IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON ADVANCED DATA MINING METHODS FOR SOCIAL COMPUTING

Received July 12, 2019, accepted July 29, 2019, date of publication August 2, 2019, date of current version September 9, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2932784

A Multiview Representation Framework for Micro-Expression Recognition

TIANHUAN HUANG, LEI CHEN, YUNCONG FENG, XIANYE BEN, (Member, IEEE), RUIXUE XIAO, AND TIANLE XUE

School of Information Science and Engineering, Shandong University, Qingdao 266237, China

Corresponding author: Xianye Ben (benxianye@gmail.com)

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFC0831001, in part by the Natural Science Foundation of China under Grant 61571275 and Grant 61971468, in part by the Key Research and Development Program of Shandong Province, and in part by the Young Scholars Program of Shandong University.

ABSTRACT Multiview representation has become important due to its good performance for machine learning problems. In this paper, a multiview representation framework based on transfer learning is proposed for micro-expression recognition. The framework takes macro-expression as the auxiliary domain and micro-expression as the target domain, and assists the identification of micro-expressions by transferring the rich information extracted from the auxiliary domain, which effectively addresses the small sample problem of micro-expression recognition. The proposed algorithm mainly consists of three parts. Firstly, the features of the two domains are projected into a common space and the dictionaries of each domain are studied respectively. Then the dictionary of micro-expression domain is linearly reconstructed. Finally, in order to improve the comprehensive utilization of feature information, the most representative features from four different micro-expression feature sets are selected by multiview representation. The experiments and evaluation are carried out on three different databases, and the performance comparison of the proposed algorithm with other advanced methods are given. The experimental results show that the proposed algorithm has the better performance than other related methods.

INDEX TERMS Multiview representation, transfer learning, sparse dictionary learning, micro-expression recognition.

I. INTRODUCTION

Micro-expression is an involuntary facial expression during a very short time when people are in a suppressed state or try to hide their real feelings. In 1966, Hagrid and Issacs [1] first discovered this subtle expression. Compared with the traditional expression, micro-expression is more likely to expose a person's real inner activity, so it has been identified as a reliable clue to detect lies. In recent years, micro-expressions have shown good application prospects in security monitoring, case investigation, network security, military affairs, business and even entertainment field. However, the duration of the micro-expression is very short, which makes it extremely difficult to accurately recognized with naked eyes. In 2002, Endres and Laidlaw [2] developed a micro-expression training tool (METT), which effectively improved the accuracy of micro-expression recognization.

The associate editor coordinating the review of this article and approving it for publication was Shirui Pan.

With the development of facial expression algorithms [3], [4], great progresses have been made in micro-expression research [5]–[7]. However, there are still two issues to be addressed. Firstly, micro-expression is an expression sequence, but there is no conclusion on what kind of features should be extracted to improve the recognition rate. Secondly, the number of correctly labeled micro-expressions is very limited because of extensive manpower and time costs. As a result, the performances of most existing micro-expression recognition methods are far from perfect.

Recently, a few feature extraction methods have been proposed to extract more effective and representative features. Jia *et al.* [9] introduced some novel macro-expression and micro-expression feature descriptors, which include dual-cross patterns from three orthogonal planes, multiple order dual-cross patterns, hot wheel patterns, etc. Huang *et al.* [8] proposed a new feature selection method based on Laplacian algorithm to extract the discriminative

information for facial micro-expression recognition. These methods have achieved good performance in specific domain. In addition to selecting effective features, the design of classifier also affects the recognition rate. Jian and Lam [10], [11] proposed a framework based on singular value decomposition (SVD) for performing both face hallucination and recognition simultaneously, which greatly improved the accuracy of classification. Gong *et al.* [12] comprehensively considered data distribution information and proposed a new discriminant positive and unlabeled learning (PU) classifier, which effectively improved the performance of the algorithm.

Although, the recognition rate of micro-expressions has increased due to the improvement of various classification algorithms, the small sample problem is still a big challenge. It is encouraging that transfer learning can use the knowledge of source domains to solve the problems in related target domains, which may be helpful to deal with some small sample problems. Transfer learning can be divided into three cases: inductive learning, transductive learning and unsupervised transfer learning [13]. Transfer learning has been widely used in data mining, computer vision, natural language processing and other fields, and has achieved inspiring results. Yeh *et al.* [14] proposed a field adaptive algorithm which uses canonical correlation analysis (CCA) to project all the data into a common space. Yang *et al.* [15] explored abundant well-tagged images to facilitate the video tagging, which aimed at finding an optimal kernel space to minimize the distance between images and videos.

Inspired by ideas above, we try to apply transfer learning to micro-expression recognition problem. However, how to choose an effective auxiliary domain and explore its correlations with micro-expressions to the greatest degree still need to be further researched. Compared with traditional facial macro-expression, micro-expression is significantly different in terms of duration and range of motion. But it is undeniable that they are very similar in nature. Besides, the classification criteria for them are both based on the facial muscle movement unit. Therefore, it is reasonable to use macro-expressions as auxiliary domain to assist the recognition of micro-expression.

Taking macro-expressions as auxiliary domain and transferring rich information learned from macro-expressions to assist micro-expression recognition belongs to cross-domain transfer learning. However, using cross-domain transfer learning to identify micro-expressions comes with another problem. Features extracted with different methods can represent different characteristics of the data. In the cross-domain transfer learning, we can not subjectively determine what micro-expression features are more representative and the simple cascade of multiple features can easily lead to information redundancy and noise. To address this problem, we first introduced the multiview representation. We extracted different feature sets of the target domain for cascading and then sparsed the multiple feature sets with F norm to select the most representative features. After sparse feature selection, the sparse dictionary learning was used to process a large number of high-dimensional features, which could avid feature redundancy and noise to a certain extent.

Based on the above research and analysis, we proposed a multiview representation framework based on transfer learning for micro-expression recognition. The main contributions of this paper can be summarized as follows:

- We proposed a novel framework based on transfer learning for micro-expression recognition to solve small sample problems. Moreover, the sparse dictionary learning was used to process high-dimensional features, which could effectively avoid information redundancy and noise interference.
- To further reduce domain shift, we linearly reconstructed the dictionary of the target domain with auxiliary domain dictionary to achieve high correlations between the two domains.
- We introduced multiview representation into the framework for multiple feature selection. Though multiview representation, the most representative features were selected and the advantages of different feature sets were combined, which was conducive to improve the recognition results of micro-expression.

The rest of this paper is structured as follows: In section II, the research status of micro-expression recognition, multiview representation and transfer learning are introduced in detail. The proposed multiview representation framework and the optimal solution of it are given in section III and section IV respectively. Experimental validation and algorithm performance evaluation are presented in section V. Finally, section VI concludes the entire paper.

II. RELATED WORK

A. THE RESEARCH STATUS OF MICRO-EXPRESSION RECOGNITION

In recent years, micro-expression recognition has attracted considerable attention with the development of machine learning and face recognition technologies. Micro-expression recognition mainly includes three parts: micro-expression detection, feature extraction and recognition. In the aspect of micro-expression detection, Polikovsky *et al.* [16] collected simulated micro-expressions under high-speed cameras and used the k-means algorithm to locate the start frame, vertex frame and end frame of a posed micro-expression sequence.

Feature extraction methods include local feature extraction and global feature extraction. Pfister *et al.* [6] firstly extended the features of the Local Binary Pattern (LBP). They added time series to the traditional LBP and realized the analysis of dynamic micro-expressions. Zhang *et al.* [17] proposed a normalized difference vector (NDV) for texture representation. Different from descriptors based on local binary patterns, this descriptor makes full use of local differences and can be flexibly expanded to cover larger local areas. However, many methods either based on local feature or global feature are heavily dependent on facial

coordinate localization. Besides, many feature extraction methods have difficulty explaining the mechanism of facial micro-expression [18], [19]. In response to this situation, Liu *et al.* [20] proposed a simple and effective main direction mean optical flow (MDMO) for micro-expression extraction. MDMO is a kind of normalized statistical feature based on the region of interest. It considers both the local statistical information and spatial information, and has a good performance in micro-expression recognition.

B. MULTIVIEW REPRESENTATION

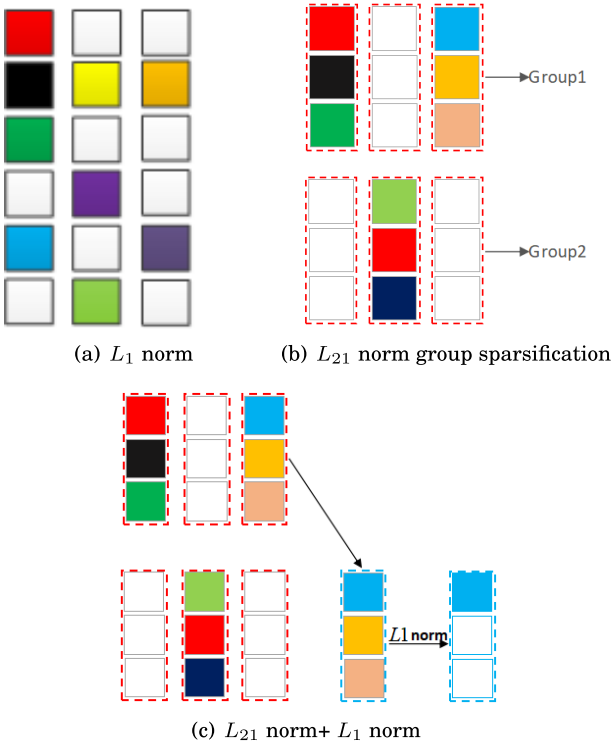
Multiview representation is a method of representing a topic by using multiple sets of different features, and has been applied to various branches of machine learning [21]. Compared with a single feature, an ontology can be described more specially by using different feature sets simultaneously. However, directly using different feature sets will increase the computing cost and is probably to produce information redundancy or even noise interference.

In machine learning, we usually introduce different kinds of regularizations according to different requirements to produce different sparse methods. Inspired by this fact, we incorporated regularization items into the multiple feature selection. The sparse methods can be summarized into two categories: separable sparse learning and joint sparse learning [22].

Separate sparse learning performs sparsification to only one sample each time. The L_1 norm, L_{21} norm and the combination of the two norms are commonly used for separate sparse learning. Since the L_1 norm sparses each sample independently [23], [24], its sparse results are reflected in each feature sample, and the effect is shown in Figure 1-(a). Unlike the L_1 norm, the L_{21} norm performs group sparsification [25] to samples, which can preserve the data structure better, and its principle is shown in Figure 1-(b). L_1 norm and L_{21} norm can be cascaded for the combination of their characteristics. As showed in Figure 1(c), it firstly uses L_{21} norm to perform group sparsification to samples and then uses L_1 norm for further sparsification to the samples of different feature groups.

The joint sparse method performs sparse to all the samples simultaneously. Different from separable sparse, the joint sparse partitions the samples and performs sparsification to the samples within the same block simultaneously. The common sparse patterns include L_{21} norm sparse, F norm sparse, etc. The L_{21} norm regularization penalizes each row of the feature matrix as a whole and enforces sparsity among the rows, which can select the most representative features [26]. The specific principle is shown in Figure 2-(a). And the F norm can perform block sparse towards the feature data, which is showed in Figure 2-(b).

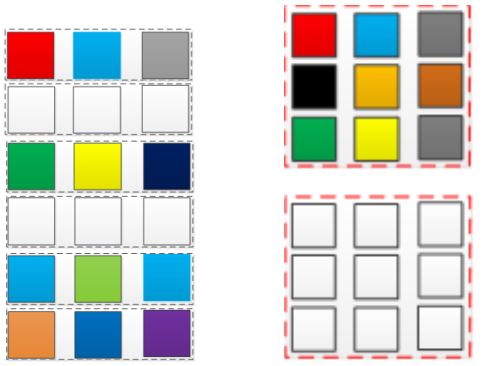
Based on the above analysis, we introduced the concept of multiview representation. We firstly extracted multiple feature sets of micro-expressions for cascading and then performed F norm sparseness on them.



(a) L_1 norm (b) L_{21} norm group sparsification

(c) L_{21} norm+ L_1 norm

FIGURE 1. Three kinds of separate sparse patterns. Each column represents a sample and the row represents feature dimension. The white block is the sparse element while the colored one is the dense element.



(a) L_{21} norm line sparse (b) F norm block sparse

FIGURE 2. Two different patterns of joint sparse.

C. TRANSFER LEARNING

Recently, the research of cross-domain transfer learning has been developed rapidly. Shekhar *et al.* [27] proposed a coupling adaptive dictionary learning approach, with which the data from different domains were projected into a common subspace where the coupling dictionary was studied. Ren *et al.* [28] used coupling margin fisher analysis for transfer learning. It not only maintained each domain's individual characteristics, but also explored the commonality between the source domain and target domain. Masci *et al.* [29] mapped the training data to the hash space, which ensured a small intra-class distance and a large inter-class distance, and

also built a bridge between the data from different domains. Yang *et al.* [30] proposed a common dictionary learning method instead of training the dictionaries separately, which can ensure the common distribution of data from different domains to a certain extent.

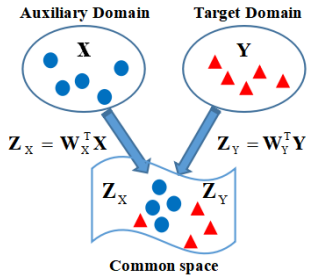
It can be concluded from the above methods that transfer learning can be applied for micro-expression recognition and it is very important to find the relevance of different domains for cross-domain micro-expression identification. However, there are two problems to be solved in feature extraction of different domains. Firstly, how to choose the most effective features from multiple feature sets extracted with differernt methods. Secondly, each feature set of different domains may have a lot of redundancy and noise, which will seriously lower the recognition performance. The first problem has been discussed in Section II-B and the feature dimension reduction will be introduced in this section.

When dealing with the features of different domains, there are a lot of redundant information and noise in high-dimensional features, which will annihilate the effective information. However, if the feature dimension is reduced directly, the natural structure of data may be destroyed. In order to address this problem, we applied sparse coding for feature dimension reduction. When processing high-dimensional features, sparse dictionary learning is an effective technique, which can obtain a common representation of the features of two different domains. Yang *et al.* [30] proposed two methods for the sparse dictionary learning of different domains. One approach is to perform sparse description towards data in each domain separately, which takes consideration of the natural features of each domain but ignores the relevance of different domains. The other one is to perform common dictionary learning towards data in different domains, which can obtain the common feature representation of different domains, but the structure of each domain will be lost.

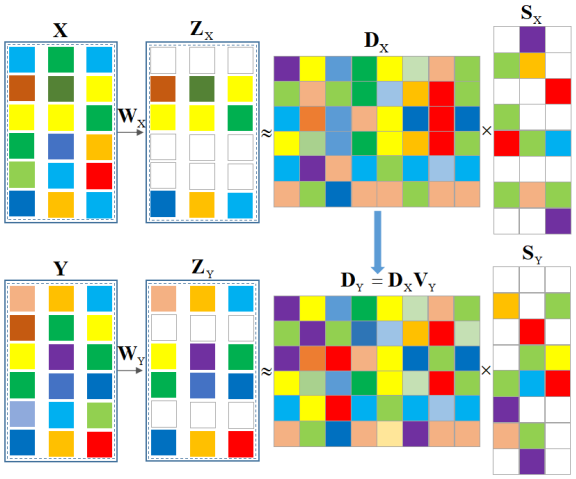
In order to balance the integrity of the domain structure and the relevance of different domains, this paper first carried out sparse dictionary learning on the auxiliary domain and target domain respectively. And then the dictionary of the target domain was restructured linearly with the auxiliary domain dictionary. This method can not only effectively reduce the feature dimension, but also keep the integrity of each domain and the correlations between different domains.

III. THE PROPOSED ALGORITHM

The proposed multiview representation framework based on transfer learning takes macro-expression as the auxiliary domain and micro-expression as the target domain. Firstly, we project these two domains into a common space, and then learn their sparse dictionary representations respectively. Secondly, the target domain dictionary is reconstructed using the auxiliary domain dictionary to obtain higher correlations between two domains. Thirdly, the multiview representation is introduced to select the most effective micro-expression features from different features sets by F norm. The detailed framework is shown in Figure 3.



(a) Projection of auxiliary domain and target domain



(b) Sparse dictionary representation based on transfer learning

FIGURE 3. The framework of the proposed algorithm.

A. FEATURE PROJECTION AND SPARSE DICTIONARY LEARNING

The main notations and their descriptions are given in Table 1. The feature sets of the auxiliary domain and the target domain are divided into the training set ($\mathbf{X}^{Tr}$, $\mathbf{Y}^{Tr}$) and the test set ($\mathbf{X}^{Te}$, $\mathbf{Y}^{Te}$). In order to simplify the writing, the subscripts of training sets $\mathbf{X}^{Tr}$ and $\mathbf{Y}^{Tr}$ are omitted. Thus, the feature sets can be rewritten as $\mathbf{X} = \left[\mathbf{x}_1, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_{n_x}\right] \in \mathbf{R}^{m_x \times n_x}$ and $\mathbf{Y} = \left[\mathbf{y}_1, \ldots \mathbf{y}_i, .., \mathbf{y}_{n_y}\right] \in \mathbf{R}^{m_y \times n_y}$, where $\mathbf{x}_i, \mathbf{y}_i$ are the features of a sample in the auxiliary domain and target domain, m_x, m_y denote the feature dimensions and n_x, n_y denote the sample numbers.

If the data of two domains are classified directly after learning the dictionary, it would be difficult to obtain an ideal recognition result because of the big differences between the two domains.

In order to address domain shift, the data from both auxiliary domain and target domain are mapped into a common space before sparse dictionary learning. As shown in Figure 3-(a) $\mathbf{Z}_X = \mathbf{W}_X^\top\mathbf{X}$, $\mathbf{Z}_Y = \mathbf{W}_Y^\top\mathbf{Y}$, $\mathbf{Z}_X \in \mathbf{R}^{d \times n_x}$ and $\mathbf{Z}_Y \in \mathbf{R}^{d \times n_y}$ are the data of the auxiliary domain and target domain after projection; $\mathbf{W}_X \in \mathbf{R}^{m_x \times d}$ and $\mathbf{W}_Y \in \mathbf{R}^{m_y \times d}$

TABLE 1. List of important mathematical notations.

Symbol	Description	Symbol	Description
$\mathbf{X}^{Tr}$	The training set of the auxiliary domain.	$\mathbf{W}_X$	The projection matrix of the auxiliary domain.
$\mathbf{X}^{Te}$	The test set of the auxiliary domain.	$\mathbf{W}_Y$	The projection matrix of the target domain.
$\mathbf{Y}^{Tr}$	The training set of the target domain.	$\mathbf{Z}_X$	The projection data of the auxiliary domain feature set.
$\mathbf{Y}^{Te}$	The test set of the target domain.	$\mathbf{Z}_Y$	The projection data of the target domain feature set.
$\mathbf{x}_i$	One sample from the auxiliary domain.	$\mathbf{D}_X$	The dictionary of the auxiliary domain.
$\mathbf{y}_i$	One sample from the target domain.	$\mathbf{D}_Y$	The dictionary of the target domain.
$\mathbf{S}_X$	The sparse coefficient matrix of the auxiliary domain.	$\mathbf{V}_Y$	The reconstruction matrix of the target domain dictionary.
$\mathbf{S}_Y$	The sparse coefficient matrix of the target domain.		

are the projection matrices of auxiliary domain and target domain, where d represents the dimension after two domains projected to the common space. Then the sparse dictionary learning is performed towards the data of the two domains. The basic form of loss function is as follows:

$$\begin{aligned} &\min \left\|\mathbf{W}_X^{\top}\mathbf{X}-\mathbf{D}_X\mathbf{S}_X\right\|_F^2+\left\|\mathbf{W}_Y^{\top}\mathbf{Y}-\mathbf{D}_Y\mathbf{S}_Y\right\|_F^2 \\ &\text{s.t. } \mathbf{W}_X^{\top}\mathbf{W}_X=\mathbf{I}_X,\quad \mathbf{W}_Y^{\top}\mathbf{W}_Y=\mathbf{I}_Y,\ \|\mathbf{d}_{xi}\|_2\le 1, \\ &\quad \left\|\mathbf{d}_{yj}\right\|_2\le 1,\quad \|\mathbf{S}_X\|_1\le\sigma,\ \|\mathbf{S}_Y\|_1\le\sigma \end{aligned} \tag{1}$$

where $\sigma = 0.001$, $\mathbf{D}_X \in \mathbf{R}^{d\times p_x}$ and $\mathbf{D}_Y \in \mathbf{R}^{d\times p_y}$ are the dictionaries of the auxiliary domain and target domain, and p_x, p_y denote the dictionary sizes; $\mathbf{d}_{xi} \in \mathbf{R}^d (i = 1, 2, \ldots, p_x)$ is the i-th column vector of $\mathbf{D}_X$ and $\mathbf{d}_{yj} \in \mathbf{R}^d (j = 1, 2, \ldots, p_y)$ is the j-th column vector of $\mathbf{D}_Y$. $\mathbf{S}_X \in \mathbf{R}^{p_x\times n_x}$ and $\mathbf{S}_Y \in \mathbf{R}^{p_y\times n_y}$ are the sparse coefficient matrices of the auxiliary domain and target domain. $\mathbf{I}_X \in \mathbf{R}^{d\times d}$ and $\mathbf{I}_Y \in \mathbf{R}^{d\times d}$ are the unit matrices of the two domains.

B. SPARSE DICTIONARY RECONSTRUCTION

In order to make the projected features of the two domains in a similar distribution, the target domain dictionary is linearly reconstructed by the auxiliary domain dictionary based on the sparse dictionary learning of each domain, the specific formula is as follows:

$$\mathbf{D}_Y = \mathbf{D}_X\mathbf{V}_Y \tag{2}$$

where $\mathbf{D}_X \in \mathbf{R}^{d\times p_x}$, $\mathbf{D}_Y \in \mathbf{R}^{d\times p_y}$, and $\mathbf{V}_Y \in \mathbf{R}^{p_x\times p_y}$ is the reconstruction matrix.

After domain projection, sparse dictionary learning and dictionary reconstruction, the objective function J can be expressed as follows:

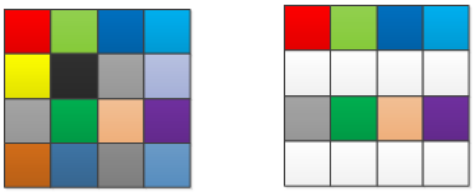
$$\begin{aligned} J = \min_{\mathbf{W}_X,\mathbf{W}_Y,\mathbf{D}_X,\mathbf{D}_Y,\mathbf{V}_Y,\mathbf{S}_X,S_Y} &\left\|\mathbf{W}_X{}^{\top}\mathbf{X}-\mathbf{D}_X\mathbf{S}_X\right\|_F^2 \\ &+\left\|\mathbf{W}_Y{}^{\top}\mathbf{Y}-\mathbf{D}_Y\mathbf{S}_Y\right\|_F^2+\|\mathbf{D}_Y-\mathbf{D}_X\mathbf{V}_Y\|_F^2 \\ \text{s.t. } \mathbf{W}_X^{\top}\mathbf{W}_X=\mathbf{I}_X,&\quad \mathbf{W}_Y^{\top}\mathbf{W}_Y=\mathbf{I}_Y, \\ \|\mathbf{d}_{xi}\|_2\le 1,&\quad \left\|\mathbf{d}_{yj}\right\|_2\le 1,\quad \|\mathbf{V}_Y\|_1\le\tau,\ \|\mathbf{S}_X\|_1\le\sigma, \\ \|\mathbf{S}_Y\|_1\le\sigma& \end{aligned} \tag{3}$$

where τ=0.001 and $\sigma = 0.001$. The first two terms of J guarantee the sparse dictionary learning based on domain projection. And the last term uses the linear representation of the auxiliary domain dictionary to reconstruct the target domain dictionary, which further reduces the differences between the two domains.

C. FEATURE SELECTION BASED ON MULTIVIEW REPRESENTATION

Different feature sets can be obtained with different extracting methods. If all feature sets are simply integrated together, the information redundancy would cause much computing cost and a poor recognition result. Therefore, it is critical to select the most effective features.

In order to fully exploit the features of target domain, the idea of multiview representation is introduced to extract the most effective micro-expression features from different feature sets. The sparse feature selection is carried out by F norm sparse.



(a) Multiple sets of dense features (b) Sparse feature selection

FIGURE 4. Sparse feature selection with *F* norm.

The traditional method simply concatenates the different feature sets, as shown in Figure 4-(a), where each column represents a sample and different rows represent different features. This simple feature connection will cause information redundancy, high calculation and unsatisfactory recognition results.

In this paper, we use F norm sparse to perform feature selection. Redefine the feature sets of target domain $\mathbf{Y} = \left[\mathbf{Y}^1; \ldots; \mathbf{Y}^G; \ldots; \mathbf{Y}^N\right] \in \mathbf{R}^{M \times n_y}$, where $\mathbf{Y}^G$ represents the G-th feature set of the target domain. $M = m_y \times N$, m_y and n_y denote the dimension of the feature sets and the sample number of the target domain respectively. N denotes the number of the feature descriptors, that is, N different feature sets are extracted. Therefore, the projection matrix of Y can be rewritten as $\mathbf{W}_Y = \left[\mathbf{W}_Y^1, \ldots \mathbf{W}_Y^G, \ldots, \mathbf{W}_Y^N\right] \in \mathbf{R}^{M \times d}$, where $\mathbf{W}_Y^G \in \mathbf{R}^{m_y \times d}$ is the G-th part of $\mathbf{W}_Y$, corresponding to $\mathbf{Y}^G$.

Then the feature projection of the target domain is modified as $\sum_{G=1}^{N} (\mathbf{W}_Y^G)^\top \mathbf{Y}^G$, and the restriction of F norm is introduced by $\min \sum_{G=1}^{N} \left\| W_Y^G \right\|_F^2$. In order to minimize the overall $\mathbf{W}_Y$, we minimize each $\mathbf{W}_Y^G$ of $\mathbf{W}_Y$ according to $\|\mathbf{W}_Y\|_F = \sqrt{\sum_{G=1}^{N} \left\| \mathbf{W}_Y^G \right\|_F^2}$. When $\mathbf{W}_Y^G$ is small enough and is approximated to zero, the corresponding $\mathbf{Y}^G$ will be removed. Then the feature selection is realized and the result is shown in Figure 4-(b), where the color block represents the selected feature and the white ones are the eliminated features. The projection for the target domain should satisfy the following equation:

$$\min \left\| \sum_{G=1}^{N} (\mathbf{W}_Y^G)^\top \mathbf{Y}^G - \mathbf{D}_Y \mathbf{S}_Y \right\|_F^2 \tag{4}$$

where $\sum_{G=1}^{N} (\mathbf{W}_Y^G)^\top \mathbf{Y}^G = \mathbf{W}_Y^\top \mathbf{Y}$. To obtain the same form as Equation (1), the above formula can be simplified as $\min \left\| \mathbf{W}_Y^\top \mathbf{Y} - \mathbf{D}_Y \mathbf{S}_Y \right\|_F^2$. Finally, the objective function can be defined as follows:

$$\begin{aligned} J = \min_{\mathbf{W}_X, \mathbf{W}_Y, \mathbf{D}_X, \mathbf{D}_Y, \mathbf{V}_Y, \mathbf{S}_X, S_Y} & \left\| \mathbf{W}_X^\top \mathbf{X} - \mathbf{D}_X \mathbf{S}_X \right\|_F^2 \\ & + \left\| \mathbf{W}_Y^\top \mathbf{Y} - \mathbf{D}_Y \mathbf{S}_Y \right\|_F^2 + \|\mathbf{D}_Y - \mathbf{D}_X \mathbf{V}_Y\|_F^2 \\ & + \lambda \sum_{G=1}^{N} \left\| \mathbf{W}_Y^G \right\|_F^2 \\ \text{s.t. } & \mathbf{W}_X^\top \mathbf{W}_X = \mathbf{I}_X, \quad \mathbf{W}_Y^\top \mathbf{W}_Y = \mathbf{I}_Y, \\ & \|\mathbf{d}_{xi}\|_2 \le 1, \\ & \left\| \mathbf{d}_{yj} \right\|_2 \le 1, \quad \|\mathbf{V}_Y\|_1 \le \tau, \ \|\mathbf{S}_X\|_1 \le \sigma, \\ & \|\mathbf{S}_Y\|_1 \le \sigma \end{aligned} \tag{5}$$

where τ=0.001, $\sigma = 0.001$, and λ is a penalty coefficient on the projection matrix $\mathbf{W}_Y$.

IV. OPTIMAL SOLUTION

The objective function Eq.(5) can be solved by variable alternating optimization method. The derivatives of the objective function with respect to each variable are given as follows.

A. UPDATE $\mathbf{V}_Y$, GIVEN $\mathbf{D}_X, \mathbf{D}_Y$

For $\mathbf{V}_Y$, when $\mathbf{W}_X, \mathbf{W}_Y, \mathbf{D}_X, \mathbf{D}_Y, \mathbf{S}_X, \mathbf{S}_Y$ are fixed, the relevant part of $\mathbf{V}_Y$ in objective function is $\|\mathbf{D}_Y - \mathbf{D}_X \mathbf{V}_Y\|_F^2$. Combined with the constraint term $\|\mathbf{V}_Y\|_1 \le \tau$, the iterated function can be expressed as:

$$\min_{\mathbf{V}_Y} \|\mathbf{D}_Y - \mathbf{D}_X \mathbf{V}_Y\|_F^2 + \beta \|\mathbf{V}_Y\|_1 \tag{6}$$

where β is the penalty coefficient of $\mathbf{V}_Y$. Equation (6) is in accordance with lasso's objective function and lasso method [31] is adopted to update $\mathbf{V}_Y$ in this paper.

B. UPDATE $\mathbf{D}_X(\mathbf{D}_Y)$, GIVEN $\mathbf{W}_X(\mathbf{W}_Y), \mathbf{D}_Y(\mathbf{D}_X), \mathbf{S}_X(\mathbf{S}_Y), \mathbf{V}_Y$

For $\mathbf{D}_X$, when $\mathbf{D}_Y, \mathbf{W}_X, \mathbf{W}_Y, \mathbf{V}_Y, \mathbf{S}_X, \mathbf{S}_Y$ are fixed, the objective function of $\mathbf{D}_X$ becomes a convex function. So it is possible to derive $\mathbf{D}_X$ directly and let the derivative be 0. Then we can obtain the iterative expression of $\mathbf{D}_X$. The iterative expression of $\mathbf{D}_Y$ can be obtained in a similar way. The derivatives toward $\mathbf{D}_X$ and $\mathbf{D}_Y$ are as follows:

$$\begin{aligned} \frac{\partial J}{\partial \mathbf{D}_X} &= \frac{\partial \left[\begin{array}{l} tr\left(\left(\mathbf{W}_X^\top \mathbf{X} - \mathbf{D}_X \mathbf{S}_X \right)^\top \left(\mathbf{W}_X^\top \mathbf{X} - \mathbf{D}_X \mathbf{S}_X \right) \right) + \\ tr\left(\left(\mathbf{D}_Y - \mathbf{D}_X \mathbf{V}_Y \right)^\top \left(\mathbf{D}_Y - \mathbf{D}_X \mathbf{V}_Y \right) \right) \end{array} \right]}{\partial \mathbf{D}_X} \\ &= -2\mathbf{W}_X^\top \mathbf{X} \mathbf{S}_X^\top + 2\mathbf{D}_X \mathbf{S}_X \mathbf{S}_X^\top - 2\mathbf{D}_Y \mathbf{V}_Y^\top + 2\mathbf{D}_X \mathbf{V}_Y \mathbf{V}_Y^\top \end{aligned} \tag{7}$$

$$\begin{aligned} \frac{\partial J}{\partial \mathbf{D}_Y} &= \frac{\partial \left[\begin{array}{l} tr\left(\left(\mathbf{W}_Y^\top \mathbf{Y} - \mathbf{D}_Y \mathbf{S}_Y \right)^\top \left(\mathbf{W}_Y^\top \mathbf{Y} - \mathbf{D}_Y \mathbf{S}_Y \right) \right) + \\ tr\left(\left(\mathbf{D}_Y - \mathbf{D}_X \mathbf{V}_Y \right)^\top \left(\mathbf{D}_Y - \mathbf{D}_X \mathbf{V}_Y \right) \right) \end{array} \right]}{\partial \mathbf{D}_Y} \\ &= -2\mathbf{W}_Y^\top \mathbf{Y} \mathbf{S}_Y^\top + 2\mathbf{D}_Y \mathbf{S}_Y \mathbf{S}_Y^\top + 2\mathbf{D}_Y - 2\mathbf{D}_X \mathbf{V}_Y \end{aligned} \tag{8}$$

C. UPDATE $\mathbf{W}_X(\mathbf{W}_Y)$, GIVEN $\mathbf{D}_X(\mathbf{D}_Y), \mathbf{S}_X(\mathbf{S}_Y)$

For $\mathbf{W}_X$, when $\mathbf{W}_Y, \mathbf{D}_X, \mathbf{D}_Y, \mathbf{V}_Y, \mathbf{S}_X, \mathbf{S}_Y$ are fixed, the objective function of $\mathbf{W}_X$ is a convex function. However, in order to obtain better iterative result, the constraint term of $\mathbf{W}_X$ is given as follows: $\mathbf{W}_X^\top \mathbf{W}_X = \mathbf{I}_X$, so it cannot be directly derived. Thus, we used a generalized gradient descent method on the Grassman manifold to achieve the optimization [32], [33]. $\mathbf{W}_Y$ can be operated in a similar approach. After introducing multiview representation, the derivatives toward $\mathbf{W}_X$ and $\mathbf{W}_Y$ in the objective function are slightly modified, the specific form is as follows:

$$\begin{aligned} \frac{\partial J}{\partial \mathbf{W}_X} &= \frac{\partial \left[tr\left(\left(\mathbf{W}_X^\top \mathbf{X} - \mathbf{D}_X \mathbf{S}_X \right)^\top \left(\mathbf{W}_X^\top \mathbf{X} - \mathbf{D}_X \mathbf{S}_X \right) \right) \right]}{\partial \mathbf{W}_X} \\ &= 2\mathbf{X}\mathbf{X}^\top \mathbf{W}_X - 2\mathbf{X}\mathbf{S}_X^\top \mathbf{D}_X^\top \end{aligned} \tag{9}$$

$$\frac{\partial J}{\partial \mathbf{W}_Y} = \frac{\partial \begin{bmatrix} tr\left(\left(\mathbf{W}_Y^\top \mathbf{Y} - \mathbf{D}_Y \mathbf{S}_Y\right)^\top \left(\mathbf{W}_Y^\top \mathbf{Y} - \mathbf{D}_Y \mathbf{S}_Y\right)\right) + \\ tr\left(\lambda \sum_{G=1}^{N} \left(\mathbf{W}_Y^G\right)^\top \mathbf{W}_Y^G\right) \end{bmatrix}}{\partial \mathbf{W}_Y}$$
$$= 2\mathbf{Y}\mathbf{Y}^\top \mathbf{W}_Y - 2\mathbf{Y}\mathbf{S}_Y^\top \mathbf{D}_Y^\top + 2\lambda \mathbf{W}_Y \mathbf{C} \quad (10)$$

where λ is the same penalty coefficient as in Eq.(5), $\mathbf{C} = diag(\mathbf{C}^1, \ldots, \mathbf{C}^G)$, each $\mathbf{C}^g (g = 1, \ldots, G)$ is a diagonal matrix, and the j-th diagonal element of C^g is:

$$C_{j,j} = \begin{cases} 0, & if \ \left\|\mathbf{W}_Y^G\right\|_F = 0 \\ \frac{1}{2\left\|\mathbf{W}_Y^G\right\|_F}, & otherwise \end{cases} \quad (11)$$

The algorithm can be summarized in Algorithm 1.

Algorithm 1 A Multiview Representation Framework for Micro-Expression Recognition

Input:

Training Set:

The feature set of the auxiliary domain : $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_n] \in \mathbf{R}^{m_x \times n}$;

The feature set of the target domain : $\mathbf{Y} = [\mathbf{Y}_1; \ldots; \mathbf{Y}_G; \ldots; \mathbf{Y}_n] \in \mathbf{R}^{M \times n}$.

Test Set:

The feature set of the auxiliary domain : $\mathbf{X}_{Te} = [\mathbf{x}_1, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_n] \in \mathbf{R}^{m_x \times n}$;

The feature set of the target domain : $\mathbf{Y}_{Te} = [\mathbf{Y}_1; \ldots; \mathbf{Y}_G; \ldots; \mathbf{Y}_n] \in \mathbf{R}^{M \times n}$.

Output:

The Sparse coefficient matrices of the auxiliary domain and target domain : $\mathbf{S}_{X,Te}, \mathbf{S}_{Y,Te}$.

Algorithm steps:

1: Initialize parameters: $\mathbf{D}_X, \mathbf{D}_Y, \mathbf{W}_X, \mathbf{W}_Y, \mathbf{S}_X, \mathbf{S}_Y, \mathbf{V}_Y, T$.
2: For $iter = 1 : T$
3: Get the derivative of $\mathbf{D}_X$ as Eq.(7) with the values of $\mathbf{W}_X$, $\mathbf{D}_Y$, $\mathbf{S}_X$, $\mathbf{V}_Y$ being fixed and get the derivative of $\mathbf{D}_Y$ as Eq.(8) with the values of $\mathbf{W}_Y$, $\mathbf{D}_X$, $\mathbf{S}_Y$, $\mathbf{V}_Y$ being fixed.
4: Let Eqs. (7) and (8) equal zero to update $\mathbf{D}_X$ and $\mathbf{D}_Y$.
5: Get the derivative of $\mathbf{W}_X$ as Eq.(9) with the values of $\mathbf{D}_X, \mathbf{S}_X$ being fixed and get the derivative of $\mathbf{W}_Y$ as Eq.(10) with the values of $\mathbf{D}_Y, \mathbf{S}_Y$ being fixed.
6: Let Eqs. (9) and (10) equal zero to update $\mathbf{W}_X$ and $\mathbf{W}_Y$.
7: Use the orthogonal matching pursuit (OMP) algorithm to update the sparse coefficient matrices $\mathbf{S}_X, \mathbf{S}_Y$.
8: Use the lasso algorithm to update the reconstruction matrix $\mathbf{V}_Y$.
9: $iter = iter + 1$
10: The final $\mathbf{D}_X, \mathbf{D}_Y, \mathbf{W}_X, \mathbf{W}_Y, \mathbf{S}_Y, \mathbf{S}_X, \mathbf{V}_Y$ are obtained by the above iteration process. We use the obtained projection matrices $\mathbf{W}_X, \mathbf{W}_Y$ to project the test data: $\mathbf{Z}_{X,Te} = \mathbf{W}_X^\top \mathbf{X}_{Te}$; $\mathbf{Z}_{Y,Te} = \mathbf{W}_Y^\top \mathbf{Y}_{Te}$.
11: Obtain sparse coefficient matrices $\mathbf{S}_{X,Te}, \mathbf{S}_{Y,Te}$ according to the dictionaries $\mathbf{D}_X, \mathbf{D}_Y$.

V. EXPERIMENTS

In this section, we first introduce three expression databases and the experimental settings. Subsequently, the experiments of the multiview representation's performance validation, the influence of dictionary size on recognition rate and the comparison between the proposed algorithm and other advanced algorithms are given.

A. DATABESE:CK+ & CASME & CASME II

The Extended Cohn-Kanade Dataset (CK+) [34] released in 2010 is an extension of the Cohn-Kanade Dataset. This database contains 123 subjects and 593 image sequences. The last frame of each image sequence has an action unit label, and of the 593 image sequences, 327 sequences have emotion labels. CK+ database is commonly used for facial expression recognition and it divides human emotions into eight categories. CASME II [35] is an extension of the Chinese Academy of Sciences Micro-Expression (CASME). It consists of 26 subjects and 255 micro-expression segments with seven categories. In this paper, the CK+ database was selected as the macro-expression database, CASME and CASME II were selected as the micro-expression databases. Since the three databases' categories are not consistent, to keep the data of different databases in similar distribution, we selected three types of samples with a comparative large number of samples from these three databases, namely, happiness, sadness, and surprise. A total of 60 samples were selected for testing.

B. EXPERIMENTAL SETTINGS

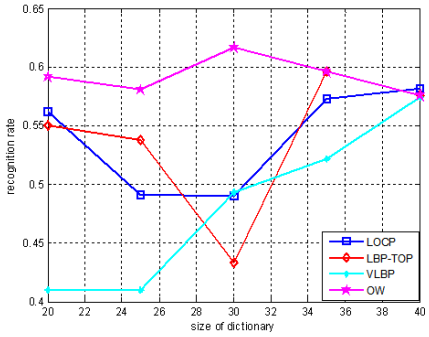
Considering the numbers of corresponding samples in these three databases are small, all experiments in this paper used a leave-one verification method and the nearest neighbor algorithm was used as the classifier of the micro-expression recognition. Since the KSVD toolbox [33] was used to obtain the sparse coefficient matrices, the experimental results varied each time. Thus, each test was repeated 20 times and the average was used as the final result.

The experiments were carried out on two sets of databases. The first set of databases includes the macro-expression database CK+ and the micro-expression database CASME. The second set of databases consists of the macro-expression database CK+ and the micro-expression database CASME II.

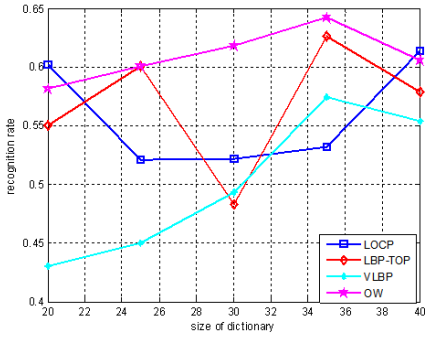
In this paper, LBP feature of macro-expression was extracted, LBP-TOP feature [36], LOCP-TOP feature [36], VLBP feature [37] and optical flow (OW) [20] of micro-expression were extracted.

C. RECOGNITION OF FOUR DIFFERENT SINGLE FEATURE SETS

In this group of experiments, the basic performance of the proposed algorithm was evaluated on two sets of



(a) CK+&CASME



(b) CK+&CASME II

FIGURE 5. Recognition results of four different single feature sets on the two sets of databases.

different databases. The recognition results of using four different single micro-expression feature sets were compared on each database set. For the different feature sets, the dictionary size was setted to range from 20 to 40. From Figure 5 we can draw the following conclusions:

a) The recognition rate can be up to 61.7% in the CK+&CASME database; and in the CK+&CASME II database, the recognition rate can reach 64.2%.

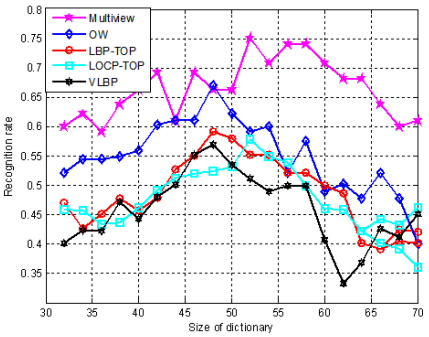
b) The change of dictionary size has a significant impact on the micro-expression recognition and the sparse effect varies with different dictionary sizes. If the dictionary is too large, it will result in excessive sparseness of features, which may have a detrimental effect on the final recognition rate.

c) Optical flow (OW) has the best and relatively stable performance of four different feature sets on the two sets of databases. This is because optical flow is a normalized feature based on the region of interest and in the process of optical flow feature extraction, the problem of human face alignment can be better solved. Due to the number of VLBP features is small and the description of facial motion mechanism is insufficient, so VLBP is less effective than OW.

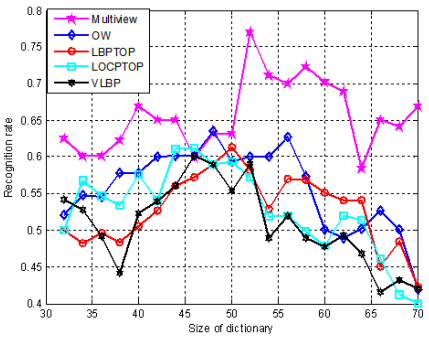
d) Since the CASME II is an extension of CASME, the experiment results of CK+&CASME II database are better than CK+&CASME database. Compared with CASME, CASME II has a larger number of samples, and a greatly increased video frame rate (60fps to 200fps). Besides, the pixel resolution has been increased from 150 × 190 to 280 × 340. These enhancements are beneficial to improve the detailed expression of the micro-expression, which can help increase the final recognition rate.

D. RECOGNITION OF MULTIVIEW REPRESENTATION

Based on four different feature sets, the most effective features were selected with the idea of multiview representation. We validated the performance of the proposed multiview representation algorithm based on transfer learning and examined multiview representation's ability in utilizing different features comprehensively. Similar to Section V-C, experiments were performed on the two sets of databases. The multiview representation's result was compared with that using the four single feature sets respectively in Figure 6.



(a) CK+&CASME



(b) CK+&CASME II

FIGURE 6. Comparison between multiview representation and four single feature sets.

In order to further validate the effectiveness of multiview representation, we connected the four different features sets in series, and merged them into a high-dimensional feature group without feature selection. For different dictionary sizes, multiple experiments were conducted with the features directly connected in series. Then its optimal results were compared with the optimal results of multiview representation in Table 2. After analyzing Figure 6 and Table 2, we conclude that:

a) The advantages of comprehensive utilization of feature information are demonstrated with multiview representation. As is shown in Table 2, the highest recognition rate is 77.0%

TABLE 2. Comparison of recognition rates between multiview representation and feature series.

Feature	Multiview		Feature Series	
Database	CASME	CASMEII	CASME	CASMEII
Recognition Rate	75.0%	77.0%	57.4%	59.6%

on CK+&CASME II and the average performance of multiview representation is better than that of a single feature regardless of dictionary size as shown in Figure 6.

b) As can be seen in Figure 6, the recognition rate of multiview representation is higher than that of the OW feature. For multiview representation, the F norm sparse is not simply selecting a feature or removing the remaining features completely. It is a integration way to the feature information. It also shows that a single kind of features can not adequately represent expressions.

c) Table 2 also shows that simple feature series cannot effectively improve the recognition results. Stacking different feature sets simply without feature selection will cause a large amount of redundant information and noise, which will annihilate the effective features. Besides, the computational costs are greatly reduced in multiview representation compared with simple feature series.

E. EXPERIMENTS ON THE EFFECT OF DICTIONARY SIZE

It can be seen from the above two experiments that the recognition results varies significantly when the dictionaries are different. The recognition results cannot be improved by simply increasing or decreasing the size of dictionary. Based on this fact, this experiment explores the response of different samples when the dictionary size is not fixed.

TABLE 3. Recognition results of the two database sets corresponding to the optimal dictionaries.

Database	CASME	CASME II
Recognition Rate	94%	92%

The experimental results in Table 3 show that for non-fixed dictionary, the sizes of dictionaries chosen for different samples is inconsistent and the experimental results are much better than the previous results. However, it is worth of exploring how to make the samples autonomously select the optimal dictionaries in the feature training stage, which is a follow-up problem.

F. COMPARISON WITH OTHER MICRO-EXPRESSION RECOGNITION ALGORITHMS

In this section, we have compared our proposed algorithm with some recently proposed micro-expression recognition algorithms. The algorithms are JFSSL [38], LBP-TOP [36], DTSA [38] and FDM [20]. The experimental results are shown in Table 4.

TABLE 4. Comparison between our algorithm and other several advanced algorithms.

Method		JFSSL [38]	LBP-TOP [36]	DTSA [38]	FDM [20]	Ours
Recognition Rate	CK+& CASME	67.3%	62.9%	52.4%	61.5%	74.6%
	CK+& CASME II	65.7%	59.6%	54.1%	68.5%	78.2%

Based on the comprehensive analysis of the experimental results in Table 4, we have drawn the following two main conclusions:

a) JFSSL and our algorithm perform better than the LBP-TOP, DTSA, and FDM algorithms. The recognition rates of JFSSL transfer algorithm on CK+&CASME and CK+&CASME II can reach 67.3% and 65.7% respectively. And our proposed algorithm has the highest recognition rates, which are 74.6% and 78.2% on CK+&CASME and CK+&CASME II respectively. The reason is that the transfer learning algorithms can make full use of the rich information of macro-expressions, and the knowledge theory is transferred to the domain of micro-expression.

b) JFSSL has better performance than other methods, because JFSSL makes good use of the sample labels. It projects the feature sets of different domains directly to the label space, which has superiority over unsupervised learning. Our proposed method performs better than JFSSL, because multiview representation has screened out multiple sets of features to avoid the interference of noise and combined the advantages of different feature sets.

VI. CONCLUSION

In this paper, we proposed a multiview representation framework based on transfer learning for micro-expression recognition. The data of the auxiliary domain and the target domain are projected into a common space, and the dictionaries are learned in each domain respectively. In order to increase the data correlations in two different domains, the dictionary of the target domain is linearly represented by the dictionary in the auxiliary domain. In order to fully exploit the features of micro-expressions, multiple feature sets are extracted from micro-expressions. The optimal features are selected from multiple feature sets according to the multiview representation, which effectively improves the recognition accuracy. For future work, more effective models should be explored, and the autonomous selection of optimal dictionary can be taken into consideration.

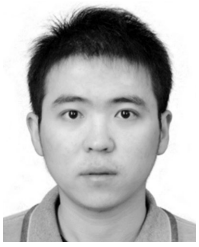
REFERENCES

[1] E. A. Haggard and K. S. Isaacs, "Micromomentary facial expressions as indicators of ego mechanisms in psychotherapy," in *Methods of Research in Psychotherapy* (The Century Psychology). Boston, MA, USA: Springer, 1966.

[2] J. Endres and A. Laidlaw, "Micro-expression recognition training in medical students: A pilot study," *BMC Med. Edu.*, vol. 9, no. 1, p. 47, 2009.

[3] X. Ben, M. Yang, P. Zhang, and J. Li, "Survey on automatic micro expression recognition methods," *J. Comput.-Aided Design Comput. Graph.*, vol. 26, no. 9, pp. 1385–1395, 2014.
[4] C. Gong, D. Tao, S. J. Maybank, W. Liu, G. Kang, and J. Yang, "Multi-modal curriculum learning for semi-supervised image classification," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3249–3260, Jul. 2016.
[5] W. Yan, X. Li, S.-J. Wang, G. Zhao, Y.-J. Liu, Y.-H. Chen, and X. Fu, "CASME II: An improved spontaneous micro-expression database and the baseline evaluation," *PLoS ONE*, vol. 9, no. 1, p. e86041, 2014.
[6] T. Pfister, X. Li, G. Zhao, and M. Pietikäinen, "Recognising spontaneous facial micro-expressions," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 1449–1456.
[7] Z. Xia, X. Feng, J. Peng, X. Peng, and G. Zhao, "Spontaneous micro-expression spotting via geometric deformation modeling," *Comput. Vis. Image Understand.*, vol. 147, pp. 87–94, Jun. 2016.
[8] X. Huang, S. Wang, X. Liu, G. Zhao, X. Feng, and M. Pietikäinen, "Discriminative spatiotemporal local binary pattern with revisited integral projection for spontaneous facial micro-expression recognition," *IEEE Trans. Affective Comput.*, vol. 10, no. 1, pp. 32–47, Jan. 2019.
[9] X. Ben, X. Jia, R. Yan, X. Zhang, and W. Meng, "Learning effective binary descriptors for micro-expression recognition transferred by macro-information," *Pattern Recognit. Lett.*, vol. 107, pp. 50–58, May 2018.
[10] M. Jian and K.-M. Lam, "Simultaneous hallucination and recognition of low-resolution faces based on singular value decomposition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 11, pp. 1761–1772, Nov. 2015.
[11] M. Jian, K.-M. Lam, and J. Dong, "A novel face-hallucination scheme based on singular value decomposition," *Pattern Recognit.*, vol. 46, no. 11, pp. 3091–3102, 2013.
[12] C. Gong, T. Liu, J. Yang, and D. Tao, "Large-margin label-calibrated support vector machines for positive and unlabeled learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, to be published.
[13] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
[14] Y.-R. Yeh, C.-H. Huang, and Y.-C. F. Wang, "Heterogeneous domain adaptation and classification by exploiting the correlation subspace," *IEEE Trans. Image Process.*, vol. 23, no. 5, pp. 2009–2018, May 2014.
[15] Y. Yang, Y. Yang, Z. Huang, and H. Shen, "Effective transfer tagging from image to video," *Trans. Multimedia Comput. Commun. Appl.*, vol. 9, no. 2, pp. 1–20, 2013.
[16] S. Polikovsky, Y. Kameda, and Y. Ohta, "Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor," in *Proc. Int. Conf. Crime Detection Prevention*, 2010, pp. 1–6.
[17] W. Zhang, W. Zhang, K. Liu, and J. Gu, "A feature descriptor based on local normalized difference for real-world texture classification," *IEEE Trans. Multimedia*, vol. 20, no. 4, pp. 880–888, Apr. 2018.
[18] W. Qi, X. Shen, and X. Fu, "The machine knows what you are hiding: An automatic micro-expression recognition system," in *Proc. Int. Conf. Affective Comput. Intell. Interact.*, 2011, pp. 152–162.
[19] F. Xu, J. Zhang, and J. Z. Wang, "Microexpression identification and categorization using a facial dynamics map," *IEEE Trans. Affect. Comput.*, vol. 8, no. 2, pp. 254–267, Apr. 2017.
[20] Y.-J. Liu, J.-K. Zhang, W.-J. Yan, S.-J. Wang, G. Zhao, and X. Fu, "A main directional mean optical flow feature for spontaneous micro-expression recognition," *IEEE Trans. Affect. Comput.*, vol. 7, no. 4, pp. 299–310, Oct. 2016.
[21] S. Sun, "A survey of multi-view machine learning," *Neural Comput. Appl.*, vol. 23, nos. 7–8, pp. 2031–2038, 2013.
[22] X. Zhu, X. Li, and S. Zhang, "Block-row sparse multiview multilabel learning for image Classification," *IEEE Trans. Cybern.*, vol. 46, no. 2, pp. 450–461, Feb. 2016.
[23] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Ann. Statist.*, vol. 32, no. 2, pp. 407–499, 2004.
[24] X. Zhu, H. Zi, C. Hong, J. Cui, and H. Shen, "Sparse hashing for fast multimedia search," *ACM Trans. Inf. Syst.*, vol. 31, no. 2, pp. 1–24, 2013.
[25] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Roy. Statist. Soc. B, Stat. Methodol.*, vol. 68, no. 1, pp. 49–67, 2006.
[26] F. Nie, H. Huang, C. Xiao, and C. H. Q. Ding, "Efficient and robust feature selection via joint $\ell_{2,1}$-norms minimization," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2010, pp. 1813–1821.
[27] S. Shekhar, V. M. Patel, H. Van Nguyen, and R. Chellappa, "Coupled projections for adaptation of dictionaries," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 2941–2954, Oct. 2015.
[28] C. X. Ren, D. Q. Dai, K. K. Huang, and Z. R. Lai, "Transfer learning of structured representation for face recognition," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5440–5454, Dec. 2014.
[29] J. Masci, M. M. Bronstein, A. M. Bronstein, and J. Schmidhuber, "Multimodal similarity-preserving hashing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 4, pp. 824–830, Apr. 2014.
[30] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Fisher discrimination dictionary learning for sparse representation," in *Proc. IEEE Int. Conf. Comput. Vis., (ICCV)*, Barcelona, Spain, Nov. 2011.
[31] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Stat. Soc. B, Stat. Methodol.*, vol. 58, no. 1, pp. 267–288, 1996.
[32] J. Hu, W. Zheng, J. Lai, and J. and Zhang, "Jointly learning heterogeneous features for RGB-D activity recognition," in *Proc. Int. Comput. Vis. Pattern Recognit.*, 2015, pp. 5344–5352.
[33] M. Shreve, S. Godavarthy, V. Manohar, D. Goldgof, and S. Sarkar, "Towards macro- and micro-expression spotting in video using strain patterns," in *Proc. Appl. Comput. Vis.*, 2009, pp. 1–6.
[34] 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. Comput. Vis. Pattern Recognit. Workshops*, 2010, pp. 1–6.
[35] S. Wang, W. Yan, X. Li, G. Zhao, C. Zhou, X. Fu, M. Yang, and J. Tao, "Micro-expression recognition using color spaces," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 6034–6047, Dec. 2015.
[36] C. H. Chan, B. Goswami, J. Kittler, and W. Christmas, "Local ordinal contrast pattern histograms for spatiotemporal, lip-based speaker authentication," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 2, pp. 602–612, Apr. 2012.
[37] G. Zhao and M. Pietikäinen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 915–928, Jun. 2007.
[38] K. Wang, R. He, L. Wang, W. Wang, and T. Tan, "Joint feature selection and subspace learning for cross-modal retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 10, pp. 2010–2023, Oct. 2016.

TIANHUAN HUANG received the B.E. degree in electronic information engineering from the School of Physical Science and Technology, Nanjing Normal University, Nanjing, China, in 2018. She is currently pursuing the M.S. degree with the School of Information Science and Engineering, Shandong University, Qingdao, China. Her current research interests include machine learning, image processing, and transfer learning.



LEI CHEN received the B.E. and M.E. degrees from Shandong University, China, in 2010 and 2013, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Ottawa, Canada, in 2018. His research interests include image processing and computer vision, image and video watermarking, and visual quality assessment.

YUNCONG FENG received the B.E. degree in electronic science and technology from the Harbin Institute of Technology, Weihai, China, in 2015, and the M.E. degree in integrated circuit engineering from Shandong University, Jinan, China, in 2018. His research interests include machine learning, image processing, and deep learning.

XIANYE BEN received the Ph.D. degree in pattern recognition and intelligent system from the College of Automation, Harbin Engineering University, Harbin, in 2010. She is currently an Associate Professor with the School of Information Science and Engineering, Shandong University, Qingdao, China. She has published more than 80 papers in major journals and conferences, such as IEEE T-IP, IEEE T-CSVT, PR, and so on. Her current research interests include pattern recognition, digital image processing and analysis, and machine learning. She received the Excellent Doctoral Dissertation awarded by Harbin Engineering University. She was also enrolled by the Young Scholars Program of Shandong University.

RUIXUE XIAO received the B.E. degree in electronic information science and technology from Shandong Agricultural University, Tai'an, China, in 2017. She is currently pursuing the M.S. degree with the School of Information Science and Engineering, Shandong University, Qingdao, China. Her current research interests include machine learning, transfer learning, and image processing.

TIANLE XUE is currently pursuing the B.E. degree with the School of Information Science and Engineering, Shandong University, Qingdao, China. Her current research interests include machine learning, data mining, and micro-expression recognition.