

Delineating Buildings by Grouping Lines with MRFs ¹

Santhana Krishnamachari

Rama Chellappa

Department of Electrical Engineering

Center for Automation Research and

Institute for Advanced Computer Studies

University of Maryland

College Park, MD 20742

Abstract: Traditionally, Markov Random Field (MRF) models have been used in low level image analysis. This correspondence presents an MRF based scheme to perform object delineation. The proposed edge-based approach involves extracting straight lines from the edge map of an image. Then an MRF model is used to group these lines to delineate buildings in aerial images.

1 Introduction

Markov random field models are used in image analysis to characterize prior beliefs about various image features. MRF models exploit the local statistical dependence of image features and enable global optimizations to be performed in terms of iterative local computations. These models have been extensively used in low level image processing applications, such as texture synthesis and segmentation, image restoration, image compression, to name a few. In all the aforementioned applications MRF models are used on the pixels of the image. The efficacy of these models has been demonstrated in low level processing, but their applicability to high level problems such as object recognition is yet to be fully explored. In [2], [10] MRF models have been used for some high level vision problems. In this correspondence,, we explore the applicability of MRFs to object delineation. We have chosen the problem of delineating buildings in aerial images, which in its own right is an important problem, because buildings are the perceptually dominant structures in aerial images; hence building detection has been an important task in aerial image understanding. Region and edge based methods presented in [5], [9], [6], [4] and [11] use a large set of rules and are computationally very expensive. Also the rules are implicitly associated with

¹Supported in part by grant MIP-91-00655 from the National Science Foundation

several parameters. The goal of this correspondence is to present a fast and systematic method to group lines based on an MRF model to delineate buildings of specific geometric shapes. Our approach is as follows: (1) Straight lines are extracted from images by using an edge detector followed by a line extractor [12]. (2) An MRF model is used on these extracted lines with a suitable neighborhood. The probabilistic model is chosen to support the properties of the shapes of buildings (rectangular, L-shaped). (3) The energy function associated with the MRF is minimized, resulting in the grouping of lines.

2 An MRF on Lines

We model the probabilistic dependence of lines in an image by an MRF, because in an attempt to delineate buildings, it is not necessary for each line in the image to interact with every other line. Only lines that are spatially close to each other and with specific angular orientations need to interact.

Definition: Let l_i and l_j be two lines in the image. Let $\Phi(l_i, l_j)$ be a measure of the angle between l_i and l_j , and $L(l_i, l_j)$ be a measure of the distance between them. Exact definitions of Φ and L are given in Section 3. The definitions of Φ and L are such that

$$\begin{aligned} 1. \quad & L(l_i, l_i) = 0; \quad 2. \quad L(l_i, l_j) \geq 0; \\ 3. \quad & L(l_i, l_j) = L(l_j, l_i); \quad 4. \quad \Phi(l_i, l_j) = \Phi(l_j, l_i). \end{aligned} \tag{1}$$

Let N_{l_i} be the neighbor set of l_i . Then $l_j \in N_{l_i}$ iff

$$L(l_i, l_j) \leq \tau_1; \quad |\Phi(l_i, l_j)| \leq \tau_2, \tag{2}$$

where the thresholds τ_1 and τ_2 have to be chosen. From the definition of L and Φ it is easy to note that if $l_j \in N_{l_i}$, then $l_i \in N_{l_j}$. In essence, we have built a symmetric graph on the lines.

The probabilistic model on the lines is chosen as follows. Let $\Lambda = (b_1, b_2, \dots, b_k)$ be the set of unknown buildings in the image and (l_1, l_2, \dots, l_n) be the set of lines extracted from the edge map of the image. We use the following notation for the conditional probability of (l_1, l_2, \dots, l_n) given Λ ,

$$\begin{aligned} P(l_1, \dots, l_n / b_1, \dots, b_k) &= P_\Lambda(l_1, \dots, l_n) \\ P_\Lambda(l_1, \dots, l_n) &= \frac{1}{Z} e^{-\beta E(l_1, \dots, l_n)} \end{aligned}$$

$$E(l_1, \dots, l_n) = \alpha_1 \sum_i U_1(l_i) + \alpha_2 \sum_i \sum_{j \in N_{l_i}} U_2(l_i, l_j) \quad (3)$$

where U_1 and U_2 are the potentials associated with the first and second order cliques and Z is a normalizing constant. A discussion of MRFs and the proof of equivalence between MRFs and Gibbs distributions can be found in [3], [1].

Maximizing the above probability results in the *maximum likelihood* estimate of the buildings in the image. But the quantities b_1, \dots, b_k do not directly appear in the energy function. So instead of maximizing with respect to b_1, \dots, b_k , we attach a parameter p_i to every l_i , where p_i is a measure of l_i being selected in the final grouping (if $p_i=1.0$, then l_i is selected). The energy function is minimized while each p_i converges to either 0 or 1.

3 The Energy Function

In this section we elaborate on the details of the energy function and the optimization technique. As mentioned in the previous section, the measure p_i on each line is “pushed” to either 0 or 1, to minimize the energy function. If the total number of lines is n , then $\{p_i, i = 1, \dots, n\}$ has 2^n possible values. Thus the global minimum of the energy function can be obtained by exhaustively searching through the exponential number of combinations. But, as we show in our experiments, it suffices to use a gradient descent minimization technique.

We use the following notation:

$l_i = (a_i, b_i)$ where a_i and b_i are the starting and ending points of the line l_i ;

d_i is the length of l_i ;

θ_{ij} be the angle between the l_i and l_j ;

$\text{dist}(a_i, b_i)$ be the distance between the points a_i and b_i ;

$D_{ij} = \min(\text{dist}(a_i, a_j), \text{dist}(a_i, b_j), \text{dist}(b_i, a_j), \text{dist}(b_i, b_j))$.

The L and Φ introduced in the previous section are selected as follows:

$$\begin{aligned} L(l_i, l_j) &= \min(D_{ij}, \min_{k \in N_{l_i}, N_{l_j}} \frac{(D_{ik} + D_{kj})}{p_k}), \\ \Phi(l_i, l_j) &= \sin(2\theta_{ij}). \end{aligned} \quad (4)$$

From the definition of Φ , it is easy to observe that $\Phi(l_i, l_j) = 0$ if the lines are either parallel or perpendicular. The selection of L is such that the lines corresponding to two parallel sides of a building, l_i and l_j , may not be neighbors of each other initially, but if there is a line l_k joining

these two lines, and if p_k is close to 1, then eventually the two parallel lines will become neighbors and get grouped together.

The energy function E consists of the following four terms:

$$\begin{aligned}
E_1 &= -\alpha \sum_{i=1}^n \frac{(1-p_i)}{d_i} \\
E_2 &= -\beta \sum_{i=1}^n \sum_{j \in N_{l_i}} p_i p_j |\cos(2\theta_{ij})| \\
E_3 &= \omega \sum_{i=1}^n \sum_{j \in N_{l_i}} p_i p_j \min(D_{ij}, \min_{k \in N_{l_i}, N_{l_j}} \frac{(D_{ik} + D_{kj})}{p_k}) \\
E_4 &= -\gamma \sum_{i=1}^n p_i \ln p_i + (1-p_i) \ln(1-p_i)
\end{aligned} \tag{5}$$

E_1 favors long lines, avoiding the grouping of very small lines parallel or perpendicular to the other selected lines. E_2 favors lines which are parallel or perpendicular to each other. E_3 supports grouping of lines that are spatially close to each other. E_4 is the entropy of the probabilities p_i and hence pushes the p_i s to either 0 or 1, thus ensuring convergence to one of the vertices of the n -dimensional binary hypercube in which the relaxation is performed.

The p_i s are initialized to 0.5 and then a gradient descent algorithm is used to minimize the energy function.

$$\{p_i\}_{t+1} = \{p_i\}_t - \epsilon \nabla E \tag{6}$$

3.1 Discussion

The following aspects of the energy function and grouping process need to be observed:

(1) Observe that the neighborhood defined by Φ and L is dynamic (due to the definition of $L(l_i, l_j)$), i.e., as the p_i s change, the neighborhood also changes. If two lines are neighbors of each other at the beginning of the gradient descent optimization (with the initial values of p_i s), then they are called *direct neighbors*. If two lines are not neighbors initially, but become neighbors as iterations proceed they are termed as *indirect neighbors*.

(2) We have chosen the dynamic neighborhood on purpose. It could have been avoided by choosing $L(l_i, l_j) = D_{ij}$, which would have resulted in a fixed neighborhood. But we have chosen this dynamic neighborhood to allow the interaction to spread to *indirect neighbors*, at the same

time obviating the use of higher order cliques in the MRF which would result in more parameters and rules, reducing which is the focus of this approach.

(3) We also want to emphasize that all computations are local and fast. If the number of lines is n , there is an initial overhead of $O(n^2)$, to check all pairs of lines to see if they are *direct neighbors*. However, during each iteration, as the neighborhood changes it is *not* necessary to make $O(n^2)$ search to decide neighbors, explained as follows. At the end of an iteration, if a line l_i has a new neighbor l_j which was not in its neighborhood during the previous iteration, then l_j must be an *indirect neighbor* of l_i , i.e., l_j must have become a neighbor of l_i through some already existing neighbor of l_i . We maintain a list of neighbors for each line. At the end of each iteration to decide the neighbors of a line l_i , we search through the neighbor list of all lines that are present in the neighbor list of l_i . Typically no line has more than 6–8 neighbors and most lines have less than 3–4 neighbors, hence the computation during each iteration is only $O(n)$. Also during each iteration the gradient needs to be computed, again, here all computations are local.

(4) Even though the energy function has four free parameters, in [8] we have shown interdependencies between these parameters, hence it is not necessary to choose all these parameters independently.

4 Delineating Buildings

In delineating buildings one aspect that has to be taken into consideration is the presence of shadows. Shadows give rise to edges very close to the ones corresponding to the buildings. During grouping, it is necessary to identify and isolate the lines corresponding to shadows. The shadow edge lines are parallel and very close to that of the buildings and hence tend to get grouped with the building edge lines. To prevent shadow edge lines from being grouped with building edge lines, we make use of the assumption that buildings are brighter than the surroundings, together with the fact that shadows are “dark”. For each line, image gradient in two perpendicular directions to the line are calculated and the direction that has the maximum gradient is used to label the gradient direction as pointing either “inwards” or “outwards”. During grouping, we require that lines can be grouped together only if their gradient directions are toward each other as shown in Figure 1.

4.1 Partial Shape Completion

In real images, lines corresponding to all sides of a building are not always obtained. This is often due to poor contrast. Thus the grouping may result in some partial boundary shapes of the buildings. We use deformable contours (snakes) [7] for partial shape completion. Deformable contours are active contours of deformable shape, the final shape being determined by minimization of an energy function. In any approach using deformable contours, the initialization is important, which is two-folds, the initial shape and its spatial position. The spatial position of the initialization is very crucial, while the initial shape itself is not. In our approach, initialization is not a problem, since we have already obtained the partial shapes. The deformable contours are initialized at the locations of the partial shapes and then allowed to deform according to the following energy function:

Let (x_s, y_s) , $s=1, 2, \dots, N$, be the initial position of the snake. The energy function associated with the deformable contour is given by:

$$\begin{aligned}
 E &= E_{snake} + E_{image} + E_{shape}; & E_{snake} &= \alpha_1 \sum_{s=1}^N (x'_s{}^2 + y'_s{}^2); \\
 E_{image} &= -\beta_1 \sum_{s=1}^N |\nabla f(x_s, y_s)|; & E_{shape} &= \gamma_1 \sum_{s=1}^N \rho_s^2 (1 - C_s) + \omega_1 C_s
 \end{aligned} \tag{7}$$

where $f(x_s, y_s)$ is the image intensity at (x_s, y_s) , ρ_s is the curvature at the site s and C_s is a corner process defined between the snake sites (x_s, y_s) and (x_{s+1}, y_{s+1}) ,

$$\rho_s = \frac{x_s y'_s{}'' - y_s x'_s{}''}{(y'_s{}{}^2 + x'_s{}{}^2)^{\frac{3}{2}}}; \quad C_s = \{0, 1\}.$$

4.2 Results

Figure 2 shows the LAX image which has many buildings of different sizes. Figure 3 shows the lines with gradient directions. Gradient directions are represented by short projections at the centers of the lines. The result of the grouping is shown in Figure 4. As seen in Figure 4 many of the groupings display only partial shapes. The deformable contours approach is used to complete the partial shapes. Figure 5 shows the final result after partial shape completion. Figure 6 is a section of the aerial image made available in connection with the ARPA RADIUS project. Figure 7 shows the lines with their gradients, Figure 8 shows the result of grouping and Figure 9 shows the partial shape completion result.

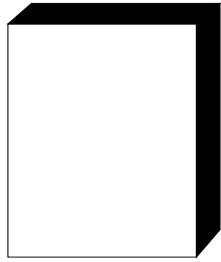
5 Summary

MRF models have been successfully used for many low level image processing problems, but not many attempts have been made to use MRFs for high level problems. Here, we have presented an MRF model for a high level delineation problem. Using MRF models for high level problems enables the specification of prior beliefs in a probabilistic framework. Our experiments show that the grouping based on MRFs provides good results on real aerial images.

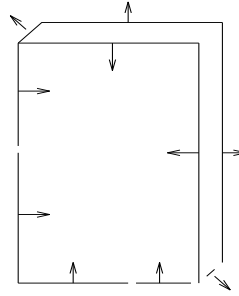
References

- [1] J. Besag, "Spatial Interaction and the Statistical Analysis of Lattice Systems," *Journal of the Royal Statistical Society*, Vol. 36, pp. 192–326, 1974.
- [2] P. R. Cooper, "Parallel Object Recognition from Structure," Tech. Rep. 301, Dept. of Computer Science, University of Rochester, 1989.
- [3] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distribution and the Bayesian Restoration of Images," *IEEE Trans. Patt. Anal. Mach. Intell.*, Vol. 6, pp. 721–741, 1984.
- [4] M. Herman and T. Kanade, "Incremental Reconstruction of 3D Scenes from Multiple, Complex Images," *Artificial Intelligence*, Vol. 30, pp. 289–341, 1986.
- [5] A. Huertas and R. Nevatia, "Detecting Buildings in Aerial Images," *Computer Vision, Graphics and Image Processing*, Vol. 41, pp. 131–152, 1988.
- [6] R. B. Irvin and D. M. McKeown, "Method for Exploiting the Relationship Between Buildings and Their Shadows in Aerial Imagery," *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 19, pp. 1564–1575, 1989.
- [7] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes : Active Contour Models," *International Journal of Computer Vision*, Vol. 4, pp. 321–331, 1988.
- [8] S. Krishnamachari and R. Chellappa, "Delineating Buildings by Grouping Lines," Tech. Rep. CAR-TR 682, University of Maryland, 1993.
- [9] T. Matsuyama and V. S. S. Hwang, *SIGMA: A Knowledge Based Aerial Image Understanding System*, New York: Plenum Press, 1990.

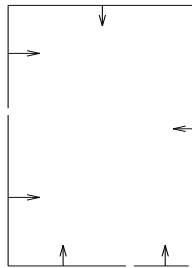
- [10] J. W. Modestino and J. Zhang, "A Markov Random Field Model-Based Approach to Image Interpretation," *IEEE Trans. Patt. Anal. Mach. Intell.*, Vol. 14, pp. 606–615, 1992.
- [11] V. Venkateswar and R. Chellappa, "A Hierarchical Approach to Detection of Buildings in Aerial Images," Tech. Rep. CAR-TR 567, University of Maryland, 1991.
- [12] Y. T. Zhou, V. Venkateswar, and R. Chellappa, "Edge Detection and Linear Feature Extraction Using a 2-D Random Field Model," *IEEE Trans. Patt. Anal. Mach. Intell.*, Vol. 11, pp. 84–94, 1989.



Building with shadows



Gradient directions for lines



Grouped Lines

Figure 1: Effect of shadows



Figure 2: LAX Image

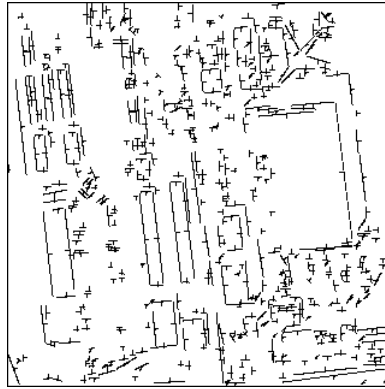


Figure 3: Lines with gradients

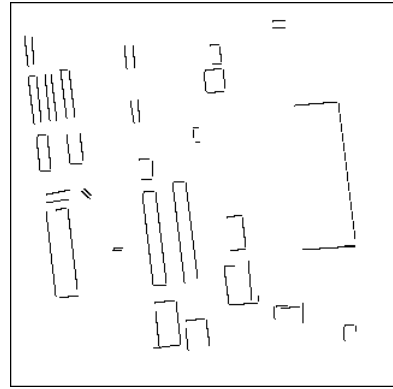


Figure 4: Result of grouping

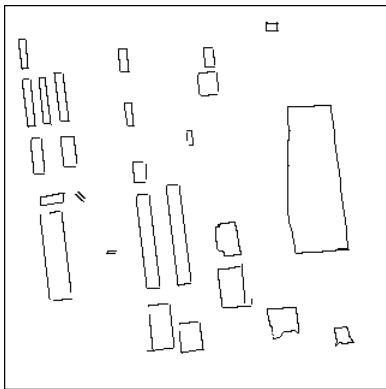


Figure 5: Final result



Figure 6: Ft. Hood Image

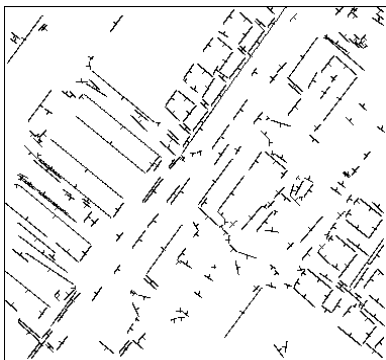


Figure 7: Lines with gradients

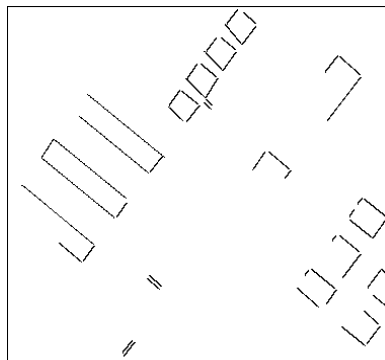


Figure 8: Result of grouping

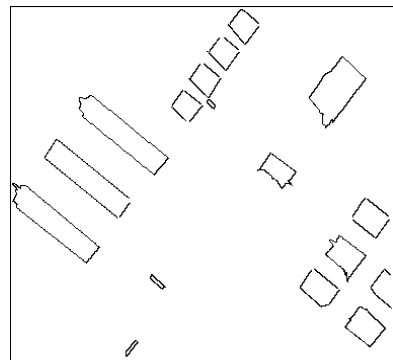


Figure 9: Final result