Probabilistic Retrieval with a Visual Grammar

Selim Aksoy, Giovanni Marchisio, Krzysztof Koperski, Carsten Tusk
Insightful Corporation
1700 Westlake Ave. N., Suite 500
Seattle, WA, 98109-3044
{saksoy,giovanni,krisk,ctusk}@insightful.com

Abstract—**We describe a system for content-based retrieval and classification of multispectral images. Our system models images on pixel, region and scene levels. To reduce the gap between low-level features and high-level user semantics, and to support complex query scenarios that consist of many regions with different feature characteristics, we propose a probabilistic visual grammar that includes automatic identification of region prototypes and modeling of their spatial relationships. A Bayesian framework is used to automatically classify scenes based on these models. We demonstrate our system with query scenarios that cannot be expressed by traditional region or scene level approaches but where the visual grammar provides accurate classifications and effective retrieval.**

I. Introduction

Automatic content extraction, classification and retrieval are highly desired goals in intelligent remote sensing databases. Most of the proposed approaches use low-level features like spectral values and texture features to index images and then use distance measures in these feature spaces to find similarities between them. However, there is a large semantic gap between the low-level features and the high-level user expectations and search scenarios.

The VisiMine system [1] supports interactive classification and retrieval of multispectral images by modeling them on pixel, region and scene levels. Pixel level characterization includes spectral bands, spectral unmixing for surface reflectance, Gabor, co-occurrence and Laws texture features, line-angle-ratio statistics, and DEM information. After the features are computed for each pixel, an automatic region segmentation algorithm is used to compute an approximate polygon decomposition of each scene. Then, region level features are computed using moments for shape and orientation information, and statistics of pixel features within and around individual regions.

To reduce the gap between low-level features and high-level user semantics, Schroder *et al.* [2] developed a Bayesian label training algorithm that operates on individual pixels. Traditional region or scene level search algorithms assume that the regions or scenes consist of uniform pixel feature distributions. However, complex query scenarios usually contain many pixels and regions that have different feature characteristics. Furthermore, two images with similar regions can have very different interpretations if the regions have different spatial arrangements. Therefore, we need a higher level visual grammar to describe these scenarios.

This work is supported by NASA SBIR contracts NAS5-98053 and NRA2-37143.

Previous approaches for modeling spatial relationships of regions [3] include manual delineation by experts and construction of graph models that are powerful representations but are not usable due to the infeasibility of manual annotation in large remote sensing databases. In this paper we describe an automatic probabilistic framework that includes prototypes of primitive regions, their spatial relationships, and automatic and supervised algorithms to use them for content-based retrieval and classification.

II. Prototype Regions

The first step to construct a visual grammar is to find meaningful regions in an image. To mimic the identification of regions by experts, we define the concept of prototype regions. A prototype region is a region that has a relatively uniform low-level pixel feature distribution and describes a simple scene or part of a scene. Ideally, a prototype is frequently found in a specific class of scenes and differentiates this class of scenes from others. Also, using prototypes reduces the number of associations between regions and makes the combinatorial problem more tractable.

VisiMine uses unsupervised model-based clustering to automate the process of finding prototypes. We use a Gaussian mixture model where each component corresponds to a prototype. The maximum *a posteriori* probability (MAP) rule is used to assign a prototype label to each region with the degree of match being the posterior probability of the prototype given the feature vector of that region. Interesting prototypes in remote sensing images can be cities, rivers, lakes, residential areas, tidal flats, forests, fields, snow, clouds, etc. An extension of this prototype framework can be to use supervised algorithms like Bayesian label training or relevance feedback to learn and improve the prototype models.

III. Region Relationships

A. Second-order Region Relationships

Second-order region relationships consist of the relationships between region pairs. These pairs can occur in the image in many possible ways. However, the regions of interest are usually the ones that are close to each other. Representations of spatial relationships depend on the representations of regions. In VisiMine, regions are represented by their boundary pixels. Other possible representations include minimum bounding rectangles, centroid-based and graph-based approaches [3].

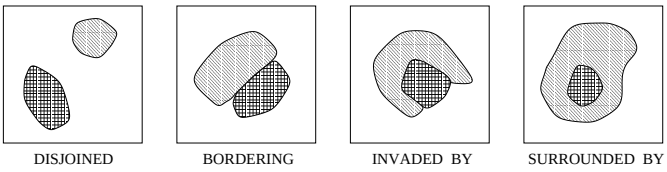

Fig. 1. Spatial relationships of region pairs: *disjoined*, *bordering*, *invaded_by* and *surrounded_by*.

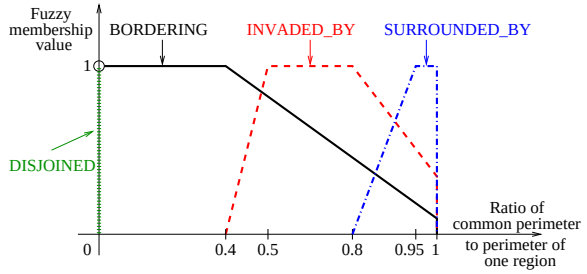

Fig. 2. Fuzzy membership functions for spatial relationships based on region perimeter ratios.

The spatial relationships between all region pairs in an image can be represented by a region relationship matrix. For a pair of regions, we first compute

- perimeter of the first region, π_i
- perimeter of the second region, π_j
- common perimeter between two regions π_{ij}

where $i, j \in \{1, \ldots, n\}$ and n is the number of regions in the image. The $n \times n$ region relationship matrix is defined as $R = \{r_{ij} = \frac{\pi_{ij}}{\pi_i} \,|\, i, j = 1, \ldots, n, \ \forall i \neq j\}$. Then, relationships can be derived by quantizing the r_{ij} values. Quantization gives crisp (Boolean) decisions about r_{ij} which may have limited expressiveness. A more flexible method is to define the relationships as relationship classes. Each region pair can be assigned a degree of their spatial relationship using fuzzy class membership functions. The pairwise relationships used in VisiMine are shown in Fig. 1. The class membership functions are denoted as Ω_c where c is one of *disjoined*, *bordering*, *invaded_by* and *surrounded_by*. Then, the value $\Omega_c(r_{ij})$ represents the degree of membership of regions i and j to class c. We use the trapezoidal membership functions shown in Fig. 2.

The motivations for the choice of these functions are as follows. Two regions are disjoined when they are not touching each other. They are bordering each other when they have a common perimeter. When the common perimeter gets closer to 50%, the larger region starts invading the smaller one. When the common perimeter goes above 80%, the relationship is considered an almost complete invasion, i.e. surrounding.

The class membership functions are chosen so that only one of them is the largest for a given perimeter ratio. To have the relationship between the region pair i and j uniquely defined, we label them as having the relationship

$$c_{ij} = \arg\max_c \Omega_c(\max\{r_{ij}, r_{ji}\}) \tag{1}$$

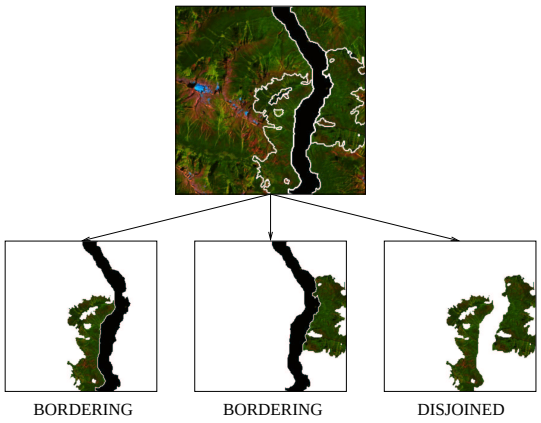

Fig. 3. Spatial relationships of three regions decomposed into second-order relationships. This particular example is used to recognize mountain lakes with surrounding hills with trees.

with the degree

$$d_{ij} = \Omega_{c_{ij}}(\max\{r_{ij}, r_{ji}\}). \tag{2}$$

B. Higher-order Region Relationships

Higher-order region relationships can be decomposed into multiple second-order relationships. The equivalent of the Boolean "and" operation in fuzzy logic is the "min" operation. The relationship between a combination of k regions can be represented as a list of $\binom{k}{2}$ pairwise relationships using (1) as

$$c_{1\ldots k} = \{c_{ij} \,|\, i, j = 1, \ldots, k, \ \forall i < j\} \tag{3}$$

with its degree computed using (2) as

$$d_{1\ldots k} = \min_{\substack{i,j=1,\ldots,k \\ i<j}} d_{ij}. \tag{4}$$

See Fig. 3 for an example decomposition.

IV. Image Retrieval

Users can compose queries for complex scenarios by giving a set of example regions. VisiMine encodes and searches for a query scenario using the proposed visual grammar as follows:

1. Let k be the number of regions selected by the user. Find the prototype label for each of the k regions.
2. Find the perimeter of each of the k regions and the common perimeter for each of the $\binom{k}{2}$ possible region pairs.
3. Find the spatial relationship and its degree among these k regions using (3) and (4). Denote them by $c^* = \{c^*_{ij} \,|\, i, j = 1, \ldots, k, \ \forall i < j\}$ and d^*, respectively.
4. For each image in the database,

(a) For each query region, find the list of regions with the same prototype label as itself. Denote these lists by $U_i, i = 1, \ldots, k$.

(b) Rank region groups $(u_1, u_2, \ldots, u_k) \in U_1 \times U_2 \times \cdots \times U_k$ according to the distance

$$\left| d^* - \min_{\substack{i,j=1,\ldots,k \\ i<j}} \Omega_{c^*_{ij}}(r_{u_i u_j}) \right|.$$

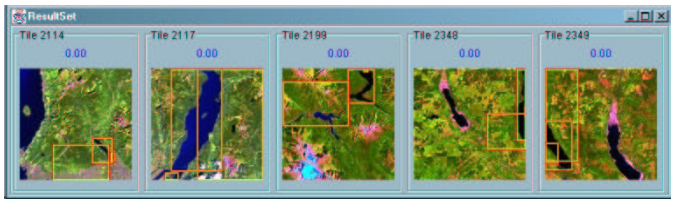

Fig. 4. Top 5 search results for the mountain lake described in Fig. 3. They all include a lake with a similar hill structure around it.

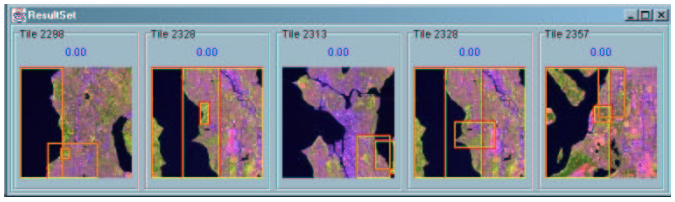

Fig. 5. Top 5 search results for a residential area with a park where both are neighboring water.

(c) The equivalent of the Boolean "or" operation in fuzzy logic is the "max" operation. To rank image tiles, use the distance

$$\left| d^* - \max_{\substack{(u_1,u_2,\ldots,u_k)\in \\ U_1\times U_2\times\cdots\times U_k}} \left\{ \min_{\substack{i,j=1,\ldots,k \\ i<j}} \Omega_{c^*_{ij}}(r_{u_i u_j}) \right\} \right|.$$

Example queries on a LANDSAT database are given in Fig. 4 and 5.

V. Image Classification

The visual grammar can also be used to classify images in a Bayesian framework. The input to the system is a set of example images for each class defined by the user. Let s be the number of classes, m be the number of relationships defined for region pairs (as in Fig. 1), k be the number of regions in a region group, and t be a threshold for the number of region groups that will be used in the classifier. Denote the classes by $w_1, \ldots, w_s$. VisiMine automatically builds classifiers from the training data as follows:

1. Count the number of times each possible region group is found in the set of training images for each class. Compute the variance of the count for each region group across all classes. A region group of interest is the one with a large variance, i.e. the one that is frequently found in a particular class of images but rarely exists in other classes.
2. Select the top t region groups with the largest variances. Let $x_1, \ldots, x_t$ be Bernoulli random variables for these region groups, where $x_j = T$ if the region group x_j is found in an image and $x_j = F$ otherwise. Let $p(x_j = T) = \theta_j$. Then, the number of times x_j is found in images from class w_i has a Binomial(v_i, θ_j) distribution where v_i is the number of training images for class w_i. Using a Beta$(1, 1)$ distribution as the conjugate prior, the Bayes estimate for θ_j is computed as

$$p(x_j = T|w_i) = \frac{v_{ij} + 1}{v_i + 2} \tag{5}$$

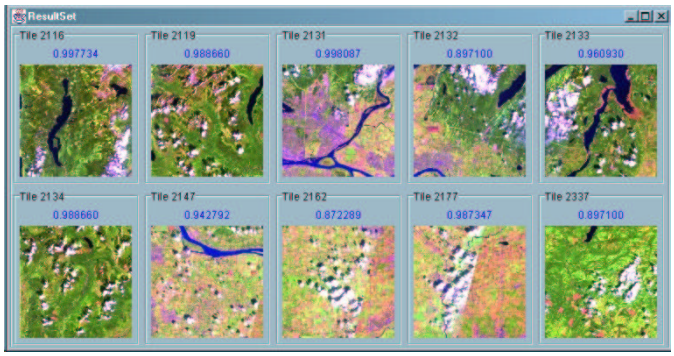

Fig. 6. Classification results for clouds that are modeled by white regions with their neighboring shadows.

where v_{ij} is the number of training images for w_i that contain x_j. Using a similar procedure, the Bayes estimate for an image belonging to class w_i is computed as

$$p(w_i) = \frac{v_i + 1}{\sum_{i=1}^{s} v_i + s}. \tag{6}$$

In other words, discrete probability tables are constructed using v_i and $v_{ij}, i = 1, \ldots, s, j = 1, \ldots, t$, and conjugate priors are used to update them when new images become available via relevance feedback.

3. For an unknown image, search for each of the t region groups and compute the probability for each class using the conditional independence assumption. Assign that image to the best matching class using the MAP rule as

$$\begin{aligned} w^* &= \arg\max_{w_i} p(w_i|x_1, \ldots, x_t) \\ &= \arg\max_{w_i} p(w_i) \prod_{j=1}^{t} p(x_j|w_i). \end{aligned} \tag{7}$$

An example classification is given in Fig. 6.

VI. Conclusions

In this paper we proposed a probabilistic framework to automatically analyze complex query scenarios using spatial relationships of regions and described algorithms to use them for content-based image retrieval and classification. Future work include using supervised methods to update prototype models and developing new spatial relationships like *near*, *far*, *right_of*, *left_of*, *above*, *below*, etc.

References

[1] K. Koperski, G. Marchisio, S. Aksoy, and C. Tusk, "Visimine: Interactive mining in image databases," in *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, Toronto, Canada, 2002.

[2] M. Schroder, H. Rehrauer, K. Siedel, and M. Datcu, "Interactive learning and probabilistic retrieval in remote sensing image archives," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 5, pp. 2288–2298, September 2000.

[3] S. Santini, *Exploratory Image Databases: Content-Based Retrieval*, Academic Press, 2001.