

Remote Sensing Image Segmentation by Combining Spectral and Texture Features

Jiangye Yuan, DeLiang Wang, *Fellow, IEEE*, and Rongxing Li, *Senior Member, IEEE*

Abstract—We present a new method for remote sensing image segmentation, which utilizes both spectral and texture information. Linear filters are used to provide enhanced spatial patterns. For each pixel location, we compute combined spectral and texture features using local spectral histograms, which concatenate local histograms of all input bands. We regard each feature as a linear combination of several representative features, each of which corresponds to a segment. Segmentation is given by estimating combination weights, which indicate segment ownership of pixels. We present segmentation solutions where representative features are either known or unknown. We also show that feature dimensions can be greatly reduced via subspace projection. The scale issue is investigated, and an algorithm is presented to automatically select proper scales, which does not require segmentation at multiple-scale levels. Experimental results demonstrate the promise of the proposed method.

Index Terms—Segmentation, singular value decomposition (SVD), spectral histogram, texture.

I. INTRODUCTION

AS EARTH observation data are available with increasingly high spatial and spectral resolution, object-based image analysis approaches receives more and more attention in analyzing remote sensing data [1]. In contrast to traditional pixel-based analysis, object-based analysis uses regions or segments of an image as basic units, which has a number of benefits, including reduced spectral variability and more spatial and contextual information such as shape and topological relationships. A key step in object-based analysis is image segmentation, which partitions an image into nonoverlapping regions so that each region is as homogeneous and neighboring ones as different as possible. Segmentation provides building blocks for object-based analysis.

Image segmentation has been extensively studied. In remote sensing, a segmentation method should leverage the advances made in data acquisition, specifically the spectral and spatial

resolution capability. Multispectral (MS) images, which are the main type acquired by remote sensing radiometers, provide much enhanced capabilities of characterizing ground objects. Meanwhile, high-resolution images contain rich texture information, which has been shown to improve segmentation results [2], [3]. Therefore, remote sensing segmentation methods are expected to make use of both spectral and texture information [4]–[6].

It is widely recognized that a visual texture is very difficult to characterize. In remote sensing image analysis, morphological transformations are often employed to deal with texture information [6]–[8]. However, morphological operations have limited forms and, thus, lack the ability to describe complex textures. Semivariograms, which quantify spatial variability, are frequently used for texture analysis in geospatial data [9], [10]. The main drawback of using a semivariogram as a texture descriptor is the high computational cost, which makes it impractical for large images. Recent work on texture analysis shows an emerging consensus that an image should be first convolved with a bank of filters tuned to various orientations and spatial frequencies [11]–[13]. Texture descriptors constructed by analyzing the local distribution of filter responses have been shown to be powerful features for texture synthesis and discrimination.

With such texture descriptors, one can develop a combined spectral–texture segmentation framework by feeding integrated features to clustering approaches to produce segmentation. However, there are two main problems associated with such framework. First, applying multiple filters to spectral bands generates high-dimensional features. As a result, not only is the computational cost high, but many clustering methods also fail to work for high-dimensional data. The second problem stems from texture descriptors generated from the image windows crossing multiple regions [14], which cause difficulty in localizing region boundaries.

In this paper, we use local spectral histogram representation [15], which consists of histograms of filter responses in a local window. This representation provides an effective feature to capture both spectral and texture information. However, as a form of texture descriptors, local spectral histograms also suffer from the problems of high dimensionality and boundary localization. To address these problems, we employ a recently proposed segmentation method [16], which formulates segmentation as multivariate linear regression. This method works across different bands in a computationally efficient way and accurately localizes boundaries.

With remote sensing images, segmentation is inextricably linked to the scale issue. Conceptually, scale is a “window of

Manuscript received February 14, 2012; revised June 25, 2012 and October 8, 2012; accepted December 9, 2012. This work was supported in part by the National Geospatial-Intelligence Agency University Research Initiatives under Grant HM 1582-07-1-2027 and a summer Graduate Research Associate award from the Center for Cognitive Science at The Ohio State University.

J. Yuan is with Oak Ridge National Laboratory, Oak Ridge, TN 37831 USA (e-mail: yuanj@ornl.gov).

D. L. Wang is with the Department of Computer Science and Engineering and the Center for Cognitive Science, The Ohio State University, Columbus, OH 43210 USA (e-mail: dwang@cse.ohio-state.edu).

R. Li is with the Mapping and GIS Laboratory, Department of Civil and Environmental Engineering and Geodetic Science, The Ohio State University, Columbus, OH 43210 USA (e-mail: li.282@osu.edu).

Digital Object Identifier 10.1109/TGRS.2012.2234755

perception” [17]. It is well known that meaningful structures and objects exist over a certain range of scales. In image processing, a scale usually refers to the size of the operators or measurement probes used to extract information from image data. Improper scales can lead to oversegmentation, where segments correspond to portions of regions, or undersegmentation, where one segment contains multiple land-cover classes. Due to the inherent multiscale nature of real-world objects, many multiscale segmentation algorithms have been proposed [8], [18]. However, manual interpretation is typically needed in order to utilize the segmentation results at multiple levels, which inevitably involve subjectivity. Moreover, it has been shown that, in specific cases, single-scale representation might be sufficient and more straightforward [19]. In this paper, we focus on selecting a single scale: Based on our new formulation of the segmentation problem, we propose a scale selection method to appropriately characterize spatial patterns and give a controlled smoothing effect.

The rest of this paper is organized as follows: In Section II, we introduce the local spectral histogram representation. Section III presents our segmentation algorithm in detail. Section IV describes the scale selection method. In Section V, we show experimental results and comparisons. Section VI concludes this paper.

II. COMBINED FEATURES BASED ON LOCAL SPECTRAL HISTOGRAMS

Given an input image window \mathbf{W} and a bank of filters $\{F^{(\alpha)}, \alpha = 1, 2, \dots, K\}$, we can compute a subband image $\mathbf{W}^{(\alpha)}$ for each filter $F^{(\alpha)}$ through convolution. For $\mathbf{W}^{(\alpha)}$, we have the corresponding histogram, which is denoted by $H_{\mathbf{W}}^{(\alpha)}$. In this paper, we use 11 equal-width bins for each filter response. The spectral histogram is defined as the concatenation of the histograms of different filter responses [21], i.e.,

$$H_{\mathbf{W}} = \frac{1}{|\mathbf{W}|} \left(H_{\mathbf{W}}^{(1)}, H_{\mathbf{W}}^{(2)}, \dots, H_{\mathbf{W}}^{(K)} \right) \quad (1)$$

where $|\cdot|$ denotes cardinality. The size of the window is referred to as the integration scale. The spectral histogram characterizes both local patterns via filtering and global patterns through a histogram. It has been shown that, with properly selected filters, the spectral histogram is sufficient to capture texture appearance [15].

A local spectral histogram is computed over a window centered at a pixel location, which is essentially a feature vector consisting of local distributions of filter responses. Since the integration scale has to be large enough to obtain meaningful features, it is computationally expensive to compute all the local histograms. The integral histogram approach [20] can speed up this process considerably. With integral histogram computed for each subband image, we can obtain any local spectral histogram via three vector arithmetic operations regardless of integration scales. As a result, computing local spectral histograms at all pixel locations takes linear time with respect to the pixel number. A detailed description of the fast approach can be found in [21].

To specify a spectral histogram, one needs to choose a set of filters. Based on previous studies, three types of filters are used in this paper, namely, the intensity filter, Laplacian of Gaussian (LoG) filters, and Gabor filters. The intensity filter is the Dirac delta function, which gives the intensity value at a pixel location. LoG filters are given by

$$\text{LoG}(x, y|\sigma) = (x^2 + y^2 - 2\sigma^2)e^{-\frac{x^2+y^2}{2\sigma^2}}. \quad (2)$$

We use an even-symmetric Gabor filter, which has the following form:

$$\begin{aligned} \text{Gabor}(x, y|\sigma, \theta) = & e^{-\frac{1}{2\sigma^2}[(x \cos \theta + y \sin \theta)^2 + (-x \sin \theta + y \cos \theta)^2]} \\ & \times \cos \left[\frac{2\pi}{\lambda}(x \cos \theta + y \sin \theta) \right] \end{aligned} \quad (3)$$

where θ defines the orientation of the filter, and the ratio σ/λ is set to 0.5. For both types of filters, σ determines the scale.

Local spectral histograms are capable of capturing both spectral and texture information for remote sensing images. We apply such filters to each spectral band, where the intensity filter gives spectral intensities, and other linear filters generate subband images that enhance certain spatial structures. Local spectral histograms are computed from local windows across all the bands, which collectively define a region appearance based on spectral and spatial properties.

Due to limits in the radiation energy received by the sensor and the data storage capacity [22], it is difficult to acquire satellite images with high resolutions in both spectral and spatial domains. To mitigate this problem, most Earth observation satellites, such as SPOT, IKONOS, and QuickBird, provide both a panchromatic (PAN) image at a higher spatial resolution and MS images at a lower spatial resolution. To accommodate multiresolution data, we use the following approach to produce features at the PAN image resolution. The MS images are first resampled to the size of the PAN image using bicubic interpolation, and they are then combined with the PAN image to generate local spectral histograms. Despite less detail, the resized images preserve basic shapes of local distributions, which are essential for the discriminative power of spectral histograms. An alternative is to use pan-sharpening techniques [22], which integrate two types of data to produce high-resolution MS images. However, as spectral histograms are built on local distributions, sharpened images, which give more detailed distributions, can make local features less similar to neighboring ones. Therefore, our method does not use pan-sharpening to deal with multiresolution data.

III. SEGMENTATION ALGORITHM

Here, we present the algorithm that utilizes local spectral histogram features to produce accurate segmentation. We refer the reader to [16] for a more detailed analysis.

A. Segmentation Using Linear Regression

Without loss of generality, let us consider an image consisting of three agricultural fields placed in a way shown in

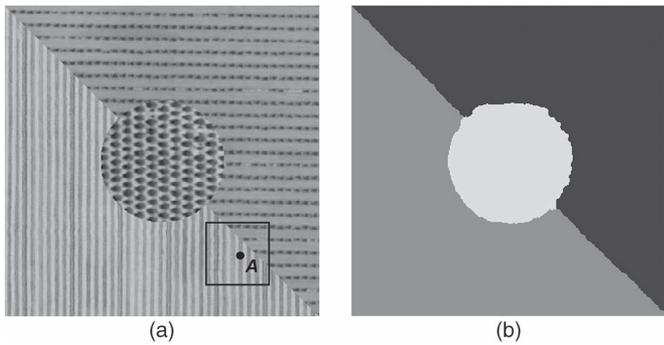


Fig. 1. Image segmentation via linear regression. (a) Synthesized image with size 320×320 . The feature at pixel A can be approximated by the weighted sum of two neighboring representative features. (b) Segmentation result using least squares estimation.

Fig. 1(a). The local spectral histogram at a pixel location A is computed using a square window that crosses two regions. As discussed in Section I, this feature does not carry discriminative information, and thus, the corresponding pixel is difficult to correctly classify.

We assume that local spectral histograms within a homogeneous region are approximately constant. We then have a representative feature for each region. Let us consider only intensity filters for the time being, which simplifies local spectral histograms to local histograms of pixel intensity. The local histogram of pixel A can be approximated by a linear combination of two histograms representing two neighboring regions, and the combination weights correspond to the area coverage within the window. Hence, we can assign this pixel to the region whose histogram has a larger weight. Such a linear relationship between a boundary feature and representative features holds for other filter responses, except when the scales of filters are very large, which can significantly distort histograms near the boundaries. Since filtering in spectral histograms aims to capture basic spatial patterns, the use of large-scale filters is discouraged. As a result, although a filter may have strong responses to region boundaries, it does not have a significant effect on the local spectral histograms, which are computed from a much larger local window.

By extending the above analysis, each feature in an image can be regarded as a linear combination of all representative features weighted by the fractional area coverage in the local window. Given an image with N pixels, M -dimensional features at each pixel, and L representative features, we use a multivariate linear regression model to associate each feature to the representative features, which is expressed as

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (4)$$

where \mathbf{Y} is an $M \times N$ matrix with each column representing a feature at a pixel location, \mathbf{Z} is an $M \times L$ matrix containing L representative features, $\boldsymbol{\beta}$ is an $L \times N$ matrix containing combination weights for N pixels, and $\boldsymbol{\varepsilon}$ is an $M \times N$ matrix representing noise.

Representative features can be computed by manually selecting seeds within each region. Given the feature matrix \mathbf{Y} and the representative feature set \mathbf{Z} , we seek to estimate $\boldsymbol{\beta}$ that best models the relationship between the feature matrix and the

representative features. This can be easily solved by the least squares estimation, i.e.,

$$\hat{\boldsymbol{\beta}} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{Y}. \quad (5)$$

The segmentation result is immediately given by $\hat{\boldsymbol{\beta}}$, where the largest weight in each column indicates the segment ownership of the corresponding pixel. It is worth noting that in the solution, all the dimensions of the features are involved to estimate the weights, i.e., all the bands are taken into account to produce segmentation. Fig. 1(b) shows the segmentation result of the image in Fig. 1(a), where each segment is displayed using a distinct gray value. We can see that different regions are separated with rather accurate boundary localization. Here, we use five filters, namely, the intensity filter, LoG (0.5), LoG (1.0), Gabor(1.5, 0°), and Gabor(1.5, 90°).

The above image model bears some resemblance to the linear mixing model used to analyze hyperspectral imagery [23], which assumes that an observed spectrum is the linear combination of endmember spectrum weighted by the correspondent abundance fractions. However, there are major differences. The mixture in the linear mixing model is the measured spectrum of each pixel location. In contrast, the mixture in our model is the local spectral histogram extracted from an image patch. More importantly, the purpose of decomposing a mixed pixel in hyperspectral imagery is to identify material constituents at the pixel location, whereas we decompose local spectral histograms in order to locate region boundaries.

B. Segmentation Algorithm With Dimensionality Reduction

The image model presented above gives a semiautomatic segmentation method, where a human operator needs to choose seeds to obtain representative features. The choice of seeds can be subjective and reliant on expert knowledge. To address this problem, an unsupervised method is presented to estimate representative features. Along a similar line, features are projected onto low-dimensional space, which reduces noise and speeds up the algorithm significantly.

To ensure a unique solution in (5), \mathbf{Z} needs to have the full column rank so that $(\mathbf{Z}^T \mathbf{Z})^{-1}$ is not singular. Hence, representative features have to be linearly independent. Here, we assume that \mathbf{Z} has more rows than columns. That is, the number of feature dimensions is larger than the number of representative features. This assumption generally holds because the number of feature dimensions corresponds to the product of the number of filters, the number of spectral bands, and the number of bins and, hence, tends to be relatively large. Since each feature is a linear combination of representative features, the rank of feature matrix \mathbf{Y} should be the same as the rank of \mathbf{Z} , i.e., the number of representative features. However, due to image noise, \mathbf{Y} tends to be of full rank. Due to the Eckart–Young theorem [24], we can estimate the underlying feature matrix with the expected rank.

The low-rank approximation is based on singular value decomposition (SVD). The feature matrix can be decomposed in the following form:

$$\mathbf{Y} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T \quad (6)$$

where \mathbf{U} is an $M \times M$ matrix with each column representing an eigenvector of $\mathbf{Y}\mathbf{Y}^T$, and \mathbf{V} is an $N \times N$ matrix with each column representing an eigenvector of $\mathbf{Y}^T\mathbf{Y}$. Σ is an $M \times N$ rectangular diagonal matrix with nonnegative real numbers on its diagonal and zero elsewhere. The diagonal elements are called singular values, which are the square roots of the eigenvalues of $\mathbf{Y}\mathbf{Y}^T$, or $\mathbf{Y}^T\mathbf{Y}$.

Assume r representative features. According to the Eckart–Young theorem, the best least squares rank- r approximation to \mathbf{Y} has the same form as (6), except that the smallest $M - r$ singular values are replaced by zeros. Thus, we can write the approximated matrix as

$$\mathbf{Y}' = \mathbf{U}'\Sigma'\mathbf{V}'^T \quad (7)$$

where \mathbf{U}' and \mathbf{V}' consist of the first r columns of \mathbf{U} and \mathbf{V} in (6), respectively. Σ' is an $r \times r$ matrix with the largest r singular values on the diagonal.

Let $\beta' = \Sigma'\mathbf{V}'^T$. We note that \mathbf{U}' and β' have the same sizes as \mathbf{Z} and β , respectively, which gives the minimum least squares error guaranteed by the Eckart–Young theorem. Hence, \mathbf{U}' and β' can be the solution for the linear regression model in (4). However, the decomposition is not unique because of the following fact:

$$\mathbf{Y}' = \mathbf{U}'\beta' = \mathbf{U}'\mathbf{Q}\mathbf{Q}^{-1}\beta' \quad (8)$$

where \mathbf{Q} is any invertible square matrix, and thus, $(\mathbf{U}'\mathbf{Q})$ and $(\mathbf{Q}^{-1}\beta')$ can be another possible solution.

Although the above analysis cannot immediately give the final solution, it reveals that the representative features should be \mathbf{U}' multiplied by some \mathbf{Q} , i.e., some linear transformation of \mathbf{U}' . In other words, the representative features should lie in an r -dimensional subspace spanned by the columns of \mathbf{U}' , which are orthogonal vectors. Note that, in general, the original feature dimension M is much larger than r , which is the number of representative features or segments. Consequently, we can project all features onto the subspace, which greatly reduce the feature dimension and remove the noise outside the subspace. For example, if three filters are applied to a four-band MS image, the dimension of features is 132 (11-bin histograms are used). Subspace projection can reduce the feature dimension to the number of representative features, which is typically smaller than 10. Because features inside homogeneous regions are close to the representative features, we apply k -means clustering to the features in the subspace, where the cluster centers correspond to the representative features. The representative features can be easily obtained by projecting the cluster centers back to the original space.

A problem can occur when performing k -means clustering. The features near boundaries can form small clusters, which make clustering very sensitive to initialization. To address this problem, we measure edgeness for each pixel and discard the features of pixels with high edgeness before clustering. Specifically, at pixel location (x, y) , we compute the feature distance between pixel locations $(x + d, y)$ and $(x - d, y)$ and that between $(x, y + d)$ and $(x, y - d)$, where d is half of the side length of \mathbf{W} . Edgeness is the sum of the two distances. Since features are in a low-dimensional space and

generally well grouped, we use the Euclidean distance as the distance metric for both edgeness measurement and k -means clustering.

IV. AUTOMATIC SCALE SELECTION

Local spectral histograms involve two types of scale parameters, namely, filter scales and integration scales, both of which have an influence on segmentation results. With proper filter scales, spatial patterns can be enhanced, which is important for characterizing region appearances. Integration scales need to be sufficiently large to capture meaningful features, but too large scales will result in overly smooth segmentation [16]. Without any prior knowledge, it is a challenging problem to find scales that lead to optimal segmentation results. Here, we tackle this problem by studying singular values of a feature matrix, which does not require segmentation at different scale levels.

In the low-rank approximation, the approximation error is related to the singular values of the original matrix in the following way:

$$\|\mathbf{Y} - \mathbf{Y}'_r\| = \sqrt{\sum_{i=r+1}^M \sigma_i^2} \quad (9)$$

where \mathbf{Y}'_r is the rank- r approximation of matrix \mathbf{Y} , $\|\cdot\|$ denotes the Frobenius norm¹, and $\sigma_1, \sigma_2, \dots, \sigma_M$ are singular values in a nonincreasing order. Thus, we have

$$\sigma_r^2 = \sum_{i=r}^M \sigma_i^2 - \sum_{i=r+1}^M \sigma_i^2 = \|\mathbf{Y} - \mathbf{Y}'_{r-1}\|^2 - \|\mathbf{Y} - \mathbf{Y}'_r\|^2. \quad (10)$$

That is, σ_r^2 is equal to the reduced approximation error by increasing the rank from $r - 1$ to r . Likewise, σ_{r+1}^2 is equal to the reduced error by increasing the rank from r to $r + 1$. Assume that the number of representative features is r . σ_r should be as large as possible so that the feature matrix cannot be well approximated by a rank- $(r - 1)$ matrix, whereas σ_{r+1} should be as small as possible because it corresponds to noise. Hence, we select the scales based on these two singular values.

We first determine filter scales. For a rigorous solution, we need to select both filter types and filter scales, which give rise to a very large search space. To make the problem more tractable, we assume that for an image, a filterbank is known to be sufficient to discriminate texture appearances, where the filter scales are the only tunable parameters. Corresponding filters in two filterbanks have scales of the same ratio; hence, all filter scales in the filterbank depend on a single value. We compute feature matrices for preselected scales $s_n = \gamma^n s_0$, where γ is the ratio between successive scale levels. s_0 and γ are set to 0.5 and 1.2, respectively, in our experiments. A fixed integration scale that is equal to the largest filter scale is used for all scale levels. We use the ratio σ_r / σ_{r+1} of a feature matrix as an indicator, and the scale with the maximum ratio is chosen.

¹Frobenius norm is defined as the square root of the sum of the squares of all matrix entries.

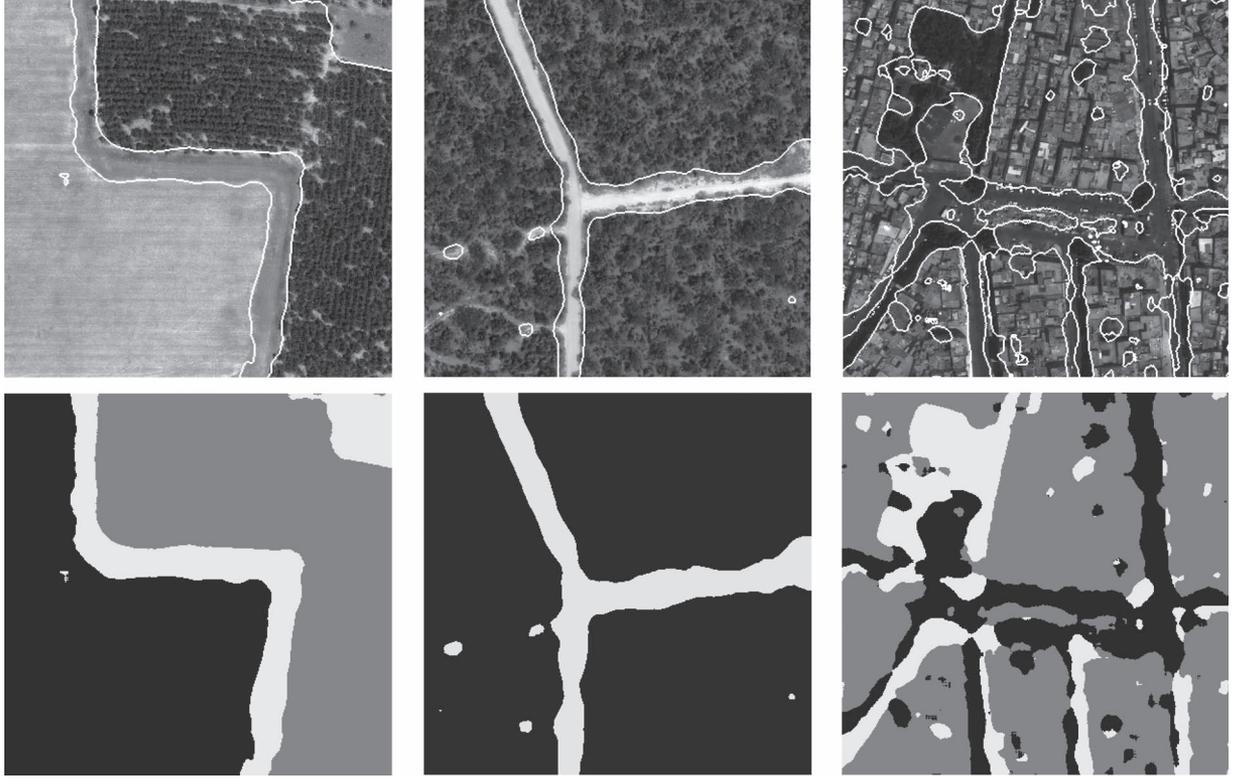


Fig. 2. Experiments on three GeoEye-1 images with size 400×400 . In each column, the top is the red band of images with segment boundaries marked in white, and the bottom is region labeling, where different segments correspond to different gray levels.

When selecting filter scales, the fixed integration scale is generally larger than necessary. After the filter scales are determined, we decrease the integration scale to diminish the smoothing effect. As the integration scale decreases, features tend to be less similar to the neighboring features, which increases the noise level and, thus, reduces the gap between σ_r and σ_{r+1} . We again examine the ratio σ_r / σ_{r+1} . For an integration scale h , a ratio can be computed, which is denoted by R_h . The integration scale is determined by

$$h = \max\{h : R_h < \omega\} \quad (11)$$

where ω is a threshold. Although ω can be tuned to achieve the desired smoothness of segmentation results, in our experiments, a fixed value of 1.8 is used, and it works well.

V. EXPERIMENTAL RESULTS AND COMPARISONS

We first test our method on a set of GeoEye-1 images with a spatial resolution of 0.5 m. The images have three bands (red, green, and blue), and the red band is shown in the top row in Fig. 2. For each image, we use three filters: the intensity filter, $\text{LoG}(s)$, and $\text{LoG}(2s)$. Filter scale s and the integration scale are determined using the automatic method described in Section IV. The only free parameter is the number of segments, which is set to 3, 2, and 3, respectively, for the three images. The results are presented in Fig. 2, where the top row shows the segment boundaries overlaid on the red-band images, and the bottom row shows region labeling. In the left and middle images, vegetation areas with heavy texture are well separated

from other areas. The third image contains an urban residential area, which is challenging for segmentation due to high spectral variance and irregular patterns. In our result, the main areas, including buildings, roads, and vegetation, are clearly identifiable. Fig. 3 illustrates the estimated representative feature corresponding to the building segment and five local spectral histograms within building areas. For this image, the chosen integration scale is 19×19 . It can be seen that the representative feature can describe the features from the same region. We can observe that many inaccurate boundaries occur in shadow areas. Addressing this problem would require prior knowledge. These as well as other results not shown demonstrate that the proposed method is successful in segmenting remote sensing images based on spectral and texture features.

To systematically evaluate the performance of the proposed method, we conduct experiments on an IKONOS image set covering a complex scene, i.e., a $2 \text{ km} \times 2 \text{ km}$ section of the city of San Diego (CA) [25]. The data set includes a PAN image of size 2004×2004 , which is commensurate with the 1-m spatial resolution, and four-band (red, green, blue, and near infrared) MS images of size 501×501 . The PAN image is shown in Fig. 4(a). This two-resolution data set is processed using the approach described in Section II, which first resizes the low-resolution images and then use all the bands as input. We choose a filterbank consisting of five filters, namely, the intensity filter, $\text{LoG}(s)$, $\text{LoG}(2s)$, $\text{Gabor}(s, 0^\circ)$, and $\text{Gabor}(s, 90^\circ)$. Filter scale s and the integration scale are automatically selected. Because the PAN image contains the most detailed spatial variations, only the PAN image is convolved with LoG and Gabor filters to provide texture bands, for reducing memory

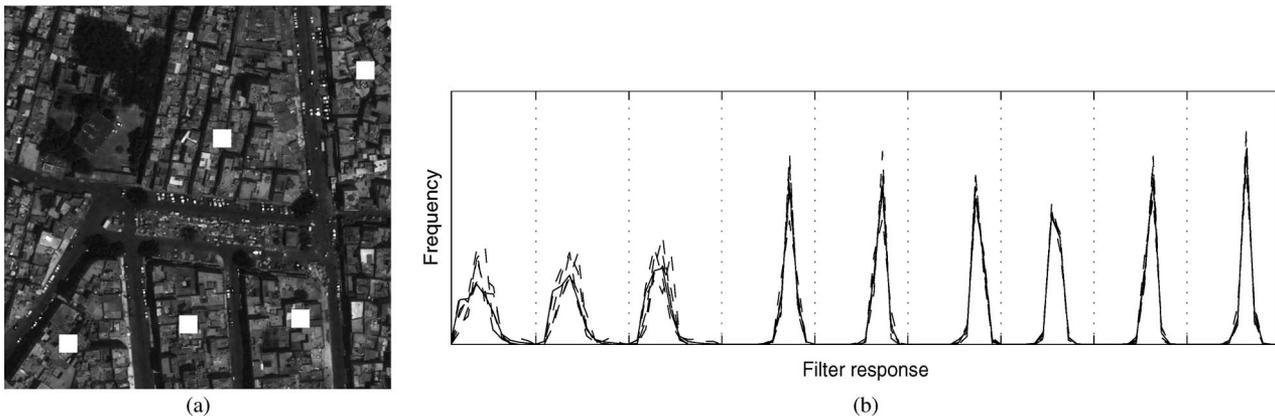


Fig. 3. Illustration of local spectral histograms. (a) The third image in Fig. 2. White boxes indicate the windows for computing local spectral histograms. (b) The solid line stands for the representative feature corresponding to the building segment, and dashed lines stand for the local spectral histograms computed from the windows in Fig. 3(a). The spectral histogram includes the responses of nine filters. (The intensity filter and two LoG filters are applied to each of the three bands.) Their corresponding histograms are separated by dotted lines.

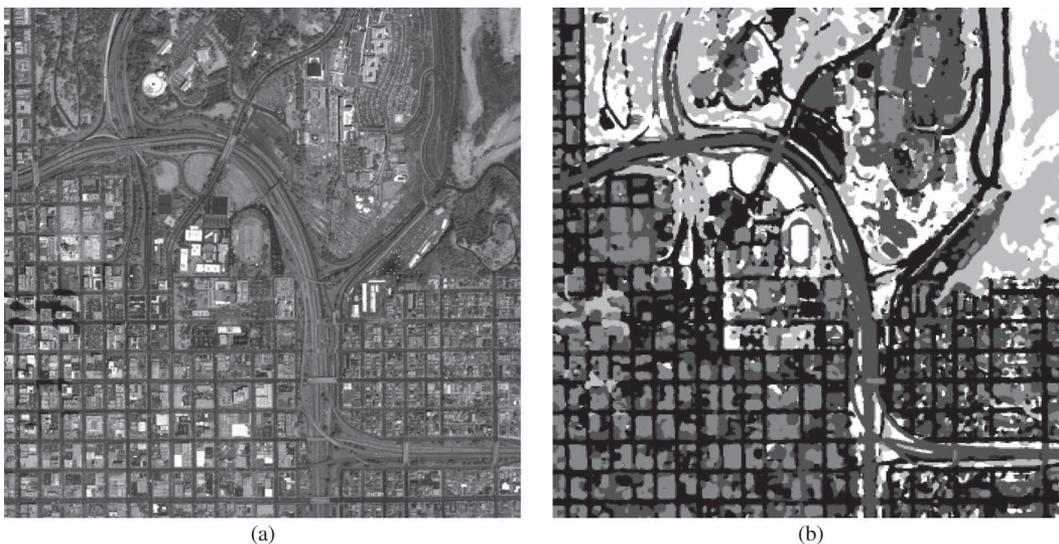


Fig. 4. Experiments on IKONOS imagery. (a) A 1-m resolution PAN image with size 2004×2004 . (b) Segmentation result using the proposed method.

requirements and computational time. Based on preliminary trials, we set the number to 5, corresponding to the five major land-cover types: roads, parking lots, buildings, trees, and grass. The segmentation result is displayed in Fig. 4(b), where each gray level indicates a distinct segment. As we can see, the major areas of the image are clearly visible.

We compare our method with a recent segmentation method, which is based on a hierarchical multiple Markov chain (H-MMC) model [25]. This method first segments an image into elementary regions with similar spatial and spectral characteristics and then produces a sequence of nested segmentation maps through merging the regions. The method is tested on the same data set, which gives good results. For quantitative evaluation, Gaetano *et al.* [25] generated the ground truth by visual inspection. The original ground truth contains seven classes: roads, large buildings, parking lots, small buildings, green spots, grass, and trees. Since our method does not take object sizes into consideration, we do not differentiate between large buildings and small buildings or green spots and grass.

Thus, we merge seven classes in the ground truth into five classes.

By matching the resulting segments with those in the ground truth, we obtain a confusion matrix shown in Table I. Column labels correspond to machine-generated segments, and row labels correspond to classes in the ground truth. For example, the cell in column 1 row 4 indicates that 3185 pixels in the road segment correspond to (incorrectly) trees in the ground truth. Diagonal entries give correctly labeled pixel numbers. Each entry in the last column shows the corresponding diagonal entry divided by the total pixel number in the corresponding row, indicating how complete a particular class is identified. The last row shows diagonal entries divided by the total pixel numbers in the corresponding columns, indicating the correctness rate for one segment. We also present the five-class result from the H-MMC method,² shown in parentheses; for conciseness, we

²The seven-class confusion matrix is reported in [25], from which the five-class one can be easily obtained.

TABLE I
CONFUSION MATRIX FOR FIVE-CLASS SEGMENTATION FROM THE PROPOSED METHOD AND THE H-MMC METHOD SHOWN IN PARENTHESES

| | Roads | Parking lots | Buildings | Trees | Grass | Completeness |
|--------------|--------------------|--------------------|--------------------|--------------------|--------------------|------------------|
| Roads | 382078 (269471) | 151893 | 8923 | 17212 | 1500 | 68.0% |
| Parking lots | 147565 | 114856 (109289) | 20680 | 15798 | 400 | 38.4% |
| Buildings | 49052 | 153660 | 320034 (371321) | 5246 | 1742 | 60.4% |
| Trees | 3185 | 1117 | 168 | 266400 (279043) | 46391 | 84.0% |
| Grass | 8889 | 14904 | 1337 | 16999 | 140097 (123650) | 76.9% |
| Correctness | 64.7% | 26.3% | 91.1% | 82.8% | 73.7% | 64.7% (61.0%) |

TABLE II
OVERALL ACCURACY OF SEGMENTATION RESULTS USING DIFFERENT SUBSETS OF BANDS

| Excluded band | PAN | Red | Green | Blue | Near infrared | LoG | Gabor | LoG + Gabor |
|---------------|--------|--------|--------|--------|---------------|--------|--------|-------------|
| Accuracy | 61.09% | 62.21% | 62.61% | 62.46% | 62.85% | 63.56% | 63.93% | 63.25% |

only show the number of correctly segmented pixels and the overall accuracy. As we can see, our method attains a higher overall accuracy (64.7% versus 61.0%). The accuracy values for roads and grass are particularly high, which have uniform spectral and texture characteristics and, thus, are well segmented by our method.

To verify that the better accuracy provided by the proposed method is statistically significant, we conduct a statistical significance test. We divide our segmentation result into 36 equal-sized blocks and compute the accuracy rate of each block. The null hypothesis is that the accuracy rate of the H-MMC method (61.0%) is the mean of the 36 accuracy rates. We obtain the p -value of 0.001, which is much smaller than 0.05 (5% significance level), providing strong evidence for rejecting the null hypothesis. Therefore, the increased accuracy rate from the proposed method is statistically significant.

Our method produces segmentation based on spectral and texture features. The results can be easily improved by incorporating additional information (e.g., shape and context). For example, in our result, a large number of confusions occur between parking lots and roads because of similar spectral and texture features. Particularly, the highways in the middle of the image, although well segmented, have the same segment labels as the areas mostly corresponding to parking lots. This can be corrected by a simple technique, which is to compute the medial axis of each connected segment and assign those with long medial axes and small associated radii to the road class [26]. By doing so, the overall accuracy rate is boosted to 75.1%, showing substantial improvement over the H-MMC method.

To show whether each spectral and texture band contributes to the final segmentation, we exclude certain bands and apply the same method. The overall accuracy rates in different cases are summarized in Table II. In all cases, the accuracy is lower than that when using all the bands. This result confirms that our

method makes effective use of the information from all bands to produce segmentation.

As for computational time, the feature extraction step in our algorithm, including filtering and spectral histogram computation, takes linear time with respect to the number of pixels. In our case, we do not need to perform complete SVD. After the eigenvalue decomposition of $\mathbf{Y}\mathbf{Y}^T$, which is an $M \times M$ matrix (M is the feature dimension), we only need the first several eigenvalues and the corresponding eigenvectors to construct \mathbf{U}' . Estimating representative features using k -means clustering is also fast because the features are projected onto a low-dimensional subspace, and the clusters are well grouped. Scale selection needs to compute feature matrices and corresponding SVD at different scale levels, which can be sped up by using subsampled pixel locations. We implement the whole system using MATLAB, and the experiment on the IKONOS image set takes 150 s on a 2.6-GHz Intel processor.

VI. CONCLUSION

We have presented a new method for segmenting remote sensing images based on spectral and texture features. We use local spectral histograms to provide combined features. By regarding each feature as a linear combination of several representative features, we formulate the segmentation problem as a multivariate linear regression, which can be solved by least squares estimation. We have also proposed methods based on SVD to automatically estimate representative features and select proper scales. The experimental results on different image sets are encouraging.

The performance of our method can degrade when choosing a large segment number to deal with a complex scene. The reason is that with more representative features, the least squares solution is more sensitive to noise, and the estimated

combination weights might not correspond to actual coverage fractions. In most cases, this problem can be tackled by a divide-and-conquer strategy: We divide an image into blocks containing fewer regions; segmentation of each block is individually generated, and the final result is obtained by merging the segments with similar representative features. However, such a strategy would fail if a block still contains a multitude of regions. A more general approach is to impose constraints on a least squares solution, making it more robust to noise. We are currently investigating this issue.

ACKNOWLEDGMENT

We would like to thank the Image Processing Research Group, University of Naples Federico II, Naples, Italy, for providing us with the IKONOS imagery and the ground truth. This manuscript has been authored by employees of UT-Battelle, LLC, Oak Ridge, TN, USA, under Contract DE-AC05-00OR22725 with the U.S. Department of Energy. Accordingly, the United States Government retains, and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a nonexclusive, paid-up, irrevocable, and worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for the United States Government's purposes.

REFERENCES

- [1] T. Blaschke, "Object based image analysis for remote sensing," *ISPRS J. Photogramm. Remote Sens.*, vol. 65, no. 1, pp. 2–16, Jan. 2010.
- [2] S. Ryherd and C. Woodcock, "Combining spectral and texture data in the segmentation of remotely sensed images," *Photogramm. Eng. Remote Sens.*, vol. 62, no. 2, pp. 181–194, Feb. 1996.
- [3] M. Kim, M. Madden, and B. Xu, "GEOBIA vegetation mapping in Great Smoky Mountains National Park with spectral and non-spectral ancillary information," *Photogramm. Eng. Remote Sens.*, vol. 76, no. 2, pp. 137–149, 2010.
- [4] X. Hu, C. V. Tao, and B. Prenzel, "Automatic segmentation of high-resolution satellite imagery by integrating texture, intensity and color features," *Photogramm. Eng. Remote Sens.*, vol. 71, no. 12, pp. 1399–1406, 2005.
- [5] R. Trias-Sanz, G. Stamon, and J. Louchet, "Using colour, texture, and hierarchical segmentation for high-resolution remote sensing," *ISPRS J. Photogramm. Remote Sens.*, vol. 63, no. 2, pp. 156–168, Mar. 2008.
- [6] H. G. Akcay 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.
- [7] J. A. Benediktsson, M. Pesaresi, and K. Arnason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 1940–1949, Sep. 2003.
- [8] A. Tzotsos, K. Karantzalos, and D. Argialas, "Object-based image analysis through nonlinear scale-space filtering," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 1, pp. 2–16, Jan. 2011.
- [9] C. Coburn and A. C. B. Roberts, "A multiscale texture analysis procedure for improved forest stand classification," *Int. J. Remote Sens.*, vol. 25, no. 20, pp. 4287–4308, 2004.
- [10] J. R. Carr and F. P. Miranda, "The semivariogram in comparison to the co-occurrence matrix for classification of image texture," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 6, pp. 1945–1952, Nov. 1998.
- [11] N. Li, H. Huo, and T. Fang, "A novel texture-preceded segmentation algorithm for high-resolution imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 7, pp. 2818–2828, Jul. 2010.
- [12] D. Martin, C. Fowlkes, and J. Malik, "Learning to detect natural image boundaries using local brightness, color, and texture cues," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 5, pp. 530–549, May 2004.
- [13] S. K. Meher, B. Uma Shankar, and A. Ghosh, "Wavelet-feature-based classifiers for multispectral remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1881–1886, Jun. 2007.
- [14] J. Malik, S. Belongie, T. Leung, and J. Shi, "Contour and texture analysis for image segmentation," *Int. J. Comput. Vis.*, vol. 43, no. 1, pp. 7–27, Jun. 2001.
- [15] X. Liu and D. L. Wang, "A spectral histogram model for texton modeling and texture discrimination," *Vis. Res.*, vol. 42, no. 23, pp. 2617–2637, Oct. 2002.
- [16] J. Yuan, D. L. Wang, and R. Li, "Image segmentation using local spectral histograms and linear regression," *Pattern Recognit. Lett.*, vol. 33, no. 5, pp. 615–622, Apr. 2012.
- [17] D. J. Marceau and G. J. Hay, "Remote sensing contribution to scale issue," *Can. J. Remote Sens.*, vol. 25, no. 4, pp. 357–366, 1999.
- [18] B. Johnson and Z. Xie, "Unsupervised image segmentation evaluation and refinement using a multi-scale approach," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 4, pp. 473–483, Jul. 2011.
- [19] S. Lang and T. Langanke, "Object-based mapping and object-relationship modeling for land use classes and habitats," *Photogramm., Fernerkundung, Geoinf.*, vol. 10, no. 1, pp. 5–18, 2006.
- [20] F. Porikli, "Integral histogram: A fast way to extract histograms in Cartesian spaces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2005, pp. 829–836.
- [21] X. Liu and D. L. Wang, "Image and texture segmentation using local spectral histograms," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 3066–3077, Oct. 2006.
- [22] Y. Zhang, "Understanding image fusion," *Photogramm. Eng. Remote Sens.*, vol. 59, no. 1, pp. 657–661, Jun. 2004.
- [23] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.
- [24] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218, Mar. 1936.
- [25] 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.
- [26] J. Yuan, D. L. Wang, B. Wu, L. Yan, and R. Li, "LEGION-based automatic road extraction from satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4528–4538, Nov. 2011.



Jiangeye Yuan received the M.S. degree in computer science and engineering and the Ph.D. degree in geodetic science from The Ohio State University, Columbus, USA, in 2009 and 2012, respectively.

He is currently a Postdoctoral Researcher with Oak Ridge National Laboratory, Oak Ridge, TN. His research interests include image segmentation, neural networks, and pattern recognition with applications in remote sensing and geospatial analysis.



DeLiang Wang (M'90–SM'01–F'04) received the B.S. and M.S. degrees from Peking University, Beijing, China, in 1983 and 1986, respectively, and the Ph.D. degree from the University of Southern California, Los Angeles, USA, in 1991, all in computer science.

From July 1986 to December 1987, he was with the Institute of Computing Technology, Academia Sinica, Beijing. Since 1991, he has been with the Department of Computer Science and Engineering and the Center for Cognitive Science, The Ohio State University, Columbus, USA, where he is a Professor. From October 1998 to September 1999, he was a Visiting Scholar with the Department of Psychology, Harvard University, Cambridge, MA, USA. From October 2006 to June 2007, he was a Visiting Scholar with Oticon A/S, Denmark. His research interests include machine perception and neurodynamics.

Dr. Wang received the Office of Naval Research Young Investigator Award in 1996, the 2005 Outstanding Paper Award from the IEEE TRANSACTIONS ON NEURAL NETWORKS, and the 2008 Helmholtz Award from the International Neural Network Society. He is a Coeditor-in-Chief of Neural Networks.



Rongxing (Ron) Li (M'09–SM'11) received the B.S. and M.S. degrees in surveying engineering from Tongji University, Shanghai, China, in 1982 and 1984, respectively, and the Ph.D. degree in photogrammetry and remote sensing from the Technical University of Berlin, Berlin, Germany, in 1990.

From 1994 to 1996, he was an Assistant and Associate Professor with the Department of Geomatics Engineering, The University of Calgary, Calgary, AB, Canada. In 1996, he moved to The Ohio State University, Columbus, USA, where he has been a Full Professor since 2001. He is a Participating Scientist of the Mars Exploration Rover and the Lunar Reconnaissance Orbiter National Aeronautics and Space Administration (NASA) missions and a Member of the Mars Exploration Program Geodesy and Cartography Working Group. In addition, he is a Science Team Member of the Panoramic Camera (PanCam) system of the European Space Agency ExoMars mission. His research interests include photogrammetric engineering and geomatics applications, planetary exploration, digital mapping, spatial data structures, and coastal and marine GIS.

Dr. Li is an American Society of Civil Engineering Fellow. He received numerous academic and professional awards, including several NASA Group Achievement Awards, the Pacific Congress on Marine Science and Technologies Service Award, and best paper awards from the International Society for Photogrammetry and Remote Sensing, the American Society of Photogrammetry and Remote Sensing, and Chinese Professionals of Geographic Information System. In addition, he was a recipient of the Duane C. Brown Photogrammetry Award, the Lumley Research Award, and the Innovators Award presented by The Ohio State University. He is the Editor-in-Chief of the international journal *Marine Geodesy*.