

DETECTION OF COMPOUND STRUCTURES USING MULTIPLE HIERARCHICAL SEGMENTATIONS

H. Gökhan Akçay, Selim Aksoy

Department of Computer Engineering
Bilkent University
Bilkent, 06800, Ankara, Turkey
{akçay,saksoy}@cs.bilkent.edu.tr

ABSTRACT

In this paper, our aim is to discover compound structures comprised of regions obtained from hierarchical segmentations of multiple spectral bands. A region adjacency graph is constructed by representing regions as vertices and connecting these vertices that are spatially close by edges. Then, dissimilarities between neighboring vertices are computed using statistical and structural features, and are assigned as edge weights. Finally, the compound structures are detected by extracting the connected components of the graph whose edges with relatively large weights are removed. Experiments using WorldView-2 images show that grouping of these vertices according to different criteria can extract high-level compound structures that cannot be obtained using traditional techniques.

Index Terms— Object detection, hierarchical segmentation, graph-based representation, alignment detection, hierarchical clustering

1. INTRODUCTION

Object recognition has been an important problem in remote sensing image analysis. Many popular algorithms in the computer vision literature assume a moderate number of homogeneous objects in images. However, this assumption does not hold for high-resolution remote sensing images that contain a large number of intrinsically heterogeneous structures. We call these structures *compound structures*. Examples of compound structures include different types of residential areas, commercial areas, industrial areas, and agricultural areas that are comprised of different spatial arrangements of various *primitive objects* such as buildings, roads, and trees (see Figure 1 for an illustration). In this paper, we describe our work on the modeling and unsupervised detection of such compound structures.

Hierarchical segmentation has emerged as a promising approach for the detection of compound structures. Further-



Fig. 1. Compound structures in three 500×500 pixel multi-spectral WorldView-2 images of Ankara, Turkey.

more, given a hierarchical segmentation, meaningful and interesting objects can be extracted [1]. A common method for constructing the hierarchy is splitting and/or merging based on spectral homogeneity. However, compound structures that consist of multiple parts with different spectral characteristics often do not appear in such hierarchies. As an alternative, Gaetano et al. [2] performed hierarchical texture segmentation assuming that frequent neighboring regions are strongly related. In order to find the strongly related regions, they clustered the image regions to compute the frequencies of quantized region pairs. However, these frequencies may be very sensitive to the number of clusters which is determined heuristically. Similarly, Zamalieva et al. [3] found the significant relations between neighboring regions as the modes of a probability distribution estimated using the continuous features of region co-occurrences. The resulting modes were used to construct the edges of a graph where a graph mining algorithm was used to find subgraphs that may correspond to compound structures. However, these frequency-based algorithms are not usually sufficient for modeling the complex characteristics of compound structures. In [4], we described a procedure for the detection of compound structures that combined statistical characteristics of primitive objects modeled using spectral, shape, and position information with structural characteristics encoded using spatial alignments of neighboring object groups.

In this paper, we detect compound objects whose primitive objects are found in a set of hierarchical segmentations.

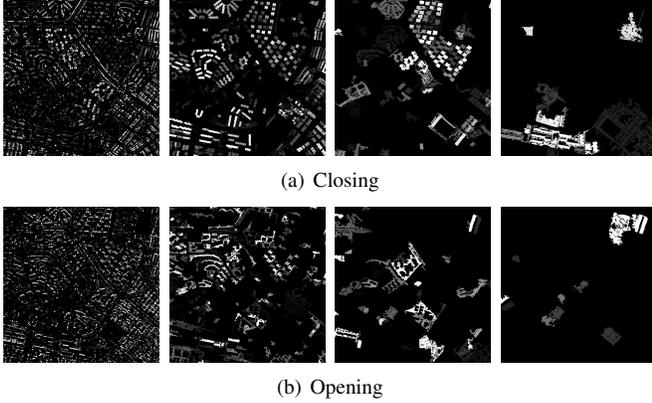


Fig. 2. Morphological profile using structuring element sizes 3, 6, 9, 12.

First, we obtain multiple hierarchical segmentations by applying morphological opening and closing operations to individual spectral bands using structuring elements with increasing sizes (Section 2). These operations produce a set of regions forming a hierarchy for each band. Then, grouping of these regions according to different criteria produces different compound structures (Section 3). The proposed algorithms are illustrated in proof-of-concept experiments using a WorldView-2 image of Ankara, Turkey (Section 4).

2. HIERARCHICAL REGION EXTRACTION

The hierarchical segmentation algorithm uses morphological operations to exploit structural information in each spectral band. First, morphological opening and closing by reconstruction operations are applied to individual spectral bands using structuring elements (SE) in increasing sizes to generate morphological profiles. For each opening and closing profile, through increasing SE sizes, each morphological operation reveals connected components that are contained within each other in a hierarchical manner (see Figure 2 for an illustration). These connected components form a hierarchy of regions for each band.

An important observation is that different structures are extracted more clearly in different spectral bands. For example, buildings with red roofs are detected more accurately in the hue band of the HSV color space but industrial buildings are detected more accurately in the red band of the particular example image shown in Figure 3.

3. COMPOUND OBJECT DETECTION

The input to the detection algorithm is a set of hierarchical segmentations corresponding to different spectral bands. The goal is to find region groups that correspond to compound structures. In each segmentation scale, we construct a re-



Fig. 3. Example segmentation results for different spectral bands of the second image in Figure 1. The left and right images show the regions extracted in the hue and red bands where the red-roof and the industrial buildings are detected more clearly, respectively.

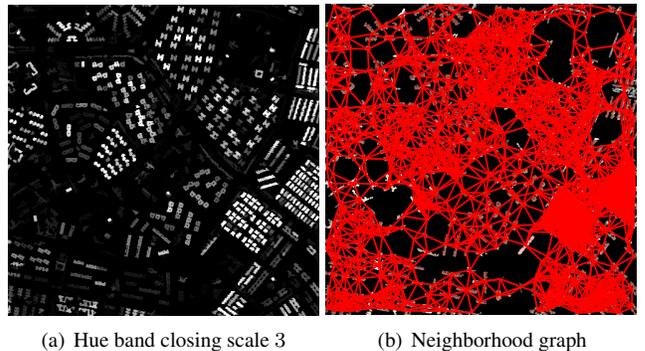


Fig. 4. Examples of graph construction. The vertices that are considered as neighbors based on proximity analysis are connected with red edges in (b).

gion adjacency graph (RAG) where the individual primitive objects correspond to the vertices. We assume that neighboring regions can be related, and connect every neighboring vertex pair with an edge. The neighborhood information is obtained by proximity analysis where a threshold on the distance between the centroids of each object pair is used to determine the neighbors. Figure 4 shows an example graph where the regions in the third scale of the closing profile are used as vertices of interest and edges are drawn using a distance threshold of 30 pixels.

We assign edge weights in the RAG according to statistical and structural dissimilarities between vertices in the same scale of a segmentation hierarchy. Once edge weights are assigned between vertices, we obtain an attributed relational graph (ARG) of the scene. Finally, meaningful compound object candidates can be detected by grouping the vertices of ARG using graph cuts. The sections below describe how the edge weights are computed for different types of structures of interest.

3.1. Statistical dissimilarity

The statistical features for vertices represent the properties of individual objects. In particular, we propose to model each region using a Gaussian distribution in feature and spatial domains. Given an image with d spectral bands, spectral information of each region v is represented using the mean values of the pixels within the region for each spectral band, i.e., $\mu_v^{spec} = \{\mu_{vj}^{spec} : j = 1, \dots, d\}$, and the covariance matrix of the spectral features of the pixels within the region, i.e., Σ_v^{spec} . Similarly, the shape of each region v is represented using the covariance matrix of the spatial locations (coordinates) of the pixels within the region, i.e., Σ_v^{shape} . Then, the spectral and shape dissimilarities between two vertices (and their corresponding distributions) v_1 and v_2 can be measured using the Kullback-Leibler (KL) divergence as

$$D_{KL}^{spec} = \frac{1}{2} \left(\log \frac{|\Sigma_{v_2}^{spec}|}{|\Sigma_{v_1}^{spec}|} + \text{Tr}((\Sigma_{v_2}^{spec})^{-1} \Sigma_{v_1}^{spec}) + (\mu_{v_1}^{spec} - \mu_{v_2}^{spec})^T (\Sigma_{v_2}^{spec})^{-1} (\mu_{v_1}^{spec} - \mu_{v_2}^{spec}) - d \right) \quad (1)$$

and

$$D_{KL}^{spat} = \frac{1}{2} \left(\log \frac{|\Sigma_{v_2}^{spat}|}{|\Sigma_{v_1}^{spat}|} + \text{Tr}((\Sigma_{v_2}^{spat})^{-1} \Sigma_{v_1}^{spat}) - d \right), \quad (2)$$

respectively, where $|\Sigma|$ denotes the determinant of the matrix Σ and Tr represents trace. A larger KL value corresponds to a higher dissimilarity between two regions.

Heterogeneous structures that are brighter/darker than their surroundings (e.g., industrial buildings) may not be accurately represented with a Gaussian distribution. These structures usually correspond to regions that appear in upper scales of the hierarchy. Therefore, each such region is modeled using the statistical summary of its pixel content. In the experiments, these summaries are obtained by quantizing the feature values of the pixels that appear in the first scale of the hierarchy using the k -means algorithm, and by representing the distribution of these quantized values in a histogram [1]. Then, the dissimilarity between two region histograms is measured using the L_1 distance. Different histogram distance measures can also be used.

3.2. Structural dissimilarity

The structural features represent the spatial layout of each region with respect to its neighbors, and are extracted using the relationships among the neighboring regions. An important structural information is the amount of alignment among the regions. In [4], we proposed a method for the detection of aligned object groups using a depth-first search on the graph that is constructed as described above. At the end of the search procedure, the set of structural features computed for each object group corresponding to each alignment group

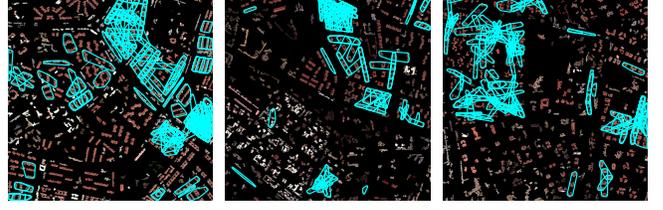


Fig. 5. Example results for alignment detection. The detected groups are marked by their convex hulls.

consists of the orientation of the line fitted to the centroids of the individual objects in that group, θ_i , and the mean of the distances computed between the closest object pairs in the group, μ_i , where $i = 1, \dots, m$ and m is the number of detected aligned object groups. Both structural features are normalized to the unit range by using the respective minimum and maximum values. Finally, each vertex in the graph is assigned a list of aligned object groups that it belongs to as its structural features. Figure 5 shows example results for alignment detection.

The dissimilarity between two alignment groups is computed as the sum of squared differences between the corresponding features of these groups. The dissimilarity for two objects is computed as the minimum of the distances between all pairs of alignment groups where one group in a pair is associated with one of the objects and the other group is associated with the other object. The structural dissimilarity will be small if two objects belong to alignment groups whose orientations and object spacing are similar. If at least one of the objects is not found to belong to any alignment group, the dissimilarity of that object to any other object is set to ∞ .

3.3. Grouping

To cluster the regions into groups based on statistical and structural dissimilarities, we remove some edges of the RAG by thresholding the edge weights to obtain a similarity graph. Connected components of the similarity graph correspond to compound objects. These connected components can also be obtained by hierarchical clustering using the single linkage criterion. In single linkage-based clustering, two vertices v and v' are in the same cluster if there exists a chain $v, v_1, v_2, \dots, v_k, v'$ such that v is similar to v_1 , v_1 is similar to v_2 , and so on, for the whole chain. Thus, the clustering corresponds to the connected components of the similarity graph.

4. EXPERIMENTS

We performed experiments on the WorldView-2 images of Ankara shown in Figure 1 to illustrate the grouping framework proposed in this paper. Experiments were done using the regions that were extracted by applying morphological

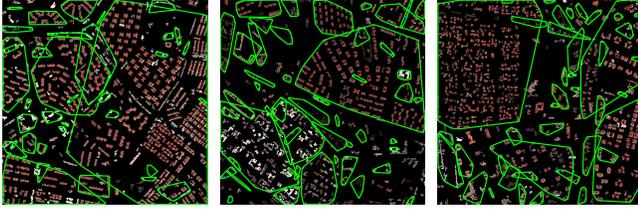


Fig. 6. Groups formed by clustering the graphs according to spectral features.

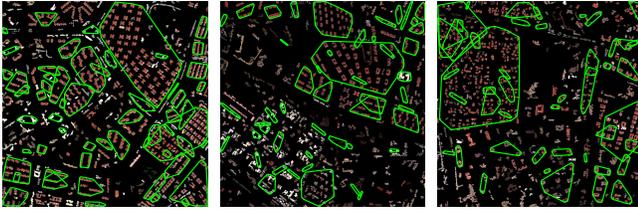


Fig. 7. Groups formed by clustering the graphs according to both spectral and shape features.

opening and closing by reconstruction operations to different spectral bands.

The first set of experiments consists of grouping regions obtained from the third scale of the closing profile of the hue band. The grouping results using only spectral features with (1) are shown in Figure 6. This clustering resulted in relatively large groups that contain neighboring vertices with similar color content. Figure 7 shows groups that are obtained by using both spectral and shape features by adding (1) and (2). We can observe that, by adding shape features, groups that contain vertices with similar color content but different shapes can be separated into more meaningful smaller groups. Moreover, the groups that exploit additional structural properties (e.g., alignments) are more meaningful as a whole compared to using only statistical properties of individual vertices. For example, Figure 8 shows the grouping results using structural features. The results show successful extraction of groups with three or more buildings that are part of similar linearly aligned groups. The groups that do not satisfy this strict definition of alignment remain separated. The last set of experiments aims to group intrinsically heterogeneous regions obtained from the fifth and seventh scales of the opening profile of the yellow band in Figure 9. These regions are comprised of pixels with different properties, and single Gaussians are usually not sufficient to represent their characteristics. The results show that complex industrial buildings are successfully grouped using the $L1$ distance between their histogram features.

5. CONCLUSIONS

In this paper, we described a method that aims to group regions that appear in different hierarchical segmentations ob-

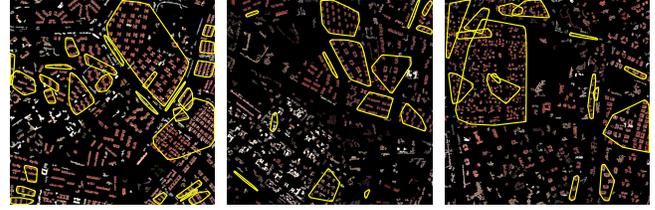


Fig. 8. Groups formed by clustering the graphs according to structural features.

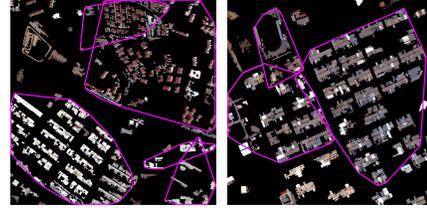


Fig. 9. Groups formed by clustering the graphs according to histogram features.

tained from multiple spectral bands for the detection of compound structures. Our method models statistical characteristics of regions by assuming Gaussian spectral and shape distributions, and structural characteristics are encoded using spatial alignments. We evaluated the proposed approach qualitatively on three images. The experiments showed that the proposed method is able to group regions belonging to different compound structures. As a result, such compound structures can be used in new semantic classification, annotation, indexing, and retrieval applications.

6. REFERENCES

- [1] H. G. Akcay and S. Aksoy, “Automatic detection of geospatial objects using multiple hierarchical segmentations,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 7, pp. 2097–2111, July 2008.
- [2] R. Gaetano, G. Scarpa, and G. Poggi, “Hierarchical texture-based segmentation of multiresolution remote-sensing images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 7, pp. 2129–2141, July 2009.
- [3] D. Zamalieva, S. Aksoy, and J. C. Tilton, “Finding compound structures in images using image segmentation and graph-based knowledge discovery,” in *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, Cape Town, South Africa, July 13–17, 2009.
- [4] H. G. Akcay and S. Aksoy, “Detection of compound structures using hierarchical clustering of statistical and structural features,” in *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, Vancouver, Canada, July 25–29, 2011, pp. 2385–2388.