

A NOVEL APPROACH FOR SATELLITE IMAGE CLASSIFICATION USING LOCAL SELF-SIMILARITY

Zheng Huaxin, Bai Xiao and Zhao Huijie

Beihang University

37 Xueyuan Road, Haidian District, Beijing, China, 100191.

E-mail:zhenghx007@gmail.com

ABSTRACT

Extracting man-made objects in satellite images which are generated from the meter to sub-meter resolution plays an important role in remote satellite image analysis. However, spectral characteristics of urban land objects are so similar. So the classification accuracies are far from satisfactory by using only spectral information. As a result, researchers turn to incorporate geometrical information into satellite image classification. In this paper, we introduce a new local feature, namely local self-similarity(LSS) which captures internal geometric layouts of local self-similarities, into high spatial resolution images classification application. Our method captures self-similarity of color, edges, repetitive patterns and complex textures in a single unified way. With the help of Bag-of-Visual Words and SVMs, the proposed method performs well. Experimental results on Quickbird-image data set show that the proposed local self-similarity representation yields better classification performance than the low-level features, such as the spectral and texture features.

Index Terms— Classification, High resolution imagery, Local self-similarity

1. INTRODUCTION

The very high spatial resolution(VHR) remote sensing images provide valuable spatial and textural information for land cover classification. Improvements in spatial resolution of optical sensors opened a wide range of opportunities for remote sensing image analysis. With the fine resolution, there is an increased interest in the detection and identification of a variety of man-made structures such as roads and buildings. However, the difficulty of using VHR imagery such as IKONOS [1] and QuickBird [2] is that the classification accuracy for such purposes is far from satisfactory [3]. The classes such as road, parking lots, and open areas are hard to separated using only spectral information cause of their similar spectral characteristics. As a result, researchers turn to incorporate geometrical information into image classification.

For automated land cover classification, it can be accomplished using either pixel-based or object-based approaches.

Pixel-based methods classify each pixel individually, whereas object-based methods first group pixels in a meaningful way by image segmentation. Compared with the pixel-based methods, the object-based approach provides a straight way to incorporate geometrical information. However, it is a difficulty task to get well segmentations.

Many low-level object features have been applied for object-based remote sensing image analysis [4], such as spectral, texture, and structure features. In [5] [6], texture is regarded as a very useful feature of spatial structure information in high resolution images analysis. In addition to spectral and texture features, structural information extracted by mathematical morphological operations is used for detection as well [3] [7] [8].However, it is still an open problem on extract and represent useful structure information for classification.

In this paper, we demonstrate a new local feature, called local self-similarity(LSS, [9]), which integrated with the contextual and shape information [10]. We demonstrate the usefulness of LSS in urban-area land cover classification in VHR remote sensing images.Instead of measuring features such as gradients or color of a pixel, LSS measures the difference between the pixel and its neighbors. The LSS descriptor captures the internal geometric layout of local regions and can be compared across images which appear substantially different at the pixel level.

In this paper, we combine LSS with an object-based classification approach, which provides a straightforward method to incorporate geometrical information. The experimental results show that it performs well.The proposed framework has been tested for high spatial resolution images from QuickBird. The paper is structured as follows. Section 2 explains the local self-similarity feature and exhibits the process of extracting the LSS feature descriptor. Section 3 shows the classification process where Section 3.1 is for extracting local region and LSS descriptor, Section 3.2 is for visual-vocabulary construction and histogram computing and Section is 3.3 for classification using semi-supervised SVMs. Section 4 presents the experimental results which including the comparison between classification using other different

feature. Finally, conclusions are outlined in Section 5.

2. THE LOCAL SELF-SIMILARITY FEATURE

In this part, we briefly describe local self-similarity and how to use it for remote image analysis. LSS captures self-similarity of color, edges, repetitive patterns and complex textures in a single unified way. A textured region in one image can be clustered with a uniformly colored region as long as they have a similar spatial layout. These self-similarity descriptors are estimated on a dense grid of points in image, at multiple scales.

The process of extracting the LSS descriptor is exhibited as follows. We first associate a local self-similarity descriptor d_p with every pixel p . This is done by correlating the image patch centered at p with a larger surrounding image region (e.g., of radius 40 pixels), resulting in a local internal correlation surface. We use the term local to denote a small portion of the image as opposed to the entire image. The correlation surface is then transformed into a binned log-polar representation. It is computed as follows:

(1) Determine the $N \times N$ correlation surface ζ_p of the $\omega \times \omega$ patch t_p with the surrounding $N \times N$ region R_p . Both R_p and t_p are centered on p . $\zeta_p(x)$ is the correlation of t_p with a patch t_x centered on x :

$$\zeta_p(x) = \exp\left(-\frac{SSD(t_p, t_x)}{\delta}\right) \quad (1)$$

(2) Discretize the correlation surface ζ_p on a log-polar grid and store the maximal value of ζ_p within each grid bin:

$$d_p(p, d) = \max_{x \in BIN(p, d)} \{\zeta_p(x)\} \quad (2)$$

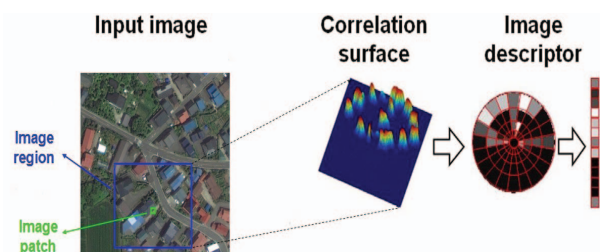


Fig. 1. The process of extracting the LSS descriptor.

Figure 1 shows the procedure of extracting the LSS descriptors. From Figure 1, LSS feature is a local feature which can be extended to a wide range of larger areas. The use of patches (at different scales) as the basic unit for measuring internal self-similarities captures more meaningful image

patterns than individual pixels, especially for the object classification and detection. This can be done without any explicit segmentation or edge detection, thus can handle regions without clear boundaries such as some vegetation region. The log-polar representation accounts for local affine deformations in the self-similarities.

3. CLASSIFICATION USING LSS

In this section, we give the the main classification process. We combine the bag-of-visual-word(BOV, [11]) and semi-supervised support vector machines(SVMs) in our proposed approach.

Given an object from the sample data set, patch detection and description are used to form a set of feature vectors from an object. In the training phase, the k-means method is applied to train k classes whose center is named as visual word. In the testing phase, a single visual word is built to represent all implausible patches which are not close enough to warrant representation by any relevant visual words. Based on the visual words, the histogram is generated by counting their occurrence numbers. Here, such histogram is defined as a BOV representation. Thus, this representation is fed for classification using semi-supervised SVMs.

3.1. Local Region Detection and LSS Descriptors

We use two different ways to extract local regions.

Evenly regular grid: Evenly regular grids are extracted at different scales, where each grid is spaced at 11×11 pixels for a given object. The size of interest points in the patch is randomly sampled between the scale of 10 to 30 pixels.

DoG detector: A set of local regions (patches) which are stable and affine invariant over different scales are extracted using the DoG detector [12]. And in each local region we can get salient points as interest points using DoG detector.

Thus, the conspicuous points are located, and their neighborhoods are regarded as the detected patches for further description. Traditionally, the BOV approach models the distribution of low-level local features such as the scale-invariant feature transformation (SIFT) [12], which computes the orientation and gradient of the keypoints in gray-level information. Here we extract LSS descriptor instead of SIFT descriptor in the conspicuous points. We discuss the comparison results in the experiments.

3.2. Visual-Vocabulary Construction and Histogram Computing

Owing to its simplicity, k-means is exploited in this work for visual-word encoding. During the local region detection and description, the test image is separated into a set of patches, each of which is described by a set of feature vectors. Hence, k clustering classes are learned from the data set, and their

centers are defined as the visual words. Note that the visual-vocabulary construction is an unsupervised method without any references to the class label of each patch, which make the generation of visual words label free. Here, k-means algorithm is implemented by an expectation-maximization (EM) scheme. However, there are two problems during this process. First, the k-means algorithm converges only to the local optima. Various initial settings would lead to different solutions. Therefore, the k-means algorithm should be repeated with random initial values, and the approximate optimal solution is obtained by averaging over the results of each run. Second, the parameter k (the number of cluster center) could not be computed automatically. In this paper, we choose $k = 100$.

Based on the clustering algorithm, a visual vocabulary is constructed to describe the object contents. Thus, each descriptor is assigned to the closest visual word by using the Euclidean distance, and a patch can be represented as a histogram by counting the occurrence numbers of the visual words.

3.3. Classification by SVMs

Here, every object is described by a set of patches, each patch is described by a histogram computed and we randomly sample patches in different objects to form a training data set. Then the semi-supervised SVMs are employed for the classification.

SVM is among the most used kernel learning algorithms. It performs robust nonlinear classification of samples using the kernel trick. The idea is to find a separating hyperplane in some feature space induced by the kernel function while all the computations are done in the original space. A good introduction to SVM for pattern recognition can be found in [13].

Despite the good performance of some types of kernels (e.g., RBF kernel), it is obvious that both the unlabeled information (data structure) and the geometrical relationship between labeled and unlabeled samples are obviated. Including the unlabeled information in the regression method may improve the results, which is the focus of semi-supervised learning (SSL). A simple yet effective way to estimate the marginal data distribution and then include this information into any kernel method consists of deforming the structure of the kernel matrix according to the unlabeled data structure. The deformation can be designed either with cluster kernels computed from the solution offered by clustering algorithms or by deforming a valid kernel with graph-based methods that account for the geometrical relations between labeled and unlabeled. In this letter, we choose the first strategy.

The selected semi-supervised SVM [14] cluster the full patches to build a bagged kernel and then modify the base kernel. A bagged kernel is a kernel function encoding the similarity between unlabeled samples. Such a kernel can be defined by counting the occurrences of two patches in the same

cluster over several runs of an unsupervised algorithm. The algorithm is defined in the following steps.

(1) Compute the base SVM kernel K_{SVM} (e.g., using the RBF kernel).

(2) Run t times the k -means algorithm with different initializations but with the same number of clusters k . This results in $p = 1, \dots, t$ cluster assignments $c_p(x_i)$ for each sample x_i .

(3) Build a bagged kernel K_{bag} based upon the fraction of times that x_i and x_j are assigned to the same cluster

$$K_{bag}(x_i, x_j) = \frac{1}{t} \sum_{p=1}^t [c_p(x_i) = c_p(x_j)] \quad (3)$$

where operator $[c_p(x_i) = c_p(x_j)]$ returns 1 if samples x_i and x_j belong to the same cluster according to the p th realization of the clustering $c_p(\cdot)$ and 0 otherwise.

(4) Take the sum or the product between the original and bagged kernels

$$K(x_i, x_j) \leftarrow K_{bag}(x_i, x_j) + K_{SVM}(x_i, x_j) \quad (4)$$

$$K(x_i, x_j) \leftarrow K_{bag}(x_i, x_j) \cdot K_{SVM}(x_i, x_j) \quad (5)$$

(5) Train an SVM with the modified kernel $K(x_i, x_j)$.

4. EXPERIMENTAL RESULTS

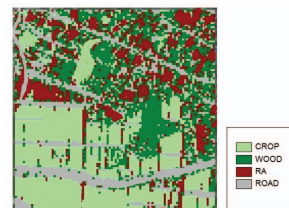
The results are presented on the satellite images by Quickbird with 0.6-m resolution. Four classes are considered, which are roads, resident area (RA), crop and wood.



(a) Experiment area.



(b) Reference map.



(c) Classification result.

Fig. 2. Experiment result.

The classification results are presented in Figure 2. From these data, we can see that our proposed methods can achieve reasonable well accuracy on objects classification. The efficiency of our methods can be accepted since our experimental are performed on desktop PC with high speed.

Table 1. Classification accuracies using different features.

Accuracy	LSS	GLCM	SIFT	Spectral feature
Overall Accuracy	88.63%	86.43%	82.08%	80.17%
Average Accuracy	92.12%	90.72%	89.11%	85.26%
Kappa Coefficient	83.06%	80.37%	77.49%	74.61%

We also compare our methods with Gray-Level Co-occurrence Matrices (GLCM, [11]), Scale Invariant Feature Transform (SIFT) and Spectral features. The results are given in Table 1. Among all different features, LSS performed best. So LSS feature can provide better structure information for satellite image analysis.

5. CONCLUSION

To conclude, a new shape feature LSS for urban area land cover classification in remote sensing images is presented and tested. We combine LSS with BOV and semi-supervised SVMs. The whole process achieved good performance in satellite image analysis. In the future, we plan to further explore semi-supervised learning methods in satellite image analysis. Hence, the information of unlabeled data can be extracted and used. Meanwhile, traditional spectral features will be incorporated in our system.

6. ACKNOWLEDGMENT

This work is supported Fundamental Research Funds for the Cental Universities and Open Projects of National Laboratory of Pattern Recognition(201001102).

7. REFERENCES

- [1] L. Wang, WP Sousa, and P. Gong, "Integration of object-based and pixel-based classification for mapping mangroves with IKONOS imagery," *International Journal of Remote Sensing*, vol. 25, no. 24, pp. 5655–5668, 2004.
- [2] D. Stow, A. Lopez, C. Lippitt, S. Hinton, and J. Weeks, "Object-based classification of residential land use within Accra, Ghana based on QuickBird satellite data," *International journal of remote sensing*, vol. 28, no. 22, pp. 5167–5173, 2007.
- [3] R. Bellens, S. Gautama, L. Martinez-Fonte, W. Philips, J.C.W. Chan, and F. Canters, "Improved classification of VHR images of urban areas using directional morphological profiles," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 10, pp. 2803–2813, 2008.
- [4] T. Blaschke, S. Lang, and G. J. Hay, *Object-Based Image Analysis*, New-York: Springer-Verlag, 2008.
- [5] P.C. Smits and A. Annoni, "Updating land-cover maps by using texture information from very high-resolution space-borne imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 3, pp. 1244–1254, 1999.
- [6] SE Franklin, RJ Hall, LM Moskal, AJ Maudie, and MB Lavigne, "Incorporating texture into classification of forest species composition from airborne multispectral images," *International Journal of Remote Sensing*, vol. 21, no. 1, pp. 61–79, 2000.
- [7] M. Fauvel, J.A. Benediktsson, J. Chanussot, and J.R. Sveinsson, "Spectral and spatial classification of hyper-spectral data using SVMs and morphological profiles," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 46, no. 11, pp. 3804–3814, 2008.
- [8] D. Tuia, F. Pacifici, M. Kanevski, and W.J. Emery, "Classification of very high spatial resolution imagery using mathematical morphology and support vector machines," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 11, pp. 3866–3879, 2009.
- [9] E. Shechtman and M. Irani, "Matching local self-similarities across images and videos," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.
- [10] B. Xiao, E. Hancock, and H. Yu, "Manifold embedding for shape analysis," *Neurocomputing*, vol. 73, no. 10-12, pp. 1606–1613, 2010.
- [11] L. Weizman and J. Goldberger, "Urban-Area Segmentation Using Visual Words," *IEEE Geoscience and Remote Sensing Letters*, vol. 6, no. 3, pp. 388–392, 2009.
- [12] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [13] C.J.C. Burges, "A tutorial on support vector machines for pattern recognition," *Data mining and knowledge discovery*, vol. 2, no. 2, pp. 121–167, 1998.
- [14] D. Tuia and G. Camps-Valls, "Semisupervised remote sensing image classification with cluster kernels," *Geoscience and Remote Sensing Letters, IEEE*, vol. 6, no. 2, pp. 224–228, 2009.