IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 26, 2019, accepted July 10, 2019, date of publication July 22, 2019, date of current version August 9, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2930288

Supervised Dictionary Learning With Regularization for Near-Infrared Spectroscopy Classification

LINGQIAO LI[1,2], XIPENG PAN[1], HUIHUA YANG[1,2], TAO ZHANG[3], AND ZHENBING LIU[2]

[1]School of Automation, Beijing University of Posts and Telecommunications, Beijing 100876, China
[2]School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin 541004, China
[3]Department of Computer Science and Technology, Jiangnan University, Wuxi 214122, China

Corresponding author: Huihua Yang (yhh@bupt.edu.cn)

This work was supported in part by the NSFC under Grant 61702226, Grant 61562013, and Grant 61672263, in part by the Key Research and Development Program of Guangxi Province under Grant AB16380293, in part by the Natural Science Foundation of Guangxi Province under Grant 2017GXNSFDA198025, and in part by the Natural Science Foundation of Jiangsu Province under Grant BK20170200.

ABSTRACT Near-infrared spectroscopy (NIRS) has been widely used in many fields due to its advantages with fast analysis speed, non-destructive testing, and on-site detection. However, NIRS has some shortcomings, such as low signal-to-noise ratio, weak absorption intensity, and overlapping peaks. The research of near-infrared spectral modeling method becomes the core of analyzing NIRS. In order to improve the accuracy of prediction model for NIRS, this paper proposes a novel sparse classification mechanism by designing appropriate regularization factors. The existing supervised dictionary learning methods have been proposed for classification aim and increasing its accuracy, the proposed method addresses some defects existing in this area through designing the representation-constrained term and the coefficients incoherence term, and the added two terms can get the reconstruction error of coding coefficients and correlations between similar samples by sharing dictionary under more stable control. Then, based on the proposed model, a supervised class-specific dictionary learning algorithm is developed by choosing appropriate samples with class labels. Finally, a classification scheme integrating the novel sparse model is designed to exploit such discriminative information. The proposed method is evaluated by conducting experiments on drug and tobacco leaves NIRS datasets. The experimental results show that the proposed sparse classification mechanism is promising for classifying NIRS and may be an alternative method to the traditional ones.

INDEX TERMS Near-infrared spectroscopy, sparse classification, regularization factors, representation-constrained, coefficients incoherence.

I. INTRODUCTION

Spectroscopy is more and more widely used in various fields of pattern recognition [1]. NIRS is an electromagnetic wave with a wavelength of 780 $\sim$ 2526nm. NIRS has many advantages, such as speed, accuracy, easy-use and non-destructiveness, and hence it has been used to classify some objects, especially with a large number of classes [2]. NIRS usually involves multiple kinds of samples, there is a lot of relationship between these features of samples. Therefore, it makes sense to treat with this complex data with some machine learning algorithms [3], [4].

NIRS has been widely used in the fields of food detection, petroleum industry, pharmacy and so on. Shen *et al.* [5] developed a model to rapidly discriminate the waste oil and four normal edible vegetable oils through combining NIRS with support vector machine (SVM) and the result of which was proved better than that of K-means and Linear Discriminant Analysis. Zhou *et al.* [6] employed training dictionary and sparse representation to test the cetane number, boiling point and total aromatics fast of diesel, but the model was the combination of singular value decomposition algorithm (K-SVD) and the orthogonal matching pursuit algorithm (OMP). Risoluti *et al.* [7] used NIRS coupled to chemometrics calibration to detect new psychoactive substances in street samples. The capabilities of this approach in forensic chemistry were assessed in the determination of new molecules appeared in the illicit market and often claimed to contain non-illegal compounds. Due to the differences in production process, packaging, raw materials and other

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

forms, the quality of the same kind of medicine produced by different manufacturers also varies to some extent. The identification of these differences is of great significance in drug supervision. Yang *et al.* [8] employed dropout and deep belief network to construct the Dropout-DBN classifier to identify near-infrared spectroscopy of erythromycin ethylsuccinate, which effectively alleviated the over-fitting phenomenon caused by the lacking of training data.

The problem to be solved in our paper is how to accurately predict the classes of spectral data from NIRS. Considering the scarcity of drug data, to verify the effectiveness of our proposed model, we mainly conducted corresponding research on the classification of drug and tobacco leaves datasets. The identification of metformin tablets and other drugs by near-infrared diffuse reflectance spectroscopy is still limited because the multi-classification problem is more complex than the two-category classification problem. On the other hand, the most important properties of tobacco are the flavor and aroma of the leaf. When tobacco leaves are graded well, different grades of tobacco leaves are easier to mix and meet numerous and strict specifications in terms of chemical composition, smoking characteristics, flavor, aroma and other aspects. Generally speaking, tobacco leaves are graded mainly according to the place of origin and color.

Therefore, what we are facing is a classification problem. The main work that we are interested in is to extract the feature of tobacco leaves and design a robust classification algorithm, which can recognize unknown samples. This pattern recognition task [9], [10] is important because the classification of tobacco leaves is mainly depending on human work, which is time-consuming. Therefore, it is necessary to develop a robust classification algorithm using machine learning algorithm. Recently proposed sparse representation classification (SRC) algorithms have obtained impressive performance in image classification. [11]–[17]. A full description of sparse representation classification (SRC) can be seen in [18]–[20]. SRC algorithms can be adopted at ascertaining the inherent similitude of objects embedded in high-dimensional target data to obtain good performance. Some researchers have shown great interests in this area, as a result of the competitive performance of the SRC scheme in face recognition [20], it triggers the interest of researchers in sparsity-based classification. However, learning a discriminative dictionary for a robust classification model is still very difficult.

Sparse representation has received a lot of attention in machine learning area, but most classification models mainly consider the reconstruction and representation error, and little attention is paid to the normalization of representation term and the incoherence of samples. Based on the predefined relationship between dictionary atoms and class labels, we can divide these existing dictionary learning approaches into three kinds: shared dictionary learning, class-specific dictionary learning and hybrid dictionary learning [21]–[23].

Shared dictionary learning approaches aimed at getting a shared dictionary from all classes [24]–[28]. Marial *et al.* [27] designed a novel sparse model through training a linear classifier. Inspired by the KSVD [29], Zhang and Li [25] proposed a joint learning framework to learn a dictionary for face recognition, and named it as discriminative KSVD (DKSVD). Following the work in [25], Jiang *et al.* [26] constructed a label consistent term to enhance the discriminative power of model. However, the shared dictionary does not consider the relationship between the dictionary atoms and the class labels, and the discriminant ability of the model has not been fully exploited. Class-specific dictionary learning approaches aimed at learning a dictionary that corresponds to class labels [23], [30]–[34]. Mairal *et al.* [32] proposed a novel dictionary learning approach by designing a reconstruction penalty term. Yang *et al.* [31] constructed a structural dictionary learning algorithm by adding non-negative penalty terms. Castrodad and Sapiro [33] designed a set of action-specific dictionaries to enhance the discriminant ability of model. Ramirez *et al.* [30] improved the sparse representation classification model by adding an incoherence promotion term. Hybrid dictionary learning approaches aimed at combining shared dictionary learning and class-specific dictionary learning, such as the methods in [35], [36] and [37]. These dictionary learning methods try to increase the between-class difference, while the within-class variation is not still accounted well.

In this paper, a novel sparse classification model and the effective feature representation method for classifying NIRS of drugs and tobacco leaves are proposed. First, considering the NIRS classification task, a robust classification framework is designed. Second, a modified sparse model has been proposed to learn a discriminative dictionary for classification. In the modified sparse model, the representation-constrained term and the coefficient incoherence term have been introduced to ensure the learned dictionary to obtain a powerful discriminative ability. At last, based on the proposed model, we have presented a corresponding classification scheme.

Our main contributions in this paper are three-fold: 1) a novel NIRS classification framework is designed; 2) a modified sparse model that incorporates the representation-constrained term and the coefficients incoherence term for classification is proposed; 3) A classification loss function is proposed to optimize class-specific dictionaries. The rest of the paper is organized as follows. Sect. II briefly presents the related work. Sect. III introduces our proposed algorithm. Sect. IV describes the optimization procedure of modified sparse model. Sect. V presents the modified sparse model based classifier. Sect. VI presents experimental results to demonstrates the robustness of our model. Sect. VII gives the conclusion of this paper.

II. RELATED WORK

A. ORIGINAL SRC MODEL

A common sparse representation based classification (SRC) scheme was presented in [20]. The whole class training

samples were utilized to constitute a dictionary that represented the query face image. And the image was categorized according to evaluation of which class caused the minimum error to reconstruct it.

There are K classes of subjects, $A = [A_1, A_2, \cdots, A_K]$ is the dictionary formed by A_i, where A_i $(i = 1, 2, \cdots, K)$ is the subset of training samples of class i, y is a test sample. The conclusions of the SRC algorithm are as below.

(1) Normalize each training sample A_i, $i = 1, 2, \cdots, K$.

(2) Define and solve the l_1-minimization problem: $\hat{x} = \arg\min_x\{\|y - Ax\|_2^2 + \gamma \|x\|_1\}$, where γ is a scalar constant.

(3) Label the test sample y by: $Label(y) = \arg\min_i\{e_i\}$, where $e_i = \left\|y - A_i\hat{\alpha}_i\right\|_2^2$, with $\hat{\alpha}_i$ representing the coefficient vector associated with class i.

Apparently, this scheme is based on the potential assumption that a weighted linear combination of training samples, which are attached only to the same category, can stand for a test sample. Wright *et al.* [20] reported that it achieved impressive performance which indicates the sparse representation has natural distinctiveness.

B. CLASS-SPECIFIC DICTIONARY LEARNING

For the dictionary learning, the atoms we used are denoted as $D = [D_1, D_2, \ldots, D_K]$ have class label correspondences to the subject classes, where D_i is the sub-dictionary corresponding to class i. Once the representation vector $\hat{\alpha} = [\hat{\alpha}_1; \hat{\alpha}_2; \ldots; \hat{\alpha}_K]$ is computed, the class-specific representation residual $\left\|y - D_i\hat{\alpha}_i\right\|_2$ could be used for classification. The sub-dictionary D_i could be learned class by class algorithm [38]: $\arg\min_{D_i,Z_i}\{\|A_i - D_iZ_i\|_F^2 + \lambda \|Z_i\|_1\}$, where Z_i is the representation matrix of A_i on D_i.

We regard it to be the fundamental model of dictionary learning in specific classes. It should be noted that the fundamental model mentioned above trains the class-specific sub-dictionaries respectively, regardless of the relationship between different sub-dictionaries. In order to ensure the discontinuity among the sub-dictionaries and the distinction of the entire class-specific dictionary, Ramirez *et al.* [30] designed incoherent promotion terminology to prompt the highly independence of the sub-dictionaries.

III. PROPOSED SPARSE CLASSIFICATION MODEL

Two terms are proposed in the model: the representation-constrained term and coefficients incoherence term. The representation-constrained term with good ability is to reconstruct query data which use training samples with the same class label by enforcing the class-specific sub-dictionary. The coefficients incoherence term with poor ability are to reconstruct query data that use training samples having different class labels by enforcing the class-specific sub-dictionary.

In the class-specific dictionary learning, each dictionary atom in the learned dictionary $D = [D_1, D_2, \ldots, D_K]$ have class label corresponding to the subject classes, where D_i is the sub-dictionary corresponding to class i. By representing a test sample over the learned dictionary D, the representation residual associated with each class can be naturally used to classify it.

Given $a_{i,j}$, $i = 1, \ldots K, j = 1, \ldots, n_i$ represents a training sample in class i, where K is the sum of classes, and n_i is the number of samples with respect to class i. Let $A = [A_1, A_2, \cdots, A_i] \in \mathbb{R}^{n\times N}$, where $A_i = [a_{i1}, a_{i2}, \cdots, a_{in}]$, $i = 1, 2, \cdots, K$ and n is the achieved feature dimension. We can learn the dictionary D from the following extended sparse model:

$$\begin{aligned} < D, X > = \arg\min_{D,X} \sum_{i=1}^{K} \{&\|A_i - DX_i\|_F^2 + \lambda_1\|X_i\|_1 \\ &+\lambda_2\|A_i - D_iX_i^i\|_F^2 + \xi_1 \sum_{j\neq i} \|X_j^T X_i\|_F^2 \\ &+\xi_2 \sum_{j\neq i} \|D_jX_i^j\|_F^2\} \\ s.t. \|d_n\|_2 = 1, \forall n \end{aligned} \tag{1}$$

where X_i is the sub-matrix containing the coding coefficients of A_i over D. X_i can be written as $X_i = [X_i^1; \ldots; X_i^j; \ldots; X_i^K]$, where X_i^j represents the coefficients of A_i over D_j. X_j is $X_j = [X_{j,1}, X_{j,2}, \ldots, X_{j,n_j}]$. And $\lambda_1, \lambda_2, \xi_1, \xi_2$ are the scalars controlling the relative contributions of the corresponding terms.

Unlike the traditional sparse model SRC in [20], the representation-constrained term $\|A_i - D_iX_i^i\|_F^2 + \sum_{j\neq i} \|X_j^T X_i\|_F^2$ and coefficients incoherence term $\sum_{j\neq i} \|D_jX_i^j\|_F^2$ are introduced in Eq.(1).

A. REPRESENTATION-CONSTRAINED TERM

X_i represents the sparse coefficients of A_i over dictionary D, so $A_i \approx DX_i$. X_i can be well represented by only A_i in class i because X_i is related to class i, which is in natural expectation. Therefore, there should exist a X_i^i such that $\|A_i - D_iX_i^i\|_F^2$ is small. This term is able to keep the reconstruction error of coefficients X_i. On the other hand, X_i is the similarity adapter that gives different freedom for each basis vector proportional to its similarity to the input signal A_i, so $\sum_{j\neq i} \|X_j^T X_i\|_F^2$ represents the classification error, and minimizing this term is to minimize the classification error.

Therefore, minimizing the representation-constrained term defined by $\|A_i - D_iX_i^i\|_F^2 + \sum_{j\neq i} \|X_j^T X_i\|_F^2$ can control both the reconstruction error and classification error.

B. COEFFICIENTS INCOHERENCE TERM

The test sample given in the SRC scheme proposed by Wright *et al.* [20] can be classified accurately in accordance with the maximum coefficients related to the training samples, which fall into the same class as the test sample. This means when the test sample is sparsely represented by its own training samples, the reconstruction error is minimized. Likewise, in the class-specific dictionary learning, it is expected that the largest coefficients of A_i are associated with the sub-dictionary D_i. In Eq. (1), minimizing the coefficients incoherence term $\sum_{j\neq i} \|D_jX_i^j\|_F^2$ encourages that

for the A_i and A_j, the largest coefficients are associated with the corresponding different sub-dictionary D_i and D_j. This means that similar samples over dictionary D have similar coefficients and samples belonging to different classes over dictionary D have absolutely different coefficients. Thus, when a sample is sparsely represented by a dictionary atom in its own sub-dictionary, the value of the object function Eq.(1) is minimized. Minimizing the coefficients incoherence term $\sum_{j\neq i}\|D_jX_i^j\|_F^2$ encourages samples from different classes are reconstructed by different sub-dictionaries

Overall, minimizing the representation-constrained term ensures that learned sub-dictionary has the powerful representation ability to the training samples from the corresponding class and minimizing the coefficients incoherence term encourages that maximum coefficients are related to the corresponding different sub-dictionary. It shows that similar samples have similar coefficients and samples belonging to different classes have completely different coefficients. Thus, when a sample is sparsely represented by a dictionary atom in its own sub-dictionary, the value of the object function Eq.(1) is minimized. By combining the representation-constrained term and coefficients incoherence term, this sparse representation model will be more effective for classification.

IV. OPTIMIZATION OF PROPOSED SPARSE MODEL

Although the objective function in Eq.(1) is not jointly convex to (D, X). it is convex with respect to each of D and X when the other one parameter is fixed.

A. UPDATE OF X

When the dictionary D is fixed, the objective function in Eq.(1) can be regarded as sparse coding problem for handling $X = [X_1, X_2, \ldots, X_K]$. When X_i changes, all $X_j (j \neq i)$ are also fixed. Thus, the objective function in Eq.(1) can be rewritten by:

$$\min_{X_i}\{\|A_i - DX_i\|_F^2 + \lambda_1\|X_i\|_1 + \lambda_2\|A_i - D_iX_i^i\|_F^2 + \xi_1\sum_{j\neq i}\|X_j^TX_i\|_F^2 + \xi_2\sum_{j\neq i}\|D_jX_i^j\|_F^2\} \quad (2)$$

It can be proved that $\varphi_i(X_i) = \|A_i - DX_i\|_F^2 + \lambda_2\|A_i - D_iX_i^i\|_F^2 + \xi_1\sum_{j\neq i}\|X_j^TX_i\|_F^2 + \xi_2\sum_{j\neq i}\|D_jX_i^j\|_F^2$ is convex with Lipschitz continuous gradient. In this paper, a novel fast iterative projection method (IPM) [39] is used to solve Eq.(2), which is introduced in Algorithm 1.

B. UPDATE OF D

Then we will give the detail of updating $D = [D_1, D_2, \ldots, D_K]$. When updating D_i, all $D_j, j \neq i$, are fixed and $D_i = [d_1, d_2, \ldots, d_{p_i}]$ is updated class by class. Objective function in Eq.(1) can be modified as:

$$\min_{D_i}\{\|\bar{A} - D_iX^i\|_F^2 + \lambda_2\|A_i - D_iX_i^i\|_F^2 + \xi_2\sum_{j\neq i}\|D_jX_i^j\|_F^2\}\ s.t.\ \|d_l\|_2 = 1, \quad l = 1, \ldots, p_i \quad (3)$$

Algorithm 1 Learning Sparse Code X_i.

Require:
A training subset A_i from class i; the dictionary D; the parameters $\rho, \tau > 0$.
Initialize:
$\hat{X}_i^{(1)} = 0$ and $t = 1$;
while convergence or the maximal iteration step is not reached **do**
$t = t + 1$; $u^{t-1} = \hat{X}_i^{(t-1)} - 1/2\rho\nabla\varphi_i(\hat{X}_i^{(t-1)})$,
where $\nabla\varphi_i(\hat{X}_i^{(t-1)})$ is the derivative of $\varphi_i(\hat{X}_i^{(t-1)})$ w.r.t. $\hat{X}_i^{(t-1)}$;
$\hat{X}_i^{(t)} = soft(u^{(t-1)}, \tau/\rho)$, where $soft(u, \tau/\rho)$ is defined by:
$soft(u, \tau/\rho) = 0$, if $\|u_j| \leq \tau/\rho$; $soft(u, \tau/\rho) = u_j - sign(u_j)\tau/\rho$, otherwise
end while
Ensure:
$\hat{X}_i = \hat{X}_i^{(t)}$.

where $\bar{A} = A - \sum_{j=1, j\neq i}^{K} D_jX^j$; X^i denotes the coefficient matrix of A over D_i. Eq.(3) can be further rewritten as:

$$\min_{D_i}\|\Lambda_i - D_iZ_i\|_F^2 \quad s.t.\ \|d_l\|_2 = 1, \quad l = 1, \ldots, p_i \quad (4)$$

where $\Lambda_i = [\bar{A}A_i0\ldots0]$, $Z_i = [X^iX_i^iX_1^i\ldots X_{i-1}^i\ldots X_k^i]$. 0 is a zero matrix. Eq.(4) can be efficiently solved by updating each dictionary atom one by one by the algorithm in [38], as is shown in Algorithm 2.

Algorithm 2 Learning Dictionary D_i.

Require:
A training subset A_i from class i; the coefficients X_i.
Let $Z_i = [z_1; z_2; \ldots; z_{p_i}]$ and $D_i = [d_1; d_2; \ldots; d_{p_i}]$, where $z_j, j = 1, 2, \ldots, p_i$,
is the row vector of Z_i and d_j is the jth column vector of D_i;
for $j = 1$ to p_i **do**
Fix all $d_l, l \neq j$ and update d_j. Let $Y = \Lambda_i - \sum_{l\neq j} d_lz_l$.
The minimization of Eq.(4) becomes: $\min_{d_j}\|Y - d_jz_j\|_F^2$ s.t. $\|d_j\|_2 = 1$
By solving this objective function, we could get the solution
$d_j = Xz_j^T/\|Yz_j^T\|_2$.
end for
Ensure:
Update all d_i, and hence the whole dictionary D will change.

C. COMPLETE DICTIONARY D LEARNING ALGORITHM

Algorithm 3 gives the complete learning flow. Due to the cost function in Eq.(1) has a lower bound and can merely reduce in the following two alternating minimization phases (i.e., updating X and updating D).

Algorithm 3 The Complete Algorithm of Dictionary D Learning.

Initialize D**.**

We initialize the atoms of D_i as the eigenvectors of A.

Update coefficients X**.**

Fix D and solve X_i, $i = 1, 2, \ldots, K$, one by one by solving Eq.(2) via Algorithm 1.

Update coefficients D**.**

Fix Z and update each D_i, $i = 1, 2, \ldots, K$, by solving Eq.(3) via Algorithm 2.

return

Update D and X until the objective function values in adjacent iterations are close enough or the maximum number of iterations is reached.

Ensure:

X and D

V. PROPOSED SPARSE MODEL-BASED CLASSIFIER

When the training process has finished, the dictionary D can be used to denote the query sample y and complete the classification task. On the basis of different ways to learn the dictionary D, we can use different information to carry out the classification task.

In our constructed classification model, considering both the representation-constrained term and coefficients incoherence term are discriminative, so we incorporate them into the classification mode to achieve more accurate classification results. As is described in the following equation:

$$\hat{\alpha} = \arg\min_{\alpha}\{\|y - D\alpha\|_2^2 + \gamma\|\alpha\|_1\} \tag{5}$$

where, γ constant.

Let $\hat{\alpha} = [\hat{\alpha}^1, \hat{\alpha}^2, \ldots, \hat{\alpha}^K]$, where $\hat{\alpha}^i$ is the coefficient sub-vector associated with sub-dictionary D_i. In the training stage, we have enforced the representation residual to be discriminative. Therefore, if y is from class i, the residual $\|y - D_i\hat{\alpha}^i\|_2^2$ should be small while $\|y - D_j\hat{\alpha}^j\|_2^2$, $j \neq i$, should be large. In addition, the representation sub-vector $\hat{\alpha}^j$ should be far different from the representation vector of other classes. Once we get X_i, the mean coefficient vector of class A_i, denoted by m_i, could be calculated. The mean vector m_i can be viewed as the center of class A_i in the transformed space spanned by the dictionary D. Taking into account the discriminative ability of the representation-constrained term and the coefficients incoherence term, the following metric can be defined to classify:

$$e_i = \|y - D_i\hat{\alpha}_i\|_2^2 + \beta_1 \sum_{j \neq i} \|X_j^T \hat{\alpha}\|/n_j + \beta_2\|\hat{\alpha}_i - m_i\|_2^2 \tag{6}$$

where β_1 and β_2 are preset weight to balance the contribution of the two terms for classification. The classification rule can be defined as $identity(y) = \arg\min_i\{e_i\}$.

TABLE 1. Details of the pharmaceutical samples.

Manufacturer	kinds	Non-aluminum plastic packed
Shanghai xinyi	metformin	90
shenzhen zhonglian	metformin	76
Tianjin Pacific	metformin	63
Bristol-Myers Squibb	metformin	138
Total	metformin	367

VI. EXPERIMENTAL RESULTS AND DISCUSSION

A. DATASETS AND EXPERIMENTS SETTINGS

In our paper, considering the scarcity of drug data, to verify the effectiveness of our proposed model, we also conducted more experiments on some tobacco leaves datasets.

Drug data were collected by the National Institutes for Food and Drug Control, including non-aluminum-plastic packaged metformin tablets produced by Shanghai xinyi pharmaceutical company and other pharmaceutical companies. The spectral curve of each sample at different wavelengths was determined by the Bruker spectrometer. The wavelength range of each spectral data was 4000-11995 cm^{-1}, with an interval of $4cm^{-1}$. There are 2074 absorption points at one complete spectrum. The drug is metformin hydrochloride tablets, non-aluminum-plastic packaging. Four manufacturers are selected, namely Shanghai xinyi pharmaceutical co., LTD., shenzhen zhonglian pharmaceutical co., LTD., Tianjin Pacific pharmaceutical co., LTD., and Bristol-Myers Squibb Pharmaceutical co., LTD. See the table1 below for the specific information.

While for the tobacco leaves dataset, tobacco leaves contain three kinds: Lugs, cutters and leaves. Based on the color difference, tobacco leaves can be also divided into three kinds: lemon, orange and red-brown. In reference [2], the authors combined the position and color feature and forming more categories, such as lemon lugs group, orange lugs group, lemon cutters group, orange cutters group, lemon leaves group, orange leaves group and redbrown leaves group. Each group also contains 3 or 4 levels. In our proposed sparse representation model, there are two stages: Dictionary learning stage and classification stage. In dictionary learning stage we set weight coefficients by experiment: $\lambda_1 = 0.005$, $\lambda_2 = 0.01$, $\xi_1 = 0.01$, $\xi_2 = 0.02$; In classification stage we set $\gamma = 0.001$, $\beta_1 = 0.05$, $\beta_2 = 0.005$. In the proposed model, the number of atoms in D_i, denoted by p_i, is important and it is set as the number of training samples by default. All of the experiments are executed on a workstation with 2.8GHz CPU and 16GB RAM. In this paper, we have constructed two experimental data sets. One group contains 207 tobacco leaf samples: It contains three different types of tobacco leaves, such as orange lugs grade 2, orange cutters grade 3 and orange leaves grade 2. The other group includes 200 tobacco leaf samples: it contains four categories, such as orange lugs grade 2, orange cutters grade 3, orange leaves grade 2 and red-brown leaves grade 2. The spectra in the near-infrared range of 1000∼2500 nm are saved in three parts using a Nicolet Nexus Fourier transform spectrometer. Through averaging the spectra of three parts, then we can

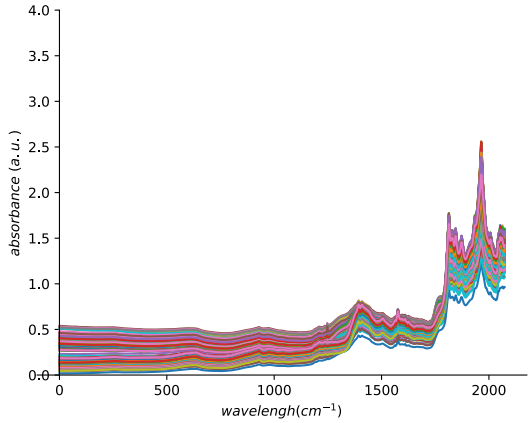


FIGURE 1. NIR spectra of phamaceutical samples.

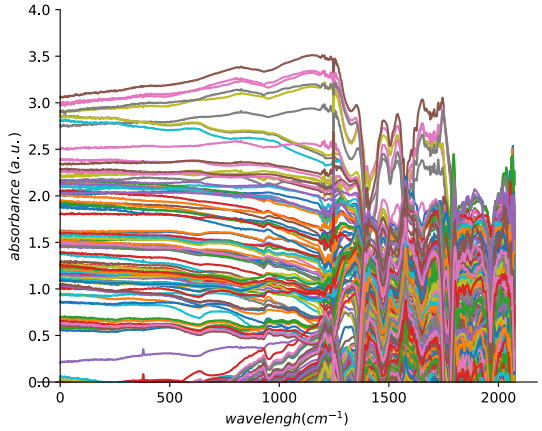


FIGURE 2. Normalized NIR spectra of pharmaceutical samples

get a mean spectrum for each sample. we adopt the same pre-processing method in [1] to process near-infrared spectral data.

B. EXPERIMENTS ON DRUG DATASET

In order to verify the proposed model in the identification of NIR of drugs, this model is used to carry out two and multiple classification experiments on the same drug with different packaging forms produced by different manufacturers, and is compared and evaluated with other methods. To verify the effectiveness of our proposed sparse model for drug data classification, we make comparisons with other classifiers, such as SRC [20], SVM, label consistent KSVD (LCKSVD) [26], FDDL approach [34], and our proposed algorithm.

First, the spectral deviation caused by drift and shift is eliminated by OPUS software, and 367 consistent drug spectra are obtained. As shown in Fig. 1, the spectra have high similarity and severe overlap, making it difficult to analyze the information. By means of spectral normalization processing, eliminate the order of magnitude difference between spectral data, and it can avoid the large difference between input and output data, normalized drug spectra are shown in Fig. 2.

TABLE 2. Size configuration of training sample set for two-category discrimination.

Total number of training sets	Number of negative samples	Number of positive samples
60	20	40
80	25	55
100	30	70
120	35	85
140	40	100
160	45	115
180	50	130
200	55	145

1) EXPERIMENTS ON TWO-CATEGORY

First, rough classification (typed two-category recognition) of the same drug in the same packaged form produced by the designated manufacturer and the other three manufacturers are used to verify the predictive power of our proposed model. As shown in table 1,a total of 367 spectral samples of drugs are collected. In this experiment, the data set is designed as bellows: 90 spectral samples in the form of aluminum-plastic packaging produced by the Shanghai xinyi pharmaceutical manufacturer are taken as the negative sample set; A total of 277 non-aluminum-plastic packaged spectral samples produced by the other three manufacturers are taken as the positive sample set. Also 10 training sets and test sets are constructed independently, and they were configured separately. Finally, we take average based on the running results of ten times.

In order to verify the performance of all algorithms under datasets, the positive and negative samples are randomly extracted according to the size of data in table2 and the proportion of positive and negative samples.

The SVM parameters are estimated by 5-fold cross-validation on each training data. The result is shown in Table 3. Through analyzing the experimental results, it is observed that the accuracy of our proposed model is higher than LCKSVD and FDDL. With the increase of training samples, the classification accuracy of our algorithm can reach 100%. Because the introduction of similarity constrained improves the degree of linear discrimination of data and the practical application ability of classification model, our algorithm can effectively improve the classification accuracy of the model. LCKSVD outperforms FDDL, which shows that representing the query sample on the whole dictionary is more effective than representing it on each class-specific sub-dictionary in the application. SRC and SVM are worse than other three algorithms, which implies that the ability of non-linear modeling is poor. Therefore, it proves that the proposed supervised sparse model is more discriminative, and by incorporating the representation-constrained term and coefficients incoherence term, our proposed sparse representation model is more effective for classification. The stability of our model is better than that of LCKSVD and FDDL, especially when the training sample is small. With the increase of training samples, our model achieves optimal stability. As is shown in Fig 3.

TABLE 3. The performance comparison on drug dataset (Testing accuracy:%).

Train/Test data	SRC	SVM	LCKSVD	FDDL	Ours
60/307	93.44	92.13	97.35	96.13	99.25
80/287	95.97	94.52	98.13	97.53	99.65
100/267	96.78	93.62	99.32	98.19	99.96
120/247	96.65	95.25	100	98.83	100
140/227	97.21	96.17	99.25	99.13	100
160/207	97.97	97.08	98.78	98.57	100
180/187	98.23	96.58	97.85	97.26	100
200/167	97.98	96.96	99.15	98.39	100

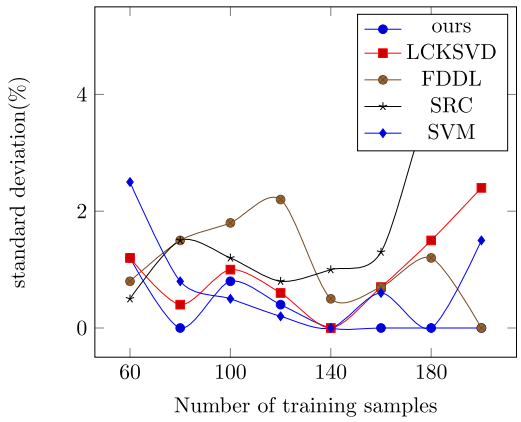


FIGURE 3. Standard deviations of accuracy for different binary classification models.

TABLE 4. Size configuration of training samples for multi-category discrimination.

Total number of training sets	Shanghai xinyi	Shenzhen zhonglian	Tianjin Pacific	Bristol-Myers
60	20	10	8	22
80	25	15	13	27
100	30	20	18	32
120	35	25	23	37
140	40	30	28	42
160	45	35	33	47
180	50	40	38	52
200	55	45	43	57

2) EXPERIMENTS ON MULTI-CATEGORY

Secondly, fine classification (typed multi-category recognition) of the same drug in the same packaged form produced by four different manufacturers is used to verify the predictive power of our proposed model. As shown in table 4,a total of 367 spectral samples of drugs are collected. In this experiment, the data set is divided into four categories. Also 10 training sets and test sets are constructed independently, and they are configured separately. Finally, we take an average based on the running results of ten times.

LCKSVD outperforms FDDL, which shows that representing the query sample on the whole dictionary is more effective than representing it on each class-specific sub-dictionary in the application. Compared with SRC, SVM, LCKSVD and FDDL, in the application of multi-classification of the drug, our proposed model shows stronger classification ability due to the introduction of proposed two additional terms, which improves the linear separability of data, while the accuracy of other algorithms own poor performance. When the number of training samples is small, our algorithm achieves high classification accuracy. With the increase of training samples, the classification accuracy of our model is stable at 100%. FDDL and LCKSVD are worse than our proposed algorithm, which implies that the similarity constrained is more powerful than the representation residual in this dataset. It proves that the proposed supervised sparse model is more discriminative, and by incorporating the representation-constrained term and coefficients incoherence term, our proposed sparse representation model is more effective for classification.

TABLE 5. Multi-classification accuracy on different ratios of training samples (Testing accuracy:%).

Train/Test data	SRC	SVM	LCKSVD	FDDL	Ours
60/307	93.44	92.13	96.39	95.73	98.65
80/287	90.01	84.52	95.93	96.13	99.27
100/267	94.78	93.62	97.12	96.85	99.69
120/247	96.65	95.25	97.79	97.08	100
140/227	97.21	96.17	98.25	97.51	100
160/207	97.81	97.11	98.81	97.93	100
180/187	98.29	97.87	99.07	98.12	100
200/167	98.52	98.01	100	98.61	100

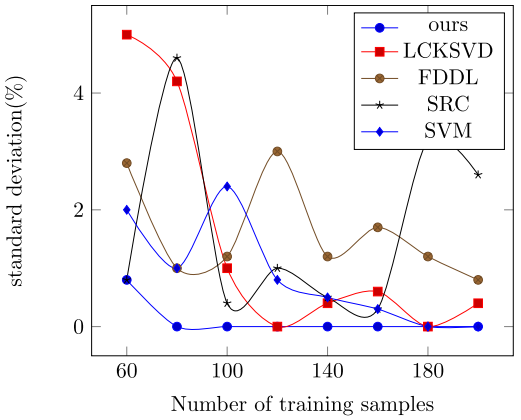


FIGURE 4. Standard deviations of accuracy for different multi-class classification.

As shown in Fig 4, our model delivers good stability and is superior to the other four methods in the case of small training samples and increased training samples. With the increase of training samples, our model obtains the optimal stability.

C. EXPERIMENTS ON TOBACCO LEAVES: DATASET 1

To verify the effectiveness of our proposed sparse model for tobacco leaves classification, we make comparisons with other classifiers. We use the same features exacted from NIRS as the input of SRC [20], SVM, label consistent KSVD (LCKSVD) [26], FDDL approach [34], and our proposed algorithm.

We evaluate our algorithm on data set 1 with cross-validation: 10∼39 random samples are used for training, and the remaining ones are selected for testing. For each size of training images, we process 10 times with our method and the results are averaged. The SVM parameters are estimated by 5-fold cross-validation on each training data. The result is shown in Table 6.

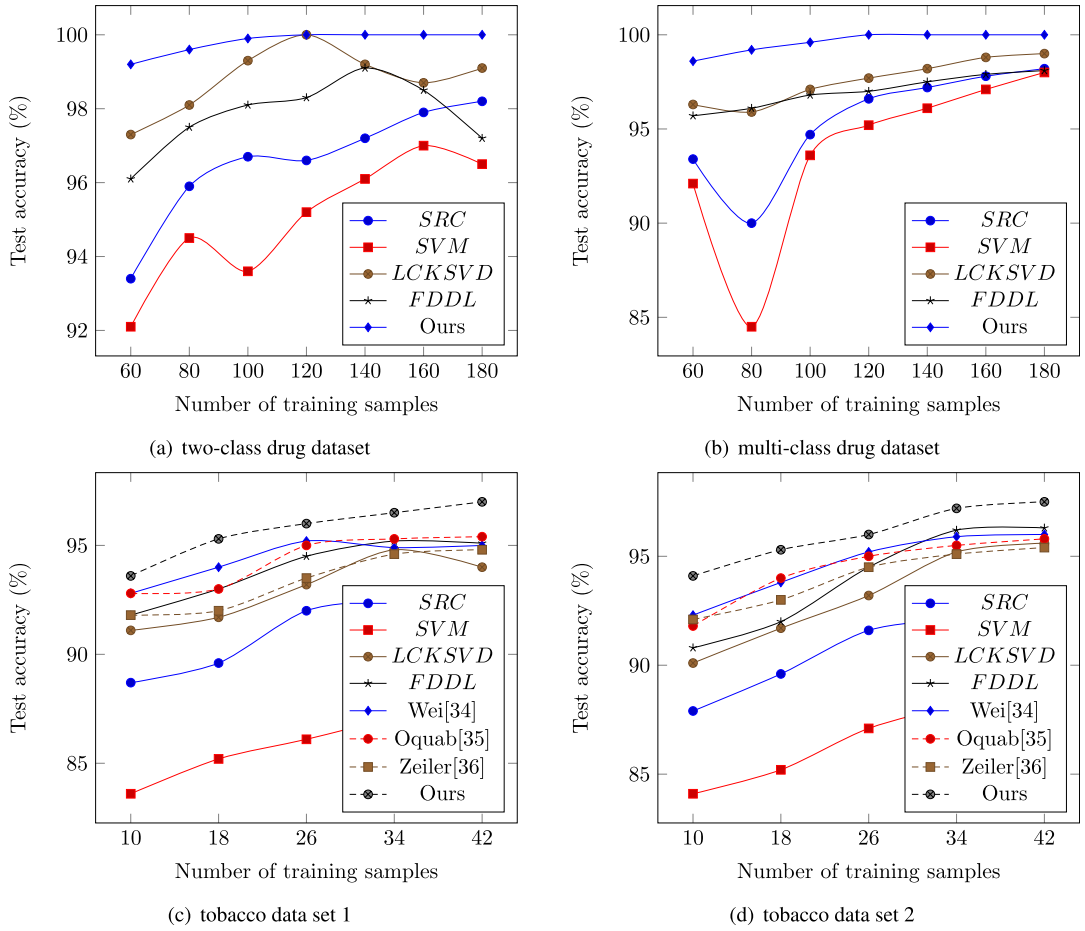


FIGURE 5. Performance comparisons of existing methods on all datasets.

TABLE 6. The performance comparison on data set 1(Accuracy:%).

Class	Orange lugs	Orange cutters	Orange leaves	Overall accuracy
SRC	92.38	93.44	92.13	92.65
SVM	82.07	90.01	84.52	85.53
LCKSVD	93.97	94.78	93.62	94.12
FDDL	94.58	96.65	95.25	95.49
Ours	95.38	97.21	96.17	96.25

TABLE 7. The performance comparison on data set 2 (Accuracy:%).

Category	Wei [41]	Oquab [42]	Zeiler [43]	SRC [20]	FDDL [35]	Ours
Orange lugs	94.7	93.5	92.0	91.3	94.2	95.6
Orange cutters	98.0	96.4	96.1	94.7	97.8	98.5
Orange leaves	97.2	96.7	96.4	94.9	97.5	98.2
Red-brown leaves	94.2	93.9	93.1	91.8	94.4	95.7
Overall accuracy	96	95.1	94.4	93.2	95.9	97

Through the above experiments, it is observed that the accuracy of SVM is higher than that of SRC which only uses the original training samples as dictionary. LCKSVD outperforms FDDL, which shows that representing the query sample on the whole dictionary is more effective than representing it on each class-specific sub-dictionary in the application. FDDL and LCKSVD are worse than our proposed algorithm, which implies that the similarity constrained is more powerful than the representation residual in this dataset. It proves that the proposed supervised sparse model is more discriminative, and by incorporating the representation-constrained term and coefficients incoherence term, our proposed sparse representation model is more effective for classification.

D. EXPERIMENTS ON TOBACCO LEAVES: DATASET 2

To further verify the effectiveness of our algorithm, we make experiments on data set 2. In this dataset, 25 samples of each class are used for training and remaining samples are selected for testing. We compare our model with the state-of-art methods, including multi-label classification [40], mid-level image representations [41], SRC [20], FDDL [34] and the convolutional network models in [42]. The experimental results are shown in Table 7. The overall accuracy we obtain on this dataset is 97%. For most of the categories, the results reach more than 95%. Our proposed method outperforms others with similar CNN features in most of the categories.

To demonstrate statistical significance, the classification performance is evaluated in terms of average accuracy and standard deviation (acc $\pm$ std), as is shown in Table 8 and Table 9. The corresponding test precision comparisons of experimental results are given in Fig 5. Comparing with the existing approaches, our proposed model can effectively

TABLE 8. Comparisons of classification results on drug dataset (Accuracy:%).

Algorithms	acc±std (two-class drug dataset)	acc±std (multi-class drug dataset)
SRC	96.16±2.11	96.73±2.06
SVM	95.23±1.99	96.45±1.98
LCKSVD	97.89±1.96	98.68±1.91
FDDL	98.53±1.91	98.02±1.89
Ours	**99.76±1.51**	**99.89±1.47**

TABLE 9. Comparisons of classification results on tobacco dataset (Accuracy:%).

Algorithms	acc±std (tobacco data set 1)	acc±std (tobacco data set 2)
SRC	92.51±3.05	92.13±3.06
SVM	87.53±3.22	88.11±3.97
LCKSVD	94.81±2.15	95.51±2.61
FDDL	95.13±1.97	96.38±2.19
Wei [41]	95.13±1.97	96.15±2.35
Oquab [42]	95.31±1.92	95.78±2.89
Zeiler [43]	94.68±2.07	95.33±2.94
Ours	**96.51±1.71**	**97.26±1.67**

improve the classification performance on the two benchmark datasets. In most cases, the precision of our constructed model is at least 3% better than the others, except LCKSVD and FDDL. To give more statistical analysis, we perform $t-test$ for the precision obtained by LCKSVD and FDDL and by the proposed method on the four datasets, under the null hypothesis using a significance level of 0.05. $p-value$ is found as 0.0016, 0.00052, 0.0018, 0.00021, respectively, indicating that accuracy rates achieved by the proposed model are indeed significantly better than LCKSVD and FDDL.

VII. CONCLUSION

In this paper, we have presented a novel method for drug and tobacco leaves classification based on NIRS and class-specific sparse model. Our proposed scheme incorporates NIRS information which represents samples more accurately. We have also introduced a supervised classification model with the dictionary learning method and designed a classification loss function. Our framework achieves comparable performance on the datasets above. The experiments have validated that our proposed framework works very well, and the proposed sparse model incorporating NIRS information is effective for classification.

REFERENCES

[1] J. Zhang, W. Liu, Y. Hou, C. Qiu, S. Yang, C. Li, and L. Nie, "Sparse representation classification of tobacco leaves using near-infrared spectroscopy and a deep learning algorithm," *Anal. Lett.*, vol. 51, no. 7, pp. 1029–1038, 2018.

[2] J. Bin, F.-F. Ai, W. Fan, J.-H. Zhou, Y.-H. Yun, and Y.-Z. Liang, "A modified random forest approach to improve multi-class classification performance of tobacco leaf grades coupled with NIR spectroscopy," *RSC Adv.*, vol. 6, no. 36, pp. 30353–30361, 2016.

[3] H. Lu, Y. Li, M. Chen, H. Kim, and S. Serikawa, "Brain intelligence: Go beyond artificial intelligence," *Mobile Netw. Appl.*, vol. 23, no. 2, pp. 368–375, Apr. 2018.

[4] H. Lu, D. Wang, Y. Li, J. Li, X. Li, H. Kim, S. Serikawa, and I. Humar, "CONet: A cognitive ocean network," *IEEE Wireless Commun.*, vol. 26, no. 3, pp. 90–96, 2019.

[5] X. Shen, X. Zheng, Z. Song, D. He, and P. Qi, "Rapid identification of waste cooking oil with near infrared spectroscopy based on support vector machine," in *Proc. Int. Conf. Comput. Comput. Technol. Agricult.*, vol. 92, 2012, pp. 11–18.

[6] Y. Zhou, S. Dai, J. Lv, T. Liu, and Y. Shi, "Application of training dictionary and sparse representation in diesel property detection using near-infrared spectroscopy," *Acta Chim. Sinica-Chin. Ed.*, vol. 70, no. 18, pp. 1969–1973, 2012.

[7] R. Risoluti, S. Materazzi, A. Gregori, and L. Ripani, "Early detection of emerging street drugs by near infrared spectroscopy and chemometrics," *Talanta*, vol. 153, pp. 407–413, Jun. 2016.

[8] H. Yang, B. Hu, X. Pan, S. Yan, Y. Feng, X. Zhang, L. Yin, and C. Hu, "Deep belief network-based drug identification using near infrared spectroscopy," *J. Innov. Opt. Health Sci.*, vol. 10, no. 2, 2016, Art. no. 1630011.

[9] R. Lan, Y. Zhou, Z. Liu, and X. Luo, "Prior knowledge-based probabilistic collaborative representation for visual recognition," *IEEE Trans. Cybern.*, to be published. doi: 10.1109/TCYB.2018.2880290.

[10] R. Lan, H. Lu, Y. Zhou, Z. Liu, and X. Luo, "An LBP encoding scheme jointly using quaternionic representation and angular information," *Neural Comput. Appl.*, pp. 1–7, Jan. 2019. doi: 10.1007/s00521-018-03968-y.

[11] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[12] B. Han, F. Wu, and D. Wu, "Image representation by compressive sensing for visual sensor networks," *J. Vis. Commun. Image Represent.*, vol. 21, no. 4, pp. 325–333, 2010.

[13] Z. Fan, M. Ni, Q. Zhu, C. Sun, and L. Kang, "L_0-norm sparse representation based on modified genetic algorithm for face recognition," *J. Vis. Commun. Image Represent.*, vol. 28, pp. 15–20, Apr. 2015.

[14] Y. Xu, B. Zhang, and Z. Zhong, "Multiple representations and sparse representation for image classification," *Pattern Recognit. Lett.*, vol. 68, pp. 9–14, Dec. 2015.

[15] J. Sánchez and J. Redolfi, "Exponential family Fisher vector for image classification," *Pattern Recognit. Lett.*, vol. 59, pp. 26–32, Jul. 2015.

[16] H. Lu, Y. Li, S. Nakashima, and S. Serikawa, "Turbidity underwater image restoration using spectral properties and light compensation," *IEICE Trans. Inf. Syst.*, vol. E99-D, no. 1, pp. 219–227, 2016.

[17] R. Lan, J. He, S. Wang, T. Gu, and X. Luo, "Integrated chaotic systems for image encryption," *Signal Process.*, vol. 147, pp. 133–145, Jun. 2018.

[18] Z. Zhang, Y. Xu, J. Yang, X. Li, and D. Zhang, "A survey of sparse representation: Algorithms and applications," *IEEE Access*, vol. 3, pp. 490–530, 2015.

[19] H. Cheng, Z. Liu, L. Yang, and X. Chen, "Sparse representation and learning in visual recognition: Theory and applications," *Signal Process.*, vol. 93, no. 6, pp. 1408–1425, 2013.

[20] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[21] P. Dai, H. Di, L. Dong, L. Tao, and G. Xu, "Group interaction analysis in dynamic context," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 38, no. 1, pp. 275–282, Feb. 2008.

[22] D. Damen and D. Hogg, "Recognizing linked events: Searching the space of feasible explanations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 927–934.

[23] H. Wang, C. Yuan, W. Hu, and C. Sun, "Supervised class-specific dictionary learning for sparse modeling in action recognition," *Pattern Recognit.*, vol. 45, no. 11, pp. 3902–3911, 2012.

[24] J. Mairal, F. Bach, and J. Ponce, "Task-driven dictionary learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 791–804, Apr. 2012.

[25] Q. Zhang and B. Li, "Discriminative K-SVD for dictionary learning in face recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2691–2698.

[26] Z. Jiang, Z. Lin, and L. S. Davis, "Label consistent K-SVD: Learning a discriminative dictionary for recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2651–2664, Nov. 2013.

[27] J. Mairal, F. R. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Supervised dictionary learning," in *Proc. NIPS*, 2009, pp. 1033–1040.

[28] S. Bengio, F. Pereira, Y. Singer, and D. Strelow, "Group sparse coding," in *Proc. NIPS*, 2009, pp. 82–89.

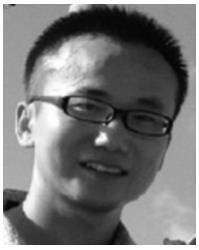
[29] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[30] I. Ramirez, P. Sprechmann, and G. Sapiro, "Classification and clustering via dictionary learning with structured incoherence and shared features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 3501–3508.
[31] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Fisher discrimination dictionary learning for sparse representation," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 543–550.
[32] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Learning discriminative dictionaries for local image analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2008, pp. 1233–1240.
[33] A. Castrodad and G. Sapiro, "Sparse modeling of human actions from motion imagery," *Int. J. Comput. Vis.*, vol. 100, no. 1, pp. 1–15, Jan. 2012.
[34] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Sparse representation based Fisher discrimination dictionary learning for image classification," *Int. J. Comput. Vis.*, vol. 109, no. 3, pp. 209–232, 2014.
[35] N. Zhou, Y. Shen, J. Peng, and J. Fan, "Learning inter-related visual dictionary for object recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 17, Jun. 2012, pp. 3490–3497.
[36] S. Kong and D. Wang, "A dictionary learning approach for classification: Separating the particularity and the commonality," in *Proc. Eur. Conf. Comput. Vis.*, vol. 100, 2012, pp. 186–199.
[37] L. Shen, S. Wang, G. Sun, S. Jiang, and Q. Huang, "Multi-level discriminative dictionary learning towards hierarchical visual categorization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 16, Jun. 2013, pp. 383–390.
[38] A. Y. Yang, Z. Zhou, A. G. Balasubramanian, S. S. Sastry, and Y. Ma, "Fast ℓ_1-minimization algorithms for robust face recognition," *IEEE Trans. Image Process.*, vol. 22, no. 8, pp. 3234–3246, Aug. 2013.
[39] L. Rosasco, A. Verri, M. Santoro, S. Mosci, and S. Villa, "Iterative projection methods for structured sparsity regularization," Massachusetts Inst. Technol., Cambridge, MA, USA, Tech. Rep. CBCL-282, 2009.
[40] Y. Wei, W. Xia, J. Huang, B. Ni, J. Dong, Y. Zhao, and S. Yan, "CNN: Single-label to multi-label," *Comput. Sci.*, vol. 6, no. 1, pp. 1–14, 2014.
[41] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1717–1724.
[42] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 818–833.

LINGQIAO LI is currently pursuing the Ph.D. degree with the School of Automation, Beijing University of Posts and Telecommunications, China. He is also an Assistant Researcher with the Guilin University of Electronic and Technology, China. His research interests include pattern recognition and large spectral data analysis.

XIPENG PAN is currently pursuing the Ph.D. degree with the School of Automation, Beijing University of Posts and Telecommunications, China. His research interests include machine learning and medical image processing. He is a member of CCF.

HUIHUA YANG received the Ph.D. degree from the East China University of Science and Technology, China, in 2005. He was a Postdoctoral Research Fellow of Tsinghua University, from 2005 to 2007. He is currently a Professor with the School of Automation, Beijing University of Posts and Telecommunications, China. He has published more than 40 papers. His research interests include machine learning, spectrum analysis, and optimization. He is a Senior Member of CCF and a member of ACM. He serves as the Director of the China Instrument and Control Society (CICS) and the Vice Director of the NIR Division of CICS.



TAO ZHANG received the bachelor's degree from Henan Polytechnic University, China, in 2008, and the Ph.D. degree from the Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, China. He is currently an Associate Professor with the Department of Computer Science, Jiangnan University. His major research interests include visual surveillance, object detection, and pattern analysis.

ZHENBING LIU received the Ph.D. degree in computer science from the Huazhong University of Science and Technology, China, in 2010. He was a Visiting Research Fellow with Pennsylvania University, in 2015. He is currently a Professor with the Guilin University of Electronic Technology, China. He has published more than 30 papers. His research interests include machine learning and medical image processing.