Spectral-spatial rotation forest for hyperspectral image classification

Junshi Xia, Lionel Bombrun, Yannick Berthoumieu, Christian Germain, Peijun Du

SPECTRAL-SPATIAL ROTATION FOREST FOR HYPERSPECTRAL IMAGE CLASSIFICATION

Junshi Xia [1]*, Lionel Bombrun* [1]*, Yannick Berthoumieu* [1]*, Christian Germain* [1] *and Peijun Du*[2]

[1] Université de Bordeaux and CNRS, IMS, UMR 5218, F-33405 Talence, France

[2] Key Laboratory for Satellite Mapping Technology and Applications of State Administration of Surveying, Mapping and Geoinformation of China, Nanjing University, 210093 Nanjing, China

ABSTRACT

Rotation Forest (RoF) is a decision tree ensemble classifier, which uses random feature selection and data transformation techniques to improve both the diversity and accuracy of base classifiers. Traditional RoF only considers data transformation on spectral information. In order to further improve the performance of RoF, we introduce spectral-spatial data transformation into RoF and thus propose a spectral-spatial Rotation Forest (SSRoF). The proposed method is experimentally investigated on a hyperspectral remote sensing image collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor. Experimental results indicate that the proposed methodology achieves excellent performance.

Index Terms— Rotation Forest, Spectral-spatial, Classification, Hyperspectral

1. INTRODUCTION

Multiple classifier systems (MCSs) or classifier ensemble, which produce the final output based on the decisions made by a set of individual classifiers according to certain rules, have been a hot topic for image classification in the hyperspectral remote sensing community [1, 2]. This is because a set of classifiers provide complementary and diverse information, thus enhancing the classification performance [3].

Rotation forest (RoF) is one of the current state-of-the-art decision tree ensemble classifier [4]. RoF is an extension of random forest (RF) classifier. In contrast to RF, RoF first splits the features into several disjoint subsets and apply data transformation to each subset. Second, new training set for the decision tree (DT) is formed by concatenating the linear extracted features contained in each subset. Thus, RoF enhances both accuracy and diversity within the ensemble [4].

Studies on the use of RoF dealing with hyperspectral classification problems have been recently published [5–9]. RoF has proven to be effective not only for hyperspectral data analysis, but also for very high spatial resolution and SAR images [10,11]. Although RoF obtains remarkable performance, data transformation in RoF is performed only by measuring the similarity between the samples using spectral-domain Euclidean distance [12]. However, this is insufficient to reveal the intrinsic the relationship between different samples [13]. Therefore, the spatial correlations should be considered in measuring the sample similarity [13].

In order to further improve the performance of RoF ensemble, we introduce spectral-spatial data transformation into RoF and thus proposed a spectral-spatial RoF (SSRoF) as a new classifier. We expect that SSRoF improves the performance of RoF by introducing further diversity by performing a spectral-spatial data transformation. The experimental analysis, including a comparison with RoF based on spectral or spatial information, is carried out on the Indian Pines test site.

The reminder of this paper is organized as follows. Section 2 introduces spectral-spatial data transformation. The proposed SSRoF is described in Section 3. Section 4 presents the experimental results. Conclusions are drawn in Section 5.

2. SPECTRAL-SPATIAL DATA TRANSFORMATION

Let us denote $\{\mathbf{X}, \mathbf{Y}\} = \{(\mathbf{x}_1, y_1), ..., (\mathbf{x}_n, y_n)\}$ as the training samples, where $\mathbf{x}_i \in \mathbb{R}^D$ is a pixel and y_i is a scalar with classes of interest $\mathcal{C} = \{1, ..., C\}$, where C is the total number of classes. For data transformation, we often assume that there exists a mapping function $f : \mathbb{R}^D \rightarrow \mathbb{R}^d, d \leq D$, which can transform each data point $\mathbf{x}_i$ to $\mathbf{z}_i = f(\mathbf{x}_i)$. This mapping is always represented by a $D \times d$ matrix $\mathbf{V}$:

$$\mathbf{z}_i = f(\mathbf{x}_i) = \mathbf{V}^\top \mathbf{x}_i \tag{1}$$

For many feature extraction methods, the projection matrix $\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, ...\mathbf{v}_d)$ are obtained as the d eigenvectors corresponding to the d largest eigenvalues $\{\lambda_1, \lambda_2, ..., \lambda_d\}$, by solving the following eigenvalue decomposition equation:

$$\mathbf{S}_1\mathbf{v} = \lambda\mathbf{S}_2\mathbf{v} \tag{2}$$

where, $\mathbf{S}_1$ and $\mathbf{S}_2$ are matrices which depend on the data transformation approach.

2.1. Spectral-based data transformation

In this work, local Fisher discriminant analysis (LFDA) is used to extract the spectral-domain local similarity. LFDA effectively combines the ideas of Fisher discriminant analysis (FDA) and locality-preserving projection (LPP). Hence, LFDA maximizes the between-class separability and preserves the with-class local structure [12]. Practically, LFDA is obtained by solving the following eigenvalue decomposition equation:

$$\mathbf{S}^{lb}\mathbf{v} = \lambda \mathbf{S}^{lw}\mathbf{v} \tag{3}$$

where $\mathbf{S}_{lb}$ and $\mathbf{S}_{lw}$ denote respectively the local between-class and the within-class scatter matrix.

$$\mathbf{S}^{lb} = \frac{1}{2}\sum_{i,j=1}^{n} \omega_{i,j}^{lb}(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^\top \tag{4}$$

$$\mathbf{S}^{lw} = \frac{1}{2}\sum_{i,j=1}^{n} \omega_{i,j}^{lw}(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^\top \tag{5}$$

$$\omega_{i,j}^{lb} = \begin{cases} A_{i,j}(\frac{1}{n} - \frac{1}{n_{y_i}}) & \text{if } y_i = y_j \\ \frac{1}{n} & \text{otherwise} \end{cases} \tag{6}$$

$$\omega_{i,j}^{lw} = \begin{cases} \frac{A_{i,j}}{n_{y_i}} & \text{if } y_i = y_j \\ \frac{1}{n} & \text{otherwise} \end{cases} \tag{7}$$

$$A_{i,j} = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{\sigma_i \sigma_j}\right) \tag{8}$$

$$\sigma_i = \left\|\mathbf{x}_i - \mathbf{x}_i^k\right\| \tag{9}$$

where, $\mathbf{x}_i^k$ is the k-th nearest neighbor of $\mathbf{x}_i$ (k is set to 7). n_{y_i} is the number of labeled samples in class $y_i \in \mathcal{C}$.

2.2. Spatial-based data transformation

In this subsection, spatial-based data transformation used in [13] is presented. The neighboring pixels in a spatial local homogeneous region belong to the same class [13]. Under this situation, the spatial information is used to learn the projections. Assume a training pixel $\mathbf{x}_i$ with its spatial neighbors in $\mathcal{N}(\mathbf{x}_i)$ form a local pixel patch: $\{\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{im}\}$. The local pixel neighborhood preserving matrix is defined as

$$\mathbf{H} = \sum_{i=1}^{n}\sum_{j=1}^{m} \frac{\mu_j}{\sum_{k=1}^{m}\mu_k}\left(\mathbf{x}_i - \mathbf{x}_{ij}\right)\left(\mathbf{x}_i - \mathbf{x}_{ij}\right)^\top \tag{10}$$

where $\mu_k = \exp(-\tau\left\|\mathbf{x}_i - \mathbf{x}_{ik}\right\|^2)$ is the spectral similarity between the neighboring pixels to the central pixel (τ should be tuned by the user). The total scatter matrix is defined as:

$$\mathbf{S} = \sum_{i=1}^{n}\left(\mathbf{x}_i - \mathbf{X}_{mean}\right)\left(\mathbf{x}_i - \mathbf{X}_{mean}\right)^\top \tag{11}$$

where $\mathbf{X}_{mean}$ is the mean of $\mathbf{X}$. Thus, the projection of spatial-domain data transformation can be obtained by solving the following eigenvalue problem:

$$\mathbf{S}\mathbf{v} = \lambda \mathbf{H}\mathbf{v} \tag{12}$$

2.3. Spectral-spatial data transformation

In order to take account of spectral and spatial information, spectral-spatial data transformation method is introduced into RoF. It preserves not only the spectral-domain local Euclidean neighborhood class relations but also the spatial-domain local pixel neighborhood structures. Finally, the projection is achieved by solving the following eigenvalue problem:

$$\left(\alpha\mathbf{S}^{lb} + (1-\alpha)\mathbf{S}\right)\mathbf{v} = \lambda\left(\alpha\mathbf{S}^{lw} + (1-\alpha)\mathbf{H}\right)\mathbf{v} \tag{13}$$

where α is the control parameter of spectral and spatial information.

Algorithm 1 SSRoF

Training phase

Input: $\{\mathbf{X}, \mathbf{Y}\} = \{\mathbf{x}_i, y_i\}_{i=1}^n$: training samples, T: number of classifiers, K: number of subsets (M: number of features in each subset), L: base classifier. The ensemble $\mathcal{L} = \varnothing$. $\mathbb{F} \in \mathbb{R}^D$: Feature set

Output: The ensemble $\mathcal{L}$

1: **for** $i = 1 : T$ **do**
2: Randomly split the features $\mathbb{F}$ into K subsets $\mathbb{F}_j^i$
3: **for** $j = 1 : K$ **do**
4: Extract from $\mathbf{X}$ the new training set $\mathbf{X}_{i,j}$ with the corresponding features $\mathbb{F}_j^i$
5: Transform $\mathbf{X}_{i,j}$ by (13) to get the coefficients $v_{i,j}^{(1)}, ..., v_{i,j}^{(M_k)}$
6: **end for**
7: Sparse rotation matrix $\mathbf{R}_i$ is composed of the above coefficients

$$\mathbf{R}_i = \begin{bmatrix} v_{i,1}^{(1)}, ..., v_{i,1}^{(M_1)} & 0 & \cdots & 0 \\ 0 & v_{i,2}^{(1)}, ..., v_{i,2}^{(M_2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & v_{i,j}^{(1)}, ..., v_{i,j}^{(M_K)} \end{bmatrix}$$

8: Rearrange $\mathbf{R}_i$ to $\mathbf{R}_i^a$ with respect to the original feature set,
9: Obtain the new training samples $\left\{\mathbf{X}^\top\mathbf{R}_i^a, \mathbf{Y}\right\}$
10: Build the DT classifier L_i using $\left\{\mathbf{X}^\top\mathbf{R}_i^a, \mathbf{Y}\right\}$
11: Add the classifier to the current ensemble, $\mathcal{L} = \mathcal{L} \cup L_i$.
12: **end for**

Prediction phase

Input: The ensemble $\mathcal{L} = \{L_i\}_i^T$. A new sample $\mathbf{x}_*$. Rotation matrix: $\mathbf{R}_i^a$.

Output: class label y_*

1: get the output ensemble with $\mathbf{x}_*^\top\mathbf{R}_i^a$.
2: the label is assigned to the class with maximum number of votes. $y_* = \underset{i\in\{1,2,...,C\}}{\operatorname{argmax}} \sum_{j:L_j(\mathbf{x}_*^\top\mathbf{R}_i^a)=i}^{T} 1$

Table 1. Overall, average, κ and class-specific accuracies obtained for the Indian Pines AVIRIS image

Class	Train	Test	No WMF				WMF			
			RF	SpeRoF	SpaRoF	SSRoF	RF	SpeRoF	SpaRoF	SSRoF
Alfalfa	10	44	74.32	**84.77**	83.64	82.50	82.73	84.32	84.77	**86.36**
Corn-no till	10	1514	28.80	50.48	56.26	**61.00**	48.17	59.02	61.41	**64.96**
Corn-min till	10	824	34.15	45.55	47.39	**51.84**	**72.16**	70.16	70.67	70.76
Bldg-Grass-Tree-Drives	10	224	43.16	62.90	63.88	**64.42**	63.88	71.29	71.43	**73.44**
Grass/pasture	10	487	64.76	73.86	74.31	**75.07**	83.37	88.60	89.34	**89.96**
Grass/trees	10	737	59.18	80.56	82.08	**83.47**	91.32	**95.41**	94.97	94.41
Grass/pasture-mowed	10	16	86.88	**91.88**	91.25	91.25	96.88	**98.13**	96.88	97.50
Corn	10	479	71.96	79.54	84.32	**84.45**	91.23	93.61	94.20	**94.72**
Oats	10	10	90.00	**97.00**	**97.00**	96.00	**100**	**100**	**100**	99.00
Soybeans-no till	10	958	42.48	62.21	64.67	**66.45**	63.56	74.02	**74.37**	73.28
Soybeans-min till	10	2458	39.99	43.93	43.83	**45.72**	59.32	54.69	54.49	**61.22**
Soybeans-clean till	10	604	27.84	45.81	47.45	**48.18**	63.92	61.99	62.75	**63.97**
Wheat	10	202	89.55	95.69	96.14	**96.49**	94.85	95.94	**97.77**	**97.77**
Woods	10	1284	69.87	71.83	72.01	**72.73**	86.43	87.63	87.13	**88.50**
Hay-windrowed	10	370	36.22	46.51	**49.27**	48.24	**78.35**	77.81	73.76	76.95
Stone-steel towers	10	85	92.47	96.00	96.59	**97.06**	**97.76**	97.52	**97.76**	**97.76**
OA			46.92	59.80	60.10	**62.33**	69.85	72.02	72.20	**74.65**
AA			59.48	71.76	71.80	**72.87**	79.62	81.92	81.98	**83.16**
κ			40.88	55.18	55.50	**57.59**	66.17	68.74	68.94	**71.51**

3. SPECTRAL-SPATIAL ROTATION FOREST

Spectral-spatial Rotation Forest (SSRoF) is a variant of RoF, which uses spectral-spatial data transformation. The main training and prediction steps are presented in Algorithm 1.

In the training phase, the feature space is firstly divided into K disjoint subsets. Spectral-spatial data transformation (13) is performed on each subset. A transformed training set is generated by rotating with a sparse matrix $\mathbf{R}_i^a$ the original training set. An individual DT classifier is trained on this rotated training set.

In the prediction phase, a new sample $\mathbf{x}^*$ is rotated by $\mathbf{R}_i^a$. Then, the transformed set, i.e., $\mathbf{x}_*^\top \mathbf{R}_i^a$, is classified by the ensemble and the class with the maximum number of votes is chosen as the final class.

SSRoF can be reduced to spectral-based RoF (SpeRoF) or spatial-based RoF (SpaRoF) by setting respectively $\alpha = 1$ and $\alpha = 0$ in (13).

4. EXPERIMENTAL RESULTS

In this section, the proposed approach is evaluated using real hyperspectral data, which is recorded by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over the Indian Pines in Northwestern Indiana, USA. This scene, which comprises 220 spectral bands in the wavelength range from 0.4 to 2.5 μm with spectral resolution 10 nm, is composed of 145 $\times$ 145 pixels, and the spatial resolution is 20 m/pixel (seen in **Fig**.1). This dataset has 16 classes of interest.

(a) (b)

Fig. 1. (a) Three-band color composite of AVIRIS image. (b) Ground truth.

In this experiment, we randomly select 10 samples per class to form the training set and the rest of the pixels are used for testing. Number of classifiers (T), number of features in a subset (M) and α are set to be 20, 110 and 0.5 respectively. The results are obtained after 10 Monte Carlo runs. The proposed SSRoF is compared to the RF, SpeRoF and SpaRoF. The new features smoothed by the fast weight median filter (WMF) [14] are also used to evaluate the proposed method. The window size is set to be 5×5.

Table 1 gives the overall, average and class-specific accuracies obtained for RF, SpeRoF, SpaRoF and SSRoF using only 10 samples per class when applied to the spectral information and the features smoothed by WMF. From this table, it is clear that SSRoF provides the best results in terms of global and individual class accuracies. **Fig**. 2 shows the visualization of the thematic maps generated by the classification methods (with WMF).

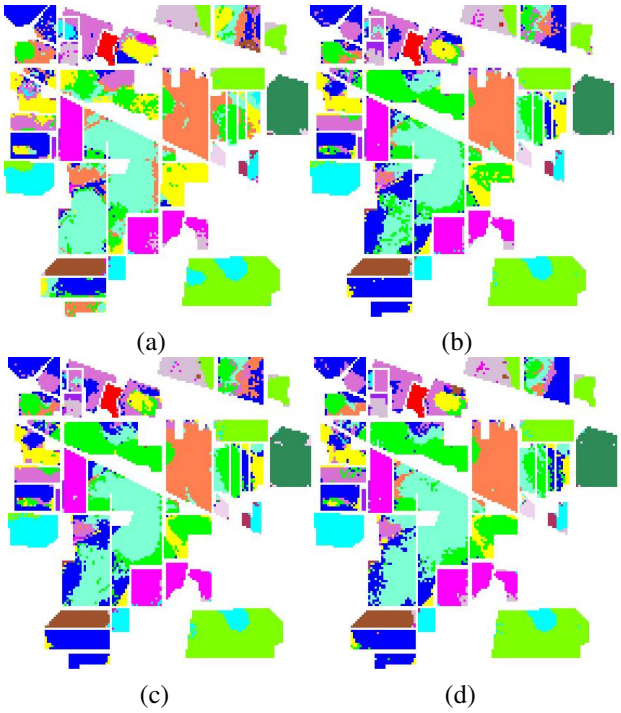

Fig. 2. Classification results of Indiana Pines AVIRIS image (with WMF). (a) RF, OA = 68.61%. (b) SpeRoF, OA = 70.39%. (c) SpaRoF, OA = 72.36%. (d) SSRoF, OA = 74.03%.

T, M and α are known as the important parameters in the construction of SSRoF. The sensitivity analysis of T and M can be found in our previous studies [6, 7, 9]. The additional experiment indicates that the method we propose is not sensitive to α. The user may consider a value for alpha between 0.1 and 0.9.

5. CONCLUSION

This paper presents a novel spectral-spatial rotation forest (SSRoF) for hyperspectral image classification. It has been tested and compared to the RF, spectral- and spatial-based RoF on the well-known Indian Pines AVIRIS hyperspectral image. Experimental results demonstrate the excellent performance of the proposed SSRoF, which captures both spectral and spatial information in the construction of ensemble.

6. ACKNOWLEDGMENT

The authors would like to thank Prof D. Landgrebe from Purdue University for providing the data set and Prof. Peng for sharing the codes in [13]. This study has been carried out with financial support from the French State, managed by the French National Research Agency (ANR) in the frame of the "Investments for the future" Programme IdEx Bordeaux-CPU (ANR-10-IDEX-03-02).

7. REFERENCES

[1] J. A. Benediktsson, J. Chanussot, and M. Fauvel, "Multiple classifier systems in remote sensing: from basics to recent developments," in *Proceedings of the 7th International Workshop on Multiple Classifier Systems, Prague, Czech Republic, May 23-25*, 2007, pp. 501–512.

[2] 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.

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

[4] J. J. Rodriguez, L.I. Kuncheva, and C.J. Alonso, "Rotation forest: A new classifier ensemble method.," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 28, no. 10, pp. 1619–1630, 2006.

[5] J. Xia, P. Du, X. He, and J. Chanussot, "Hyperspectral remote sensing image classification based on rotation forest," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 239 – 243, 2014.

[6] 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.

[7] 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.

[8] J. Xia, J. Chanussot, P. Du, and X. He, "Rotation-based support vector machines in classification of hyperspectral data with limited training samples," *IEEE Trans. Geosci. Remote Sens.*, in press.

[9] J. Xia, M. Dalla Mura, J. Chanussot, P. Du, and X. He, "Random subspace ensembles for hyperspectral image classification with extended morphological attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 4768–4786, 2015.

[10] T. Kavzoglu, I. Colkesen, and T. Yomralioglu, "Object-based classification with rotation forest ensemble learning algorithm using very-high-resolution worldview-2 image," *Remote Sens. Letters*, vol. 6, no. 11, pp. 834–843, 2015.

[11] P. Du, A. Samat, B. Waske, S. Liu, and Z. Li, "Random forest and rotation forest for fully polarized sar image classification using polarimetric and spatial features," *ISPRS J. Photogramm Remote. Sens.*, vol. 105, pp. 38 – 53, 2015.

[12] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis," *J. Mach. Learn. Res.*, vol. 27, no. 8, pp. 1021–1064, 2007.

[13] Y. Zhou, J. Peng, and C.L.P. Chen, "Dimension reduction using spatial and spectral regularized local discriminant embedding for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 1082–1095, 2015.

[14] Q. Zhang, L. Xu, and J. Jia, "100+ times faster weighted median filter (WMF)," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, Ohio, USA, June 24-27*, 2014.