An Unsupervised Domain Adaptation Approach For Cross-Domain Visual Classification

Cheng-An Hou[1], Yi-Ren Yeh[2], and Yu-Chiang Frank Wang[1]

[1]Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan

[2]Department of Mathematics, National Kaohsiung Normal University, Kaohsiung, Taiwan

Abstract

For cross-view action recognition and many real-world visual classification problems, one needs to recognize test data at a particular target domain of interest, while training data are collected at a different source domain. Without eliminating such domain differences, recognition of test data using classifiers trained in the source domain will not be expected to produce satisfactory performance. In this paper, we propose a novel domain adaptation approach, which is able to learn a common feature space relating cross-domain data. In particular, we not only aim at matching cross-domain data marginal distributions during adaptation, we also exploit the structure of target domain data and update class-conditional distributions accordingly. Experiments on various cross-domain visual classification tasks would verify the effectiveness and robustness of our proposed method.

1. Introduction

Conventional pattern recognition algorithms assume that training and test data for addressing a particular classification task exhibit the same (or similar) data distribution. As a result, one typically learn classifiers using training data for recognizing test data accordingly. However, one often needs to perform recognition of data across different datasets and domains (e.g., images with different resolutions). In such cases, it is typically not possible to collect a sufficient amount of labeled training data in the domain/dataset of interest (otherwise the recognition would not be performed across domains). It is not surprising that, as depicted in Figure 1, if we directly apply classifiers learned from training images of one dataset for recognizing test images of a different dataset, the performance will be poor due to domain or dataset mismatches [17, 4].

In order to address the above problem, the idea of *domain adaptation* in transfer learning has been proposed to associate data across different domains (i.e., from *source* to *target* domains), so that the tasks of recognition can be performed accordingly [14]. In such domain adaptation problems, one typically observes labeled training data in the source domain, while few or no labeled training data can be collected at the target domain. In this paper, we particularly consider the problem of *unsupervised domain adaptation*, which is known to be very challenging due to no labeled training data but only test data can be observed in the target domain [11, 3].

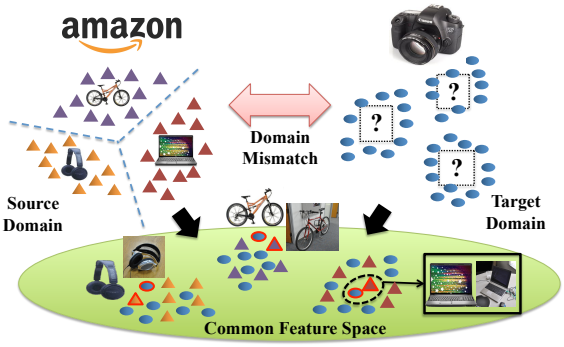

Figure 1. Illustration of eliminating domain mismatches for cross-domain visual classification.

Several domain adaptation methods have been proposed for cross-domain classification [15, 7, 3, 12]. Most approaches focus on determining a common feature space for relating cross-domain data [15, 13, 11]. Once this feature space is determined, one can simply perform classification using projected data accordingly. However, for unsupervised domain adaptation, it is very difficult to associate cross-domain data due to the lack of labeled training data in the target domain. In our work, we need to eliminate domain mismatch by observing/predicting class-conditional distributions of the test data in the target domain. We propose a novel learning algorithm for addressing the above problem (see Section 3 for detailed discussions). The contributions of this paper can be summarized summarized as follows:

- A novel instance selection algorithm for cross-domain

classification is proposed, which is able to exploit the structure of unlabeled data in the target domain for improved domain adaptation and cross-domain visual classification.

- We conduct experiments on multiple real-world visual classification datasets. By comparing our results with those reported by state-of-the-art unsupervised domain adaptation approaches, the effectiveness of our proposed method can be successfully verified.

2. Related Works

Recently, the research topic of domain adaptation has attracted the attention from researchers in the areas of computer vision [11, 6, 3, 15, 7, 5] and machine learning [10, 18, 1]. Domain adaptation methods can be generally divided into two different categories based on the availability of target labeled data. For *semi-supervised domain adaptation* [15, 1], one can observe source-domain labeled data in advance, but only a small amount of target labeled data can be acquired. Given such cross-domain data and label information, the task is to recognize the remaining target-domain data. On the other hand, *unsupervised domain adaptation* [6, 11] deals with a more challenging scenario, in which only labeled data are available in the source domain, while the target-domain data are totally unsupervised. This paper focuses on unsupervised domain adaptation.

For unsupervised domain adaptation approaches, we particularly focus on the MMD-based feature adaptation method [13, 11, 3, 5, 10], which aim at matching cross-domain information (e.g., marginal and/or class-conditional distributions) for adaptation and recognition purposes. However, since only unlabeled data can be observed in the target domain, modeling of the associated class-conditional distributions is a nontrivial task. Previously, researchers chose to model and match cross-domain marginal distributions only. For example, the method of kernel mean matching proposed by Huang et al. [10] aims to minimize the mean difference between cross-domain data in a predetermined kernel space by weighting source-domain data. To provide additional robustness, Pan et al. [13] proposed Transfer Component Analysis (TCA) to determine low-dimensional embeddings for cross-domain data, so that the matching of cross-domain marginal distributions can be performed in the derived lower-dimensional feature space. Based on TCA, Long et al. [12] further proposed Transfer Feature Matching (TJM) integrating the idea of instance reweighting and distribution adaptation techniques for further improving the adaptation ability.

However, only adapting marginal distributions would not be sufficient for relating cross-domain data, especially if recognizing target-domain data of different classes is desirable. To address this concern, Long et al. [11] proposed Joint Distribution Adaptation (JDA) to match both marginal and conditional feature distributions. To address the unsupervised domain adaptation settings, they applied the prediction of source-domain classifiers as the pseudo labels of target-domain data, and thus the conditional distributions can be observed accordingly. However, the direct use of such pseudo labels might not be preferable due to the error induced by the possible domain mismatch. In this paper, we propose a novel target instance selection algorithm for improving conditional distribution estimation in the target domain. By the proposed selection algorithm, we are able to exploit the local structure of the target domain and estimate the conditional distribution in confidence. As a result, the improved recognition performance can be expected in the target domain.

3. Our Proposed Method

3.1. Motivation

We first start from the problem definition and introduce the notations used in this paper. Let $\mathcal{D}_S = \{\mathbf{x}_{S_m}, y_{S_m}\}_{m=1}^{M} = \{X_S, Y_S\}$ as M d-dimensional labeled instances in the source domain, and $\mathcal{D}_T = \{\mathbf{x}_{T_n}\}_{n=1}^{N} = \{X_T\}$ as N unlabeled ones (with the same feature dimension) in the target domain. With the assumption that both source and target domains contain data of the same C classes of interest, our goal is to determine the label of each instance in the target domain (i.e., Y_T).

In our work, we aim at determining a common feature space in which the joint distributions of projected source and target domain data will be similar to each other, allowing cross-domain classification. To be more specific, let $P_S(X_S, Y_S) = P_S(X_S)P_S(Y_S|X_S)$ and $P_T(X_T, Y_T) = P_T(X_T)P_T(Y_T|X_T)$ as joint distributions of source and target domain data in the original space, respectively. Note that $P(X)$ indicates the marginal distributions, and $P(Y|X)$ indicates the posterior distributions. To match the joint distributions, we aim at determining a transformation to adapt both marginal and posterior distributions. However, since the modeling of the posterior distribution $P(Y|X)$ is not applicable during training, we choose to adapt the class-conditional distribution $P(X|Y)$ instead, based on the sufficient statistics of $P(Y|X)$. As a result, our goal is to find a transformation Φ, so that $P_S(\Phi(X_S)) \approx P_T(\Phi(X_T))$ and $P_S(\Phi(X_S)|Y_S) \approx P_T(\Phi(X_T)|Y_T)$ can be satisfied. We now detail our proposed algorithm in the following subsections.

3.2. Adaptation of Cross-Domain Marginal Distributions

To adapt cross-domain marginal distributions for eliminating domain differences, we utilize the rule of MMD [8]

as suggested in [13]. Let $\mathbf{X} = \{X_S, X_T\} \in \mathbb{R}^{d\times(M+N)}$, our goal is to learn an orthonormal projection matrix $\mathbf{W} \in \mathbb{R}^{d\times k}$ which minimizes the difference between cross-domain marginal distributions. In other words, the objective function can be described as follows:

$$\min_{\mathbf{W}^\top\mathbf{W}=\mathbf{I}} \| \frac{1}{M} \sum_{x_i \in D_S} \mathbf{W}^\top \mathbf{x}_i - \frac{1}{N} \sum_{x_j \in D_T} \mathbf{W}^\top \mathbf{x}_j \|^2, \quad (1)$$

which can be simplified and rewritten as:

$$\min_{\mathbf{W}^\top\mathbf{W}=\mathbf{I}} tr(\mathbf{W}^\top \mathbf{X}\mathbf{L}\mathbf{X}^\top \mathbf{W}), \quad (2)$$

where

$$L_{ij} = \begin{cases} \frac{1}{M^2}, & \text{for} \quad \mathbf{x}_i, \mathbf{x}_j \in \mathcal{D}_S \\ \frac{1}{N^2}, & \text{for} \quad \mathbf{x}_i, \mathbf{x}_j \in \mathcal{D}_T \\ \frac{-1}{MN}, & \text{otherwise.} \end{cases}$$

By solving (2), MMD solves the domain mismatch problems by minimizing the distance between the global means of the projected cross-domain data. As noted and verified in [8, 11], this would associate the projected marginal distributions of data from each domain. However, since our goal is to perform cross-domain classification, suppressing the domain mismatch by simply adapting cross-domain marginal data distributions will *not* be sufficient. This is because that, without taking label (i.e., class) information into consideration, the adapted information and features will not be able to exhibit discriminating capabilities. In order to address this issue, we further propose to adapt class-conditional distributions across domains, which will be described below.

3.3. Adaptation of Cross-Domain Class Conditional Distributions

Recall that, we have labeled training data in the source domain, but only unlabeled data are available in the target domain for training. Therefore, one cannot perform adaptation of class-conditional distribution $P(X|Y)$ across domains as we did in Section 3.2.

In [11], it has been proposed to utilize target-domain pseudo labels predicted by source-domain classifiers for describing the class-conditional data distributions in the target domain. However, the direct use of such pseudo labels would not be preferable, since mismatch between source and target domains will be presented, and thus the use of source-domain classifiers for predicting target-domain data might not generalize well. To overcome the above concern, we propose a novel target instance selection algorithm, which takes the structure of target-domain data into consideration for better predicting their pseudo labels during adaptation.

In our proposed algorithm, we first perform k-means clustering with $k = C$ on unlabeled target-domain data. At the same time, the pseudo label of each target-domain instance will be determined by source-domain classifiers as did in [11] (in our work, we consider nearest neighbor classifiers for simplicity). As a result, we have target-domain data initialized as $\hat{\mathcal{D}}_T = \{\mathbf{x}^u_{T_n}, \hat{y}_{T_n}\}_{n=1}^N$, in which u indicates the cluster of the instance belongs to ($1 \leqslant u \leqslant C$), and $\hat{y}$ denotes the pseudo label of the instance ($1 \leqslant \hat{y} \leqslant C$).

With $\hat{\mathcal{D}}_T$, we perform the following step for instance selection in the target domain:

$$\alpha(\mathbf{x}^u_{T_n}) = \begin{cases} 1, & \text{for} \quad \hat{y}_{T_n} = mode(u) \\ 0, & \text{otherwise,} \end{cases} \quad (3)$$

where $mode(u)$ indicates the majority of the pseudo labels observed in cluster u. Based on (3), we see that only the instances whose pseudo labels are the same as the mode of the associated cluster will be selected. This process allows us to observe data structure in the target domain, while the label of each cluster can still be properly predicted using source-domain classifiers. Thus, to adapt cross-domain class-conditional distributions for each class $c \in \{1, 2, ..., C\}$, we propose to solve the following problem:

$$\min_{\mathbf{W}^\top\mathbf{W}=\mathbf{I}} \| \frac{1}{M^{(c)}} \sum_{x_i \in \mathcal{D}_S^{(c)}} \mathbf{W}^\top \mathbf{x}_i - \frac{1}{\sum \alpha^{(c)}} \sum_{\mathbf{x}_j \in \hat{\mathcal{D}}_T^{(c)}} \alpha(\mathbf{x}_j) \mathbf{W}^\top \mathbf{x}_j \|^2, \quad (4)$$

where $\mathcal{D}_S^{(c)} = \{\mathbf{x}_{S_i} : y_{S_i} = c\}$ and $M^{(c)} = |\mathcal{D}_S^{(c)}|$. We note that $\hat{\mathcal{D}}_T^{(c)} = \{\mathbf{x}_{T_j} : \hat{y}_{T_j} = c\}$, and $\sum \alpha^{(c)}$ calculates the number of selected instances in pseudo class c. Similarly, (4) can be further simplified as below:

$$\min_{\mathbf{W}^\top\mathbf{W}=\mathbf{I}} tr(\mathbf{W}^\top \mathbf{X}\mathbf{L_c}\mathbf{X}^\top \mathbf{W}), \quad (5)$$

with

$$(L_c)_{ij} = \begin{cases} \frac{1}{M^{(c)2}}, & \mathbf{x}_i, \mathbf{x}_j \in \mathcal{D}_S^{(c)} \\ \frac{1}{(\sum \alpha^{(c)})^2}, & \mathbf{x}_i, \mathbf{x}_j \in \hat{\mathcal{D}}_T^{(c)}, \alpha(\mathbf{x}_i) = \alpha(\mathbf{x}_i) = 1 \\ \frac{-1}{M^{(c)}(\sum \alpha^{(c)})}, & \begin{cases} \mathbf{x}_i \in \mathcal{D}_S^{(c)}, \mathbf{x}_j \in \hat{\mathcal{D}}_T^{(c)}, \alpha(\mathbf{x}_j) = 1 \\ \mathbf{x}_i \in \hat{\mathcal{D}}_T^{(c)}, \mathbf{x}_j \in \mathcal{D}_S^{(c)}, \alpha(\mathbf{x}_i) = 1 \end{cases} \\ 0, & \text{otherwise.} \end{cases}$$

It can be seen that, by optimizing (2) and (5), we will be able to transfer labels from the source to the target domain, which essentially adapts both marginal and class-conditional data distributions for cross-domain classification.

3.4. Optimization

To jointly solve (2) and (5) for cross-domain classification, we formulate a unified optimization problem as follow:

$$\begin{aligned} &\min tr(\mathbf{W}^\top \mathbf{X}\mathbf{L}\mathbf{X}^\top \mathbf{W}) + \sum_{c=1}^{C} tr(\mathbf{W}^\top \mathbf{X}\mathbf{L_c}\mathbf{X}^\top \mathbf{W}) + \lambda \|\mathbf{W}\|_F^2 \\ &\text{s.t. } \mathbf{W}^\top \mathbf{X}\mathbf{C}\mathbf{X}^\top \mathbf{W} = \mathbf{I}. \end{aligned} \quad (6)$$

Table 1. Comparisons of recognition rates on the IXMAS dataset. Note that each row corresponds to a source camera view of interest, and each column indicates a target camera view (and the method to be evaluated).

	cam0				cam1				cam2				cam3				cam4			
	NN	[13]	[11]	Ours	NN	[13]	[11]	Ours	NN	[13]	[11]	Ours	NN	[13]	[11]	Ours	NN	[13]	[11]	Ours
cam0	–	–	–	–	33.33	42.78	44.44	**42.78**	27.22	54.44	58.89	**61.11**	32.78	50.56	55.56	**57.78**	30.00	42.78	45.00	**45.78**
cam1	31.67	39.22	27.22	**42.22**	–	–	–	–	26.67	32.00	27.22	**33.33**	26.67	37.22	28.89	**44.44**	24.44	38.33	35.00	**48.89**
cam2	27.78	47.80	57.78	**58.90**	21.67	28.33	19.44	**29.44**	–	–	–	–	22.22	25.56	22.78	**37.78**	27.78	37.78	37.78	**39.44**
cam3	28.89	48.33	51.11	**61.11**	25.11	26.54	21.11	**36.11**	16.11	30.56	32.56	**33.33**	–	–	–	–	25.56	43.89	31.67	**44.44**
cam4	27.78	48.89	51.22	**57.31**	23.33	40.56	51.11	**52.22**	24.44	37.79	32.22	**40.56**	24.44	50.00	57.78	**57.78**	–	–	–	–

It can be seen that, the first two terms in (6) address the adaptation of marginal and conditional distributions, respectively. The third term in (6) regularizes the learned projection matrix $\mathbf{W}$, while it is weighted by the parameter λ. The centering matrix $\mathbf{C}$ in the constraint of (6) is defined as $\mathbf{C} = \mathbf{I} - \frac{1}{MN}\mathbf{1}$ with $\mathbf{1}$ indicating the matrix of ones. As noted in [13, 11], adding such a constraint would allow us to preserve the data variance after adaptation, which implies and introduces additional data discriminating ability into the proposed model $\mathbf{W}$.

We note that (6) can be rewritten by Lagrange techniques:

$$tr(\mathbf{W}^\top(\mathbf{X}\mathbf{L}\mathbf{X}^\top + \mathbf{X}\sum_{c=1}^{C}\mathbf{L}_c\mathbf{X}^\top + \lambda\mathbf{I})\mathbf{W}) + tr((\mathbf{I} - \mathbf{W}^\top\mathbf{X}\mathbf{C}\mathbf{X}^\top\mathbf{W})\mathbf{\Psi}), \quad (7)$$

where $\mathbf{\Psi}$ is a diagonal matrix with Lagrange Multipliers (i.e., $\mathbf{\Psi} = diag(\psi_1, ..., \psi_k) \in \mathbb{R}^{k\times k}$). Let the derivative of (7) with respective to $\mathbf{W}$ equal to zero, the generalized eigen-decomposition problem can be derived as follows:

$$(\mathbf{X}\mathbf{L}\mathbf{X}^\top + \mathbf{X}\sum_{c=1}^{C}\mathbf{L}_c\mathbf{X}^\top + \lambda\mathbf{I})\mathbf{W} = \mathbf{X}\mathbf{C}\mathbf{X}^\top\mathbf{W}\mathbf{\Psi}. \quad (8)$$

As a result, the optimal solution of the projection matrix $\mathbf{W}$ can be determined by finding the $k-$smallest eigenvectors for solving (8). Since the initial pseudo labels predicted by source-domain classifiers without any adaptation might not be accurate, we practically iterate the above process until the solution converges (we observe that the iteration number smaller than 5 is typically sufficient).

3.5. Performing Recognition

Once the projection $\mathbf{W}$ is derived, one can project labeled training source-domain data into the observed common feature space, and the projected (unlabeled) test data can be recognized using conventional classification techniques accordingly. In our work, we apply the nearest neighbor (NN) classifier due to its simplicity [6].

4. Experiments

We consider one video dataset and five different image datasets for evaluating our performance on cross-domain visual classification: (1) *IXMAS Multi-View*, (2) *Office* [6], (3) *Caltech-256* [9], (4) *COIL-20*, (5) *MNIST*, and (6) *USPS* datasets. The IXMAS dataset is used for action recognition, in which each action is synchronically captured by five cameras. The first three image datasets are used for object recognition. We note that, the object images of Office and Caltech-256 are generally presented in real-world backgrounds, while those in COIL-20 exhibit significant aspect view variations (with plain background). MNIST and USPS are datasets used for handwritten digit recognition.

4.1. Cross-View Action Recognition

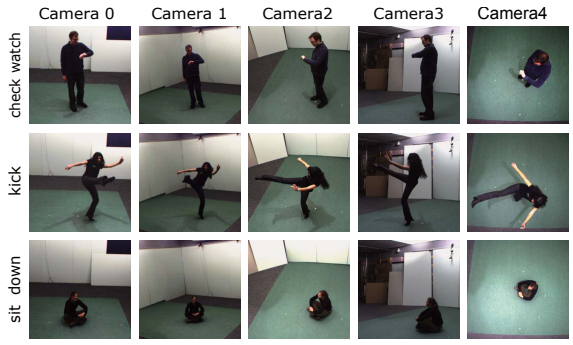

Figure 2. Example images of different actions in the IXMAS dataset. Each row represents an action at five different views.

The IXMAS Multi-View dataset contains videos of eleven action classes. Each action video is performed three times by twelve actors, and the actions are synchronically captured by five cameras (i.e., five views/domains), as shown in Figure 2. In our experiments, we use five actions, including check watch, cross arms, scratch head, sit down, and get up, for the cross-domain recognition. For feature representation, we extract descriptors considered by [2] and describe each action video as a group of spatio-temporal cuboids (at most 200). For each video, these cuboids are quantized into $N = 1000$ visual words. That is, each video is represented by a 1000-dimensional space for domain adaptation. For parameter selection, we set the $k = 100$ for our learned model $\mathbf{W}$ as the dimensionality of the derived common feature space, and the regularization parameter $\lambda = 1$.

For the IXMAS dataset, we first compare our method

Table 2. Comparisons of recognition rates on *Office* + *Caltech256*.

S → T	C → A	C → W	C → D	A → C	A → W	A → D	W → C	W → A	W → D	D → C	D → A	D → W	Average
NN	23.70	25.76	25.48	26.00	29.83	25.48	19.86	22.96	59.24	26.27	28.50	63.39	31.37
PCA	36.95	32.54	38.22	34.73	35.59	27.39	26.36	31.00	77.07	29.65	32.05	75.93	39.79
GFK [6]	41.02	40.68	38.85	40.25	38.89	36.31	30.72	29.75	80.89	30.28	32.05	75.59	42.94
TSL [16]	44.47	34.24	43.31	37.58	33.90	26.11	29.83	30.27	87.26	28.50	27.56	85.42	42.37
TCA [13]	38.20	38.64	41.40	37.76	37.63	33.12	29.30	30.06	87.26	31.70	32.15	86.10	43.61
JDA [11]	44.78	41.69	45.22	39.36	37.97	39.49	31.17	32.78	89.17	31.52	33.09	89.49	46.31
TJM [12]	46.76	38.98	44.59	39.45	42.03	**45.22**	30.19	29.96	89.17	31.43	32.78	85.42	46.33
Ours	**47.39**	**46.56**	**48.41**	**41.41**	**43.05**	42.04	**32.41**	**34.45**	**91.08**	**34.19**	**34.24**	**90.85**	**48.84**

with the nearest neighbor method (i.e., direct recognition without domain adaptation). In addition, we also compare our approach with recent unsupervised domain adaptation approaches of TCA [13] and JDA [11]. The results are shown in Table 1. It can be seen that, not only significant recognition improvements over the baseline method can be observed, our proposed method also outperformed state-of-the-art domain adaptation approaches for this particular task.

4.2. Cross-Domain Object Recognition

The *Office* dataset contains 31 objects categories, which are collected from three different domains: Amazon, DSLR and webcam. We note that, images of Amazon are collected from the Internet, those of DSLR are with high/sufficient resolution, while the webcam images are often over exposed or blurred with lower resolution. The *Caltech-256* dataset consists of object images of 256 categories (at least 80 images per category). Following the settings of [11, 3], we combine *Office* and *Caltech-256* datasets and select the 10 overlapping object categories from each domain: Caltech (C), Amazon (A), DSLR (D), and webcam (W). As a result, there are a total of 12 different cross-domain pairs for experiments (e.g., C→A, C→W, etc.). Detailed settings such as the number of training images per category can be found in [11, 3].

To represent each object image, we extract SURF features from images of Amazon and construct a codebook with 800 visual words (as did in [6, 11, 3]). Each image is thus converted into a 800-dimensional normalized Bag-of-Words (BoW) model for domain adaptation and recognition. Following [11], we set $k = 100$ and $\lambda = 1$.

We compare our approach with several state-of-the-art cross-domain classification methods: GFK [6], TSL [16], TCA [13], JDA [11] and TJM [12]. We also consider the direct uses of nearest neighbor (NN) and PCA as baseline approaches. Note that NN performs direct matching and thus does not consider domain adaptation, while PCA jointly projects source and target domain data into the PCA subspace for performing NN accordingly. For fair comparisons, NN is performed by all methods in each resulting feature space. The recognition results of each method on each source-target domain pair are listed in Table 2. From this table, it is clear that our proposed method outperformed all other approaches for cross-domain object recognition tasks.

4.3. Cross-View Object Recognition

The *COIL-20* dataset consists of 1440 images of 20 object categories. There are 72 images for each object class which were taken on a turntable for 5 degrees apart. Each image is of size 32×32 pixels. Since there is no background presented, the pixel values are directly applied as features for representing each image (as did in [11]). For our method, we set $k = 100$ for $\mathbf{W}$ and the regularization parameter $\lambda = 0.1$ (as [11] did).

To create cross-domain pairs, we follow the setting of [11]. To be more specific, the dataset is partitioned into *COIL A* and *COIL B*, in which *COIL A* contains images taken in the directions of the first and third quadrants (e.g., $[0°, 80°] \bigcup [180°, 265°]$), while those in *COIL B* are in the directions of the second and fourth quadrants (e.g., $[90°, 175°] \bigcup [270°, 355°]$). As a result, images of *COIL A* and *COIL B* would exhibit significantly different data/feature distributions due to camera view variations.

When performing cross-domain classification, we first consider the domain pair of *COIL A* vs. *COIL B*. More precisely, we use all 720 labeled images in *COIL A* as source domain training data, and all 720 images in *COIL B* as unlabeled target domain data to be recognized. Then, we switch source/target domain data to form other round of experiments. We use the same baseline and state-of-the-art methods as in Section 4.2 for comparisons, and we list the resulting recognition rates in Table 3. It can be seen that, based on the results shown in this table, our proposed method generalize well on the task of cross-view object classification.

4.4. Cross-Dataset Hand-Written Digit Recognition

Finally, we address cross-domain handwritten digit recognition using *MNIST* and *USPS* datasets. *MNIST* contains a training set of 60,000 images, and a test set of 10,000 images (images are of size 28×28 pixels). On the other hand, *USPS* has a total of 7291 training and 2007 test images, while each image is of size 16×16 pixels. The visual appearances of images in these two datasets are very different and thus exhibit distinct distributions. We randomly sample 2000 and 1800 samples from MNIST and USPS for experiments, respectively. Obviously, both datasets share the same ten digit classes.

Table 3. Recognition comparisons on *MNIST, USPS*, and *COIL-20*.

S → T	MNIST → USPS	USPS → MNIST	COIL A → B	COIL B → A
NN	65.94	44.70	83.61	82.78
PCA	66.22	44.95	84.72	84.03
GFK	67.22	46.45	72.50	74.17
TSL	66.06	53.75	88.06	87.92
TCA	56.28	51.05	88.47	85.83
JDA	67.28	59.65	89.31	88.47
TJM	63.28	52.25	89.02	88.89
Ours	**70.33**	**59.70**	**91.81**	**90.28**

For fair comparisons, we follow the setting of [11] performing cross-domain classification on these two datasets, and we also set $k = 100$ and $\lambda = 0.1$. We scale images from both dataset into 16×16 pixels, and use pixel values as the image features. With the same baseline and state-of-the-art approaches considered, Table 3 lists the recognition rates of each method on the two source-target domain pairs. From this table, we see that our approach also outperformed all others on the cross-domain handwritten digit recognition problems.

5. Conclusion

In this paper, we proposed a domain adaptation approach for cross-domain visual classification. In particular, we considered the scenario in which labeled source domain training data can be collected in advance, but only unlabeled data (to be recognized) are available in the target domain of interest. Since the direct use of classifiers learned in the source domain would not generalize well to the target domain, our method aims at eliminating such domain mismatches. By exploiting the data structure information in the target domain, our algorithm focuses on automatically adapting cross-domain marginal and class-conditional data distributions for improved classification. Our experiments on cross-domain object classification, cross-view object recognition, and cross-domain handwritten digit classification successfully verified the effectiveness and robustness of our approach.

Acknowledgement

This work is supported in part by the Ministry of Science and Technology of Taiwan via MOST 103-2221-E-001 -021 -MY2 and MOST 103-2221-E-034 -007.

References

[1] H. Daumé III, A. Kumar, and A. Saha. Co-regularization based semi-supervised domain adaptation. In *NIPS*, 2010.

[2] P. Dollár, V. Rabaud, G. Cottrell, and S. Belongie. Behavior recognition via sparse spatio-temporal features. In *ICCV joint workshop on VS-PETS*, 2005.

[3] B. Gong, K. Grauman, and F. Sha. Connecting the dots with landmarks: Discriminatively learning domain-invariant features for unsupervised domain adaptation. In *ICML*, 2013.

[4] B. Gong, K. Grauman, and F. Sha. Reshaping visual datasets for domain adaptation. In *NIPS*, 2013.

[5] B. Gong, K. Grauman, and F. Sha. Learning kernels for unsupervised domain adaptation with applications to visual object recognition. *IJCV*, 2014.

[6] B. Gong, Y. Shi, F. Sha, and K. Grauman. Geodesic flow kernel for unsupervised domain adaptation. In *IEEE CVPR*, 2012.

[7] R. Gopalan, R. Li, and R. Chellappa. Domain adaptation for object recognition: An unsupervised approach. In *IEEE ICCV*, 2011.

[8] A. Gretton, K. M. Borgwardt, M. Rasch, B. Schölkopf, and A. J. Smola. A kernel method for the two sample problem. In *NIPS*, 2007.

[9] G. Griffin, A. Holub, and P. Perona. Caltech-256 object category dataset. 2007.

[10] J. Huang, A. J. Smola, A. Gretton, K. M. Borgwardt, and B. Schölkopf. Correcting sample selection bias by unlabeled data. In *NIPS*, 2007.

[11] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu. Transfer feature learning with joint distribution adaptation. In *IEEE ICCV*, 2013.

[12] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu. Transfer joint matching for unsupervised domain adaptation. In *IEEE CVPR*, 2014.

[13] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang. Domain adaptation via transfer component analysis. *IEEE Trans Neural Networks*, 2011.

[14] S. J. Pan and Q. Yang. A survey on transfer learning. *IEEE TKDE*, 2010.

[15] K. Saenko, B. Kulis, M. Fritz, and T. Darrell. Adapting visual category models to new domains. In *ECCV*. 2010.

[16] S. Si, D. Tao, and B. Geng. Bregman divergence-based regularization for transfer subspace learning. *IEEE TKDE*, 2010.

[17] A. Torralba and A. A. Efros. Unbiased look at dataset bias. In *IEEE CVPR*, 2011.

[18] K. Zhang, K. Muandet, and Z. Wang. Domain adaptation under target and conditional shift. In *ICML*, 2013.