

Modeling Object Classes in Aerial Images Using Texture Motifs

Sitaram Bhagavathy, Shawn Newsam and B.S. Manjunath
Department of Electrical and Computer Engineering
University of California, Santa Barbara, CA 93106
{sitaram,snewsam,manj}@ece.ucsb.edu

Abstract

We propose a canonical model for object classes in aerial images. This model is motivated by the observation that geographic regions of interest are characterized by collections of texture motifs corresponding to the geographic processes that generate them. We show that this model is effective in learning the common texture themes, or motifs, of the object classes.

1. Introduction

Researchers have shown significant interest in using texture descriptors for the automated analysis of aerial images [1-3]. The size of large aerial image collections precludes manual annotation, so being able to use automatically extracted descriptors is very appealing. Homogeneous texture descriptors [4] have proven to be effective in characterizing a variety of basic earth-components, such as water, agricultural fields, etc.

Despite this, texture remains underutilized in the automated analysis of remotely sensed imagery. The work presented in this paper is progress toward using texture descriptors for image analysis at the object level. Distinct spatial signatures result from many natural and man-made geographic processes that create objects of interest. For example, the rows of vines in a vineyard or cars in a parking lot appear as homogeneous textures in aerial images. A major challenge in using homogeneous texture features is that the objects in spatial datasets usually consist of multiple textures. Most golf courses consist of grass-covered fairways lined by trees. Grass and trees each result in distinctive textures but neither feature by itself characterizes a golf course. The fairways and the trees are thus texture themes, or motifs, that characterize the class of golf courses.

The major contribution of this work is a model that utilizes multiple textures to characterize image regions. In particular, texture motifs are used to discover and

characterize the set of geographic processes that create the objects of interest. Our main objective is developing an effective characterization of objects of interest in aerial images.

The geographic processes that create object classes are statistically modeled as mixtures of Gaussians. The models are trained using instances of the object classes. Experimental results show the technique can characterize many objects of interest in spatial datasets, such as airports, harbors, etc. In particular, we show that the models learn the common texture motifs of the object classes.

1.1. Related work

Mixtures of Gaussians have been used to model image feature distributions for a variety of research objectives. In [5], texture-based image segmentation is performed by clustering texture feature vectors using mixtures of Gaussians. Spatial coherence is taken into account to group those texture clusters that correspond to the same image region. This allows the resulting image segmentation to contain regions with multiple textures.

In the Blobworld system [6], mixtures of Gaussians are used to derive image descriptors for content-based retrieval. The Expectation-Maximization (EM) algorithm is used to discover the feature vector groupings that correspond to the visual blobs in an image. A joint color-texture descriptor is extracted from each blob and used to perform similarity retrieval in a database of images.

2. Texture Motif Analysis of Objects

We start out with the basic assumption that geographic regions of interest are characterized by a collection of texture motifs corresponding to the geographic processes that generate the class. Often, users of geographic image collections need to retrieve information on semantically relevant classes of regions, which we term *objects*, e.g. airports and trailer parks in aerial images.

Researchers have created systems that perform region-level [6] and crude object-level [7-9] image content retrieval using properties of homogeneous region segmentations. However, the problem of modeling a general set of semantic classes (such as airports and trailer parks in aerial images) is still unsolved.

2.1. Problem Statement

There are three stages in building an object-based querying and retrieval system for large aerial image collections, namely, (1) modeling the object classes, (2) characterizing objects based on the class models, and (3) defining similarity measures between objects.

In this paper, we focus on the first stage. Given a training set of object instances from several semantic classes, we model the classes based on their underlying texture motifs.

3. Modeling Object Classes

A good model for an object class should capture the texture motifs that characterize it. We model an object class as a mixture of Gaussians, one for each texture motif that characterizes the class. We call our model the *canonical class model* for object classes.

3.1. The Canonical Class Model

Homogeneous texture feature vectors [4] are extracted by applying a set of Gabor-wavelet filters (at 5 scales and 6 orientations) to the aerial images. Let $\mathbf{c}(x)$ denote the 30-dimension feature vector extracted from the neighborhood of pixel x . Assuming that the pixels in an object class are generated by one of N possible texture motifs modeled as Gaussians, the probability density function of \mathbf{c} can thus be expressed as a mixture distribution,

$$p(\mathbf{c}) = \sum_{j=1}^N P(j) p(\mathbf{c} | j) \quad (1)$$

where $p(\mathbf{c}|j)$ is the conditional likelihood of the feature \mathbf{c} being generated by the motif j , and $P(j)$ is the prior probability of the motif j . The number of motifs N along with the distribution means and covariance matrices are the parameters that specify the class model.

We use the EM algorithm to estimate the parameters of the Gaussian mixture model (GMM). A K-means clustering process is used to bootstrap the EM algorithm. Once we determine a GMM for an object class, we can go back to an instance in that class and use a Maximum A Posteriori (MAP) classifier,

$$i^* = \arg \max_{1 \leq i \leq N} [P(i | c(x))] \quad (2)$$

to label each pixel x according to its generating texture motif. The posterior probabilities $P(i|\mathbf{c}(x))$ are obtained using Bayes's rule.

3.2. Training the model

The training set of object instances \mathcal{O}_c from a class c , is denoted as follows:

$$\mathcal{O}_c = \{O_{1c}, O_{2c}, \dots, O_{Mc}\}, \quad (3)$$

where M is the number of instances.

Given \mathcal{O}_c , the step-by-step process of constructing the canonical class model for class c is as follows:

1. For each object in the training set, extract the 30-dimensional texture features described in Section 3.1.
2. Randomly choose P training feature points within each object in Equation (3). We choose P according to the cardinality of the training set M , thus keeping the computations tractable.
3. Choose the number of texture motifs N for the model.
4. Train the GMM in Equation (1) using the union of all the training points from the objects in the training set. Each Gaussian in the mixture corresponds to one texture motif characterizing the object class.
5. Using the model from Step 4, and the MAP classifier in Equation (2), we identify the texture motifs that generate the pixels of any given instance of class c .

3.3. Merits of the model

The canonical class model captures the global characteristics of an object class in terms of the texture motifs that characterize the whole class. By training a GMM using several instances from an object class, the model learns the texture motifs that are important in characterizing the class. Because of this capability, the canonical class model has the following merits:

1. It enables the semantic analysis of the texture motifs that characterize a class. We discuss this concept in Section 5.
2. Each texture motif characterizes the varying orientations of a particular texture. The GMM clusters together textures that are similar but occur at different orientations.
3. Because the model globally analyses the class, we expect it to de-emphasize irrelevant textures that occur in a few instances, but do not characterize the class in general.

4. Experiments and Observations

The dataset contains objects from the Digital Orthophoto Quarter-Quadrangle (DOQQ) coverage of California. For ease of availability, we choose the following six object classes: airports, golf courses, harbors, high schools, mobile home parks, and vineyards. The geo-referencing information in the Alexandria Digital Library Gazetteer [10] is used to approximately locate several instances from each object class. In each case, we extract a rectangular image region containing the object, and manually create a binary mask that defines the object boundary.

Two experiments are conducted to understand (1) the usefulness of the canonical model, (2) the effect of increasing the training set size, and (3) the concept of rotation invariance. For simplicity, $N=5$ in Equation (1), for both experiments. The issue of choosing the number of motifs to model a given class is discussed in Section 5.

4.1. Experiment 1

In this case, the training set for a class consists of a single object instance, i.e. $M=1$ in Equation (3). P is chosen to be equal to 25,000 in Step 2 of the training process. Figures 1(a) and 2(a) show two object instances from the harbor class. Each of these instances is used in turn to create two separate models for the harbor class. The two important texture motifs in this class are the moored boats and the water. Figures 1(b) and 2(b) show the texture motif assignments for both instances, with a common color for each label.

Observations:

1. The moored-boats motif is mostly mapped to red in Figure 1(b), and to yellow in Figure 2(b). Since the object instances are modeled separately, there is no correspondence between the motif assignments for different instances of a class.
2. In Figure 1(b), a portion of the moored-boats motif is mapped to yellow because the boats have a different orientation. In this case, the model is sensitive to the orientations of the texture motifs.
3. A small training set is ineffective for class modeling because a few object instances do not represent a class well.

4.2. Experiment 2

In this case, $M=10$, and $P=10,000$. Thus each class is trained using 100,000 representative feature points. Figures 1(c) and 2(c) show the texture motif assignments for the harbor instances in Figures 1(a) and 2(a), respectively.

Observations:

1. In Figure 1(c) and 2(c), the moored-boats and water motifs are almost uniformly mapped to the same

color for both instances, irrespective of orientation. This leads to the following two observations.

2. As the training set size is increased, the canonical model characterizes the class globally, and thus assigns the same label to a particular texture motif for all instances of a class. This illustrates the potential that this model has for semantic analysis of object classes (see Section 5).
3. As the training set size is increased, the model characterizes the motifs in a more rotation invariant manner.

5. Conclusion and Future Work

This paper proposes a canonical class model for object classes in aerial images. Initial experiments show that the model succeeds in capturing the texture motifs that characterize object classes. The model characterizes the motifs in a more rotationally invariant manner as the training set size is increased. The effect of training set size on the model is being investigated as well as whether the model generalizes well to other object classes.

The next level of abstraction in the canonical class model is to analyze the texture motifs semantically. This would involve mapping the motif labels to named motifs like grassy area, road, water, housing development, etc. This analysis will help in tackling research problems like automatic object segmentation, object-based image retrieval, etc.

An open issue in using this model is determining the number of texture motifs that most effectively characterize a class. One solution for this problem is to have human observers look at the class data and estimate the number of motifs. Alternatively, there are schemes [11] that automatically determine the number of mixtures for a given dataset.

An interesting extension to this work is to additionally characterize the spatial layout of the texture motifs of the object classes. This “texture of textures” approach would help disambiguate objects that consist of the same motifs but are semantically different. For example, the spatial layout of grass and trees in a golf course is different from that in a park due to the elongated fairways.

6. Acknowledgements

This research was supported in part by the following grants and awards: NSF grant #IIS-9817432; ONR/AASERT award #N00014-98-1-0515; ONR grant #N00014-01-0391; NSF Instrumentation grant #EIA-9986057; and NSF Infrastructure grant #EIA-0080134.

References

- [1] S. Newsam, S. Bhagavathy, C. Kenney, B. S. Manjunath, and L. Fonseca, "Object-based Representations of Spatial Images," in *Acta Astronautica*, Vol. 48, No. 5-12, 2001, pp. 567-577.
- [2] B. S. Manjunath and W. Y. Ma, "Browsing Large Satellite and Aerial Photographs," in *Proceedings of the Third IEEE International Conference on Image Processing*, 1996, Vol. II, pp. 765-768.
- [3] B. Zhu, M. Ramsey, and C. Hsinchun, "Creating a Large-Scale Content-Based Airphoto Image Digital Library," *IEEE Transactions on Image Processing*, Vol. 9, (No. 1), Jan. 2000, pp.163-167.
- [4] B. S. Manjunath and W. Y. Ma, "Texture Features for Browsing and Retrieval of Image Data," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 18, No. 8, pp. 837-842, Aug 1996.
- [5] R. Manduchi, "A Cluster Grouping Technique for Texture Segmentation," in *Proceedings of the 15th International Conference on Pattern Recognition*, 2000, pp. 1060-1063.
- [6] C. Carson, M. Thomas, S. Belongie, J. M. Hellerstein, and J. Malik, "Blobworld: a System for Region-based Image Indexing and Retrieval," in *Proceedings of the Third International Conference on Visual Information Systems*, 1999, pp. 509-516.
- [7] L. Jia and L. Kitchen, "Object-based Image Content Characterization for Semantic-level Image Similarity Calculation," in *Pattern Analysis and Applications*, (2001)4:215-216.
- [8] P. Duygulu and F. Yarman-Vural, "Multi-level Segmentation and Object Representaion for Content-based Image Retrieval," in *Proceedings of the SPIE - The International Society for Optical Engineering*, Vol. 4315, 2001, pp. 460-469.
- [9] Y. Chahir and L. Chen, "Efficient Content-based Image Retrieval Based on Color Homogeneous Objects Segmentation and Their Spatial Relationship Characterization," in *Proceedings of the IEEE International Conference on Multimedia Computing and Systems*, Vol. 2, 1999, pp 705-709.
- [10] L. Hill, J. Frew, and Q. Zheng, "Geographic names: The Implementation of a Gazetteer in a Georeferenced Digital Library," in *D-Lib Magazine*, Corporation for National Research Initiatives, Jan 1999.
- [11] X. Q. Li and I. King, "Gaussian Mixture Distance for Information Retrieval," in *Proceedings of the International Conference on Neural Networks*, 1999, pp. 2544-2549.

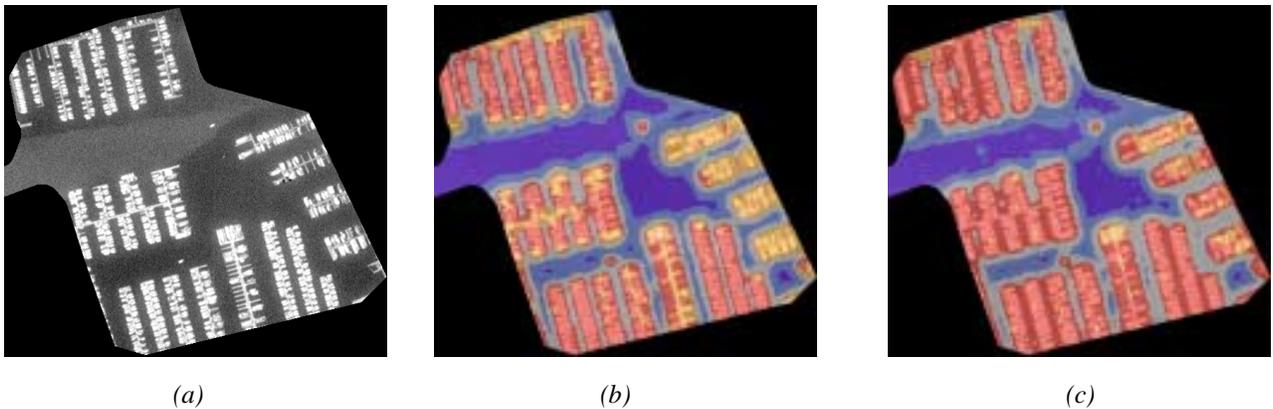


Figure 1. (a) One instance of the harbor class, (b) texture motif assignments using the canonical class model with training set size $M=1$, and (c) texture motif assignments with $M=10$.

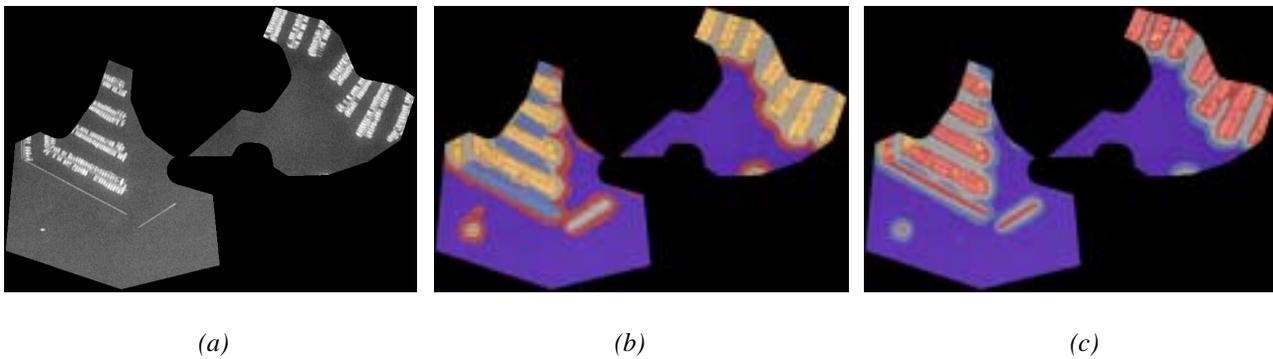


Figure 2. (a) Second instance of the harbor class, (b) texture motif assignments using the canonical class model with $M=1$, and (c) texture motif assignments with $M=10$.