An Energy-Driven Total Variation Model for Segmentation and Classification of High Spatial Resolution Remote-Sensing Imagery

Qian Zhang, Xin Huang, and Liangpei Zhang

***Abstract*—An energy-driven total variation (TV) formulation is proposed for the segmentation of high spatial resolution remote-sensing imagery. The TV model is an effective tool for image processing operations such as restoration, enhancement, reconstruction, and diffusion. Due to the relationship between the TV model and the segmentation problem, in this letter, a TV-based approach is investigated for segmentation of high-spatial-resolution remote-sensing imagery. Subsequently, an object-based classification method, i.e., majority voting, is used to classify the segmented results. In experiments, the proposed TV-based method is compared with the widely used fractal net evolution approach and the clustering segmentation methods such as the expectation–maximization and *k*-means. The performances of the segmentation and the classification are evaluated based on both thematic and geometric indices.**

***Index Terms*—Classification, high resolution, object based, segmentation, total variation (TV).**

I. Introduction

HIGH spatial resolution images show small interclass but large intraclass variance, leading to inadequacy of the traditional spectral-based image interpretation. It is widely agreed that methods which integrate the spatial and spectral information can improve the classification accuracy of high spatial resolution images [1]. Object-based image analysis (OBIA) is an effective approach for simultaneously taking into account the spectral and contextual information in an image. In recent years, the OBIA approach has received much attention, and a large amount of literature concerning OBIA has been published in the remote-sensing community. Tian and Chen [2] proposed a framework based on the object-based image segmentation for artificial feature recognition from IKONOS multispectral images. Li and Shu [3] used object-oriented structural features for distinguishing between the artificial objects and natural objects. Tan *et al.* [4] used the objected-oriented classification for building extraction from IKONOS images with light detection and ranging data. Huang and Zhang [5] proposed a morphological building index for automatic building extraction from high-resolution images. Korting *et al.* [6] proposed a resegmentation approach for detection of rectangular objects. Weizman and Goldberger [7] presented a new approach for object-based urban-area detection using visual words. Zheng *et al.* [8] proposed a new local feature, namely, local self-similarity, for man-made object extraction from QuickBird images.

Segmentation is a critical preprocessing technique for the subsequent object-oriented classification of high spatial resolution images. With the development of the OBIA method, segmentation has received increasing interest. Vincent and Soille [9] proposed the watershed method, which has been widely used in medical and remote-sensing image segmentation. Baatz and Sch [10] proposed the fractal net evolution approach (FNEA) segmentation, which is embedded in the well-known object-oriented analysis software eCognition. FNEA is a bottom-up region growing method, which integrates both the spectral and geometric information. Comaniciu and Meer [11] proposed the mean-shift (MS) segmentation based on the probability density estimation of feature space. Huang and Zhang [12] proposed an adaptive MS approach for object extraction and classification from urban hyperspectral imagery. In addition, spectral clustering methods, such as k-means, ISODATA [13], and expectation–maximization (EM), can also be used for image segmentation. Tarabalka *et al.* [14] proposed a spectral–spatial classification for hyperspectral imagery based on EM clustering algorithm. Gaetano *et al.* [15] proposed a hierarchical segmentation based on the texture features of multiresolution images, e.g., IKONOS panchromatic image and four multispectral bands.

The total variation (TV) regularization model was originally proposed by Rudin *et al.* [16] for image restoration. The main benefit of the TV model is that it has no particular bias toward a discontinuous or smooth solution. In other words, the TV model is able to simplify the image and, at the same time, preserve edges and structures. This property makes the TV model appropriate for segmentation of remotely sensed imagery. It has been shown that TV is a very effective tool for segmentation of medical and natural images [17], [18]. The traditional TV segmentation [17], [19] was implemented by tracking the isolevel sets obtained by TV regularization through the scale-space stack. Subsequently, Petrovic and Vandergheynst [18] proposed a new TV segmentation approach based on the graph theory,

Manuscript received November 9, 2011; revised February 25, 2012; accepted April 2, 2012. Date of publication May 4, 2012; date of current version September 7, 2012. This work was supported in part by the National Basic Research Program of China (973 Program) under Grant 2011CB707104, by the National Natural Science Foundation of China under Grants 40930532 and 41101336, by the Program for New Century Excellent Talents in University under Grant NCET-11-0396, and by the Research Fund for the Doctoral Program of Higher Education of China under Grant 20110141120072.

The authors are with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: huang_whu@163.com).

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.2012.2194694

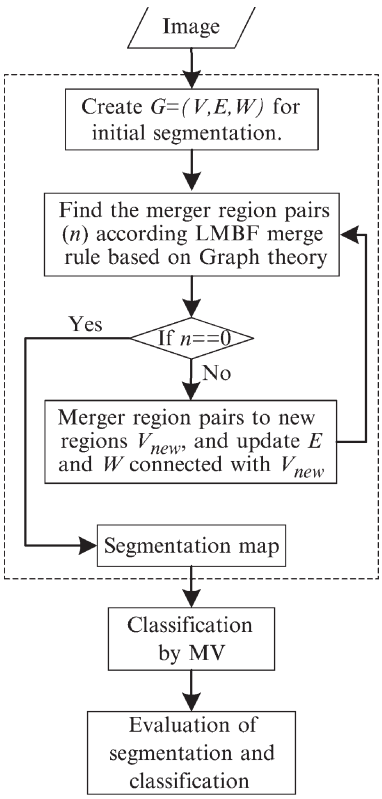

Fig. 1. Processing chain of the proposed TV segmentation and classification model (n is the number of mutual best-fitting region pairs).

to overcome the problem proposed in [17]. The TV-based segmentation for medical and natural images has been reported by several studies [17], [18], but few studies have applied it to remote-sensing image segmentation and object-based classification. In this context, the objective of this study is to propose an energy-driven TV model for the object-based image segmentation and classification of high-resolution remote-sensing images. The energy-driven TV model is resolved and optimized by a local mutual best fitting (LMBF) merge rule based on the graph theory, which is detailed in Section II. The object-based classification using the majority voting (MV) method and the assessment of the segmentation quality are also described in Section II. Sections III and IV present the results and conclusions for the experiments.

II. Methodology

As shown in Fig. 1, the processing chain of the proposed TV-based segmentation and classification method consists of four blocks: 1) TV segmentation; 2) LMBF rule for region merging; 3) object-based classification based on the TV segmentation; and 4) evaluation of results.

A. TV Model for Segmentation

The TV model was originally proposed by Rudin *et al.* [16] for image denoising. The main idea is to minimize the TV in the image, subject to constraints involving goodness of fit with the original data. The constraints are resolved using Lagrange multipliers. Therefore, the problem is formulated as an unconstrained problem [18], as shown in

$$\min \frac{1}{2}\int_{\Omega}(u-u_0)^2dx + \lambda \int_{\Omega}|\nabla u|dx \tag{1}$$

where n is the number of pixels in image u_0 and u represents the reconstructed or enhanced image. The first term $\int_{\Omega}(u-u_0)^2dx$ is the fidelity term that measures the similarity between u and u_0. λ is the regularization parameter. The second term $\int_{\Omega}|\nabla u|dx$ is the TV or regularization term, where $|\nabla u|$ is the pixel divergence. The divergence of each pixel $|\nabla u_{x,y}|$ (x and y denote the position of a pixel) is obtained based on its neighborhood, as expressed in

$$|\nabla u_{x,y}| = \sqrt{(u_{x,y}-u_{x+1,y})^2 + (u_{x,y}-u_{x,y+1})^2}. \tag{2}$$

Compared with other regularization techniques, the main advantages of the TV model are that it has no particular bias toward a discontinuous or smooth solution and it does not penalize edges.

Segmentation can be defined as dividing an image into a series of nonoverlapping regions. Pixels within a region have similar spectral and textural characteristics. Essentially, the image segmentation problem can be seen as a constrained energy minimization problem. As shown in

$$E(u) = E_A(u) + \lambda E_R(u) \tag{3}$$

$E(u)$ represents the energy of a simplified or segmented image u. E_A and E_R are the approximation and regularization terms, respectively. E_A describes the difference between the segmentation (or simplified) image and the original image. E_R represents the difference between adjacent regions in the segmented image.

Some [17], [18] have proved that TV minimization is consistent with image segmentation. Combining (1) and (3), the derived TV segmentation formulation can be expressed in

$$E(V_i, V_j) = \frac{1}{2}\sigma_i^2 + \lambda|m_i - m_j|. \tag{4}$$

The approximation term E_A represents the spectral variance σ_i^2 for a region i. The regularization term E_R can be expressed as the difference of the spectral means between adjacent regions i and j (m_i and m_j).

B. LMBF Rule

In addition to the regularization parameter λ and energy $E(V)$, the optimization of the energy-driven TV model is significantly affected by the rule of region merging. In this study, the LMBF merge rule [10], based on the framework of the graph theory, is used to generate segments. As shown in the dotted rectangle in Fig. 1, the proposed TV segmentation method can be described as follows.

1) Create the graph $G = (V, E, W)$ with initial segmentation. Each pixel is viewed as a separate region V. Each region V connects with four spatially adjacent regions (or pixels) by four connected edges E, with the corresponding weights W. W is obtained according to (4).

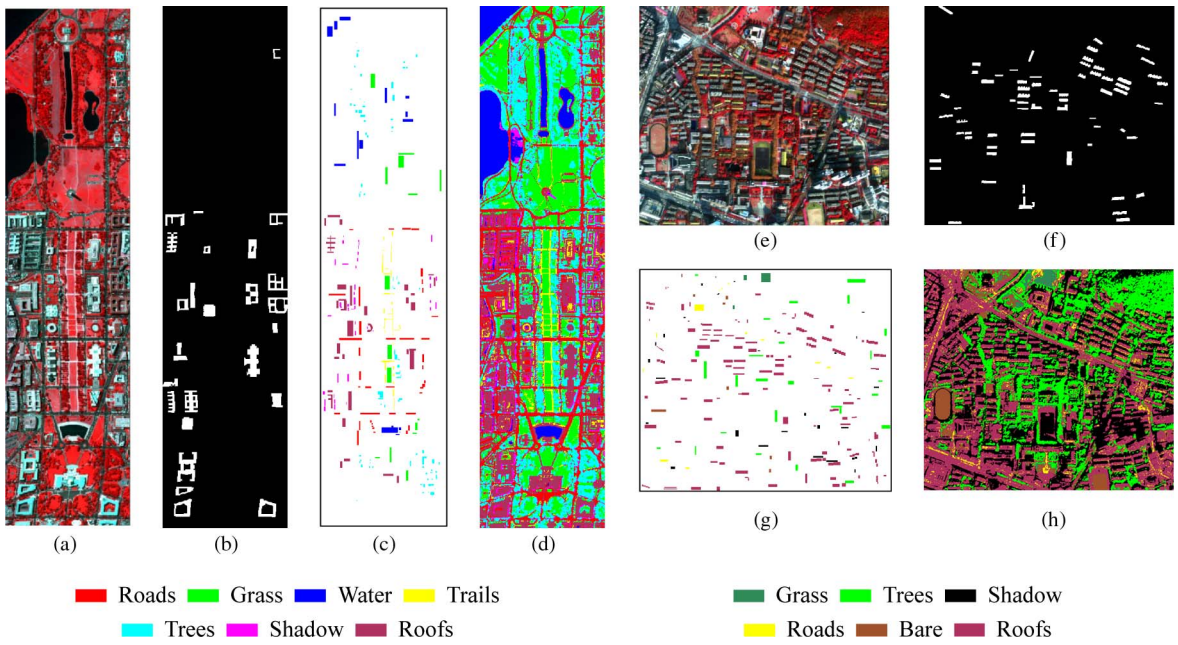

Fig. 2. (a)–(d) First test image HYDICE Washington DC data set. (e)–(h) Second test image GeoEye-1 Wuhan data set. (a) and (e) present the test images, (b) and (f) are their test samples of buildings for segmentation, (c) and (f) are the test samples for classification, and (d) and (h) show the pixel-based SVM classification maps of (a) and (e).

2) Find the region merging pairs

$$j = \arg\min_{j \in N_i} (E(V_i, V_j))$$
$$k = \arg\min_{k \in N_j} (E(V_j, V_k)) \quad (5)$$
$$E(V_i, V_j) < E_{\text{th}} \qquad E(V_j, V_k) < E_{\text{th}} \quad (6)$$

where E_{th} is the minimum energy that a region can contain. For a region i, we first search for the most similar region j in its neighborhood N_i and subsequently find the most similar neighbor k for the region j in the neighborhood N_j. If both energies fulfill the homogeneity criterion defined in (6) and i and k are the same regions, the regions i and j are viewed as the mutual best-fitting pair of regions. Subsequently, regions i and j are merged into one region V_{ij}, and the edge E_{ij} between them is removed.

3) Update all edges (E) and the corresponding weights (W) connected with region V_{ij}.
4) Go to the next region. If each region in the image is processed according to the aforementioned steps, then go to the first region of G.
5) Repeat steps 2)–4), until no adjacent regions satisfy the merging conditions.

C. Object-Oriented Classification Based on MV

In this study, an MV method is used to classify the segmented image, based on the pixel-based classification map of the original image [14]. The MV algorithm can be written as

$$k = \arg\max_{k \in \{1,\ldots,K\}} (p_k(V_i)) \quad (7)$$

where K is the number of classes and $P_k(V_i)$ is the number of pixels in region i belonging to class k. First of all, the pixelwise classification result for the original image is obtained. The region is then labeled with the class that has the largest proportion of the all classes within the region. In this letter, the pixelwise classification map is obtained by a support vector machine (SVM) classifier [20].

D. Assessment for the Segmentation

Segmentation quality is commonly indirectly evaluated through the accuracy of the classification obtained from the segmentation result. Some studies have presented object-oriented methods for direct assessment of segmentation, including both supervised [2], [21], [22] and unsupervised [23], [24] ones. In this letter, both thematic and geometric indices are used for the evaluation of segmentation. On the one hand, for the evaluation of classification, two thematic indices are extracted from the confusion matrix: overall accuracy (OA) and kappa coefficient (Ka). On the other hand, segmentation performance is evaluated based on manually delineated building objects $O = \{O_1, O_2, \ldots, O_d\}$, where d is the number of reference building objects. Two geometric indices, namely, oversegmentation (OS) and undersegmentation (US) [22], are used for the assessment of segmentation quality (small values indicate better segmentation quality). The OS refers to the subdivision of a single object into several distinct regions. The US indicates the segmentation error that a group of pixels belonging to different objects are merged into a single region. Readers can refer to [22] for the details about the segmentation quality assessment.

III. Experiments

A. Data Sets

The first test image is the HYDICE Washington DC image [12]. It contains 1280 lines and 307 columns, with 191 bands after removal of the water absorption bands. Four spectral

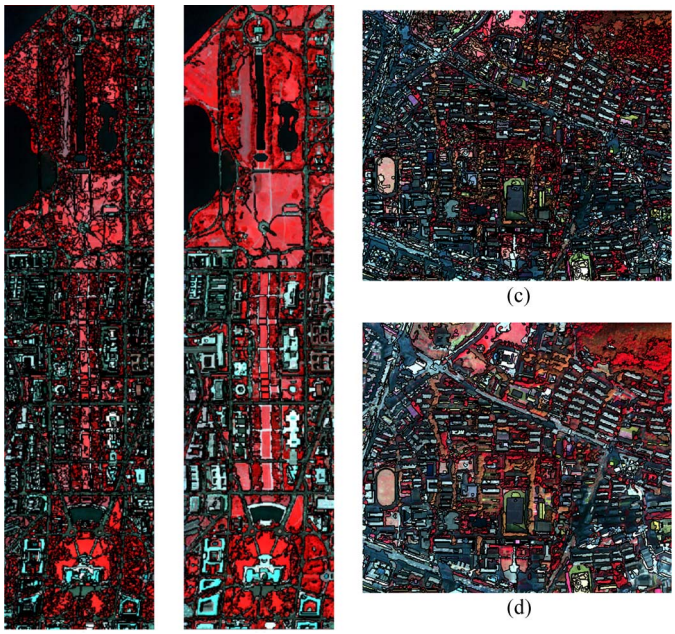

Fig. 3. TV segmentation. (a) and (b) are the boundaries of regions obtained by the TV segmentation of the HYDICE data set, with $E_{\rm th} = 50$ and $E_{\rm th} = 1500$ ($\lambda = 5$), respectively. (c) and (d) are the boundaries of the GeoEye image, with $E_{\rm th} = 30$ and $E_{\rm th} = 500$ ($\lambda = 5$), respectively.

principal components containing over 99% of the image information are used in this experiment. The false-color imagery is shown in Fig. 2(a). The test samples of building objects for the segmentation evaluation are shown in Fig. 2(b). Fig. 2(c) shows the test samples for classification. Fig. 2(d) shows the pixel-based SVM classification map. The second test image is a GeoEye-1 image of Wuhan in central China (acquired on December 22, 2009), with four multispectral bands of 2.0-m spatial resolution [5]. The test samples for segmentation and classification are shown in Fig. 2(f) and (g), respectively. Fig. 2(h) shows the pixel-based SVM classification map. The radial basis function (RBF) function is used as the SVM kernel, and its parameters are tuned manually ($penalty\ parameter = 100$, and the Gamma parameter of the RBF kernel is set to 0.25 in this study).

B. TV Segmentation Performance

Fig. 3 shows the TV segmentation results that are superimposed by the boundaries of regions of the images with different energy parameters ($E_{\rm th}$). It should be noted that the variances of spectral bands of the GeoEye-1 image are smaller than those of the HYDICE image. As a result, the appropriate range of $E_{\rm th}$ for the GeoEye-1 data set is smaller than that for the DC data set. It can be observed that a smaller energy parameter $E_{\rm th}$ [50 for Fig. 3(a) and 30 for Fig. 3(c)] leads to an oversegmented result, where small objects are extracted well but large objects are segmented into several regions. When $E_{\rm th}$ is large [1500 for Fig. 3(b) and 500 for Fig. 3(d)], US is observed. Large objects are extracted well, such as grasslands in Fig. 3(b) and (d). However, small objects are merged into their adjacent objects, e.g., trails in Fig. 3(b) and roofs in Fig. 3(d).

In order to investigate the impact of the parameters of the TV segmentation model, the segmentation and classification results

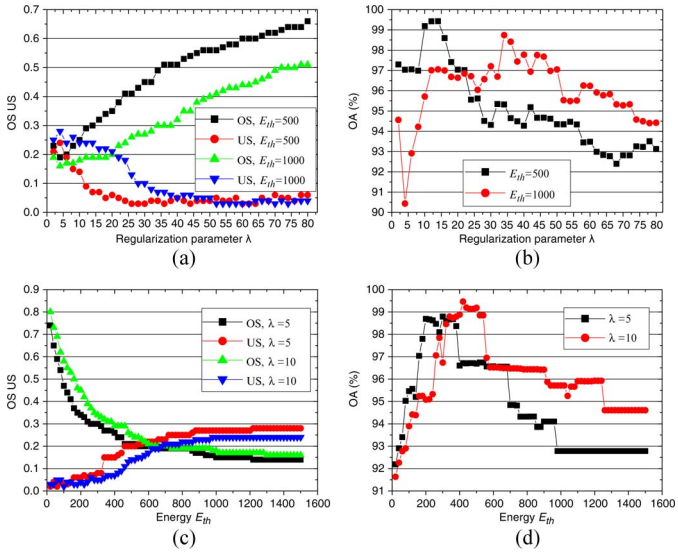

Fig. 4. Evaluation of TV segmentation for HYDICE DC data set: (a) and (b) show the segmentation and classification accuracies for a series of λ with $E_{\rm th} = 500$ and 1000, respectively, and (c) and (d) are the segmentation and classification accuracies for a series of $E_{\rm th}$ with $\lambda = 5$ and 10, respectively.

were assessed using a series of regularization parameters λ and energy thresholds $E_{\rm th}$. Fig. 4 shows the quantitative evaluation of TV segmentation for the DC data set. Fig. 4(a) and (b) shows the impact of λ on the segmentation and classification results, respectively. Fig. 4(c) and (d) shows the impact of $E_{\rm th}$.

From Fig. 4(a), it can be seen that the OS increases and US decreases when the value of λ becomes larger. It can be said that a small value of λ leads to US and a large value leads to OS. From Fig. 4(b), it can be found that there is an optimal value range for λ (in this experiment, from 10 to 15) and the optimal value of λ is small when $E_{\rm th}$ is small. Fig. 4(c) and (d) are related to the energy parameter. It can be seen from Fig. 4(c) that the OS decreases and US increases with the increase of $E_{\rm th}$. Large energy parameters refer to small OS errors but slightly large US errors. Similarly, an optimal parameter value range can be observed for $E_{\rm th}$.

C. Comparison With FNEA, EM, and k-Means

FNEA, which is a widely used object-based segmentation algorithm, is used as a benchmark for the evaluation of the proposed TV-based model. In order to further verify the effectiveness of the proposed method, the clustering-based segmentation methods, such as EM [14] and k-means, are also used for comparison.

Table I shows the quantitative evaluation of different object-based methods for both HYDICE and GeoEye data sets. All the object-based methods, i.e., FNEA, EM, k-means, and TV, show their significant superiority to the pixel-based classification. The improvements in OA and Ka are about 6%–10% and 0.07–0.12 for the HYDICE data set and 1.6%–6% and 0.02–0.07 for the GeoEye data set, respectively. TV + MV outperforms the other object-based classifications such as FNEA + MV, EM + MV, and k-means + MV. The accuracy increments in OA achieved by the TV model are 2%–4% and 1.4%–4.1% for the HYDICE and GeoEye data sets, respectively.

The quantitative indices of segmentation for the HYDICE data set also support the conclusion that the TV model is better

TABLE I
SEGMENTATION AND CLASSIFICATION ACCURACIES OF PIXEL-BASED SVM, FNEA, EM, k-MEANS, AND TV WITH THE HYDICE DC AND GEOEYE WUHAN DATA SETS

Method	Segmentation						Classification				
	HYDICE DC			GeoEye Wuhan				HYDICE DC		GeoEye Wuhan	
	Parameter	OS	US	Parameter	OS	US		OA (%)	Ka	OA (%)	Ka
Pixel-based	NA	NA	NA	NA	NA	NA	SVM	89.49	0.87	85.98	0.82
FNEA	H=30	0.47	0.03	H=15	0.33	0.07	FNEA + MV	95.65	0.95	88.85	0.85
	H=35	0.45	0.04	H=17	0.28	0.08		95.46	0.96	88.26	0.84
	H=40	0.41	0.12	H=20	0.22	0.10		95.45	0.94	88.31	0.85
EM	Cluster=9	0.47	0.07	Cluster=8	0.26	0.08	EM + MV	96.67	0.96	89.66	0.86
K-means	Cluster=11	0.39	0.03	Cluster=13	0.51	0.02	K-means + MV	96.82	0.96	87.60	0.84
TV	λ=5, E_{th}=300	0.3	0.08	λ=10, E_{th}=40	0.48	0.05	TV + MV	98.79	0.99	91.05	0.88
	λ=10, E_{th}=420	0.29	0.08	λ=44, E_{th}=200	0.45	0.06		99.46	0.99	91.71	0.89

than the FNEA algorithm since the OS error of TV is lower than that of FNEA by about 0.10–0.18. Compared with the HYDICE data set, the size of roofs in the GeoEye image is more uniform. As a result, the *OS* error of FNEA segmentation is lower than that of TV since the FNEA algorithm is more appropriate for objects with uniform size.

IV. CONCLUSION

In this letter, an energy-driven TV model is proposed for the segmentation and classification of high spatial resolution remotely sensed imagery. To the best of our knowledge, few studies have reported on TV-based remote-sensing image segmentation and classification. Our experiments show that the TV model is more effective in extracting homogeneous regions and, at the same time, preserves details (i.e., small regions) compared to the commonly used FNEA algorithm and other segmentation techniques. We also analyzed the impact of the key parameters (energy threshold $E_{\rm th}$ and regularization λ) on both the segmentation and classification results. With a small value of λ, the segmentation result of the TV model is not sensitive to spectral value since a small regularization parameter signifies a small contribution to the energy cost function. Future research should focus on adaptive parameter selection since the regularization parameter can be adaptively determined according to the local image structure.

REFERENCES

[1] Q. Jackson and D. A. Landgrebe, "Adaptive Bayesian contextual classification based on Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 11, pp. 2454–2463, Nov. 2002.

[2] J. Tian and D. M. Chen, "Optimization in multi-scale segmentation of high-resolution satellite images for artificial feature recognition," *Int. J. Remote Sens.*, vol. 28, no. 20, pp. 4625–4644, Oct. 2007.

[3] L. Li and N. Shu, "Object-oriented classification of high-resolution remote sensing image using structural feature," in *Proc. 3rd Int. Congr. CISP*, 2010, pp. 2212–2215.

[4] Q. Tan, Q. Wei, and F. Liang, "Building extraction from VHR multispectral images using rule-based object-oriented method: A case study," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2010, pp. 2754–2756.

[5] X. Huang and L. Zhang, "A multidirectional and multiscale morphological index for automatic building extraction from multispectral GeoEye-1 imagery," *Photogramm. Eng. Remote. Sens.*, vol. 77, no. 7, pp. 721–732, Jul. 2011.

[6] T. S. Korting, L. V. Dutra, and L. M. G. Fonseca, "A resegmentation approach for detecting rectangular objects in high-resolution imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 621–625, Jul. 2011.

[7] L. Weizman and J. Goldberger, "Urban-area segmentation using visual words," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 3, pp. 388–392, Jul. 2009.

[8] H. Zheng, X. Bai, and H. Zhao, "A novel approach for satellite image classification using local self-similarity," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2011, pp. 2888–2891.

[9] L. Vincent and P. Soille, "Watersheds in digital spaces: An efficient algorithm based on immersion simulations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 6, pp. 583–598, Jun. 1991.

[10] M. Baatz and A. Sch, "Multiresolution segmentation: An optimization approach for high quality multi-scale image segmentation," in *Proc. Angewandte Geographische Informationsverarbeitung XII. Beiträge zum AGIT-Symp, Salzburg*, Karlsruhe, Germany, 2000, pp. 12–23.

[11] 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.

[12] X. Huang and L. Zhang, "An adaptive mean-shift analysis approach for object extraction and classification from urban hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 12, pp. 4173–4185, Dec. 2008.

[13] J. T. Tou and R. C. González, *Pattern Recognition Principles*. Reading, MA: Addison-Wesley, 1974.

[14] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, "Spectral–spatial classification of hyperspectral imagery based on partitional clustering techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2973–2987, Aug. 2009.

[15] 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.

[16] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenom.*, vol. 60, no. 1–4, pp. 259–268, Nov. 1992.

[17] A. Petrovic, O. D. Escoda, and P. Vandergheynst, "Multiresolution segmentation of natural images: From linear to nonlinear scale-space representations," *IEEE Trans. Image Process.*, vol. 13, no. 8, pp. 1104–1114, Aug. 2004.

[18] A. Petrovic and P. Vandergheynst, An Adaptive Total Variation Model for Image Segmentation, Oct. 2005. [Online]. Available: http://infoscience.epfl.ch/record/87186/files/Petrovic2005_1402.pdf

[19] Y- W. Wang, Y.-F. Wang, Y. Xue, and W. Gao, "A new algorithm for remotely sensed image texture classification and segmentation," *Int. J. Remote Sens.*, vol. 25, no. 19, pp. 4043–4050, Oct. 2004.

[20] C.-C. Chang and C.-J. Lin, LIBSVM: A Library for Support Vector Machines, 2001. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm

[21] A. P. Carleer, O. Debeir, and E. Wolff, "Assessment of very high spatial resolution satellite image segmentations," *Photogramm. Eng. Remote Sens.*, vol. 71, no. 11, pp. 1285–1294, Nov. 2005.

[22] C. Persello and L. Bruzzone, "A novel protocol for accuracy assessment in classification of very high resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 3, pp. 1232–1244, Mar. 2010.

[23] P. Corcoran, A. Winstanley, and P. Mooney, "Segmentation performance evaluation for object-based remotely sensed image analysis," *Int. J. Remote Sens.*, vol. 31, no. 3, pp. 617–645, Apr. 2010.

[24] X. Zhang, P. Xiao, and X. Feng, "An unsupervised evaluation method for remotely sensed imagery segmentation," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 2, pp. 156–160, Mar. 2012.