Title: Interactive Domain Adaptation for the Classification of Remote Sensing Images using Active Learning

This paper appears in: IEEE Geoscience and Remote Sensing Letters

Date of Publication: 2013

Author(s): Claudio Persello

Volume: 10, Issue: 4

Page(s): 736 - 740

DOI: 10.1109/LGRS.2012.2220516

Interactive Domain Adaptation for the Classification of Remote Sensing Images using Active Learning

Claudio Persello, *Member, IEEE*

***Abstract*—This paper presents a novel interactive domain-adaptation technique based on active learning for the classification of remote sensing (RS) images. The proposed method aims at adapting the supervised classifier trained on a given RS *source image* to make it suitable for classifying a different but related *target image*. The two images can be acquired in different locations and/or at different times. The proposed approach iteratively selects the most informative samples of the target image to be labeled by the user and included in the training set, while the source-image samples are re-weighted or possibly removed from the training set on the basis of their disagreement with the target-image classification problem. In this way, the consistent information available from the source image can be effectively exploited for the classification of the target image and for guiding the selection of new samples to be labeled, whereas the inconsistent information is automatically detected and removed. This approach can significantly reduce the number of new labeled samples to be collected from the target image. Experimental results on both a multispectral Very High Resolution (VHR) and a hyperspectral dataset confirm the effectiveness of the proposed method.**

***Index Terms*—Domain Adaptation, Active Learning, Image Classification, Support Vector Machine.**

I. Introduction

THE continuously growing availability of RS images gives us the opportunity to develop several important applications related to land-cover monitoring and mapping. In order to exploit such an opportunity, it is necessary to develop adequate classification systems capable to produce accurate land-cover maps at reasonable cost and time. At the present, the most common approach to obtain land cover maps is based on supervised learning methods that require a new set of labeled training samples every time that a new RS image has to be classified, leading to high costs for the acquisition of additional reference information. This is due to possible differences in the image acquisition conditions (e.g., illumination, viewing angle), ground conditions (e.g., soil moisture, topography) or in the phenological stages of the vegetation that may affect the observed spectral signatures of the land-cover classes. Therefore, the labeled samples of a given RS image cannot in general be directly used for: 1) classifying another image of a different area (with similar characteristics), or 2) updating a land-cover map given a new image acquired on the same geographical area. However, both problems can be modeled in the framework of domain adaptation (DA), whose goal is to adapt a classifier initially trained with examples coming from a *source domain* to produce good predictive performances on samples coming from a different but related *target domain*.

C. Persello is with the Max Planck Institute for Intelligent Systems, Spemannstrasse 41, 72076 Tübingen, Germany, and also with the Department of Information Engineering and Computer Science, University of Trento, 38123 Trento, Italy (e-mail: claudio.persello@tuebingen.mpg.de)

In this paper, we propose an interactive DA technique for the classification of RS images that allows one to exploit the consistent information of the source image to classify the target image. In this way, the amount of target samples to be labeled can be significantly reduced. The proposed method is interactive, since the user is guided by the classification system by means of an active learning (AL) technique [1]–[4] that iteratively select the most informative samples from the target image to be labeled.

The main novel contributions of the present work are: 1) the use of a *query+* function that considers both uncertainty and diversity criteria for addressing DA problems, 2) the introduction of a *re-weighting* mechanism for source-domain samples based on the cosine-angle similarity measure in the kernel space, 3) the definition of a *query-* that adaptively selects the inconsistent samples to be discarded.

II. Problem formulation and state of the art

In order to statistically characterize the variation between source and target domain, i.e., the data-set shift between two RS images, let $P^s(\mathbf{x}, y) = P^s(\mathbf{x})P^s(y|\mathbf{x})$ and $P^t(\mathbf{x}, y) = P^t(\mathbf{x})P^t(y|\mathbf{x})$ denote the joint distributions of the feature vector $\mathbf{x}$ (e.g., the spectral signatures) and the class label y for the source and the target domain, respectively. Several works at the state of the art (e.g., [3], [5], [6]) assume (explicitly or implicitly) that the conditional probabilities for the different domains are approximately equal, while only the marginal distributions are allowed to change, that is $P^t(y|\mathbf{x}) \approx P^s(y|\mathbf{x})$ and $P^t(\mathbf{x}) \neq P^s(\mathbf{x})$. This means that it is believed that the classification problems defined on the two domains are actually the same or very similar, but the estimated classification functions (learned from the data) may be different due to a different sampling of the feature space. Under such an assumption, the problem is usually referred to as *covariate shift*. In [5], a method for addressing the covariate shift problem by re-weighting source-domain samples is proposed. The weights for the source-domain samples are obtained by minimizing the discrepancy between the distributions of the unlabeled samples in the source and target domain. In [6], an AL technique to address data-set shift problems in the classification of RS images under covariate shift is proposed. Finally, an AL method to address transfer learning problems in the classification of hyperspectral data has been proposed in [3], employing a sample re-weighting method based on the TrAdaBoost algorithm [7].

It is worth noting that the covariate shift assumption is very strong and in real DA problems it is usually not possible to asses its validity. Moreover, in our setting we are explicitly considering that the spectral signatures of the classes may change across RS images, this implies that such assumption is generally not satisfied. Therefore, in this work we assume that both marginal and conditional probabilities are allowed to vary. An important consequence of this assumption is that part of the source-domain samples may not be consistent with the target-domain classification problem (i.e., their class labels can be wrong for the target problem). For this reason, the DA technique should be able to automatically detect and remove such inconsistent samples from the training set, preventing them to reduce prediction accuracy on the target image. To this aim, the concept of *query-* function was introduced in [8]. However, the query- presented in [8] removes a fixed amount of samples at any iteration of the AL procedure, which might be not an optimal strategy. In this work, we propose a method for automatically detecting inconsistent source-domain samples to be removed from the training set.

Another important observation regards the query+ function used by the AL algorithm to select the most informative samples from the target domain. Traditional AL query functions based on the *uncertainty* criterion only are not optimal for the considered DA problem due to the biased estimation of the decision boundaries especially in the initial iterations. Indeed, we expect that the decision boundaries learned by the classifier may shift significantly from the source toward the target domain problem. The proposed method adopts a query+ function that considers both the *uncertainty* and *diversity* criteria for the selection of non-redundant batches [4]. This allows the query+ to both reduce redundancy among selected samples and improve *exploration* of the feature space.

III. Proposed Interactive Domain-Adaptation Method

The main goal of the proposed interactive domain-adaptation (IDA) method is to exploit the consistent information of the source image to classify the target image and for guiding the user in the selection of the most informative samples to be labeled. The proposed approach consists in an iterative procedure based on AL. At the first step, a supervised algorithm is trained using only source-domain training samples. For all the subsequent iterations, a query function selects the most informative samples of the target image, which the user is requested to annotate. The new-labeled samples are added to the training set for re-training the supervised classification algorithm. The classifier is trained considering different weights for instances of the source and target domain. Target-domain samples are considered fully reliable and are therefore associated to weight one. Source-domain samples are re-weighed according to their agreement with the target-domain problem (considering the difference between the class-conditional densities in the two domains) and associated to weights in the range $(0, 1]$. In addition, inconsistent samples from the source domain are automatically detected and removed (i.e., associated to weight zero) in order to prevent them to mislead the classification on the target domain. The proposed system consists of the following main components:

A) *Query+*: selects the most informative samples of the target domain to be labeled by the user;
B) *Re-weighting*: re-weights source-domain samples according to their agreement with target-domain samples;
C) *Query-*: removes source-domain samples that are inconsistent for the target-domain problem.

Using these three components, the proposed system iteratively adapts the classifier to the target-domain problem. If the two classification problems are highly related, the number of samples of the target image to be annotated can be strongly reduced, by exploiting most of the source-domain samples. If the classification problems are less similar, the proposed system will nevertheless allow the classifier to adapt to the target domain. In the proposed approach, the supervised classification is performed using support vector machines (SVM) [9], which proved very effective in the classification of both multispectral and hyperspectral images [10]. In particular, we adopt a formulation that considers different weights for source-domain instances in the learning phase. More precisely, we solve the following constrained minimization problem:

$$
\begin{aligned}
&\min_{\mathbf{w},\xi^s,\xi^t,b} \frac{1}{2}\|\mathbf{w}\|^2 + C\left(\sum_{j=1}^{m}\xi_j^t + \sum_{i=1}^{n}\beta_i\xi_i^s\right)\\
&\text{subject to: } y_j^t[\mathbf{w}\cdot\phi(\mathbf{x}_j^t)+b] \geq 1-\xi_j^t \qquad j=1,...,m\\
&\qquad\qquad\quad y_i^s[\mathbf{w}\cdot\phi(\mathbf{x}_i^s)+b] \geq 1-\xi_i^s \qquad i=1,...,n\\
&\qquad\qquad\quad \xi_j^s, \xi_i^t \geq 0
\end{aligned} \tag{1}
$$

where $\mathbf{w}$ is a vector orthogonal to the separating hyperplane, b is a bias term such that $b/\|\mathbf{w}\|$ represents the distance of the hyperplane from the origin, C is the regularization parameter, ϕ is function mapping the data into the feature space, ξ_i^s and ξ_i^t are the slack variables associated with source and target-domain samples, respectively, m and n are the number of target and source-domain samples at a given iteration, respectively, and β_i are the weights for source samples obtained according to the procedures that will be detailed in subsections B and C.

A) Query+: The aim of the query+ is to select a batch of the most informative samples from a pool U of unlabeled samples, which are taken from the target domain. Once selected, such samples are manually labeled by the user, and added to the training set. In our approach, we adopted the batch-mode query function MCLU-ECBD proposed in [4]. Such a technique selects a batch of informative samples from the pool by considering both uncertainty and diversity. The uncertainty criterion is associated to the confidence of the supervised algorithm in correctly classifying the considered samples, while the diversity aims at selecting a set of unlabeled samples that are as more diverse (distant one another) as possible, thus reducing the redundancy among the selected samples. The MCLU (namely *multiclass-level uncertainty*) technique evaluates the uncertainty of the samples for multiclass classification problems considering the *one-against-all* architecture of binary SVMs. The ECBD (namely *enhanced clustering-based diversity*) technique provide diversity by applying kernel k-means clustering to the u most uncertain samples selected by MCLU to identify $h = k$ clusters, and finally selects the most uncertain sample from each cluster. The combination of the two criteria results in the selection of the h potentially most informative samples of the target domain at any iteration. Moreover, the ECBD technique prevent the query+ from selecting only samples that are close to the decision boundary, thus obtaining a better exploration of the feature space. Such property is fundamental in our DA setting, especially in the first iterations of the AL procedure, where the estimated decision boundary will be biased toward the source-domain problem and therefore the uncertainty criterion would result in the selection of suboptimal samples. For a detailed description of the MCLU-ECBD AL technique we refer the reader to [4].

B) Re-weighting: In order to take into account the difference between the class-conditional densities $P^s(\mathbf{x}|y)$ and $P^t(\mathbf{x}|y)$ we adopt a strategy that re-weights source-domain samples. The weight for each source-domain sample is computed by considering its similarity to the target-domain samples of the same class according to the mean cosine-angle similarity, defined as:

$$\beta_i = \frac{1}{m_{y_i^s}} \sum_{j:y_j^t=y_i^s} \frac{k(\mathbf{x}_i^s, \mathbf{x}_j^t)}{\sqrt{k(\mathbf{x}_i^s, \mathbf{x}_i^s)k(\mathbf{x}_j^t, \mathbf{x}_j^t)}} \tag{2}$$

where $m_{y_i^s}$ is the number of target-domain samples $\mathbf{x}_j^t$ associated to the same class $y_j^t = y_i^s$ of the source-domain sample and $k(\cdot,\cdot)$ is a positive semidefinite kernel function. In our implementation we adopted an RBF kernel function (the same as for the SVM classifier). The weights β_i assume therefore values in the range $(0, 1]$. The rationale of this re-weighting procedure is to reduce the weight of source-domain samples that a far apart from the samples of the same class in the target domain, as they are considered not in agreement (or less reliable) for the target-domain classification problem.

C) Query-: Since $P^t(y|\mathbf{x})$ can be different from $P^s(y|\mathbf{x})$, some source-domain samples may not be consistent for the target-domain classification problem (i.e., their class labels may be wrong for the target problem). It is therefore very important to identify and remove the source-domain samples that bring misleading information for the classification of the target image. In [8], a query- function for removing misleading samples from the training set was proposed. However, the query- in [8] removes a fixed amount of source-domain samples at each iteration of the AL process. Here, we adopt instead a simple heuristic to remove the inconsistent source-domain samples from that training set that does not require fixing a priori the amount of samples to be removed at each iteration. In the proposed methodology, we remove at each iteration the source-domain samples that are misclassified by the SVM classifier. This is done by setting $\beta_i = 0$ in correspondence with the misclassified source-domain samples.

Summarizing, at each iteration of the AL process, the new-labeled samples selected by the query+ are included in the training set, the weights β_i of source-domain samples are re-computed considering the re-weighting procedure and the query- function, and the SVM algorithm is re-trained according to (1).

TABLE I
NUMBER OF LABELED SAMPLES AVAILABLE FOR THE TWO MULTISPECTRAL IMAGES QB_1 AND QB_2.

Class	Number of Samples			
	QB_1		QB_2	
	T_1	VAL_1	T_2	TS_2
Vineyard	658	314	848	6677
Water	98	32	266	1180
Agriculture Fields	105	45	260	620
Forest	272	146	332	2434
Apple Tree	3060	1523	2712	3273
Urban Area	234	116	250	1780
Total/Average	4427	2176	4668	15964

IV. EXPERIMENTAL EVALUATION

We carried out different experiments in order to assess the effectiveness of the proposed technique and compare it with state-of-the-art techniques. The experiments are carried out on both a multispectral VHR and a hyperspectral data set. The description of the two data sets and the design of experiments are given below.

A) VHR data set: The first data set is made up of two VHR multispectral images acquired by the Quickbird satellite (named QB_1 and QB_2 hereafter) over agricultural areas in the south of the city of Trento, Italy. The spatial resolution of the multispectral channels is 2.8 m, while the panchromatic band has a geometric resolution of 0.7 m. The first image QB_1 consists of 2066 $\times$ 2983 pixels, while the size of the second image QB_2 is 3100 $\times$ 2066 pixels. The available labeled samples for the two images (detailed for each land cover class) are reported in Table I. The experiments were carried out in order to adapt the SVM classifier trained on QB_1 (considered as source image) to the classification of QB_2 (considered as target image). From the original training set T_1 of the source domain, ten different initial training sets of 965 samples were derived. The ten initial training sets were uses for training the classifier at the first iteration in ten different trials. The values for the C parameter of the SVM classifier and the variance of the RBF kernel were selected according to a grid-search approach in order to maximize the overall accuracy (OA) on the validation set VAL_1. The set of labeled samples T_2 of the target image was used as pool U for the query+ function. We also performed the classification of QB_2 applying AL directly to the target domain (ignoring the source-domain information). Ten different trials were performed starting AL from initial training sets obtained by randomly selecting 60 samples from T_2 (10 samples per class). The rest of the samples of T_2 were used as pool. The tuning of the free parameters of the SVM was done by performing cross-validation on the initial training samples. The accuracy on the target image was computed using the test set TS_2.

B) Hyperspectral data set: The second data set is a hyperspectral image acquired by the Hyperion sensor of the EO-1 satellite in an area of the Okavango Delta, Botswana. The considered image has a spatial resolution of 30 m over a 7.7 km strip in 145 bands. For greater details on this data set, we refer the reader to [2]. Reference labeled samples of 14

TABLE II
NUMBER OF AVAILABLE LABELED SAMPLES FOR THE HYPERSPECTRAL DATA SET.

Class	Number of Samples			
	Area 1		Area 2	
	T_1	VAL_1	T_2	TS_2
Water	69	57	213	57
Hippo Grass	81	81	83	18
Floodplain Grasses 1	83	75	199	52
Floodplain Grasses 2	74	91	169	46
Reeds	80	88	219	50
Riparian	102	109	221	48
Firescar	93	83	215	44
Island Interior	77	77	166	37
Acacia Woodlands	84	67	253	61
Acacia Shrublands	101	89	202	46
Acacia Grasslands	184	174	243	62
Short Mopane	68	85	154	27
Mixed Mopane	105	128	203	65
Exposed Soil	41	48	81	14
Total	1242	1252	2621	627

land-cover classes are available for two different and spatially disjoint areas, which are referred in the following as Area 1 and Area 2, representing two different geographical areas with the same set of land-cover classes characterized by slightly different distributions. The labeled samples taken from Area 1 were randomly partitioned into two sets T_1 and VAL_1 and the samples of Area 2 were similarly partitioned into a training set T_2 and a test set TS_2, as in [8] (see Table II for detailed information). The experiments on the hyperspectral data set were carried out in order to adapt the classifier trained on the Area 1 to the spatially separate Area 2. Also for the hyperspectral data set, ten different initial training sets made up of 739 samples were selected. These training sets were used for training the classifier at the first iteration in ten different trials. The set VAL_1 was used as validation set for the model selection. As pool U for the AL process we considered T_2. As done for the VHR data set, we also applied AL directly to Area 2 starting from initial training sets made up of 70 samples (5 samples per class) randomly selected from T_2. TS_2 was used as test set to evaluate the classification accuracy on the target domain.

V. EXPERIMENTAL RESULTS

For both data sets, we compared the results obtained by the proposed IDA method with those obtained by using: 1) random selection, 2) the standard MCLU AL method [4], 3) a method that combines the MCLU query+ with the re-weighting procedure proposed in [3], and 4) AL directly applied to the target domain using MCLU-ECBD. For all the methods, the query+ was set to select $h = 5$ samples per iteration.

A) VHR data set: Fig. 1 shows the OA on the target domain (averaged over the ten trials) obtained with the considered methods versus the number of pool samples added to the training set. The obtained results show that the proposed technique lead to significantly higher accuracies than standard methods, confirming its effectiveness in exploiting the consistent information of the source image and in removing the inconsistent

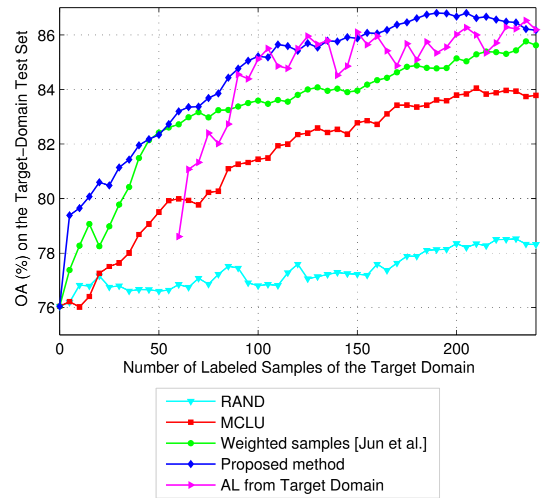

Fig. 1. OA obtained on TS_2 versus the number of samples of U labeled and added to the training set. The different curves correspond to: 1) random selection, 2) the MCLU AL method, 3) a method that combines MCLU and the re-weighting procedure presented in [3], 4) the proposed method, and 5) the MCLU-ECBD AL method applied directly to the target image (multispectral VHR data set).

one for classifying the target image. Table III reports *per-class* classification results obtained by the proposed method and the method using the re-weighting heuristic presented in [3] at different iterations (i.e., at the first step and with 100 and 200 target-domain samples included in the training set). In brackets is reported the corresponding average number of source-domain samples in the training set (not removed by the query-). The results are reported in terms of producer accuracy (PA) for all the classes, OA and mean PA. The proposed method resulted in higher PA with respect to the compared method for all the classes (except one) in both considered iterations. It is worth noting the important improvement of the proposed method with respect to the compared one in the classification of the two most critical classes “Agriculture Fields” and “Forest”. Moreover, we observe that the proposed IDA method leads to higher OA compared with the AL method directly applied to QB_2. The advantage given by the proposed method with respect to the standard approach not based on DA is particularly evident for limited amount of target samples included in the training set, where the standard approach cannot either be used or leads to poorer accuracies.

B) Hyperspectral data set: The averaged learning curves obtained by the considered methods on the hyperspectral data set are shown in Fig. 2. Also with this data set, the proposed method resulted in higher classification accuracy with respect to the other considered methods. Table IV reports *per-class* classification results at different iterations, as done for the previous data set. The proposed method lead to higher classification accuracy for most of the classes. Worth noting is the significant improvement in the classification of the most critical class “Hippo Grass”, whose gain in the PA is more than 20% in the case of 50 target-domain samples included in the training set. We observe that the proposed method results in higher OA compared with AL applied to the target domain, when limited amount of labeled target samples are included

TABLE III
CLASSIFICATION RESULTS AT DIFFERENT ITERATIONS (INCLUDING DIFFERENT NUMBER OF TARGET-DOMAIN SAMPLES IN THE TRAINING SET) OBTAINED WITH THE PROPOSED METHOD AND A METHOD THAT COMBINES MCLU AND THE RE-WEIGHTING PROCEDURE PRESENTED IN [3] (MULTISPECTRAL VHR DATA SET).

Class	Number of target and (source)-domain samples				
	0 (965)	100 (849)		200 (837)	
	-	Jun et al.	Prop.	Jun et al.	Prop.
Vineyard	81.4	89.1	90.1	88.1	89.5
Water	100	100	100	100	100
Agriculture Fields	1.5	11.5	15.3	17.2	17.8
Forest	19.9	45.2	52.2	56.6	62.5
Apple Tree	99.4	99.7	99.8	99.9	99.9
Urban Area	100	99.9	99.9	99.9	99.8
OA	76.0	83.6	85.2	85.1	86.7
Mean PA	67.0	74.2	76.2	76.9	78.3

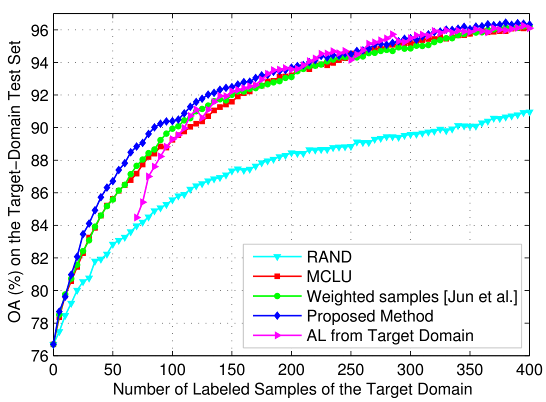

Fig. 2. OA obtained on TS_2 versus the number of samples of U labeled and added to the training set. The different curves correspond to: 1) random selection, 2) the MCLU AL method, 3) a method that combines MCLU and the re-weighting procedure presented in [3], 4) the proposed method, and 5) the MCLU-ECBD AL method applied directly to the target domain (hyperspectral data set).

in the training set. As the number of labeled target samples increases, their effect tends to dominate the one of the source samples, and the gain of the IDA method (given by the use of source-domain information) decreases.

VI. CONCLUSION

In this paper, an IDA method for the classification of RS images has been proposed. The proposed method allows the user to effectively exploit the consistent information of a source image for the classification of a different but related target image. This can result in a significant reduction of the number of new target-domain samples to be labeled, thus reducing the cost associated with the classification of the target image. In operative scenarios when the budget for acquiring new labeled samples is limited, the user may decide to stop the IDA procedure at early iterations as soon as the desired level of accuracy is reached. This allows the user to select among different tradeoff solutions between cost and accuracy of the classification map. The experimental results obtained in the classification of both a multispectral VHR and a hyperspectral image confirm the effectiveness of the proposed technique.

TABLE IV
CLASSIFICATION RESULTS AT DIFFERENT ITERATIONS (INCLUDING DIFFERENT NUMBER OF TARGET-DOMAIN SAMPLES IN THE TRAINING SET) OBTAINED WITH THE PROPOSED METHOD AND A METHOD THAT COMBINES MCLU AND THE RE-WEIGHTING PROCEDURE PRESENTED IN [3] (HYPERSPECTRAL DATA SET).

Class	Number of target and (source)-domain samples				
	0 (739)	50 (722)		150 (706)	
	-	Jun et al.	Prop.	Jun et al.	Prop.
Water	92.6	93.2	96.1	94.4	99.1
Hippo Grass	32.8	54.4	75.0	93.3	95.0
Floodplain Grasses 1	63.7	88.3	89.8	98.5	99.2
Floodplain Grasses 2	95.2	95.7	95.9	96.1	96.3
Reeds	62.4	67.0	69.2	77.8	78.0
Riparian	82.5	80.6	80.8	83.1	83.1
Firescar	98.2	98.2	98.2	99.1	99.5
Island Interior	71.6	95.7	97.8	99.5	100
Acacia Woodlands	64.4	83.3	80.7	93.1	93.3
Acacia Shrublands	90.0	86.1	83.5	85.9	82.0
Acacia Grasslands	68.2	83.1	86.8	91.9	93.4
Short Mopane	100	97.0	95.9	94.8	94.1
Mixed Mopane	63.8	78.6	77.7	90.5	90.5
Exposed Soil	95.0	98.6	100	100	100
OA	76.7	85.6	86.7	92.0	92.5
Mean PA	77.2	85.7	87.7	92.7	93.1

ACKNOWLEDGMENT

The work of Dr. Claudio Persello has been supported by the Autonomous Province of Trento and the European Community in the framework of the project "Trentino - PCOFUND-GA-2008-226070 (call 3 - post-doc 2010 Outgoing)". The author would like to thank Prof. M. Crawford for kindly providing the hyperspectral data set and Dr. F. Dinuzzo for valuable discussions.

REFERENCES

[1] P. Mitra, B. U. Shankar, and S. K. Pal, "Segmentation of multispectral remote sensing images using active support vector machines," *Pattern Recognition Letters*, vol. 25, no. 9, pp. 1067 – 1074, 2004.

[2] S. Rajan, J. Ghosh, and M. Crawford, "An active learning approach to hyperspectral data classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 4, pp. 1231–1242, april 2008.

[3] G. Jun and J. Ghosh, "An efficient active learning algorithm with knowledge transfer for hyperspectral data analysis," in *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2008)*, vol. 1, july 2008, pp. I–52 –I–55.

[4] B. Demir, C. Persello, and L. Bruzzone, "Batch-mode active-learning methods for the interactive classification of remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 3, pp. 1014–1031, March 2011.

[5] J. Huang, A. Gretton, B. Schölkopf, A. J. Smola, and K. M. Borgwardt, "Correcting sample selection bias by unlabeled data," in *Proc. Advances in Neural Information Processing Systems*. MIT Press, 2007.

[6] D. Tuia, E. Pasolli, and W. J. Emery, "Using active learning to adapt remote sensing image classifiers," *Remote Sensing of Environment*, 2011.

[7] W. Dai, Q. Yang, G. Xue, and Y. Yu, "Boosting for transfer learning," in *Proc. International Conference on Machine Learning*, 2007.

[8] C. Persello and L. Bruzzone, "Active learning for domain adaptation in the supervised classification of remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, in press.

[9] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2001.

[10] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 8, pp. 1778 – 1790, aug. 2004.