

Hierarchical Remote Sensing Image Analysis via Graph Laplacian Energy

Zhang Huigang, Bai Xiao, Zheng Huaxin, Zhao Huijie, Zhou Jun, Cheng Jian, and Lu Hanqing

Abstract—Segmentation and classification are important tasks in remote sensing image analysis. Recent research shows that images can be described in hierarchical structure or regions. Such hierarchies can produce the state-of-the-art segmentations and can be used in the classification. However, they often contain more levels and regions than required for an efficient image description, which may cause increased computational complexity. In this letter, we propose a new hierarchical segmentation method that applies graph Laplacian energy as a generic measure for segmentation. It reduces the redundancy in the hierarchy by an order of magnitude with little or no loss of performance. In the classification stage, we apply local self-similarity feature to capture the internal geometric layouts of regions in an image. By incorporating advantages from both semantic hierarchical segmentation and local geometric region description, we have achieved better performance than those from the methods being compared. In the experimental section, we validate the effectiveness of our method by showing results on QuickBird and GeoEye-1 image data sets.

Index Terms—Classification, graph Laplacian energy (LE), high-resolution imagery, local self-similarity (LSS).

I. INTRODUCTION

IMPROVEMENTS in spatial resolution of optical sensors have opened new opportunities for remote sensing image analysis. These sensors produce very high spatial resolution (VHR) remote sensing images that contain valuable spatial and textural information for land-cover classification. It leads to increased interests in the detection and identification of a variety of man-made structures such as roads and buildings.

Land-cover classification can be accomplished using either pixel-based or object-based approaches. The pixel-based methods classify each pixel individually, whereas the object-based

methods require image segmentation that groups pixels into regions before the classification step. Compared to the pixel-based methods, the object-based counterparts provide straightforward solutions that incorporate geometrical information from remote sensing images. Therefore, they can handle the salt-and-pepper noise that is inherent in the images and can extract meaningful and stable features for region description. These properties have generated great advantages in analyzing high-resolution remote sensing data [1], [2].

Image segmentation is a crucial step for the object-based remote sensing imagery analysis. Several segmentation algorithms describe the remote sensing images using hierarchical structures [3], [4]. Akcay and Aksoy [5] applied morphological operations on individual spectral bands to form a hierarchy segmentation. Ons and Tebourbi [6] presented a hierarchical image segmentation method based on region merging according to spatial and spectral criteria. In [7], Gaetano *et al.* introduced an improved image segmentation method for multiresolution remote sensing images. In [8], a top-down analysis method is introduced for pattern extraction. Such hierarchies can produce the state-of-the-art segmentation results and can be used in the classification. However, they often contain more levels and regions than required for an efficient image description, which may cause increased computational complexity.

In this letter, we propose a new hierarchical image analysis method that applies the graph Laplacian energy (LE) [10] as a generic measure for segmentation. With segmentation results available, we proceed to the classification step using local self-similarity (LSS) [9] to incorporate the local contextual and shape information. We demonstrate the effectiveness of the proposed classification method in urban-area land-cover classification using VHR remote sensing images.

The contributions of this letter are threefold. First, we use the graph LE to describe the VHR remote sensing image in a hierarchical structure. It highlights the high-level semantic structure of an image. Second, we introduce a method to extract local contextual and shape information in local regions using LSS. Finally, we demonstrate the feasibility of a classification system for remote sensing images by combining both high-level and low-level information.

This letter is organized as follows. Section II explains the segmentation method using the graph LE. Section III describes the classification step by extracting LSS features and training support vector machines (SVMs). Finally, in Section IV, we present the experimental results, which include comparisons between the proposed method and alternative classification approaches in the literature. Conclusions of this letter are drawn in Section V.

Manuscript received November 5, 2011; revised March 1, 2012 and May 17, 2012; accepted June 27, 2012. Date of publication August 21, 2012; date of current version October 22, 2012. This work was supported in part by National Science Foundation of China under Grant 61105002, by the Open Projects of National Laboratory of Pattern Recognition under Grant 201001102, and by the Fundamental Research Funds for the Central Universities.

H. Zhang and X. Bai are with the School of Computer Science and Engineering, Beihang University, Beijing 100191, China (e-mail: huigangzhang@buaa.edu.cn; baixiao.buaa@googlegmail.com).

H. Zheng and H. Zhao are with the School of Instrumentation Science and Optoelectronics Engineering, Beihang University, Beijing 100191, China (e-mail: huaxinzheng@buaa.edu.cn; huijiezhao@buaa.edu.cn).

J. Zhou is with the School of Information and Communication Technology, Griffith University, Nathan, QLD 4111, Australia (e-mail: jun.zhou@anu.edu.au).

J. Cheng and H. Lu are with the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: jcheng@nlpr.ia.ac.cn; hlu@nlpr.ia.ac.cn).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/LGRS.2012.2207087



Fig. 1. Example of watershed segmentation. The left hand is the original image. The right hand shows the corresponding watershed segmentation results.

II. HIERARCHICAL REPRESENTATION USING GRAPH LE

Image segmentation is a crucial step for the object-oriented remote sensing imagery analysis. In this section, we describe how we use the graph LE defined by Gutman and Zhou [10] as a segmentation measure to analyze the remote sensing images.

Let G be an (n, m) graph with n vertices and m edges and A be its adjacency matrix. Let d_i be the degree of the i th vertex of G and D be the corresponding degree matrix, where $D(i, i) = d_i$. Then, $L = D - A$ is the Laplacian, and the graph LE is defined as

$$LE(G) = \sum_{i=1}^n \left| \lambda_i - \frac{2m}{n} \right| \quad (1)$$

where λ_i denotes eigenvalues of the Laplacian matrix and $2m/n$ is the average vertex degree.

Based on this definition, we propose a new method to extract hierarchical structure from remote sensing images. First, we calculate the image gradient and cut the remote sensing image into small connected regions using watershed segmentation [11]. An example of the segmentation results is shown in Fig. 1. This step generates the bottom level of the hierarchy and transforms the input image into a region adjacency graph.

The next step is to merge neighboring regions from bottom up to create a hierarchical tree description. In each merging iteration, we merge the most similar pairs of neighboring regions and treat the newly merged regions as parent nodes. The merging continues until there is only one region left. The initial watershed segmentation results are set as level 1. Each merging iteration generates a higher level of segmentation over the previous one.

The merging step should be able to handle regions in arbitrary shape. Therefore, we use simple statistical properties, i.e., a vector containing the average *red*, *green*, and *blue* (*RGB*) colors and the average intensity gradient value, to describe a region for similarity comparison. Each region is modeled as a single Gaussian $\Omega = (N, \mu, C)$, in which Ω is a region, N indicates the number of pixels in this region, and μ and C are the mean and covariance of the property vectors at each pixel, respectively. If two regions (Ω_i and Ω_j) can be merged, the new region ($\Omega = \Omega_i + \Omega_j$) is represented as

$$\Omega = \left(N_i + N_j, \frac{N_i \mu_i + N_j \mu_j}{N_i + N_j}, \frac{N_i C_i + N_j C_j}{N_i + N_j} + e \right). \quad (2)$$

The “error” term e compensates the eigenspace for the difference between the means of the model, which is

$$e(i, j) = \frac{N_i N_j}{N_i + N_j} (\mu_i - \mu_j)^T (\mu_i - \mu_j). \quad (3)$$

To calculate the graph LE of each merged region, we need to get the adjacency matrix at each merging step. The adjacency matrix is weighted using the error measure. If regions Ω_i and Ω_j are connected, the weight is $w(i, j) = \exp(-e(i, j)/e_{\max})$, in which e_{\max} is the maximum value of $e(i, j)$ over all connected pairs.

In each merging iteration, we merge the most similar pair of neighboring regions, i.e., the pair with the minimum value of $e(i, j)$. This merging step generates a full hierarchy tree. In the next step, we analyze this tree via graph LE. Our principle is to choose those tree levels that are lower in complexity than their adjacent levels.

Note that the standard graph LE in (1) is normally defined on a single connected graph. However, in our case, we treat the merged regions as connected graphs so that each initial watershed region is a node in a graph. At each level except the highest one, there are more than one connected graphs. Therefore, we extend the original graph LE definition to make it suitable for computing LE at each level of the tree. We introduce the normalized graph LE (nGLE). For a level with K connected graphs, we define the nGLE as

$$E = \frac{n}{K} \sum_{i=1}^K \frac{LE(G_i)}{|n_i|} \quad (4)$$

where G_i is the i th connected graph of $|n_i|$ nodes and n is the number of nodes of all the connected graphs. In the case of a single connected graph, our expression reduces to the original graph energy.

We compute the nGLE at every level in the hierarchy independently and use the nGLE as a function of level index. Fig. 2 shows a typical nGLE curve. In this curve, local minima are met when graphs at particular levels exhibit homogeneous node degree, which means that the graphs are close to regular graphs. They correspond to levels that are less complex compared to the adjacent levels. We choose the highest level partition that gives the local minimum, which breaks the image into the fewest large components.

Through the aforementioned process, we can get the main components of a remote sensing image by using the graph LE index as shown in Fig. 2. This first segmentation can produce some main components of the image such as roads, lands, and buildings. In order to classify the image, we need to further segment the existing main components into finer parts. We find these parts by recursively applying the graph Laplacian method on each component. This means that we compute the nGLE for each component again and segment the component by searching for the local minima as described previously. In this step, each component is treated as a hierarchy tree, and the level corresponding to the largest local minimum is segmented again. Our experiments verified that two levels of recursion are enough to achieve good performance. Thus, our method only applies

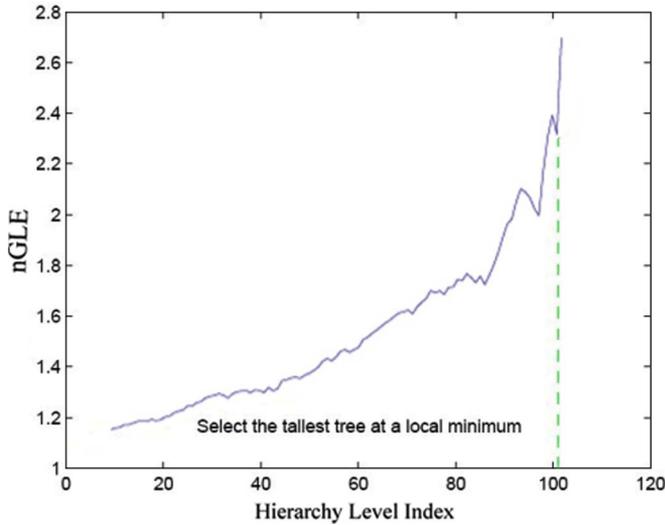


Fig. 2. Graph energy as a function of level index. The plot is computed using (4) on the image in Fig. 1.

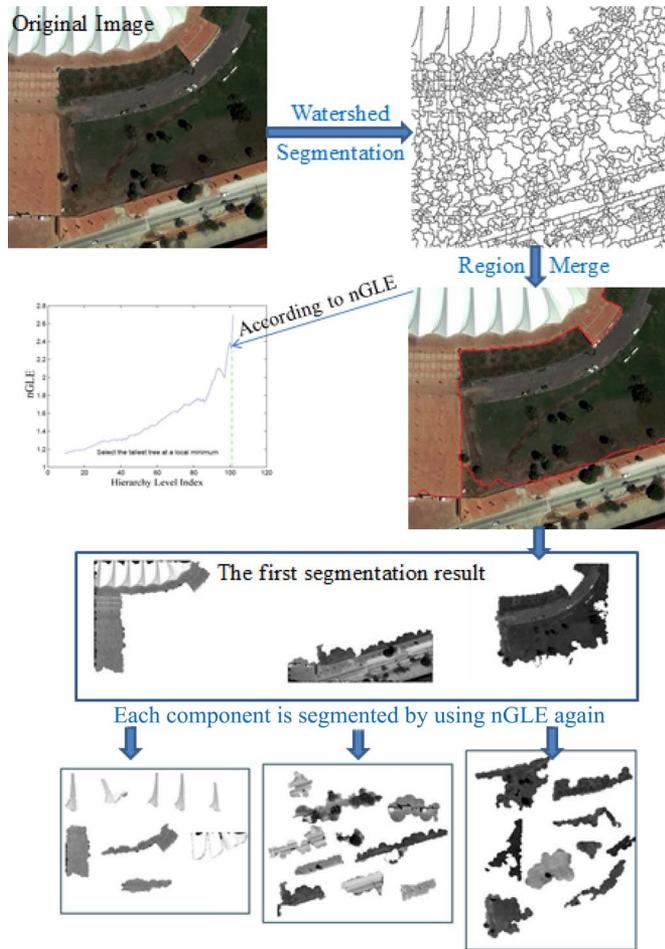


Fig. 3. Segmentation process.

two levels of division, as shown in Fig. 3. The merging and segmentation processes are shown in Fig. 3. In this example, the high-resolution remote sensing image has been segmented in a hierarchical manner. From this process, a set of parts

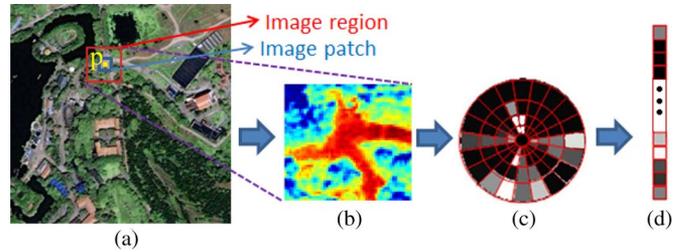


Fig. 4. Process of extracting the LSS descriptor of pixel p . (a) Image patch and its close neighborhood region. (b) Local internal correlation surface. (c) Binned log-polar representation. (d) Normalized log-polar vector.

for each image can be produced, as shown at the bottom row in Fig. 3.

III. CLASSIFICATION USING LSS

The hierarchical description of the image forms the basis to the classification step. To extract features from each segmented parts, the LSS method [9] is used. The LSS describes the similarity between a patch and its neighboring region in an image. It offers a single unified way to describe the internal relations in an image. The process of extracting the LSS descriptor is shown in Fig. 4. It is computed as follows.

- 1) Calculate an $N \times N$ correlation surface ζ_p of an $\omega \times \omega$ patch t_p with the surrounding $N \times N$ region R_p using sum of squared differences (SSD) method. In this letter, $N = 30$ and $\omega = 5$. Both R_p and t_p are centered at p . $\zeta_p(x)$ is the correlation of t_p with an $\omega \times \omega$ patch t_x centered at x : $\zeta_p(x) = \exp(-(\text{SSD}(t_p, t_x)/\delta))$, where δ represents the maximal variance of the difference between all patches within a very small neighborhood of p and the patch centered at p .
- 2) Discretize the correlation surface ζ_p on a log-polar grid, and store the maximal value of ζ_p within each grid bin $d_p(p, d) = \max_{x \in \text{BIN}(p, d)} \{\zeta_p(x)\}$.
- 3) Normalize the binned log-polar vector by linearly stretching its values to the range [0, 1].

From each image part generated in Section II, a set of LSS descriptors can be generated. In order to describe each part using a single vector, the bag-of-visual-words model [18] is adopted. We exploited k -means clustering method for visual-word encoding ($k = 300$ in our experiments), which groups the LSS descriptors into k clusters. The cluster centers are defined as the visual words, and a visual vocabulary is constructed to describe the content of objects. After assigning each descriptor to the closest visual word, all image parts can be represented as a histogram by counting the occurrence numbers of the visual words.

In the classification step, the SVM is used as the classifier. The choice is mainly based on the fact that it is one of the state-of-the-art classifiers on bag-of-visual-words image representation. We adopted the C-SVC in LIBSVM [15] with an *RBF* kernel. The parameters for the SVM are obtained by cross-validation on a subset of manually labeled training parts. For example, the kernel parameter γ and cost parameter C are set to 2 and 0.762, respectively, in the 50 training number case in Table I.

TABLE I
CLASSIFICATION ACCURACIES WITH
DIFFERENT TRAINING SET SIZES

Land-Cover Category	Number of training samples per class				
	50	100	150	200	250
RA	88.45%	90.51%	91.47%	92.93%	93.36%
WOOD	86.76%	89.32%	91.16%	92.67%	93.03%
LAWN	83.24%	88.49%	90.71%	91.87%	92.52%
ROAD	79.95%	83.11%	86.98%	88.34%	90.66%
LAND	85.00%	87.45%	90.02%	90.73%	91.58%
Overall	85.27%	87.18%	90.02%	91.23%	91.89%

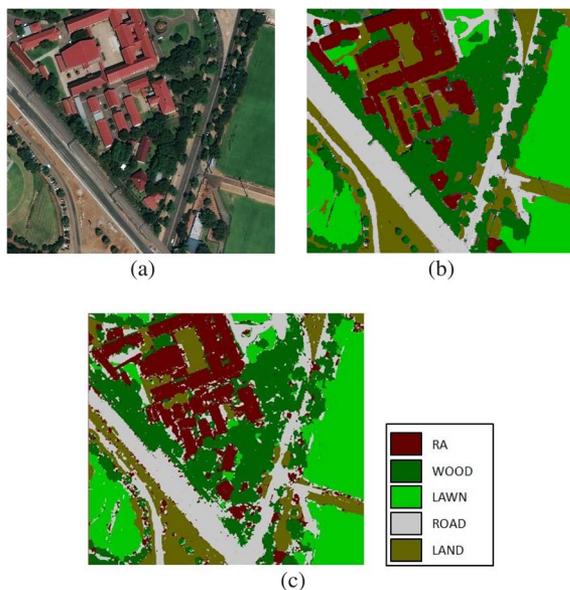


Fig. 5. Example on classification result. (a) Experiment area. (b) Reference map. (c) Classification result.

IV. EXPERIMENTS AND RESULTS

In this section, we turn our attention to demonstrating the utility of the proposed hierarchical image analysis method. We also compare our methods with alternatives in the literature. The experiments were performed on a desktop with an Intel Core 2 Duo 2.40-GHz processor.

A. Classification Performance

Two types of high-resolution satellite images, including 0.6-m QuickBird and 0.5-m GeoEye-1 images, have been employed to evaluate the performance of the proposed method. We have collected more than 500 images with different scenes, each with a spatial dimension of 640×640 pixels.

In our experiments, we used random sampling approach [17] to validate the classification results. We first segmented the 500 images into hierarchical structures to get the segmented parts. Then, different numbers of image parts (i.e., 50, 100, 150, 200, and 250) were randomly selected as training sets, while the rest were used as the testing sets. After that, LSS features were extracted, and each part was represented by a vector using the bag-of-visual-words model. Finally, we classified these parts using the SVM and got the final result.

An example of the classification result is shown in Fig. 5. In this image, five classes are considered, which are lawn, resident



Fig. 6. Comparison of segmentation methods. (Top left) RGB image. (Top right) Results from our method. (Down left) Results from normalized cuts. (Down right) Results from mean shift.

area (RA), road, land, and wood. Manual interpretation has been used as a reference for the classification algorithm. It can be seen in Fig. 5 that our method has achieved classification results that are close to manual labels, with all the five classes being segmented and classified well. The results suggest that the graph LE method can segment the image into refined parts. Furthermore, the LSS descriptor provides abundant local similarity information, which is useful for the local part classification.

Table I presents the quantitative evaluation of our classification method. The size of the training sets varies from 50 to 250. The classification accuracy is computed by using 1000 parts as the testing set. It can be observed that most experiments have achieved accuracy over 85%. The proposed method works well with small training set. When the number of the training samples is 50, the approach can reach 80% accuracy. Meanwhile, when the training parts are 250 per category, the accuracy is over 90%.

B. Comparison With Other Methods

We performed two experiments to compare our method with several alternative methods in the literature. In the first experiment, we compared the hierarchical segmentation method with normalized cuts [12] and mean shift [13] which are two commonly applied segmentation algorithms in remote sensing image processing. In this experiment, we set the number of segments used for normalized cuts algorithm as 50. For mean-shift algorithm, we set the window size to 3×3 and adopted a Gaussian kernel. Fig. 6 shows that our method gives more reasonable results in object-level segmentation with moderate number of blocks. The normalized cuts can separate the main regions of a remote sensing image but is not successful on the

TABLE II
CLASSIFICATION ACCURACIES USING DIFFERENT FEATURES

Accuracy	LSS	GLCM	SIFT
Overall Accuracy	88.63%	86.43%	82.08%
Average Accuracy	92.12%	90.72%	89.11%
Kappa Coefficient	83.06%	80.37%	77.49%

detailed parts. For example, it cannot separate the “lawn” and “wood” very well. Note that it is possible to improve some of the segments by tuning the parameters of the normalized cuts and mean-shift algorithms. However, we observe that different parameters needed to be manually selected for different data sets and the parameters that performed well for an image may generate bad results for another image. On the contrary, the proposed segmentation algorithm can automatically select suitable segments without parameter tuning. Thus, the proposed graph LE-based method can provide better solution for satellite image analysis.

We have also evaluated the efficiency of our algorithm. For the proposed image segmentation method based on Laplacian energy, the average time for segmenting an image is 10.6 s. The normalized cuts and mean-shift algorithms take 14.5 and 15.3 s, respectively, to finish the same task. Therefore, the proposed method is more efficient than the two alternatives.

In the second experiment, we compare the LSS feature with scale-invariant feature transform (SIFT) [14] and gray-level co-occurrence matrices (GLCMs) [16]. The reason for choosing SIFT and GLCM is that both of them are classical and widely adopted methods in remote sensing image analysis. For the SIFT method, we have used bag-of-visual-words model to describe each segmented region. As to the GLCM method, for each region, we first calculate four GLCMs, with directions of 0° , 45° , 90° , and 135° , respectively. From each co-occurrence matrix, we then calculate the corresponding measures, i.e., energy, entropy, contrast, and homogeneity. The mean and standard deviation of the four measures are calculated based on the four co-occurrence matrices. Finally, each region is represented as an 8-D feature vector composed of the four means and four standard deviations. To ensure a fair comparison, we also use SVM classifiers to get the classification results of these two methods.

The training set contains 50 samples per class, while the testing set contains 1500 samples in total. Overall, accuracy, average accuracy, and kappa coefficient were used to compare the performances of the methods under consideration. The average accuracy is the average of the accuracies for each class. The overall accuracy is a similar measure weighted by the proportion of testing samples for each class. The kappa coefficient value represents how much accuracy improvements can be achieved over the classification which resulted from a random unsupervised assignment. The classification results are given in Table II. The results indicate that the LSS features are a better choice than the alternatives in the proposed approach. It has achieved higher classification accuracy in all three evaluation criteria than those from the SIFT and GLCM.

V. CONCLUSION

In this letter, we have introduced graph LE as a complexity measure for remote sensing image analysis. With the selected hierarchies using this energy, we can get a sensible semantic interpretation in terms of objects and object parts which helps to achieve more robust classification. We also introduced the LSS for urban-area land-cover classification in remote sensing images. The proposed method has achieved performance on satellite image analysis that is better than those from alternative methods. In the future, we plan to further explore the semisupervised learning methods in satellite image analysis. This allows the information in unlabeled data to be extracted and used. Furthermore, traditional spectral features will be incorporated into our system.

REFERENCES

- [1] L. Wang, W. P. Sousa, and P. Gong, “Integration of object-based and pixel-based classification for mapping mangroves with IKONOS imagery,” *Int. J. Remote Sens.*, vol. 25, no. 24, pp. 5655–5668, 2004.
- [2] T. Blaschke, S. Lang, and G. J. Hay, *Object-Based Image Analysis*. New York: Springer-Verlag, 2008.
- [3] J. M. Duarte-Carvajalino, G. Sapiro, M. Vélez-Reyes, and P. E. Castillo, “Multiscale representation and segmentation of hyperspectral imagery using geometric partial differential equations and algebraic multigrid methods,” *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2418–2434, Aug. 2008.
- [4] Y. Wu, K. Ji, W. Yu, and Y. Su, “Region-based classification of polarimetric SAR images using Wishart MRF,” *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 668–672, Oct. 2008.
- [5] H. G. Akçay and S. Aksoy, “Automatic detection of geospatial objects using multiple hierarchical segmentations,” *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2097–2111, Jul. 2008.
- [6] G. Ons and R. Tebourbi, “Object oriented hierarchical classification of high resolution remote sensing images,” in *Proc. 16th IEEE ICIP*, Nov. 7–10, 2009, pp. 1681–1684.
- [7] R. Gaetano, G. Scarpa, and G. Poggi, “Hierarchical texture-based segmentation of multiresolution remote-sensing images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2129–2141, Jul. 2009.
- [8] C. Kurtz, N. Passat, P. Gançarski, and A. Puissant, “Extraction of complex patterns from multiresolution remote sensing images: A hierarchical top-down methodology,” *Pattern Recognit.*, vol. 45, no. 2, pp. 685–706, Feb. 2012.
- [9] E. Shechtman and M. Irani, “Matching local self-similarities across images and videos,” in *Proc. IEEE CVPR*, Jun. 17–22, 2007, pp. 1–8.
- [10] I. Gutman and B. Zhou, “Laplacian energy of a graph,” *Linear Algebra Appl.*, vol. 414, no. 1, pp. 29–37, Apr. 2006.
- [11] J. M. Gauch, “Image segmentation and analysis via multiscale gradient watershed hierarchies,” *IEEE Trans. Image Process.*, vol. 8, no. 1, pp. 69–79, Jan. 1999.
- [12] J. Shi and J. Malik, “Normalized cuts and image segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.
- [13] D. Comaniciu and P. Meer, “Mean shift: A robust approach toward feature space analysis,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.
- [14] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
- [15] C.-C. Chang and C.-J. Lin (2011, Apr.). LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* [Online]. 2(3), pp. 27:1–27:27. Available: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- [16] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, “Textural features for image classification,” *IEEE Trans. Syst., Man Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.
- [17] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification and Scene Analysis*, 2nd ed. New York: Wiley, 2000.
- [18] L. Fei-Fei, R. Fergus, and A. Torralba, “Recognizing and learning object categories,” in *Proc. CVPR Short Course*, 2007.