

OBJECT LOCALIZATION USING TEXTURE MOTIFS AND MARKOV RANDOM FIELDS

S. Newsam, S. 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

This work presents a novel approach to object localization in complex imagery. In particular, the spatial extents of objects characterized by distinct spatial signatures at multiple scales are estimated by using statistical models to control a simple region growing process. Texture motifs are used to model the spatial signatures at the smallest, or pixel, scale. Markov random fields are used to model the spatial signatures at the larger, or motif, scale. These models are used to iteratively expand a bounding box to approximate the spatial extent of an object. The approach is applied to localizing geo-spatial objects in high-resolution panchromatic aerial imagery.

1. INTRODUCTION

The human visual system is highly adept at localizing objects characterized by distinct spatial signatures. For example, even a casual observer can determine the spatial extents of golf courses and housing tracts in panchromatic aerial imagery. While progress has been made in developing effective spatial descriptors, such as the texture features recently standardized by MPEG-7 [1], automatically estimating spatial extents remains a challenge for many classes of objects.

Many applications not only require that the presence of an object be detected but also that its spatial extent be localized. For example, knowing the spatial extents of geo-spatial objects in geographic datasets is necessary for supporting spatial queries. The Alexandria Digital Library (ADL) [2] contains an extensive gazetteer that catalogues the spatial locations of over 5 million instances of over 200 types of geo-spatial objects. However, each instance is represented by only a single point location so that extending the gazetteer to include even bounding boxes has been identified by the ADL development team as essential for supporting a broader range of spatial queries [3]. The proposed technique automates this using readily available aerial imagery.

Texture features based on the outputs of scale- and orientation-selective Gabor filters [4,5] have been shown to effectively characterize a variety of basic land-cover types in high-resolution aerial imagery, such as forests, grasslands, agricultural fields, and urban development [6].

Gaussian mixture models (GMMs) can be used to represent the distribution of texture features for objects that consist of multiple spatial signatures [7]. The components of the GMMs correspond to the textures common to the object classes. These textures are termed *texture motifs* and examples include the rows of moored boats and water in harbors, and the grass and trees in golf courses.

While the texture motifs capture the spatial distribution of the pixel intensities, it is often the spatial distribution of the motifs themselves that is the distinguishing characteristic for an object class. For example, it is the spatial distribution of grass and trees that distinguishes golf courses from parks. Spatial signatures need to be modeled at both the pixel and motif scale.

The main contribution of this paper is a technique that uses the *configuration of the texture motifs to localize objects characterized by spatial signatures at multiple scales*. The spatial distribution of the texture motifs is modeled as a Markov random field (MRF). This model is used to optimally grow a bounding box that represents an approximation of the spatial extent. The proposed method is applied to localizing geo-spatial objects in high-resolution panchromatic aerial imagery.

The rest of the paper is organized as follows. Section 2 describes the salient components of the proposed method, Section 3 presents experimental results, and Section 4 concludes with a discussion.

2. THE PROPOSED METHOD

The proposed method consists of three steps. The first step uses texture motifs to represent the textures common to an object class. The second step models the spatial distribution of the motifs as an MRF. And, the third step uses the MRF to optimally expand a bounding box to estimate the spatial extent of an object instance.

2.1. Texture Motifs

Gaussian mixture models have been shown to effectively characterize the texture motifs, or common textures, of a variety of object classes that are characterized by multiple spatial signatures [7]. The Gabor texture features corresponding to a motif are assumed to be Gaussian distributed in the high-dimensional feature space and the GMM parameters—the prior probabilities and the class

conditioned means and covariances—are estimated using the Expectation-Maximization algorithm applied to a set of training images. The GMM model is used to label the texture motifs of a novel object instance using a maximum a posteriori classifier. Figure 1 displays an example of a golf course and its corresponding motif labeling. More details concerning texture motifs can be found in [7].

2.2. Markov Random Fields

The result of the previous step is an image in which each pixel, or site, is labeled as one of the M motifs. The spatial arrangement of the motifs is often the distinguishing characteristic for an object class. If the simplifying assumptions are made that only neighboring sites have direct interactions with each other—i.e., Markovianity—and that these interactions do not vary with the locations of the sites—i.e., homogeneity—then a tractable way to represent such a spatial distribution is to model the motif labels as a homogeneous MRF. The Markov-Gibbs equivalence properties can then be used to compute the joint probability of a region by evaluating the energy function at all sites.

Figure 2 shows the first order neighborhood defined for a site i . This neighborhood contains one single-site clique and four pair-site cliques that are taken to be equivalent since the MRF is also assumed to be isotropic. Since each site is labeled as one of M motifs, a multi-level logistic model (MLL) is adopted. If site i has label f_i then the potential functions for the single-site and pair-wise cliques are

$$V_1(f_i) = \alpha_j \quad \text{if } f_i = j \in \{1, \dots, M\} \quad (1)$$

and

$$V_2(f_i, f_{i'}) = \begin{cases} \beta & \text{if } f_i = f_{i'} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where $i' \in N_i$, the 4-neighborhood of site i . The joint probability of f , a labeling of a set of sites S , is then

$$P(f) = \frac{1}{Z} e^{-U(f)} \quad (3)$$

where the energy function $U(f)$ is

$$U(f) = \sum_{i \in S} V_1(f_i) + \sum_{i \in S} \sum_{i' \in N_i} V_2(f_i, f_{i'}). \quad (4)$$

The partition function Z is

$$Z = \sum_{f \in \mathbb{F}} e^{-U(f)} \quad (5)$$

where \mathbb{F} represents all possible labelings of S .

The MLL parameters α_j , $j=1, \dots, M$, and β , can be estimated from a training set using a Markov Chain Monte Carlo (MCMC) approach [8].

2.3. Object Localization

The result of the previous step is an MRF model for an object class. The model is completely specified by the MLL parameters α_j , $j=1, \dots, M$, and β . This model is used to estimate the spatial extent of a novel object instance with the assumption that an interior point is known.

The guiding premise is to expand a bounding box in the direction that is most similar to the object class *with respect to the MRF model*. The likelihood function in (3) cannot be used to calculate this similarity since the partition function cannot be computed except in simple cases. An alternate approach is needed. The proposed technique determines how similar a direction is to the object class by estimating the MLL parameters for the adjacent region for that direction and then comparing these estimates with the class model parameters.

Let the current bounding box consist of sites S and let the region adjacent to side k be ΔS^k where $k \in K = \{1, 2, 3, 4\}$, as shown in Figure 3. (The labeling K is arbitrary.) The parameters for region ΔS^k , α_j^k , $j=1, \dots, M$, and β^k , are estimated, again using an MCMC approach.

Let D^k be the L1 distance between the parameters for side k and the model parameters:

$$D^k = |\beta^k - \beta| + \sum_{j=1}^M |\alpha_j^k - \alpha_j|. \quad (6)$$

The bounding box is then extended by the row or column of sites adjacent to the side corresponding to the minimum D^k . That is, the sites ΔS_{edge}^{k*} are added where

$$k^* = \arg \min_{k \in K} D^k. \quad (7)$$

Any point inside the object can be used to initialize the algorithm. The ADL Gazetteer provides such a point.

The procedure for iteratively expanding the bounding box is summarized as follows:

1. Initialize the bounding box as a point.
2. Estimate the MLL parameters for regions ΔS^k where $k=1, 2, 3, 4$.
3. Expand the box by adding the sites $\Delta S_{edge}^{k^*}$ corresponding to the region whose parameters minimize D^k .
4. If the stopping criterion is met then finish. Otherwise, go to step 2.

A logical choice for the stopping criterion is when none of the adjacent regions are sufficiently similar to the object class; i.e., when D^k for all sides is greater than a predetermined threshold.

3. EXPERIMENTS AND RESULTS

The proposed technique is applied to two classes of objects, golf courses and mobile home parks. Modeling the spatial arrangement of the texture motifs for these classes using MRFs is appropriate since the arrangements are distinguishing characteristics. The dataset for each class consists of five aerial images in which the object boundaries have been manually estimated and represented as image masks.

The proposed technique is evaluated using a leave-one-out approach in which four images are used for training and the remaining image for testing. The GMM for the texture motifs is trained using the masked regions in the four training images as described in Section 2.1. This model is used to assign motif labels to all five images. An example of this labeling is shown in Figure 1(b). The parameters for the MRF model are estimated using the masked regions in the four motif-labeled training images as described in Section 2.2. This model is used to iteratively expand a bounding box in the motif-labeled test image as described in Section 2.3. The bounding box is initiated as a random point inside the object as this simulates the information available in the ADL Gazetteer.

Figures 4(a) and 4(b) show the growth of bounding boxes for a mobile home park and a golf course, respectively. The image regions outside the object masks are dimmed to indicate the manually specified spatial footprints that serve as the ground truth. The bounding boxes are shown at 75-iteration intervals from random starting points until the stopping criterion is met. In this case this criterion is when D^k is larger than an empirically chosen threshold for $k=1,2,3,4$; i.e. when the distance between the estimated MLL parameters and the object parameters exceeds the threshold in all four directions. These examples demonstrate that the proposed technique effectively estimates the spatial extent of the objects even when the starting points are near the boundaries.

Receiver operating characteristic (ROC) curves can be used to evaluate the expansion of the bounding box. Since the object boundary is known for the test image, the true-positive ratio can be computed as the percentage of the object that is contained in the bounding box and the false-positive ratio as the percentage of the bounding box that is not part of the object. Figures 5(a) and (b) show the ROC curves averaged over the five mobile home park images and five golf course images, respectively. The proposed technique is shown to maintain a high true-positive to false-positive ratio and outperforms a naïve approach that

begins at the same random location but does not consider the MRF model and uniformly expands the sides of the bounding box.

4. CONCLUSION

This work describes a technique for estimating the spatial extents of objects characterized by spatial signatures at multiple scales by optimally expanding a bounding box with respect to a model that uses texture motifs to represent the spatial configuration of the pixel intensities and MRFs to represent the spatial configuration of the motifs. Examples are provided for localizing golf courses and mobile home parks in panchromatic aerial imagery.

A future research direction is improving the stopping criterion to make the region growing procedure more robust. Another direction is to model the MLL parameters for an object class as a GMM rather than a single set of values to account for the lack of homogeneity. This would result in a more stable region growing procedure.

Acknowledgements

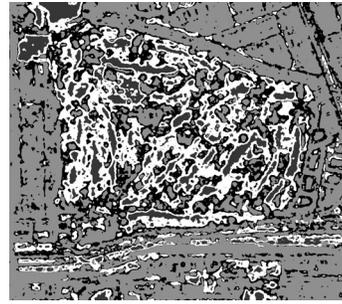
This work is supported by the following grants: NASA California Space Grant, USDOT Research and Special Programs Administration Contract DTRS-00-T-0002, ONR# N00014-01-1-0391, NSF Instrumentation #EIA-9986057, NSF Infrastructure #EIA-0080134, and NSF #IIS-9817432.

5. REFERENCES

- [1] B.S. Manjunath, P. Salembier, and T. Sikora, Eds., *Introduction to MPEG7: Multimedia Content Description Interface*, John Wiley & Sons, first edition, 2002.
- [2] Alexandria Digital Library Project homepage: <http://www.alexandria.ucsb.edu>.
- [3] L.L. Hill, J. Frew, and Q. Zheng, "Geographic names: The implementation of a gazetteer in a georeferenced digital library," *D-Lib*, January 1999.
- [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.837-42, August 1996.
- [5] P. Wu, B.S. Manjunath, S. Newsam, and H.D. Shin, "A texture descriptor for browsing and similarity retrieval," *Journal of Signal Processing: Image Communication*, Volume 16, Issue 1-2, page 33-43, September 2000.
- [6] B.S. Manjunath and W.Y. Ma, "Browsing large satellite and aerial photographs," *IEEE International Conference on Image Processing*, Lausanne, Switzerland, Sep 1996.
- [7] S. Bhagavathy, S. Newsam, and B.S. Manjunath, "Modeling object classes in aerial images using texture motifs," *International Conference on Pattern Recognition*, Quebec, Canada, August 2002.
- [8] L. Wang, J. Liu and S.Z. Li, "MRF Parameter Estimation by MCMC Method," *Pattern Recognition*, Vol. 33, No. 11, pp. 1919-1925, 2000.



(a)



(b)

Figure 1. (a) An example of a golf course in an aerial image. (b) The texture motif labeling for this golf course.

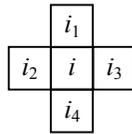


Figure 2. First order neighborhood for site i .

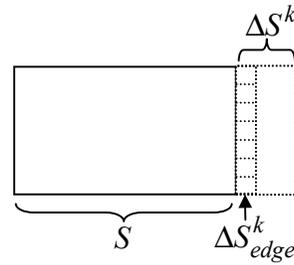


Figure 3. The bounding box S is expanded by adding the row of sites ΔS^k_{edge} to the side k for which the adjacent region ΔS^k best matches the MRF model for the object.

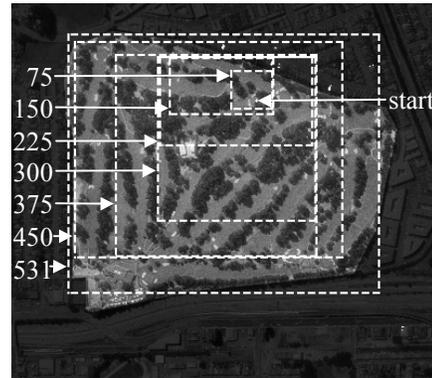
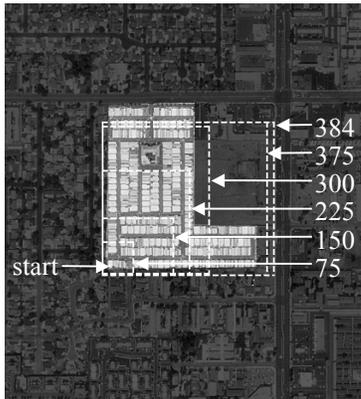
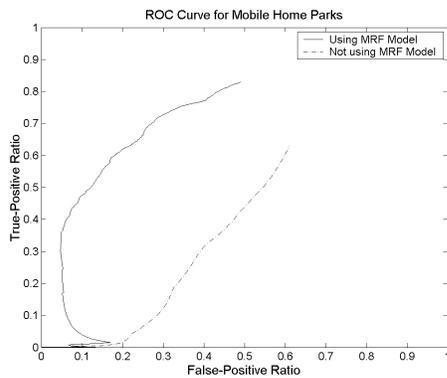
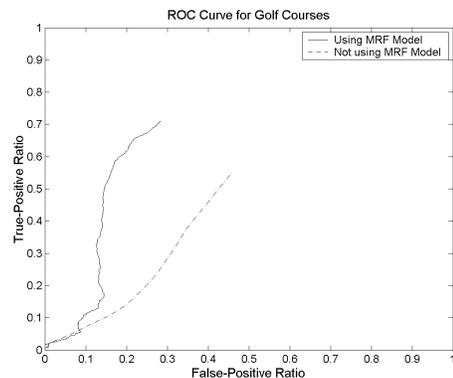


Figure 4. Bounding box every 75 iterations until stopping for (a) mobile home park, and (b) golf course.



(a)



(b)

Figure 5. ROC curves for (a) mobile home parks, and (b) golf courses.