MODELING OBJECT CLASSES IN AERIAL IMAGES USING HIDDEN MARKOV MODELS

Shawn Newsam, Sitaram Bhagavathy, and *B. S. Manjunath*

Department of Electrical and Computer Engineering
University of California, Santa Barbara, CA 93106.
{snewsam, sitaram, manj}@ece.ucsb.edu

Abstract

A canonical model is proposed 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 geographic processes. Furthermore, the spatial arrangement of the motifs is an important discriminating characteristic. In our approach, the states of a Hidden Markov Model (HMM) correspond to the texture motifs and the state transitions correspond to the spatial arrangement of the motifs. A one-dimensional approach reduces the computational complexity. The model is shown to be effective in characterizing objects of interest in spatial datasets in terms of their underlying texture motifs. The potential of the model for identifying the classes of unlabeled objects is demonstrated.

1. INTRODUCTION

Researchers have shown significant interest in using texture descriptors for the automated analysis of aerial images [1, 2, 3]. Homogeneous texture descriptors [4] have been shown to be effective in characterizing a variety of basic land-cover types, such as water, agricultural fields, etc. The work presented in this paper is progress toward using texture descriptors for image analysis at the object level.

Textures have the capacity to describe distinct spatial signatures resulting from many natural and man-made geographic processes that create objects of interest. For example, a distinct texture results from grassy areas which constitute golf courses and parks. A major challenge in using homogeneous texture features is that the objects in spatial datasets usually consist of multiple textures. Grass and trees each result in distinct textures but neither feature by itself characterizes a golf course. Hence, objects must be characterized by sets of texture themes, or *motifs*.

Importantly, the geographic processes that generate objects have a structured spatial arrangement. Therefore, object models should consider the spatial arrangement of the texture motifs. Both golf courses and parks have grass and trees but it is the arrangement of these features that differentiates one from the other. Since analyzing the spatial arrangement of the entire object region is computationally challenging, analysis is often restricted to context or adjacency, such as Markov frameworks for statistical methods.

The major contribution of this work is a model that utilizes multiple textures to characterize image regions. In particular, texture motifs and their spatial arrangement are used to discover and characterize the set of geographic processes that create the objects of interest. Experimental results show that the technique characterizes many objects of interest in spatial datasets, such as airports, harbors, etc.

Statistical methods have been used in the past to represent the spatial arrangement of image features. In [5], 2-D HMMs are used to perform binary classification of image blocks. The block feature vectors and spatial context are used to estimate the parameters of a 2-D HMM. The model is then used to classify unlabeled blocks. Results are presented for classifying aerial images into man-made and natural regions and for classifying document images into text and graphic regions. In [6], 2-D HMMs are used to learn the statistical models of individual images. A statistical distance measure between images, based on the similarity of their models, is used for classification and retrieval tasks.

In Section 2, we motivate the analysis of texture motifs in object modeling. In Section 3, we describe an HMM-based object model, along with some applications. We present some experimental results in Section 4 and the concluding arguments in Section 5.

2. TEXTURE MOTIF ANALYSIS OF OBJECTS

We start with the basic assumption that geographic regions of interest are characterized by collections of texture motifs corresponding to geographic processes. Users of geographic image collections often need to retrieve information on semantically relevant classes of regions, which we term *objects*, e.g. airports, mobile home parks, etc.

There exist several systems that perform region-level [7] and crude object-level [8, 9, 10] image content retrieval using properties of homogeneous region segmentations. However, the problem of modeling a general set of semantic classes is still unsolved.

3. MODELING OBJECT CLASSES

A good model for an object class should (1) capture the texture motifs that characterize it, and (2) identify objects that belong to that class. We model an object class as a combination of horizontal and vertical 1-D HMMs, where a many-to-one mapping may exist between the states and the texture motifs that characterize the class. We call our model, the *canonical class model* (CCM) for object classes.

3.1. Previous Work

In previous work [11], we only consider the statistical characteristics of the texture motifs and not the spatial arrangement. Gaussian Mixture Models (GMMs) are used to characterize the object classes. Homogeneous texture feature vectors are extracted from object instances. The model parameters, namely the distribution means and covariances, are estimated using the Expectation-Maximization algorithm. The models are then used to characterize novel object instances.

3.2. The Canonical Class Model Using HMMs

Homogeneous texture feature vectors [4] are extracted by applying a set of Gabor-wavelet filters (at 5 scales and 6 orientations) to the object images. To reduce computational complexity, we divide the images into 4×4 pixel blocks and observe only the averaged feature vectors from these blocks. The details pertaining to how we obtain and represent objects are provided in Section 4.

The canonical model consists of a horizontal HMM (HHMM) and a vertical HMM (VHMM), separately trained using the observations from the rows and columns, respectively, of the objects. To construct the HHMM and VHMM, we make the following assumptions.

The first assumption is that each row and column of observations (feature vectors) are first order Markov chains, i.e.,

$$
\begin{aligned}
P[q_{i,j}^H = S_n^H | q_{i-1,j}^H &= S_m^H, q_{i-2,j}^H = S_l^H, ...] \\
&= P[q_{i,j}^H = S_n^H | q_{i-1,j}^H = S_m^H] \\
&= a_{m,n}^H, \quad\quad\quad\quad\quad\quad\quad (1)
\end{aligned}
$$

$$
\begin{aligned}
P[q_{i,j}^V = S_n^V | q_{i,j-1}^V &= S_m^V, q_{i,j-2}^V = S_l^V, ...] \\
&= P[q_{i,j}^V = S_n^V | q_{i,j-1}^V = S_m^V] \\
&= a_{m,n}^V, \quad\quad\quad\quad\quad\quad\quad (2)
\end{aligned}
$$

where $q_{i,j}^H$ and $q_{i,j}^V$ are the horizontal and vertical states, respectively, of the the block at the intersection of the ith row and j^{th} column; $\{S_m^H; m = 1, 2, ..., M_H\}$ and $\{S_n^V, n = 1, 2, ..., M_V\}$, are the possible states for the horizontal chains (rows) and vertical chains (columns), respectively; and $a_{m,n}^H$ and $a_{m,n}^V$ are the horizontal and vertical transition probabilities, respectively.

The second assumption is that for every state s, the observations y follow a Gaussian distribution,

$$
b_s(y) = \frac{1}{\sqrt{(2\pi)^d |\Sigma_s|}} e^{-\frac{1}{2}(y-\mu_s)^T \Sigma_s^{-1} (y-\mu_s)} \quad (3)
$$

where d is the dimensionality of the data, Σ_s is the covariance matrix, and μ_s is the mean vector.

3.2.1. Training the Model

Given a training set of object instances from a class, the step-by-step process of constructing the canonical class model is as follows:

1. Divide each object in the training set into 4×4 pixel blocks and obtain the averaged feature vector for each block.

2. Estimate the model parameters $\{a_{m,n}^H, \Sigma_m, \mu_m; \forall S_m^H, S_n^H\}$ of the HHMM using the horizontal observation chains from the rows of all objects. The number of states for the model M_H is manually chosen depending on the visual complexity of the object class. Initialization is random.

3. Choose M_V. Estimate the model parameters of the VHMM using the vertical observation chains from the columns of all objects.

3.2.2. Texture Motif Identification

After training the canonical model for a class, we use the model to identify the texture motifs that characterize this class. Given a test object from a class, and the model for this class, the following novel method is used to identify the texture motifs:

1. Divide the test object into 4×4 pixel blocks and obtain the averaged feature vector for each block.

2. For each horizontal observation chain, determine the state path with maximum a posteriori probability, using the Viterbi algorithm. This step gives us the horizontal state assignments $q_{i,j}^H$ for each block (i,j) in the test object.
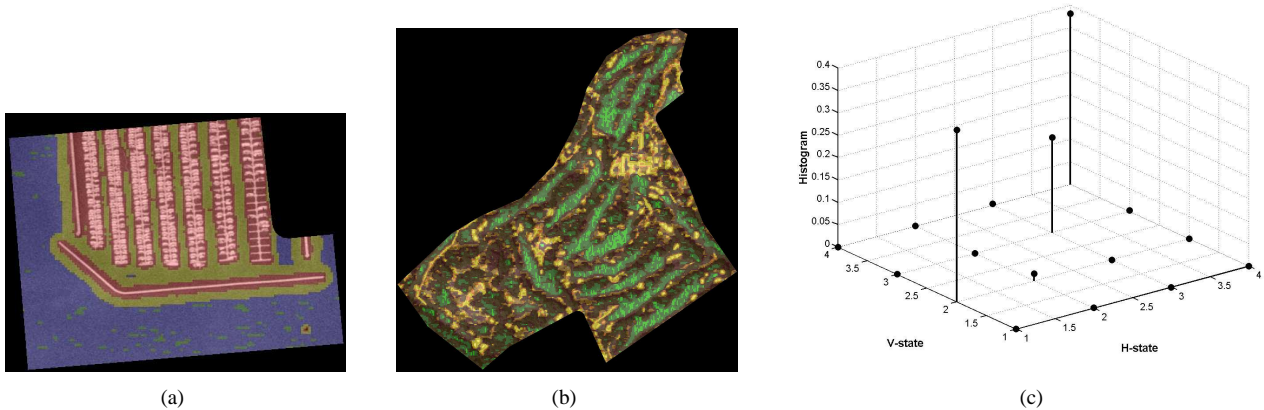
Fig. 1. Instance of (a) the harbor class, and (b) the golf course class, both showing texture motif assignments; and (c) the co-occurrence histogram of the state assignments for the harbor instance. The two tallest spikes correspond to the water and moored boats motifs.

3. Repeat Step 2 for the vertical observation chains, to get the vertical state assignments $q_{i,j}^V$.

4. The state assignment of each block is then given by an ordered pair,

$$q_{i,j} = \{q_{i,j}^H, q_{i,j}^V\}. \tag{4}$$

5. For the given object, construct the co-occurrence histogram of the M_H horizontal states and M_V vertical states,

$$p_{m,n} = P[q_{i,j}^H = S_m^H, q_{i,j}^V = S_n^V]. \tag{5}$$

The $p_{m,n}$ corresponding to spikes in the histogram values (greater than a fixed threshold) identify the texture motifs that occur frequently in that class, and therefore characterize it.

The spikes in the co-occurrence histogram are due to the multiplicative nature of the classifications from the horizontal and vertical HMMs. Note that this procedure also empirically determines the number of texture motifs that are predominant in a class.

3.2.3. Object Classification

Given a test object, the models for N classes, and the knowledge that the test object belongs to one of these classes, the following method is used to classify the test object:

1. Divide the test object into 4×4 pixel blocks and obtain the averaged feature vector for each block.

2. For each ith horizontal observation chain, and using the HHMM for the nth class ($n = 1, ..., N$), determine the log-likelihood $L_{n,i}^H$, that the observation belongs to the model.

3. Repeat Step 2 for the vertical observation chains, using the VHMMs, to get $L_{n,j}^V$.

4. Calculate the classwise average of the log-likelihoods $L_{n,i}^H$ and $L_{n,j}^V$ over the whole object,

$$l_n = \frac{1}{R} \sum_{i=1}^{R} L_{n,i}^H + \frac{1}{C} \sum_{j=1}^{C} L_{n,j}^V \tag{6}$$

where R is the number of horizontal chains and C is the number of vertical chains.

5. The class of the object is that with the minimum absolute value of l_n; $n = 1, ..., N$.

4. EXPERIMENTAL RESULTS

We present here the preliminary results of our work. The dataset contains five classes of objects from the Digital Orthophoto Quarter-Quadrangle (DOQQ) coverage of California: airports, golf courses, harbors, mobile home parks, and vineyards. For each object instance, we extract a rectangular bounding region and manually create a binary mask to define the object boundary.

We train the HHMM and VHMM for each class using 6 object instances from that class. We choose $M_H = M_V = 4$ for simplicity. The test set for each class has 4 objects which are not present in the training set. Figure 1(a) shows a test harbor object in which the identified texture motifs are shaded with different colors. The model captures the significant motifs, in this case the moored boats and the water. Figure 1(b) shows a test golf course object. The significant motifs are now the grassy fairways, trees, and sand-traps and paths. Figure 1(c) shows the co-occurrence histogram of the state assignments for the object in Figure 1(a). We observed that the two tallest spikes correspond to the water

and moored boats motifs, which are the characteristic texture motifs for harbors.

Table 1 shows the results of the object classification method described in Section 3.2.3. The columns stand for the models, and the rows for test classes. The entry in the mth row and n^{th} column denotes the normalized, absolute value of the average l_n using the test set for class m. The minimum value for each row is shown in bold font. Note that the test classes give minimum values for their corresponding class models.

5. CONCLUSION AND FUTURE WORK

We have proposed a canonical class model for object classes in aerial images. Initial experiments show that the model succeeds in capturing the texture motifs that globally characterize object classes. We are investigating whether the model generalizes well to other object classes. Because the model globally analyzes the class, we expect it to de-emphasize irrelevant textures that occur in a few instances, but do not characterize the class in general.

A merit of our approach is that by using horizontal and vertical 1-D HMMs in conjunction, we incorporate spatial context information in the model while avoiding the complexity of a 2-D HMM.

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. We expect that this analysis will be useful to research problems like automatic object segmentation, object-based image retrieval, etc.

Table 1. Object Classification Results: The entries are the average, absolute, normalized log-likelihoods of test class m belonging to class model n (*A*-airports, *B*-golf courses, *C*-harbors, *D*-mobile home parks, and *E*-vineyards).

	MODEL				
	A	B	C	D	E
Class *A*	**0.7502**	0.8250	0.9176	1.5910	0.9872
Class *B*	0.6921	**0.5855**	0.6666	1.0202	0.7349
Class *C*	0.5998	0.5199	**0.3564**	0.7838	0.6188
Class *D*	0.6810	0.5223	0.4733	**0.3594**	0.6155
Class *E*	0.1873	0.2006	0.2240	0.2631	**0.1639**

References

[1] S. Newsam, S. Bhagavathy, C. Kenney, B. S. Manjunath, and L. Fonseca, "Object-based representations of spatial images," *Acta Astronautica*, vol. 48, no. 5-12, pp. 567–577, 2001.

[2] B. S. Manjunath and W. Y. Ma, "Browsing large satellite and aerial photographs," in *Proceedings of the Third International Conference on Image Processing*, 1996, vol. II, pp. 765–768.

[3] B. Zhu, M. Ramsey, and H. Chen, "Creating a large-scale content-based airphoto image digital library," *IEEE Transactions on Image Processing*, vol. 9, no. 1, pp. 163–167, January 2000.

[4] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 837–842, August 1996.

[5] J. Li, A. Najmi, and R. M. Gray, "Image classification by a two-dimensional hidden Markov model," *IEEE Transactions on Signal Processing*, vol. 48, no. 2, pp. 517–533, February 2000.

[6] D. DeMenthon, D. Doermann, and M. V. Stückelberg, "Image distance using hidden Markov models," in *Proceedings of the 15th International Conference on Pattern Recognition*, 2000, vol. 3, pp. 143–146.

[7] 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.

[8] L. Jia and L. Kitchen, "Object-based image content characterization for semantic-level image similarity calculation," *Pattern Analysis and Applications*, vol. 4, no. 2-3, pp. 215–226, 2001.

[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*, 1999, vol. 2, pp. 705–709.

[10] 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, 2001, vol. 4315, pp. 460–469.

[11] S. Bhagavathy, S. Newsam, and B. S. Manjunath, "Modeling object classes in aerial images using texture motifs," in *Proceedings of the International Conference on Pattern Recognition*, Quebec city, August 2002 (to appear).