Grouping with Bias *

Stella X. Yu
Robotics Institute
Carnegie Mellon University
Center for the Neural Basis of Cognition
Pittsburgh, PA 15213-3890
stella.yu@cs.cmu.edu

Jianbo Shi
Robotics Institute
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA 15213-3890
jshi@cs.cmu.edu

Abstract

We present a graph partitioning method to integrate prior knowledge in data grouping. We consider priors represented by three types of constraints: unitary constraints on labelling of groups, partial *a priori* grouping information, external influence on binary constraints. They are modelled as biases in the grouping process. We incorporate these biases into graph partitioning criteria. Computationally this formulation leads to a constrained eigenproblem. We demonstrate the effectiveness of this algorithm on image segmentation with priors and object detection with spatial attention.

1 Introduction

The grouping problem emerges from several practical applications including image segmentation, text analysis and data mining. The goal is to find the overall structural organization that can be used to summarize the data. The view of grouping as a process of extracting global impression from local structures is characterized by a number of computational approaches that evaluate the goodness of grouping based on local cue interactions. These approaches, such as Markov random fields (MRF) [7], variational formulations [2, 13], and graph partitioning approaches [1, 6, 15, 14, 17, 18], often formulate grouping as a global optimization problem based on local cost functions.

One of the major merits of MRF and variational formulations is that they can be naturally cast in a Bayesian framework. In this framework, the task is to find the maximum *a posteriori* estimate of some underlying quantity. The posterior probability incorporates all local clique potentials evaluated on nearby data points, which can encode a variety of configuration constraints and probability distributions [20]. However, such models are usually solved by Markov Chain Monte Carlo methods, which often find local optima with slow convergence. The lack of efficient computational solutions severely limits the application of these approaches.

The same goal of grouping is achieved in the graph partitioning formulation, where a global decision is made based on local pairwise relationships. These algorithms often have efficient computational solutions and it has been demonstrated that they work successfully on the segmentation of complex natural images [11]. Recently a probabilistic interpretation [12] has provided an insight into how and why these algorithms work.

*CMU-RI-TR-01-22, July 2001

There have also been direct attempts in solving MRF by graph partitioning algorithms [8, 4, 3, 16, 9]. This is typically done through the introduction of instantiating multiple valuation of the data or label nodes. With an appropriate choice of graph edge weights, a subclass of MRF problems can be translated into a graph partitioning framework. However, negative constraints in MRF can not be easily handled in most graph algorithms, and have thus been mostly avoided. We note that this barrier in computation is not crucial. Recent work on repulsion and asymmetrical relationships [19] has expanded graph partitioning on nonnegative symmetrical matrices to arbitrary Hermitian matrices, where both positive and negative correlations, undirected and directed relationships can be dealt with in one framework. This advance brings one step closer to the direct application of spectral graph partitioning to MRF formulations.

However, a gap does exist between MRF and graph partitioning approaches. The former seek a meaningful interpretation of the data with a generative model, whereas the latter find partitioning with a discriminative model. What lacks in graph algorithms is the capability to incorporate prior knowledge on data to guide grouping.

In this paper, we model such prior as biases in the grouping process formulated in a graph framework. We consider three types of biases: 1) unitary constraints on labelling of groups; 2) partial grouping information available *a priori*; 3) external influence on binary constraints. We incorporate such constraints into a graph partitioning criterion, which result in a constrained eigenproblem. Though we consider normalized cuts criteria in particular [18], the same derivation can be applied to other criteria as well. We demonstrate the use of the model on spatial attention and image segmentation.

The rest of the paper is organized as follows. Section 2 expands our model in detail. Section 3 shows our experimental results. Section 4 concludes the paper.

2 Model

The key principles of grouping can often be illustrated in the context of image segmentation. In graph methods for image segmentation, all pairwise relationships between two pixels can be captured in two graph representations $\mathsf{G} = \{\mathsf{G_A}, \mathsf{G_R}\}$: $\mathsf{G_A} = (\mathsf{V}, \mathsf{E_A}, A)$, $\mathsf{G_R} = (\mathsf{V}, \mathsf{E_R}, R)$, encoding pairwise attraction and repulsion relationships respectively. Each pixel is a vertex in V. For the edge between pixel i and j, weight A_{ij} describes the likelihood of two pixels belonging to one group, whereas R_{ij} describes the likelihood of them belonging to different groups. A weight of 0 in both cases means neutral, i.e., no information biases the two pixels toward attraction or repulsion relationships.

After an image is transcribed into a graph, image segmentation becomes a vertex partitioning problem. A good segmentation is some optimal partitioning scheme according to an energy function, evaluating how tightly each group is internally connected (associations) and/or how loosely between-group connections (cuts) are. Such segmentation algorithms are solely based on pairwise relationships contained in the original data. Moreover, in the standard image segmentation, we are not interested in the particular labelling of the groups, i.e. we have no discrimination between groups. This is not the case for figure-ground segregation [14], where the information in image may suggest unequal likelihood for a pixel to belong in figure or ground [1].

These issues can be addressed by the introduction of bias into grouping. We consider three types of biases in this paper. They all reflect some aspect of our prior knowledge in grouping data points and can come from sources unavailable in the given data. The first type characterizes the preference of vertices belonging to a particular group. They are essentially unitary constraints, which for example can arise from sensor models in MRF [7]. The second type imposes grouping assignment of some vertices *a prior*, which for example can arise from human computer interac-

tion [10]. The third type regulates the binary constraints obtained from given data, which for example arises from top-down gating control in spatial attention. Among these three types of biases, only the last one can be modelled by modulating the information obtained from data without any change on the grouping engine. The first two impose new constraints to the grouping process and are the focus of our discussion below.

2.1 Representation

A vertex bipartitioning $(\mathsf{V}_1, \mathsf{V}_2)$ on undirected graph G has $\mathsf{V} = \mathsf{V}_1 \cup \mathsf{V}_2$ and $\mathsf{V}_1 \cap \mathsf{V}_2 = \varnothing$. We assume that both A and R are *nonnegative* symmetric weight matrices. We denote by $N = |\mathsf{V}|$ the number of vertices in the graph. Let B_a (B_r) denote the $N \times 2$ unitary bias weight matrix of attraction (repulsion) nature, the two columns of which describes the likeness (dislikeness) of vertices belong to V_1 and V_2 respectively.

First, we introduce dummy nodes to turn unitary constraints into binary constraints. The two dummy nodes represent anchors of group V_1 and V_2 and are labelled as vertex $N+1$ and $N+2$ respectively. We expand the vertex set to include dummy nodes and expand the weight matrix accordingly. With slight abuse of notation, we define:

$$n := N+2, \quad \mathsf{V} := \{1:n\}, \quad A := \begin{bmatrix} (1-\gamma)A & \gamma B_a \\ \gamma B_a^T & \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \end{bmatrix}, \; R := \begin{bmatrix} (1-\gamma)R & \gamma B_r \\ \gamma B_r^T & \gamma \begin{bmatrix} 0 & r \\ r & 0 \end{bmatrix} \end{bmatrix},$$

where r is a parameter we can choose to enforce the separation of the two dummy nodes and $0 \leq \gamma \leq 1$ is a parameter determining the relative importance of bias and original data: no bias when $\gamma = 0$, full bias when $\gamma = 1$.

Second, we handle two kinds of explicit vertex pre-assignment: some vertices are known *a priori* to belong together, and still others are known to belong in V_l, $l = 1, 2$. The rest vertices are so called *free* nodes. Let P_l, $l = 1, 2$ denote the set of nodes that have been assigned to V_l. By definition, $N+1 \in P_1$ and $N+2 \in P_2$. Let P_l, $l = 3, \cdots, n_p$ denote a set of nodes we only know that they belong together, but we are not sure about their relationships to V_l, $l = 1, 2$. Let $p_l = |P_l|$ be the number of nodes in P_l, $l = 1, \cdots, n_p$. An example is illustrated in Fig. 1. With this representation of unitary constraints and pre-set grouping, we turn the problem into an ordinary grouping problem with additional linear constraints.

2.2 Criteria

Given weight matrix W and two vertex sets P and Q, let $\mathcal{C}_W(P,Q) = \sum_{j \in P, k \in Q} W(j,k)$ denote the total W connections from P to Q, In particular, $\mathcal{C}_W(\mathsf{V}_1, \mathsf{V}_2)$ is the total weights cut by the bipartitioning, whereas $\mathcal{C}_W(\mathsf{V}_l, \mathsf{V}_l)$ is the total connections among vertices in V_l, $l = 1, 2$. Let $\mathcal{D}_W(P) = \mathcal{C}_W(P, \mathsf{V})$ denote the total outdegree of P, which is the total weights connected to all vertices in a set P. Let $\mathcal{S}_W(P,Q) = \frac{\mathcal{C}_W(P,Q)}{\mathcal{D}_W(P)}$ denote the connection ratio from P to Q, In particular, $\mathcal{S}_W(\mathsf{V}_l, \mathsf{V}_l)$ is called the *normalized association* of vertex set V_l; $\mathcal{S}_W(\mathsf{V}_1, \mathsf{V}_2)$ is called the *normalized cuts* between V_1 and V_2. Summing up normalized association by attraction with normalized cuts by repulsion, or normalized cuts by attraction with normalized association by repulsion, weighted by their total degrees of connections, we generalize the normalized cuts criteria [18] as follows:

$$\epsilon_a = \sum_{l=1}^{2} \frac{\mathcal{S}_A(\mathsf{V}_l, \mathsf{V}_l)\mathcal{D}_A(\mathsf{V}_l) + \mathcal{S}_R(\mathsf{V}_l, \mathsf{V} \setminus \mathsf{V}_l)\mathcal{D}_R(\mathsf{V}_l)}{\mathcal{D}_A(\mathsf{V}_l) + \mathcal{D}_R(\mathsf{V}_l)},$$

$$\epsilon_c = \sum_{l=1}^{2} \frac{\mathcal{S}_A(\mathsf{V}_l, \mathsf{V} \setminus \mathsf{V}_l)\mathcal{D}_A(\mathsf{V}_l) + \mathcal{S}_R(\mathsf{V}_l, \mathsf{V}_l)\mathcal{D}_R(\mathsf{V}_l)}{\mathcal{D}_A(\mathsf{V}_l) + \mathcal{D}_R(\mathsf{V}_l)}.$$

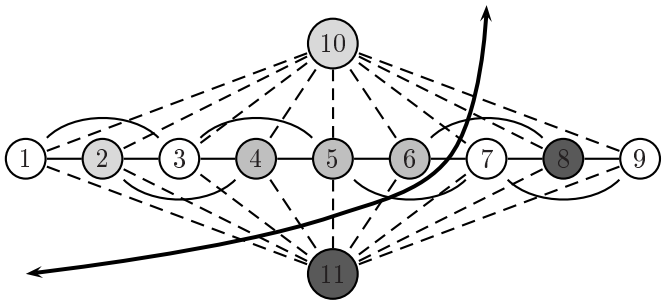

Figure 1: Represent unitary constraints and *a priori* grouping information. Here we have a $N = 5$ node graph. We introduce dummy nodes 10 and 11 as anchor nodes for V_1 and V_2 respectively. The dashed lines are connections from original nodes to dummy nodes, the weights of which indicate the preference of each node belonging to the two groups. The *a priori* grouping includes $P_1 = \{2, 10\}$, $P_2 = \{8, 11\}$, $P_3 = \{4, 5, 6\}$, which says $2 \in \mathsf{V}_1$, $8 \in \mathsf{V}_2$, and $4, 5, 6$ must be together. For all connections, we have both attraction and repulsion weights for positive and negative pairwise correlations respectively. For a bisection indicated by the thick line, our criteria minimize the attraction weights but maximize the repulsion weights cut by the partitioning.

Note that the duality between ϵ_a and ϵ_c is maintained as $\epsilon_a + \epsilon_c = 2$. The maximization of ϵ_a automatically leads to the minimization of ϵ_c and vice versa.

It turns out that these criteria can be written as Rayleigh quotients of some equivalent matrices. Let X_l be a membership indicator vector for V_l, $l = 1, 2$, where $X_l(j)$ assumes 1 if vertex j belongs to V_l and 0 otherwise. Let D_W be the diagonal degree matrix: $D_W(j,j) = \sum_k W_{jk}$, $\forall j$. Let $\mathbf{1}$ be the all-one vector. Let k denote degree ratio of V_1: $k = \frac{X_1^T D_{eq} X_1}{\mathbf{1}^T D_{eq} \mathbf{1}}$. With $y = (1-k)\, X_1 - k\, X_2$, we have:

$$\begin{array}{rclcrcl} \epsilon_a & = & \frac{y^T W_{eq} y}{y^T D_{eq} y} + \frac{\mathbf{1}^T D_A \mathbf{1}}{\mathbf{1}^T D_{eq} \mathbf{1}} & , \epsilon_c & = & \frac{y^T (D_{eq} - W_{eq}) y}{y^T D_{eq} y} + \frac{\mathbf{1}^T D_R \mathbf{1}}{\mathbf{1}^T D_{eq} \mathbf{1}}, \\ D_{eq} & = & D_A + D_R, & W_{eq} & = & A - R + D_R. \end{array}$$

2.3 Constraints for Bias

Some of the biases in grouping has been taken care of in the connection weights to dummy nodes. Others are reflected in constraints on the vector y we seek. The first is the zero-sum constraint: $y^T D_{eq} \mathbf{1} = 0$. The rest are to enforce prior assignment: $X_1(j) = 1$, $X_2(j) = 0$, $\forall j \in P_1$; $X_1(j) = 0$, $X_2(j) = 1$, $\forall j \in P_2$; and $X_1(i) = X_1(j)$, $\forall i, j \in P_l$, $l = 3, \cdots, n_p$. This is done by making $y_i - y_j = 0$, for any $i, j \in P_l$, $l = 1, \cdots, n_p$. We create an $n \times \sum_{l=1}^{n_p} (p_l - 1)$ full rank matrix C, such that each column of C has only two none-zero weights, 1 and -1 respectively, for a (e.g. subsequent) pair of vertices $i, j \in P_l$. $C^T y = 0$ guarantees that all pre-set vertices in P_l get the same labelling in a feasible solution. To make sure that vertices in P_1 have different valuation from those in P_2, we need *one* more constraint: $c^T y = 1$, which says that $y_i - y_j = 1$ for some $i \in P_1$ and $j \in P_2$. Note c and C columns are linearly independent.

We summarize these constraints in one matrix equation:

$$M^T y = 0,\ M = [\ D_{eq}\mathbf{1} \quad C\ ];\qquad c^T y = 1.$$

Since D_{eq} is full rank, M is also full rank. We denote by $m = \sum_{l=1}^{n_p} (p_l - 1) + 1$ the rank of M. The ideal solution we seek for the biased grouping is then

$$y_{opt} = \arg\max \frac{y^T W_{eq} y}{y^T D_{eq} y},\ \text{s. t. } M^T y = 0,\ c^T y = 1;\ y_j \in \{1-k, -k\},\ \forall j.$$

2.4 Computational Solutions

The above graph bisectioning problem is NP-complete. We find an approximate solution by relaxing the discreteness constraints. Such constrained Rayleigh quotient optimization arises in many applications and has been studied extensively [5]. However, we further drop $c^T y = 1$ for two reasons. The first reason is computational. When we relax the problem into the continuous domain, we do not necessarily know how far apart these two groups are in our normalized criteria, so it is impossible to set a margin in advance. The second reason is that we can actually enforce it through other parameters, such as the r in the augmented R, which describes the repulsion strength between two dummy nodes. For homogeneous linear constraints, the trick is to cast the original problem into an unconstrained optimization problem using the null space kernel of the constraint matrix.

We first turn the Rayleigh quotient of (W_{eq}, D_{eq}) into a standard form. Let $x = D_{eq}^{\frac{1}{2}} y$, $\bar{M} = D_{eq}^{-\frac{1}{2}} M$ and $\bar{W} = D_{eq}^{-\frac{1}{2}} W_{eq} D_{eq}^{-\frac{1}{2}}$. We have:

$$x_{opt} = \arg\max_x \frac{x^T \bar{W} x}{x^T x}, \qquad \text{s.t. } \bar{M}^T x = 0; \qquad \text{with } y_{opt} = D_{eq}^{-\frac{1}{2}} x_{opt}$$

We find the null vectors of $n \times m$ matrix $\bar{M}$ by its QR decomposition, where Q is an orthogonal matrix and Z is an upper triangular matrix of rank m:

$$\bar{M} = Q \begin{bmatrix} Z_{m\times m} \\ 0_{(n-m)\times m} \end{bmatrix} = \begin{bmatrix} Q_{1\,(n\times m)} & Q_{2\,(n\times(n-m))} \end{bmatrix} \begin{bmatrix} Z_{m\times m} \\ 0_{(n-m)\times m} \end{bmatrix}.$$

Q_2 contains the set of basis vectors for feasible solutions for $\bar{M}^T x = 0$. Define $(n-m) \times 1$ vector $z = Q_2^T x$ and $(n-m) \times (n-m)$ matrix $W = Q_2^T \bar{W} Q_2$. Now the problem can be formulated as an unconstrained one:

$$z_{opt} = \arg\max_z \frac{z^T W z}{z^T z}, \text{ with } x_{opt} = Q_2\, z_{opt}.$$

The optimal solution to this problem is given by the eigenvector corresponding to the largest eigenvalue λ of W. It can be shown that $|\lambda(W)| \leq 2$.

2.5 Complete Algorithm

We summarize the algorithm as follows. Given $N \times N$ pairwise attraction matrix A and repulsion matrix R for vertex set $\{1 : N\}$, $N \times 2$ unitary weight matrices B_a and B_r, parameter r and γ, pre-assigned vertex set P_l for group $l = 1 : n_p$, we compute a graph bisection as follows.

Step 1: $A := \begin{bmatrix} (1-\gamma)A & \gamma B_a \\ \gamma B_a^T & \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \end{bmatrix}$, $R := \begin{bmatrix} (1-\gamma)R & \gamma B_r \\ \gamma B_r^T & \gamma \begin{bmatrix} 0 & r \\ r & 0 \end{bmatrix} \end{bmatrix}$.

Step 2: $W_{eq} = A - R + D_R$, $D_{eq} = D_A + D_R$.

Step 3: $C_{it} = 1$, $C_{jt} = -1$, $i < j$, $i, j \in P_l$, $l = 1, \cdots, n_p$, $t = 1 : m-1$.

Step 4: $\bar{M} = D_{eq}^{-\frac{1}{2}} \begin{bmatrix} D_{eq}\mathbf{1} & C \end{bmatrix}$.

Step 5: QR decomposition: $\bar{M} = [Q_1, Q_2] \begin{bmatrix} Z_{m\times m} \\ 0 \end{bmatrix}$.

Step 6: Let $Q = D_{eq}^{-\frac{1}{2}} Q_2$ and $W = Q^T W_{eq} Q$.

Step 7: Find the first largest eigenvector z_{opt} of W.

Step 8: $y_{opt} = Q z_{opt}$.

3 Results

We demonstrate the use of bias in grouping of point sets (Fig. 2), synthetic images (Fig. 3) and figure detection in real images (Fig. 4). We also model spatial attention

in (Fig. 5). The details are given in the captions.

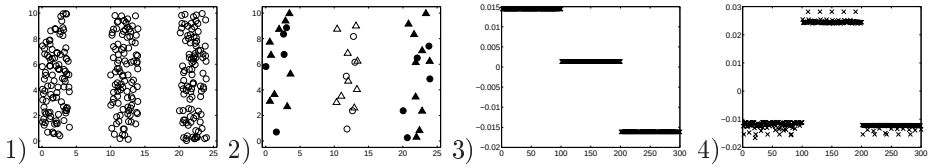

Figure 2: Point clustering with bias. 1)Data consist of three spatially separated sets of 100 uniformly distributed points. Points are numbered continuously from left to right. 2)Points in filled (empty) circles are pre-labelled to $V_1(V_2)$. Points in filled (empty) triangles are biased in B connections to $V_1(V_2)$. These constraints bias leftmost and rightmost point sets to be one group. We compute the affinity using Gaussian function of pairwise distance with $\sigma = \sqrt{2}$. 3)Segmentation without bias gives three separate groups for the three sets. 4)Segmentation with bias groups leftmost and rightmost into one class.

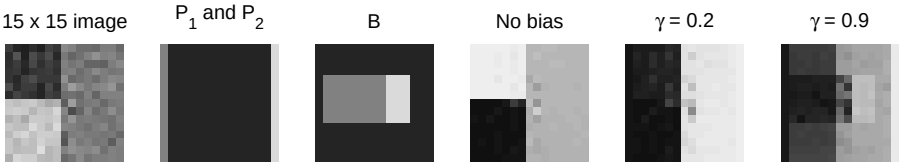

Figure 3: Image segmentation with bias. 1)An image has three intensity levels: 0, 0.5, 1, added with Gaussian noise of $\sigma = 0.1$. 2)Pixels in the leftmost(rightmost) column are pre-labelled as V_1 (V_2) . 3)Pixels in gray(white) are biased in B connections to V_1 (V_2). Notice that the bias is incorrect for two columns near the boundary. 4)Segmentation result without bias. 5)Segmentation with bias binds two dissimilar patches on the left into one group. The wrong bias is corrected based on structures in the data. 6)Large weight on bias emphasizes the mis-information in 3).

Figure 4: Figure detection with bias. 1)Background image. 2)Foreground image. 3)Image difference of 2) and 1). 4)Thresholded map of 3), which is used for B connections. The bright(dark) pixels are biased in B connections to $V_1(V_2)$. 5)Segmentation result without bias on foreground image using intervening contours [11]. 6)Segmentation result with bias. Compared to 5), holes in figure are filled in while extra blobs in ground are removed.

4 Conclusions

We propose a model in a graph framework to handle prior knowledge for data grouping. These priors can be encoded in unitary constraints, partial grouping information and weight modulation on binary constraints. Computationally, our model leads to a constrained eigenproblem. We give examples on the three types of constraints in representing various priors. This work provides a platform to explore the incorporation of generative models into graph-based grouping algorithms.

Acknowledgements: This research is supported by (DARPA HumanID) ONR N00014-00-1-0915 and NSF IRI-9817496. We thank Tai Sing Lee for valuable comments.

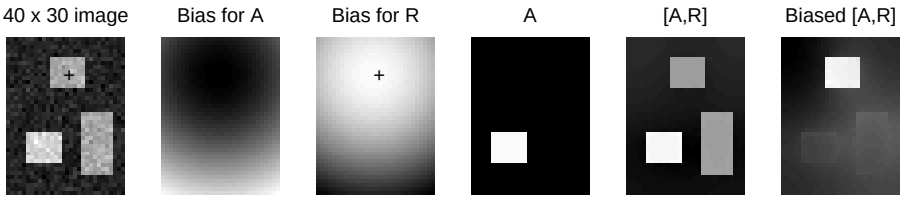

Figure 5: Modelling spatial attention in a graph framework. 1)This image has three objects of average intensity 0.8, 0.9, 0.8 against background of 0.5. Gaussian noise with $\sigma = 0.03$ is added. A and R are computed using difference of Gaussians [19] on distance and intensity difference, with σ's of 0.1, 10 respectively and two Gaussian parameter ratio of 3. To model spatial attention centered at the cross in (1), we introduce two Gaussian modulation fields of $\sigma = 30$. 2)This is an inverted Gaussian. When this bias is added to the original A, pixels far from the attention hotspot have more affinity with the background. 3)This is a normal Gaussian. When this bias is multiplied with the original R, pixels away from the hotspot get attenuated repulsion. 4)With attraction, the high contrast object stands out in segmentation. 5)With repulsion, all objects stand out against the common background. 6)With bias, only the object at the center of spatial attention pops out.

References

[1] A. Amir and M. Lindenbaum. Quantitative analysis of grouping process. In *ECCV*, pp. 371–84, 1996.

[2] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, 1987.

[3] Y. Boykov, O. Veksler, and R. Zabih. Markov random fields with efficient approximations. In *CVPR*, 1998.

[4] P. A. Ferrari, A. Frigessi, and P. G. D. SA. Fast approximate maximum a posteriori restoration of multicolour images. *J. Royal Stat. Society, B*, 57(3):485–500, 1995.

[5] W. Gander, G. H. Golub, and U. von Matt. A constrained eigenvalue problem. *Linear Algebra and its applications*, 114/5:815–39, 1989.

[6] Y. Gdalyahu, D. Weinshall, and M. Werman. A randomized algorithm for pairwise clustering. In *NIPS*, pp. 424–30, 1998.

[7] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *PAMI*, 6(6):721–41, 1984.

[8] D. M. Greig, B. T. Porteous, and A. H. Seheult. Exact maximum A Posteriori estimation for binary images. *J. Royal Stat. Society, B*, 51(2):271–9, 1989.

[9] H. Ishikawa and D. Geiger. Segmentation by grouping junctions. In *CVPR*, 1998.

[10] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *IJCV*, pp. 321–331, 1988.

[11] J. Malik, S. Belongie, T. Leung, and J. Shi. Contour and texture analysis for image segmentation. *IJCV*, 2001.

[12] M. Meila and J. Shi. Learning segmentation with random walk. In *NIPS*, 2001.

[13] D. Mumford and J. Shah. Boundary detection by minimizing functionals. In *CVPR*, IEEE Publ. 85CH2145-1., pp. 22–6. IEEE, 1985.

[14] P. Perona and W. Freeman. A factorization approach to grouping. In *ECCV*, 1998.

[15] J. Puzicha, T. Hofmann, and J. Buhmann. Unsupervised texture segmentation in a deterministic annealing framework. *PAMI*, 20(8):803–18, 1998.

[16] S. Roy and I. J. Cox. A maximum-flow formulation of the n-camera stereo correspondence problem. In *ICCV*, 1998.

[17] E. Sharon, A. Brandt, and R. Basri. Fast multiscale image segmentation. In *CVPR*, pp. 70–7, 2000.

[18] J. Shi and J. Malik. Normalized cuts and image segmentation. In *CVPR*, 1997.

[19] S. X. Yu and J. Shi. Grouping with directed relationships. In *EMMCVPR*, 2001.

[20] S. C. Zhu, Y. N. Wu, and D. Mumford. Filters, random field and maximum entropy: — towards a unified theory for texture modeling. *IJCV*, 27(2):1–20, 1998.