# Learn Multiple-Kernel SVMs for Domain Adaptation in Hyperspectral Data

Zhuo Sun, Cheng Wang, *Member, IEEE*, Hanyun Wang, and Jonathan Li, *Senior Member, IEEE*

***Abstract*—This letter presents a novel semisupervised method for addressing a domain adaptation problem in the classification of hyperspectral data. To overcome the influence of distribution bias between the source and target domains, we introduce the domain transfer multiple-kernel learning to simultaneously minimize the maximum mean discrepancy criterion and the structural risk functional of support vector machines. Then, the pairwise binary classifiers are merged as the multiclass classifier for solving the classification problem in hyperspectral data. Both bias and nonbias sampling strategies are introduced to evaluate the robustness of the proposed method against the spectral distribution bias. The results obtained from real data sets show that the proposed method can achieve higher classification accuracy even with cross-domain distribution bias and provide robust solutions with different labeled and unlabeled data sizes.**

***Index Terms*—Domain adaptation (DA), hyperspectral image classification, maximum mean discrepancy (MMD), remote sensing, sample selection bias, support vector machines (SVMs).**

## I. Introduction

HIGH-spectral-resolution images obtained from hyperspectral sensors offer the potential to improve land cover classification performance. In practice, only a small fraction of the training samples within a limited area can be labeled as ground truth to develop land cover classifiers [1], [2]. However, between different areas in hyperspectral data, there exists some shift in the spectral distribution, which can be attributed to the differences in the local illumination or atmospheric conditions, in the photological state, in the shadowing effects, etc. Due to the intraclass variance between the source domain (training area) and the target domain (testing area), the classifiers learned based on the sample domain may not fit in the target domain, even though large amounts of labeled data are trained in the source domain. This phenomenon is called the sample selection bias problem.

The domain adaptation (DA) framework (also known as transfer learning) is aimed at taking advantage of the available knowledge on a given source domain to develop a classifier built on the target domain where *a priori* information is not available [3], [4]. In this letter, we focus on the investigation of learning the kernel function from different areas in hyperspectral images. Recently, active learning [5], [6] has been applied to learn the data set shift and perform sample queries over the target domain. To help classify the data obtained from spatially and temporally different areas, the contextual information in the classifiers trained from source domain data is also exploited [1]. In addition, several new approaches have been performed using the data from the source domain and target domain based on the following: DA support vector machines (SVMs) [7], clustering technology [8], and maximum *a posterior* classifiers [9].

Maximum mean discrepancy (MMD) was proposed [10] to measure the distribution distance across different domains in a reproducing kernel Hilbert space (RKHS). The MMD criterion is learned as a kernel matrix and then applied for feature reduction [11]. This technique, called transfer component analysis, has been introduced for feature extraction in remote sensing data by Matasci *et al.* [12] and reveals improvements for DA. In this letter, we introduce the domain transfer multiple-kernel learning (DTMKL) to hyperspectral image classification, which combines the MMD criterion with SVM learning [13], [14]. In contrast to the two-step approach in [12], DTMKL follows a unified convex optimization procedure to learn the multiple-kernel classifiers by minimizing both the SVM structural risk function and the distribution mismatch between source and target domains in one step. Based on the binary DTMKL classifiers, we adapt the one-against-one approach [15] to solve the multiclass classification problem in hyperspectral data. To closely examine the influence of the sample selection bias problem, we introduce two sampling strategies to simulate both the nonbias sampling and bias sampling. Comparison experiments with classic SVM, Laplacian SVM (LapSVM) [16], [17], simple multiple-kernel learning (SimpleMKL) [18], transductive multiple-kernel learning (TMKL) [19], [20], and the proposed method are carried out on two hyperspectral data sets.

The rest of this letter is outlined as follows. Section II reviews the framework of DTMKL and the definition of MMD. Section III presents the multiple-kernel learning for DTMKL and the solving algorithm. Section IV shows the experimental results. Finally, Section V gives the conclusions.

Manuscript received July 23, 2012; revised November 20, 2012; accepted December 10, 2012. This work was supported in part by the Natural Science Foundation of China under Grant 40971245 and in part by the European Space Agency–Ministry of Science and Technology Dragon 3 Cooperation Project 10689. *(Corresponding author: C. Wang.)*

Z. Sun and C. Wang are with the School of Information Science and Engineering, Xiamen University, Xiamen 361005, China (e-mail: azuring@gmail.com; cwang@xmu.edu.cn).

H. Wang is with the National Key Laboratory of Automatic Target Recognition, College of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, China (e-mail: hanyun.wang1986@gmail.com).

J. Li is with the School of Information Science and Engineering, Xiamen University, Xiamen 361005, China and also with the Waterloo Laboratory for GeoSpatial Technology and Remote Sensing, Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: junli@uwaterloo.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2012.2236818

## II. DTMKL Framework

The DTMKL framework is originally designed to develop a binary classifier. To resolve the multiclass problem for hyperspectral data, the one-against-one strategy is used to combine all the binary DTMKL classifiers to output the final decision. This section describes the formulation of the binary DTMKL classifier.

Let $\{X^S, Y^S\} = \{(\mathbf{x}_i^S, y_i^S)\}_{i=1}^{n_s}$ be the $n_s$ labeled source training data and $\{X^T\} = \{\{\mathbf{x}_j^T\}\}_{j=1}^{n_T}$ be the $n_T$ unlabeled target testing data, with samples $\mathbf{x}_i^S, \mathbf{x}_j^T \in R^d \ \forall i, j$ and labels $y_i^S \in \Omega = \{-1, +1\}$. The data are assumed to be transformed for better representation. Let the mapping be as follows for both domains: $X^S \to \psi(X^S) = X^{S^*}$ and $X^T \to \psi(X^T) = X^{T^*}$. The divergence of the two data distributions after transformation is expected to be minimized, so that $P(X^{S^*}) \approx P(X^{T^*})$.

DTMKL minimizes the distance of data distribution between the two domains, as well as the structural risk function of SVM. The optimization problem can be formulated as

$$[k, f] = \arg\min \left[\Phi\left(\mathrm{Dist}_k(X^S, X^T)\right) + \mu \mathrm{SVM}_{k,f}(X^S)\right] \quad (1)$$

where $\Phi(.)$ is any monotonic increasing function and $\mu > 0$ is a weight parameter to balance the difference of data distribution from the two domains and the structural risk functional SVM for labeled samples. The kernel function $k$ and the SVM decision function $f$ can be learned in one optimization problem.

*Divergence Term:* The first term in (1) is used to measure the mismatch of data distribution between the source and target domains. The MMD, proposed by Borgwardt *et al.* [10], is often used as an indicator for calculating the distribution distance in the RKHS, namely

$$\mathrm{Dist}_K(X^S, X^T) = \left\| \frac{1}{n_S} \sum_{i=1}^{n_S} \psi\left(\mathbf{x}_i^S\right) - \frac{1}{n_T} \sum_{i=1}^{n_T} \psi\left(\mathbf{x}_i^T\right) \right\|^2. \quad (2)$$

By virtue of the well-known kernel trick, the distance in (2) can be rewritten as

$$\mathrm{Dist}_K(X^S, X^T) = \mathrm{tr}(\mathbf{KL}) \quad (3)$$

where

$$\mathbf{K} = \begin{pmatrix} \mathbf{K}^{S,S} & \mathbf{K}^{S,T} \\ \mathbf{K}^{T,S} & \mathbf{K}^{T,T} \end{pmatrix} \in R^{(n_S+n_T)\times(n_S+n_T)} \quad (4)$$

with $\mathbf{K}^{S,S}$, $\mathbf{K}^{T,T}$, $\mathbf{K}^{S,T}$, and $\mathbf{K}^{T,S}$ being the kernel matrices [of elements $K_{i,j} = \psi(\mathbf{x}_i)'\psi(\mathbf{x}_j)$] obtained from the data of the source domain, target domain, and cross-domains, respectively, and $\mathbf{L} = [L_{ij}]$ with $L_{ij} = 1/n_S^2$ if $x_i, x_j \in \mathbf{X}^S$, $L_{ij} = 1/n_T^2$ if $x_i, x_j \in \mathbf{X}^T$, and $L_{ij} = -1/n_S n_T$ if otherwise.

*Discriminative Term:* The objective of the second term in (1) is to minimize the structural risk function of SVM for better classification performance in the target domain. SVM is usually solved by its dual problem that is in the form of the quadratic programming problem. Let $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_n]'$ be a vector of the dual variables $\alpha_i$ for each labeled sample, $\mathbf{y} = [y_1, \ldots, y_n]'$ be the label vector, and $\mathbf{K}^{S,S} = [k(\mathbf{x}_i, \mathbf{x}_j)] \in R^{n_S \times n_S}$ be the kernel matrix of labeled samples with $k(\mathbf{x}_i, \mathbf{x}_j) = \psi(\mathbf{x}_i)'\psi(\mathbf{x}_j)$. Summing up the two terms simultaneously, we obtain the following saddle-point minimax problem:

$$\min_{\mathbf{K}\geq 0} \left( \Phi(\mathrm{tr}(\mathbf{KL})) + \mu \max_{\boldsymbol{\alpha}\in A} \left( \boldsymbol{\alpha}'\mathbf{1} - \frac{1}{2}(\boldsymbol{\alpha}\circ\mathbf{y})'\mathbf{K}^{S,S}(\boldsymbol{\alpha}\circ\mathbf{y}) \right) \right) \quad (5)$$

where $A = \{\boldsymbol{\alpha} \in R^{n_S} | C\mathbf{1} \geq \boldsymbol{\alpha} \geq 0, \boldsymbol{\alpha}'\mathbf{y} = 0\}$ is the feasible set of $\boldsymbol{\alpha}$, with $C$ being the regularization parameter for SVM. The elementwise product vector between two vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ is represented as $\boldsymbol{\alpha} \circ \boldsymbol{\beta}$.

## III. Multiple-Kernel Learning in DTMKL

### A. Objective Function Formulation

The standard SVM builds a learning model using only a single kernel and fixed kernel parameters. The disadvantage is that it lacks the generalization capability of dealing with multidimensional data. Recent research [19]–[21] has revealed that multiple-kernel learning can provide a more flexible framework to mine the information more effectively. In this letter, in order to simplify the kernel learning and facilitate the usage of the existing SVM solver, the kernel function $k$ is assumed as a linear combination of a set of $M$ base kernel functions $\{k_m\}$, i.e., $k = \sum_{m=1}^{M} d_m k_m$, where $d_m \geq 0$ and $\sum_{m=1}^{M} d_m = 1$. The two kernel matrices are changed to multiple-kernel version: $\mathbf{K} = \sum_{m=1}^{M} d_m \mathbf{K}_m$ and $\mathbf{K}^{S,S} = \sum_{m=1}^{M} d_m \mathbf{K}_m^{S,S}$, where $\mathbf{K}_m \in R^{(n_S+n_T)\times(n_S+n_T)}$ and $\mathbf{K}_m^{S,S} \in R^{n_S \times n_S}$ are the $m$th base kernel matrices defined for both domains and for the labeled patterns, respectively. The $\Phi$ function in (5) is defined as $\Phi(x) = \frac{1}{2}(x)^2$, which is strictly convex and can help to accelerate the convergence speed in kernel learning. Applying the kernel trick, the optimization problem can be simplified as follows:

$$\min_{\mathbf{d}\in D} \left( \frac{1}{2} \left( \mathrm{tr} \left( \sum_{m=1}^{M} d_m \mathbf{K}_m \mathbf{L} \right) \right)^2 + \mu \max_{\alpha\in A} J(\mathbf{d}) \right) \quad (6)$$

where

$$J(\mathbf{d}) = \boldsymbol{\alpha}'\mathbf{1} - \frac{1}{2}(\boldsymbol{\alpha}\circ\mathbf{y})' \left( \sum_{m=1}^{M} d_m \mathbf{K}_m^{S,S} \right) (\boldsymbol{\alpha}\circ\mathbf{y})$$

and $D = \{\mathbf{d} | \mathbf{d} \geq 0, \mathbf{d}'\mathbf{1} = 1\}$ is the feasible set of $\mathbf{d}$.

### B. Optimization Algorithm

Solving (6) is a saddle-point minimax problem. Let $p_m = \mathrm{tr}(\mathbf{K}_m\mathbf{L})$ and $\mathbf{p} = [p_1, \ldots, p_M]'$; then, (6) can be transformed as

$$\min_{\mathbf{d}\in D} h(\mathbf{d}) = \min_{\mathbf{d}\in D} \frac{1}{2}\mathbf{d}'\mathbf{p}\mathbf{p}'\mathbf{d} + \mu \max_{\alpha\in A} J(\mathbf{d}). \quad (7)$$

The different variables ($\boldsymbol{\alpha}$ and $\mathbf{d}$) can be updated iteratively to obtain the global optimization solution. With a fixed $\mathbf{d}$, a standard SVM solution is used to solve the dual problem to maximize $J(\mathbf{d})$. When SVM parameter $\boldsymbol{\alpha}$ is held constant, (7) can be updated using second-order Newton method to obtain the kernel weight vector $\mathbf{d}$. The initialization value for $\mathbf{d}$ is often set to $1/M$.

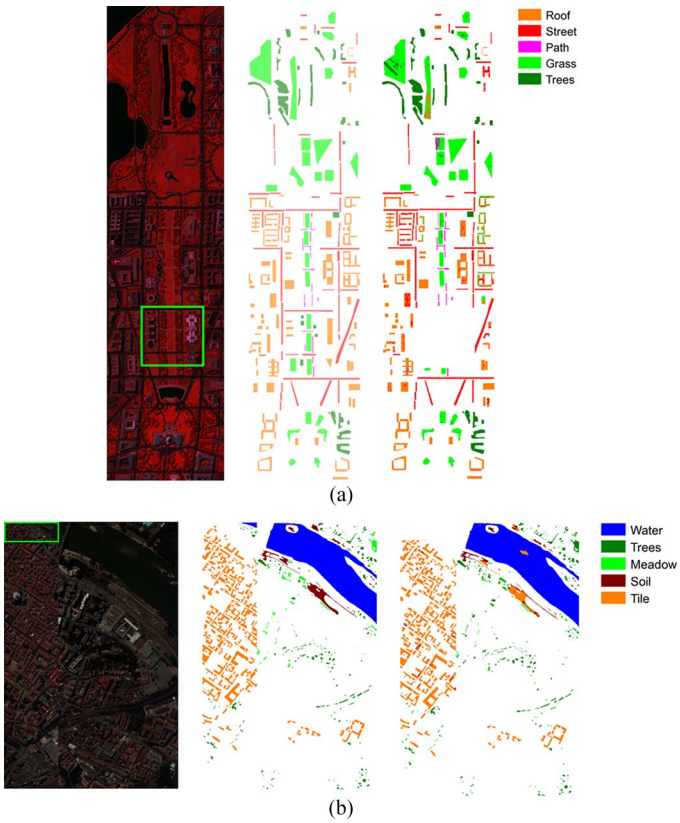


Fig. 1. Pseudocolor image, ground truth map, and DTMKL classification map for hyperspectral data. The green rectangles in the pseudocolor images represent the source domain area. (a) WDC data set. (b) PC data set.

## IV. Experiments and Results

### A. Experimental Setup

Two hyperspectral data sets were used in our experiments. The first hyperspectral data set, shown in Fig. 1(a), is an image of 1280 × 307 pixels taken over the Washington DC (WDC) mall by the HYperspectral Digital Imagery Collection Experiment sensor. The five land cover classes—which are confined in the source domain area (green rectangle in the figure)—chosen are as follows: roof, street, path, grass, and trees. The second data set, shown in Fig. 1(b), is a 1.3-m-spatial-resolution hyperspectral image acquired by the ROSIS-03 optical sensor over the city center of Pavia [Pavia center (PC)]. The image contains 1096 × 715 pixels. The five land cover classes—which exist in the source domain area—considered are as follows: water, trees, meadow, soil, and tiles.

The results of the proposed method were compared with those of the classic SVM, LapSVM, SimpleMKL, and TMKL methods. To analyze the influence of multiple kernels in DTMKL, a single-kernel version for domain transfer (noted as DT-SK-SVM) was also compared, which selects one fixed kernel that minimizes the objective function (5). For the classic SVM and LapSVM, we use a Gaussian kernel [i.e., $k(x_i, x_j) = \exp(-\lambda\|x_i - x_j\|^2)$] as the default kernel. A grid search has been run to get the optimal SVM parameters using a tenfold cross-validation, with $C \in \{0.5, 1, 2, 4\}$ and $\lambda \in \{0.0625, 0.125, 0.25, 0.5\}$. For the other four methods, we add three other types of kernels: Laplacian kernel [i.e., $k(x_i, x_j) = \exp(-\sqrt{\lambda}\|x_i - x_j\|)$], inverse square distance kernel [i.e., $k(x_i, x_j) = 1/(\lambda\|x_i - x_j\|^2 + 1)$], and inverse distance kernel [i.e., $k(x_i, x_j) = 1/(\sqrt{\lambda}\|x_i - x_j\| + 1)$]. For each kind of kernel, the $\lambda$ has four possible values $\{2^{-1}, 2^{-2}, 2^{-3}, 2^{-4}\}$. The selection of base kernel types and parameters is based on practical experiments by testing various combinations of kernels. All the experiments are done on a 2.53-GHz Core 2 CPU with 8-GB RAM.

TABLE I
Data Sets Used in the Experiments

| | Washington DC | | Pavia Center | |
|---|---|---|---|---|
| | Training | Testing | Training | Testing |
| #samples | 5845 | 55805 | 1777 | 124292 |
| #bands | 191 | 191 | 102 | 102 |
| #classes | 5 | 5 | 5 | 5 |

### B. Sampling Strategies for Performance Validation

In traditional evaluation experiments, the whole image is considered as both the source and target areas so that, from a statistical point of view, there is no distribution bias between the source and target domains. In this traditional case, we consider the random sampling strategy over the whole image to be "nonbias sampling."

To investigate the influence of the sample selection bias problem, we adopt, for comparison, a second sampling strategy—"bias sampling." In the second case, a segment of the hyperspectral image, represented by the area within the green rectangles in the pseudocolor images in Fig. 1, is designated the source domain area; the remainder of the image is designated the target domain area. The pixels within the green rectangles, i.e., the source domain, are considered to be labeled samples and are sampled for training data; the pixels outside the green rectangles, i.e., the target domain, are considered to be unlabeled samples and are sampled to help training and used for validation data.

To compare the results more fairly, we set the numbers of training and testing samples equal in both the two strategies. Table I summarizes the following three properties of the data sets: the number of samples (#samples), the number of spectral bands (#bands), and the number of classes (#classes). In the experiment in Section IV-C, the training and testing data are sampled in ratio of 50% for every class. In the experiment in Section IV-D, we varied the rate of both labeled and unlabeled samples independently in the set {10%, 30%, 50%, 70%, 90%, 100%} for the classification robustness validation. To reliably evaluate the performance of the different methods, all the results are averaged over five different randomly selected training and validation data sets for each experiment.

### C. Classification Performance Assessment

The classification accuracies and the kappa statistics of the proposed method and the other five methods are summarized in Table II with the maximum values in every row in bold. Under the nonbias sampling strategy, we can see that SimpleMKL, TMKL, and DTMKL outperform the other methods in terms of overall accuracy (OA). The bias is artificially removed by the

TABLE II
OA AND KAPPA RESULTS WITH DIFFERENT METHODS AND SAMPLING STRATEGIES ON WDC AND PC DATA SETS

| | Non-Bias Sampling | | | | Bias Sampling | | | |
|---|---|---|---|---|---|---|---|---|
| | WDC | | PC | | WDC | | PC | |
| | OA | Kappa | OA | Kappa | OA | Kappa | OA | Kappa |
| SVM | 0.946±0.006 | 0.925 | 0.983±0.005 | 0.959 | 0.859±0.006 | 0.807 | 0.930±0.004 | 0.878 |
| LapSVM | 0.939±0.009 | 0.912 | 0.979±0.006 | 0.951 | 0.847±0.014 | 0.781 | 0.934±0.007 | 0.887 |
| SimpleMKL | **0.952±0.003** | **0.933** | 0.985±0.002 | 0.961 | 0.875±0.005 | 0.829 | 0.935±0.004 | 0.889 |
| TMKL | 0.951±0.003 | 0.930 | **0.990±0.002** | **0.966** | 0.862±0.012 | 0.812 | 0.928±0.011 | 0.861 |
| DT-SK-SVM | 0.943±0.008 | 0.921 | 0.981±0.007 | 0.956 | 0.863±0.009 | 0.812 | 0.932±0.006 | 0.885 |
| DTMKL | 0.951±0.004 | 0.930 | 0.985±0.003 | 0.962 | **0.885±0.007** | **0.843** | **0.944±0.005** | **0.927** |

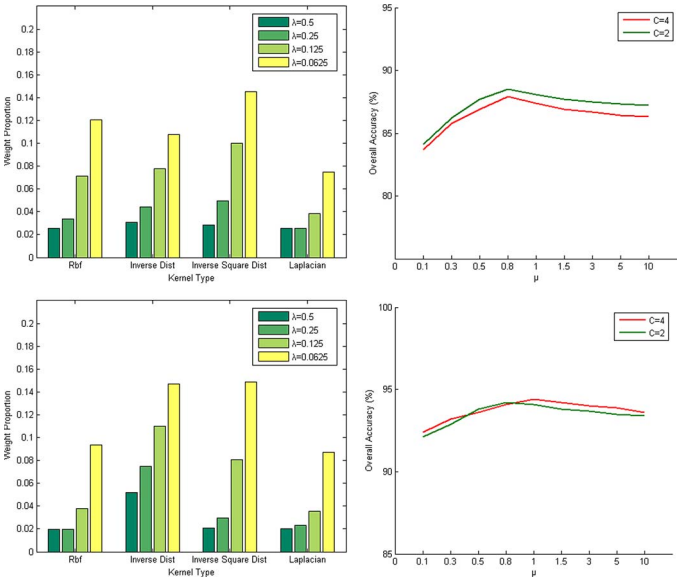


Fig. 2. Weight proportion values of different kernels and $\mu$ sensitive analysis for the (top row) WDC data set and (bottom row) PC data set.

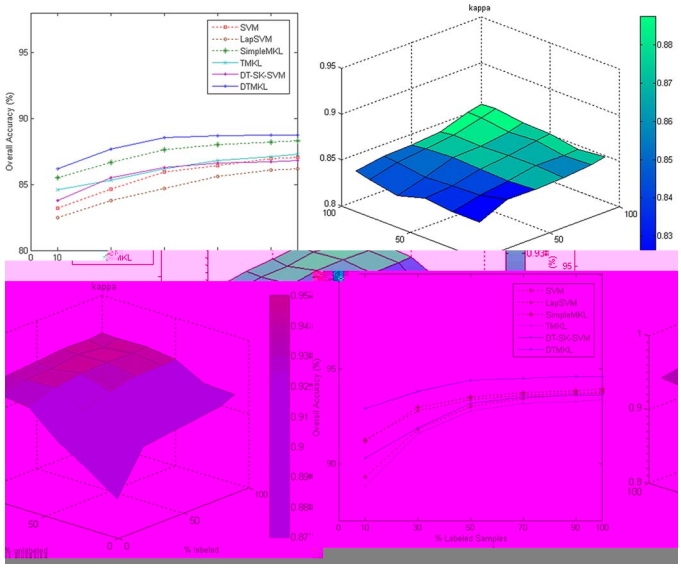


Fig. 3. Results for the (top row) WDC data set and (bottom row) PC data set. OA over the validation set as a function of the ratio of labeled samples using different methods is presented in the left side. Kappa statistic surface is presented in the right side.

first sampling strategy. Thus, all the three methods can achieve excellent classification accuracies. However, the unreasonable assumption (no bias) is often invalid in practice. When the training data sampling is constrained within the source domain area under the bias sampling strategy, the classification accuracies of all the classifiers obviously decrease due to the bias sample selection problem. DTMKL performs the best among the six methods under the bias sampling strategy. For the WDC data set, DTMKL has gains of $+2.6\%$, $+2.3\%$, and $+1.0\%$ over SVM, TMKL, and SimpleMKL, respectively. For the PC data set, DTMKL has gains of $+1.2\%$, $+1.0\%$, and $+0.9\%$ over DT-SK-SVM, LapSVM, and SimpleMKL, respectively. DTMKL learns the kernel matrix to map different domain data "closer" in the RHKS by utilizing the unlabeled data in the target domain. In the same table, we report the standard deviation of OA exhibited by each classification method. Low sensitivity is obtained by the DTMKL and SimpleMKL classifiers compared with those achieved by the other semisupervised methods like LapSVM and TMKL.

For insight of the importance of different kernels in DTMKL, we exhibit in Fig. 2 the weight proportion of all used kernels. We observe that the inverse distance kernel and inverse square distance kernel have higher weights than the other kernels. The facts demonstrate that the two kernel types have more capacity to match the data distribution of the two domains as well as minimize the structural risk of SVM.

We also investigate the performance variation of DTMKL with respect to the different balance parameters $\mu \in [0.1, 10]$, in which we set the regularization parameter $C = 2$ or 4. As shown in Fig. 2, the classification accuracy decreases when $\mu$ is too large or too small. DTMKL obtains the best performance when $\mu$ is set between 0.5 and 1.5. In that range, the classification accuracy is not sensitive to the parameter $\mu$.

### *D. Evaluate the Robustness to Different Sizes of Labeled and Unlabeled Data*

The validation results of the six methods with different labeled sample sizes are presented in Fig. 3. The corresponding DTMKL classification maps are shown in Fig. 1. Several conclusions can be drawn from the experimental results. First, DTMKL classifiers produce better classification results than SVM and SimpleMKL in all cases. Second, this gain is particularly noticeable when only a low number of labeled samples are available. This phenomenon can be explained that the unlabeled data play a more important role when the labeled data are not adequate. Third, the classification accuracy of the proposed method increases as the percentage of labeled samples increases. The kappa statistic surface shown in the right side of Fig. 3 also confirms, in general terms, the importance of the labeled information in this problem.

## V. Conclusion

The distribution bias between source domain and target domains often leads to poor classification performance of hyperspectral images. DTMKL is an effective framework to solve the cross-domain classification problem for hyperspectral data. The results show that the proposed method using DTMKL can effectively improve the classification by learning cross-domain knowledge and is robust to different sizes of training data.

## Acknowledgment

The authors would like to thank Prof. D. A. Landgrebe and Prof. P. Gamba for providing the hyperspectral data and the three anonymous reviewers and M. McAllister for giving valuable comments to improve this letter.

## References

[1] S. Rajan, J. Ghosh, and M. M. Crawford, "Exploiting class hierarchies for knowledge transfer in hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 3408–3417, Nov. 2006.

[2] B. Huang, C. Xie, R. Tay, and B. Wu, "Land-use-change modeling using unbalanced support-vector machines," *Environ. Plann. B, Plann. Des.*, vol. 36, no. 3, pp. 398–416, 2009.

[3] L. Bruzzone and D. F. Prieto, "Unsupervised retraining of a maximum likelihood classifier for the analysis of multitemporal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 456–460, Feb. 2001.

[4] H. Daumé, III and D. Marcu, "Domain adaptation for statistical classifiers," *J. Artif. Intell. Res.*, vol. 26, no. 1, pp. 101–126, May 2006.

[5] D. Tuia, F. Ratle, F. Pacifici, M. F. Kanevski, and W. J. Emery, "Active learning methods for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2218–2232, Jul. 2009.

[6] D. Tuia, E. Pasolli, and W. J. Emery, "Dataset shift adaptation with active queries," in *Proc. Urban Remote Sens. Event*, 2011, pp. 121–124.

[7] 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 Trans. Geosci. Remote Sens.*, vol. 47, no. 4, pp. 1108–1122, Apr. 2009.

[8] L. Gómez-Chova, G. Camps-Valls, L. Bruzzone, and J. Calpe-Maravilla, "Mean map kernel methods for semisupervised cloud classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 1, pp. 207–220, Jan. 2010.

[9] K. Bahirat, F. Bovolo, L. Bruzzone, and S. Chaudhuri, "A novel domain adaptation Bayesian classifier for updating land-cover maps with class differences in source and target domains," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 7, pp. 2810–2826, Jul. 2012.

[10] K. M. Borgwardt, A. Gretton, M. J. Rasch, H. P. Kriegel, B. Schölkopf, and A. J. Smola, "Integrating structured biological data by kernel maximum mean discrepancy," *Bioinformatics*, vol. 22, no. 14, pp. e49–e57, Jul. 2006.

[11] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Trans. Neural Netw.*, vol. 22, no. 2, pp. 199–210, Feb. 2011.

[12] G. Matasci, M. Volpi, D. Tuia, and M. Kanevski, "Transfer component analysis for domain adaptation in image classification," in *Proc. 17th SPIE, Image Signal Process. Remote Sens.*, 2011, vol. 8180, pp. 81 800F-1–81 800F-9.

[13] L. Duan, I. W. Tsang, D. Xu, and S. J. Maybank, "Domain transfer SVM for video concept detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 1375–1381.

[14] L. Duan, I. W. Tsang, and D. Xu, "Domain transfer multiple kernel learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 3, pp. 465–479, Mar. 2012.

[15] L. Bottou, C. Cortes, J. S. Denker, H. Drucker, I. Guyon, L. D. Jackel, Y. LeCun, U. A. Muller, E. Sackinger, P. Simard, and V. Vapnik, "Comparison of classifier methods: A case study in handwritten digit recognition," in *Proc. Pattern Recognit.*, Oct. 9–13, 1994, vol. 2, pp. 77–82.

[16] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *J. Mach. Learn. Res.*, vol. 7, pp. 2399–2434, Dec. 2006.

[17] L. Gómez-Chova, G. Camps-Valls, J. Muñoz-Marí, and J. Calpe, "Semisupervised image classification with Laplacian support vector machines," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 336–340, Jul. 2008.

[18] A. Rakotomamonjy, F. Bach, S. Canu, and Y. Grandvalet, "SimpleMKL," *J. Mach. Learn. Res.*, vol. 9, pp. 2491–2521, Nov. 2008.

[19] X. Tian, G. Gasso, and S. Canu, "A multiple kernel framework for inductive semi-supervised SVM learning," *Neurocomputing*, vol. 90, pp. 46–58, Aug. 2012.

[20] D. Tuia, G. Camps-Valls, G. Matasci, and M. Kanevski, "Learning relevant image features with multiple-kernel classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3780–3791, Oct. 2010.

[21] Y. Gu, C. Wang, D. You, Y. Zhang, S. Wang, and Y. Zhang, "Representative multiple kernel learning for classification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 7, pp. 2852–2865, Jul. 2012.