ENSEMBLE OF TRANSFER COMPONENT ANALYSIS FOR DOMAIN ADAPTATION IN HYPERSPECTRAL REMOTE SENSING IMAGE CLASSIFICATION

Junshi Xia, Naoto Yokoya, Akira Iwasaki

Research Center for Advanced Science and Technology, The University of Tokyo,
153-0041 Tokyo, Japan

ABSTRACT

In this work, we address the problem of unsupervised domain transfer learning via an ensemble strategy in the context of classification between multiple hyperspectral images. The objective of domain adaption is to assign the label to an image of interest (the target image) using the labeled samples in the source image. The proposed method is based on the rotation-based ensemble and transfer component analysis (TCA). In this method, the feature space in both source and target image is divided into several disjoint feature subsets. Then, the features induced by the TCA technique in the source domain are used as the input space to a random forest (RF) classifier. Finally, the results achieved by each step are fused by a majority vote. We compare the proposed method, ensemble of TCA (E-TCA), with a regular RF and an RF with the reduced features by the TCA. Experiments on the real hyperspectral image acquired over a Japanese mixed forest show remarkable cross-image classification performances.

Index Terms— Domain adaptation (DA), transfer component analysis (TCA), ensemble learning, hyperspectral image classification.

1. INTRODUCTION

The quality of reference samples plays an important role in providing high precision classification results of remotely sensed images. However, collecting and labeling sufficient data are very labor-intensive and tedious [1]. An effective technique is the domain adaptation and transfer learning in which we can use the labeled samples from another dataset to enhance the performance. In particular, classifiers trained on the *source* image with the labeled datasets are used to predict the labels of the *target* image [2].

In general, the data in both source and target domains share the same task but with different distributions. According to whether labeled information used or not in the target domain, transfer learning techniques are divided into two groups: unsupervised and supervised [3]. In this work, we focus on unsupervised transfer learning because of the common settings of real-world applications.

In the last decade, researchers have investigated a variety of approaches to domain adaptation into two categories: 1) modifying the trained classifier and 2) making a feature-based representation. Considering the first category, Bruzzone *et al.* [4, 5] re-trained a maximum likelihood classifier or multiple classifier systems by adjusting the parameters to update the land cover maps. The second category aims at learning a common feature representation that minimized the cross-domain difference. Representative work that follows the second strategy is the transfer component analysis (TCA) [6]. Matasci *et al.* [7] investigated the performance of TCA and extended the TCA to the semi-supervised version for cross-domain hyperspectral and high spatial resolution remote sensing images. Experimental results demonstrate the clear improvements over the TCA and semi-supervised TCA (SSTCA) when compared to other methods.

Ensemble learning combines multiple based learners to show a significantly better performance that the individual learner [8]. Benefiting from the superiority, the ensemble learning methods have attracted increasing attention in the remote sensing community over the last several years [9]. In order to make a successful ensemble learning method, we should increase the accuracy of the base learners and the diversity within the ensemble. Such approach was exploited in the rotation-based ensemble for hyperspectral image classification, obtaining remarkable classification results [10, 11].

Inspired by the merits of rotation-based ensemble and TCA, we proposed a novel ensemble of TCA (E-TCA) for domain adaptation in hyperspectral remote sensing image classification.

The remainder is organized as follows. In Section II, we give the main description of TCA. The proposed ensemble of TCA (E-TCA) is presented in Section III. The experiments and the conclusion are given in Section IV and V, respectively.

2. TRANSFER COMPONENT ANALYSIS

Let us denote $\mathcal{D}_S = \{\mathbf{X}_S, \mathbf{Y}_S\} = \{\mathbf{x}_{S_i}, y_{S_i}\}_i^{n_{S_i}}$ be the set of source training samples and $\mathbf{X}_T == \{\mathbf{x}_{T_i}\}_i^{n_{T_i}}$ be the set of unlabeled target samples. The objective of domain adaptation classifier is to assign a label y_{T_i} to each pixel of $\mathbf{X}_T$.

TCA [6] aims at learning a shared subspace between the source and target domains based on a reproducing kernel Hilbert space (RKHS) using maximum mean discrepancy (MMD) [12].

TCA finds a projection matrix $\mathbf{W}$ by solving the following eigenvalue decomposition problem

$$\mathbf{KHKw} = \lambda(\mathbf{KLK} + \mu\mathbf{I})\mathbf{w} \tag{1}$$

where $\mathbf{K} = \begin{pmatrix} \mathbf{K}_{S,S} & \mathbf{K}_{S,T} \\ \mathbf{K}_{T,S} & \mathbf{K}_{T,T} \end{pmatrix}$, $\mathbf{K}_{S,S}, \mathbf{K}_{S,T}, \mathbf{K}_{T,S}, \mathbf{K}_{T,T}$ are the kernel matrices (with the elements of $K_{i,j} = exp(-\theta||\mathbf{x}_i - \mathbf{x}_j||^2)$). $\mathbf{H} = \mathbf{I} - \mathbf{1}\mathbf{1}^\top/(n_S + n_T)$. The elements of matrix $\mathbf{L}$ is defined as

$$L_{i,j} = \begin{cases} 1/n_S^2 & \text{if } \mathbf{x}_i, \mathbf{x}_j \in \mathbf{X}_S \\ 1/n_T^2 & \text{if } \mathbf{x}_i, \mathbf{x}_j \in \mathbf{X}_T \\ -1/n_S n_T & \text{otherwise} \end{cases}$$

The projection matrix $\mathbf{W}$ can be obtained from the d eigenvectors corresponding to the d largest eigenvalues achieved by (1).

The d transfer components for a new test sample $\mathbf{x}_{test}$ is computed by $\mathbf{K}_{test}\mathbf{W}$, where $\mathbf{K}_{test}$ is the kernel matrix between the test sample and the $(n_S + n_T)$ samples.

3. ENSEMBLE OF TCA

The performance of ensemble learning method relies on two essential components: the accuracy of base classifier and the diversity within the ensemble [8]. Rotation-based ensemble is a recent effective strategy to construct the ensemble [10, 11].

In this work, we introduce this strategy into the domain adaption in hyperspectral remote sensing image classification. Thus, we propose the ensemble of TCA (E-TCA), which combines the advantages of TCA and rotation-based ensemble together.

The training steps of the proposed method is summarized as follows (seen in Algorithm 1)

- the feature space in both source and target domain is randomly split into K ($K = D/M$) disjoint subspace and each space contain M features.
- TCA is applied to each subspace to achieve the projection matrix $\mathbf{W}_k$ and kernel matrices.
- the new training set of the source domain is formed by concatenating M extracted features that are obtained by rotating the source training set using the aforementioned kernel matrices and the projection matrix.
- an individual random forest (RF) classifier is trained on the new source training set.
- the above process is repeated T times and the ensemble with T RFs are obtained.

In the prediction phase, the kernel matrices between the target and the sum of source and target training samples is generated firstly. Then, the new transformed dataset of target samples is classified by the ensemble, and the final result is assigned to the corresponding class by using a *majority voting* rule.

Algorithm 1 E-TCA

Training phase

Input: $\{\mathbf{X}_S, \mathbf{Y}_S\} = \{\mathbf{x}_{S_i}, y_{S_i}\}_i^{n_{S_i}}$: source training samples. $\mathbf{X}_T$: target samples. T: number of classifiers, K: number of subsets, M: number of features in a subset, L: base classifier. The ensemble $\mathcal{L} = \varnothing$.

Output: The ensemble $\mathcal{L}$

for $i = 1 : T$ **do**
  randomly split the features in the source and target domains into K subsets.
  for $j = 1 : K$ **do**
    form the new source training set $\mathbf{X}_{S_{i,j}}$ and target samples $\mathbf{X}_{T_{i,j}}$ corresponding to the subset of features
    using TCA to transform $[\mathbf{X}_{S_{i,j}} \mathbf{X}_{T_{i,j}}]$ with the aim of obtaining the coefficients $\mathbf{R}_{i,j} = [\mathbf{w}_{i,j}^1, \cdots, \mathbf{w}_{i,j}^M]$
    calculate the kernel matrices by $\mathbf{X}_{S_{i,j}}$, $\mathbf{Ktrain}_{i,j} = \mathbf{K}(\mathbf{X}_{S_{i,j}}, [\mathbf{X}_{S_{i,j}} \mathbf{X}_{T_{i,j}}])$
  end for
  the features extracted will be given by: $\mathbb{F}_i^{new} = \left[\mathbf{Ktrain}_{i,1}^\top \mathbf{R}_{i,1}, \cdots, \mathbf{Ktrain}_{i,K}^\top \mathbf{R}_{i,K}\right]$
  train a RF classifier L_i using $\{\mathbb{F}_i^{new}, \mathbf{Y}\}$
  add the classifier to the current ensemble, $\mathcal{L} = \mathcal{L} \cup L_i$.
end for

Prediction phase

Input: The ensemble $\mathcal{L} = \{L_i\}_i^T$. $\mathbf{X}_T$: target samples. Rotation matrix: $\mathbf{R}$.

Output: class label $\mathbf{Y}_T$

for $i = 1 : T$ **do**
  for $j = 1 : K$ **do**
    generate the kernel matrices $\mathbf{Ktest}_{i,j} = \mathbf{K}(\mathbf{X}_{T_{i,j}}, [\mathbf{X}_{S_{i,j}} \mathbf{X}_{T_{i,j}}])$
    generate the test features of target samples, $\mathbb{F}_i^{test} = [\mathbf{Ktest}_{i,1}^\top \mathbf{R}_{i,1}, \cdots, \mathbf{Ktest}_{i,k}^\top \mathbf{R}_{i,K}]$
  end for
  run the classifier L_i using $\mathbb{F}_i^{test}$ as input
end for
calculate the confidence of each sample $\mathbf{x}$ in $\mathbf{X}_T$ for each class and assign the class label $p(y|\mathbf{x}) = \frac{1}{T}\sum_{i=1}^T p(y_i|\mathbb{F}_i^{test})$ to the class with the largest confidence.

4. DATASETS AND RESULTS

A Japanese mixed forest, namely Tama Forest Science Garden located in the western region of Tokyo, is selected as the

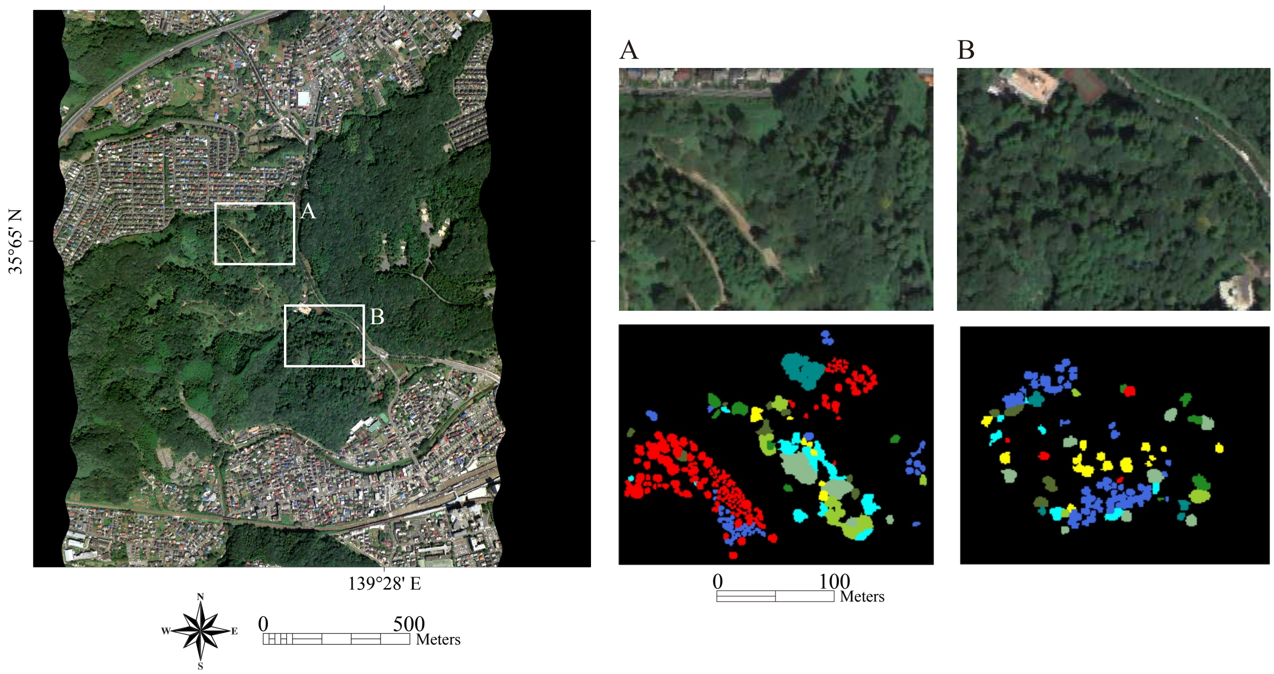
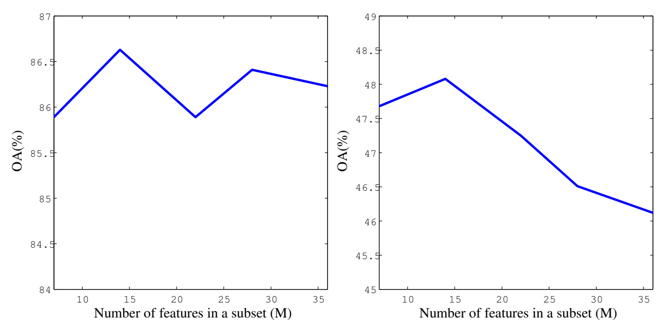


Fig. 1. (Left) The false color composite image of hyperspectral data. (Right) Enlarged RGB images of (A) Source and (B) target images with the corresponding ground references

Table 1. Class name and number of reference samples in the source and target images.

	No.	Name	Source	Target
Conifer	1	Japanese cypress	875	2120
	2	Japanese cedar	3327	165
	3	Momi fir	282	813
Broadleaf	4	American sweetgum	733	286
	5	Japanese bigleaf magnolia	380	382
	6	Painted maple	403	352
	7	Chinese evergreen oak	948	256
	8	Japanese blue oak	1083	422
	9	Konara oak	987	768



(a) (b)

Fig. 2. RF classification performance on the target image by using the E-TCA with different numbers of features in a subset (M). (a) Two major species. (b) 9 classes.

Table 2. Classification results with two major species.

Name	RF	TCA	E-TCA
Conifer	84.25	87.12	**88.77**
Broadleaf	83.86	82.16	**88.57**
OA	84.08	85.19	**86.63**
AA	84.05	84.87	**86.38**

study area. The hyperspectral dataset with 72 bands was acquired by the CASI-3 sensor. The ground sampling distance is 1 m [13]. The left image of Fig. 1 is the false color composite image obtained by the hyperspetral image (R: 654 nm; G: 552 nm; B: 449 nm). The right image of Fig. 1 is the enlarged RGB images of source (A) and target (B) areas and the respective distributions of the reference with two major species (nine classes).

For the TCA and E-TCA, θ is set to be the median Euclidean distance among all used samples. The label samples in both source and target images are used to find the projection achieved by TCA. Random Forest (RF) classifier is used to train on the labeled samples in the source domain and predict the labels in the target image. For the RF, the number of trees is set to be 50 and the number of features in a subset is set to be the square root of the number of the used features. For the TCA, the range of extracted components is from 2 to 30 and we only report the best results.

Fig. 2 depicts the influence of the number of features in

Table 3. Classification results with 9 classes.

No.	RF	TCA	E-TCA
1	61.56	73.54	**79.25**
2	48.48	90.91	**92.12**
3	7.38	11.44	**12.14**
4	36.71	36.36	**48.25**
5	31.94	39.01	**39.79**
6	7.10	6.25	**10.23**
7	**31.25**	28.52	29.69
8	17.54	**25.12**	24.88
9	30.60	28.52	**30.86**
OA	37.49	44.48	**48.08**
AA	30.28	37.74	**40.80**

a subset (M) on the RF classification performance for two major species and nine classes, respectively. The optimum classification accuracy is reached at the small values of M. Thus, $M = 15$ is used to present the class individual accuracies, which is shown in Tables 2 and 3. It can be seen that the proposed E-TCA shows better performance than RF and TCA in both global accuracies (OA and AA) and class-specific accuracies (7 out of 9). For two major species, E-TCA achieves 2.55% and 1.44% improvements over RF and TCA, respectively. For nine classes, the improvements are 10.59% and 3.60%, respectively.

5. CONCLUSION

In this paper, we propose a novel transfer learning method for hyperspectral image classification. We cast the transfer learning problem as an ensemble strategy with transfer component analysis (TCA). Experimental results show that our method outperforms RF and TCA approaches. In the future, we will extend our method to semi-supervised scenarios.

6. ACKNOWLEDGMENT

The authors would like to that KAKENHI Grant Number 24360347 and 16H04587, and JSPS KAKENHI Grant Number 16F16053 for supporting this work.

7. REFERENCES

[1] J. Xia, J. Chanussot, P. Du, and X. He, "(Semi-) supervised probabilistic principal component analysis for hyperspectral remote sensing image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2224–2236, June 2014.

[2] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct 2010.

[3] Ming Shao, Dmitry Kit, and Yun Fu, "Generalized transfer subspace learning through low-rank constraint," *International Journal of Computer Vision*, vol. 109, no. 1, pp. 74–93, 2014.

[4] L. Bruzzone and D. Fernández-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, 2001.

[5] L. Bruzzone and R. Cossu, "A multiple-cascade-classifier system for a robust and partially unsupervised updating of land-cover maps," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 9, pp. 1984–1996, 2002.

[6] 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.

[7] G. Matasci, M. Volpi, M. Kanevski, L. Bruzzone, and D. Tuia, "Semisupervised transfer component analysis for domain adaptation in remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, July 2015.

[8] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*, Wiley-Interscience, 2004.

[9] P. Du, J. Xia, W. Zhang, K. Tan, Y. Liu, and S. Liu, "Multiple classifier system for remote sensing image classification: A review," *Sensors*, vol. 12, no. 4, pp. 4764–4792, 2012.

[10] J. Xia, J. Chanussot, P. Du, and X. He, "Rotation-Based Ensemble Classifiers for High-Dimensional Data," in *Fusion in Computer Vision*, Bogdan Ionescu, Jenny Benois-Pineau, Tomas Piatrik, and Georges Quénot, Eds., pp. 135–160. Springer, 2014.

[11] J. Xia, J. Chanussot, P. Du, and X. He, "Spectral-spatial classification for hyperspectral data using rotation forests with local feature extraction and Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2532–2546, 2015.

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

[13] T. Matsuki, N. Yokoya, and A. Iwasaki, "Hyperspectral tree species classification of Japanese complex mixed forest with the aid of LiDAR data,," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 5, pp. 2177–2187, May 2015.