Image Mining Using Directional Spatial Constraints

Selim Aksoy, *Member, IEEE* and R. Gökberk Cinbiş

Abstract—Spatial information plays a fundamental role in building high-level content models for supporting analysts' interpretations and automating geospatial intelligence. We describe a framework for modeling directional spatial relationships among objects and using this information for contextual classification and retrieval. The proposed model first identifies image areas that have a high degree of satisfaction of a spatial relation with respect to several reference objects. Then, this information is incorporated into the Bayesian decision rule as spatial priors for contextual classification. The model also supports dynamic queries by using directional relationships as spatial constraints to enable object detection based on the properties of individual objects as well as their spatial relationships to other objects. Comparative experiments using high-resolution satellite imagery illustrate the flexibility and effectiveness of the proposed framework in image mining with significant improvements in both classification and retrieval performance.

Index Terms—Spatial relationships, image classification, image retrieval, object detection, mathematical morphology

I. INTRODUCTION

The goal of image information mining in geospatial data archives is to automate the content extraction and exploitation process by building high-level subjective content models by combining low-level features and supporting classification and content-based retrieval in terms of semantic queries. In addition to a large number of content-based retrieval systems proposed in the computer vision literature, several systems have been specifically designed for mining Earth observation data. 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. Shyu *et al.* [2] developed an extensive system that supports both tile-based and object-based indexing.

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, the GeoIRIS system in [2] supports the retrieval of tiles according to the spatial configuration of the objects they contain. The VisiMine system we developed includes automatic methods for

extraction of topological, distance-based and relative position-based relationships between region pairs [3] where such relationships can be successfully used for image classification and retrieval in scenarios that cannot be expressed by traditional pixel- and region-based approaches.

Both location and direction play a fundamental role in the modeling and analysis of geospatial information. In this paper, we describe a framework for contextual classification and retrieval in geospatial data where the main goal is to enable object detection based on the properties of individual objects as well as their directional spatial relationships to other objects. The paper builds on our work on morphological modeling of relative position-based spatial relationships [4] (Section II). Most of the existing methods for defining relative positions rely on angle measurements between points of objects of interest. The angle between object centroids or the histogram of angles between all pairs of points have been used. However, the former can give quite counterintuitive results when the objects do not have compact shapes, and the latter is often computationally expensive. The morphological models we developed define a fuzzy landscape where each image point is assigned a value that quantifies its relative position according to a reference object. Mathematical morphology provides a strong basis for such formulation to incorporate the influence of the shape of the object. Furthermore, the fuzzy representation enables flexibility to the imprecision and subjectivity inherent in the definitions of the relationships.

Our main contributions in this paper include extending the relationship model to multiple reference objects, incorporating the spatial information into the Bayesian decision rule as spatial priors for contextual classification, and enabling dynamic queries by using directional relationships as spatial constraints with support for the visibility of image areas that are partially enclosed by reference objects (Section III). We illustrate the effectiveness of the proposed methods using quantitative and qualitative results on contextual classification and retrieval of high spatial resolution satellite imagery (Section IV).

II. DIRECTIONAL SPATIAL RELATIONSHIPS

Position-based spatial relationships describe the spatial arrangements of objects relative to each other. These relationships can be modeled with respect to a direction of interest. In this section, we describe how image areas that have a high degree of satisfaction of a particular directional relationship relative to a reference object can be identified.

Given a reference object B and a direction specified by the angle α, the landscape $\beta_\alpha(B)$ around the reference object along the given direction can be defined as a fuzzy function from the image space $\mathcal{I}$ into $[0, 1]$. The fuzzy membership value $\beta_\alpha(B)(x)$ of an image point $x \in \mathcal{I}$ corresponds to the

The authors are with Department of Computer Engineering, Bilkent University, Ankara, 06800, Turkey. Email: saksoy@cs.bilkent.edu.tr.

degree of its satisfaction of the spatial relation. Given the unit vector along the direction α with respect to the horizontal axis, Bloch [5] suggested that the angle $\theta_\alpha(x, b)$ measured between this vector and the vector from a point b in the reference object to the image point x corresponds to the visibility of the image point from the reference object in the direction α. In [5], the fuzzy landscape is computed as

$$\beta_\alpha(B)(x) = \max\left\{0, 1 - \frac{2}{\pi}\min_{b \in B}\theta_\alpha(x, b)\right\} \quad (1)$$

using a function linearly decreasing with the smallest such angle by considering all points in the reference object. It can be shown that (1) can be computed using the morphological dilation of B,

$$\beta_\alpha(B)(x) = (B \oplus \nu_\alpha)(x) \cap B^c, \quad (2)$$

using the fuzzy structuring element

$$\nu_\alpha(x) = \max\left\{0, 1 - \frac{2}{\pi}\theta_\alpha(x, o)\right\} \quad (3)$$

where o is the origin (center) of the structuring element. B is removed from the result of dilation in (2).

Figure 1(c) shows the landscape corresponding to the east of a building detected in an Ikonos image. (More examples using synthetic images can be found in [6].) The linear function in (3) often leads to a large spread and unintuitive transitions when the angle departs from α especially at points that are farther away from the reference object. We developed a more intuitive and flexible structuring element using a nonlinear function with the shape of a Bézier curve

$$\nu_{\alpha,\lambda}(x) = g_\lambda\left(\frac{2}{\pi}\theta_\alpha(x, o)\right) \quad (4)$$

where $\lambda \in (0, 1)$ determines the inflection point of the curve (see [6] for the derivation). Increasing λ increases the spread around α. This definition can be further extended to decrease the degree of a point's spatial relation to a reference object according to its distance to that object by introducing a new linear term

$$\nu_{\alpha,\lambda,\tau}(x) = g_\lambda\left(\frac{2}{\pi}\theta_\alpha(x, o)\right)\max\left\{0, 1 - \frac{\|\vec{ox}\|}{\tau}\right\} \quad (5)$$

where $\|\vec{ox}\|$ is the Euclidean distance of point x from the structuring element's center and τ is a threshold corresponding to the distance where a point is no longer visible from the reference object. The definition in (5) provides a structuring element that is tunable along both angular and radial dimensions. As can be seen in Figure 1, the landscapes obtained using (4) and (5) are more intuitive and have more compact support compared to the one obtained using (3).

Another important issue especially for remote sensing images that contain complex natural and man-made structures is the handling of image areas that are partially or fully enclosed by the reference object and are not visible from image points along the direction of interest. The landscape definition (2) using any of the structuring elements (3), (4) and (5) can give high values at such areas as shown in Figure 2. The visibility of these areas can be correctly handled as

$$\beta_{\alpha,\lambda,\lambda',\tau,\tau'}(B)(x) = (B \oplus \nu_{\alpha,\lambda,\tau})(x) \cap (B \oplus \nu_{\alpha+\pi,\lambda',\tau'})(x)^c \quad (6)$$
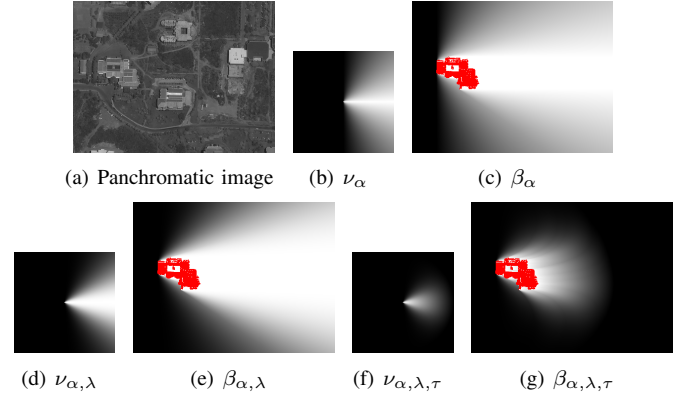

Fig. 1. An Ikonos panchromatic image and the directional landscapes to the east of a detected building using the parameters $\alpha = 0$, $\lambda = 0.3$ and $\tau = 150$. β_α in (c) produces a large spread and unintuitive transitions when the angle departs from α especially at points that are farther away from the reference object. $\beta_{\alpha,\lambda}$ and $\beta_{\alpha,\lambda,\tau}$ in (e) and (g), respectively, result in more intuitive landscapes with more compact support.
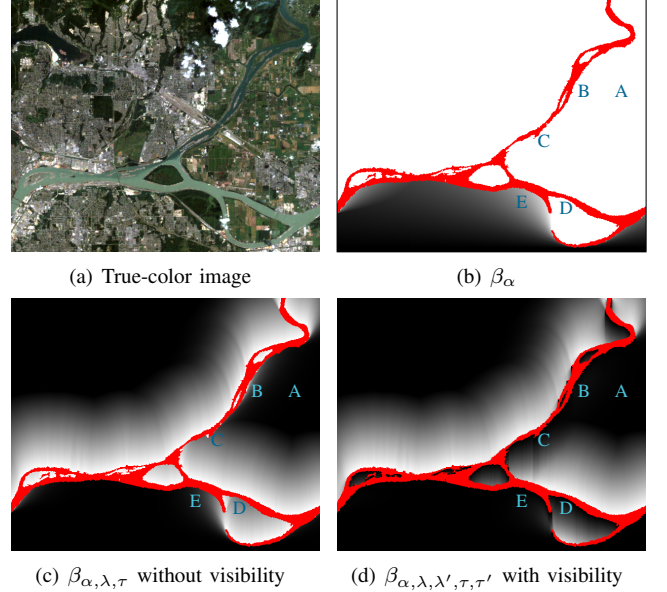

Fig. 2. A Landsat true-color image and the directional landscapes without and with visibility to the north of a detected river object using the parameters $\alpha = \pi/2$, $\lambda = 0.3$, $\lambda' = 0.001$, $\tau = 150$ and $\tau' = 100$. β_α in (b) produces a maximum value of 1 for the points A–D and a large value for E even though these points are more to the south than to the north of a river segment. $\beta_{\alpha,\lambda,\tau}$ in (c) gives a correct value for A. The values for B–E are closer to 0 but are still positive due to the spread of the structuring element. $\beta_{\alpha,\lambda,\lambda',\tau,\tau'}$ in (d) gives the most intuitive results for all points.

where the fuzzy intersection is computed using multiplication and the fuzzy complement is computed by subtracting the values from 1. λ' can be set to a very small number to consider only the image points along $\alpha+\pi$, and τ' can be set to a value less than τ to allow a positive landscape value at the enclosed areas that are closer to one part of the object along α than other parts (see Figure 2 for examples).

Examples in Figures 1 and 2 show that the fuzzy landscape definitions in (2) and (6) using the structuring elements in (4) and (5) provide intuitive and flexible methods for distinguishing image areas for which the directional spatial relationships relative to a reference object hold.

III. Contextual Classification and Retrieval

Once the fuzzy directional landscape β_α is obtained for a reference object B,[1] the degree of satisfaction of this relation by a target object A can be quantified by integrating the landscape over the support of the target object as

$$\mu(A) = \frac{1}{area(A)} \sum_{a \in A} \beta_\alpha(B)(a). \tag{7}$$

Furthermore, the relationship model can easily be extended when there is more than one reference object. Given the landscapes $\beta_{\alpha_1}(B_1), \ldots, \beta_{\alpha_n}(B_n)$ for n reference objects $B_1, \ldots, B_n$ with n possibly different directions of interest $\alpha_1, \ldots, \alpha_n$, the combined relationship can be obtained as

$$\beta_{\alpha_1, \ldots, \alpha_n}(B_1, \ldots, B_n)(x) = \min_{i=1,\ldots,n} \beta_{\alpha_i}(B_i)(x). \tag{8}$$

The "min" operator is used as the equivalent of the Boolean "and" in fuzzy logic. The degree of satisfaction of the combined relation by another object can be computed as in (7).

The computed landscape can be used as contextual information in classification and retrieval for automatically improving the accuracy of image mining. In the remote sensing literature, classification is conventionally done using pixels or objects (regions) with their spectral or textural features. Even though both object-based classification and textural features make use of spatial information through neighboring pixels, they are still far from exploiting any high-level contextual information. Therefore, a significant amount of commission is still unavoidable among the classes with similar low-level features.

Let $\mathbf{x}$ denote the feature vector of a pixel or an object at location x in a binary classification problem with two classes w_1 and w_2. As a widely used solution, the Bayesian classifier makes a decision using the posterior probabilities as

$$\text{Decide} \quad \begin{cases} w_1 & \text{if } \frac{P(w_1|\mathbf{x})}{P(w_2|\mathbf{x})} > 1 \\ w_2 & \text{otherwise} \end{cases} \tag{9}$$

which is equivalent to

$$\text{Decide} \quad \begin{cases} w_1 & \text{if } \frac{P(\mathbf{x}|w_1)}{P(\mathbf{x}|w_2)} > \frac{P(w_2)}{P(w_1)} \\ w_2 & \text{otherwise} \end{cases} \tag{10}$$

using the Bayes rule with the class-conditional and prior probabilities. The equal priors assumption ($P(w_1) = P(w_2)$) is often used when no additional information is available.

Assume that there is a third class w_3 that is related to w_2. The pixels/objects that are assigned to w_3 can be used as spatial constraints for improving the discrimination between w_1 and w_2. First, the directional landscape $\beta_\alpha(w_3)$ is computed for the whole scene by using w_3 as the reference. Then, the fuzzy landscape value in the range $[0,1]$ at each image location is used as the spatial prior for w_2 at that location, i.e., $P(w_2) = \beta_\alpha(w_3)(x)$ and $P(w_1) = 1 - P(w_2)$. The resulting contextual decision rule becomes

$$\text{Decide} \quad \begin{cases} w_1 & \text{if } \frac{P(\mathbf{x}|w_1)}{P(\mathbf{x}|w_2)} > \frac{\beta_\alpha(w_3)(x)}{1 - \beta_\alpha(w_3)(x)} \\ w_2 & \text{otherwise} \end{cases} \tag{11}$$

---

[1]The landscape is denoted as β_α to simplify the notation in this section. Any definition and structuring element from Section II can be used for β.

using these spatial priors. We illustrate the use of (11) for the classification of asphalt (w_1) versus shadow (w_2) using buildings and trees as the spatial reference (w_3) in Section IV-A. The extension of (11) for multi-class classification with multiple reference classes is straightforward but is not included in this paper due to space constraints.

The directional landscapes can also be used for image retrieval for geospatial intelligence in both civilian and military applications. Existing methods with pre-computed spatial relationships within fixed partitions (tiles) cannot handle dynamic queries formulated as a search for objects with certain properties (spectral, textural, shape, etc.) at a particular relative position with respect to other reference objects. Methods for answering dynamic queries can be found in GIS and spatial database literatures but the former is often limited to topological (adjacency) and distance constraints, and the latter often assumes that the objects are represented using single points (e.g., centroids) or bounding rectangles. Such assumptions are often violated by complex natural and man-made structures in remote sensing images.

The proposed directional models are promising solutions for dynamic queries due to their flexibility for any type of objects with support for the notion of visibility as described in Section II. Given the objects detected in a data set with possibly the confidence of detection and a list of attributes (features) for each object, dynamic queries for objects having certain attributes and satisfying several additional spatial relationship constraints with respect to multiple reference objects can be answered as follows. First, the fuzzy landscape is computed as in (8). Then, the degrees of satisfaction of these relationships by objects that satisfy the attribute criteria are found as in (7). Finally, the objects are ranked according to a combined measure (e.g, product, sum, weighted sum) that involves the confidence of detection, the attribute values, and the spatial constraints. We illustrate such queries in Section IV-B.

IV. Experiments

Performance analysis of high-level spatial relationship models is a very difficult and subjective task where synthetic images [5] or rotated or scaled versions of real images [2] were used for evaluation in the literature. We present proof-of-concept experiments to illustrate the use of directional landscape models as spatial contextual constraints for image classification and retrieval. The models described in this paper were implemented in Matlab. Parallel and faster implementations are possible but are beyond the scope of this paper.

A. Classification Experiments

We used a well-known hyperspectral image of Pavia, Italy obtained by the ROSIS sensor. A pixel-based Bayesian classification was performed for 9 classes using spectral and textural features as described in [7]. The output of the Bayesian classifier was a probability value for each class at each pixel.

Figures 3(a) and 3(b) show a 490×199 pixel section of this image and the corresponding classification map. Table I(a) shows the confusion matrix for the pixels where either the asphalt class (w_1) or the shadow class (w_2) gave the highest

TABLE I
CONFUSION MATRICES FOR ASPHALT VERSUS SHADOW CLASSIFICATION.

(a) Using the decision rule (9) without spatial information. Overall accuracy is %63.66.

		Assigned	
		asphalt	shadow
True	asphalt	3,302	2,000
	shadow	1,028	2,003

(b) Using the decision rule (11) with spatial information. Overall accuracy is %86.16.

		Assigned	
		asphalt	shadow
True	asphalt	5,054	248
	shadow	905	2,126

(a) True-color image

(b) Classification map using decision rule (9) without spatial information

(c) Directional landscape with respect to the detected tiles

(d) Directional landscape with respect to the detected trees

(e) Ground truth map

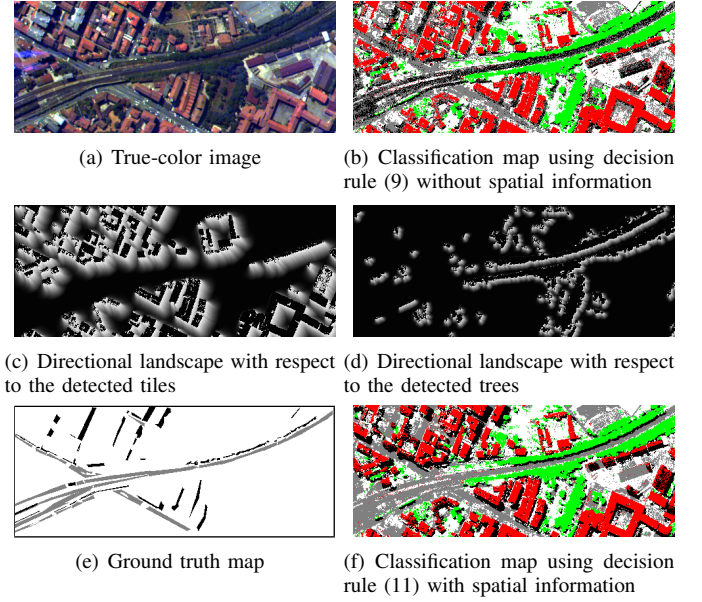(f) Classification map using decision rule (11) with spatial information

Fig. 3. Classification of the Pavia image without and with using spatial contextual information. The classes in the classification and ground truth maps are asphalt (gray), shadow (black), tiles (red) and trees (green). The ground truth is produced by visual inspection.

probability. Figure 3(e) was used as the ground truth. The %63.66 accuracy shows a significant amount of commission between these two classes when only pixel-based information is used. This is a common problem in the classification of images with high spatial resolution. It occurs because of the mismatches between the land cover/use classes in the ground truth used for training the classifiers and the land cover/use observed in the image being classified. These mismatches that are caused by external factors such as the position of the sun and clouds at the time of image capture make the classification time-dependent where the classes with relatively similar spectral values (e.g., water versus shadow, asphalt versus shadow, snow versus cloud) are often misclassified.

The spatial context can be incorporated into the decision process by using the tiles (building roofs) and trees as additional information. The directional landscape was computed for the pixels classified as tiles using the parameters $\alpha = -50°$, $\lambda = 0.3$ and $\tau = 25$. The α value measured from the horizontal axis in counter-clockwise direction was visually determined from the image to approximate the sun angle (can be obtained from the image metadata if available). The λ and τ values were determined empirically. Similarly, the directional landscape for the trees class was computed using the parameters $\alpha = -50°$, $\lambda = 0.3$ and $\tau = 10$. The two landscapes, shown in Figures 3(c) and 3(d), were combined using the "max" operator (which is the equivalent of the Boolean "or" operator). Then, the contextual decision rule in (11) was used to update the classification at each pixel by using tiles and trees as reference (w_3). Figure 3(f) and Table I(b) show the classification results when spatial information was used. The updated contextual decision gave a %86.16 accuracy that corresponds to a net %22.50 improvement by classifying a pixel with shadow-like feature values as shadow only when it also has a high degree of directional spatial relationship with respect to buildings or trees at a particular angle.

B. Retrieval Experiments

Retrieval experiments were done using an Ikonos panchromatic image of Antalya, Turkey. Figure 4(a) shows a $1,000 \times 700$ pixel section that is part of a university campus. First, derivative of the morphological opening and closing profiles (DMP) were computed using disk structuring elements with radii from 3 to 15. Then, objects were extracted by simple thresholding of the DMP levels. No pre- or post-processing of the results were considered as the aim was not to assess the performance of object detection in detail but to illustrate the potential use of the spatial relationship models for retrieval.

Figure 4(b) shows the extracted objects grouped into four categories, namely: building, road, parking lot, and tree. Further processing would have improved the detection results but we considered them as sufficient for retrieval experiments.

Next, several complex queries were constructed to search for different objects when two or more objects were used as spatial constraints as described in Section III. The object types in the grouping in Figure 4(b) were used as the object attributes. For each query, the fuzzy landscape corresponding to multiple reference objects was computed as in (8), and the objects that satisfied both the attribute criterion and the spatial constraints were included in the result set of that query. An object was considered as satisfying the spatial constraints if its degree of satisfaction of the spatial relationship computed using (7) was greater than 0.5.

Retrieval performance was evaluated using precision (percentage of the correctly detected objects among all detections in the result set) and recall (percentage of the correctly detected objects among all objects in the ground truth) using a ground truth that was constructed by manually identifying the objects satisfying each query. Figures 4(c) and 4(d) show 10 queries and the corresponding precision values when the fuzzy directional landscape was computed using the definition in (6) with the structuring element in (4) (proposed model) and using the definition in (1) (model in [5]) for comparison. The α values were determined from the query descriptions in Figure 4(c) and the λ parameter in (4) was fixed at 0.3. The structuring element in (5) with τ was not used in order not to increase the number of parameters during evaluation.

Both models achieved perfect recall for all queries. However, our model resulted in significantly better precision than the one by [5]. The errors by the latter were due to the large spread by the structuring element in (3) and the missing

(a) Panchromatic image

(b) Extracted objects and their IDs

Q1	Find buildings between 68 and 71
Q2	Find buildings to the east of 8 and south of 74
Q3	Find buildings to the east of 19 and north of 22
Q4	Find buildings to the east of 19, north of 22 and west of 46
Q5	Find buildings between 60 and 80
Q6	Find buildings to the east of 16, north of 23 and west of 46
Q7	Find parking lots between 31, 32 and 57
Q8	Find buildings between 37 and 8
Q9	Find parking lots to the east of 46 and northwest of 80
Q10	Find tree groups between 17 and 46

(c) Queries

(d) Precision

(e) Results for Q2

(f) Results for Q6
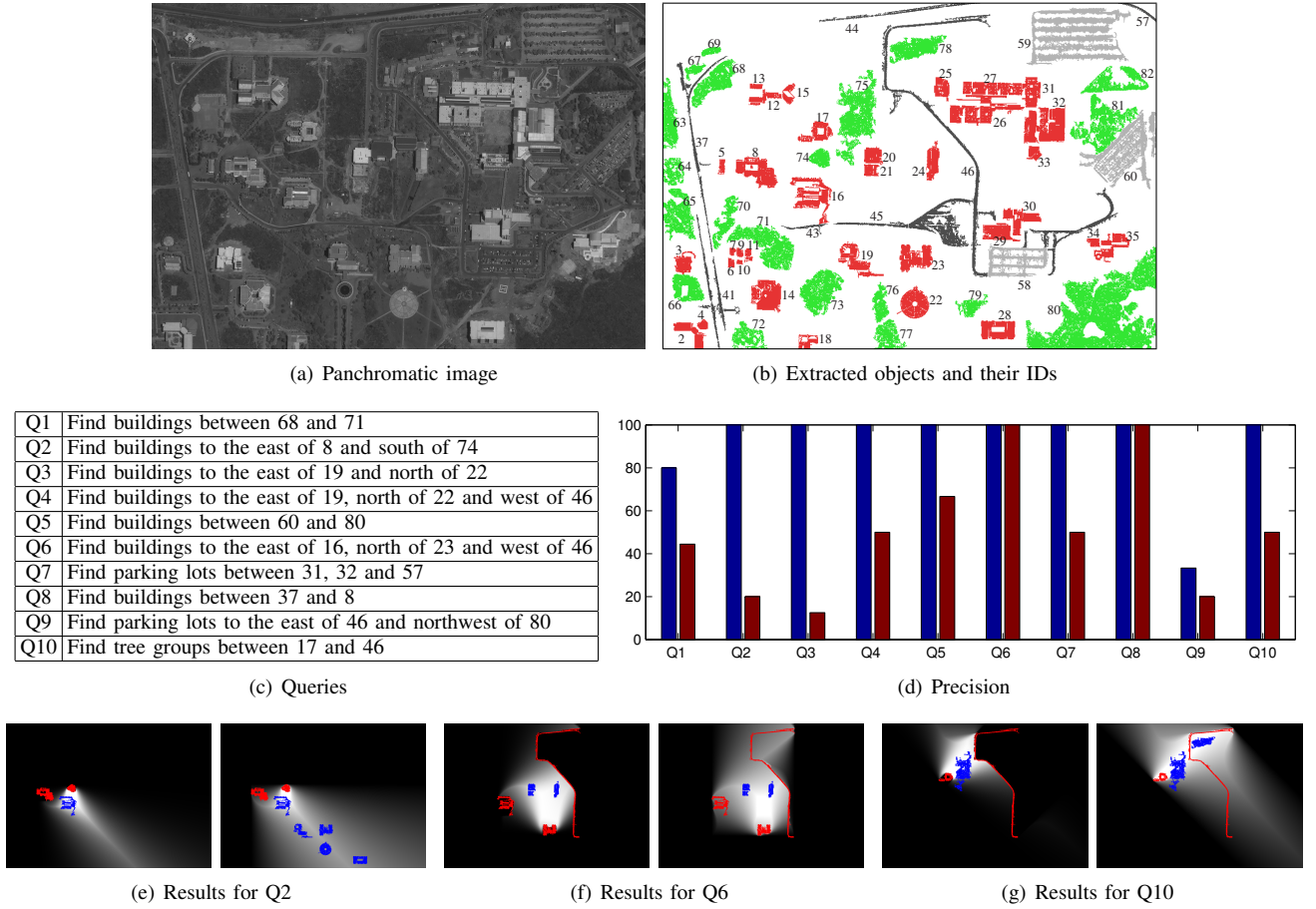

(g) Results for Q10

Fig. 4. Retrieval experiments using the Antalya image. The extracted objects in (b) are labeled as building (red), road (dark gray), parking lot (light gray) and tree (green). The blue (left) bars in (d) show the precision using the directional relationship model proposed in Section II, and the red (right) bars correspond to the model in [5]. The query results in (e)-(g) show the combined landscape, the reference objects (red) and the detected objects (blue). For each query, the image on the left shows the result for the proposed model and the one on the right shows the result using the model in [5].

support for visibility to handle the areas that were partially enclosed by complex structures (e.g., roads) (see Figures 4(e)–4(g) for examples). Overall, the directional landscapes obtained by the proposed model using multiple reference objects were more intuitive than the ones by the compared method as also reflected in the precision values. Even further improvements can be obtained by using a "between" model that uses the definition in (6) but also handles the cases where one object is significantly spatially extended relative to others by taking spatial proximity into consideration [4], [6].

V. Conclusions

This paper presented our work on modeling directional spatial relationships by automatically identifying image areas for which such relationships relative to several reference objects hold, and using this information as spatial constraints for contextual classification and retrieval. Experiments using high-resolution satellite imagery showed that the Bayesian decision rule that incorporated spatial information significantly decreased the amount of commission among spectrally similar classes. Retrieval experiments also showed that the proposed models produced more intuitive results and higher precision than other approaches in dynamic query scenarios with spatial constraints.

References

[1] M. Datcu, H. Daschiel, A. Pelizzari, M. Quartulli, A. Galoppo, A. Co-lapicchioni, 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] 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.

[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] R. G. Cinbis and S. Aksoy, "Relative position-based spatial relationships using mathematical morphology," in *Proceedings of IEEE International Conference on Image Processing*, vol. II, San Antonio, Texas, September 16–19, 2007, pp. 97–100.

[5] I. Bloch, "Fuzzy relative position between objects in image processing: A morphological approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 7, pp. 657–664, July 1999.

[6] R. G. Cinbis and S. Aksoy, "Modeling spatial relationships in images," Department of Computer Engineering, Bilkent University, Ankara, Turkey, Tech. Rep. BU-CE-0702, January 2007, http://retina.cs.bilkent.edu.tr/papers/BU-CE-0702.pdf.

[7] S. Aksoy, "Spatial techniques for image classification," in *Signal and Image Processing for Remote Sensing*, C. H. Chen, Ed. Taylor & Francis Books, 2006, pp. 489–511.