

MINING OF REMOTE SENSING IMAGE ARCHIVES USING SPATIAL RELATIONSHIP HISTOGRAMS

Firat Kalaycilar, Aslı Kale, Daniya Zamalieva, Selim Aksoy

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

ABSTRACT

We describe a new image representation using spatial relationship histograms that extend our earlier work on modeling image content using attributed relational graphs. These histograms are constructed by classifying the regions in an image, computing the topological and distance-based spatial relationships between these regions, and counting the number of times different groups of regions are observed in the image. We also describe a selection algorithm that produces very compact representations by identifying the distinguishing region groups that are frequently found in a particular class of scenes but rarely exist in others. Experiments using Ikonos scenes illustrate the effectiveness of the proposed representation in retrieval of images containing complex types of scenes such as dense and sparse urban areas.

Index Terms— Image retrieval, spatial relationships, feature selection

1. INTRODUCTION

Image information mining is a relatively new field of research for automating the content extraction and exploitation processes in large Earth observation data archives where the goal is to build high-level subjective content models by combining low-level features, and support classification and content-based retrieval of image content in terms of semantic queries. For example, Datcu *et al.* [1] developed a system where users can train Bayesian classifiers for a particular concept (e.g., water) using positive and negative examples of pixels, and can have image tiles ranked according to the coverage of this concept estimated using pixel level models. Li and Narayanan [2] described a system where images are divided into tiles and are retrieved using spectral and textural statistics. Systems that support object extraction and modeling of image content based on these objects have also been developed [3, 4].

Even though correct identification of pixels and regions improve the processing time for content extraction, manual

interpretation is often necessary for many applications because two scenes with similar regions can have very different interpretations if the regions have different spatial arrangements. Therefore, modeling spatial information to understand the context has been an important and challenging research problem. A structural method for modeling context is through the quantification of spatial relationships. For example, Shyu *et al.* [4] developed a method that generates a spatial signature of the configuration of the objects in an image tile. In previous work [3], we developed automatic methods for extraction of topological, distance-based and relative position-based relationships between region pairs, and successfully used such relationships for image classification and retrieval in scenarios that cannot be expressed by traditional pixel- and region-based approaches. Then, in [5], we modeled image scenes using attributed relational graphs that combine region class information and spatial arrangements, and formulated image retrieval as a relational graph matching problem.

Attributed relational graphs (ARG) are very general and powerful representations of image content. In our ARG model, for an image with n regions, the regions are represented by n graph nodes and the $\binom{n}{2}$ pairwise spatial relationships between them are represented by the edges between these nodes. However, finding similarities between graphs can easily become intractable for large images in large data sets, and image mining that is formulated as a graph searching problem can become infeasible when these data sets are concerned. Furthermore, these graphs can be too detailed, and the result set of a search session can be quite small when these detailed representations are compared.

In this paper, we propose new models for image content representation using spatial relationship histograms. These histograms are more powerful representations than commonly used tile-based spectral or textural feature histograms [2, 4] but are not as complex as the full graph-based representations. In other words, they provide a summary of the full scene graph while enabling complex queries that cannot be modeled using histograms of pixels' spectral or textural feature values.

This work was supported in part by the TUBITAK CAREER grant 104E074 and grant 105E065.

The computation of the spatial relationship histograms starts with image segmentation and region classification (Section 2). Given the extracted regions with their associated class labels, topological and distance-based spatial relationships between all region pairs are computed (Section 3). Then, this relationship information is encoded using histograms that count the number of times different groups of regions are observed in the image (Section 4). As the size of the region groups is increased, the detail of the content representation also increases but the histograms become sparser. Therefore, a novel selection algorithm is proposed to find important region groups that are more informative in distinguishing one type of scene from the others (Section 5). Experiments using Ikonos scenes illustrate the effectiveness of the spatial relationship histograms in retrieval of images containing complex types of scenes such as dense and sparse urban areas (Section 6).

2. REGION SEGMENTATION AND CLASSIFICATION

Segmentation and classification are done jointly by using Bayesian classifiers. Spectral values and Gabor texture features are used for pixel representation, and binary classifiers are trained using positive and negative examples for the following classes: roof, water, tree, bare soil, grass, street, path and shadow. Then, each pixel is assigned to a class according to the maximum posterior probability given by these classifiers. The final segmentation is obtained using an iterative split-and-merge algorithm that combines contiguous groups of pixels that are assigned to the same class. Details of the segmentation and classification algorithm can be found in [3].

3. REGION SPATIAL RELATIONSHIPS

Topological (e.g., disjointed, bordering, invading, surrounding) and distance-based (e.g., near, far) spatial relationships (Figure 1) can be computed using overlaps and distances between region boundaries, respectively, for each region pair in an image [3, 5]. In this paper, the coarse-to-fine search strategy described in [5] is used to compare all region pairs in each image according to the region boundaries, and fuzzy membership functions [3] are used to convert the computed quantitative relationship information into semantic labels.

4. SPATIAL RELATIONSHIP HISTOGRAMS

In [5], the image content is modeled by using attributed relational graphs of labeled regions where the regions are represented by the graph nodes and their pairwise spatial relationships are represented by the edges between these nodes. Although these graphs are very powerful representations, the graph similarity that is computed as the minimum cost taken over all sequences of operations that transform one graph to

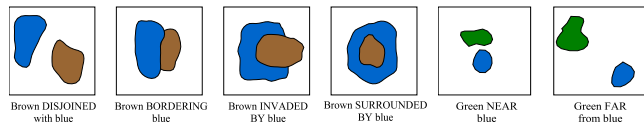


Fig. 1. Spatial relationships of region pairs.

the other can lead to very high computational complexity. Furthermore, the detailed representation can produce a small result set for a search session. Due to these practical issues, we propose to use spatial relationship histograms that can be easily obtained from ARGs. These histograms are not as complex as full graph models, but are still more powerful than commonly used low-level representations.

The spatial relationship histograms can be computed at different levels of detail. The complexity of the histogram is determined by the size (order) of the region groups considered. For example, when only region pairs are taken into account, the histograms encode second-order region relationships, and when groups of three regions are considered, the histograms encode third-order relationships. To compute these histograms for a given order (i.e., for a given number of regions to be considered), first, we generate all possible relationships between all possible region classes. This combinatorial problem is solved recursively. Then, the histogram for an image is computed by counting the number of times each possible region group is observed in the ARG of that image. For example, a sample bin of a second-order histogram can correspond to the number of “(grass BORDERING street)” observations found in an image. A sample bin of a third-order histogram can correspond to the number of “(roof INVADDED.BY NEAR grass) & (grass BORDERING NEAR street) & (street DISJOINED NEAR roof)” observations. We also compute first-order histograms that simply count the number of pixels belonging to each region class without considering any spatial relationships for comparison.

5. FEATURE SELECTION

When the order of region groups is increased, the detail of the content representation also increases but the representation may become too specific and the problem of sparsity can also become more significant. In other words, the histograms may become sparser because not all possible region groups are observed in an image. An interesting problem is the identification of the important region groups for a given set of scene classes because not all region groups are equally informative in distinguishing one type of scene from the others.

Given example images for a user-defined set of scene classes, the goal of the selection process is to identify the region groups that are frequently found in a particular class of scenes, consistently occur together in the same type of scenes, but rarely exist in other scenes. We formulate the selection process as a multi-subset search problem that is solved using

the sequential forward selection algorithm that we recently developed for image classification [6]. The goal of this algorithm is to find a set of subsets (called a multi-subset) for which a given goodness criterion is maximized.

The smallest component in this procedure is a group of regions with their class labels and spatial relationships (in other words, a component corresponds to a bin in the spatial relationship histogram). Each subset consists of several components that are determined to be the best set of region groups for a particular type of scene, and the multi-subset represents the region groups selected for the whole data set. The particular goodness criterion used here consists of two parts where the first part quantifies the importance of each component for a particular scene class and the second part measures the importance of each pair of components with respect to different scene classes. The sequential forward selection algorithm iteratively finds the components (region groups) that maximize this criterion (details can be found in [6]). Note that this procedure performs selection using only the frequencies of region groups in example images, and does not depend on a specific classifier unlike most of the supervised selection algorithms. After feature selection, only the selected set of region groups are used in the spatial relationship histogram.

6. EXPERIMENTS

The performances of the spatial relationship histogram representation and the selection algorithm were evaluated using a retrieval system that finds images with content similar to the query image. The data set used consisted of an Ikonos scene of Istanbul with pan-sharpened red, green and blue bands and 14416×11946 pixels. The whole scene was divided into 250×250 pixel tiles and a spatial relationship histogram was computed for each tile. A subset of these tiles were assigned high level class labels as ground truth. The high level classes were chosen to be *dense urban*, *sparse urban* and *very sparse urban*. The number of tiles for each class were 46, 62 and 74, respectively. During retrieval, a tile was accepted as a true match if it belonged to the same high level class as the query.

The histograms were computed at three levels (orders) of detail using regions labeled with 8 classes listed in Section 2. Different settings were used as shown in Table 1. Only the region pairs that were near each other according to the computed distance-based relationship were considered in all settings. Topological relationships of bordering, invading and surrounding were used only in setting 2. These settings determined the size of the histogram. For example, 8^2 possible types of nearby region pairs resulted in 64 bins for setting 1, adding 3 possible topological relationships resulted in 192 bins for setting 2, and 3 pairs of regions with 8^2 possible types for each pair resulted in 262144 bins for setting 3. Setting 4 consisted of the first-order histograms computed as the baseline method without using any spatial information.

Example histograms are shown in Figure 2. As expected,

Table 1. Settings and the corresponding histogram sizes used in the experiments.

Setting	Order	Topological	Distance	Size
1	2	—	near	64
2	2	bordering, invading, surrounding	near	192
3	3	—	near	262144
4	1	—	—	8

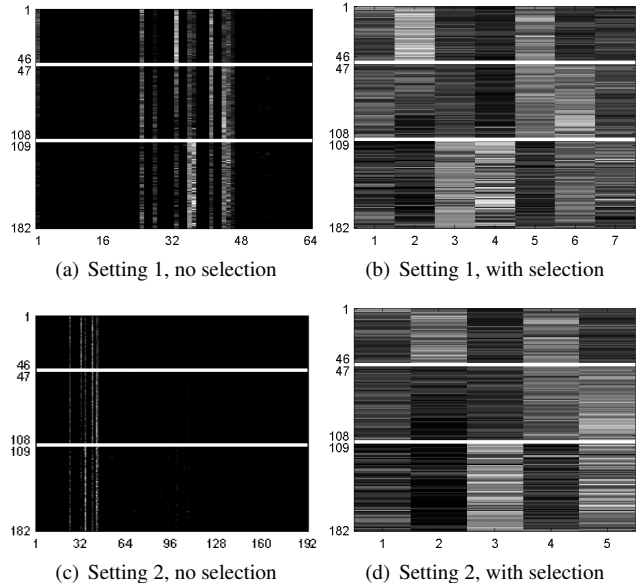


Fig. 2. Example spatial relationship histograms. Rows correspond to image tiles (grouped as dense urban, sparse urban, very sparse urban from top to bottom) and columns correspond to histogram bins. Brighter values correspond to larger values in the histogram.

the sparseness problem was encountered when the number of relationships and the order used increased. Feature selection was incorporated for automatic selection of the most discriminant histogram bins. We observed that the selected bins generally corresponded to meaningful spatial relationships related to their associated classes. The bins selected for setting 1 are listed in Table 2. As can be seen from this list, the subset for the *dense urban* class contained the relationships depending on roofs, including the (roof NEAR roof) relationship, whereas the *sparse urban* class had the relationship (bare soil NEAR street) in addition to the relationships including roofs. When the *very sparse urban* class was considered, no relationship containing a roof was selected. We can conclude that,

Table 2. Region groups selected for setting 1.

Dense urban	Sparse urban	Very sparse urban
(roof NEAR roof)	(roof NEAR bare soil)	(bare soil NEAR grass)
(roof NEAR grass)	(bare soil NEAR street)	(grass NEAR street)
(roof NEAR street)		

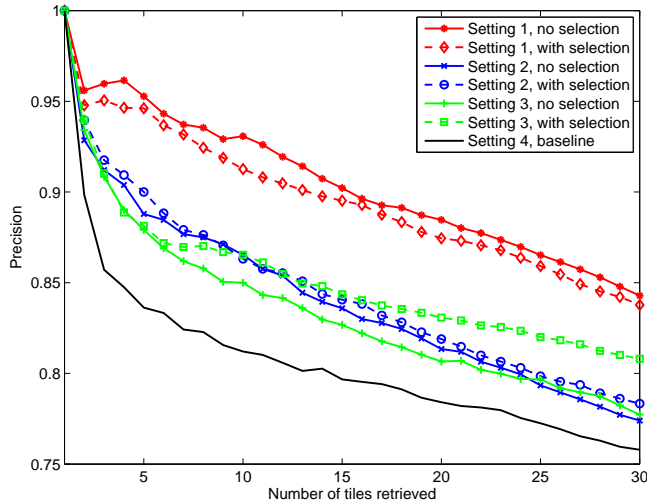


Fig. 3. Average precision for different settings.

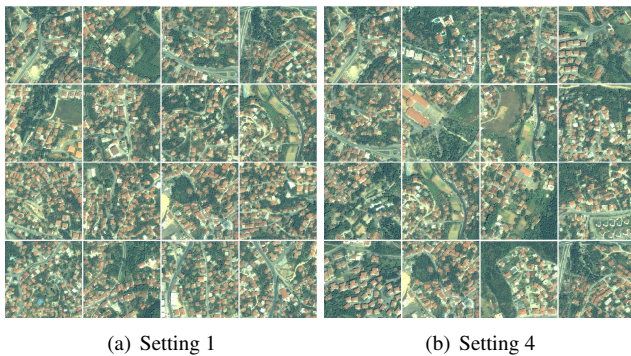


Fig. 4. Retrieval examples for the dense urban class.

when the degree of urbanization decreased, the importance of the relationships regarding roof regions diminished.

The retrieval performance was evaluated by using each image tile as a query and ranking all tiles in increasing order of the Euclidean distance between their histograms and the histogram of the query. Precision, which is defined as the percentage of the correctly retrieved tiles among all tiles retrieved, was computed for quantitative performance analysis. The results for different settings are shown in Figure 3. All settings that encoded spatial relationships outperformed the baseline method that did not use any spatial information. Selection also had a positive effect when the amount of sparseness in the histogram increased. Figure 4 shows two example retrievals for the dense urban class. The results obtained by the spatial relationship histogram were almost all correct, but the baseline method returned some tiles belonging to the sparse urban class because it could not distinguish a large number of small buildings from a smaller number of large buildings, and several small grass areas scattered around the buildings from larger areas of grass.

7. CONCLUSIONS

We described a new image content representation using spatial relationship histograms that were computed by counting the number of times different groups of regions were observed in an image. The high level information encoded in each group consisted of the class labels for all regions and the topological and distance-based spatial relationships between these regions. We also described an algorithm for finding distinguishing region groups for different types of scenes. The selection process produced very compact but very effective representations by significantly reducing the dimensionality of the histograms and the corresponding computational cost of image mining. Image retrieval experiments using Ikonos images showed that the new model resulted in better precision values compared to the traditional representations that did not use any spatial information.

8. REFERENCES

- [1] M. Datcu, H. Daschiel, A. Pelizzari, M. Quartulli, A. Galoppo, A. Colapicchioni, M. Pastori, K. Seidel, P. G. Marchetti, and S. D’Elia, “Information mining in remote sensing image archives: system concepts,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 12, pp. 2923–2936, December 2003.
- [2] J. Li and R. M. Narayanan, “Integrated spectral and spatial information mining in remote sensing imagery,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 3, pp. 673–685, March 2004.
- [3] 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.
- [4] C.-R. Shyu, M. Klaric, G. J. Scott, A. S. Barb, C. H. Davis, and K. Palaniappan, “GeoIRIS: Geospatial information retrieval and indexing system — content mining, semantics modeling, and complex queries,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 4, pp. 839–852, April 2007.
- [5] S. Aksoy, “Modeling of remote sensing image content using attributed relational graphs,” in *Proceedings of 11th IAPR International Workshop on Structural and Syntactic Pattern Recognition*, Hong Kong, August 17–19, 2006, pp. 475–483, Lecture Notes in Computer Science, vol. 4109.
- [6] D. Gokalp and S. Aksoy, “Scene classification using bag-of-regions representations,” in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Beyond Patches Workshop*, Minneapolis, Minnesota, June 23, 2007.