Locality Constraint Dictionary Learning With Support Vector for Pattern Classification

HE-FENG YIN[ID][1,2], XIAO-JUN WU[ID][1,2], AND SU-GEN CHEN[ID][3]

[1]School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China
[2]Jiangsu Provincial Engineering Laboratory of Pattern Recognition and Computational Intelligence, Jiangnan University, Wuxi 214122, China
[3]School of Mathematics and Computational Science, Anqing Normal University, Anqing 246133, China

Corresponding author: Xiao-Jun Wu (wu_xiaojun@jiangnan.edu.cn)

ABSTRACT Discriminative dictionary learning (DDL) has recently gained significant attention due to its impressive performance in various pattern classification tasks. However, the locality of atoms is not fully explored in conventional DDL approaches which hampers their classification performance. In this paper, we propose a locality constraint dictionary learning with support vector discriminative term (LCDL-SV), in which the locality information is preserved by employing the graph Laplacian matrix of the learned dictionary. To jointly learn a classifier during the training phase, a support vector discriminative term is incorporated into the proposed objective function. Moreover, in the classification stage, the identity of test data is jointly determined by the regularized residual and the learned multi-class support vector machine. Finally, the resulting optimization problem is solved by utilizing the alternative strategy. Experimental results on benchmark databases demonstrate the superiority of our proposed method over previous dictionary learning approaches on both hand-crafted and deep features. *The source code of our proposed LCDL-SV is accessible at https://github.com/yinhefeng/LCDL-SV.*

INDEX TERMS Dictionary learning, support vector discriminative term, locality constraint, pattern classification.

I. INTRODUCTION

Dictionary learning (DL) has aroused considerable interest during the past decade and has been adopted in a wide range of applications, such as face recognition [1], image fusion [2] and person re-identification [3], [4]. According to the characteristic of the learned dictionary, existing DL approaches for pattern classification can be divided into three categories: synthesis dictionary learning (SDL), analysis dictionary learning (ADL) and dictionary pair learning (DPL). In SDL, the dictionary is employed to represent the input data as a linear superposition of atoms. ADL aims to yield the sparse representation by exploiting the dictionary as a transformation matrix. DPL, also referred to as analysis-synthesis dictionary learning (ASDL), can jointly learn synthesis dictionary and analysis dictionary. According to whether the dictionary is class-shared or not, SDL can be further divided into three different types, i.e., shared SDL, class-specific SDL

The associate editor coordinating the review of this manuscript and approving it for publication was Madhu S. Nair[ID].

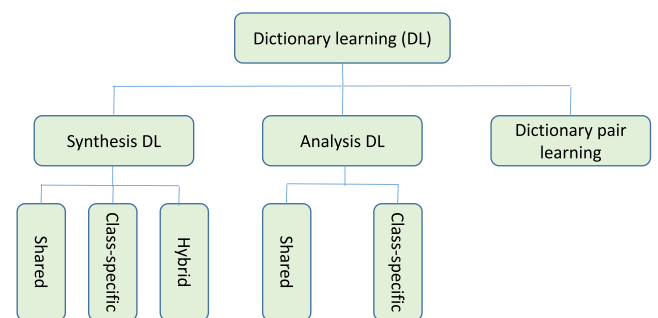**FIGURE 1.** A taxonomy of dictionary learning approaches for pattern classification.

and hybrid SDL. Similarly, ADL can be classified into two categories, i.e., shared ADL and class-specific ADL. Fig. 1 presents a taxonomy of dictionary learning approaches for pattern classification.

In class-specific SDL, sub-dictionary for each class is independently learned, then all the sub-dictionaries are

concatenated to form the final dictionary. Ramirez *et al.* [5] presented a dictionary learning with structured incoherence (DLSI) method by imposing incoherence constraint on sub-dictionaries so as to encourage dictionaries correspond to different classes to be as independent as possible. Yang *et al.* [6] proposed a metaface learning (MFL) algorithm which learns a set of metafaces for each class. Yang *et al.* [7] developed a Fisher discrimination dictionary learning (FDDL) method which imposes the Fisher discrimination criterion on the coding coefficients to learn class-specific sub-dictionaries. By considering the fact that different training samples contribute unequally to the dictionary, Liu *et al.* [8] proposed a class specific dictionary learning (CSDL) approach. Akhtar *et al.* [9] developed a joint discriminative Bayesian dictionary and classifier learning (JBDC) approach which associates the dictionary atoms with the class labels using Bernoulli distributions. By employing the directions of coefficients to promote the discriminative capability of representation, Wang *et al.* [10] presented a unidirectional representation dictionary learning (URDL) algorithm. Ling *et al.* [11] proposed a class-oriented discriminative DL (CODDL) method, in which the class-specific sub-dictionaries are learned in a classwise fashion.

In shared SDL, a universal dictionary shared by all classes is learned. The most classic SDL approach is the K-SVD algorithm [12] which has been successfully applied to image compression and denoising. However, KSVD mainly focuses on the representational ability of the dictionary without considering its capability for classification. To address this problem, Zhang and Li [13] proposed a discriminative K-SVD (D-KSVD) method by introducing the classification error into the framework of K-SVD. Jiang *et al.* [14] further incorporated a label consistency constraint into K-SVD and presented a label consistent K-SVD (LC-KSVD) algorithm. The ℓ_0-norm sparse regularization term is used in LC-KSVD, which is difficult to find the optimum sparse solution. To overcome this limitation, Shao *et al.* [15] explored a label embedded dictionary learning (LEDL) method which utilizes the ℓ_1-norm as the sparse regularization term. By jointly learning a multi-class support vector machine (SVM) classifier, Cai *et al.* [16] developed a support vector guided dictionary learning (SVGDL) model. Zhang *et al.* [17] designed class relatedness oriented discriminative DL (CRO-DDL) method which utilizes the $\ell_{1,\infty}$ norm constraint [18] on the coding coefficient matrix. By integrating multiple classifiers training into dictionary learning process, Quan *et al.* [19] presented a multiple classifiers based dictionary learning (MCDL) method. Dong *et al.* [20] proposed an orthonormal DL method by exerting an orthonormal constraint on the learned dictionary to enforce the dictionary atoms to be as dissimilar as possible. Min *et al.* [21] constructed a Laplacian regularized locality-constrained coding (LapLLC) algorithm for image classification, in which the similarity matrix is defined on the training data. To fully exploit the locality and label information of the learned dictionary, Li *et al.* [22]

constructed a locality-constrained and label embedding dictionary learning (LCLE-DL) algorithm. Song *et al.* [23] presented a class-wise discriminative DL (CW-DDL) method which introduces a label-aware constraint and graph regularization into the framework of SDL. By employing the profiles (row vectors of coding coefficient matrix) to construct discriminative terms in SDL, Li *et al.* [24] proposed an interactively constrained discriminative DL (IC-DDL) algorithm for image classification.

In hybrid SDL, a dictionary that contains several class-specific sub-dictionaries and a shared dictionary is learned. Kong and Wang [25] proposed a DL approach dubbed DL-COPAR which explicitly learns the shared patterns (the commonality) and the class-specific dictionaries (the particularity). Gao *et al.* [26] developed a category-specific and shared dictionary learning (CSDL) method for fine-grained image categorization. Sun *et al.* [27] presented a discriminative group sparse dictionary learning (DGSDL) model which learns a class-specific sub-dictionary for each class as well as a common sub-dictionary shared by all classes. By introducing a cross-label suppression constraint and group regularization term into the framework of SDL, Wang *et al.* [28] designed a cross-label suppression discriminative DL (CLS-DDL) approach. Lin *et al.* [29] proposed a robust, discriminative and comprehensive dictionary learning (RDCDL) model which learns a class-shared dictionary, class-specific dictionaries and a disturbance dictionary to represent the commonality, particularity and disturbance components in the data. In order to tackle corrupted samples, Vu and Monga [30] developed a low-rank shared dictionary learning (LRSDL) framework which simultaneously learns a set of common patterns and class-specific features for classification. By integrating the low-rank matrix recovery technique with the class-specific and class-shared dictionary learning, Rong *et al.* [31] explored a low-rank double dictionary learning (LRD²L) approach. Du *et al.* [32] proposed a low-rank graph preserving discriminative dictionary learning (LRGPDDL) method which incorporates the low-rank constraint on the class-specific dictionaries, graph preserving criterion and the dictionary incoherence term into the framework of SDL. Readers can refer to [33] for a survey of SDL approaches.

Recently, ADL has received increasing attention due to its efficacy and efficiency, and shared ADL has been widely studied. Rubinstein *et al.* [34] presented analysis K-SVD which is parallel to the synthesis K-SVD [12]. Afterwards, Shekhar *et al.* [35] applied ADL to image classification tasks and obtained comparable or better recognition performance than conventional SDL models. To enhance the classification performance of ADL, Guo *et al.* [36] proposed discriminative ADL (DADL) method. By introducing a synthesis-linear-classifier-based error term into the basic ADL model, Wang *et al.* [37] presented a synthesis linear classifier based ADL (SLC-ADL) algorithm. By solving a joint learning of ADL and a linear classifier through K-SVD based technique, Wang *et al.* [38] designed a synthesis

K-SVD based ADL (SK-SVDADL) method. Similar to LC-KSVD [14], Tang *et al.* [39] incorporated the label consistency term and classification error term into the framework of ADL and developed a structured ADL (SADL) approach. Maggu *et al.* [40] proposed label consistent transform learning (LCTL) for hyperspectral image classification. In essence, transform learning and ADL have similar formulation. For class-specific ADL, Wang *et al.* [41] proposed a class-aware ADL model which learns a discriminative analysis sub-dictionary for each class.

In DPL, a pair of synthesis dictionary and analysis dictionary is learned from the input data. Gu *et al.* [42] presented a projective dictionary pair learning (PDPL) framework which jointly learns a synthesis dictionary and an analysis dictionary. To further enhance the discriminative ability of DPL, Chen *et al.* [43] developed a discriminative DL approach called DPL-SV which introduces a differentiable support vector discriminative term into the DPL model. DPL does not impose sparse constraint on the representation matrix, which may lose discriminative power of sparse property. To alleviate this problem, Zhang *et al.* [44] designed a joint label consistent embedding and dictionary learning (JEDL) model which explicitly exploit a sparse constraint on the representation matrix. To preserve the locality property of learned atoms in the synthesis dictionary, Zhang *et al.* [45] proposed a locality constrained projective dictionary learning (LC-PDL) method. By jointly learning a classifier with the dictionary pair, Yang *et al.* [46] explored a discriminative analysis-synthesis dictionary learning (DASDL) model. To preserve the local geometry structure of input data, Chang *et al.* [47] presented a graph-regularized discriminative analysis-synthesis dictionary pair learning (GDASDL) model to enhance the classification performance of DASDL. To integrate structured dictionary learning, analysis representation and analysis classifier training into a unified framework, Zhang *et al.* [48] proposed an analysis discriminative dictionary learning (ADDL) algorithm. Inspired by the superiority of $\ell_{1,\infty}$ norm [18], Wei *et al.* [49] developed a fast DDL (FaDDL) method for synthetic aperture radar (SAR) image classification. The ordinal locality of analysis dictionary is not fully exploited in the above DPL and its variants, to tackle this problem, Li *et al.* [50] proposed a discriminative low-rank analysis-synthesis dictionary learning (LR-ASDL) algorithm with the adaptively ordinal locality.

In addition to the above DL approaches, to deal with multi-view data, some multi-view DL methods have been presented recently. Wu *et al.* [51] offered a multi-view low-rank dictionary learning (MLDL) method for image classification. Wu *et al.* [52] proposed a multi-view discriminant dictionary learning via learning view-specific and shared structured dictionaries (MDVSD) for image classification, in which a structured dictionary shared by all views and multiple view-specific structured dictionaries are simultaneously learned. Ma *et al.* [53] developed a multi-view coupled dictionary pair learning (MVCDL) framework for person re-identification. Wu *et al.* [54] presented a multi-view synthesis

and analysis dictionaries learning (MSADL) approach for pattern classification.

However, ADL often requires enormous atoms to achieve satisfactory results when applied to pattern classification. For hybrid SDL, how to choose the optimal number of shared atoms remains unresolved. Moreover, the optimization process of class-specific SDL is time-consuming, especially when the number of classes is large. In this paper, we propose a locality constraint dictionary learning with support vector discriminative term (LCDL-SV) for pattern classification, which belongs to the shared SDL category. A support vector discriminative term is introduced to promote the discrimination of coding coefficients. Since the original training data may contain noise or outliers, graph Laplacian matrix constructed on the original training samples cannot faithfully describe the manifold structure. To alleviate this problem, we employ a locality constraint on atoms. The atoms are updated in the dictionary learning procedure, thus the graph Laplacian matrix defined on the atoms is also updated. More importantly, to further enhance the classification performance of our proposed method, the regularized residual and the learned multi-class SVM classifier are jointly exploited to classify the test data. The flowchart of our proposed method for classification is illustrated in Fig. 2. Firstly, features are extracted from the original training and test samples, respectively. Then the training data is fed into our proposed dictionary learning algorithm, when the dictionary learning process is completed, a compact dictionary and multi-class SVM are obtained. Finally, the test data is classified based on the learned multi-class SVM and the regularized residual. Our main contributions are summarized as follows,

- A locality constraint of atoms is introduced in our approach, and this term can intrinsically inherit the manifold structure of training data.
- In addition to the learned SVM classifier, we take the regularized residual into account to further promote the classification performance.
- The resulting problem is solved elegantly by employing an alternative optimization technique.

The remainder of this paper is structured as follows. Section II reviews related work on SDL. In Section III, we present our proposed approach, and detailed optimization procedures are given in Section IV. Section V reports experimental results on five benchmark datasets. Finally, conclusions are drawn in Section VI.

II. RELATED WORK

In this section, we will briefly review some related work, including the basic K-SVD [12] and its two discriminative variants, i.e., D-KSVD [13] and LC-KSVD [14]. Additionally, support vector guided dictionary learning (SVGDL) method [16] is also introduced. To begin with, we first give an introduction to the notations used throughout this paper. Let $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_C] = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n] \in \mathbb{R}^{m \times n}$ be the data matrix of n training samples belonging to C classes,

of K-SVD,

$$\min_{\mathbf{D},\mathbf{W},\mathbf{Z}} \|\mathbf{X} - \mathbf{DZ}\|_F^2 + \beta \|\mathbf{H} - \mathbf{WZ}\|_F^2 + \lambda \|\mathbf{W}\|_F^2,$$
$$\text{s.t. } \|z_i\|_0 \le T_0 \qquad (2)$$

where $\mathbf{H} = [\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_n] \in \mathbb{R}^{C \times n}$ is the label matrix of training data, $\boldsymbol{h}_i = [0, 0, \ldots, 1, \ldots, 0, 0]^T \in \mathbb{R}^{C \times 1}$ is the label vector of $\boldsymbol{x}_i$, and $\mathbf{W}$ is the parameters for a linear classifier. As can be seen from (2), dictionary and a linear classifier are jointly learned in D-KSVD. Afterwards, Jiang *et al.* [14] presented LC-KSVD by solving the following optimization problem,

$$\min_{\mathbf{D},\mathbf{W},\mathbf{A},\mathbf{Z}} \|\mathbf{X} - \mathbf{DZ}\|_F^2 + \alpha \|\mathbf{Q} - \mathbf{AZ}\|_F^2 + \beta \|\mathbf{H} - \mathbf{WZ}\|_F^2,$$
$$\text{s.t. } \|z_i\|_0 \le T_0 \qquad (3)$$

where $\mathbf{Q} = [\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_n] \in \mathbb{R}^{K \times n}$ is an ideal representation matrix and $\mathbf{A}$ is a linear transformation matrix.

B. SVGDL

To promote the discriminative ability of coding vectors, Cai *et al.* [16] introduced a multi-class SVM regularization term into the framework of SDL. The regularization term is defined as follows,

$$L(\mathbf{Z}) = 2 \sum_{c=1}^{C} f(\mathbf{Z}, \boldsymbol{y}^c, \boldsymbol{u}_c, b_c) \qquad (4)$$

where $\boldsymbol{u}_c$ is the normal vector associated with the c-th class hyperplane of SVM, b_c is the corresponding bias, and $\boldsymbol{y}^c = [y_1^c, y_2^c, \ldots, y_n^c]$ is defined as $y_i^c = 1$ if class labels $y_i = c$ and otherwise $y_i^c = -1$. Concretely, the discrimination term is $f(\mathbf{Z}, \boldsymbol{y}, \boldsymbol{u}, b) = \|\boldsymbol{u}\|_2^2 + \theta \sum_{i=1}^{n} l(z_i, y_i, \boldsymbol{u}, b)$, where $l(z_i, y_i, \boldsymbol{u}, b)$ is the hinge loss function, and θ is a penalty parameter.

The objective function of SVGDL is formulated as follows,

$$\min_{\mathbf{D},\mathbf{Z},\mathbf{U},b} \|\mathbf{X} - \mathbf{DZ}\|_F^2 + 2\lambda_2 \sum_{c=1}^{C} f(\mathbf{Z}, \boldsymbol{y}^c, \boldsymbol{u}_c, b_c) + \lambda_1 \|\mathbf{Z}\|_F^2,$$
$$\text{s.t. } \|\boldsymbol{d}_k\|^2 \le 1 \qquad (5)$$

where $\mathbf{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_C]$ and $\boldsymbol{b} = [b_1, b_2, \ldots, b_C]$.

III. PROPOSED METHOD

In this section, our proposed LCDL-SV is presented. First we will introduce a locality constraint on the atoms of the learned dictionary. Then by incorporating the locality constraint and the support vector discriminative term into the framework of SDL, we will present the formulations of our proposed LCDL-SV.

A. LOCALITY CONSTRAINT ON ATOMS

As mentioned earlier, $\mathbf{Z}$ is the coding coefficient matrix of training data $\mathbf{X}$ over the dictionary $\mathbf{D}$, and $z_i = [z_{1,i}, z_{2,i}, \ldots, z_{K,i}]^T, (i = 1, 2, \ldots, n)$ is the coding vector of $\boldsymbol{x}_i$ on $\mathbf{D}$. The input training data can be represented as a linear
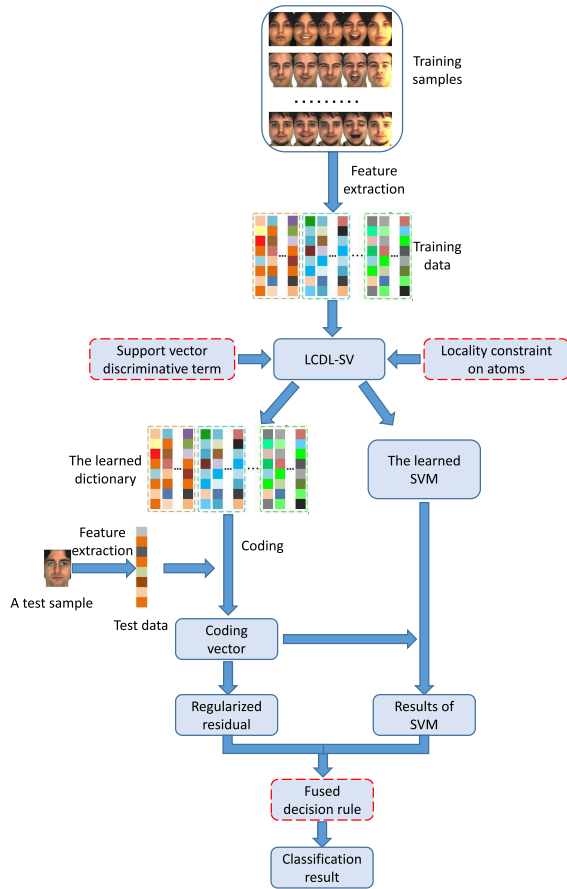

FIGURE 2. Flowchart of our proposed method for classification. In our LCDL-SV, a locality constraint on atoms and a support vector discriminative term are introduced in the formulation of LCDL-SV (red dashed rectangles in the middle part). Moreover, a fused decision strategy (red dashed rectangle at the bottom) which employs both the regularized residual and the learned multi-class SVM is designed to classify the input test sample.

where m is the dimension of vectorized data and n is the total number of training samples, $\mathbf{D} = [\boldsymbol{d}_1, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_K] \in \mathbb{R}^{m \times K}$ is the learned dictionary which has K atoms, $\mathbf{Z} = [z_1, z_2, \ldots, z_n] \in \mathbb{R}^{K \times n}$ is the coding coefficients matrix of $\mathbf{X}$ on the dictionary $\mathbf{D}$.

A. K-SVD AND ITS DISCRIMINATIVE VARIANTS

By generalizing the K-means clustering process, Aharon *et al.* [12] developed K-SVD to learn an overcomplete dictionary that best suits given data. The objective function of K-SVD is formulated as follows,

$$\min_{\mathbf{D},\mathbf{Z}} \|\mathbf{X} - \mathbf{DZ}\|_F^2, \quad \text{s.t. } \|z_i\|_0 \le T_0 \qquad (1)$$

where $\mathbf{D} \in \mathbb{R}^{m \times K}$ is the dictionary that is to be learned, $\mathbf{Z} \in \mathbb{R}^{K \times n}$ is the coding coefficient matrix, and T_0 is a given sparsity level. (1) can be solved by alternatively updating $\mathbf{D}$ and $\mathbf{Z}$. Although K-SVD yields impressive results in image compression and denoising, it is not tailored for classification. To make K-SVD suitable for classification problems, Zhang *et al.* [13] proposed D-KSVD algorithm by introducing the classification error term into the framework

$$x_1 = d_1 z_{1,1} + d_2 z_{2,1} + \ldots + \boxed{d_j z_{j,1}} + \ldots + d_K z_{K,1}$$
$$x_2 = d_1 z_{1,2} + d_2 z_{2,2} + \ldots + \boxed{d_j z_{j,2}} + \ldots + d_K z_{K,2}$$
$$\ldots$$
$$x_n = d_1 z_{1,n} + d_2 z_{2,n} + \ldots + \boxed{d_j z_{j,n}} + \ldots + d_K z_{K,n}$$

FIGURE 3. Linear representation of training data by atoms in the dictionary.

combination of atoms in the dictionary, and the formulation is illustrated in Fig. 3.

In [55] and [22], the j-th row vector of $\mathbf{Z}$ is coined the profile of atom d_j. Thus, $\hat{z}_j = [z_{j,1}, z_{j,2}, \ldots, z_{j,n}]^T (j = 1, 2, \ldots, K)$ is the profile of atom d_j, and the red rectangle in Fig. 3 depicts profile $\hat{z}_j$. So the profile matrix can be constructed as $\mathbf{Z}^T = [\hat{z}_1, \hat{z}_2, \ldots, \hat{z}_K] \in \mathbb{R}^{n \times K}$, which is the transpose matrix of $\mathbf{Z}$. Based on the definition of profile, the linear representations in Fig. 3 can be reformulated as follows,

$$\mathbf{X} = d_1(\hat{z}_1)^T + \ldots + d_j(\hat{z}_j)^T + \ldots + d_K(\hat{z}_K)^T \quad (6)$$

From (6), one can see that the profile $\hat{z}_j$ and atom d_j have a one-to-one correspondence. In this paper, instead of preserving locality information of the original training data, we introduce a locality constraint on the atoms of the learned dictionary, which has proven to be more effective and robust [22]. A viable way to encourage similar atoms to have similar profiles is to minimize the following problem,

$$\frac{1}{2} \sum_{i=1}^{K} \sum_{j=1}^{K} (\hat{z}_i - \hat{z}_j)^2 \mathbf{M}_{ij} \quad (7)$$

where $\mathbf{M}$ is a similarity matrix which can be defined as,

$$\mathbf{M}_{ij} = \begin{cases} \exp(-\dfrac{\|d_i - d_j\|_2}{\delta}), & \text{if } d_j \in k\text{NN}(d_i) \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

where $k\text{NN}(d_i)$ represents the k-nearest neighbors of atom d_i and δ is a parameter. After some deductions, we can obtain the following equivalent formulation of (7),

$$\frac{1}{2} \sum_{i=1}^{K} \sum_{j=1}^{K} (\hat{z}_i - \hat{z}_j)^2 \mathbf{M}_{ij} = tr(\mathbf{Z}^T \mathbf{L} \mathbf{Z}) \quad (9)$$

where $\mathbf{L} = \mathbf{T} - \mathbf{M}$ is a graph Laplacian matrix, $\mathbf{T} = \text{diag}(t_1, \ldots, t_K)$ and $t_i = \sum_{j=1}^{K} \mathbf{M}_{ij}$. We can observe that the graph Laplacian matrix $\mathbf{L}$ is defined on the learned dictionary $\mathbf{D}$. As a result, the graph Laplacian matrix $\mathbf{L}$ is updated due to the fact that $\mathbf{D}$ is updated in the dictionary learning process. Therefore, the graph Laplacian matrix $\mathbf{L}$ can inherit the manifold structure of the training samples.

B. LCDL-SV MODEL
Apart from the locality constraint on the atoms, to facilitate the subsequent classification stage, a support vector

discriminative term is incorporated into our proposed method. The purpose of this term is to enforce the coefficients of different classes to be separated by a max-margin. Intuitively, when the coefficients are separated by a hyperplane, the large margin of different classes can promote the confidence of classification. Moreover, the parameters of SVM (i.e., $\mathbf{U}$ and b) can be learned in our dictionary learning process. The support vector discriminative term has the same formulation in (4). Thus, the objective function of our proposed LCDL-SV is formulated as follows,

$$\min_{\mathbf{D}, \mathbf{Z}, \mathbf{L}, \mathbf{U}, b} \|\mathbf{X} - \mathbf{D}\mathbf{Z}\|_F^2 + 2\lambda_2 \sum_{c=1}^{C} f(\mathbf{Z}, y^c, \boldsymbol{u}_c, b_c)$$
$$+ \lambda_1 tr(\mathbf{Z}^T \mathbf{L} \mathbf{Z}),$$
$$\text{s.t. } \|d_k\|^2 = 1, \quad k = 1, \ldots, K \quad (10)$$

where λ_1 and λ_2 are two balancing parameters.

IV. OPTIMIZATION
In this section, we adopt an alternative strategy to solve the LCDL-SV model. The alternative minimization scheme can be partitioned into the following three sub-problems,

Update $\mathbf{Z}$: Fix the other variables and update $\mathbf{Z}$ by solving the following problem:

$$\min_{\mathbf{Z}} \|\mathbf{X} - \mathbf{D}\mathbf{Z}\|_F^2 + 2\lambda_2 \theta \sum_{c=1}^{C} l(z_i, y_i^c, \boldsymbol{u}_c, b_c)$$
$$+ \lambda_1 tr(\mathbf{Z}^T \mathbf{L} \mathbf{Z}) \quad (11)$$

The optimization of $\mathbf{Z}$ in (11) can be performed by columns, which is formulated as,

$$\min_{z_i} \|\boldsymbol{x}_i - \mathbf{D}z_i\|_2^2 + 2\lambda_2 \theta \sum_{c=1}^{C} l(z_i, y_i^c, \boldsymbol{u}_c, b_c) + \lambda_1 tr(z_i^T \mathbf{L} z_i) \quad (12)$$

To facilitate the optimization process, we employ the quadratic hinge loss function to approximate the original one. The quadratic hinge loss function is defined as,

$$l(z_i, y_i^c, \boldsymbol{u}_c, b_c)$$
$$= \begin{cases} \|y_i^c(\boldsymbol{u}_c^T z_i + b_c) - 1\|_2^2, & y_i^c(\boldsymbol{u}_c^T z_i + b_c) - 1 > 0 \\ 0, & t = 1 \text{ or } y_i^c(\boldsymbol{u}_c^T z_i + b_c) - 1 \leq 0 \end{cases} \quad (13)$$

where t denotes the iteration number. When $t = 1$, (12) is degenerated into the following problem,

$$\min_{z_i} \|\boldsymbol{x}_i - \mathbf{D}z_i\|_2^2 + \lambda_1 tr(z_i^T \mathbf{L} z_i) \quad (14)$$

(14) has the following closed-form solution,

$$z_i = (\mathbf{D}^T \mathbf{D} + \lambda_1 \mathbf{L})^{-1} \mathbf{D}^T \boldsymbol{x}_i \quad (15)$$

When $t \geq 2$, (12) can be rewritten as,

$$\min_{z_i} \|x_i - \mathbf{D}z_i\|_2^2 + \lambda_1 tr(z_i^T \mathbf{L} z_i)$$

$$+ 2\lambda_2 \theta \sum_{c \in \phi} \left\| y_i^c(u_c^T z_i + b_c) - 1 \right\|_2^2 \quad (16)$$

where $\phi = \left\{ c | 1 \leq c \leq C, y_i^c(u_c^T z_i + b_c) - 1 > 0 \right\}$, (16) also has closed-form solution which is given by,

$$z_i = (\mathbf{D}_1)^{-1} \mathbf{D}_2 \quad (17)$$

where $\mathbf{D}_1 = \mathbf{D}^T\mathbf{D} + \lambda_1\mathbf{L} + 2\lambda_2\theta \sum_{c \in \phi} u_c u_c^T$ and $\mathbf{D}_2 = \mathbf{D}^T x_i + 2\lambda_2\theta \sum_{c \in \phi} u_c(y_i^c - b_c)$

Update $\mathbf{D}$ and $\mathbf{L}$: To update $\mathbf{D}$, we fix variables other than $\mathbf{D}$ and minimize (10), which leads to

$$\min_{\mathbf{D}} \|\mathbf{X} - \mathbf{D}\mathbf{Z}\|_F^2, \quad \text{s.t.} \ \|d_k\|^2 = 1, \ k = 1, \dots, K \quad (18)$$

We can see that (18) becomes a least squares problem with quadratic constraints. Here we employ the Lagrange dual function [56] to solve (18), and the Lagrange dual function of (18) is formulated as,

$$g(\delta) = \inf_{\mathbf{D}} \left(\|\mathbf{X} - \mathbf{D}\mathbf{Z}\|_F^2 + \sum_{k=1}^{K} \delta_k(\|d_k\|^2 - 1) \right) \quad (19)$$

where $\delta = [\delta_1, \delta_2, \dots, \delta_K]$ and δ_k is the Lagrange multiplier corresponds to the k-th equality constraint ($\|d_k\|^2 - 1 = 0$). We can define a diagonal matrix Δ whose diagonal element $\Delta_{kk} = \delta_k$, then (19) can be reformulated as,

$$L(\mathbf{D}, \delta) = \|\mathbf{X} - \mathbf{D}\mathbf{Z}\|_F^2 + tr(\mathbf{D}^T\mathbf{D}\Delta) - tr(\Delta) \quad (20)$$

By setting the first-order derivative of (20) to zero, we can obtain the following solution to $\mathbf{D}$,

$$\mathbf{D} = \mathbf{X}\mathbf{Z}^T(\mathbf{Z}\mathbf{Z}^T + \Delta)^{-1} \quad (21)$$

To speed up the optimization process, we discard Δ in the final formulation, which is given by,

$$\mathbf{D} = \mathbf{X}\mathbf{Z}^T(\mathbf{Z}\mathbf{Z}^T)^{-1} \quad (22)$$

When $\mathbf{D}$ is updated, we update the graph Laplacian matrix $\mathbf{L}$ by using (8).

Update $\mathbf{U}$ and b: When the other variables are fixed, (10) with respect to $\mathbf{U}$ and b is boiled down to the following problem,

$$\min_{\mathbf{U}, b} \sum_{c=1}^{C} \left\{ \|u_c\|_2^2 + \theta \sum_{i=1}^{n} l(z_i, y_i^c, u_c, b_c) \right\} \quad (23)$$

(23) is a multi-class linear SVM problem which can be solved by the SVM solver presented in [57]. Due to the fact that the objective function proposed in (10) is non-convex, the algorithm cannot converge to the global minimum. However, satisfactory solutions can be obtained with the decreasing of the objective function. The convergence curve of LCDL-SV on the Extended Yale B database is plotted in Fig. 4. Algorithm 1 outlines the optimization process of our proposed LCDL-SV.
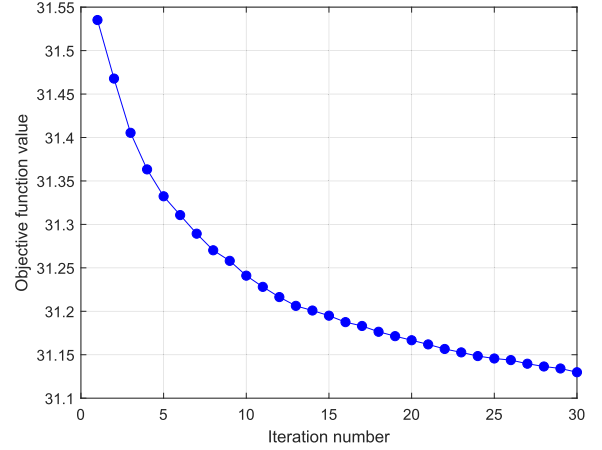

FIGURE 4. Convergence curve of LCDL-SV on the Extended Yale B database.

Algorithm 1 Optimization Procedure of LCDL-SV

Input: Training data matrix $\mathbf{X}$, parameters λ_1, λ_2 and θ
1: Initialize $\mathbf{D}$, $\mathbf{Z}$, $\mathbf{U}$ and b
2: **while** not converged **do**
3:      Construct the graph Laplacian matrix $\mathbf{L}$ by using (8)
4:      **for** $i = 1$ to n **do**
5:          Update z_i by using (15) and (17)
6:      **end for**
7:      Update the dictionary by using (22)
8:      **for** $c = 1$ to C **do**
9:          Update $\mathbf{U}_c$ and b_c by solving (23)
10:     **end for**
11: **end while**
Output: $\mathbf{D}$, $\mathbf{U}$ and b

When the dictionary learning process is completed, we perform classification as follows. For a test sample x_{new}, first we obtain its coding vector by $z = \mathbf{P}x_{new}$, where $\mathbf{P} = (\mathbf{D}^T\mathbf{D} + \eta_1\mathbf{I})^{-1}\mathbf{D}^T$. Then the regularized residual for the c-th class can be obtained by,

$$r_c = \frac{\|x_{new} - \mathbf{D}_c z_c\|_2}{\|z_c\|_2} \quad (24)$$

where $\mathbf{D}_c$ and z_c are the sub-dictionary and coding vector associated with the c-th class, respectively. Moreover, the result produced by the learned SVM classifier is formulated as,

$$s_c = u_c^T z + b_c \quad (25)$$

Finally, the identity of x_{new} is given by,

$$\text{identity}(x_{new}) = \min_c (r_c - \eta_2 s_c) \quad (26)$$

where η_2 is a weighting parameter.

V. EXPERIMENTAL RESULTS
In this section, we conduct experiments on five publicly available databases, i.e., the Extended Yale B

TABLE 1. Optimal parameters of LCDL-SV on each dataset.

	EYaleB	AR	Scene 15	Caltech 101	LFW (VGG16)	LFW (VGG19)
λ_1	1e-3	1e-3	1e-5	1e-1	1e-2	1e-2
λ_2	1e-6	1e-6	1e-4	1e-6	1e-6	1e-6
η_1	1e-2	1e-3	1e-5	1	1e-2	1e-2
η_2	5	50	10	600	600	80

database [58], AR database [59], Scene 15 dataset [60], Caltech 101 dataset [61] and LFW database [62]. We compare LCDL-SV with SRC [63], D-KSVD [13], LC-KSVD [14], FDDL [7], SVGDL [16] and two recently proposed ADL approaches, i.e., CADL [41] and SADL [39]. To validate the effectiveness of employing both the regularized residual and SVM, we also report the results of LCDL-SV only using regularized residual for classification and LCDL-SV only employing the learned multi-class SVM for classification, which are denoted by LCDL-SV (Res) and LCDL-SV (SVM), respectively. Besides the classification accuracy, we also record the training time and testing time of these competing methods in our experiments. SRC directly employs all the training data as the dictionary, thus we do not report its training time. The difference between LCDL-SV (Res), LCDL-SV (SVM) and LCDL-SV lies in the classification rule, thus they have the same training time but different testing time. To construct the similarity matrix in (8), k and δ are set to 1 and 4, respectively. All experiments are conducted with MATLAB R2019b under Windows 10 on a PC equipped with Intel i9-8950HK 2.90 GHz CPU and 32 GB RAM.

There are five parameters in our proposed method, i.e., θ, λ_1, λ_2, η_1 and η_2. In all experiments, θ is set to be 0.2, the other four parameters are determined by cross-validation, and λ_1 and λ_2 are selected from $10^{-6}, 10^{-5}, \ldots, 10^{-1}$. The optimal values on each dataset are recorded in Table 1. For fair comparison, we tune the parameters of competing approaches to achieve their best performance.

A. EXTENDED YALE B

The Extended Yale B database contains 2414 frontal face images of 38 subjects, each person has about 64 images, and some example images are shown in Fig. 5. Following the experimental setting in [16], in our experiments, all images are cropped to 54×48, then they are reduced to a dimension of 300 by PCA. We randomly select 20 images per person as training set and the remaining as testing set. The dictionary has 380 atoms, which corresponds to an average of 10 atoms per subject. Experimental results are summarized in Table 2. We can observe that the proposed LCDL-SV achieves the highest recognition accuracy. Moreover, LCDL-SV is much faster than FDDL in the training phase, and the training time of LCDL-SV is comparable to that of SVGDL. Thanks to the framework of ADL, CADL and SADL are efficient on this dataset. Due to the jointly learning dictionary and multi-class SVM, the testing time of SVGDL and LCDL-SV (SVM) is less than our proposed LCDL-SV. Nevertheless, by fusing the regularized residual

FIGURE 5. Example images from the Extended Yale B database.

TABLE 2. Recognition accuracy (%) and computing time on the Extended Yale B database.

Methods	Accuracy	Training time(s)	Testing time(s)
SRC [63]	90.0	No Need	3.2
D-KSVD [13]	75.3	7.8	0.2
LC-KSVD [14]	90.6	11.5	0.2
FDDL [7]	91.9	592.6	3.4
SVGDL [16]	96.1	57.8	0.01
CADL [41]	96.3	23.8	0.01
SADL [39]	95.2	27.2	0.01
LCDL-SV (Res)	91.5	57.2	0.8
LCDL-SV (SVM)	97.1	57.2	0.06
LCDL-SV	**97.8**	57.2	0.9

FIGURE 6. Example images from the AR database.

and the learned multi-class SVM, LCDL-SV outperforms all the competing approaches in recognition accuracy, and it is more efficient than SRC and FDDL in terms of testing time. It should be noted that, in [39], the reported accuracy of SADL and SRC is 96.35% and 96.51%, respectively. The differences lie in the following two aspects. On the one hand, in [39], each image of 192×168 pixels is projected onto a 504-dimensional vector by random projection, while we use the cropped image of 54×48 pixels and employ PCA to reduce the image to a dimension of 300. On the other hand, half of the images (i.e., 32 images) per subject are used for training in [39], while 20 images per person are employed for training in our experiments.

B. AR

The AR database has more than 4000 face images of 126 subjects with variations in facial expression, illumination conditions and occlusions. Fig. 6 shows example images from the database. In our experiments, we use a subset of 2600 images of 50 male and 50 female subjects from the database. As in [14], each 165×120 face image is projected onto a

TABLE 3. Recognition accuracy (%) and computing time on the AR database.

Methods	Accuracy	Training time(s)	Testing time(s)
SRC [63]	66.5	No Need	2.3
D-KSVD [13]	88.8	25.3	0.08
LC-KSVD [14]	93.7	38.2	0.1
FDDL [7]	97.4	822.4	2.9
SVGDL [16]	98.8	358.9	0.01
CADL [41]	98.5	288.3	0.01
SADL [39]	97.2	183.1	0.01
LCDL-SV (Res)	94.7	360.3	0.6
LCDL-SV (SVM)	99.0	360.3	0.01
LCDL-SV	**99.2**	360.3	0.62

TABLE 4. Recognition accuracy (%) and computing time on the Scene 15 dataset.

Methods	Accuracy	Training time(s)	Testing time(s)
SRC [63]	91.8	No Need	3.5
D-KSVD [13]	89.1	54.1	0.4
LC-KSVD [14]	92.9	72.2	0.4
FDDL [7]	97.5	2470.5	208.2
SVGDL [16]	98.4	159.0	0.1
CADL [41]	98.6	5365.8	0.67
SADL [39]	98.5	314.7	0.2
LCDL-SV (Res)	97.8	159.2	6.5
LCDL-SV (SVM)	98.4	159.2	0.4
LCDL-SV	**99.0**	159.2	6.7

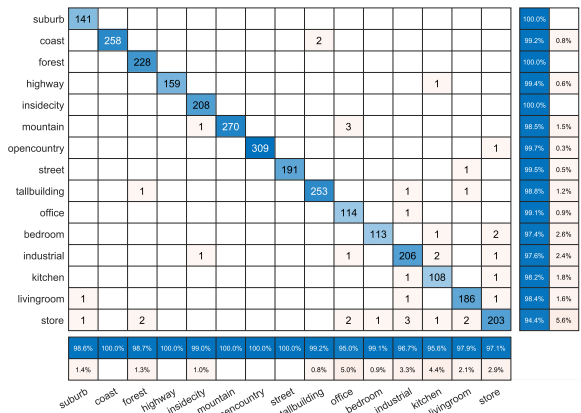**FIGURE 7.** Example images from the Scene 15 dataset.

540-dimensional vector by random projection. For each person, 20 images are randomly selected for training and the remaining for testing. The learned dictionary has 500 atoms, namely five atoms per person. Table 3 lists the recognition accuracy and computing time of all compared methods. Notice that 20 atoms per class are exploited in [41], while only five atoms per class are used in our experiments. One can see that LCDL-SV has the best performance in recognition accuracy and is more efficient than FDDL in both training and testing phases. Moreover, on this database and the Extended Yale B database, LCDL-SV (SVM) achieves better results than SVGDL, which demonstrates that locality constraint of atoms does promote the classification performance of SDL approaches on these two face databases.

C. SCENE 15

Scene 15 dataset contains 15 natural scene categories, which comprises a wide range of indoor and outdoor scenes, such as bedroom, office and mountain, example images from this dataset are shown in Figure 7. For fair comparison, we employ the 3000-dimensional SIFT-based features used in LC-KSVD [14]. Following the common experimental settings, we randomly select 100 images per category as training data and use the remaining for testing. The learned dictionary has 450 atoms. Experimental results are shown in Table 4. The recognition accuracy of LCDL-SV is 99.0%, which outperforms all the compared methods. CADL performs the second best on this dataset, followed by SADL and SVGDL. Moreover, LCDL-SV is 15 times faster than FDDL in the training stage. The confusion matrix for LCDL-SV is depicted in Fig. 8, in which diagonal elements are well-marked. It can be seen that LCDL-SV attains 100% recognition accuracy for the categories of suburb, forest and inside-city.

FIGURE 8. Confusion matrix on the Scene 15 dataset.

D. CALTECH 101

Caltech101 dataset is a widely used dataset for object classification with 102 classes (i.e., 101 object classes and one background class). The number of images in each category is unbalanced, varying from 31 to 800, and in total this dataset contains 9144 images. For fair comparison, we also employ the 3000-dimensional SIFT-based features used in LC-KSVD [14]. Following the common experimental protocol, we randomly choose 5, 10, 15, 20, 25, and 30 samples per category for training and test on the remaining images. We repeat this process 10 times with different splits of training and test images and record the averaged classification accuracy. Table 5 summarizes the classification results and Table 6 lists the training time and testing time when 5 samples per category are used for training. As can be seen from Table 5, LCDL-SV consistently outperforms the other competing approaches in all cases. Compared with SVGDL, LCDL-SV (SVM) does not always achieve better accuracy (e.g., when the number of training samples per category is 25). This indicates that using only the locality constraint cannot guarantee the best performance. Combining with the proposed classification scheme, LCDL-SV exhibits its advantage over other dictionary learning approaches. From Table 6, we can observe that the training time of LCDL-SV is only one-sixteenth of that of FDDL and LCDL-SV is faster than SRC in the testing phase.

TABLE 5. Recognition accuracy (%) on the Caltech 101 dataset.

number of train. samp.	5	10	15	20	25	30
SRC [63]	48.8	60.1	64.9	67.7	69.2	70.7
D-KSVD [13]	49.6	59.5	65.1	68.6	71.1	73.0
LC-KSVD [14]	54.0	63.1	67.7	70.5	72.3	73.6
FDDL [7]	53.6	63.6	66.8	69.8	71.7	73.1
SVGDL [16]	55.3	64.3	69.6	72.3	**75.1**	76.7
CADL [41]	55.6	63.9	65.7	68.1	70.1	75.0
SADL [39]	46.6	59.4	62.4	68.4	70.3	74.4
LCDL-SV (Res)	52.6	61.1	62.2	63.9	62.2	60.7
LCDL-SV (SVM)	56.9	65.6	69.2	72.6	74.5	76.5
LCDL-SV	**57.0**	**65.9**	**69.7**	**73.1**	**75.1**	**76.8**

FIGURE 9. Example images from the Caltech 101 dataset.

TABLE 6. Computing time on the Caltech 101 dataset when 5 samples per category are used for training.

Methods	Training time(s)	Testing time(s)
SRC [63]	No Need	33.9
D-KSVD [13]	35.3	1.3
LC-KSVD [14]	43.3	1.6
FDDL [7]	2112.7	177.2
SVGDL [16]	162.3	0.3
CADL [41]	183.6	1.2
SADL [39]	176.2	0.2
LCDL-SV (Res)	139.4	23.2
LCDL-SV (SVM)	139.4	2.1
LCDL-SV	139.4	24.2

E. DEEP FEATURES

In this subsection, a subset of LFW database is used to evaluate our proposed LCDL-SV and other competing approaches on deep features. This subset contains 1251 images of 86 subjects, each person has 11-20 images. All images are converted to grayscale images and cropped and resized to 32×32, some example images are shown in Fig. 10. Five images per subject are randomly selected as training samples and the remaining as test samples. The pre-trained VGG16 and VGG19 [64] models are employed to extract deep features, and FC6 in both VGG16 and VGG19 is used for feature extraction. The dimension of deep features extracted by VGG16 and VGG19 is 4096. In order to obtain more compact representations, we apply principal component analysis (PCA) to the 4096-dimensional features (keeping 98% of the variance) and the dimensions of reduced features for VGG16 and VGG19 are 270 and 264, respectively. Finally, the reduced features are fed into the compared approaches. Experimental results are shown in Table 7, and the training time and testing time are recorded for the VGG19 feature. From Table 7, we can see that LCDL-SV delivers the best result on both

FIGURE 10. Example images from the LFW database.

TABLE 7. Recognition accuracy (%) and computing time on the LFW database.

Methods	VGG16	VGG19	Training time(s)	Testing time(s)
SRC [63]	42.4	36.4	No Need	2.3
D-KSVD [13]	35.7	35.8	5.9	0.2
LC-KSVD [14]	41.5	36.9	8.7	0.1
FDDL [7]	45.4	39.4	203.7	2.3
SVGDL [16]	46.9	42.8	85.5	0.01
CADL [41]	46.5	41.2	26.8	0.01
SADL [39]	41.0	35.1	7.1	0.01
LCDL-SV (Res)	25.6	22.9	66.0	0.5
LCDL-SV (SVM)	47.6	45.3	66.0	0.03
LCDL-SV	**48.0**	**46.5**	66.0	0.6

the VGG16 and VGG19 features. This demonstrates that LCDL-SV is not only superior to its competing methods on hand-crafted features, but on the deep features as well. Similar to the observations on the other datasets used in our experiments, LCDL-SV is more efficient than FDDL in terms of training and testing time.

F. PARAMETER SENSITIVITY ANALYSIS

As mentioned earlier, five parameters should be determined in our proposed method, i.e., θ, λ_1, λ_2, η_1 and η_2. For the two parameters λ_1 and λ_2, we find that relatively small values (e.g., 1e-5) can guarantee our proposed method to achieve satisfactory results for pattern classification tasks. η_1 is used to obtain the coding coefficients of test samples and it is usually set to 1e-3. For diverse datasets, η_2 has a relatively wide range in fusing the regularized residual and the results of multi-class SVM. Therefore, the above observations can be treated as a rule of thumb for selecting parameters of the proposed LCDL-SV. To investigate the sensitivity of parameters, we carry out experiments on the Extended Yale B database, and the experimental settings are the same as that in Section V-A. When analyzing one parameter, we fix the other two parameters. Firstly, we fix the parameters λ_2 and η_2, and examine how the performance changes with varying λ_1. Fig. 11 (a) plots the recognition accuracy with varying λ_1. Similarly, Figs. 11 (b) and 11 (c) plot the results of varying λ_2 and η_2. As can be seen from Fig. 11 (a), when the value of λ_1 increases from 10^{-5} to 10^{-3}, the accuracy of LCDL-SV is gradually increasing. However, the performance of LCDL-SV will degrade when the value
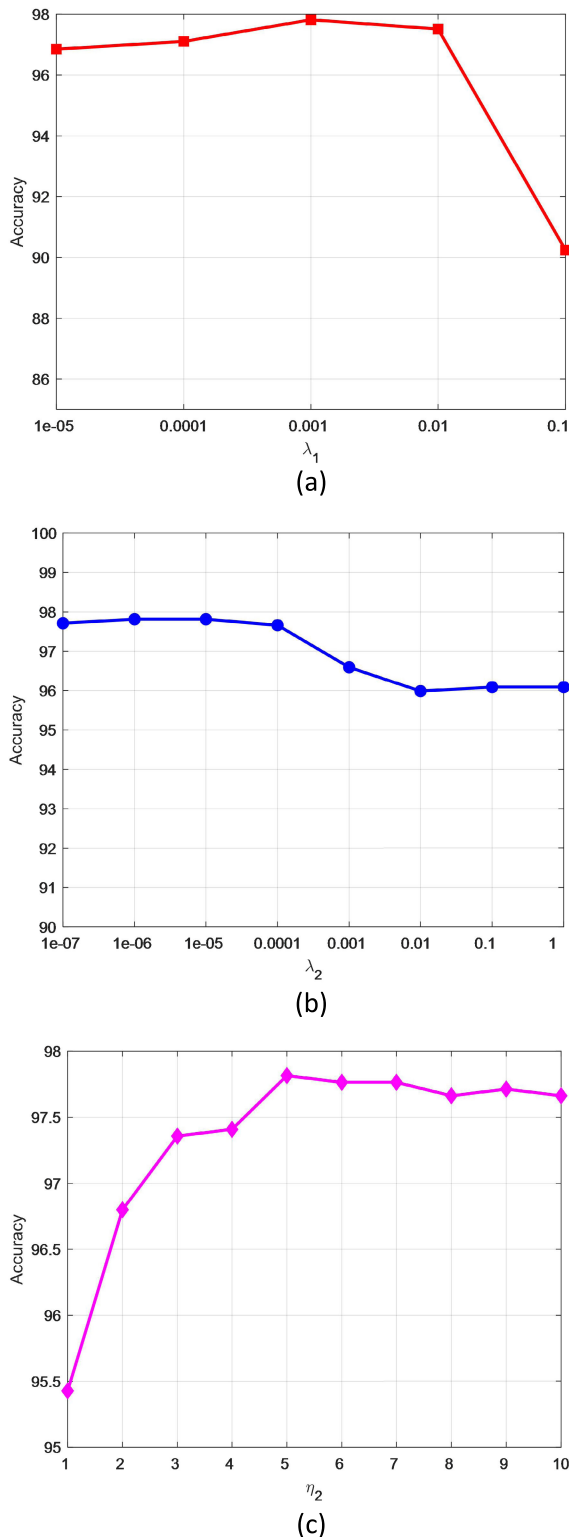
FIGURE 11. Classification accuracy (%) of LCDL-SV with varying parameters on the Extended Yale B database. (a) λ_1 changes when λ_2 and η_2 are fixed to 1e-6 and 5, respectively. (b) λ_2 changes by fixing $\lambda_1 = 1e\text{-}3$ and $\eta_2 = 5$ and (c) η_2 varies when λ_1 and λ_2 are fixed to 1e-3 and 1e-6, respectively.

of λ_1 is larger than 0.01. From Fig. 11 (b), we can see that LCDL-SV achieves stable performance when the value of λ_2

is in the range of $[10^{-7}, 10^{-4}]$. With the increasing of λ_2, the performance drops to some extent. A larger λ_2 will reduce the discriminative ability of the support vector term, leading to degenerated performance. From Fig. 11 (c), one can see that accuracy of LCDL-SV has an increase with η_2 from 1 to 5, and then has a decline when η_2 continues increasing. On the Extended Yale B database, LCDL-SV has the best performance when η_2 is set to be 5.

VI. CONCLUSION
In this paper, we propose a locality constraint dictionary learning with support vector discriminative term (LCDL-SV) for pattern classification. In contrast with traditional methods in which the graph Laplacian matrix is derived from the original training data, we preserve the locality of atoms on the basis of the learned dictionary. By introducing a support vector discriminative term into the formulation of LCDL-SV, a classifier can be jointly learned in our dictionary learning procedures. More importantly, the regularized residual and multi-class SVM are simultaneously employed to classify the test samples. Experimental results on face databases, scene dataset and object dataset validate the effectiveness of LCDL-SV, and it outperforms some state-of-the-art dictionary learning approaches, e.g., FDDL, SVGDL, CADL, and SADL.

Discriminative analysis dictionary learning methods have aroused considerable interest due to their efficiency and efficacy. In future work, we will develop new discriminative ADL approach and apply it to other classification scenarios, such as action recognition and texture classification.

REFERENCES
[1] X. Song, Y. Chen, Z.-H. Feng, G. Hu, T. Zhang, and X.-J. Wu, "Collaborative representation based face classification exploiting block weighted LBP and analysis dictionary learning," *Pattern Recognit.*, vol. 88, pp. 127–138, Apr. 2019.

[2] H. Li and X.-J. Wu, "Multi-focus image fusion using dictionary learning and low-rank representation," in *Proc. Int. Conf. Image Graph.*, Shanghai, China, 2017, pp. 675–686.

[3] S. Li, M. Shao, and Y. Fu, "Person re-identification by cross-view multi-level dictionary learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 12, pp. 2963–2977, Dec. 2018.

[4] X.-Y. Jing, X. Zhu, F. Wu, X. You, Q. Liu, D. Yue, R. Hu, and B. Xu, "Super-resolution person re-identification with semi-coupled low-rank discriminant dictionary learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, Jun. 2015, pp. 695–704.

[5] 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.*, San Francisco, CA, USA, Jun. 2010, pp. 3501–3508.

[6] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Metaface learning for sparse representation based face recognition," in *Proc. IEEE Int. Conf. Image Process.*, Hong Kong, Sep. 2010, pp. 1601–1604.

[7] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Fisher discrimination dictionary learning for sparse representation," in *Proc. Int. Conf. Comput. Vis.*, Barcelona, Spain, Nov. 2011, pp. 543–550.

[8] B.-D. Liu, B. Shen, L. Gui, Y.-X. Wang, X. Li, F. Yan, and Y.-J. Wang, "Face recognition using class specific dictionary learning for sparse representation and collaborative representation," *Neurocomputing*, vol. 204, pp. 198–210, Sep. 2016.

[9] N. Akhtar, A. Mian, and F. Porikli, "Joint discriminative Bayesian dictionary and classifier learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Honolulu, HI, USA, Jun. 2017, pp. 3919–3928.

[10] X. Wang, Y. Lid, S. You, H. Li, and S. Wang, "Unidirectional representation based efficient dictionary learning," *IEEE Trans. Circuits Syst. Video Technol.*, to be published, doi: 10.1109/tcsvt.2018.2886600.

[11] J. Ling, Z. Chen, and F. Wu, "Class-oriented discriminative dictionary learning for image classification," *IEEE Trans. Circuits Syst. Video Technol.*, to be published, doi: 10.1109/tcsvt.2019.2918852.

[12] 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.

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

[14] Z. Jiang, Z. Lin, and L. S. Davis, "Learning a discriminative dictionary for sparse coding via label consistent K-SVD," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* Colorado Springs, CO, USA: Springer, 2011, pp. 1697–1704.

[15] S. Shao, Y.-J. Wang, B.-D. Liu, W. Liu, and R. Xu, "Label embedded dictionary learning for image classification," 2019, *arXiv:1903.03087*. [Online]. Available: https://arxiv.org/abs/1903.03087

[16] S. Cai, W. Zuo, L. Zhang, X. Feng, and P. Wang, "Support vector guided dictionary learning," in *Proc. Eur. Conf. Comput. Vis.*, Zürich, Switzerland, 2014, pp. 624–639.

[17] D. Zhang, P. Liu, K. Zhang, H. Zhang, Q. Wang, and X. Jing, "Class relatedness oriented-discriminative dictionary learning for multiclass image classification," *Pattern Recognit.*, vol. 59, pp. 168–175, Nov. 2016.

[18] A. Quattoni, X. Carreras, M. Collins, and T. Darrell, "An efficient projection for $\ell_{1,\infty}$ regularization," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, Montreal, QC, Canada, 2009, pp. 857–864.

[19] Y. Quan, Y. Xu, Y. Sun, and Y. Huang, "Supervised dictionary learning with multiple classifier integration," *Pattern Recognit.*, vol. 55, pp. 247–260, Jul. 2016.

[20] Z. Dong, M. Pei, and Y. Jia, "Orthonormal dictionary learning and its application to face recognition," *Image Vis. Comput.*, vol. 51, pp. 13–21, Jul. 2016.

[21] H. Min, M. Liang, R. Luo, and J. Zhu, "Laplacian regularized locality-constrained coding for image classification," *Neurocomputing*, vol. 171, pp. 1486–1495, Jan. 2016.

[22] Z. Li, Z. Lai, Y. Xu, J. Yang, and D. Zhang, "A locality-constrained and label embedding dictionary learning algorithm for image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 2, pp. 278–293, Feb. 2017.

[23] J. Song, X. Xie, G. Shi, and W. Dong, "Exploiting class-wise coding coefficients: Learning a discriminative dictionary for pattern classification," *Neurocomputing*, vol. 321, pp. 114–125, Dec. 2018.

[24] Z. Li, Z. Zhang, Z. Fan, and J. Wen, "An interactively constrained discriminative dictionary learning algorithm for image classification," *Eng. Appl. Artif. Intell.*, vol. 72, pp. 241–252, Jun. 2018.

[25] S. Kong and D. Wang, "A dictionary learning approach for classification: Separating the particularity and the commonality," in *Proc. Eur. Conf. Comput. Vis.*, Florence, Italy, 2012, pp. 186–199.

[26] S. Gao, I. W.-H. Tsang, and Y. Ma, "Learning category-specific dictionary and shared dictionary for fine-grained image categorization," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 623–634, Feb. 2014.

[27] Y. Sun, Q. Liu, J. Tang, and D. Tao, "Learning discriminative dictionary for group sparse representation," *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 3816–3828, Sep. 2014.

[28] X. Wang and Y. Gu, "Cross-label suppression: A discriminative and fast dictionary learning with group regularization," *IEEE Trans. Image Process.*, vol. 26, no. 8, pp. 3859–3873, Aug. 2017.

[29] G. Lin, M. Yang, J. Yang, L. Shen, and W. Xie, "Robust, discriminative and comprehensive dictionary learning for face recognition," *Pattern Recognit.*, vol. 81, pp. 341–356, Sep. 2018.

[30] T. H. Vu and V. Monga, "Fast low-rank shared dictionary learning for image classification," *IEEE Trans. Image Process.*, vol. 26, no. 11, pp. 5160–5175, Nov. 2017.

[31] Y. Rong, S. Xiong, and Y. Gao, "Low-rank double dictionary learning from corrupted data for robust image classification," *Pattern Recognit.*, vol. 72, pp. 419–432, Dec. 2017.

[32] H. Du, L. Ma, G. Li, and S. Wang, "Low-rank graph preserving discriminative dictionary learning for image recognition," *Knowl.-Based Syst.*, vol. 187, Jan. 2020, Art. no. 104823, doi: 10.1016/j.knosys.2019.06.031.

[33] Y. Xu, Z. Li, J. Yang, and D. Zhang, "A survey of dictionary learning algorithms for face recognition," *IEEE Access*, vol. 5, pp. 8502–8514, 2017.

[34] R. Rubinstein, T. Peleg, and M. Elad, "Analysis K-SVD: A dictionary-learning algorithm for the analysis sparse model," *IEEE Trans. Signal Process.*, vol. 61, no. 3, pp. 661–677, Feb. 2013.

[35] S. Shekhar, V. M. Patel, and R. Chellappa, "Analysis sparse coding models for image-based classification," in *Proc. IEEE Int. Conf. Image Process.*, Paris, France, Oct. 2014, pp. 5207–5211.

[36] J. Guo, Y. Guo, X. Kong, M. Zhang, and R. He, "Discriminative analysis dictionary learning," in *Proc. AAAI Conf. Artif. Intell.*, Phoenix, AZ, USA, 2016, pp. 1617–1623.

[37] J. Wang, Y. Guo, J. Guo, M. Li, and X. Kong, "Synthesis linear classifier based analysis dictionary learning for pattern classification," *Neurocomputing*, vol. 238, pp. 103–113, May 2017.

[38] Q. Wang, Y. Guo, J. Guo, and X. Kong, "Synthesis K-SVD based analysis dictionary learning for pattern classification," *Multimedia Tools Appl.*, vol. 77, no. 13, pp. 17023–17041, Jul. 2018.

[39] W. Tang, A. Panahi, H. Krim, and L. Dai, "Analysis dictionary learning based classification: Structure for robustness," *IEEE Trans. Image Process.*, vol. 28, no. 12, pp. 6035–6046, Dec. 2019.

[40] J. Maggu, H. K. Aggarwal, and A. Majumdar, "Label-consistent transform learning for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 9, pp. 1502–1506, Sep. 2019.

[41] J. Wang, Y. Guo, J. Guo, X. Luo, and X. Kong, "Class-aware analysis dictionary learning for pattern classification," *IEEE Signal Process. Lett.*, vol. 24, no. 12, pp. 1822–1826, Dec. 2017.

[42] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Projective dictionary pair learning for pattern classification," in *Proc. Adv. Neural Inf. Process. Syst.*, Montreal, QC, Canada, 2014, pp. 793–801.

[43] B. Chen, J. Li, B. Ma, and G. Wei, "Discriminative dictionary pair learning based on differentiable support vector function for visual recognition," *Neurocomputing*, vol. 272, pp. 306–313, Jan. 2018.

[44] Z. Zhang, F. Li, T. W. S. Chow, L. Zhang, and S. Yan, "Sparse codes auto-extractor for classification: A joint embedding and dictionary learning framework for representation," *IEEE Trans. Signal Process.*, vol. 64, no. 14, pp. 3790–3805, Jul. 2016.

[45] Z. Zhang, W. Jiang, Z. Zhang, S. Li, G. Liu, and J. Qin, "Scalable block-diagonal locality-constrained projective dictionary learning," 2019, *arXiv:1905.10568*. [Online]. Available: https://arxiv.org/abs/1905.10568

[46] M. Yang, H. Chang, and W. Luo, "Discriminative analysis-synthesis dictionary learning for image classification," *Neurocomputing*, vol. 219, pp. 404–411, Jan. 2017.

[47] H. Chang, H. Tang, F. Zhang, Y. Chen, and H. Zheng, "Graph-regularized discriminative analysis-synthesis dictionary pair learning for image classification," *IEEE Access*, vol. 7, pp. 55398–55406, 2019.

[48] Z. Zhang, W. Jiang, J. Qin, L. Zhang, F. Li, M. Zhang, and S. Yan, "Jointly learning structured analysis discriminative dictionary and analysis multiclass classifier," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 8, pp. 3798–3814, Aug. 2018.

[49] Y. Wei, L. Jiao, F. Liu, S. Yang, Q. Wu, and G. Sanga, "Fast DDL classification for SAR images with $l_{1,\infty}$ constraint," *IEEE Access*, vol. 7, pp. 68991–69006, 2019.

[50] Z. Li, Z. Zhang, J. Qin, S. Li, and H. Cai, "Low-rank analysis–synthesis dictionary learning with adaptively ordinal locality," *Neural Netw.*, vol. 119, pp. 93–112, Nov. 2019.

[51] F. Wu, X.-Y. Jing, X. You, D. Yue, R. Hu, and J.-Y. Yang, "Multi-view low-rank dictionary learning for image classification," *Pattern Recognit.*, vol. 50, pp. 143–154, Feb. 2016.

[52] F. Wu, X.-Y. Jing, and D. Yue, "Multi-view discriminant dictionary learning via learning view-specific and shared structured dictionaries for image classification," *Neural Process. Lett.*, vol. 45, no. 2, pp. 649–666, 2017.

[53] F. Ma, X. Zhu, Q. Liu, C. Song, X.-Y. Jing, and D. Ye, "Multi-view coupled dictionary learning for person re-identification," *Neurocomputing*, vol. 348, pp. 16–26, Jul. 2019.

[54] F. Wu, X. Dong, L. Han, X.-Y. Jing, and Y.-M. Ji, "Multi-view synthesis and analysis dictionaries learning for classification," *IEICE Trans. Inf. Syst.*, vol. E102-D, no. 3, pp. 659–662, Mar. 2019.

[55] M. Sadeghi, M. Babaie-Zadeh, and C. Jutten, "Learning overcomplete dictionaries based on atom-by-atom updating," *IEEE Trans. Signal Process.*, vol. 62, no. 4, pp. 883–891, Feb. 2014.

[56] M. Zheng, J. Bu, C. Chen, C. Wang, L. Zhang, G. Qiu, and D. Cai, "Graph regularized sparse coding for image representation," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1327–1336, May 2011.

[57] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Miami, FL, USA, Jun. 2009, pp. 1794–1801.

[58] K.-C. Lee, J. Ho, and D. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 684–698, May 2005.

[59] A. M. Martínez and R. Benavente, "The AR face database," Univ. Autònoma de Barcelona, Bellaterra, Spain, CVC Tech. Rep. 24, Jun. 1998. [Online]. Available: http://www2.ece.ohiostate.edu/~aleix/ARdatabase.html

[60] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, New York, NY, USA, Jun. 2006, pp. 2169–2178.

[61] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories," in *Proc. Conf. Comput. Vis. Pattern Recognit. Workshop*, Washington, DC, USA, Jun./Jul. 2004, p. 178.

[62] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," in *Proc. Workshop Faces Real-Life Images, Detection, Alignment, Recognit.*, Marseille, France, 2008, pp. 1–14.

[63] 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.

[64] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: https://arxiv.org/abs/1409.1556

XIAO-JUN WU received the B.S. degree in mathematics from Nanjing Normal University, Nanjing, China, in 1991, and the M.S. and Ph.D. degrees in pattern recognition and intelligent system from the Nanjing University of Science and Technology, Nanjing, in 1996 and 2002, respectively. From 1996 to 2006, he was with the School of Electronics and Information, Jiangsu University of Science and Technology, where he was involved in teaching and promoted to a Professor. From 1999 to 2000, he was a Fellow of the International Institute for Software Technology, United Nations University. From 2003 to 2004, he was a Visiting Researcher with the Centre for Vision, Speech, and Signal Processing, University of Surrey, U.K. Since 2006, he has been with the School of Information Engineering, Jiangnan University, where he is currently a Professor of computer science and technology. He has published over 200 articles. His current research interests include pattern recognition, computer vision, and computational intelligence. He received the Most Outstanding Postgraduate Award from the Nanjing University of Science and Technology.

HE-FENG YIN received the B.S. degree from the School of Computer Science and Technology, Xuchang University, Xuchang, China, in 2011. He is currently pursuing the Ph.D. degree with the School of IoT Engineering, Jiangnan University, Wuxi, China. He was a Visiting Ph.D. Student with the Centre for Vision, Speech and Signal Processing (CVSSP), University of Surrey, under the supervision of Prof. Josef Kittler. His research interests include representation based classification methods, dictionary learning, and low rank representation.

SU-GEN CHEN received the B.S. degree in mathematics and applied mathematics from Anqing Normal University, Anqing, Anhui, in 2004, the M.S. degree in computational mathematics from the Hefei University of Technology, Hefei, Anhui, in 2009, and the Ph.D. degree in control science and engineering from Jiangnan University, Wuxi, Jiangsu, in 2016. Since 2015, he has been an Associate Professor with the School of Mathematics and Computational Science, Anqing Normal University. His research interests include pattern recognition and intelligent systems, and machine learning.

• • •