

# Morphological Segmentation of Urban Structures

H. Gökhan Akçay and Selim Aksoy  
Department of Computer Engineering  
Bilkent University  
Bilkent, 06800, Ankara, Turkey  
{saksoy,akcay}@cs.bilkent.edu.tr

**Abstract**— Automatic segmentation of high-resolution remote sensing imagery is an important problem in urban applications because the resulting segmentations can provide valuable spatial and structural information that are complementary to pixel-based spectral information in classification. We present a method that combines structural information extracted by morphological processing with spectral information summarized using principal components analysis to produce precise segmentations that are also robust to noise. First, principal components are computed from hyper-spectral data to obtain representative bands. Then, candidate regions are extracted by applying connected components analysis to the pixels selected according to their morphological profiles computed using opening and closing by reconstruction with increasing structuring element sizes. Next, these regions are represented using a tree, and the most meaningful ones are selected by optimizing a measure that consists of two factors: spectral homogeneity, which is calculated in terms of variances of spectral features, and neighborhood connectivity, which is calculated using sizes of connected components. The experiments show that the method is able to detect structures in the image which are more precise and more meaningful than the structures detected by another approach that does not make strong use of neighborhood and spectral information.

## I. INTRODUCTION

Due to the constantly increasing public availability of high-resolution data sets, automatic content extraction and classification on satellite images for urban applications have become important research problems. There is an extensive literature on classification of remotely sensed imagery where pixel level processing has been the common choice for remote sensing image analysis systems. However, a recent study [1] showed that there has not been any significant improvement in the performance of classification methodologies over the last 15 years. The main reason is that the use of only pixel level data often does not meet the expectations as the resolution increases. Even though high success rates have been published in the literature using limited ground truth data, visual inspection of the results can show that most of the urban structures still cannot be delineated as accurately as expected.

Pixel-based approaches assume that similar land structures will cluster together and behave similarly in terms of pixel level features. However, the assumptions for distribution models often do not hold for high-resolution data. We believe

that, in addition to pixel-based spectral data, structural information should also be used to interpret land cover and land use. A common method for incorporating structure information into classification is through the use of regions. This is also referred to as object-oriented classification in the remote sensing literature. For example, Bruzzone and Carlin [2] performed classification using the spatial context of each pixel according to a complete hierarchical multi-level representation of the scene. In a similar approach [3], we obtained a multi-resolution representation using wavelet decomposition, segmented images at each resolution, and used region-based spectral, textural and shape features for classification. Benediktsson *et al.* [4] applied morphological operators with different structuring element sizes to obtain a multi-scale representation of structural information, and used neural network classifiers to label pixels according to their morphological profiles.

In this work, our goal is to develop a segmentation algorithm for partitioning images into spatially contiguous regions so that the structural information can be modeled using the properties of these regions. Most of the segmentation work in the remote sensing literature are based on merging neighboring pixels according to user-defined thresholds on their spectral similarity. Proximity filtering and morphological operations can also be used as post-processing techniques to pixel-based classification results for segmenting regions [5]. In a related work, Pesaresi and Benediktsson [6] performed segmentation using morphological characteristic of pixels in the image. In their approach, opening and closing operations with increasing structuring element sizes were successively applied to an image to generate morphological profiles for all pixels, and the segment label of each pixel was assigned as the structuring element size corresponding to the largest derivative of these profiles. A problem with that approach is that it assumes all the pixels in a particular structure have only one significant derivative maximum occurring at the same structuring element size. However, our experiments have shown that many pixels in most structures often have more than one significant derivative maximum. Furthermore, even though morphological profiles are sensitive to different pixel neighborhoods, the segmentation decision is performed by evaluating pixels individually without considering the neighborhood information.

In this paper, we present a method that uses the neighborhood and spectral information as well as the morphological

This work was supported in part by the TUBITAK CAREER Grant 104E074 and European Commission Sixth Framework Programme Marie Curie International Reintegration Grant MIRG-CT-2005-017504.

information. We first apply principal components analysis to hyper-spectral data to obtain representative bands. Then, we extract candidate regions on each principal component by applying opening and closing by reconstruction operations. For each principal component, we represent the extracted regions by a hierarchical tree, and select the most meaningful regions in that tree by optimizing a measure that consists of two factors: spectral homogeneity, which is calculated in terms of variances of multi-spectral features, and neighborhood connectivity, which is calculated using sizes of connected components.

The rest of the paper is organized as follows. Data and features used are introduced in Section II. Morphological profiles used for modeling structural information are described in Section III. Hierarchical extraction of regions using these profiles is proposed in Section IV. Algorithm for selecting the most meaningful regions within the hierarchy is presented in Section V. Experiments are discussed in Section VI and conclusions are given in Section VII.

## II. FEATURE EXTRACTION

We will illustrate our algorithms with two data sets:

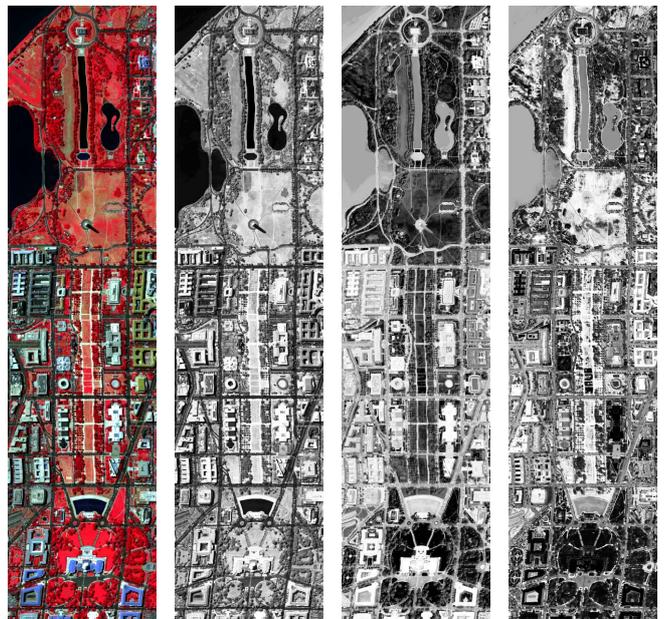
- 1) *DC Mall*: HYDICE (Hyperspectral Digital Image Collection Experiment) image with  $1,280 \times 307$  pixels and 191 spectral bands corresponding to an airborne data flightline over the Washington DC Mall area. False color image is given in Figure 1(a).
- 2) *Centre*: DAIS (Digital Airborne Imaging Spectrometer) and ROSIS (Reflective Optics System Imaging Spectrometer) data with  $1,096 \times 715$  pixels and 102 spectral bands corresponding to the city center in Pavia, Italy. False color image is given in Figure 2(a).

Since morphological operations have traditionally been defined for single band binary or gray scale images, we apply principal components analysis (PCA) and keep the top principal components that represent the 99% variance of the whole data. This corresponds to the first three bands for both data sets (shown in Figures 1(b)–1(d) and 2(b)–2(d)). Considering the fact that different structures may appear more clearly in different principal components, we analyze each PCA band separately for region extraction.

## III. MORPHOLOGICAL PROFILES

Morphological opening and closing operations are used to model structural characteristics of pixel neighborhoods. These operations are applied using increasing structuring element sizes to generate multi-scale characteristics called morphological profiles. The derivative of the morphological profile (DMP) is defined as a vector where the measure of the slope of the opening-closing profile is stored for every step of an increasing SE series [6].

In their segmentation scheme, Pesaresi and Benediktsson [6] define an image segment as a set of connected pixels showing the greatest value of the DMP for the same SE size. That is, the segment label of each pixel is assigned according to the SE size corresponding to the largest derivative of its profiles.



(a) False color (b) 1st PCA band (c) 2nd PCA band (d) 3rd PCA band

Fig. 1. False color image (generated using the bands 63, 52 and 36) and the PCA bands of the *DC Mall* data set.

Their scheme works well in images where the structures in the image are mostly flat so that all pixels in a structure have only one derivative maximum. A drawback of this scheme is that neighborhood information is not used while assigning segment labels to pixels. This results in lots of small noisy segments in images with non-flat structures where the scale with the largest value of the DMP may not correspond to the true structure (see Figure 3 for an illustration). In our approach, we do not consider pixels alone while assigning segment labels. Instead, we also take into account the behavior of the neighbors of the pixels.

## IV. HIERARCHICAL REGION EXTRACTION

Morphological opening and closing operations are known to isolate structures that are brighter and darker than their surroundings, respectively. Contrary to opening (respectively, closing), opening by reconstruction (respectively, closing by reconstruction) preserves the shape of the structures that are not removed by erosion (respectively, dilation). In other words, image structures that the SE cannot be contained are removed while others remain.

In our segmentation approach, our aim is to determine the regions by applying opening and closing by reconstruction operations. We assume that pixels with a positive DMP value at a particular SE size face a change with respect to their neighborhoods at that scale. The main idea is that a neighboring group of pixels that have a similar change for a particular SE size is a candidate region for the final segmentation. These groups can be found by applying connected components analysis to the DMP at each scale. The connected components whose average DMP values are greater than 0.5 and the

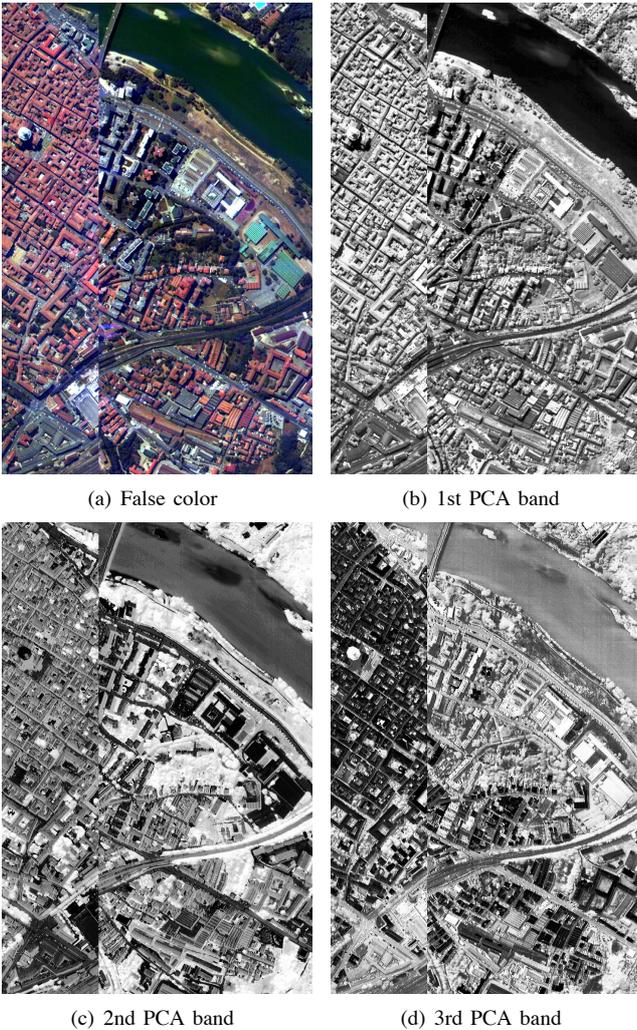
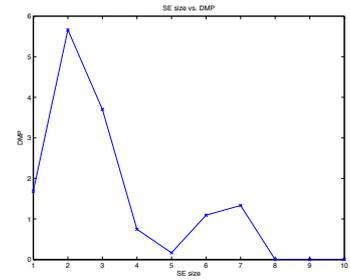


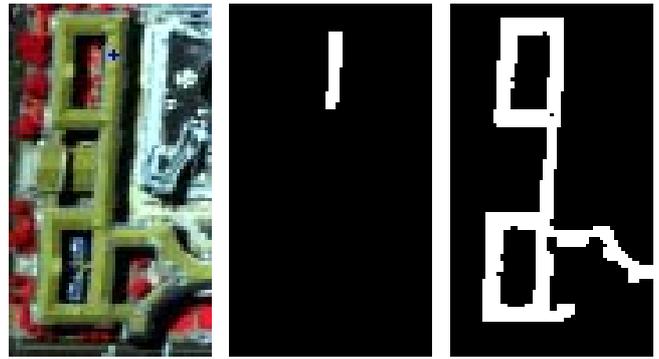
Fig. 2. False color image (generated using the bands 68, 30 and 2) and the PCA bands of the *Centre* data set. (A missing vertical section in the middle was removed.)

numbers of pixels are greater than 10 are considered in the rest of the analysis.

Considering the fact that different structures have different sizes, we apply opening and closing by reconstruction using SEs in increasing sizes from 1 to  $m$ . However, a connected component appearing for a small SE size may be appearing because heterogeneity and geometrical complexity of the scenes as well as other external effects such as shadows produce texture effects in images and result in structures that can be one to two pixels wide [6]. In this case, there is most probably a larger connected component appearing at the scale of a larger SE and to which the pixels of those noise components belong. On the other hand, a connected component that corresponds to a true structure in the final segmentation may also appear as part of another component at larger SE sizes. The reason is that a meaningful connected component may start merging with its surroundings and other connected components after the SE size in which it appears is reached. Figure 4 illustrates these cases.

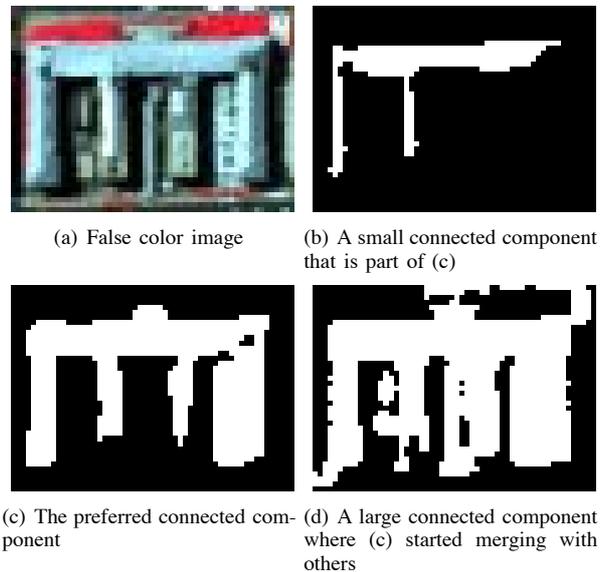


(a) DMP of the pixel marked in (b)



(b) Sample pixel (c) Region for SE size 2 (d) Region for SE size 3 marked on the image

Fig. 3. The greatest value in the DMP of the pixel marked with a blue + in (b) is obtained for SE size 2 (derivative of the opening profile of the 3rd PCA band is shown in (a)). (c) shows the region that we would obtain if we label the pixels with the SE size corresponding to the greatest DMP. The region in (d) that occurs with SE size 3 is more preferable as a complete structure but it does not correspond to the scale of the greatest DMP for all pixels inside the region.



(a) False color image (b) A small connected component that is part of (c)

(c) The preferred connected component (d) A large connected component where (c) started merging with others

Fig. 4. Example connected components for a building structure. These components appear for SE sizes 3, 5 and 6, respectively, in the derivative of the opening profile of the 2nd PCA band.

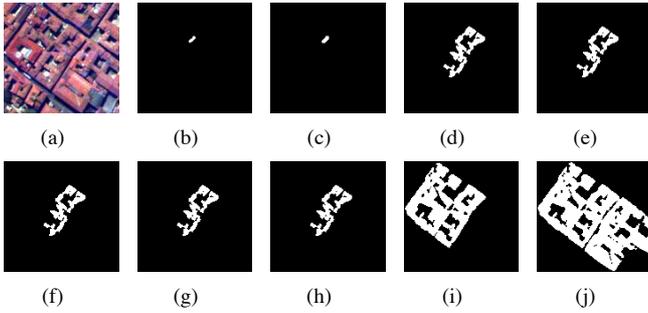


Fig. 5. Example connected components appearing for SE sizes from 2 to 10 in the derivative of the opening profile of the 3rd PCA band. These regions are contained within each other in a hierarchical manner. Note that the components do not change in some of the scales.

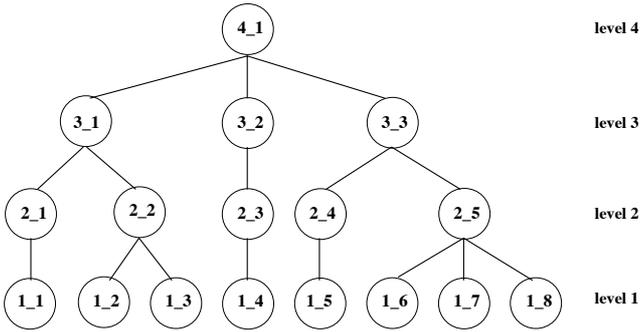


Fig. 6. An example tree. Node  $i_j$  is a connected component that exists for SE size  $i$ .  $j$  denotes the sequence number of the node from left to right in level  $i$ .

Through increasing SE sizes from 1 to  $m$ , each morphological operation reveals connected components that are contained within each other in a hierarchical manner where a pixel may be assigned to more than one connected component appearing at different SE sizes (see Figure 5). We treat each component as a candidate meaningful region. Using these candidate regions, a tree is constructed where each connected component is a node and there is an edge between two nodes corresponding to two consecutive scales (SE sizes differ by 1) if one node is contained within the other. Leaf nodes represent the components that appear for SE size 1. Root nodes represent the components that exist for SE size  $m$ . Since we use a finite number of SE sizes, there may be more than one root node. In this case, there will be more than one tree and the algorithms described in the next section are run on each tree separately.

Figure 6 shows an example tree where the nodes are labeled as  $i_j$  with  $i$  denoting the node's level and  $j$  denotes the number of the node from left to right in level  $i$ . For example, node 3\_3 has two children nodes 2\_4 and 2\_5, and its parent is node 4\_1. The reason of node 2\_3 having only one child may be that either no new connected component appears in level 2 or node 2\_3 is formed by merging of node 1\_4 with its surrounding pixels that are not included in any connected component in level 1. The same reasons also hold for node 3\_2.

## V. REGION SELECTION

After forming the tree, our aim is to search for the most meaningful connected components among those appearing at different SE sizes in the segmentation hierarchy. With a similar motivation in [7], Tilton analyzed hierarchical image segmentations and selected the meaningful regions manually. Then, Plaza and Tilton [8] investigated how different spectral, spatial and joint spectral/spatial features of regions change from one level to another in a segmentation hierarchy with the goal of automating the selection process in the future. In this paper, each node in the tree is treated as a candidate region in the final segmentation, and selection is done automatically as described below.

Ideally, we expect a meaningful region to be as homogeneous as possible. However, in the extreme case, a single pixel is the most homogeneous. Hence, we also want a region to be as large as possible. In general, a region stays almost the same (both in homogeneity and size) for some number of SEs, and then faces a large change at a particular scale either because it merges with its surroundings to make a new structure or because it is completely lost. Consequently, the size we are interested in corresponds to the scale right before this change. In other words, if the nodes on a path in the tree stay homogeneous until some node  $n$ , and then the homogeneity is lost in the next level, we say that  $n$  corresponds to a meaningful region in the hierarchy.

With this motivation, to check the meaningfulness of a node, we define a measure consisting of two factors: spectral homogeneity, which is calculated in terms of variances of spectral features, and neighborhood connectivity, which is calculated using sizes of connected components. Then, starting from the leaf nodes (level 1) up to the root node (level  $m$ ), we compute this measure at each node and select a node as a meaningful region if it is the most homogeneous and large enough node on its path in the hierarchy (a path corresponds to the set of nodes from a leaf to the root).

In order to calculate the homogeneity factor in a node, we use the fact that pixels in a correct structure should have not only similar morphological profiles, but also similar spectral features. Thus, we calculate the homogeneity of a node as the standard deviation of the spectral information of the pixels in the corresponding region where the spectral information of a pixel consists of the PCA components representing the 99% variance of the whole data. However, while examining a node from the leaf up to the root in terms of homogeneity, we do not use the standard deviation of the node directly. Instead, we consider the difference of the standard deviation of that node and its parent. What we expect is a sudden increase in the standard deviation. When the standard deviation does not change much, it usually means that small sets of pixels are added to the region or some noise pixels are cleaned. When there is a large change, it means that the structure merged with a larger structure or it merged with other irrelevant pixels disturbing the homogeneity in the node. Hence, the difference of the standard deviation in the node's parent and the standard

deviation in the node should be maximized while selecting the most meaningful nodes.

As discussed above, using only the homogeneity factor will favor small structures. To overcome this problem, the number of pixels in the region corresponding to the node is introduced as another factor to create a trade-off. As a result, the goodness measure  $M$  for a node  $n$  is defined as

$$M(n) = D(n, \text{parent}(n)) \times C(n) \quad (1)$$

where the first term is the standard deviation difference between the node's parent and itself, and the second term is the number of pixels in the node. The node that is relatively homogeneous and large enough will maximize this measure and will be selected as a meaningful region.

Given the value of the goodness measure for each node, we find the most meaningful regions as follows. Suppose  $T = (N, E)$  is the tree with  $N$  as the set of nodes and  $E$  as the set of edges. The leaf nodes are in level 1 and the root node is at level  $m$ . Let  $P$  denote the set of all paths from the leaves to the root, and  $M(n)$  denote the measure at node  $n$ . We select  $N^* \subseteq N$  as the final segmentation such that

- 1)  $\forall a, b \in N^*$ ,  
 $\forall p \in P : a \in p \rightarrow b \notin p$ ,  
 $\forall p \in P : b \in p \rightarrow a \notin p$ ,
- 2)  $\forall a \in N^*, \forall n \in N$ ,  
 $\exists p \in P : a \in p \wedge n \in p \rightarrow M(a) \geq M(n)$ .

The first condition requires that any two nodes in  $N^*$  cannot be on the same path (i.e., the corresponding regions cannot overlap). The second condition requires that any node in  $N^*$  must have the greatest measure on the paths it is included.

We use a two-pass algorithm for selecting the most meaningful nodes ( $N^*$ ) in the tree. The bottom-up (first) pass aims to find the nodes whose measure is greater than all of its descendants. The algorithm first marks all nodes in level 1. Then, starting from level 2 up to the root level, it checks whether each node in each level has a measure greater than or equal to those of all of its children. The greatest measure, seen so far in each path, is propagated to upper levels so that it is enough to check only the children, rather than all descendants, in order to find whether a node's measure is greater than or equal to all of its descendants'.

After the bottom-up pass marks all such nodes, the top-down (second) pass seeks to select the nodes whose measure is the greatest on each of their corresponding paths. It starts by marking all nodes as *selected* in the root level if they are marked by the bottom-up pass. Then, in each level until the leaf level, the algorithm checks for each node whether it is marked in the bottom-up pass while none of its ancestors is marked. If this condition is satisfied, it marks the node as *selected* if its measure is greater than those of all of its ancestors. For that purpose, we again propagate the greatest measure, seen so far in each path, to lower levels. Finally, the algorithm selects the nodes that are marked as *selected* in each level as meaningful regions.

## VI. EXPERIMENTS

We applied the proposed region selection algorithm to both data sets. The tree structure was constructed for each PCA band separately and the regions were selected from each tree individually. Figures 7 and 8 show example segmentation results for *DC Mall* and *Centre* data sets, respectively. Structuring element sizes from 1 to 10 were used for both opening and closing profiles for both data sets. We present the zoomed versions of the results for several example areas to better illustrate the details for high-resolution imagery and for clarity of the presentation on paper. The results obtained by the algorithm in [6] are also given for the same areas.

The results show that our segmentation algorithm usually finds structures as a whole but the method of [6] often oversegments them and produces small regions. These small regions occur because the segment label assignment is done for each pixel individually by only considering the greatest value in its DMP. Thus, noisy pixels that are different from their neighborhoods may produce small regions because they may have large values occurring at scales corresponding to small SE sizes. However, our algorithm considers both the morphological characteristics encoded in the DMP and the spectral information measured in terms of the standard deviation within contiguous groups of pixels. It also considers the consistency of these values within neighboring pixels forming large connected components. As a result, the combined measure that uses both spectral and neighborhood information is both robust to noise and consistent within detailed structures in high-resolution images. In all of the examples, our algorithm is able to extract many meaningful regions as whole segments.

Another important observation is that different structures are extracted more clearly in different principal components. For example, the structures in both Figures 7(a) and 7(b) are found in the second PCA band of the *DC Mall* data set like many other buildings. The structures in both Figures 8(a) and 8(c) are found in the third PCA band of the *Centre* data set but the structures in Figure 8(b) are found in the first PCA band. The reason that a particular structure being extracted better in a particular PCA band is that the pixels belonging to that structure are found lighter or darker than their surroundings on that PCA band. This motivates an important future work on merging the results from individual PCA bands as a final segmentation for an image. As a final note, we also observed that the texture effects produced by vegetation in some of the PCA bands result in small regions in those areas. We will investigate additional multi-spectral features (e.g., NDVI) to improve the segmentation for such regions.

## VII. CONCLUSIONS

We described a method for segmentation of urban structures in high-resolution images. The first step was to extract structural information using morphological opening and closing by reconstruction operators. Principal components analysis bands were used to summarize hyper-spectral data and the morphological operators were applied to each band separately. Then, candidate regions were extracted by applying connected

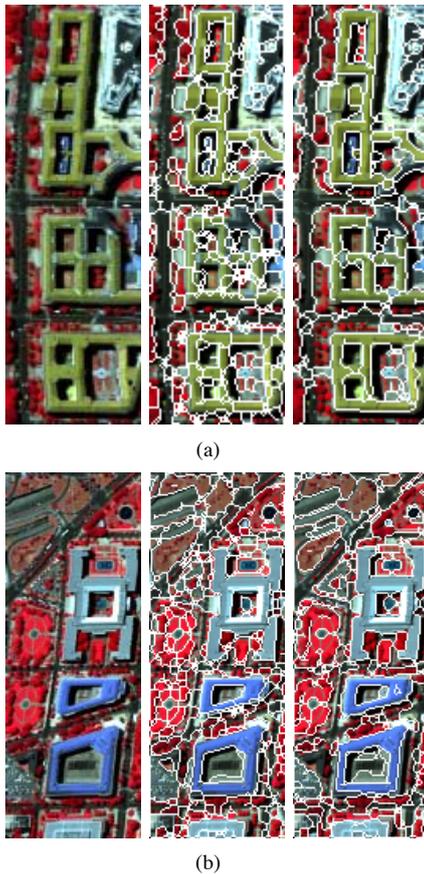


Fig. 7. Example segmentation results for the *DC Mall* data set. The left image shows the false color representation, the middle one shows the result of the algorithm in [6], and the right one shows the result of the proposed approach.

components analysis to the pixels selected according to their morphological profiles obtained using increasing structuring element sizes. Next, these regions were represented using a tree, and the most meaningful ones were selected by optimizing a measure that consisted of two factors: spectral homogeneity, which was calculated in terms of variances of spectral features, and neighborhood connectivity, which was calculated using sizes of connected components.

We evaluated the proposed approach on two data sets. The experiments showed that our method that considers morphological characteristics, spectral information, and their consistency within neighboring pixels is able to detect structures in the image which are more precise and more meaningful than the structures detected by another approach that does not make strong use of neighborhood and spectral information.

#### ACKNOWLEDGMENT

The authors would like to thank Dr. David A. Landgrebe and Mr. Larry L. Biehl from Purdue University, Indiana, U.S.A., for the *DC Mall* data set, and Dr. Paolo Gamba from the University of Pavia, Italy, for the *Centre* data set.

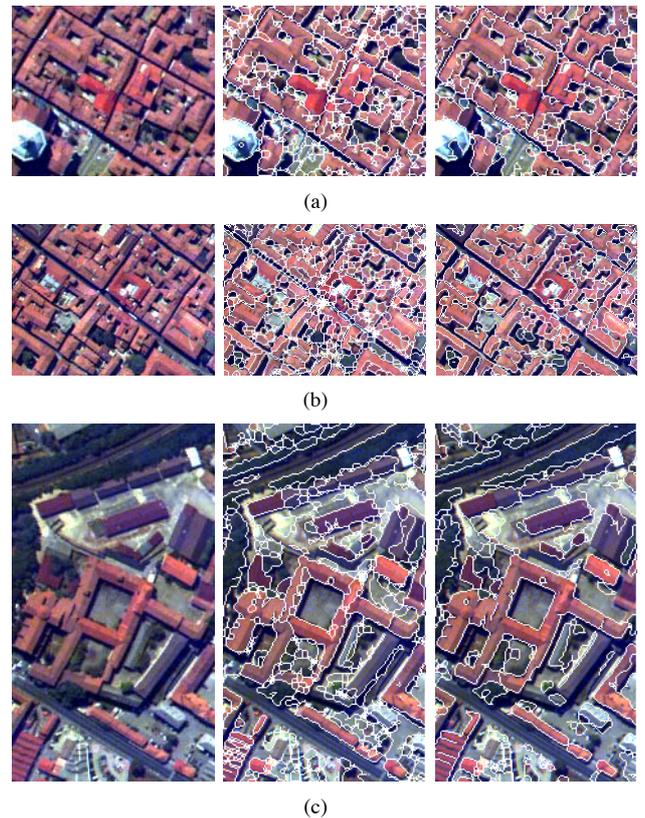


Fig. 8. Example segmentation results for the *Centre* data set. The left image shows the false color representation, the middle one shows the result of the algorithm in [6], and the right one shows the result of the proposed approach.

#### REFERENCES

- [1] G. G. Wilkinson, "Results and implications of a study of fifteen years of satellite image classification experiments," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 433–440, March 2005.
- [2] L. Bruzzone and L. Carlin, "A multilevel context-based system for classification of very high spatial resolution images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 9, pp. 2587–2600, September 2006.
- [3] S. Aksoy and H. G. Akcay, "Multi-resolution segmentation and shape analysis for remote sensing image classification," in *Proceedings of 2nd International Conference on Recent Advances in Space Technologies*, Istanbul, Turkey, June 9–11, 2005, pp. 599–604.
- [4] J. A. Benediktsson, M. Pesaresi, and K. Arnason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 9, pp. 1940–1949, September 2003.
- [5] S. Aksoy, K. Koperski, C. Tusk, G. Marchisio, and J. C. Tilton, "Learning Bayesian classifiers for scene classification with a visual grammar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 581–589, March 2005.
- [6] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 2, pp. 309–320, February 2001.
- [7] J. C. Tilton, "Analysis of hierarchically related image segmentations," in *Proceedings of IEEE GRSS Workshop on Advances in Techniques for Analysis of Remotely Sensed Data*, Washington, DC, October 27–28, 2003.
- [8] A. J. Plaza and J. C. Tilton, "Automated selection of results in hierarchical segmentations of remotely sensed hyperspectral images," in *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, vol. 7, Seoul, Korea, July 25–29, 2005, pp. 4946–4949.