

# Label consistent transform learning for pattern classification

He-Feng Yin<sup>1,2</sup> and Xiao-Jun Wu<sup>1,2</sup>

Journal of Algorithms &  
Computational Technology  
Volume 13: 1–8  
© The Author(s) 2019  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/1748302619881395  
journals.sagepub.com/home/act



## Abstract

Transform learning has been successfully applied to various image processing tasks in recent years. Nevertheless, transform learning learns the representation in an unsupervised fashion. To make transform learning suitable for pattern classification, we introduce a label consistency constraint into transform learning and propose a new label consistent transform learning to enhance the classification performance of transform learning. The resulting optimization problem can be solved elegantly by employing the alternative strategy. Experimental results on publicly available databases demonstrate that label consistent transform learning outperforms several dictionary learning approaches and the recently proposed discriminative transform learning. More importantly, label consistent transform learning has the least training time which has the potential in practical applications.

## Keywords

Transform learning, dictionary learning, label consistent, pattern classification

Received 6 August 2019; accepted 18 September 2019

## Introduction

Recent years have witnessed increasing interest in the study of dictionary learning (DL) in various domains, such as face recognition,<sup>1</sup> image fusion,<sup>2</sup> and person re-identification.<sup>3</sup> DL can be divided into two different types, i.e. synthesis dictionary learning (SDL) and analysis dictionary learning (ADL), according to how to encode the input data. SDL aims to learn a dictionary by which the input data can be reconstructed through the linear combination of atoms in the dictionary. The most classic SDL approach is the K-SVD algorithm<sup>4</sup> which has been successfully applied to image compression and denoising. However, K-SVD mainly focuses on the representational ability of the dictionary without considering its capability for classification. To address this problem, Zhang and Li<sup>5</sup> proposed a discriminative K-SVD (D-KSVD) method by introducing the classification error into the framework of K-SVD. Jiang et al.<sup>6</sup> further incorporated a label consistency constraint into K-SVD and presented a label consistent K-SVD (LC-KSVD) algorithm. Yang et al.<sup>7</sup> developed a Fisher discrimination dictionary learning method which imposes the Fisher discrimination criterion on the coding coefficients. Wang D and Kong<sup>8</sup> presented an SDL approach called COPAR which explicitly learns the shared patterns and the class-specific dictionaries.

Li et al.<sup>9</sup> proposed a locality constrained and label embedding dictionary learning algorithm to take the locality and label information of atoms into account. Many SDL employs  $\ell_0/\ell_1$  norm to promote sparsity of the coding, which will lead to high computational complexity. To alleviate this problem, Zhao et al.<sup>10</sup> designed an orthogonal collaborative dictionary learning method which can derive analytical solutions for both code learning and dictionary updating.

Although SDL achieves impressive results in classification tasks, it is time-consuming to learn the dictionary. Recently ADL has attracted lots of attention due to its efficacy and efficiency. Rubinstein et al.<sup>11</sup> presented the analysis K-SVD which is parallel to the synthesis K-SVD.<sup>4</sup> Gu et al.<sup>12</sup> developed a projective dictionary pair learning (DPL) which simultaneously learns a synthesis dictionary and an analysis dictionary.

<sup>1</sup>School of Internet of Things Engineering, Jiangnan University, Wuxi, China

<sup>2</sup>Jiangsu Provincial Engineering Laboratory of Pattern Recognition and Computational Intelligence, Jiangnan University, Wuxi, China

### Corresponding author:

Xiao-Jun Wu, School of Internet of Things Engineering, Jiangnan University, No. 1800 Lihu Avenue, Wuxi 214122, China.

Email: wu\_xiaojun@jiangnan.edu.cn



Guo et al.<sup>13</sup> explored a discriminative ADL by integrating structure preserving and discriminative properties into the basic ADL model. Yang et al.<sup>14</sup> proposed a discriminative analysis-synthesis dictionary learning, in which a linear classifier based on the coding coefficient is jointly learned with the dictionary pair. Wang et al.<sup>15</sup> designed a synthesis linear classifier-based ADL (SLC-ADL) algorithm by introducing a synthesis-linear-classifier-based error term into the basic ADL framework. Similarly, Wang et al.<sup>16</sup> also incorporated the linear classification error term into ADL. The difference between SLC-ADL<sup>15</sup> and the approach proposed by Wang et al.<sup>16</sup> is that SLC-ADL actually uses the classification error term from the analysis viewpoint.

Transform learning (TL) is a new representation learning technique which was presented by Ravishanker et al.<sup>17,18</sup>; it utilizes a transform matrix to attain the representation of input data. In fact, TL and ADL have similar formulation. TL is mainly used in signal and image denoising. And there is few work on discriminative transform learning (DTL). Very recently, Maggu et al.<sup>19</sup> developed a label consistent TL for hyperspectral image classification. Although the method proposed by Maggu et al.<sup>19</sup> is termed as label consistent TL, they essentially introduced the linear classification error term into the framework of TL, so this method is coined as DTL in this paper. To further enhance the classification capability of TL, we introduce the label consistency term into the framework of TL and present label consistent transform learning (LCTL) for pattern classification. Compared with DTL, our proposed LCTL leverages the label information of both training data and dictionary atoms to form an ideal representation matrix. Through the ideal matrix, samples belonging to the same class are forced to have similar representation vectors while those from different classes to have distinct representation vectors. LCTL is evaluated on widely used databases and the experimental results demonstrate that LCTL is superior to DTL and other SDL approaches in terms of recognition accuracy and

training time. The source code of our proposed LCTL can be downloaded from <https://github.com/yinhefeng/label-consistent-transform-learning>

## Related work

### Synthesis DL

Let  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbb{R}^{n \times N}$  be the training data matrix from  $C$  classes, where  $n$  and  $N$  denote the dimensionality and total number of training samples, respectively. The objective function of K-SVD<sup>4</sup> is formulated as follows

$$\min_{\mathbf{D}, \mathbf{Z}} \|\mathbf{X} - \mathbf{D}\mathbf{Z}\|_F^2, \text{ s.t. } \|\mathbf{z}_i\|_0 \leq T_0 \quad (1)$$

where  $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_K] \in \mathbb{R}^{n \times K}$  is the dictionary that is to be learned,  $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N] \in \mathbb{R}^{K \times N}$  is the coding coefficient matrix, and  $T_0$  is a given sparsity level. Equation (1) can be solved by alternatively updating  $\mathbf{D}$  and  $\mathbf{Z}$ . Although K-SVD delivers promising results in image compression and denoising, it is not tailored for classification. To make K-SVD applicable

**Table 1.** Recognition accuracy and training time on the Yale database.

Methods	Accuracy (%)	Training time (s)
SRC	88.0	No need
LLC	93.3	No need
K-SVD	96.0	0.27
D-KSVD	90.7	0.80
LC-KSVD	94.7	0.80
DTL	94.7	0.20
LCTL	<b>97.3</b>	<b>0.16</b>

Note: Bold values signify the best recognition accuracy and the least training time.

DTL: discriminative transform learning; LCTL: label consistent transform learning.



**Figure 1.** Example images from the Yale database.

to pattern classification, Zhang and Li<sup>5</sup> presented D-KSVD algorithm by introducing the classification error term into the framework of K-SVD

$$\begin{aligned} \min_{\mathbf{D}, \mathbf{W}, \mathbf{Z}} \|\mathbf{X} - \mathbf{D}\mathbf{Z}\|_F^2 + \beta \|\mathbf{H} - \mathbf{W}\mathbf{Z}\|_F^2 + \lambda \|\mathbf{W}\|_F^2 \\ \text{s.t. } \|\mathbf{z}_i\|_0 \leq T_0 \end{aligned} \quad (2)$$

where  $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_N] \in \mathbb{R}^{C \times N}$  is the label matrix of training data,  $\mathbf{h}_i = [0, 0, \dots, 1, \dots, 0, 0]^T \in \mathbb{R}^{C \times 1}$  is the label vector of  $\mathbf{x}_i$ , and  $\mathbf{W}$  is the parameters for a linear classifier. We can see that dictionary and classifier are jointly learned in D-KSVD. Then, Jiang et al.<sup>6</sup> developed LC-KSVD by solving the following optimization problem

$$\begin{aligned} \min_{\mathbf{D}, \mathbf{W}, \mathbf{A}, \mathbf{Z}} \|\mathbf{X} - \mathbf{D}\mathbf{Z}\|_F^2 + \alpha \|\mathbf{Q} - \mathbf{A}\mathbf{Z}\|_F^2 + \beta \|\mathbf{H} - \mathbf{W}\mathbf{Z}\|_F^2, \\ \text{s.t. } \|\mathbf{z}_i\|_0 \leq T_0 \end{aligned} \quad (3)$$

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

### Transform learning

TL was presented by Ravishankar et al.<sup>17,18</sup> and the objective function of TL is formulated as follows

$$\min_{\mathbf{T}, \mathbf{Z}} \|\mathbf{TX} - \mathbf{Z}\|_F^2 + \mu (\|\mathbf{T}\|_F^2 - \log \det \mathbf{T}) + \lambda \|\mathbf{Z}\|_1 \quad (4)$$

where  $\mathbf{T} = [\mathbf{t}_1; \mathbf{t}_2; \dots; \mathbf{t}_K] \in \mathbb{R}^{K \times n}$  is the transform matrix,  $\mathbf{t}_i \in \mathbb{R}^{1 \times n}$ .  $-\log \det \mathbf{T}$  imposes a full rank constraint to avoid the trivial solution ( $\mathbf{T} = 0, \mathbf{Z} = 0$ ),  $\|\mathbf{T}\|_F^2$  is used to address the scale ambiguity of the transform matrix. TL is an unsupervised representation learning approach, thus it is unsuitable for

classification tasks. Recently, Maggu et al.<sup>19</sup> developed a supervised version of TL by solving the following problem

$$\begin{aligned} \min_{\mathbf{T}, \mathbf{Z}, \mathbf{W}} \|\mathbf{TX} - \mathbf{Z}\|_F^2 + \mu (\|\mathbf{T}\|_F^2 - \log \det \mathbf{T}) \\ + \lambda \|\mathbf{Z}\|_1 + \beta \|\mathbf{H} - \mathbf{W}\mathbf{Z}\|_F^2 \end{aligned} \quad (5)$$

where  $\mathbf{H}$  and  $\mathbf{W}$  are the label matrix of training data and parameters of linear classifier, respectively.

### Label consistent TL

As mentioned earlier, we want to employ the label information of both training data and dictionary atoms to boost the discriminative capability of TL, and the objective function of our proposed LCTL is formulated as

$$\begin{aligned} \min_{\mathbf{T}, \mathbf{Z}, \mathbf{A}} \|\mathbf{TX} - \mathbf{Z}\|_F^2 + \mu (\|\mathbf{T}\|_F^2 - \log \det \mathbf{T}) \\ + \lambda \|\mathbf{Z}\|_1 + \alpha \|\mathbf{Q} - \mathbf{A}\mathbf{Z}\|_F^2 \end{aligned} \quad (6)$$

**Table 2.** Recognition accuracy and training time on the Extended Yale B database.

Methods	Accuracy (%)	Training time (s)
SRC	80.5	No need
LLC	82.2	No need
K-SVD	93.1	1.79
D-KSVD	94.1	32.43
LC-KSVD	95.0	38.96
DTL	94.3	3.91
LCTL	<b>95.2</b>	<b>1.27</b>

Note: Bold values signify the best recognition accuracy and the least training time.  
DTL: discriminative transform learning; LCTL: label consistent transform learning.



**Figure 2.** Example images from the Extended Yale B database.

where  $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_N] \in \mathbb{R}^{K \times N}$  is an ideal representation matrix formed by the label information of training data and dictionary atoms,  $\mathbf{q}_i = [0, 0, \dots, 1, 1, \dots, 0, 0]^T \in \mathbb{R}^{K \times 1}$ . The entries in  $\mathbf{q}_i$  are 1 when the training samples and the dictionary atoms have the same class label. For instance, suppose  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_6]$  and  $\mathbf{T} = [\mathbf{t}_1; \mathbf{t}_2; \dots; \mathbf{t}_6]$ , where  $\mathbf{x}_1, \mathbf{x}_2, \mathbf{t}_1$ , and  $\mathbf{t}_2$  belong to the first class,  $\mathbf{x}_3, \mathbf{x}_4, \mathbf{t}_3$ , and  $\mathbf{t}_4$  belong to the second class, and  $\mathbf{x}_5, \mathbf{x}_6, \mathbf{t}_5$ , and  $\mathbf{t}_6$  belong to the third class, then  $\mathbf{Q}$  can be defined as

$$\mathbf{Q} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

Equation (6) can be solved by alternatively updating  $\mathbf{T}$ ,  $\mathbf{Z}$ , and  $\mathbf{A}$ , and the optimization process is presented as follows:

*Update T:* When  $\mathbf{Z}$  and  $\mathbf{A}$  are fixed, we can obtain the following optimization problem with respect to  $\mathbf{T}$

$$\min_{\mathbf{T}} \|\mathbf{TX} - \mathbf{Z}\|_F^2 + \mu \left( \|\mathbf{T}\|_F^2 - \log \det \mathbf{T} \right) \quad (7)$$

Equation (7) has closed-form solution.<sup>18</sup> First, we factorize  $\mathbf{XX}^T + \mu \mathbf{I}$  as  $\mathbf{LL}^T$ , then let  $\mathbf{L}^{-1} \mathbf{XZ}^T = \mathbf{U}\mathbf{S}\mathbf{V}^T$  be a

full singular value decomposition (SVD). Finally, the solution to Equation (7) is given by

$$\mathbf{T} = 0.5 \mathbf{V} \left( \mathbf{\Sigma} + (\mathbf{\Sigma}^2 + 2\mu \mathbf{I})^{\frac{1}{2}} \right) \mathbf{U}^T \mathbf{L}^{-1} \quad (8)$$

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

$$\min_{\mathbf{Z}} \|\mathbf{TX} - \mathbf{Z}\|_F^2 + \lambda \|\mathbf{Z}\|_1 + \alpha \|\mathbf{Q} - \mathbf{AZ}\|_F^2 \quad (9)$$

Equation (9) can be rewritten as

$$\min_{\mathbf{Z}} \left\| \begin{pmatrix} \mathbf{TX} \\ \sqrt{\alpha} \mathbf{Q} \end{pmatrix} - \begin{pmatrix} \mathbf{I} \\ \sqrt{\alpha} \mathbf{A} \end{pmatrix} \mathbf{Z} \right\|_F^2 + \lambda \|\mathbf{Z}\|_1 \quad (10)$$

**Table 3.** Recognition accuracy and training time on the AR database.

Methods	Accuracy (%)	Training time (s)
SRC	66.5	No need
LLC	69.5	No need
K-SVD	86.5	1.97
D-KSVD	88.8	45.64
LC-KSVD	93.7	54.48
DTL	95.7	4.75
LCTL	<b>96.2</b>	<b>1.32</b>

Note: Bold values signify the best recognition accuracy and the least training time.

DTL: discriminative transform learning; LCTL: label consistent transform learning.



**Figure 3.** Example images from the AR database.



Let  $\mathbf{X}_{new} = \begin{pmatrix} \mathbf{TX} \\ \sqrt{\alpha}\mathbf{Q} \end{pmatrix}$  and  $\mathbf{A}_{new} = \begin{pmatrix} \mathbf{I} \\ \sqrt{\alpha}\mathbf{A} \end{pmatrix}$ , equation (10) can be reformulated as

$$\min_{\mathbf{Z}} \|\mathbf{X}_{new} - \mathbf{A}_{new}\mathbf{Z}\|_F^2 + \lambda\|\mathbf{Z}\|_1 \quad (11)$$

Equation (11) is the  $\ell_1$  norm constrained sparse coding problem which can be solved by the Homotopy algorithm.<sup>20</sup>

*Update A:* When fixing the other variables, equation (6) is degenerated into the following problem with respect to  $\mathbf{A}$

$$\min_{\mathbf{A}} \|\mathbf{Q} - \mathbf{AZ}\|_F^2 + \eta_1\|\mathbf{A}\|_F^2 \quad (12)$$

Equation (12) has the following closed-form solution

$$\mathbf{A} = \mathbf{QZ}^T(\mathbf{ZZ}^T + \eta_1\mathbf{I})^{-1} \quad (13)$$

When the training phase is completed, we can obtain the transform matrix  $\mathbf{T}$ , and the representation matrix of training data can be computed by  $\mathbf{Z}_{tr} = \mathbf{TX}$ . We employ a simple yet effective linear classifier to classify the test samples. A linear classifier  $\mathbf{W}$  can be learned based on the label matrix  $\mathbf{H}$  of training data and the representation matrix  $\mathbf{Z}_{tr}$ , and the following problem is utilized to learn  $\mathbf{W}$

$$\mathbf{W}^* = \arg \min_{\mathbf{W}} \|\mathbf{H} - \mathbf{WZ}_{tr}\|_F^2 + \eta_2\|\mathbf{W}\|_F^2 \quad (14)$$

Equation (14) has the following closed-form solution

$$\mathbf{W}^* = \mathbf{HZ}_{tr}^T(\mathbf{Z}_{tr}\mathbf{Z}_{tr}^T + \eta_2\mathbf{I})^{-1} \quad (15)$$

For a test sample  $\mathbf{x}$ , first we can compute its coding vector by  $\mathbf{z} = \mathbf{Tx}$ , then the identity of  $\mathbf{x}$  is given by

$$\text{label}(\mathbf{x}) = \arg \max_i \mathbf{W}^*\mathbf{z} \quad (16)$$

## Experimental results and analysis

In this section, we evaluate the classification performance of our proposed LCTL on four benchmark datasets: the Yale database, the Extended Yale B database, the AR database, and the Scene 15 dataset. To illustrate the superiority of LCTL, we compare LCTL with the following approaches: SRC,<sup>21</sup> LLC,<sup>22</sup> K-SVD,<sup>4</sup> D-KSVD,<sup>5</sup> LC-KSVD,<sup>6</sup> and DTL.<sup>19</sup> Apart from the recognition accuracy, we also present the training time (in seconds) of all the competing methods. All experiments are run with MATLAB R2019a under Windows 10 on a PC equipped with Intel i9-8950HK 2.90 GHz CPU and 16 GB RAM.

**Table 4.** Recognition accuracy and training time on the Scene 15 dataset.

Methods	Accuracy (%)	Training time (s)
SRC	91.8	No need
LLC	89.2	No need
K-SVD	91.6	<b>7.74</b>
D-KSVD	89.9	99.20
LC-KSVD	93.8	111.01
DTL	91.3	11.45
LCTL	<b>98.5</b>	8.56

Note: Bold values signify the best recognition accuracy and the least training time.

DTL: discriminative transform learning; LCTL: label consistent transform learning.



**Figure 4.** Example images from the Scene 15 dataset.

### Experiments on the Yale database

The Yale database consists of 165 images for 15 subjects, each individual has 11 images. These images have illumination and expression variations. Figure 1 shows some example images from this database. All the images are resized to  $24 \times 24$ , resulting in a 576-dimensional vector. Six images per subject are randomly selected for training and the remaining for testing. The learned dictionary contains 60 atoms, i.e. each class has four atoms. Experimental results are summarized in Table 1. We can see that LCTL has the highest recognition accuracy and the least training time. Specifically, the accuracy gain of LCTL over DTL is 2.6%.

### Experiments on the Extended Yale B database

The Extended Yale B face database contains 2414 images of 38 individuals. Each individual has 59–64 images taken under different illumination conditions; example images from this dataset are shown in Figure 2. In our experiments, each  $192 \times 168$  image is projected onto a 504-dimensional space via random projection. We randomly select half of the images per category as training and the remaining for testing.

The dictionary consists of 570 items, which corresponds to an average of 15 atoms per person. Experimental results are listed in Table 2. Although the accuracy gain of LCTL over LC-KSVD is not that significant, LCTL is about 31 times faster than LC-KSVD.

### Experiments on the AR database

The AR database has more than 4000 face images of 126 subjects with variations in facial expression, illumination conditions, and occlusions. Figure 3 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. Each  $165 \times 120$  face image is projected onto a 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 dictionary items, i.e. five atoms per person. Table 3 shows the experimental results on this database. Although the AR database consists of occluded images which pose challenge to face recognition, LCTL achieves the best recognition accuracy and the highest efficiency.

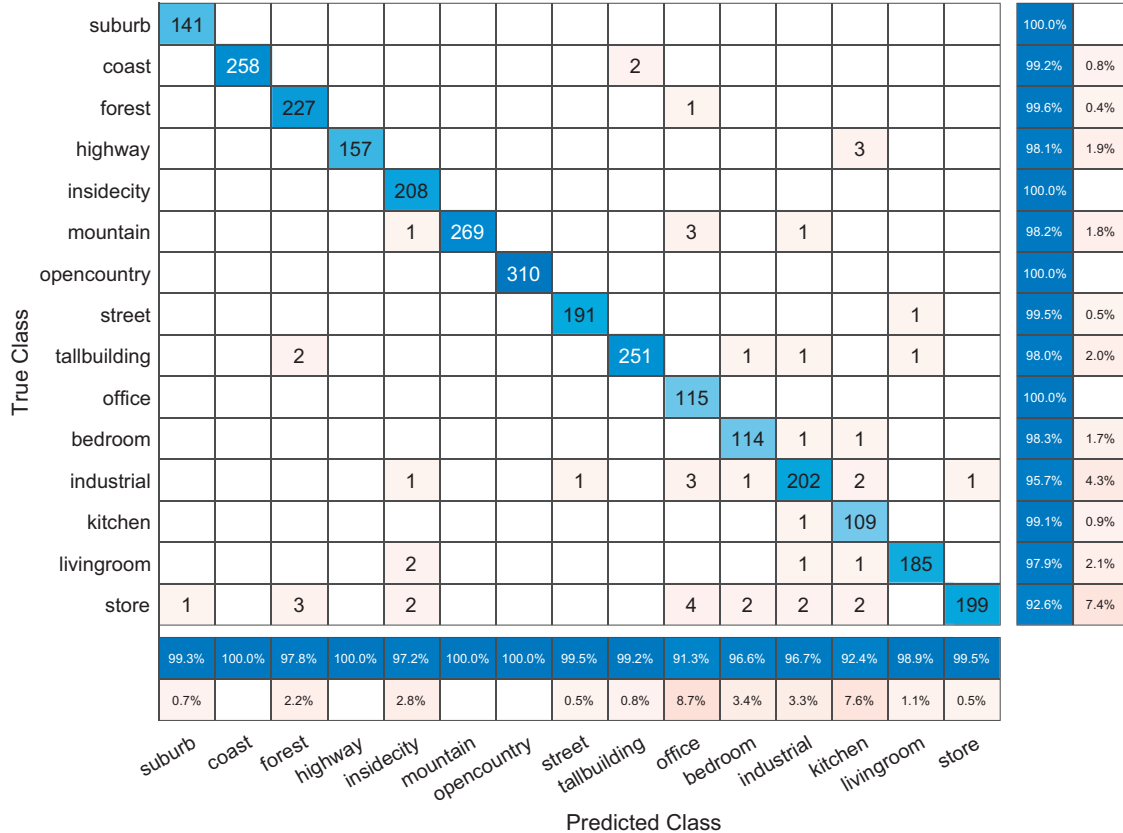
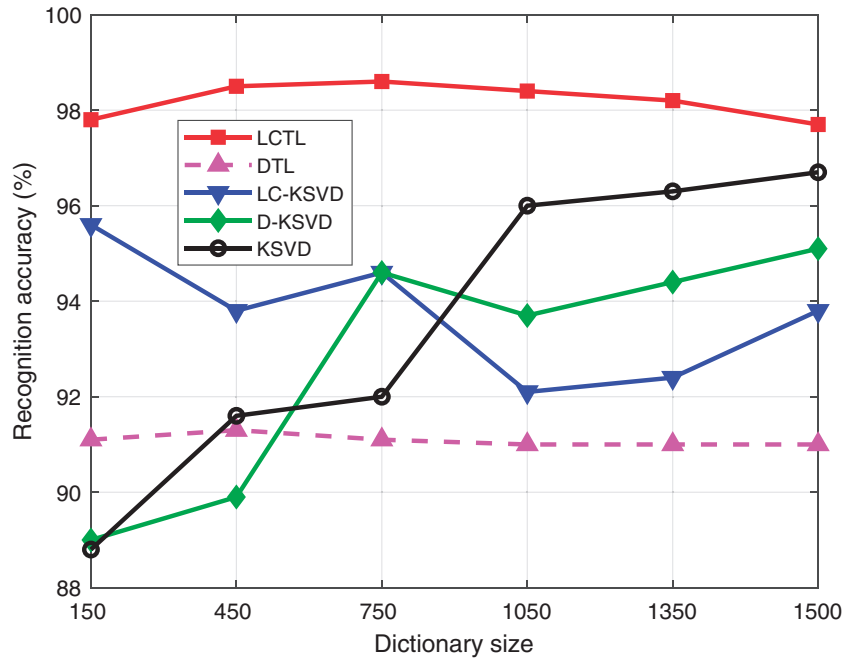


Figure 5. Confusion matrix on the Scene 15 dataset.



**Figure 6.** Recognition accuracy on the Scene 15 dataset with varying dictionary size.

### Experiments on the Scene 15 dataset

Scene 15 dataset contains 15 natural scene categories introduced by Lazebnik et al.,<sup>23</sup> 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 4. For fair comparison, we employ the 3000-dimensional SIFT-based features used in LC-KSVD.<sup>6</sup> 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 presented in Table 4. One can see that LCTL outperforms DTL by a large margin, and it makes 7.2% and 4.7% improvement over DTL and LC-KSVD, respectively. The confusion matrix for LCTL is shown in Figure 5. As can be seen from Figure 5, LCTL obtains over 99% recognition accuracy for the categories of suburb, coast, forest, insidicity, open-country, street, office, and kitchen.

To examine how the dictionary size influences the performance of DL approaches, here we compare LCTL with DTL, LC-KSVD, D-KSVD, and K-SVD. We evaluate all the competing methods under dictionary sizes of 150, 450, 750, 1050, 1350, and 1500, which correspond to 10, 30, 50, 70, 90, and 100 atoms per category, respectively. Recognition accuracy with different dictionary sizes is plotted in Figure 6. We can observe that LCTL consistently outperforms all other approaches for all dictionary sizes.

### Conclusion

In this paper, we present an LCTL algorithm. To enhance the discriminative ability of TL and apply it to pattern classification tasks, a label consistency constraint is incorporated into the framework of TL to form a unified objective function. The variables in our optimization problem are updated alternatively and some of them have closed-form solutions. Experiments are conducted on three benchmark face databases and one scene dataset, and the results validate the effectiveness of our proposed LCTL. Moreover, LCTL shows high efficiency in terms of training time, which can be used in real-time scenarios.

### Acknowledgements

The authors would like to thank Zhuolin Jiang for providing the source code of LC-KSVD at <http://users.umi.acs.umd.edu/~zhuolin/projectlksvd.html>, and Jyoti Maggu for releasing the code of DTL at [https://github.com/maggujyoti/transform\\_learning\\_hyperspectral\\_image\\_classification](https://github.com/maggujyoti/transform_learning_hyperspectral_image_classification).

### Declaration of conflicting interests


The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this

article: This work was supported by the National Natural Science Foundation of China (Grant Nos. 61672265, U1836218), the 111 Project of Ministry of Education of China (Grant No. B12018), the Postgraduate Research & Practice Innovation Program of Jiangsu Province under Grant No. KYLX\_1123, the Overseas Studies Program for Postgraduates of Jiangnan University and the China Scholarship Council (Grant No. 201706790096).

## ORCID iD

He-Feng Yin  <https://orcid.org/0000-0001-5831-1475>

## References

1. Song X, Chen Y, Feng ZH, et al. Collaborative representation based face classification exploiting block weighted LBP and analysis dictionary learning. *Pattern Recognit* 2019; 88: 127–138.
2. Li H and Wu XJ. Multi-focus image fusion using dictionary learning and low-rank representation. In: *ICIG*, Shanghai, China, 13–15 September 2017, Springer, Cham, pp.675–686.
3. Li S, Shao M and Fu Y. Person re-identification by cross-view multi-level dictionary learning. *IEEE Trans Pattern Anal Mach Intell* 2017; 40: 2963–2977.
4. Aharon M, Elad M and Bruckstein AK. SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans Signal Process* 2006; 54: 4311–4322.
5. Zhang Q and Li B. Discriminative K-SVD for dictionary learning in face recognition. In: *CVPR*, San Francisco, California, 13–18 June 2010, pp.2691–2698.
6. Jiang Z, Lin Z and Davis LS. Learning a discriminative dictionary for sparse coding via label consistent K-SVD. In: *CVPR*, Colorado Springs, Colorado, 20–25 June 2011, pp.697–1704.
7. Yang M, Zhang L, Feng X, et al. Fisher discrimination dictionary learning for sparse representation. In: *CVPR*, Colorado Springs, Colorado, 20–25 June 2011, pp.543–550.
8. Wang D and Kong S. A classification-oriented dictionary learning model: explicitly learning the particularity and commonality across categories. *Pattern Recognit* 2014; 47: 885–898.
9. Li Z, Lai Z, Xu Y, et al. A locality-constrained and label embedding dictionary learning algorithm for image classification. *IEEE Trans Neural Netw Learning Syst* 2015; 28: 278–293.
10. Zhao Z, Feng G, Zhang L, et al. Novel orthogonal based collaborative dictionary learning for efficient face recognition. *Knowledge Based Syst* 2019; 163: 533–545.
11. Rubinstein R, Peleg T and Elad M. Analysis K-SVD: a dictionary-learning algorithm for the analysis sparse model. *IEEE Trans Signal Process* 2012; 61: 661–677.
12. Gu S, Zhang L, Zuo W, et al. Projective dictionary pair learning for pattern classification. In: *NeurIPS*, Montreal, Canada, 8–13 December 2014, pp.793–801.
13. Guo J, Guo Y, Kong X, et al. Discriminative analysis dictionary learning. In: *AAAI*, Phoenix, USA, 12–17 February 2016, pp.1617–1623.
14. Yang M, Chang H and Luo W. Discriminative analysis-synthesis dictionary learning for image classification. *Neurocomputing* 2017; 219: 404–411.
15. Wang J, Guo Y, Guo J, et al. Synthesis linear classifier based analysis dictionary learning for pattern classification. *Neurocomputing* 2017; 238: 103–113.
16. Wang Q, Guo Y, Guo J, et al. Synthesis K-SVD based analysis dictionary learning for pattern classification. *Multimed Tools Appl* 2018; 77: 17023–17041.
17. Ravishankar S and Bresler Y. Learning sparsifying transforms. *IEEE Trans Signal Process* 2012; 61: 1072–1086.
18. Ravishankar S, Wen B and Bresler Y. Online sparsifying transform learning – part I: algorithms. *IEEE J Sel Top Signal Process* 2015; 9: 625–636.
19. Maggu J, Aggarwal HK and Majumdar A. Label-consistent transform learning for hyperspectral image classification. *IEEE Geosci Remote Sensing Lett* 2019; 16: 1502–1506.
20. Yang AY, Sastry SS, Ganesh A, et al. Fast  $\ell_1$ -minimization algorithms and an application in robust face recognition: a review. In: *ICIP*, Hong Kong, China, 26–29 September 2010, pp.1849–1852.
21. Wright J, Yang AY, Ganesh A, et al. Robust face recognition via sparse representation. *IEEE TPAMI* 2009; 31: 210–227.
22. Wang J, Yang J, Yu K, et al. Locality-constrained linear coding for image classification. In: *CVPR*, San Francisco, California, 13–18 June 2010, pp.3360–3367.
23. Lazebnik S, Schmid C and Ponce J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: *CVPR*, New York, USA, 17–22 June 2006, pp.2169–2178.