



2017

Semi-supervised dictionary learning with label propagation for image classification

Lin Chen

College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China.

Meng Yang

College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China. School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China. Key Laboratory of Machine Intelligence and Advanced Computing (Sun Yat-sen University), Ministry of Education, China.

Follow this and additional works at: <https://tsinghuauniversitypress.researchcommons.org/computational-visual-media>

 Part of the [Computational Engineering Commons](#), [Computer-Aided Engineering and Design Commons](#), [Graphics and Human Computer Interfaces Commons](#), and the [Software Engineering Commons](#)

Recommended Citation

Lin Chen, Meng Yang. Semi-supervised dictionary learning with label propagation for image classification. *Computational Visual Media* 2017, 3(1): 83-94.

This Research Article is brought to you for free and open access by Tsinghua University Press: Journals Publishing. It has been accepted for inclusion in *Computational Visual Media* by an authorized editor of Tsinghua University Press: Journals Publishing.

Semi-supervised dictionary learning with label propagation for image classification

Lin Chen¹, Meng Yang^{1,2,3} (✉)

© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract Sparse coding and supervised dictionary learning have rapidly developed in recent years, and achieved impressive performance in image classification. However, there is usually a limited number of labeled training samples and a huge amount of unlabeled data in practical image classification, which degrades the discrimination of the learned dictionary. How to effectively utilize unlabeled training data and explore the information hidden in unlabeled data has drawn much attention of researchers. In this paper, we propose a novel discriminative semi-supervised dictionary learning method using label propagation (SSD-LP). Specifically, we utilize a label propagation algorithm based on class-specific reconstruction errors to accurately estimate the identities of unlabeled training samples, and develop an algorithm for optimizing the discriminative dictionary and discriminative coding vectors simultaneously. Extensive experiments on face recognition, digit recognition, and texture classification demonstrate the effectiveness of the proposed method.

Keywords semi-supervised learning; dictionary learning; label propagation; image classification

1 Introduction

In recent years, sparse representation has gained

much interest in the computer vision field [1, 2] and has been widely applied to image restoration [3, 4], image compression [5, 6], and image classification [7–11]. The success of sparse representation is partially because natural images can be generally and sparsely coded by structural primitives (e.g., edges and line segments) and the images or signals can be represented sparsely by dictionary atoms from the same class.

In the task of image classification based on sparse representation, signals need to be encoded over a dictionary (i.e., a set of representation bases) with some sparsity constraint. The dictionary, which encodes the testing sample, can directly consist of the training samples themselves. For example, Wright et al. [12] firstly constructed a dictionary by using the training samples of all classes, then coded the test sample with this dictionary, and finally classified the test sample into the class with the minimal class-specific representation residual. So-called sparse representation based classification (SRC) [12] has achieved impressive performance in face recognition. However, the number of dictionary atoms used in SRC can be quite high, resulting in a large computational burden in calculating the coding vector. What is more, the discriminative information hidden in training samples cannot be exploited fully. To overcome the above problems, the problem of how to learn an effective dictionary from training data has been widely studied.

Dictionary learning methods can be divided into three main categories: unsupervised [13], supervised [14–17], and semi-supervised [11, 18–23]. K-SVD [13] is a representative unsupervised dictionary learning model, which is widely applied to image restoration tasks. Since no label information is exploited in the phase of dictionary learning,

1 College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China. E-mail: L. Chen, chen.lin@email.szu.edu.cn; M. Yang, yang.meng@szu.edu.cn (✉).

2 School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China.

3 Key Laboratory of Machine Intelligence and Advanced Computing (Sun Yat-sen University), Ministry of Education, China.

Manuscript received: 2016-09-04; accepted: 2016-12-20

unsupervised dictionary learning methods are useful for data reconstruction, but not advantageous for classification tasks.

Based on the relationship between dictionary atoms and class labels, prevailing supervised dictionary learning methods can be divided into three categories: shared, class-specific, and hybrid. In the first case, discrimination provided by shared dictionary learning is typically explored by jointly learning a dictionary and a classifier over the coding coefficients [9, 10]. Using the learned shared dictionary, the generated coding coefficients, which are expected to be discriminative, are used for classification. In class-specific dictionary learning, each dictionary atom is predefined to correspond to a unique class label so that the class-specific reconstruction error can be used for classification [14, 24]. However, the learned dictionary can be very large when there are many classes. In order to take advantage of the powerful class-specific representation ability, and to reduce the coherence between different sub-dictionaries, the hybrid dictionary learning [15, 25, 26] combines shared dictionary atoms and class-specific dictionary atoms.

Sufficient labeled training data and high quality training images are necessary for good performance in supervised dictionary learning algorithms. However, it is expensive and difficult to obtain the labeled training data due to the vast human effort involved. On the other hand, there are abundant unlabeled images that can be collected easily from public image datasets. Therefore, semi-supervised dictionary learning, which effectively utilizes these unlabeled samples to enhance dictionary learning, has attracted extensive research.

In recent years, semi-supervised learning methods have been widely studied [27–31]. One classical semi-supervised learning method is co-training [29] which utilizes multi-view features to retrain the classifiers to obtain better performance. In co-training, the multi-view features need to be conditionally independent so that one classifier can confidently select unlabeled samples for the other classifier. Another important semi-supervised learning method is the graph-based method [27]. In classification, graph-based semi-supervised learning methods can readily explore the class information in

unlabeled training data via a small amount of labeled data. A representative method based on graphs is label propagation (LP), which has been widely used in image classification and ranking. Label propagation algorithms [27, 28, 33–35] perform class estimation of unlabeled samples by propagating label information from labeled data to unlabeled data. This is done by constructing a weight matrix (or affinity matrix) based on the distance between any two samples. The basic assumption of LP algorithms is that if the weight linking two samples is high, they are likely to belong to the same class.

Semi-supervised dictionary learning [11, 18, 19, 21–23, 36] has gained considerable interest in the past several years. In semi-supervised dictionary learning, the issue of whether the unlabeled samples can be accurately estimated for use as labeled samples for training is very important. For instance, a shared dictionary and a classifier may be jointly learned by estimating the class confidence of unlabeled samples [18]. In Ref. [23], the unlabeled samples are utilized to learn a discriminative dictionary by preserving the geometrical structure of both the labeled and unlabeled data. However, the class-specific reconstruction error which carries strong discriminative ability cannot be utilized to estimate the identities of unlabeled samples in the shared dictionary. A semi-supervised class-specific dictionary has also been learned in Ref. [19]. However, its model is a little complex due to many regularizations.

By combining the label information of the labeled samples and reconstruction error of unlabeled samples over all classes, the identities of the unlabeled training samples can be estimated more accurately. In this paper, we propose a novel semi-supervised dictionary model with label propagation. In our proposed model, we design an improved label propagation algorithm to evaluate the probabilities of unlabeled data belonging to a specific class. Specifically, the proposed label propagation algorithm is based on the powerful class-specific representation provided by the reconstruction error of unlabeled samples for each sub-dictionary. Simultaneously, the label information of labeled data can be utilized better by this graph-based method via label propagation. We also well exploit the discrimination provided

by the labeled training data in dictionary learning by minimizing the within-class variance. We have conducted several experiments on face recognition, digit recognition, and texture classification, which show the advantage of our proposed SSD-LP approach.

Our main contributions are summarized as follows:

1. We propose a novel discriminative semi-supervised dictionary learning method which can effectively utilize the discriminative information hidden in both unlabeled and labeled training data.
2. By using label propagation, we estimate a more accurate relationship between unlabeled training data and classes, and enhance exploration of the discrimination provided by the unlabeled training data.
3. The discrimination provided by the labeled training data by minimizing within-class variance is explored during semi-supervised dictionary learning.
4. Experimental results show that our method has a significantly better discrimination ability using unlabeled training data in dictionary learning.

The rest of this paper is organized as follows. In Section 2, we briefly introduce related work on semi-supervised dictionary learning. Our model is presented in Section 3, and Section 4 describes the optimization procedure. Section 5 presents experimental results and Section 6 concludes the paper with a brief summary and discussion.

2 Related work

Based on the predefined relationship between dictionary atoms and class labels, semi-supervised dictionary learning approaches can be divided into two main categories: discriminative class-specific dictionary learning and discriminative shared dictionary learning.

Motivated by Ref. [24], Shrivastava et al. [19] learnt a class-specific dictionary by using Fisher discriminant analysis on the coding vectors of the labeled data. However, its model is complex: the training data is represented by a combination of all class-specific dictionaries, and the coding coefficients are regularized by both intra-class and inter-class constraints.

Another approach to semi-supervised dictionary learning is to learn a shared dictionary. Pham and

Venktesh [11] took into account the representation errors of both labeled data and unlabeled data. In addition, the classification errors of labeled data were incorporated into a joint objective function. One major drawback of the above approach is that it may fall into a local minimum due to the dictionary construction and classifier design. Wang et al. [18] utilized an artificially designed penalty function to assign weights to the unlabeled data, greatly suppressing the unlabeled data having low confidence. Zhang et al. [22] proposed an online semi-supervised dictionary learning framework which integrated the reconstruction error of both labeled and unlabeled data, label consistency, and the classification error into an objective function. Babagholami-Mohamadabadi et al. [23] integrated dictionary learning and classifier training into an objective function, and preserved the geometrical structure of both labeled and unlabeled data. Recently, Wang et al. [21] utilized the structural sparse relationships between both the labeled and unlabeled samples to learn a discriminative dictionary in which the unlabeled samples are automatically grouped into different labeled samples. Although a shared dictionary usually has a compact size, the discrimination provided by class-specific reconstruction residuals cannot be used.

3 Semi-supervised dictionary learning with label propagation (SSD-LP)

Although several semi-supervised dictionary learning approaches have been proposed, there are still some issues to be solved, such as how to build a discriminative dictionary by using unlabeled data, how to utilize the representation ability of a class-specific dictionary, and how to estimate the class probabilities of the unlabeled data. In this section, we propose a discriminative semi-supervised dictionary learning method using label propagation (SSD-LP) to address the issues mentioned above.

3.1 SSD-LP model

Let $\mathbf{A} = [\mathbf{A}_1, \dots, \mathbf{A}_i, \dots, \mathbf{A}_C]$ be the labeled training data, where \mathbf{A}_i is the i th-class training data and each column of \mathbf{A}_i is a training sample, and $\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_j, \dots, \mathbf{b}_N]$ is the unlabeled training data with unknown labels from 1 to C , where N is

the number of unlabeled training samples. Here, as in prevailing semi-supervised dictionary methods [11, 18, 19, 21–23, 36], we assume that the unlabeled training data belongs to some class of the training set.

In our proposed model, the dictionary to be learnt is $\mathbf{D} = [\mathbf{D}_1, \dots, \mathbf{D}_i, \dots, \mathbf{D}_C]$, where \mathbf{D}_i is the class-specific sub-dictionary associated with class i ; it is required to well represent the i th-class data but to have a poor representation ability for all other classes. In general, we make each column of \mathbf{D}_i a unit vector. We can write \mathbf{D}_i , the representation coefficient matrix of \mathbf{A}_i over \mathbf{D} as $\mathbf{X}_i = [\mathbf{X}_i^1; \dots; \mathbf{X}_i^j; \dots; \mathbf{X}_i^C]$, where \mathbf{X}_i^j is the coding coefficient matrix of \mathbf{A}_i on the sub-dictionary \mathbf{D}_j . Further, \mathbf{y}_j^i is the coding coefficient vector of the unlabeled sample \mathbf{b}_j on the class-specific dictionary \mathbf{D}_i .

Apart from requiring the coding coefficients to be sparse, for the labeled training data we also minimize the within-class scatter of coding coefficients, $\|\mathbf{X}_i^i - \mathbf{M}_i\|$, to make the training samples from the same class have similar coding coefficients, where \mathbf{M}_i is the mean coefficient matrix with the same size as \mathbf{X}_i^i and takes the mean column vector of \mathbf{X}_i^i as its column vectors.

We define a latent variable, $P_{i,j}$, which represents the probability that the j th unlabeled training sample belongs to the i th class. $P_{i,j}$ satisfies $0 \leq P_{i,j} \leq 1$ and $\sum_{i=1}^C P_{i,j} = 1$. If the labeled training sample k belongs to class j , then $P_{j,k} = 1$ and $P_{i,k} = 0$ for $i \neq j$.

Our proposed SSD-LP method can now be formulated as

$$\min_{\mathbf{D}, \mathbf{X}, \mathbf{P}, \mathbf{y}} \sum_{i=1}^C \left(\|\mathbf{A}_i - \mathbf{D}_i \mathbf{X}_i^i\|_F^2 + \gamma \|\mathbf{X}_i^i\|_1 - \lambda \|\mathbf{X}_i^i - \mathbf{M}_i\|_F^2 \right) + \sum_{j=1}^N \left\{ \sum_{i=1}^C P_{i,j} \|\mathbf{b}_j - \mathbf{D}_i \mathbf{y}_j^i\|_F^2 + \gamma \|\mathbf{y}_j^i\|_1 \right\} \quad (1)$$

where γ and λ are parameters, and \mathbf{P} is learned via our proposed improved label propagation (ILP) algorithm.

For the labeled training data, a discriminative representation term, i.e., $\|\mathbf{A}_i - \mathbf{D}_i \mathbf{X}_i^i\|_F^2$, and a discriminative coefficient term, i.e., $\|\mathbf{X}_i^i - \mathbf{M}_i\|_F^2$, are introduced. Since \mathbf{D}_i is associated with the i th-class, it is expected that \mathbf{A}_i should be well represented by \mathbf{D}_i but not by \mathbf{D}_j , $j \neq i$. This implies that \mathbf{X}_i^i should have some significant coefficients such

that $\|\mathbf{A}_i - \mathbf{D}_i \mathbf{X}_i^i\|_F^2$ is small, while \mathbf{X}_i^j should have nearly zero coefficients. Thus the term $\|\mathbf{D}_i \mathbf{X}_i^j\|_F^2$ is eliminated as shown in Eq. (1).

For the unlabeled training data, the probability that the sample belongs to each class is required. For instance, $P_{i,j} = 1$ indicates that the j th unlabeled training sample comes from the i th-class, and the class-specific dictionary \mathbf{D}_i should well represent the j th unlabeled training sample in that $\|\mathbf{b}_j - \mathbf{D}_i \mathbf{y}_j^i\|_F^2$ is small.

Due to the good performance of graph-based label propagation on semi-supervised classification tasks, we utilize it to select the unlabeled sample with high confidence and assign the unlabeled sample a high weight, as explained in detail in Section 4.1.

3.2 Classification scheme

Once the dictionary $\mathbf{D} = [\mathbf{D}_1, \dots, \mathbf{D}_i, \dots, \mathbf{D}_C]$ has been learned, a testing sample can be classified by coding it over the learned dictionary. Although the learned dictionary is class-specific, the testing sample is not always coded on each sub-dictionary corresponding to each class. As the discussion in Ref. [24], there are two methods of coding the testing sample.

When the number of training samples in each class is relatively small, the sample sub-space of class i cannot be supported by the learned sub-dictionary \mathbf{D}_i . Thus the testing samples \mathbf{b}^t are represented on the collaborative combination of all class-specific dictionaries. In this case, the sparse coding vector of the testing sample should be found by solving:

$$\hat{\mathbf{y}} = \arg \min_{\mathbf{y}} \{ \|\mathbf{b}^t - \mathbf{D} \mathbf{y}\|_2^2 + \gamma \|\mathbf{y}\|_1 \} \quad (2)$$

where γ is a constant for the sparsity constraint. Then the class of the testing sample \mathbf{b}^t is predicted by

$$\text{label} = \arg \min_i \|\mathbf{b}^t - \mathbf{D}_i \mathbf{y}_i\|_2^2 \quad (3)$$

where $\hat{\mathbf{y}} = [\mathbf{y}_1; \dots; \mathbf{y}_i; \dots; \mathbf{y}_C]$ and \mathbf{y}_i is the coefficient vector associated with class i .

When the number of training samples in each class is relatively large, the sub-dictionary \mathbf{D}_i , which has enough discrimination, can support the sample sub-space of class i . Thus, we can directly code testing sample \mathbf{b}^t on each sub-dictionary:

$$\hat{\mathbf{y}} = \arg \min_{\mathbf{y}} \{ \|\mathbf{b}^t - \mathbf{D}_i \mathbf{y}\|_2^2 + \gamma \|\mathbf{y}\|_1 \} \quad (4)$$

The class of testing sample \mathbf{b}^t is then predicted by

$$\text{label} = \arg \min_i \{e_i\} \quad (5)$$

where $e_i = \|\mathbf{b}^t - \mathbf{D}_i \hat{\mathbf{y}}\|_2^2$.

4 Optimization of SSD-LP

The SSD-LP objective function is not convex in the joint variables of $\{\mathbf{D}, \mathbf{X}, \mathbf{P}, \mathbf{y}\}$, but it is convex in each variable when the others are fixed. Optimization of Eq. (1) can be divided into three sub-problems: updating \mathbf{P} by fixing $\mathbf{D}, \mathbf{X}, \mathbf{y}$; updating \mathbf{X}, \mathbf{y} by fixing \mathbf{P}, \mathbf{D} ; and updating \mathbf{D} by fixing \mathbf{P}, \mathbf{X} .

4.1 Updating \mathbf{P} by improved label propagation

Unlike the approach used in Ref. [28] to construct the weight matrix, our weight matrix is constructed from the reconstruction errors of the unlabeled samples over all classes rather than the distances between any two samples. Intuitively, since sub-dictionary \mathbf{D}_i is good at representing the i th-class but is poor at representing other classes, any pair of samples is likely to belong to the same class if they achieve minimum reconstruction error in the same class.

Specifically, to compute the weight value w_{ij} (if w_{ij} is large, then sample \mathbf{b}_i and sample \mathbf{b}_j are likely to have the same class), we first compute the reconstruction errors of both training samples \mathbf{b}_i and \mathbf{b}_j over all classes. This gives $\mathbf{e}_i = [e_{i1}; \dots; e_{ik}; \dots; e_{ic}]$ and $\mathbf{e}_j = [e_{j1}; \dots; e_{jk}; \dots; e_{jc}]$ where $e_{ik} = \|\mathbf{b}_i - \mathbf{D}_k \mathbf{y}_i^k\|_2^2$ is the reconstruction error value of sample \mathbf{b}_i on the class k and \mathbf{y}_i^k is the coefficient vector for class k .

After obtaining \mathbf{e}_i and \mathbf{e}_j , we compute the distance d_{ij}^2 between them:

$$d_{ij}^2 = \|\mathbf{e}_i - \mathbf{e}_j\|_2^2 \quad (6)$$

Finally, the weight linking samples \mathbf{b}_i and \mathbf{b}_j is

$$w_{ij} = \exp(-d_{ij}^2/\sigma^2) \quad (7)$$

where σ is a constant. After finding all weight values for every pair of samples, we can get the transition matrix \mathbf{T} , which can be defined by normalizing the weight matrix using:

$$T(i, j) = \frac{w_{ij}}{\sum_{k=1}^C w_{ik}} \quad (8)$$

so that $\sum_{j=1}^C T(i, j) = 1$ and \mathbf{T} is asymmetric after normalization.

Let $n = n_l + n_u$ where n_l, n_u are the total numbers of labeled and unlabeled training samples respectively. For the multi-class problem, the

probability matrix is $\mathbf{P} = [\mathbf{P}^l; \mathbf{P}^u] \in \mathbb{R}^{n \times C}$, where C is the number of classes, \mathbf{P}^l is the probability matrix for labeled samples, and \mathbf{P}^u is the probability matrix for unlabeled samples. We set $\mathbf{P}^l(i, k) = 1$ if sample \mathbf{b}_i is a labeled sample with class k , and 0 otherwise. We initialize the probability matrix as $\mathbf{P}_0 = [\mathbf{P}_0^l; \mathbf{0}]$, i.e., the probability for the unlabeled training samples is set to zero. The improved label propagation algorithm for updating \mathbf{P} is presented in Algorithm 1. Its convergence can be seen by referring to Ref. [27]. \mathbf{P}_{t+1} denotes the next iteration of \mathbf{P}_t . Please note that the step 3.b is crucial as it ensures the label information of the labeled samples is preserved.

Compared with using a weight matrix based on the distances of any two original samples, there are two main advantages in our method. On one hand, the original method of constructing the weight matrix is a kind of single-track feedback mechanism in which the update of the probability matrix \mathbf{P} can affect the dictionary update, but the update of the latter cannot affect the former because the distances between the original samples do not change. On the other hand, a weight matrix based on reconstruction errors over all classes more realistically reflects the similarity between two samples, which is helpful in estimating the class labels of unlabeled data.

4.2 Updating \mathbf{X} and \mathbf{y}

By fixing the estimated class probabilities of the unlabeled training data (i.e., \mathbf{P}), the discriminative dictionary (i.e., \mathbf{D}) and coding coefficients (i.e., \mathbf{X} and \mathbf{y}) can now be updated.

When the dictionary \mathbf{D} is fixed, the coding coefficients of the labeled training data can be easily updated. The objective function in Eq. (1) now reduces to

$$\min_{\mathbf{X}} \sum_{i=1}^C (\|\mathbf{A}_i - \mathbf{D}_i \mathbf{X}_i^i\|_F^2 + \gamma \|\mathbf{X}_i^i\|_1 + \lambda \|\mathbf{X}_i^i - \mathbf{M}_i\|_F^2) \quad (9)$$

Algorithm 1: Improved label propagation based on reconstruction error

- 1: Construct a transition matrix \mathbf{T} by Eq. (8);
 - 2: Initialize the probability matrix $\mathbf{P}_0 = [\mathbf{P}_0^l; \mathbf{0}]$;
 - 3: Repeating the following steps until \mathbf{P} converges:
 - 3.a $\mathbf{P}_{t+1} = \mathbf{T} * \mathbf{P}_t$;
 - 3.b $\mathbf{P}_{t+1}^l = \mathbf{P}_0^l$;
 - 4: Output the probability matrix \mathbf{P} .
-

In our approach, we update \mathbf{X}_i^i for the i th-class of data by using the coding method in Ref. [24].

As discussed in Section 3.2, when the number of training samples in each class is relatively small, updating the coding coefficients of the unlabeled training data using a collaborative representation can achieve better classification performance. Conversely, we choose a local representation when there are sufficient training samples of each class. For the unlabeled training data, two coding strategies, i.e., collaborative representation and local representation, are used. In the collaborative representation, the coding coefficient is solved via

$$\min_{\mathbf{y}_j} \|\mathbf{b}_j - \mathbf{D}\mathbf{y}_j\|_F^2 + \gamma \|\mathbf{y}_j\|_1 \quad (10)$$

where $\mathbf{D} = [\mathbf{D}_1, \dots, \mathbf{D}_i, \dots, \mathbf{D}_C]$ and $\mathbf{y}_j = [\mathbf{y}_j^1; \dots; \mathbf{y}_j^i; \dots; \mathbf{y}_j^C]$; \mathbf{y}_j^i is the coding vector of the unlabeled sample \mathbf{b}_j for the sub-dictionary \mathbf{D}_i . Here the different class-specific dictionaries \mathbf{D}_i will compete with each other to represent \mathbf{b}_j . In order to ensure fair competition between different class-specific dictionaries, the encoding phase of collaborative representation ignores \mathbf{P} .

In the local representation, the SSD-LP model associated with \mathbf{y}_i changes to

$$\min_{\mathbf{y}_j^i} \sum_{i=1}^C P_{i,j} (\|\mathbf{b}_j - \mathbf{D}_i \mathbf{y}_j^i\|_F^2) + \gamma \|\mathbf{y}_j^i\|_1 \quad (11)$$

which is a standard sparse coding problem.

4.3 Updating \mathbf{D}

After updating \mathbf{P} , further unlabeled training samples are selected to train our model. If we fix the size of the learnt dictionary, the discrimination of our dictionary cannot improve. Thus, after updating the probability matrix \mathbf{P} , we should increase the size of each sub-dictionary to explore the discriminative information hidden in the unlabeled samples (i.e., an additional dictionary atom \mathbf{E}_i must be initialized and added to sub-dictionary \mathbf{D}_i).

Since the unlabeled samples provide more discrimination, \mathbf{E}_i is initialized by using unlabeled data:

$$\min_{\mathbf{E}} \sum_{j=1}^N \sum_{i=1}^C P_{i,j} \|\mathbf{b}_j - \mathbf{E}_i \tilde{\mathbf{y}}_j^i\|_F^2 \quad (12)$$

where $\tilde{\mathbf{y}}_j^i$ is the unknown coding coefficient.

We update \mathbf{E}_i class by class:

$$\min_{\mathbf{E}_i} \sum_{j=1}^N P_{i,j} \|\mathbf{b}_j - \mathbf{E}_i \tilde{\mathbf{y}}_j^i\|_F^2 \quad (13)$$

Then we combine all terms in Eq. (13):

$$\min_{\mathbf{E}_i} \|\left[\sqrt{P_{i,1}}\mathbf{b}_1, \dots, \sqrt{P_{i,j}}\mathbf{b}_j, \dots, \sqrt{P_{i,N}}\mathbf{b}_N\right] - \mathbf{E}_i \left[\sqrt{P_{i,1}}\tilde{\mathbf{y}}_1, \dots, \sqrt{P_{i,j}}\tilde{\mathbf{y}}_j, \dots, \sqrt{P_{i,N}}\tilde{\mathbf{y}}_N\right]\|_F^2 \quad (14)$$

Since we require the coding coefficients to be sparse, we compute the extended dictionary by singular-value decomposition (SVD):

$$[\mathbf{U}, \mathbf{S}, \mathbf{V}] = \text{svd}(\left(\sqrt{P_{i,1}}\mathbf{b}_1 \dots (\sqrt{P_{i,j}}\mathbf{b}_j) \dots \sqrt{P_{i,N}}\mathbf{b}_N\right)) \quad (15)$$

The extended dictionary is defined such that:

$$\mathbf{E}_i = \mathbf{U}(:, n) \quad (16)$$

where n is the number of atoms in the extended dictionary. In all experiments shown in this paper, we set $n = 1$, i.e., each sub-dictionary adds a single dictionary atom after the update of probability matrix \mathbf{P} .

The new sub-dictionary for class i is initialized using $\hat{\mathbf{D}}_i = [\mathbf{D}_i, \mathbf{E}_i]$. By fixing the coding coefficient \mathbf{X} and probability matrix \mathbf{P} , the problem in Eq. (1) is reduced to

$$\min_{\mathbf{D}} \sum_{i=1}^C (\|\mathbf{A}_i - \hat{\mathbf{D}}_i \mathbf{X}_i^i\|_F^2) + \sum_{j=1}^N \left(\sum_{i=1}^C P_{i,j} \|\mathbf{b}_j - \hat{\mathbf{D}}_i \mathbf{y}_j^i\|_F^2\right) \quad (17)$$

Dictionary updating can be easily performed atom by atom by using Metaface [8]. After updating the extended dictionary \mathbf{E} , we need several iterations to update the dictionary and coefficients to guarantee the convergence of the discriminative dictionary. In our experiment, the number of additional iterations is set to 5.

The whole algorithm of the proposed semi-supervised dictionary learning is summarized in Algorithm 2. The algorithm converges since the total objective function value in Eq. (1) decreases in each iteration. Figure 1 shows the total objective function value for the AR dataset [37]. In all the experiments mentioned in this paper, our algorithm converges in less than 10 iterations.

5 Experiment results

We have performed experiments and corresponding analysis to verify the performance of our method for image classification. We evaluate our approach on two face databases: the Extended YaleB

Algorithm 2: Semi-supervised dictionary learning with label propagation (SSD-LP)

1. Initialization

The class probabilities of the unlabeled training samples are all initialized to zero; the sub-dictionary D_i is initialized using the classes of the i th-class labeled training samples A_i ; and each column of D_i is given unit l_2 -norm.

2. Class estimation of the unlabeled training data

Update each $P_{i,j}$ for every unlabeled training sample using Algorithm 1.

3. Discriminative dictionary learning

3.1 Compute the extended dictionary E_i for each class by solving Eq. (16).

3.2 Update the coding coefficients X and y , and the dictionary D over several iterations:

3.2.1 Update the coding coefficients by solving Eqs. (9)–(11);

3.2.2 Update the dictionary atom by atom by solving Eq. (17).

4. Return to step 2 unless the values of the objective function in Eq. (1) in adjacent iterations are close enough or the maximum number of iterations is reached.

5. Output D .

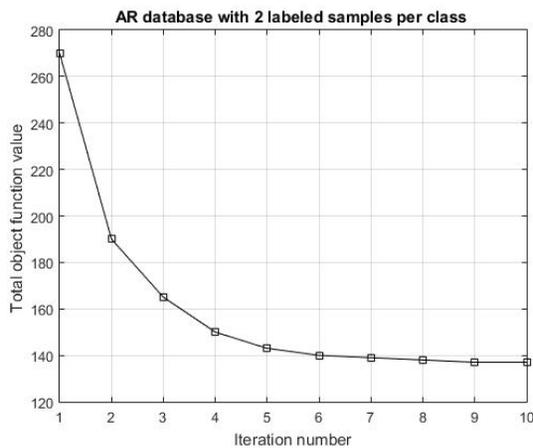


Fig. 1 Total objective function value on the AR database [37] versus number of iterations.

database [38] and the AR face database [37], two handwritten digit datasets: MNIST [39] and USPS [40], and an object category dataset: Texture-25 [41]. We compare our method with SRC [12], M-SVM [17], FDDL [24], DKSD [10], LCKSD [16], SVGDL [42], S2D2 [19], JDL [11], OSSDL [22], SSR-D [36], and the recently proposed USSDL [18] and SSP-DL [21] algorithms. The last six methods (S2D2, JDL, OSSDL, SSR-D, USSDL, and SSP-DL) are semi-supervised dictionary learning models; the others are supervised dictionary learning methods.

We repeated each experiment 10 times with different random splits of the datasets and report the average classification accuracy together with standard deviation; the best classification results are in boldface. For all approaches, we report their best results obtained after tuning their parameters.

5.1 Parameter selection and comparison with original label propagation

In our all experiments, the parameters of SSD-LP are fixed to $\gamma = 0.001$ and $\lambda = 0.01$. The number of additional iterations is set to 5 in step 3.2 of Algorithm 2. In our experiment, since the sub-dictionary D_i is initialized using the i th-class of labeled samples, the number of atoms D_i is equal to the number of labeled samples for the i th-class (e.g., in the AR database, these are 2, 3, 5 as there are 2, 3, 5 labeled samples respectively). After each update of the probability matrix P , each sub-dictionary adds an additional dictionary atom (the number of atoms of each sub-dictionary does not increase if the number of iterations exceeds the number of unlabeled training examples).

In order to show the effectiveness of our algorithm, a test was conducted on the Extended YaleB dataset. As shown in Fig. 2, we can see that for face recognition, recognition significantly improves with iteration number.

We also compare our proposed improved label propagation method with the original label propagation method (LP). As Fig. 3 shows, SSD-LP has at least 10% improvement over the performance when the images are classified directly

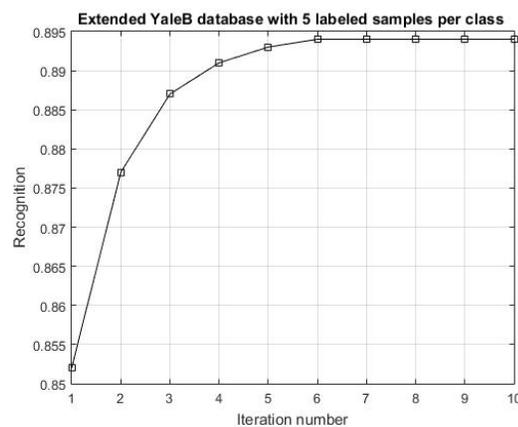


Fig. 2 Recognition rate versus iteration number for the Extended YaleB database with five labeled training samples per class.

by the original label propagation method. With an increasing number of iterations, the recognition rate of our method grows, while the performance of the original label propagation algorithm is essentially unchanged. This is because the original label propagation is dependent on the distribution structure of the input data which does not change as the dictionary is updated. This is a kind of single-track feedback mechanism between the original label propagation and dictionary learning as explained in Section 4.1.

We also compare the running time of our improved LP and the original LP using MATLAB 2015a on an Intel i7-3770 3.40 GHz machine with 16.0 GB RAM. The running time for the improved LP and original LP is 11.21 s and 7.54 s, respectively for two training examples per person (see Fig. 3 up), and 11.73 s and 7.75 s, respectively for five training examples per person (see Fig. 3 bottom). We can see that the running time of the improved LP and original LP is comparable.

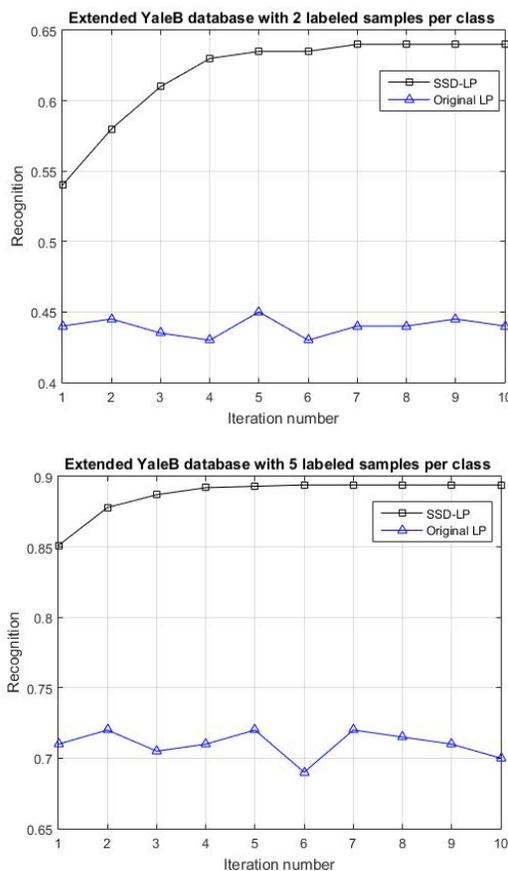


Fig. 3 Recognition rate versus iteration number for the Extended YaleB database with two labeled training samples per person (up) and five training samples per person (bottom).

5.2 Face recognition

In this section, we evaluate our method in a face recognition problem, for both AR and Extended YaleB databases, using the same experimental setting as Ref. [18]. In both face recognition experiments, the image samples are reduced to 300 dimensions by PCA.

The AR database consists of over 4000 images of 126 individuals. In the experiment we chose a subset of 50 male and 50 female subjects. Focusing on illumination and expression changes, for each subject we chose 7 images from Session 1 for training, and 7 images from Session 2 for testing. We randomly selected {2, 3, 5} samples from each class in the training set as the labeled samples, and the remaining as the unlabeled samples. Five independent evaluations were conducted for the experiment with different numbers of labeled training samples.

As shown in Table 1, when the number of labeled samples was small (2 or 3), our algorithm performed better than all other methods, especially supervised dictionary learning models. This is because supervised dictionary methods cannot utilize the discriminative information hidden in the unlabeled training samples. The semi-supervised dictionary learning methods usually perform better than supervised dictionary learning methods: for instance, USSDL performs the second best. From Table 1, we can see that USSDL has very close results to SSD-LP, but we should note that USSDL needs more information in the dictionary learning task, including classifier learning of the coding vectors. In addition, the optimization procedure of USSDL is more complex than that of SSD-LP.

We also evaluated our approach on the Extended

Table 1 Recognition rate for various methods, for different number of labeled training samples, for the AR database (Unit: %)

Method	2	3	5
SRC	72.2±1.0	79.1±0.9	88.2±0.5
M-SVM	60.1±2.2	74.3±1.2	84.9±2.0
FDDL	83.6±1.8	89.7±2.1	93.6±0.9
LC-KSVD	67.4±4.2	89.2±3.6	91.5±2.1
SVGDL	82.1±1.8	90.3±2.0	93.8±1.5
S2D2	85.3±3.1	89.2±1.9	92.1±1.1
JDL	87.2±2.0	88.2±1.8	90.7±1.2
USSDL	89.1±2.3	91.3±1.4	94.1±1.3
SSD-LP	90.9±0.9	91.6±0.6	93.7±0.5

YaleB database. The database consists of 2414 frontal face images of 38 individuals. Each individual has 64 images; we randomly selected 20 images as the training set and used the rest as the testing set. We randomly selected $\{2, 5, 10\}$ samples from each class in the training set as the labeled samples, and used the remainder as the unlabeled samples. The classification results are shown in Table 2.

It is clear that our proposed method provides better classification performance than other dictionary learning methods. Especially when a small number of label samples is involved, the SSD-LP performs significantly better than supervised dictionary learning methods which are dependent on the number of the labeled samples. It also can be seen that SSD-LP improves by at least 1.5% over the other semi-supervised dictionary learning methods. When the number of labeled samples is small, the improvement is more obvious. That is mainly because our method has strong capability to utilize the unlabeled samples by accurately determining their labels and using them as labeled samples to train our discriminative dictionary.

5.3 Digit classification

Next, we evaluated the performance of our method on both the MNIST and USPS datasets, with the same experimental setting as Ref. [21]. The MNIST dataset has 10 classes. The training set has 60,000 handwritten digital images and the test set has about 10,000 images. The dimension of each digital image is 784. We randomly selected 200 samples from each class, using 20 images as the labeled samples, 80 as the unlabeled samples, and the rest for testing.

The USPS dataset has 9298 digital images in 10 classes. We randomly selected 110 images from each

class, using 20 as the labeled samples, 40 as the unlabeled samples, and 50 as the testing samples. We used the whole image as the feature vector, and normalized the vector to have unit l_2 -norm.

The results for the ten independent tests are combined in Table 3. It can be seen that our proposed SSD-LP method can effectively utilize information from the unlabeled samples, achieving a classification accuracy clearly higher than for the other dictionary methods. Using the additional unlabeled training samples, the size of the dictionary is enlarged adaptively to better utilize the discrimination provided by the unlabeled samples, which is why we can achieve better performance than other semi-supervised dictionary methods mentioned in Table 3.

5.4 Object classification

In this experiment we used the Texture-25 dataset which contains 25 texture categories, with 40 samples of each. We used low-level features [43, 44], including PHOG [32], GIST [45], and LBP [46]. Using the experimental setting in Ref. [18], PHOG was computed with a 2-layer pyramid in 8 directions, and GIST was computed on rescaled images of 256×256 pixels, in 4, 8, and 8 orientations at 3 scales from coarse to fine. Uniform LBP were used. All the features are concatenated into a single 119-dimensional vector. In this experiment, 13 images were randomly selected for testing and we randomly select $\{2, 5, 10, 15\}$ samples from each class in the training set as labeled samples. The average accuracies together with the standard deviation in five independent tests are presented in Table 4.

It can be seen that SSD-LP improves by at least 3% over supervised dictionary learning when the number of labeled samples is 2 or 5. As the number

Table 2 Recognition rate for various methods, for different number of labeled training samples, for the Extended YaleB database (Unit: %)

Method	2	3	5
SRC	47.8±2.9	79.1±1.9	90.5±0.5
M-SVM	38.0±2.6	66.6±1.1	83.8±0.8
FDDL	52.4±2.5	82.3±0.7	92.1±0.3
LC-KSVD	48.5±2.8	69.6±3.6	84.6±3.8
SVGDL	53.4±2.2	81.1±1.0	91.7±5.8
S2D2	53.4±2.1	76.1±1.3	83.2±1.9
JDL	55.2±1.8	77.4±2.8	85.3±1.6
USSDL	60.5±2.1	86.5±2.1	93.6±0.8
SSD-LP	67.0±2.9	89.8±0.9	95.2±0.2

Table 3 Recognition rate for various methods, for digit databases USPS and MNIST (Unit: %)

Method	USPS	MNIST
SRC	68.6±2.7	72.9±2.3
DKSVD	67.5±1.8	71.4±1.7
FDDL	85.2±1.2	82.5±1.3
LC-KSVD	76.9±1.3	73.0±1.3
OSSDL	80.8±2.8	73.2±1.8
S2D2	86.6±1.6	77.6±0.8
SSR-D	87.2±0.5	83.8±1.2
SSP-DL	87.8±1.1	85.8±1.2
SSD-LP	90.3±1.3	87.8±1.6

Table 4 Recognition rate for various methods, for different number of labeled training samples, for the Texture-25 database (Unit: %)

Method	2	5	10	15
M-SVM	24.9±3.4	41.6±1.7	52.9±2.7	55.3±1.2
FDDL	31.4±4.0	48.9±1.7	52.6±3.1	56.7±1.4
LC-KSVD	28.0±4.1	38.2±1.3	48.6±2.9	54.1±2.9
SVGDL	29.8±3.9	37.9±1.3	40.3±2.3	56.8±1.3
S2D2	31.7±2.3	43.8±1.4	47.9±2.4	50.9±1.7
JDL	27.6±2.1	39.2±1.9	43.3±0.8	50.3±0.8
USSDL	34.2±3.7	51.1±2.2	54.6±1.6	57.7±1.6
SSD-LP	38.2±1.3	54.2±2.0	64.1±1.3	73.7±2.5

of labeled samples increases, the effect is clearly enhanced, by about 10%. Table 4 shows that our method also gives better results than the other three semi-supervised dictionary methods. That is because as more samples are used for training, the estimates of the labels of the unlabeled training data become more accurate. The result fully demonstrates the classification effectiveness of label propagation based on reconstruction error. In addition, adaptively adding dictionary atoms makes our learnt dictionary more discriminative. JDL, which only uses the reconstruction error of both labeled and unlabeled data, does not work well.

6 Conclusions

This paper has proposed a discriminative semi-supervised dictionary learning model. By integrating label propagation with that class-specific reconstruction error of each unlabeled training sample, we can more accurately estimate the class of unlabeled samples to train our model. The discriminative property of labeled training data is also well explored by using a discriminative representation term and minimizing within-class scatter of the coding coefficients. Several experiments, including applications to face recognition, digit recognition, and texture classification have shown the advantage of our method over supervised and other semi-supervised dictionary learning approaches. In the future, we will explore more classification questions, e.g., the case in which the training samples may not belong to any known class.

Acknowledgements

This work was partially supported by the National Natural Science Foundation for Young Scientists

of China (No. 61402289), and the National Science Foundation of Guangdong Province (No. 2014A030313558).

References

- [1] Elad, M.; Figueiredo, M. A. T.; Ma, Y. On the role of sparse and redundant representations in image processing. *Proceedings of the IEEE* Vol. 98, No. 6, 972–982, 2010.
- [2] Wright, J.; Ma, Y.; Mairal, J.; Sapiro, G.; Huang, T. S.; Yan, S. Sparse representation for computer vision and pattern recognition. *Proceedings of the IEEE* Vol. 98, No. 6, 1031–1044, 2010.
- [3] Chen, Y.-C.; Patel, V. M.; Phillips, P. J.; Chellappa, R. Dictionary-based face recognition from video. In: *Computer Vision–ECCV 2012*. Fitzgibbon, A.; Lazebnik, S.; Perona, P.; Sato, Y.; Schmid, C. Eds. Springer Berlin Heidelberg, 766–779, 2012.
- [4] Mairal, J.; Elad, M.; Sapiro, G. Sparse representation for color image restoration. *IEEE Transactions on Image Processing* Vol. 17, No. 1, 53–69, 2008.
- [5] Bryt, O.; Elad, M. Compression of facial images using the K-SVD algorithm. *Journal of Visual Communication and Image Representation* Vol. 19, No. 4, 270–282, 2008.
- [6] Bryt, O.; Elad, M. Improving the k-svd facial image compression using a linear deblocking method. In: *Proceedings of the IEEE 25th Convention of Electrical and Electronics Engineers in Israel*, 533–537, 2008.
- [7] Yang, J.; Yu, K.; Huang, T. Supervised translation-invariant sparse coding. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 3517–3524, 2010.
- [8] Yang, M.; Zhang, L.; Yang, J.; Zhang, D. Metaface learning for sparse representation based face recognition. In: *Proceedings of the IEEE International Conference on Image Processing*, 1601–1604, 2010.
- [9] Mairal, J.; Ponce, J.; Sapiro, G.; Zisserman, A.; Bach, F. R. Supervised dictionary learning. In: *Proceedings of the Advances in Neural Information Processing Systems*, 1033–1040, 2009.
- [10] Zhang, Q.; Li, B. Discriminative K-SVD for dictionary learning in face recognition. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2691–2698, 2010.
- [11] Pham, D.-S.; Venkatesh, S. Joint learning and dictionary construction for pattern recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1–8, 2008.
- [12] Wright, J.; Yang, A. Y.; Ganesh, A.; Sastry, S. S.; Ma, Y. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 31, No. 2, 210–227, 2009.
- [13] Aharon, M.; Elad, M.; Bruckstein, A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing* Vol. 54, No. 11, 4311–4322, 2006.

- [14] Mairal, J.; Bach, F.; Ponce, J.; Sapiro, G.; Zisserman, A. Discriminative learned dictionaries for local image analysis. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1–8, 2008.
- [15] Yang, M.; Dai, D.; Shen, L.; Van Gool, L. Latent dictionary learning for sparse representation based classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4138–4145, 2014.
- [16] Jiang, Z.; Lin, Z.; Davis, L. S. Learning a discriminative dictionary for sparse coding via label consistent K-SVD. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1697–1704, 2011.
- [17] Yang, J.; Yu, K.; Gong, Y.; Huang, T. Linear spatial pyramid matching using sparse coding for image classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1794–1801, 2009.
- [18] Wang, X.; Guo, X.; Li, S. Z. Adaptively unified semi-supervised dictionary learning with active points. In: Proceedings of the IEEE International Conference on Computer Vision, 1787–1795, 2015.
- [19] Shrivastava, A.; Pillai, J. K.; Patel, V. M.; Chellappa, R. Learning discriminative dictionaries with partially labeled data. In: Proceedings of the 19th IEEE International Conference on Image Processing, 3113–3116, 2012.
- [20] Jian, M.; Jung, C. Semi-supervised bi-dictionary learning for image classification with smooth representation-based label propagation. *IEEE Transactions on Multimedia* Vol. 18, No. 3, 458–473, 2016.
- [21] Wang, D.; Zhang, X.; Fan, M.; Ye, X. Semi-supervised dictionary learning via structural sparse preserving. In: Proceedings of the 30th AAAI Conference on Artificial Intelligence, 2137–2144, 2016.
- [22] Zhang, G.; Jiang, Z.; Davis, L. S. Online semi-supervised discriminative dictionary learning for sparse representation. In: *Computer Vision—ACCV 2012*. Lee, K. M.; Mu, K.; Matsushita, Y.; Rehg, J. M.; Hu, Z. Eds. Springer Berlin Heidelberg, 259–273, 2012.
- [23] Babagholami-Mohamadabadi, B.; Zarghami, A.; Zolfaghari, M.; Baghshah, M. S. PSSDL: Probabilistic semi-supervised dictionary learning. In: *Machine Learning and Knowledge Discovery in Databases*. Blockeel, H.; Kersting, K.; Nijssen, S.; Železný, F. Eds. Springer Berlin Heidelberg, 192–207, 2013.
- [24] Yang, M.; Zhang, L.; Feng, X.; Zhang, D. Fisher discrimination dictionary learning for sparse representation. In: Proceedings of the IEEE International Conference on Computer Vision, 543–550, 2011.
- [25] Zhou, N.; Shen, Y.; Peng, J.; Fan, J. Learning inter-related visual dictionary for object recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3490–3497, 2012.
- [26] Deng, W.; Hu, J.; Guo, J. Extended SRC: Undersampled face recognition via intra-class variant dictionary. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 34, No. 9, 1864–1870, 2012.
- [27] Zhu, X.; Lafferty, J.; Rosenfeld, R. Semi-supervised learning with graphs. Carnegie Mellon University, Language Technologies Institute, School of Computer Science, 2005.
- [28] Wang, B.; Tu, Z.; Tsotsos, J. K. Dynamic label propagation for semi-supervised multi-class multi-label classification. In: Proceedings of the IEEE International Conference on Computer Vision, 425–432, 2013.
- [29] Blum, A.; Mitchell, T. Combining labeled and unlabeled data with co-training. In: Proceedings of the 11th Annual Conference on Computational Learning Theory, 92–100, 1998.
- [30] Mallapragada, P. K.; Jin, R.; Jain, A. K.; Liu, Y. SemiBoost: Boosting for semi-supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 31, No. 11, 2000–2014, 2009.
- [31] Gong, C.; Tao, D.; Maybank, S. J.; Liu, W.; Kang, G.; Yang, J. Multi-modal curriculum learning for semi-supervised image classification. *IEEE Transactions on Image Processing* Vol. 25, No. 7, 3249–3260, 2016.
- [32] Bosch, A.; Zisserman, A.; Munoz, X. Image classification using random forests and ferns. In: Proceedings of the IEEE 11th International Conference on Computer Vision, 1–8, 2007.
- [33] Xiong, C.; Kim, T.-K. Set-based label propagation of face images. In: Proceedings of the 19th IEEE International Conference on Image Processing, 1433–1436, 2012.
- [34] Cheng, H.; Liu, Z.; Yang, J. Sparsity induced similarity measure for label propagation. In: Proceedings of the IEEE 12th International Conference on Computer Vision, 317–324, 2009.
- [35] Kang, F.; Jin, R.; Sukthankar, R. Correlated label propagation with application to multi-label learning. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1719–1726, 2006.
- [36] Wang, H.; Nie, F.; Cai, W.; Huang, H. Semi-supervised robust dictionary learning via efficient l-norms minimization. In: Proceedings of the IEEE International Conference on Computer Vision, 1145–1152, 2013.
- [37] Martinez, A. M. The AR face database. CVC Technical Report 24, 1998.
- [38] Lee, K.-C.; Ho, J.; Kriegman, D. J. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 27, No. 5, 684–698, 2005.
- [39] LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* Vol. 86, No. 11, 2278–2324, 1998.

- [40] Hull, J. J. A database for handwritten text recognition research. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 16, No. 5, 550–554, 1994.
- [41] Lazebnik, S.; Schmid, C.; Ponce, J. A sparse texture representation using local affine regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 27, No. 8, 1265–1278, 2005.
- [42] Cai, S.; Zuo, W.; Zhang, L.; Feng, X.; Wang, P. Support vector guided dictionary learning. In: *Computer Vision–ECCV 2014*. Fleet, D.; Pajdla, T.; Schiele, B.; Tuytelaars, T. Eds. Springer International Publishing, 624–639, 2014.
- [43] Boix, X.; Roig, G.; Van Gool, L. Comment on “Ensemble projection for semi-supervised image classification”. *arXiv preprint arXiv:1408.6963*, 2014.
- [44] Dai, D.; Van Gool, L. Ensemble projection for semi-supervised image classification. In: *Proceedings of the IEEE International Conference on Computer Vision*, 2072–2079, 2013.
- [45] Oliva, A.; Torralba, A. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision* Vol. 42, No. 3, 145–175, 2001.
- [46] Ojala, T.; Pietikainen, M.; Maenpaa, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 24, No. 7, 971–987, 2002.



Lin Chen received his B.S. degree in computer science and technology from Shenzhen University, Shenzhen, China, in 2015. He is currently pursuing his M.S. degree in the School of Computer Science & Software Engineering, Shenzhen University.



Meng Yang is currently an associate professor in the School of Computer Science & Software Engineering, Shenzhen University, Shenzhen, China. He received his Ph.D. degree from Hong Kong Polytechnic University, Hong Kong, China, in 2012. Before joining Shenzhen University, he worked as a postdoctoral fellow in the Computer Vision Lab. of ETH Zurich. His research interests include sparse coding, dictionary learning, object recognition, and machine learning. He has published 10 AAAI/CVPR/ICCV/ICML/ECCV papers, and several IJCV, IEEE TNNLS, and TIP journal papers.

Open Access The articles published in this journal are distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Other papers from this open access journal are available free of charge from <http://www.springer.com/journal/41095>. To submit a manuscript, please go to <https://www.editorialmanager.com/cvmj>.