhang and Li [19] introduced the classification error term based on the K-SVD algorithm and proposed the discriminative D-KSVD algorithm. The dictionary learned by the algorithm has the discriminative ability. Similarly, Jiang et al. [20] made full use of the label information of training samples and proposed LC-KSVD algorithm with label consistency constraints. Due to the existence of label constraints, the coding coefficients of similar training samples were similar, so as to improve the discrimination ability. Xu et al. [21] developed a within-class-similar discriminative dictionary learning algorithm. The algorithm improved the discriminative ability by using intraclass divergence restrictions in the coding coefficients.

In order to reduce the computation scale, we consider projecting the original image into a low-dimensional subspace and embedding a multiclass support vector machine into the sparse representation classification algorithm. Based on this idea, this paper proposes a sparse representation classifying embedding subspace mapping and support vector

machine (SRC-SM-SV). In detail, when learning sparse coding, the proposed algorithm uses the Laplacian regularization term and principal component analysis to mine the geometric structure information of sample features in a low-dimensional subspace. At the same time, using the label information, this algorithm further introduces a multiclass support vector machine, so that the learned sparse representation has better discriminative ability. A series of experiments on Japanese female facial expression database (JAFFE) [22] and extended Cohn Kanade (CK+) [23] datasets are carried out to compare with the proposed algorithm, which further proves the effectiveness of the proposed algorithm.

2. Related Work

The sparse representation classification algorithm based on sparse constraint comes from the classical sparse theory. With the rapid development of mathematics-related fields, sparse representation methods based on sparse constraints have been widely used in image processing and other related fields and have achieved success in practical applications such as image restoration, classification, restoration, and segmentation. Each sample is expressed as a sparse linear combination of dictionary atoms. There are n image samples with C classes. Each sample is sparse represented by the feature vector $\mathbf{y}$, and the feature dimension is m. All training samples represent in a matrix $\mathbf{Y} \in \mathbf{R}^{m\times n}$, where $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_n]$. $\mathbf{L} = [l_1, l_2, \cdots, l_n] \in \mathbf{R}^{C\times n}$ is the class label matrix of $\mathbf{Y}$, and l_i is the class label of $\mathbf{y}_i$. $\mathbf{D}_\mathbf{z}$ is the learned dictionary matrix with N atoms. For the matrix $\mathbf{Y}$, $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_n] \in \mathbf{R}^{N\times n}$ is the sparse coding matrix on the dictionary $\mathbf{D}_\mathbf{z}$. The objective function of sparse representation can be expressed as

$$\begin{aligned} &\min_{\mathbf{D}_\mathbf{z},\mathbf{Z}} \|\mathbf{Y} - \mathbf{D}_\mathbf{z}\mathbf{Z}\|_F^2 + \alpha \sum_{i=1}^{n} \|\mathbf{z}_i\|_0, \\ &s.t.\ \|\mathbf{d}_i\|^2 \le T_0,\ \forall i, \end{aligned} \tag{1}$$

where $\|\cdot\|_F$ is Frobenius norm and $\|\cdot\|_0$ is l_0 norm. The second term of Equation (1) computes the number of nonzero items of $\mathbf{z}_i$. α is an adjustable parameter that balances coding coefficients and sample reconstruction item. Given the dictionary $\mathbf{D}_\mathbf{z}$, the minimization problem on sparse coding coefficient matrix $\mathbf{Z}$ is a NP hard problem. Equation (1) can be solved by l_1 norm by replacing l_0 norm. To deal with the classification problem, some algorithms add the classification error term into the framework of sparse representation. One of the representative algorithms is D-KSVD. Its objective function is

$$\begin{aligned} &\min_{\mathbf{D}_\mathbf{z},\mathbf{Z},\boldsymbol{\omega}} \|\mathbf{Y} - \mathbf{D}_\mathbf{z}\mathbf{Z}\|_F^2 + \alpha_1 \|\mathbf{L} - \boldsymbol{\omega}\mathbf{Z}\|_F^2 + \alpha_2 \|\boldsymbol{\omega}\|_F^2, \\ &s.t.\ \|\mathbf{d}_i\|_0 \le T_0, \forall i, \end{aligned} \tag{2}$$

where $\boldsymbol{\omega}$ is the parameter for a linear classifier. α_1 and α_2 are two regularization parameters.

As can be seen from Equation (2), the D-KSVD algorithm integrates dictionary learning and linear classifier into a framework. To promote the discriminative ability of sparse representation classification algorithms, some researches embed the idea of support vector machine into the sparse representation framework. For example, the objective function of the support vector-guided dictionary learning algorithm [24] is

$$\begin{aligned} &\min_{\mathbf{D}_\mathbf{z},\mathbf{Z},\boldsymbol{\omega}} \|\mathbf{Y} - \mathbf{D}_\mathbf{z}\mathbf{Z}\|_F^2 + 2\alpha_1 \sum_{c=1}^{C} f(\mathbf{Z}, \mathbf{l}_c, \boldsymbol{\omega}_c, b_c) + \alpha_2 \|\mathbf{Z}\|_F^2, \\ &s.t.\ \|\mathbf{d}_i\|_0 \le T_0,\ \forall i, \end{aligned} \tag{3}$$

where $\mathbf{l}_c$ and $\boldsymbol{\omega}_c$ are the label matrix and classifier parameter of the cth class sample, respectively. $f(\mathbf{Z}, \mathbf{l}_c, \boldsymbol{\omega}_c, b_c)$ is the SVM term, $f(\mathbf{Z}, \mathbf{l}_c, \boldsymbol{\omega}_c, b_c) = \|\boldsymbol{\omega}_c\|_2^2 + \delta \sum_{i=1}^{n} \ell(\mathbf{z}_i, \mathbf{l}_i, \boldsymbol{\omega}_c, b_c)$. The function $\ell(\cdot)$ is the loss function in SVM. δ is the penalty parameter.

3. Sparse Representation Classify Embedding Subspace Mapping and Support Vector

3.1. The Objective Function. The features of facial expression images are mostly high-dimensional. Therefore, the SRC-SM-SV algorithm tries to find a suitable subspace to reduce the feature dimensions and redundant information, so as to obtain more effective feature representation. In the subspace, the SRC-SM-SV algorithm fully utilizes label information of training data and adopts the Laplacian regularization term and principal component analysis term. Thus, the sparse representation is obtained by maximizing the interclass separability and minimizing the intraclass discreteness.

Denoting $\boldsymbol{\Gamma} \in \mathbf{R}^{p\times m}$ as a projection matrix, the new feature representation of image dataset $\mathbf{Y}$ can be written as $\boldsymbol{\Gamma}\mathbf{Y}$. $\mathbf{D}_\mathbf{z} = [\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_N] \in \mathbf{R}^{p\times N}$ is the dictionary learned in the new feature space. The data in the real world is easily polluted by noise. In order to better mine the data structure, we establish the Laplacian regularization term based on dictionary atoms. The element of the similarity matrix $\mathbf{G}$ is expressed as

$$G_{ij} = \begin{cases} \exp\left(-\dfrac{\|\mathbf{d}_i - \mathbf{d}_j\|_2}{\sigma}\right), & \text{if } \mathbf{d}_j \in \mathrm{KNN}(\mathbf{d}_i), \\ 0, & \text{otherwise}, \end{cases} \tag{4}$$

where $\mathrm{KNN}(\mathbf{d}_i)$ means the k-nearest neighbor function and σ is the k-nearest neighbor parameter.

The Laplacian regularization term can be written as

$$\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} (\mathbf{z}_i - \mathbf{z}_j)^2 G_{ij} = Tr(\mathbf{Z}\mathbf{L}_\boldsymbol{\Omega}\mathbf{Z}^T), \tag{5}$$

where $\mathbf{L}_\boldsymbol{\Omega} = \mathbf{H} - \mathbf{G}$ is the Laplacian matrix, $\mathbf{H} = \mathrm{diag}(G_1, \cdots, G_N)$, and $G_i = \sum_{j=1}^{N} G_{ij}$.

Based on the above idea, the objective function of SRC-SM-SV is defined as

$$\begin{aligned}<\mathbf{D}_z, \boldsymbol{\Gamma}, \mathbf{Z}, \boldsymbol{\omega}, \mathbf{b}> = \arg\min\ & \|\boldsymbol{\Gamma}\mathbf{Y} - \mathbf{D}_z\mathbf{Z}\|_F^2 + \alpha Tr\left(\mathbf{Z}\mathbf{L}_{\boldsymbol{\Omega}}\mathbf{Z}^T\right) \\ & + \beta\sum_{i,j}\eta_{i,j}\left\|\boldsymbol{\Gamma}y_i - \boldsymbol{\Gamma}y_j\right\|_2^2 - \gamma Tr\left(\boldsymbol{\Gamma}\mathbf{Y}\mathbf{Y}^T\boldsymbol{\Gamma}^T\right) \\ & + \lambda\sum_{c=1}^{C} f(\mathbf{Z}, \mathbf{l}_c, \boldsymbol{\omega}_c, b_c)\end{aligned}$$

$$s.t.\ \boldsymbol{\Gamma}\boldsymbol{\Gamma}^T = \mathbf{I},$$

$$\|\mathbf{d}_i\|_2^2 \le 1, \forall i. \tag{6}$$

The first term $\|\boldsymbol{\Gamma}\mathbf{Y} - \mathbf{D}_z\mathbf{Z}\|_F^2$ in the objective function is the reconstruction error term. The second term $Tr(\mathbf{Y}\mathbf{L}_{\boldsymbol{\Omega}}\mathbf{Y}^T)$ is the Laplacian regularization term. The third term $\sum_{i,j}\eta_{i,j}\|\boldsymbol{\Gamma}y_i - \boldsymbol{\Gamma}y_j\|_2^2$ is the mapping reconstruction term. A similarity matrix with label information $\boldsymbol{\eta}$ is constructed; the element $\eta_{i,j}$ in $\boldsymbol{\eta}$ can be defined by

$$\eta_{i,j} = \begin{cases} 1, & l_i = l_j, \\ 0, & \text{otherwise}. \end{cases} \tag{7}$$

The forth term $Tr(\boldsymbol{\Gamma}\mathbf{Y}\mathbf{Y}^T\boldsymbol{\Gamma}^T)$ in the objective function is the principal component analysis (PCA) term, which is used to mine the structure information of the data. The last term $\sum_{c=1}^{C} f(\mathbf{Z}, \mathbf{l}_c, \boldsymbol{\omega}_c, b_c)$ is the SVM term.

To simplify $\sum_{i,j}\eta_{i,j}\|\boldsymbol{\Gamma}y_i - \boldsymbol{\Gamma}y_j\|_2^2$ term, we define a matrix $\mathbf{L} = \mathbf{M} - \boldsymbol{\eta}$, where $\mathbf{M} = \text{diag}(\eta_1, \cdots, \eta_N)$ and $\eta_i = \sum_{j=1}^{N}\eta_{ij}$. We obtain the following equation:

$$\sum_{i,j}\eta_{i,j}\left\|\boldsymbol{\Gamma}y_i - \boldsymbol{\Gamma}y_j\right\|_2^2 = Tr\left(\boldsymbol{\Gamma}\mathbf{Y}\mathbf{L}\mathbf{Y}^T\boldsymbol{\Gamma}^T\right). \tag{8}$$

Then, the objective function of SRC-SM-SV is re-written as

$$\begin{aligned}<\mathbf{D}_z, \boldsymbol{\Gamma}, \mathbf{Z}, \boldsymbol{\omega}, \mathbf{b}> = \arg\min\ & \|\boldsymbol{\Gamma}\mathbf{Y} - \mathbf{D}_z\mathbf{Z}\|_F^2 + \alpha Tr\left(\mathbf{Z}\mathbf{L}_{\boldsymbol{\Omega}}\mathbf{Z}^T\right) \\ & + Tr\left(\boldsymbol{\Gamma}\mathbf{Y}(\beta\mathbf{L} - \gamma\mathbf{I})\mathbf{Y}^T\boldsymbol{\Gamma}^T\right) + \lambda\sum_{c=1}^{C} f(\mathbf{Z}, \mathbf{l}_c, \boldsymbol{\omega}_c, b_c),\end{aligned}$$

$$s.t.\ \boldsymbol{\Gamma}\boldsymbol{\Gamma}^T = \mathbf{I},$$

$$\|\mathbf{d}_i\|_2^2 \le 1, \forall i. \tag{9}$$

3.2. Solution of Optimization Variables. The alternating optimization method is used to tune the variables of $\{\mathbf{D}_z, \boldsymbol{\Gamma}, \mathbf{Z}, \boldsymbol{\omega}, \mathbf{b}\}$.

(1) First, we tune the variables $\boldsymbol{\Gamma}$. Following the Proposition in [25], let

$$\boldsymbol{\Gamma} = (\mathbf{Y}\mathbf{U})^T, \tag{10}$$

$$\mathbf{D}_z = \boldsymbol{\Gamma}\mathbf{Y}\mathbf{B}, \tag{11}$$

where $\mathbf{U} \in \mathbf{R}^{n\times p}$, $\mathbf{B} \in \mathbf{R}^{n\times N}$

Equation (9) can be repressed as

$$\begin{aligned}<\mathbf{B}, \mathbf{U}, \mathbf{Z}, \boldsymbol{\omega}, \mathbf{b}> = \arg\min\ & \left\|\mathbf{U}^T\mathbf{S}(\mathbf{I} - \mathbf{B}\mathbf{Z})\right\|_2^2 + +\alpha Tr\left(\mathbf{Z}\mathbf{L}_{\boldsymbol{\Omega}}\mathbf{Z}^T\right) \\ & + Tr\left(\mathbf{U}^T\mathbf{S}(\beta\mathbf{L} - \gamma\mathbf{I})\mathbf{S}^T\mathbf{U}\right) + \lambda\sum_{c=1}^{C} f(\mathbf{Z}, \mathbf{l}_c, \boldsymbol{\omega}_c, b_c),\end{aligned}$$

$$s.t.\ \mathbf{U}\mathbf{S}\mathbf{U}^T = \mathbf{I},$$

$$\|\mathbf{d}_i\|_2^2 \le 1, \tag{12}$$

where $\mathbf{S} = \mathbf{Y}^T\mathbf{Y}$.

When the variables $\mathbf{B}$, $\mathbf{Z}$, $\boldsymbol{\omega}$, $\mathbf{b}$ are fixed, Equation (9) can be rewritten as

$$\begin{aligned}<\mathbf{U}> = \arg\min\ & \left\|\mathbf{U}^T\mathbf{S}(\mathbf{I} - \mathbf{B}\mathbf{Z})\right\|_2^2 + Tr\left(\mathbf{U}^T\mathbf{S}(\beta\mathbf{L} - \gamma\mathbf{I})\mathbf{S}^T\mathbf{U}\right), \\ s.t.\ & \mathbf{U}\mathbf{S}\mathbf{U}^T = \mathbf{I}.\end{aligned} \tag{13}$$

Equation (13) has a closed-form solution, which has the form as

$$\mathbf{U} = \boldsymbol{\Omega}\boldsymbol{\Pi}^{-1/2}\mathbf{H}. \tag{14}$$

where $\mathbf{S} = \boldsymbol{\Omega}\boldsymbol{\Pi}\boldsymbol{\Omega}^T$ and $\mathbf{H}$ is the optimal solution of the following problem,

$$\begin{aligned}\mathbf{H} = \arg\min\ & Tr\left(\mathbf{H}^T\boldsymbol{\Delta}\mathbf{H}\right), \\ s.t.\quad & \mathbf{H}^T\mathbf{H} = \mathbf{I},\end{aligned} \tag{15}$$

where $\boldsymbol{\Delta} = \boldsymbol{\Pi}^{1/2}\boldsymbol{\Omega}^T((\mathbf{I} - \mathbf{B}\mathbf{Z})(\mathbf{I} - \mathbf{B}\mathbf{Z})^T + (\beta\mathbf{L} - \gamma\mathbf{I}))\boldsymbol{\Omega}\boldsymbol{\Pi}^{1/2}$.

Then, according to Proposition in [25], $\boldsymbol{\Gamma}$ can be tuned as $(\mathbf{Y}\mathbf{U})^T$.

(2) Tune $\mathbf{D}_z$: with the other parameters fixed, Equation (9) can be rewritten as

$$\begin{aligned}<\mathbf{D}_z> = \arg\min\ & \|\boldsymbol{\Gamma}\mathbf{Y} - \mathbf{D}_z\mathbf{Z}\|_2^2, \\ s.t.\ & \|\mathbf{d}_i\|_2^2 \le 1\end{aligned} \tag{16}$$

Using the Lagrange dual method, $\mathbf{D}_z$ can be computed as

$$\mathbf{D}_z = \left(\boldsymbol{\Gamma}\mathbf{Y}\mathbf{Z}^T\right)\left(\mathbf{Z}\mathbf{Z}^T + \boldsymbol{\Theta}^*\right)^{-1}, \tag{17}$$

where $\boldsymbol{\Theta}^*$ is a diagonal matrix constructed from all the optimal dual variables. Then, matrix $\mathbf{B}$ can be computed by $(\mathbf{U}^T\mathbf{S})^{\dagger}\mathbf{D}_z$, where $(\cdot)^{\dagger}$ denotes the pseudoinverse matrix.

(3) Tune $\mathbf{Z}$: with the other parameters fixed, Equation (9) can be rewritten as

Input: labeled training data $\mathbf{Y}$, regularization parameters α,β,λ,γ, and δ.
Output: the optimal variables $\{\mathbf{D}_z^*, \mathbf{\Gamma}^*, \boldsymbol{\omega}^*, \mathbf{b}^*\}$
1. Initialize the dictionary $\mathbf{D}_z$ using K-SVD algorithm
$t = 1$
While not convergence or $t <$ maximum number of iterations
2. Compute the similarity matrix $\mathbf{G}$ via Equation (4);
3. Tune the mapping matrix $\mathbf{\Gamma}$ via Equations (10)–(15);
4. Tune the dictionary $\mathbf{D}_z$ via Equation (17);
5. Tune sparse coefficient matrix $\mathbf{Z}$ via Equation (20);
6. Tune the $\{\boldsymbol{\omega}, \mathbf{b}\}$ via Equation (21);
$t = t + 1$
end while

Algorithm 1: SRC-SM-SV algorithm

Figure 1: The example images in the JAFFE dataset.

Figure 2: The example images in the CK+ dataset.

$$<\mathbf{Z}> = \arg\min \|\mathbf{\Gamma Y} - \mathbf{D_z Z}\|_2^2 + \alpha Tr\left(\mathbf{Z L_\Omega Z}^T\right) + \lambda \sum_{c=1}^{C} f(\mathbf{Z}, \mathbf{l}_c, \boldsymbol{\omega}_c, b_c). \tag{18}$$

Equation (18) can be expanded as

$$\arg \min_{\mathbf{z}_i} \|\mathbf{\Gamma y}_i - \mathbf{D_z z}_i\|_2^2 + \alpha Tr\left(\mathbf{z}_i \mathbf{L_\Omega z}_i^T\right) + \lambda \sum_{c \in \phi} \left\| l_i^c \left(\boldsymbol{\omega}_c^T \mathbf{z}_i + b_c\right) - 1 \right\|_2^2, \tag{19}$$

where $\phi = \{c \mid 1 \le c \le C, l_i^c(\boldsymbol{\omega}_c^T \mathbf{z}_i + b_c) - 1 > 0\}$.

We can obtain a closed-form solution of $\mathbf{z}_i$ as

$$\mathbf{z}_i = \left(\mathbf{D}_\mathbf{z}^T \mathbf{D}_\mathbf{z} + \alpha \mathbf{L}_\mathbf{\Omega} + \lambda \sum_{c \in \phi} \boldsymbol{\omega}_c \boldsymbol{\omega}_c^T\right)^{-1} \left(\mathbf{D}_\mathbf{z}^T \mathbf{\Gamma y}_i + \lambda \sum_{c \in \phi} \boldsymbol{\omega}_c (l_i^c - b_c)\right), \tag{20}$$

where l_i^c is the class label of $\mathbf{z}_i$ in the *c th* SVM.

(1) Tune $\boldsymbol{\omega}$ and b: with the other variables fixed, Equation (9) is transformed to a multiclass SVM classification problem

$$\arg \min_{\boldsymbol{\omega}, \mathbf{b}} \sum_{c=1}^{C} \left\{ \|\boldsymbol{\omega}_c\|_2^2 + \lambda \sum_{i=1}^{n} f(\mathbf{z}_i, l_i^c, \boldsymbol{\omega}_c, b_c) \right\} \tag{21}$$

Here, we utilize the multiclass SVM [26] to solve Equation (21).

When the optimization procedure is completed, we obtain the optimal variables $\{\mathbf{D}_z^*, \mathbf{\Gamma}^*, \boldsymbol{\omega}^*, \mathbf{b}^*\}$. For the testing sample $\mathbf{y}_{\text{new}}$, its sparse coding vector $\mathbf{z}_{\text{new}}$ can be computed by Equation (20). Then, its classification result can be classifier by a multiclass SVM.

The optimization of the SRC-SM-SV algorithm is shown in Algorithm 1.

4. Experiment

4.1. Datasets and Experiment Setting. The experiments of FER are conducted on JAFFE [22] and CK+ [23] datasets in this study. The JAFFE has 213 facial expression images, including a variety of facial expressions of 10 Japanese women. The samples of CK+ dataset are facial expressions from different countries, nationalities, and genders. It is a relatively perfect public dataset at present. Figures 1 and 2 show some examples of facial expression images in JAFFE and CK+ databases, respectively. In the experiment, we

Table 1: Number distribution of 6 types of facial expressions in JAFFE and CK+.

Facial expression types	JAFFE	CK+
Angry	31	45
Disgust	29	59
Fear	30	25
Happy	32	69
Sad	30	28
Surprise	31	83

Table 2: Confusion matrix (%) for the results of SRC-SM-SV using LBP feature on the JAFFE dataset.

	Angry	Disgust	Fear	Happy	Sad	Surprise
Angry	94.50	2.06	0	0	3.44	0
Disgust	4.26	88.68	1.86	0	5.20	0
Fear	1.60	3.00	87.50	0	4.60	2.30
Happy	0	0	1.20	95.00	0	3.80
Sad	2.68	4.50	3.00	0	88.62	0.20
Surprise	1.06	1.40	2.04	0	0	95.50

selected 183 images in JAFFE and 210 images in CK+. Table 1 shows the number distribution of six types of expressions in the two databases. We crop and converted these facial images to 60×60 pixels and extract LBP feature. Specifically, each face image is divided into 9 (3×3) regions and extracted 2304-dimensional LBP feature. In addition, we use a fine-tuning Res Net-50 model [27] to obtain 2048-dimensional deep feature. Six basic expressions shared by the two databases, namely, anger, disgust, fear, happiness, sadness, and surprise, were selected as the objectives of the classification task. We use the enhancement method to expand the data to 3 times of the original images for JAFFE and CK+ datasets. In the experiment, the SRC-SM-SV algorithm is compared with several algorithms, including SRC [12], K-SVD [18], LC-KSVD [19], FDDL [28], SVGDL [24], SDDL [21], and LCDL-SV [29]. In the SRC-SM-SV algorithm, the dimension of mapping subspace is set to be 500, and the learned dictionary has 420 atoms. SRC-SM-SV needs to adjust the regularization parameters α, β, λ, γ, and δ. The value range of regularization parameters is set $\{10^{-3}, 10^{-2}, \cdots, 10^{3}\}$. We empirically find that there is no rule to follow for the influence of parameter changes on recognition accuracy. Therefore, we tune the regularization parameters in the strategy of grid optimization. The parameters in other comparison algorithms are set according to their default settings. All experiments are conducted with MATLAB R2019b.

4.2. Results and Analysis. Tables 2 and 3 show the confusion matrix of the proposed SRC-SM-SV algorithm in JAFFE dataset using LBP and deep features, respectively. It can be seen in Table 2 that the recognition accuracies of happy and surprise expressions in the SRC-SM-SV algorithm are the highest, reaching 95% and 95.5%, respectively. Because the facial features of these two expressions are exaggerated and have large motion range, so the model is easier to extract features. The recognition effects of the SRC-SM-SV algorithm on disgust, fear, and sad expressions are slightly poor, but the recognition rates are also about 87.5%. Because both fear and sad have the characteristics of opening their lips and tense forehead, while disgust and sad have similar eyebrow characteristics and wrinkled corners of the mouth, these three expressions have certain similarities and are prone to misclassification. In addition, the recognitions of disgust and sad are easy to interfere with each other.

Table 3: Confusion matrix (%) for the results of SRC-SM-SV using deep feature on the JAFFE dataset.

	Angry	Disgust	Fear	Happy	Sad	Surprise
Angry	94.50	2.50	0	0	3.00	0
Disgust	1.00	91.50	2.00	0	5.50	0
Fear	2.20	2.20	88.50	0	5.60	2.50
Happy	0	0	1.25	95.20	0	3.55
Sad	1.25	3.88	2.71	0	91.26	0.90
Surprise	0.45	1.50	1.85	0	0	96.20

Table 4: Confusion matrix (%) for the results of SRC-SM-SV using the LBP feature on the CK+ dataset.

	Angry	Disgust	Fear	Happy	Sad	Surprise
Angry	90.75	7.00	1.25	0	1.00	0
Disgust	1.00	95.00	0	0	4.00	0
Fear	0	1.00	87.5	5.50	3.60	2.40
Happy	0	0	0	95.5	0	4.50
Sad	1.52	4.50	1.60	0	92.38	0
Surprise	0	0	2.00	1.00	0	97.00

Table 5: Confusion matrix (%) for the results of SRC-SM-SV using deep feature on the CK+ dataset.

	Angry	Disgust	Fear	Happy	Sad	Surprise
Angry	95.50	3.00	1.00	0	0.50	0
Disgust	2.00	95.00	0	0	3.00	0
Fear	0	2.00	91.5	2.5	2.00	2.00
Happy	0	0	0	98.00	0	2.00
Sad	1.50	4.00	1.50	0	93.00	0
Surprise	0	0	2.00	1.00	0	97.00

Tables 4 and 5 show the confusion matrix of the proposed algorithm on CK+ dataset using LBP and deep features, respectively. The confusion matrixes of the model on CK+ dataset are similar to those on the JAFFE dataset. As can be seen from Tables 4 and 5, the SRC-SM-SV algorithm is the easiest to recognize happy expression, and the recognition rate reaches 98% when using deep feature. The second best recognition rate is surprise, with a recognition rate of 97% when using deep feature. Because surprise is exaggerated and its features are easier to learn, so the recognition rate of surprise expression is also higher. Disgust and sad are confused and lead to errors in recognition, because their

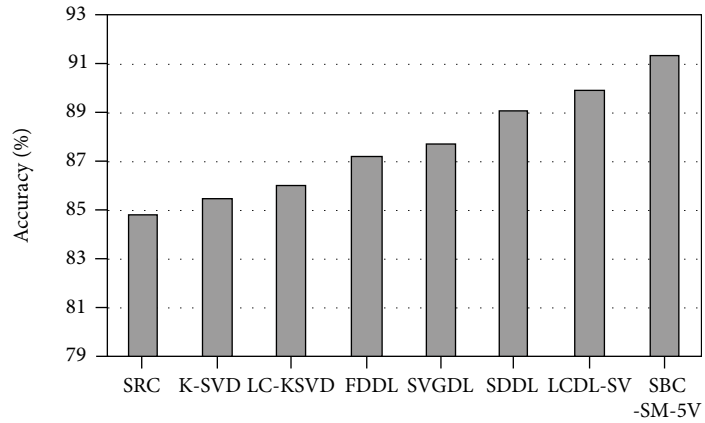

Figure 3: The comparison results of LBP feature on the JAFFE dataset.

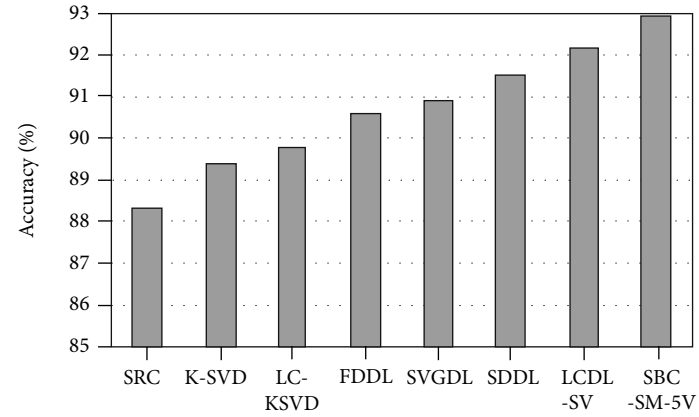

Figure 4: The comparison results of deep feature on the JAFFE dataset.

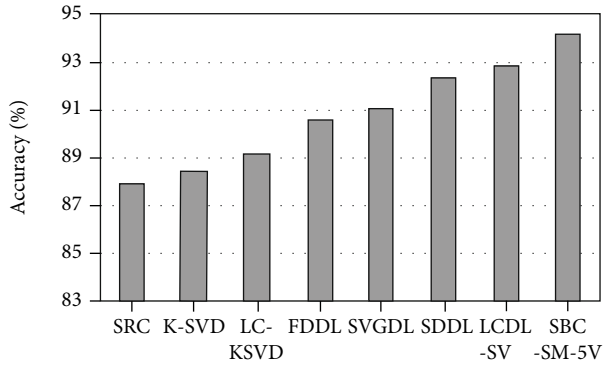

Figure 5: The comparison results of LBP feature on the CK+ dataset.

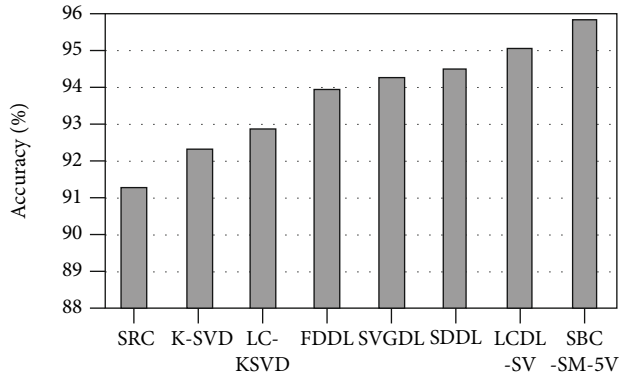

Figure 6: The comparison results of deep feature on the CK+ dataset.

expressions are similar, especially the part of the mouth. Fear expression recognition is relatively weak, mainly because the number of fear samples is small and the features that can be learned are relatively small.

The comparison results on the JAFFE dataset using LBP and deep features are shown in Figures 3 and 4, respectively. Firstly, the results in Figure 3 show that SRC-SM-SV obtains the best performance under the LBP feature. Secondly, the results in Figure 4 show that SRC-SM-SV also obtains the best performance under the deep feature. The average recognition rate of the SRC-SM-SV algorithm on all expressions is 94.18%, which improves the performance by 6.29% compared with SRC and 1.28% compared with the second best LCDL-SV algorithm. On the one hand, it shows that the deep feature of automatic learning can be well used for recognition and classification. On the other hand, it shows the advantages of the proposed SRC-SM-SV. Compared with the LCDL-SV algorithm, the objective function of the

Table 6: Performance results on the JAFFE dataset using LBP feature.

	Angry	Disgust	Fear	Happy	Sad	Surprise
SRC	89.13	83.03	83.14	88.05	82.04	83.14
K-SVD	89.85	83.86	83.99	88.52	82.75	83.61
LC-KSVD	90.80	84.57	84.24	89.11	83.01	83.68
FDDL	92.13	85.68	84.70	90.65	84.63	84.74
SVGDL	93.09	85.74	84.60	90.78	84.69	86.87
SDDL	94.29	87.01	86.13	92.87	85.28	88.65
LCDL-SV	94.66	87.30	86.68	94.03	85.61	90.46
SRC-SM-SV	**95.00**	**88.68**	**87.50**	**95.00**	**88.62**	**92.54**

The bold values in Tables 6–9 means the best classification results in the experiments.

Table 7: Performance results on the JAFFE dataset using deep feature.

	Angry	Disgust	Fear	Happy	Sad	Surprise
SRC	90.45	87.30	85.21	90.88	86.65	88.90
K-SVD	91.57	88.74	86.06	92.79	86.73	90.03
LC-KSVD	91.69	89.06	86.14	93.29	87.67	90.39
FDDL	92.39	89.91	87.56	93.88	88.43	90.99
SVGDL	93.01	90.05	87.16	94.33	89.38	91.19
SDDL	93.75	90.26	88.14	94.60	89.59	92.27
LCDL-SV	95.02	90.56	88.23	94.74	90.40	93.59
SRC-SM-SV	**96.50**	**91.50**	**88.50**	**95.20**	**91.26**	**94.20**

Table 8: Performance results on the CK+ dataset using LBP feature.

	Angry	Disgust	Fear	Happy	Sad	Surprise
SRC	86.54	86.91	85.24	91.02	86.47	88.02
K-SVD	86.63	87.28	85.87	91.75	87.08	88.86
LC-KSVD	87.54	87.98	86.40	92.44	87.77	89.47
FDDL	88.10	89.96	86.30	92.56	90.04	92.28
SVGDL	88.48	90.28	86.41	92.83	90.62	93.24
SDDL	89.33	91.82	87.16	94.54	90.91	94.35
LCDL-SV	90.17	92.32	87.16	95.16	91.24	95.09
SRC-SM-SV	**90.75**	**95.00**	**87.50**	**95.50**	**92.38**	**97.00**

proposed algorithm includes subspace mapping and PCA terms, which shows that these two terms have significant benefits to FER tasks. The comparison results on the CK+ dataset using LBP feature and deep feature are shown in Figures 5 and 6, respectively. The results are similar to those in Figures 3 and 4. The results of Figures 5 and 6 also illustrate that our proposed algorithm is effective.

To further compare the recognition accuracy of each expression, Tables 6 and 7 show the recognition accuracy of each comparison algorithm on the JAFFE dataset in six expressions using LBP and deep features, respectively. Tables 8 and 9 show the recognition accuracy of each comparison algorithm on the CK+ dataset in six expressions using LBP and deep features, respectively. From these results, we can see that SRC-SM-SV algorithm shows good recognition performance in each expression and obtains better performance than other algorithms. Especially on the JAFFE dataset, the recognition rate of SRC-SM-SV is 3 to 6 percentage higher than other algorithms, which also shows that this algorithm is suitable for FER.

Table 9: Performance results on the CK+ dataset using deep feature.

	Angry	Disgust	Fear	Happy	Sad	Surprise
SRC	90.88	90.63	87.04	92.55	90.82	92.34
K-SVD	91.61	92.06	88.91	94.34	91.41	92.44
LC-KSVD	92.23	92.70	89.37	95.31	92.09	92.57
FDDL	93.38	93.28	90.99	96.41	92.72	93.86
SVGDL	93.41	94.15	90.98	96.91	92.63	94.12
SDDL	94.04	94.28	91.30	97.28	93.05	94.16
LCDL-SV	95.01	94.51	91.47	97.68	93.12	94.82
SRC-SM-SV	**95.50**	**95.00**	**91.50**	**98.00**	**93.00**	**97.00**

5. Conclusion

As one of the most important tasks in the field of emotional computing, FER has wide applied in many practical applications, such as computer vision, multimedia entertainment, and machine intelligence. To improve the recognition performance of FER technology in practical applications, this paper explores and studies the FER system based on sparse representation classification. Then, a sparse representation classify embedding subspace mapping and support vector machine is developed in this paper. In this algorithm, the subspace learning and support vector machine are combined into the framework of sparse representation classification to obtain the discriminative sparse representation in low dimensional subspace. At the same time, this algorithm combines Laplacian regularization term and PCA term into the model to better minimize the intraclass discreteness and maximize the separability between classes. Although this study has made some achievements in FER, there are still many problems worthy of further research and exploration in practical applications. For example, the SRC-SM-SV algorithm can be extended to more complex FER tasks, such as multisource cross-database scenes, multimodal/cross-modal scenes, multiview scenes, and facial expression action unit recognition. In addition, the computational complexity of our algorithm is relatively high. At present, our algorithm is difficult to be applied in a large-scale dataset. Therefore, exploring the theory of fast sparse optimization is a direction of our research.

Data Availability

Two public datasets JAFFE and CK+ are used in this study. The JAFFE dataset can be downloaded in the hyperlink: https://zenodo.org/record/3451524#.YZT9EFVByM8. The CK+ dataset can be downloaded in the hyperlink: https://www.kaggle.com/shawon10/ckplus.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the Project of Jiangsu Education Science in the 13th five year plan in 2018 under Grant No. B-a/2018/01/41, Future Network Scientific Research Fund Project (No. FNSRFP-2021-YB-36), and Science and Technology Project of Changzhou City (No. CE20215032).

References

[1] M. Yang, Y. Ma, Z. Liu, H. Cai, X. Hu, and B. Hu, "Undisturbed mental state assessment in the 5G era: a case study of depression detection based on facial expressions," *IEEE Wireless Communications*, vol. 28, no. 3, pp. 46–53, 2021.

[2] J. Yang, T. Qian, F. Zhang, and S. U. Khan, "Real-time facial expression recognition based on edge computing," *IEEE Access*, vol. 9, pp. 76178–76190, 2021.

[3] Z. Xi, Y. Niu, J. Chen, X. Kan, and H. Liu, "Facial expression recognition of industrial internet of things by parallel neural networks combining texture features," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 4, pp. 2784–2793, 2021.

[4] T. Ni, X. Gu, C. Zhang, W. Wang, and Y. Fan, "Multi-task deep metric learning with boundary discriminative information for cross-age face verification," *Journal of Grid Computing*, vol. 18, no. 2, pp. 197–210, 2020.

[5] Y. Kortli, M. Jridi, A. Al Falou, and M. Atri, "Face recognition systems: a survey," *Sensors*, vol. 20, no. 2, 2020.

[6] I. Azam and S. A. Khan, "Feature extraction trends for intelligent facial expression recognition: a survey," *Informatica-Journal of Computing and Informatics*, vol. 42, no. 4, pp. 507–514, 2018.

[7] Y. H. Lee, "Virtual representation of facial avatar through weighted emotional recognition," *International Journal of Internet Protocol Technology*, vol. 10, no. 1, pp. 30–35, 2017.

[8] 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.

[9] P. C. Ng and S. Henikoff, "SIFT: predicting amino acid changes that affect protein function," *Nucleic Acids Research*, vol. 31, no. 13, pp. 3812–3814, 2003.

[10] C. M. Kuo, S. H. Lai, and M. Sarkis, "A compact deep learning model for robust facial expression recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 2121–2129, New York, NY, USA, 2018.

[11] S. Li and W. Deng, "Blended emotion in-the-wild: multi-label facial expression recognition using crowdsourced annotations and deep locality feature learning," *International Journal of Computer Vision*, vol. 127, no. 6-7, pp. 884–906, 2019.

[12] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on pattern analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[13] C. Li, J. Guo, and H. Zhang, "Local sparse representation based classification," in *2010 20th International Conference on Pattern Recognition*, pp. 649–652, Istanbul, Turkey, 2010.

[14] K. Hui, C. Li, and L. Zhang, "Sparse neighbor representation for classification," *Pattern Recognition Letters*, vol. 33, no. 5, pp. 661–669, 2012.

[15] E. G. Ortiz and B. C. Becker, "Face recognition for web-scale datasets," *Computer Vision and Image Understanding*, vol. 118, no. 1, pp. 153–170, 2014.

[16] X. Gu, C. Zhang, and T. Ni, "A hierarchical discriminative sparse representation classifier for EEG signal detection," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 5, pp. 1679–1687, 2021.

[17] T. Ni, C. Zhang, and X. Gu, "Transfer model collaborating metric learning and dictionary learning for cross-domain facial expression recognition," *IEEE Transactions on Computational Social Systems*, vol. 8, no. 5, pp. 1213–1222, 2021.

[18] M. Aharon, M. Elad, and A. Bruckstein, "$rm K$-SVD: an algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[19] Q. Zhang and B. Li, "Discriminative K-SVD for dictionary learning in face recognition," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2691–2698, San Francisco, CA, USA, 2010.

[20] Z. Jiang, Z. Lin, and L. S. Davis, "Label consistent K-SVD: learning a discriminative dictionary for recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2651–2664, 2013.

[21] Y. Xu, Z. Li, B. Zhang, J. Yang, and J. You, "Sample diversity, representation effectiveness and robust dictionary learning for face recognition," *Information Sciences*, vol. 375, no. 1, pp. 171–182, 2017.

[22] M. J. Lyons, J. Budynek, and S. Akamatsu, "Automatic classification of single facial images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 12, pp. 1357–1362, 1999.

[23] 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 *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 94–101, San Francisco, CA, USA, 2010.

[24] S. Cai, W. Zuo, L. Zhang, X. Feng, and P. Wang, "Support vector guided dictionary learning," in *2014 European Conference on Computer Vision*, pp. 624–639, Zurich, Switzerland, 2014.

[25] L. Qi, J. Huo, X. Fan, Y. Shi, and Y. Gao, "Unsupervised joint subspace and dictionary learning for enhanced cross-domain person re-identification," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 6, pp. 1263–1275, 2018.

[26] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1794–1801, Miami, FL, 2009.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, NV, USA, 2016.

[28] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Fisher discrimination dictionary dearning for sparse representation," in *2011 International Conference on Computer Vision*, pp. 543–550, Barcelona, Spain, 2011.

[29] H. Yin, X. Wu, and S. Chen, "Locality constraint dictionary learning with support vector for pattern classification," *IEEE Access*, vol. 7, pp. 175071–175082, 2019.