

SEMI-SUPERVISED SPARSE RELEARNING REPRESENTATION CLASSIFICATION FOR HIGH-RESOLUTION REMOTE SENSING IMAGERY

Jiayi Li, Xin Huang, and Liangpei Zhang*

The State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, and the School of Remote Sensing and Information Engineering, Wuhan University, P. R. China
Corresponding author: zlp62@whu.edu.cn

ABSTRACT

In this article, we proposed a novel semi-supervised sparse representation classification for high resolution remote sensing image. First, collaborative representation mechanism that exploits the help from whole training information rather than from only the potential associated class can enhance the class recognition ability. Second, by taking advantage of spatial occurrence and alignment of class label, the adoption of the relearning can gradually learn the flexible class-oriented spatial pattern from the label space with alleviated computational complexity to enhance the original spectral characteristics. Third, inspired by the spatial smoothing phenomenon when spatial feature stacked, a novel stable self-learning method can be designed to automatically select informative unlabeled sample to help the limited supervised set. Experiments on two hyperspectral and high-spatial resolution images validated the effectiveness and robustness of the proposed algorithm.

Index Terms—Sparse representation, relearning, self-learning, spatial co-occurrence pattern, remote sensing image classification

1. INTRODUCTION

Recent high resolution remote sensing imagery (HRRS), known as its subtle spectral characteristics that spanning the visible to infrared spectrum with hundreds of contiguous and narrow spectral bands and the abundant spatial structure, pattern description, allows for the detailed analysis of the surface of the Earth. As a general framework of classification consists of feature acquisition and classifier designing, there are still some obstacles for the classification of pixels in HRRS image.

First of all, despite such fine discriminative spatial information, constructing axillary spatial feature from such high dimensional signal should inevitably get stuck in the computational burden problem. Meanwhile, current classification methods [1], [2] are still a difficult task due to the unbalance between the high dimensionality of the data and the limited availability of labeled training samples in real analysis scenarios. In addition, the classification methods originally developed for the labeling of low-dimensional datasets, i.e. multispectral images, generally perform poorly when applied to HRRS, particularly in the case of limited training samples. Furthermore, the unstable spectral signature of the classes in the complex spatial domain of the scene can lead to an incomplete

description of the different ground-object classes. Current studies focusing on dealing with limited supervised prior obstacle can be categorized as dimensionality reduction, sparse learning, active learning, semi-supervised learning (SSL) [6], [7], transfer learning, kernel methods and the spatial smoothing [3].

In this article, the proposed algorithm proposed a novel semi-supervised framework to combine the superiorities of these techniques. First, in view of the given training information, collaborative representation (CR) mechanism [4], [5] that exploits the assistance from whole training information rather than from only the potential associated class can improve the recognition ability. Second, by taking advantage of spatial occurrence and alignment of class label, the adoption of the relearning [8] can gradually learn the flexible class-oriented spatial pattern from the label space with alleviated computational complexity to enhance the original spectral characteristics. Finally, inspired by the spatial smoothing phenomenon when spatial feature stacked, a novel stable self-learning method can be designed to automatically select informative unlabeled sample to help the limited supervised set. Details about the proposed work can be seen in Section 2, and experiments in Section 3 should indicate the effectiveness of the proposed framework, followed by some conclusion word in last section.

2. PROPOSED FRAMEWORK

2.1. CR: Posterior Probability via a Sparse Transform

In the sparse representation model, we stack the given N_i training pixels from the i th class of a dictionary \mathbf{A}_i , then the overcomplete dictionary \mathbf{A} is constructed by combining all the sub-dictionaries $\{\mathbf{A}_i\}_{i=1,\dots,M}$. In this way, the pixel s_j which belongs to the i th class can be sparsely represented as a linear combination of all the given training samples:

$$s_j = \underbrace{[\mathbf{A}_1 \dots \mathbf{A}_i \dots \mathbf{A}_M]}_{\mathbf{A}} \underbrace{[\beta_1 \dots \beta_i \dots \beta_M]}_{\beta}^T + \xi_j \quad (1)$$

where β is a sparse coefficient vector, in which only the entries of β_i are assumed to be nonzero, and ξ_j is the random noise. The coefficient vector β can be obtained by solving the following optimization problem:

$$\beta = \arg \min \|\mathbf{A}\beta - s\|_2^2 \quad \text{s.t. } \|\beta\|_0 \leq N_i \quad (2)$$

There are two effective ways of solving such a problem: the greedy pursuit based algorithms and the ℓ_1 -norm convex relaxation algorithms.

After obtaining the sparse coefficient vector β , the posterior probability can be defined with respect to the residuals associated with each label. In other words, the label with least residual $\min r_{i,j} = \min \|s_j - \mathbf{A}\delta_i(\beta)\|_2$ has the most probability to be labeled than other classes, as the principle of the maximum posterior probability. It is likely to point out that the posterior probability $p(y_{i,j} | s_j)$ is inversely proportional to the $r_{i,j}$, i.e.,

$$p(y_{i,j} | s_j) = \frac{1}{r_{i,j} \chi} \quad (3)$$

where $y_{i,j}$ refers to class i for the pixel s_j , and $\chi = \sum_m r_{m,j}$.

More detailed can be seen in [9].

2.2. Relearning: primitive co-occurrence matrix (PCM)

The relearning is referred as a label refinement in the iteration processing of sequential categorizing, which can gradually infuse the specific spatial co-occurrence to enhance the description ability of current feature. The PCM, inspired by the well-known spatial texture descriptor GLCM [10], records the spatial arrangement of the current class label for relearning, that is, next classification iteration. Analogous to the GLCM, the calculation of PCM can be described as four steps in briefly.

1. Quantization: hard labeling whole scene into M class based on the maximum posterior probability (*MAP*) mentioned above.
2. Co-occurrence statistic: Counting the number of times that the class i and j occur with distance dis (set as 1 by default and simplicity) and direction dir at the coordinate (i, j) of a $M \times M$ matrix.
3. Multi-direction summation:

$$pcm(w, dis) = \sum_{dir} pcm(w, dis, dir) \quad (4)$$

4. Multi-scale assembling:

$$PCM = \sum_w pcm(w, dis) \quad (5)$$

where w stands for size of window, and more detailed can be seen in [8].

2.3. SSL: self-learning to select informative unlabeled samples

The main assumption of SSL techniques is that the new (unlabeled) training samples [6] can be obtained from a (limited) set of available labeled samples without significant effort/cost. For remote sensing image classification, it is well-accepted that the most confusing and informative pixels are likely located at the edge of a land parcel. Meanwhile, it is noted that the major superiority of PCM without direction prior information refers to the spatial smoothing for the inner part of a land parcel, while the edge pixels should get stuck in the over-smoothing problem.

In view of this, the self-learning that iteratively selects the unlabeled samples on the edge can be implemented as follows. After $k+1$ iteration, we first take regard the unlabeled pixels tested as same class from $k-1$ th and k th iteration, which were implemented by stacking associated PCM features with original spectral feature into classifier as candidate informative ones, and followed with the judgement as:

IF $Thr < Map_k(s_p) \leq Map_{k-1}(s_p)$

THEN s_p should be selected

where Map_k and Map_{k-1} stands for the *Map* for pixel s_p at k th and $k-1$ th classification iteration, respectively, and Thr refers to the lower bound to ensure the correction labeling for this pixel.

In this proposed self-learning strategy, the *Map* decline indicates the information gain which cannot be fully represented by current training information, while the consisted label ensures the fidelity of the unlabeled samples.

3. EXPERIMENTS AND ANALYSIS

3.1. Data Sets

The first image was an airborne hyperspectral data flight line over the Washington DC Mall, which was acquired by the Hyperspectral Digital Image Collection Experiment (HYDICE) sensor, and was provided with the permission of the Spectral Information Technology Application Center of Virginia, who was responsible for its collection. The sensor system used in this case measured the pixel response in 210 bands in the 0.4 to 2.4 μm region of the visible and infrared spectrum. Bands in the 0.9 and 1.4 μm regions where the atmosphere is opaque have been omitted from the dataset, leaving 191 bands. As seen in Fig. 1(a), this image consists of 1280×307 pixels, with 19332 labeled pixels for model validation. This image contains seven reference classes, as shown in Table I. This dataset is challenging due to its complicated spatial distribution and some similar spectral response between several land-covers, such as water-shadow, trees-grass, and roofs-trails-roads.

The second scene was acquired by the Reflective Optics Systems Imaging Spectrometer (ROSIS) sensor over Pavia University, northern Italy. It consists of 610×610 pixels and 115 spectral reflectance bands. We selected 103 of the bands and cut a patch sized 610×340 . The false-color composite of the Pavia University image is shown in Fig. 2(a). The geometric resolution of this image is 1.3 m. This image contains nine reference classes, and details of the quantities and the corresponding visual map are shown in Table III and Fig. 4(b), respectively.

3.2. Experimental Setting

For the dataset, we randomly selected 10 pixels for each class as the training samples, and the rest as the test samples from the reference data to validate the performances. Each parameter of basic classifier was selected by cross-validation from a reasonable range, while the confident threshold Thr was set as 0.5 for all experiments, which is considered reasonable in real cases and the multi-scale window size of PCM was set under the suggestion given in [8]. The classification accuracies were averaged over 10 runs for each classifier to reduce the possible bias induced by the random sampling. All the experiments, except for the SVM-related work (accelerated by C++), were carried out using MATLAB R2015B on a work-station with one 2.80 GHz processor and 64.0 Gb of RAM.

In this experimental part, four benchmark algorithms and the proposed algorithm can be list as follows:

- 1) SVM: SVM with radial basis function (RBF) kernel [1]
- 2) RL-SVM: relearning-PCM with SVM in 1), which is same as that in [8].

- 3) SRC: classical sparse representation classification as in [11].
- 4) RL-SRC: relearning-PCM with SRC in 3)
- 5) SRL-SRC: the proposed semi-relearning algorithm taking SRC in 1) as basic classifier.

3.3. Experimental Results

The thematic maps of the various classifiers are visually shown in Figs. 1 and Fig. 2, respectively. In Tables I and Table II,

the best results for each quality index are labeled in bold, and the sub-optimal results for each quality index are underlined. For the latter three relearning based frameworks, the visual classification maps as well as the quantitative evaluation results were taken on 10th iteration as an agent. In the evaluation tables, the average accuracy for each class, the average overall accuracy (OA), and the average kappa coefficient (κ) with their standard deviations for the different classifiers are shown in sequence.

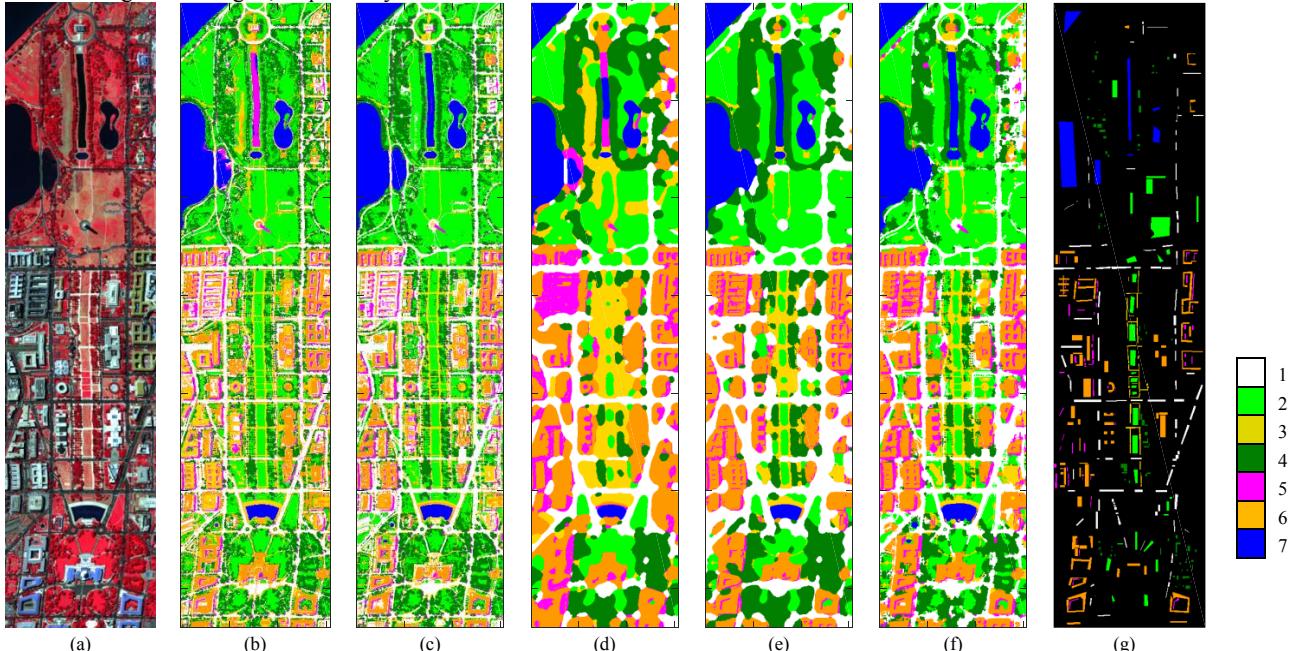


Fig. 1. Classification results for the Washington DC Mall Image: (a) false-color image (R: 63, G: 52, B: 36), (b) SVM, (c) SRC, (d) RL-SVM (10), (e) RL-SRC (10), (f) SRL-SRC (10) and (g) reference map.

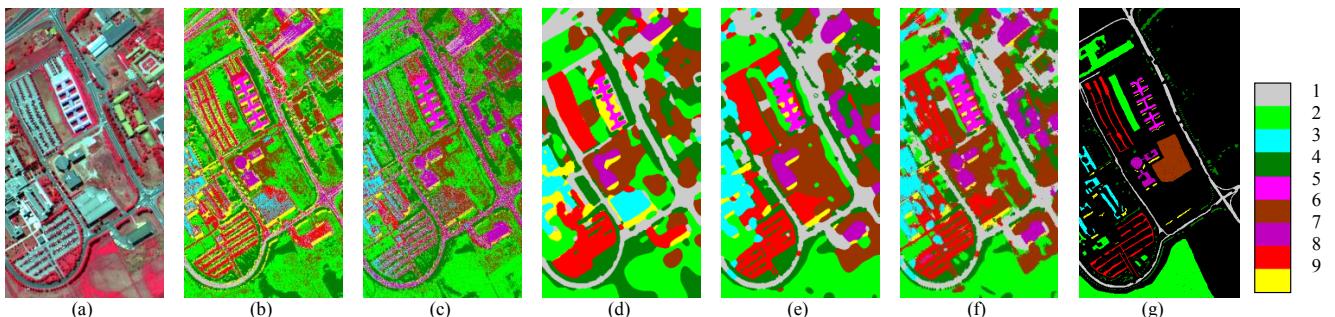


Fig. 2. Classification results for the Pavia University Image: (a) false-color image (R: 57, G: 27, B: 17), (b) SVM, (c) SRC, (d) RL-SVM (10), (e) RL-SRC (10), (f) SRL-SRC (10) and (g) reference map.

For Washington DC image, it can be observed from Table I that the proposed method yields the best accuracy and the most stable performance. With the relearning strategy, it is believed that the PCM spatial prior, which is utilized to stabilize the signal and alleviate the “salt-and-pepper” phenomenon, is effective. Compared with taking SVM as basic classifier, it is noted that the SRC with CR mechanism is more superior, as both the quantitative evaluation and the thematic map indicated. Meanwhile it is still noted that RL-SVM and RL-SRC should get stuck in the “over-smoothing” problem, even continue to slide with the growth of iteration, and the proposed framework with the SSL technique not only keeps stable accuracy, but also

maintains the dedicated spatial structure, as visually compared to the rest thematic map.

The second Pavia University data shows more spatial homogenous than the above, and all three thematic maps related to relearning technique had the “over-smoothing” problem to some extent, while the proposed work was less serious than the other two. Although the general evaluation indices as well as the iteration performance show some superiority of RL-SRC over the proposed SRL-SRC, the average accuracy for each class for four categories still show the effectiveness of the proposed work. In addition, it is noted that the almost validation samples are selected at the inner part of land-parcel, and the evaluated result cannot illustrate the full story in this case.

TABLE I
REFERENCE INFORMATION, CLASSIFICATION ACCURACY (%) FOR THE WASHINGTON DC MALL IMAGE WITH THE TEST SET

No.	CLASS Name	Num.	SVM	SRC	RL-SV M	RL-SR C	SRL-S RC
1	roads	3334	96.43	94.66	94.63	93.06	96.00
2	grass	3075	<u>90.25</u>	90.57	79.07	87.60	90.01
3	Trails	1034	90.19	95.18	93.28	95.81	98.01
4	trees	2047	98.12	96.27	<u>98.69</u>	95.18	98.74
5	shadow	1093	93.16	<u>97.45</u>	87.39	97.25	99.23
6	roofs	5867	71.37	76.61	77.36	88.87	86.32
7	water	288	91.28	98.19	90.99	<u>98.49</u>	99.48
	OA		86.47	89.42	85.72	<u>92.28</u>	93.06
			± 2.90	± 2.16	± 2.92	± 3.93	± 3.25
	κ		0.835	0.870	0.826	<u>0.905</u>	0.915
			± 0.034	± 0.026	± 0.035	± 0.048	± 0.040

TABLE II
REFERENCE INFORMATION, CLASSIFICATION ACCURACY (%) FOR THE PAVIA UNIVERSITY IMAGE WITH THE TEST SET

No.	CLASS Name	Num.	SVM	SRC	RL-SV M	RL-SR C	SRL-S RC
1	Asphalt	6631	57.98	36.55	89.42	<u>87.38</u>	81.83
2	Meadows	18649	68.63	75.94	72.33	<u>82.47</u>	82.90
3	Gravel	2099	68.15	76.77	70.02	<u>85.70</u>	89.64
4	Trees	3064	88.77	90.89	91.59	93.02	94.54
5	Metal sheet	1345	97.82	99.76	95.80	<u>99.73</u>	100.00
6	Bare soil	5029	52.21	50.17	84.89	98.23	90.35
7	Bitumen	1330	84.43	75.92	98.79	99.23	<u>98.83</u>
8	Brick	3682	68.11	29.89	92.17	<u>84.90</u>	75.45
9	Asphalt	6631	98.22	72.74	<u>91.65</u>	85.65	87.24
	OA		68.47	64.62	81.41	87.33	<u>85.25</u>
			± 2.93	± 3.70	± 6.23	± 4.12	± 4.11
	κ		0.601	0.547	0.766	0.838	<u>0.811</u>
			± 0.031	± 0.036	± 0.071	± 0.050	± 0.050

4. CONCLUSION

Fusing the collaborative representation for MAP estimation, spatial co-occurrence and alignment description and spatial smoothing inspired self-learning, this article proposed a novel semi-supervised classification framework to the HSSR image with hundreds of spectral bands and abundant spatial information. By PCM construction in the label space, more class-level spatial pattern can be gradually obtained in the relearning iteration. Experiments on several HSSR dataset have witnessed effectiveness of the proposed semi-supervised algorithm. In further agenda, more flexible spatial co-occurrence pattern should be searched.

5. REFERENCES

- [1] F. Melgani, and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778-1790, 2004.
- [2] A. Plaza, J. A. Benediktsson, J. W. Boardman *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, pp. S110-S122, 2009.
- [3] J. Li, H. Zhang, and L. Zhang, "A Nonlinear Multiple Feature Learning Classifier for Hyperspectral Images With Limited Training Samples," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, accepted for publication, 2015.
- [4] J. Li, H. Zhang, L. Zhang *et al.*, "Efficient superpixel-Level Multi-Task Joint Sparse Representation for Hyperspectral Image Classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5338-5351, 2015.
- [5] L. Zhang, M. Yang, X. Feng *et al.*, "Collaborative Representation based Classification for Face Recognition," *Arxiv preprint arXiv:1204.2358*, 2012.
- [6] I. Dopido, J. Li, P. R. Marpu, *et al.*, "Semi-supervised self-learning using soft sparse multinomial logistic regression," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4032- 4044, 2013.
- [7] B. Banerjee, F. Bovolo, A. Bhattacharya, *et al.*, "A new self-training -based unsupervised satellite image classification technique using cluster ensemble strategyS," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 4, pp. 741- 745, 2015.
- [8] X. Huang, Q. Lu, L. Zhang *et al.*, "New Postprocessing Methods for Remote Sensing Image Classification: A Systematic Study," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7140-7159, 2014.
- [9] J. Li, H. Zhang, and L. Zhang, "Supervised Segmentation of Very High Resolution Images by the Use of Extended Morphological Attribute Profiles and a Sparse Transform," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 8, pp. 1409-1413, 2014.
- [10] F. Pacifici, M. Chini, and W.J. Emery, "A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification," *Remote Sens. Environ.*, vol. 113, no. 6, pp. 1276-1292, 2009.
- [11] J. Wright, A. Y. Yang, A. Ganesh *et al.*, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210-227, 2009.

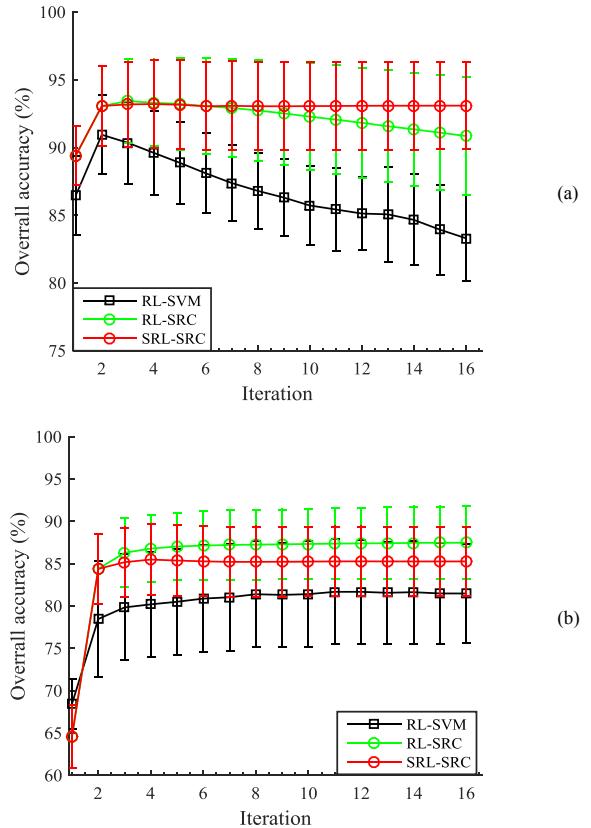


Fig. 3. Iteration analysis for (a) Washington DC image, and (b) Pavia University image