Copyright ⓒ2017 KSII

Domain Adaptation Image Classification Based on Multi-sparse Representation

Xu Zhang[1*], Xiaofeng Wang[2], Yue Du[3] and Xiaoyan Qin[4]
[1] College of Electronic Information Engineering, Suzhou Vocational University
Suzhou, 215104 China
[2] Department of Computer Science and Technology, Hefei University
Hefei, 230601 China
[3] Sixth Department, Army Officer Academy of PLA
Hefei, 230031 China
[4] Eleventh Department, Army Officer Academy of PLA
Hefei, 230031 China
[e-mail: zhangxu21cn@163.com]
*Corresponding author: Xu Zhang

Received August 10, 2016; revised September 26, 2016; accepted October 3, 2016; published May 31, 2017

Abstract

Generally, research of classical image classification algorithms assume that training data and testing data are derived from the same domain with the same distribution. Unfortunately, in practical applications, this assumption is rarely met. Aiming at the problem, a domain adaption image classification approach based on multi-sparse representation is proposed in this paper. The existences of intermediate domains are hypothesized between the source and target domains. And each intermediate subspace is modeled through online dictionary learning with target data updating. On the one hand, the reconstruction error of the target data is guaranteed, on the other, the transition from the source domain to the target domain is as smooth as possible. An augmented feature representation produced by invariant sparse codes across the source, intermediate and target domain dictionaries is employed for across domain recognition. Experimental results verify the effectiveness of the proposed algorithm.

Keywords: Image classification, domain adaptation, sparese coding, bag of visual words, dictionary learning

This research was supported by the research grant from the Anhui Natural Science Foundation(No.1608085QF144 and No.1508085QF114).

https://doi.org/10.3837/tiis.2017.05.016 **ISSN : 1976-7277**

1. Introduction

Image classification has been one of the hot topics in the field of computer vision, and plays a key role in many applications such as image retrieval[1],image annotation[2],multimedia answering[3] and so on. At present, thanks to the diffusion of image acquisition equipment, imaging system and mobile Internet technology, numerous images can be easily and quickly availed from many channels. However, due to different ways of acquisition, richness forms of expression, background of application and so on, leveraging over and across such sources has proved challenging.

In the issues of classic image classification and object recognition [4-6], the training and testing images are assumed under the same distribution. The general approach is to select a part of images as training data for classification model, while another part of images in the dataset as testing data. Namely, the training and testing data are captured from the same image dataset, subjecting to the same distribution. Yet in many practical applications, the assumption has certain limitation. Due to the influence of illumination, pose, scale transformation, occlusion and image resolution, etc. [7], images with the same labels often have large intra-class appearance variability under potential distribution. This phenomenon is also called dataset bias or domain shifts. In this case, if the classifier from the training set is directly applied to the testing set, the performance will be greatly degraded [8]. Face recognition experimental results have confirmed if the model trained by Orientals faces is applied to westerners, recognition performance decreased near 40% [9]. **Fig. 1** shows an example of dataset bias. Although labels of images in the two datasets are computer monitor, appearances of the images are obviously different.

Fig. 1. Examples of dataset bias

Therefore, it is significant to investigate how to adapt classification model to testing data with different distribution from training data. Domain adaptation [10] is an effective technique addressing thisproblem, which is also considered as a representative method of transfer learning [11]. Domain adaptation mainly refers to two related concepts: source domain and target domain. The source domain contains abundant supervision data for learning classifier and the target domain refers to the dataset assumed to have different distribution from the source domain, containing no or few labels. The main objective of domain adaptation is to apply classifiers trained in the source domain to the target domain for pursuing better performance. According to the availability of labeled data in the target domain, domain adaptation methods can be divided into unsupervised domain adaptation where no labeled data is contained in the target domain, and semi-supervised domain adaptation where a small amount of labeled data is contained. In the real-world applications, unsupervised domain adaptation is more universal and representative. But, it is no doubt more challenging.

A typically approach of unsupervised domain adaptation methods is to try to find suitable representations for reducing distribution difference between the source and the target domains. It is well known that real high-dimensional data lies in the low-dimensional manifold of high dimensional space, and can be efficiently represented in the low-dimensional subspace[12].Thus, subspace representationscan be exploited for modeling the source and target domains. Based on sparse representation [13], given an over-complete dictionary, data signals in the same subspace can be linearly reconstructed with a small number of atoms or bases. In recent years,a variety of dictionary learning algorithms have been proposed, and in those algorithms, the training data is supposed originating from the same domain with the same distribution. In domain adaptation image classification, dictionary learning should pay more attention on domain shifts. Due to the presence of domain shifts, the testing data does not lie in the linear span of training data. If the target data is decomposed by the dictionary learned from the source domain, large reconstruction error will be generated. In particular, samples of the source and target domains with the same labels do not necessarily have similar sparse coding coefficients. All of these factors lead to decrease of classification performance. Therefore, it is crucial to leverage the data of target domain in dictionary learning for adaptation.

In view of this, to eliminate the impacts in image classification while training images and testing images are captured from different domains, we propose an unsupervised domain adaptation method based on multi-sparse representation in this paper. In order to reduce distribution difference in the source and target domains, the method supposes existence of intermediate domains between the two domains, and subspace representation is utilized to model different domains. Online dictionary learning is introduced to model intermediate subspaces. Each intermediate domain is represented by a dictionary, and the source and target domains can be effectively connected through the intermediate domains. Let D_0 be the initial dictionary learned from the source domain. The intermediate domain dictionaries $D_1, D_2, ..., D_n$ can be sequentially learned by online dictionary learning scheme, where the samples in the target domain are regarded as dynamic data sequences. Starting from the source domain dictionary D_0, with continuous updating of samples in the target domain, intermediate domain dictionaries gradually adapt to the target data.

The main contributions in this paper are as follows:(a) the source and target domains are connected through intermediate domains and domain shifts are effectively mitigated in the new spanned data signal space. (b) with the target data updating, each intermediate domain is modeled by a dictionary using online dictionary learning.On the one hand, the reconstruction error of the target data is guaranteed to be minimal;on the other hand, the transition from the source domain to the target domain is as smooth as possible. (c) An augmented feature produced by invariant sparse codes across the source, intermediate and target domain dictionaries is employed for across domain recognition.

2. Related Work

The process of image classification can be briefly summarized as follows: collect training images and detect features firstly. Then learn classification model based on the relevant criteria such as empirical error minimization.Finally utilize the model for testing images. When distribution of the testing data is consistent with the training data, good classification results will be obtained. On the contrary, the poor results will be produced. Domain adaptation aims at addressing the problem.

Based on different stages of image classification process, three different types of domain adaptation methods are proposed: instance adaptation [14-15], feature adaptation [16-22] and model adaptation [23-24]. In the method of instance adaptation, instances of the source domain are weighted resampled for consistent distribution with the target domain. The kind of method is intuitive. But, if there arevisible discrepancies between instances of the source and target domains, performance of such methods is poor [14]. The model adaptation method directly conducts adaptation during training the classifier. A commonly used trick is to introduce domain distance constraint in the cost function. Based on this kind idea, Yang et al. [23] proposed an adaptative SVM algorithm, while Bruzzone et al. [24] proposed a Domain Adaptation SVM algorithm. The difference of source domain and target domain is carefully considered and the cost function in the source domain is modified in model training process.

Recently, many studies focus on the method of feature adaptation. The basic idea of this kind of methods is to learn common feature space, where distribution of the source domain and the target domain is consistent as soon as possible. The key idea of [16] is to map the source data to the target domain by learning a regularized transformation via information theoretic metric. In [17], in order toreduce the distribution, all the data of source and target domains is projected into a series of intermediate subspaces, while Gong et al. [18] defined an infinite number of subspaces based on a kernel method. However, these methods only focus on alignment of subspaces' bases rather than distribution of the projected data. Jhuo et al. [19] also pursued projection transform, where the source data is represented by linear combination of the target data exploiting low-rank and sparse decomposition. But this method belongs to semi-supervised domain adaptation, which relies on labeled data of the target domain for cross domain classification. Ni et al. [20] proposed to subspaces through dictionary learning to link the source and target domains. But during dictionary learning, the method requires all the samples of target domain all at once. Strictly speaking, the method has certain limitation. The techniques in [21-22] attempted to shorten the distance across two domains by domain similarity measuring by maximum mean discrepancy in the learned feature space.

3. Proposed Method

Sparse representation and dictionary learning have been widely applied in many applications of computer vision and receive remarkable success especially in the task of image classification. The method of image classification based on sparse representation can be simply described as the following procedure. Given a proper over-complete dictionary, images are encoded by sparse coding technique and the sparse coding coefficients are leveraged as feature representations for final classification. Thus, it can be seen that the learned dictionary determines the performance to a large extent. In order to reduce distribution difference of the source and target domains, robust dictionary learning and online dictionary learning are utilized for feature representation in the proposed method.

3.1 Robust Dictionary Learning

In sparse representation, the general framework of traditional dictionary learning can be abstracted as the following objective function minimization:

$$D = \min_{D,\{\alpha_i\}} \sum_{i=1}^{N} \|x_i - D\alpha_i\|_2^2 + \lambda \|\alpha_i\|_1 \tag{1}$$

Here $X=[x_1,x_2,...,x_N]$ and $A=[\alpha_1,\alpha_2,...,\alpha_N]$ are reconstructed data signal and corresponding sparse coding coefficients respectively. $\|x_i-D\alpha_i\|_2^2$ is reconstruction error term, and D is the learned dictionary. Each reconstructed data signal can be linearly decomposed by a small number of bases in D with corresponding coding coefficients α_i. The optimization in Eq. (1) is not convex for D and A simultaneously, but it is convex for D when A is fixed and vice versa.

Conventional dictionary learning methods employ ℓ^2-norm for measuring the reconstruction error. However, in terms of robustness, choosing ℓ^1-norm instead of ℓ^2-norm is more effective. Wagner et al. [25] proposed a robust face recognition algorithm using ℓ^1-norm for measuring the reconstruction error.Experimental results show that when the face is partially occluded, good recognition is still achieved. Similarly, Zhao et al. [26] argued it is more effective modeling the reconstruction error by a heavy-tailed iid-Laplacian distribution, which is known to handle outliers better. Therefore, it will enhance the robustness to noise and large outliers while measuring the reconstruction error using ℓ^1-norm. The problem of Eq. (1) can be reformed as the following optimization:

$$D=\min_{D,A}\|X-DA\|_1+\lambda\|A\|_1 \tag{2}$$

X is a matrix consisting of training data as the column vectors and A is coefficient matrix built with the same manner. $\|A\|_1$ denotes the ℓ^1-norm of A, which sums up the absolute values of entries in A. The optimization in Eq. (2) is also not jointly convex for D and A. A commonly used method in dictionary learning is to optimize D and A iteratively and alternately with each other frozen. The two steps are called as robust sparse coding and robust dictionary update respectively.

1) Robust sparse coding

The robust sparse coding step optimizes the following objective function for solving coding coefficients assuming the dictionary being fixed:

$$A=\min_{A}\|X-DA\|_1+\lambda\|A\|_1 \tag{3}$$

In the process of solving the problem, the reconstructed samples x_i (columns of X) are assumed to be independent from each other. So, Eq. (3) can be further broken into the following i independent subproblems:

$$\alpha_i=\min_{\alpha_i}\|x_i-D\alpha_i\|_1+\lambda\|\alpha_i\|_1 \quad i=1,...,N \tag{4}$$

2) Robust dictionary update

With the coefficient matrix being constant and discarding the second term in Eq. (2), the dictionary is updated as follows:

$$D=\min_{D}\|X-DA\|_1 \tag{5}$$

Similarly, the atoms in D are independent from each other and thus can be updated separately.

$$d_k=\min_{d_k}\|X-d_k\alpha^k\|_1 \quad k=1,...,K \tag{6}$$

d_k is the k-th atom of D and α^k is the k-th row corresponding to d_k. Because most of the coding coefficients are not exactly zero but very close to, the columns of X whose coding coefficients larger than a given small threshold are chosen for calculating, which is beneficial to updating speed and over-fitting avoiding. Besides, d_k is also normalized to avoid arbitrary small values. Zhao et al. [26] also identify with the strategy that updating d_k involving (8) at the same time will speed up the convergence rat of the formula (6). So, Eq. (7) and Eq. (8) are iteratively optimized until the convergence on d_k is reached.

$$d_k^i = \min_{d_k^i} \left\| x^i - d_k^i \alpha^k \right\|_1 \quad i = 1, \ldots, N, \ k = 1, \ldots, K \tag{7}$$

$$\alpha_j^k = \min_{\alpha_j^k} \left\| x_j - d_k \alpha_j^k \right\|_1 \quad j = 1, \ldots, M, \ k = 1, \ldots, K \tag{8}$$

3.2 Online Dictionary Learning

In traditional dictionary learning methods, all the training samples will be employed in each iteration, which methods are also called batch dictionary learning. However, in some specific applications involving large scale data and dynamic data such as video sequences or target tracking and so on, with data size growing, it is obviously not suitable to learn dictionary through traditional dictionary learning methods such as K-SVD. Pointing at this problem, online dictionary learning (ODL) method [27] is proposed, where the signals are processed one at a time or in mini-batches for online updating.

Being similar to batch dictionary learning methods, ODL also involves two steps: sparse coding and dictionary updating. The sparse coding problem with fixed dictionary is an ℓ^1 -regularized linear least-squares problem. In ODL, the decomposition α_t of x_t is computed over the dictionary D_{t-1} obtained at the previous iteration.

$$\alpha_t = \arg\min_{\alpha \in R^K} \left(\frac{1}{2} \left\| x_t - D_{t-1}\alpha \right\|_2^2\right) + \lambda \left\| \alpha \right\|_1 \tag{9}$$

During the dictionary updating step, with the coefficients being constant, the dictionary is updated by minimizing the objective function of Eq. (10):

$$D_t = \arg\min_{D \in C} \frac{1}{t} \sum_{i=1}^{t} \left(\frac{1}{2} \left\| x_i - D\alpha_i \right\|_2^2 + \lambda \left\| \alpha_i \right\|_1\right) \tag{10}$$

Choose D_{t-1} obtained at the previous iteration as beginning, and the dictionary D_t is updated by block coordinate descent algorithm. The above formula can be further expressed as:

$$D_t = \arg\min_{D \in C} \frac{1}{t} \left(\frac{1}{2} Tr(\mathrm{D}^T \mathrm{D} \mathrm{A}_t) - Tr(\mathrm{D}^T B_t) \right) \tag{11}$$

3.3 Domain Adaptation Base on Multi-sparse Coding

Although images with the same labels can be highly diversified, they also share lots of similarity and correlation due to the homogeneity. Therefore, in the paper, in order to reduce distribution difference of the source and target domains, the proposed method supposes existence of intermediate domains between the two domains, and subspace representation is utilized to model different domains. Each intermediate domain is represented by a dictionary using online dictionary learning. An augmented feature produced by invariant sparse codes across the source, intermediate and target domain dictionaries is employed for across domain recognition.

Let $X^S \in R^{n\times N_S}$ and $X^T \in R^{n\times N_T}$ be the samples from the source domain S and target domain T, where n is the dimension of data. N_S and N_T are number of samples in the source and target domains respectively. The proposed approach can be summarized in Algorithm 1.

Algorithm 1: Domain adaptation image classification based on multi-sparse coding

---

Input: samples of the source domain X^S and target domain X^T, parameter η (for BoVW representation), λ;

Output : the augmented feature vector for across domain recognition;

Step 1: Extract feature of samples $X^S = \{x_1^S, x_2^S, ..., x_{N_S}^S\}$ from the source domain and learn dictionary D through robust dictionary learning;

Step 2: Decompose the source and target data with dictionary D based on sparse coding, the BoVW vectors with fixed length of the source and target data denoted as $\{x_{BoVW1}^S, x_{BoVW2}^S, ..., x_{BoVWN_S}^S\}$ and $\{x_{BoVW1}^T, x_{BoVW2}^T, ..., x_{BoVWN_T}^T\}$ respectively;

Step 3: Learn initial dictionary D_0 in BoVW vectors of the source domain using traditional dictionary learning method such as K-SVD;

Step 4: For each $\{x_{BoVW1}^T, x_{BoVW2}^T, ..., x_{BoVWN_T}^T\}$ in the target domain, update D_0 using equation (10) and get the next intermediate dictionary $\{D_k\}_{k=1}^{N_T}$;

Step 5: Decompose the BoVW vectors of target data with D_0:

$$\alpha_i = \min_{\alpha_i} \left\| x_{BoVWi}^T - D_0\alpha_i \right\|_2^2 + \lambda \left\| \alpha_i \right\|_1 \ , \ i = 1, ..., N_T$$

α_i is the sparse coding coefficient. The target data x_{BoVWi}^T is represented as the augmented feature vector of $[(D_0\alpha_i)^T, (D_1\alpha_i)^T, ..., (D_{N_T}\alpha_i)^T]^T$, which is employed for cross domain recognition after dimension reduction such as PCA.

---

4. Experimental Results and Analysis

In this section, we evaluate the proposed approach on the Office&Caltech[18] and Bing Caltech-256 datasets[28], which are frequently used datasets for visual domain adaptation methods.

4.1 Results on Office&Caltech Dataset

Following the work of [18], the dataset of Caltech-256 is involved in the Office dataset and a modified version of the dataset is build, where the Caltech-256 is treaded as an independent domain. The Office dataset contains three independent domains: Amazon, dSLR and Webcam subsets, where different types and quantities of office supplies such as keyboard, printer, bookcase, telephone, and mobile phone and so on are provided. The images of Amazon domain are downloaded from the www.amazon.com shopping site, where the commodity images are obtained with a moderate resolution and good imaging condition. Images of the dSLR domain are captured by a digital SLR camera in realistic natural lighting conditions with high resolution and low noise level, while images of the Webcam domain are captured by a Web camera with low resolution and high noise level. Each domain consists of 31 categories, and 2790, 423 and 795 images are contained in Amazon, dSLR and Webcam respectively. **Fig. 2** shows example images from Amazon, dSLR, Webcam and Caltech domains. For fair comparisons with other methods, we follow the similar experiment protocols as the previous work [16] did. All images preprocessed into gray scale are adjusted to the same size and 64-dimensional SURF features are extracted at interest point locations. With the robust dictionary trained from a subset of Amazon images, all the images are represented as the BoVW features with 300-bin histograms.

（a） Amazon （b） dSLR

（c） Webcam （d） Caltech-256

Fig. 2. Example images of the Office&Caltech

Images in the source and target domains are represented based on multi-sparse coding in our approach. In this process, the dictionary needs be over-completed for sparsity of coding coefficients. Because of scarcity of dSLR and high similarity with Webcam, the dSLR domain is not considered. So, the Caltech, Amazon and Webcam domains are employed in this experiment, namely six different pairs of combination, where one is regarded as the source and the other as the target. The proposed method is compared with BoVW&1-NN[18]、NBNN[29]、PCA_T[18]、DA-NBNN[30]、SGF[17] and GFK[18]. The former two methods are traditional image classification methods, where the data of source domain and target domains is regarded as a whole without considering the distribution differences, while the others are state-of-the-art domain adaptation methods.

In order to describe domain adaptation of different combinations conveniently, the following intuitive representation such as $A \rightarrow C$ is used, where A and C are the first letters of Amazon and Caltech, and the training samples from Amazon are employed as the source while the testing samples from Caltech are employed as the target. Randomly select 20 labeled images per category from the source domain and 16 images per category from the target domain. Following the common benchmarking procedures, we repeat the experimental process for 5 times by randomly choosing training images and testing images. The average per-class classification accuracy of each time is recorded and the mean accuracy and its standard deviation are taken as the indices for comparison. Especially, in the two pairs experiment of $C \rightarrow A$ and $C \rightarrow W$, length of the initial dictionary trained from the source domain is 800 while in other pairs is 600.

The performances of the proposed method as well as several other methods are shown in **Table 1**, where the black bold font parts of each column denote the highest classification accuracy in each domain couple. For the six pairs of source and target domains, it can be clearly seen that the methods based on domain adaptation are superior to the traditional methods in each group and the proposed method achieves the best performance except $C \to A$、 $A \to W$ and $W \to A$. The method of BoVW&1-NN achieves the worst performance in each domain couple, where the Euclidean distance is utilized for the 1-NN classifier. It is not able to reflect the true image-to-image distance between high dimension BoVW representations using Euclidean distance. NBNN is obviously superior to BoVW&1-NN, where the image-to-class distance avoiding quantization errors is used for classification decision. In particular, NBNN achieves better performance than GFC in domain couples of $C \to A$ and $W \to A$. It also confirms that in the image-to-class setting the domain shift is intrinsically smaller, which provides good generalization for NBNN. Based on NBNN, the domain adaptation NBNN (DA-NBNN) is proposed, which iteratively learns a class metric for a large margin separation among classes. In the method of PCA_T, all the features are directly projected into the PCA subspace learned from the target domain, but this practice is too blunt, which greatly degrades the performance. The Sampling Geodesic Flow (SGF) approach employs intermediate subspaces to model domain shifts while the GFK extends SGF by leveraging the Geodesic Flow Kernel. The main idea of our method is similar to SGF and GFK, both of which assume the existence of intermediate subspaces between the source and target domains. From **Table 1**, we can see that GFK indeed outperforms SGF. Because of the smaller reconstruction error of the target data, our method provides better results than SGF and GFK (apart from $A \to W$). Especially in the experiment of $A \to C$, compared with GFK, the classification accuracy is improved by about 6.1%.

Table 1. Classification accuracy of cross domain

Methods	$C \to A$	$C \to W$	$A \to C$	$A \to W$	$W \to C$	$W \to A$
BoVW&1-NN	20.9±3.0	18.4±3.8	20.3±2.2	21.0±3.6	16.8±1.1	16.1±1.5
NBNN	41.0±3.0	28.4±3.7	31.3±1.3	31.8±2.2	26.8±1.0	37.4±1.2
PCA_T	32.8±0.7	27.7±0.4	30.7±0.5	29.8±0.7	23.5±0.4	28.1±0.5
DA-NBNN	**56.4±3.4**	33.4±2.6	40.1±3.7	36.6±2.9	32.3±3.1	**41.6±4.1**
SGF	36.8±0.5	31.7±0.4	35.3±0.5	31.0±0.7	21.7±0.4	35.5±0.7
GFK	40.4±0.7	32.5±6.8	34.2±1.5	**37.0±5.1**	27.1±1.2	31.5±0.5
Our method	44.7±0.3	**35.1±0.8**	**40.3±0.5**	36.5±0. 9	**33.1±0.4**	38.2±0.5

Given a single target domain and multiple source domains, identifying which is the best one for training classifier for the target data directly determines the final recognition. Obviously, the source domain closely related to the target is the best candidate. Gong et al. [18] proposes the concept of Rank of Domain (ROD), which is a measurement to a certain extent for domain shifts between different domains. It deems that the source domain with smaller ROD value can be well adapted to the target domain, which provides guidance for the selection of multiple source domains.

Similarly, in order to measure the similarity between the source and target domains, Ni et al. [20] puts forward the concept of Quantificaiton of Domain Shift (QDS). The QDS between the source and target domains is defined as $QDS = \frac{1}{2}(Q_{S,T} + Q_{T,S})$, where $Q_{S,T} = \left\| D_K{}^T D_0 \right\|_F$ and $Q_{T,S} = \left\| D_0{}^T D_K \right\|_F$ .The $Q_{S,T}$ characterizes the amount of shifts from the source to the target domain by measuring similarity between the source domain dictionary D_0 and the target domain dictionary D_K . By reversing the role of source and target domain, $Q_{T,S}$ characterizes the amount of shifts from the target to the source domain. Similarity to [20], the intermediate subspaces are also modeled by dictionary learning in our method. Since updating the initial dictionary trained from the source domain with continuous incensement of the target data, the final dictionary achieved by online dictionary learning can represent the target data in terms of reconstruction error. Hence, the QDS is also employed for measuring domain shifts and verifying the effectiveness of the above experiments.

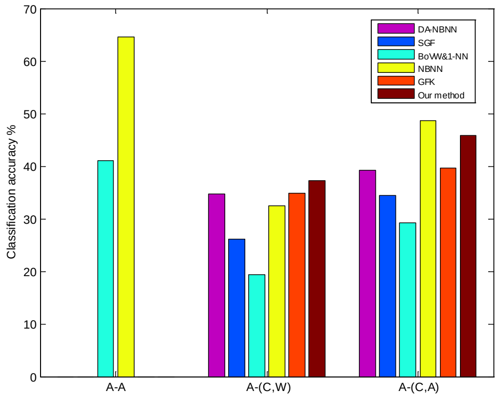
The value of QDS reflects the similarity of two domains, where a higher value indicates higher consistency. Normally, the QDS value is positive correlation with recognition rate of the target data, providing an option for the optimal source domain with least domain shift. **Table 2** shows the QDS value of different pairs of Caltech/Amazon/Webcam, where the symbol of NA represents the null value. Each column and row listed in **Table 2** indicates the target and source domains respectively. For example, in the first column, when Caltech is treated as the target domain, the QDS value of Amazon is larger than that of Webcam. Thus compared with Webcam, Amazon can be well adapted to the target domain for training classifier, which is also consistent with the experimental results in **Table 1**. The classification accuracy of $A \rightarrow C$ and $W \rightarrow C$ in our method are 40.3% and 33.1% respectively. In other words, when Caltech is treated as the target data, Amazon is the best choice for the source data. While choosing Webcam as the target domain, it can be seen that Amazon has a larger QDS value than Caltech, which indicates high similarity and adaptability between Webcam and Amazon. The experimental results in **Table 1** also confirm the conclusion, when Webcam is severed as the target domain, the accuracy in $A \rightarrow W$ is 1.4% higher than that in $C \rightarrow W$.

Table 2. QDS Value of cross domains

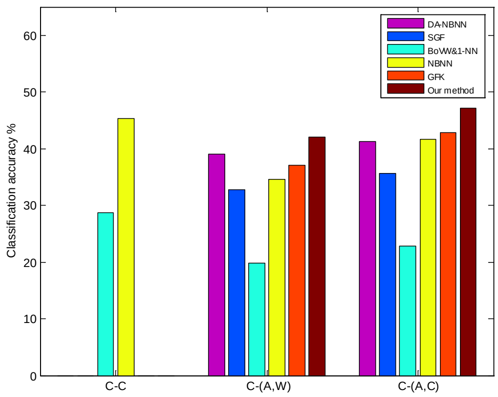
Domains	Caltech	Amazon	Webcam
Caltech	NA	12.65	10.12
Amazon	12.65	NA	10.93
Webcam	10.12	10.93	NA

Gong et al. [18] study the adaptation of multiple source domains to the single target domain, namely choosing the most suitable source domain for the target domain. However, in practical applications, the samples to be recognized with unpredictability and complexity are more common. Therefore, the adaptation from the single source domain to multiple target domains is also considered in the proposed method. For conveniently comparing and analyzing experimental results, part of the target data comes from the source domain. The following representation of $A \rightarrow (C, W)$ is used in the section, where A (Amazon) denotes the source domain, the combination of C (Caltech) and W (Webcam) denoting the target domain. Following the above setup, we carry on the experiment. In particular, when the target data is derived from two domains, eight samples are randomly selected from each domain.

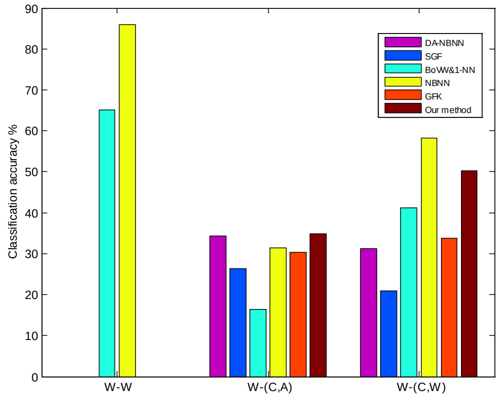
From Amazon, Caltech and Webcam, one of them is chosen as the source domain and the target data is obtained from the source domain or the other two domains. The experimental results are shown in **Fig. 3**, where Amazon, Caltech and Webcam are considered as the source domain respectively. When the source data and the target data are derived from the same dataset, such as $A \to A$、$C \to C$ and $W \to W$, the issue is transformed to the traditional image classification or in-domain problem , which does not contain any operation of domain adaptation. In these instances, good results are achieved in NBNN, which is consistent with the theoretical analysis. When the target data comes from multiple domains, our method possesses some advantages, specifically in the domain pairs of $A \to (C,W)$, $C \to (A,W)$ and $W \to (C,A)$.It also can be seen that our method improves over SGF and GFK with a significant gain in accuracy. Although the target data is derived from multiple domains, the distribution difference is mitigated in a shared feature space in our method. Therefore, the augmented feature is more robust and effective for across domain recognition. While partial data of the target domain originates from the source domain such as $C \to (A,C)$, compared with other domain adaptation methods, our method is more adaptive and achieves the highest accuracy.



（a）Amazon severing as the source domain



（b）Caltech severing as the source domain



（c）Webcam severing as the source domain

Fig. 3. Performance comparison of multiple source domains in Office&Caltech

4.2 Results on Caltech-256 & Bing Caltech-256 Dataset

In order to better study the image classification of domain adaptation, according to category labels of Caltech-256, Bergamo et al. [28] build Bing Caltech-256 dataset including 120924 images by leveraging the Bing image search engine on the Internet. Due to ambiguity of the text keywords and diversity of the Internet images, the dataset is more challenging. **Fig. 4** shows example images from Bing Caltech-256 and Caltech-256 domains. Based on suggestions of Bergamo, the Caltech-256 dataset is regarded as target domain while the source data comes from the Bing Caltech-256 dataset. In the experiment, all images of the source and target domains are preprocessed into gray scale and the max side of each image is resized to 300 pixels. The dense SIFT descriptors normalized with ℓ^2-norm are sampled from each image with a step length of 8 pixels in 16×16 pixel patches. Following the first step of the proposed method, decomposed by the robust dictionary learned from Bing Caltech-256, images of the source and target domains are represented as BoVW feature representations with 1500 dimensions. On the basis of this representation, the K-SVD algorithm is employed for the initial dictionary D_0 with the size of 4000 in the source domain.

In the proposed method, intermediate subspace dictionaries are gained with the target data updating by means of online dictionary learning. In the process, samples of the target domain are considered as a dynamic data sequence. With continuous updating of the target data, the each intermediate domain dictionary needs to be saved. In image classification and recognition task of small scale, updating the previous dictionary with a single sample of the target domain is acceptable. Nevertheless, in the large-scale dataset, increased computation and memory consumption are resulted in with data size growing.

(a) Bing Caltech-256 (b) Caltech-256

Fig. 4. Example images of the Caltech-256 & Bing Caltech-256

One of the effective ways to resolve the problem is to update the previous dictionary with mini-batch samples instead of a single sample, as shown in online dictionary learning [27] and online robust dictionary learning [31]. In order to reduce computation and memory consumption, the same trick is used in the experiment. The influence of size of mini-batch samples for dictionary updating on classification accuracy is also studied, which is shown in **Fig. 5**. From the experimental results, it can be seen that the impact of mini-batch size (p) on the accuracy is relatively small. When p is increased from 10 to 100, the classification accuracy is reduced by 1.29%. Balancing the running speed and accuracy, the value of p is set to 30 in follow-up experiment.

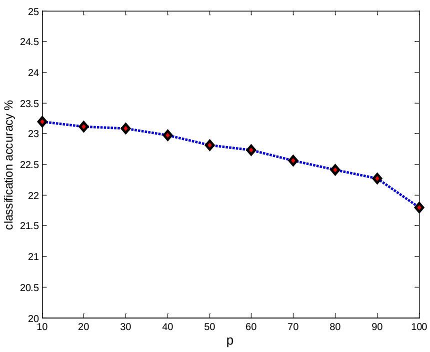


Fig. 5. Influence of mini-batch size for dictionary updating on classification accuracy

Fig. 6 shows experimental results of the proposed method and some other classification methods. It is obviously that the domain adaptation methods outperform traditional methods such as BoVW&1-NN and NBNN, where the source and target domains are considered undifferentiated. The experiment results further strengthen effectiveness of domain adaptation for image classification, while the training data and testing data have different distribution. Our method outperforms DA-NBNN, but be inferior to RDALR. In RDALR, the low-rank and sparse decomposition is utilized for connecting the source and target domains. However, RDALR is a semi-supervised domain adaptation method, which requires labeled samples in the target domain, and the classification accuracy of RDALR is significantly improved with increasement of number of labeled target data. In terms of classification accuracy, our method is less than RDALR. However, in semi-supervised domain adaptation method, it is an exceedingly laborious work to collect labels for target data in practical applications. As an unsupervised domain adaptation, our method is more applicable.

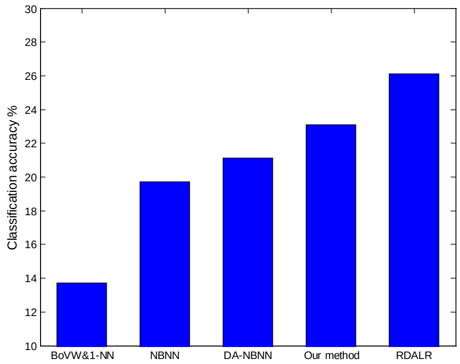


Fig. 6. Classification accuracy of different methods

5. Conclusion

In the process of image classification, if the classifier learned from training data is directly applied to the testing data without considering the difference of distribution, the accuracy of image classification will be greatly degraded. Aiming at this problem, a domain adaptation image classification method based on multi-sparse representation is proposed in the paper. In order to effectively connect the source domain and the target domain, the existence of intermediate domains is hypothesized between the two domains. Moreover, the intermediate domain subspaces are modeled by online dictionary learning method with continuous updating of the target data. On the one hand, it ensures that the reconstruction error of the target data is the smallest, and on the other, the connection between the source and target domains is as smooth as possible. A shared feature space is built through the source, intermediate and target domains, where the data distribution shift is mitigated. The augmented feature representation with more robustness is employed for across domain recognition. Experimental results show that, when the training data and testing data are derived from different domains, the proposed method achieves good performance.

References

[1] L Nie, M Wang, Z J Zha and T S Chua, "Oracle in Image Search: A Content-Based Approach to Performance Prediction," *Acm Transactions on Information Systems*, vol. 30, iss. 2, pp.1-23, May, 2012. Article (CrossRef Link).

[2] R C Hong, M Wang, Y Gao, D C Tao, X L Li and X D Wu, "Image Annotation by Multiple-Instance Learning With Discriminative Feature Mapping and Selection," *IEEE Transactions on Cybernetics*, vol. 44, no. 5, pp.669-680, June,2013. Article (CrossRef Link).

[3] L Nie, M Wang, Z Zha, G Li and T S Chua, "Multimedia answering: enriching text QA with media information," in *Proc. of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pp. 695-704, July 24-28, 2011. Article (CrossRef Link).

[4] S Lazebnik, C Schmid and J Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. of 2006 IEEE Computer Society Conference on Computer Vision & Pattern Recognition*, Vol.2, pp.2169-2178, June 17-22, 2006. Article (CrossRef Link).

[5] J Yang, K Yu, Y Gong and T Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *Proc. of 2009 IEEE Computer Society Conference on Computer Vision & Pattern Recognition*, pp. 1794-1801,June 20-25, 2009.Article (CrossRef Link).

[6] J Wang, J Yang, K Yu and F Lv, "Locality-constrained linear coding for image classification," in *Proc. of 2010 IEEE Computer Society Conference on Computer Vision & Pattern Recognition*, pp. 3360-3367, June 13-18,2010. Article (CrossRef Link).

[7] R Chellappa, J Ni and V M Patel, "Remote identification of faces: Problems, prospects, and progress," *Pattern Recognition Letters*, vol. 33, iss. 14, pp. 1849-1859, October, 2012. Article (CrossRef Link).

[8] A Torralba and A A Efros, "Unbiased look at dataset bias," in *Proc. of 2011 IEEE Computer Society Conference on Computer Vision & Pattern Recognition*, pp. 1521-1528, June 20-25, 2011. Article (CrossRef Link).

[9] M Kan, J Wu, S Shan and X Chen, "Domain Adaptation for Face Recognition: Targetize Source Domain Bridged by Common Subspace," *International Journal of Computer Vision*, vol. 109, iss. 1, pp. 94-109, August, 2014.Article (CrossRef Link).

[10] S Bendavid, J Blitzer, K Crammer, A Kulesza and F Pereira, "A theory of learning from different domains," *Machine learning*, vol. 79, iss. 1, pp. 151-175, May, 2010. Article (CrossRef Link).

[11] S. J. Pan, Q Yang, "A Survey on Transfer Learning," *IEEE Transactions on Knowledge & Data Engineering*, vol. 22, iss. 10, pp. 1345-1359, October, 2010. Article (CrossRef Link).

[12] Z c Li, J Liu, J H Tang and H Q Lu, "Robust Structured Subspace Learning for Data Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.37, iss.10, pp.2085-2098, October, 2015. Article (CrossRef Link).

[13] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, iss. 4, pp. 1289-1306, April, 2006. Article (CrossRef Link).

[14] J. J. Heckman, "Sample selection bias as a specification error," *Applied Econometrics*, vol. 31, no. 1, pp. 153-61 2013. Article (CrossRef Link).

[15] B Schölkopf, J Platt and T Hofmann, "Correcting Sample Selection Bias by Unlabeled Data," in *Proc. of Conference on Advances in Neural Information Processing Systems*, pp. 601-608, December 3-6, 2007. Article (CrossRef Link).

[16] K Saenko, B Kulis, M Fritz and T Darrell, "Adapting visual category models to new domains," in *Proc. of 11th European conference on Computer vision*, pp.213-226, September 5-11, 2010.Article (CrossRef Link).

[17] R Gopalan, R Li and R Chellappa, "Domain adaptation for object recognition: An unsupervised approach," in *Proc. of 2011 IEEE International Conference on Computer Vision*, pp.999-1006. November 6-13, 2011.Article (CrossRef Link).

[18] F Sha, Y Shi, B Gong and K Grauman, "Geodesic flow kernel for unsupervised domain adaptation," in *Proc. of 2012 IEEE Computer Society Conference on Computer Vision & Pattern Recognition*, pp. 2066-2073, June 16-21, 2012.Article (CrossRef Link).

[19] IH Jhuo, D Liu, DT Lee and SF Chang, "Robust visual domain adaptation with low-rank reconstruction," in *Proc. of 2012 IEEE Computer Society Conference on Computer Vision & Pattern Recognition*, pp. 2168-2175, June 16-21, 2012. Article (CrossRef Link).

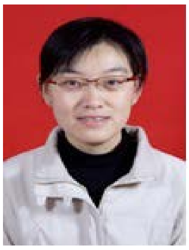
[20] J Ni, Q Qiu and R Chellappa, "Subspace interpolation via dictionary learning for unsupervised domain adaptation," in *Proc. of 2013 IEEE Computer Society Conference on Computer Vision & Pattern Recognition*, pp. 692-699 , June 25-27, 2013. Article (CrossRef Link).

[21] S J Pan, J T Kwok and Q Yang, "Transfer Learning via Dimensionality Reduction," in *Proc. of the Twenty-Third Aaai Conference on Artificial Intelligence*, pp. 677-682, July 13-17,2008. Article (CrossRef Link).

[22] S J Pan，IW Tsang，JT Kwok and Q Yang, “Domain adaptation via transfer component analysis,” *IEEE Transactions on Neural Networks*, vol.22, iss.2, pp.199-210, February, 2011. Article (CrossRef Link).
[23] J Yang, R Yan and A G Hauptmann, “Cross-domain video concept detection using adaptive svms,” in *Proc. of 15th ACM international conference on Multimedia*, pp.188-197, September 23-28, 2007. Article (CrossRef Link).
[24] L Bruzzone and M Marconcini, “Toward the automatic updating of land-cover maps by a domain-adaptation SVM classifier and a circular validation strategy,” *IEEE Transactions on Geoscience & Remote Sensing*, vol.47, iss.4, pp. 1108-1122, April, 2009. Article (CrossRef Link).
[25] A Wagner, J Wright，A Ganesh，Z Zhou，H Mobahi and Y Ma, “Toward a practical face recognition system: Robust alignment and illumination by sparse representation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.34, iss.2, pp.372-386, February, 2012. Article (CrossRef Link).
[26] C Zhao, X Wang and W K Cham, “Background subtraction via robust dictionary learning,” *EURASIP Journal on Image and Video Processing*, vol.2011, no.10, pp.2919-2929, February, 2011. Article (CrossRef Link).
[27] J Mairal, F Bach, J Ponce and G Sapiro, “Online learning for matrix factorization and sparse coding,” *Journal of Machine Learning Research*, vol.11, no.1, pp.19-60, March, 2010. Article (CrossRef Link).
[28] A Bergamo and L Torresani, “Exploiting weakly-labeled web images to improve object classification: a domain adaptation approach,” in *Proc. of Advances in Neural Information Processing Systems 23*, pp. 181-189, December 6-9, 2010.Article (CrossRef Link).
[29] O Boiman, E Shechtman and M Irani, “In defense of nearest-neighbor based image classification,” in *Proc. of 2013 IEEE Computer Society Conference on Computer Vision & Pattern Recognition*, pp. 1-8, June 23-28, 2008.Article (CrossRef Link).
[30] T Tommasi and B Caputo, "Frustratingly easy nbnn domain adaptation,” in *Proc. of 2013 IEEE International Conference on Computer Vision*, pp. 897-904, December 1-8, 2013. Article (CrossRef Link).
[31] C Lu, J Shi and J Jia, “Online robust dictionary learning,” in *Proc. of 2013 IEEE Computer Society Conference on Computer Vision & Pattern Recognition*, pp. 415-422, June 23-28, 2013. Article (CrossRef Link).

XuZhang received the Ph.D. degree from the University of Science and Technology of China, Hefei, in 2016. He is currently a lecturer in College of Electronic Information Engineering of Suzhou Vocational University.His research interests include computer vision, machine learning, and pattern recognition.

XiaoFeng Wang received the Ph.D degree in Pattern Recognition and Intelligent System from University of Science and Technology of China, Hefei, China, in 2009. He is now the head of Key Lab of Network and Intelligent Information Processing of Hefei University. His research interests include pattern recognition, image processing and data mining.



YueDu received the master's degree in Computer Applications Technology from the northwest normal university in 2008. She is a lecturer in Army Officer Academy of PLA. Her research interests include data mining, machine learning, and pattern recognition.

Xiaoyan Qin is currently a lecturer in Army Officer Academy of PLA. Her research interests include computer vision, machine learning, and pattern recognition.