Boundary Detection by Constrained Optimization

DONALD GEMAN, MEMBER, IEEE, STUART GEMAN, MEMBER, IEEE, CHRISTINE GRAFFIGNE, AND PING DONG, MEMBER, IEEE

Abstract—We use a statistical framework for finding boundaries and for partitioning scenes into homogeneous regions. The model is a joint probability distribution for the array of pixel gray levels and an array of "labels." In boundary finding, the labels are binary, zero, or one, representing the absence or presence of boundary elements. In partitioning, the label values are *generic*: two labels are the same when the corresponding scene locations are considered to belong to the same region. The distribution incorporates a measure of *disparity* between certain spatial features of pairs of blocks of pixel gray levels, using the Kolmogorov–Smirnov nonparametric measure of difference between the distributions of these features. Large disparities encourage intervening boundaries and distinct partition labels. The number of model parameters is minimized by *forbidding* label configurations that are inconsistent with prior beliefs, such as those defining very small regions, or redundant or blindly ending boundary placements. Forbidden configurations are assigned probability zero. We examine the MAP (*maximum a posteriori*) estimator of boundary placements and partitionings. The forbidden states introduce constraints into the calculation of these configurations. Stochastic relaxation methods are extended to accommodate *constrained* optimization, and experiments are performed on some texture collages and some natural scenes.

Index Terms—Annealing, Bayesian inference, boundary finding, constrained optimization, Gibbs distribution, MAP estimate, Markov random field, segmentation, stochastic relaxation, texture discrimination.

I. INTRODUCTION

MANY problems in image analysis, from signal restoration to object recognition, involve representations of the observed data, usually radiant energy or range measurements, in terms of *unobserved* attributes or label variables. These representations may be conclusive, or serve as intermediate data structures for further analysis, perhaps involving additional data, assorted "sketches," or stored models. This work concerns two such represen-

tations related to image segmentation: partition and boundary labels. These, in turn, are special cases of a general "label model" in the form of an "energy functional" involving two components, one of which expresses the interactions between the labels and the (intensity) data, and the other encodes constraints derived from general information or expectations about label patterns. The labeling we seek is, by definition, the minimum of this energy.

The partition labels do not classify. Instead they are *generic* and are assigned to blocks of pixels; the size of the blocks (or label resolution) is variable, and depends on the resolution of the data and the intended interpretations. The boundary labels are just "on"/"off," and are also of variable resolution, associated with an interpixel sublattice. In both cases, the interaction term incorporates a measure of disparity between certain spatial features of pairs of blocks of pixel gray levels, using the Kolmogorov–Smirnov nonparametric measure of difference between the distributions of these features. Large disparities encourage intervening boundaries and distinct partition labels. The number of model parameters is reduced by *forbidding* label configurations that are inconsistent with prior expectations, such as those defining very small regions, or redundant or blindly ending boundary placements. These forbidden states introduce constraints into the calculation of the optimal label configuration.

Both models are applied mainly to the problem of texture segmentation. The data is a gray-level image consisting of textured regions, such as a mosaic of microtextures from the Brodatz album [6], a patch of rug inside plastic, or radar-imaged ice floes in water. It is well-known that humans perceive "textural" boundaries between regions of approximately the same average brightness because the regions themselves, although containing sharp intensity changes, are perceived as "homogeneous" based on other properties, namely those involving the spatial distribution of intensity values. The goal, then, is to find the (visually) distinct texture regions, either by assigning categorical labels to the pixels, or by constructing a boundary map. Obviously, the problem is more difficult than texture classification, in which we are presented with only one texture from a given list; segmentation may be complicated by an absence of information about the number of textures, or about the size, shape, or number of regions.

There is no "model" for the individual textures, and hence no capacity for texture synthesis. Our approach is

therefore different from those for classification and segmentation in which textures are discriminated based on model parameters which are estimated from specific texture samples. Partitionings and boundary placements are driven by the observed spatial statistics as summarized by selected features. Still, the labeling is not unsupervised because in some cases we use "training samples" to determine feature thresholds for the disparity measures; see Sections II and III.

Since many visually distinct textures have nearly identical histograms, segmentation must rely on features, or transformations, which go beyond the raw gray levels, involving various spatial statistics. We experimented with several conventional classes of features, in particular the well-known ones based on cooccurrence matrices [16], [31], but finally adopted a new class based mainly on "directional residuals" and involving third and higher order distributions. However, our viewpoint is exactly that expressed by Zobrist and Thompson [62], Triendl and Hendersen [59], and others: instead of trying to find exactly the "right" features to convert texture to tone differences, one should find a mechanism for integrating multiple, even redundant, "cues." The label model provides a coherent method for integrating such information and, simultaneously, organizing the label patterns.

Whereas texture discrimination may be regarded as the detection of discontinuities in surface composition, we also apply the boundary model to the problem of locating sudden changes in depth (occluding boundaries) or shape (surface creases, etc.). The idea is to define contours which are faithful to the 3-D scene but avoid the "nonphysical" edges due to noise, digitization, texture, lighting, etc. In particular, the boundaries should be connected, unique, and reasonably smooth, unlike the typical output of a pure edge detector. Indeed, we formulate boundary detection as a single optimization problem, fusing the detection of edges with their pruning, linking, smoothing, and so on. Obviously, there are discontinuities, such as shadows, which are essentially impossible to distinguish from the occluding and shape boundaries, at least without information from multiple sensors or a rich knowledge base, in which case boundary classification becomes possible.

Our model enjoys some invariance properties. In particular, due to the use of the Kolmogorov–Smirnov distance, the detected labels are unaffected by linear changes in illumination, and in some cases by all monotone data transformations. Such distortions often result from sensor nonlinearities, digitizers, and film properties. In regard to rotation invariance, the estimated labeling will be independent of the relative orientation of the training samples with the image data to the extent that the features are rotation invariant. The pixel lattice precludes true invariance. However, since our collection of features is basically isotropic, our results are approximately invariant. See [38] for an approach to rotation-invariant *classification*.

Apparently, some of these ideas have been around for a while. For example, the Kolmogorov–Smirnov statistic is recommended in [51], and reference is made to still earlier papers; more recently, see [60]. Moreover, the *distributional* properties of residuals (from surface-fitting) are advocated in [29], [48] for detecting discontinuities. It is certainly our contention that the statistical warehouse is full of useful tools for computer vision.

Finally, our model may be interpreted in a Bayesian framework as a "prior" joint probability distribution for the array of pixel gray levels and the array of labels. Forbidden configurations are assigned probability zero, and the label estimate is then associated with the MAP (maximum *a posteriori*) estimate. We shall discuss this interpretation in more detail later on. Suffice to say that, whereas our formulation of the problem and definition of the "best" labeling are formally independent of the stochastic outlook, our optimization procedures are in fact strongly motivated by this viewpoint. In fact, in order to deal with the more difficult textures, the deterministic ("zero-temperature") relaxation algorithm we use for most of our experiments must be replaced by a version of stochastic relaxation extended to accommodate constrained sampling and optimization.

A. Applications

Texture is a dominant feature in remotely sensed images and regions cannot be distinguished by methods based solely on shading, such as edge detectors or clustering algorithms. Specifically, for example, one might wish to determine the concentration of ice in synthetic aperture radar images of the ocean, or analyze multispectral satellite data for land use classification. Another application is to wafer inspection: low magnification views of memory arrays appear as highly structured textures, and other geometries have a characteristic, but random, graining. In addition, the use of boundary maps as the input to further processing is ubiquitous in computer vision; for example, algorithms for stereopsis, optical flow, and simple object recognition are often based on matching boundary segments. Other applications include the analysis of medical images, automated navigation, and the detection of roads, faults, field boundaries, etc. in remotely sensed images. (Obviously no generic algorithm provides "off-the-shelf" solutions to real problems; genuine applications require substantial adaptations.)

B. Label Model

Let $x = \{x_s, s \in S\}$ and $y = \{y_{ij}, 1 \leq i, j \leq N\}$ denote, respectively, the labels and the data; thus x_s is the label at "site" $s \in S$ and y_{ij} is the gray-level at pixel (i, j). The set S of label sites is a regular lattice, distinct from that of the pixels, and typically more sparse; the coarseness depends on the *label resolution* σ. For partitioning, we associate each site $s \in S$ with a block of pixels, "sitting below it," if we were to stack the label lattice on top of the pixel lattice. In the boundary model,

pairs of nearby sites in S define boundary *segments*, and these are associated with pairs of pixel blocks, sitting "across from each other," with respect to the segments (see Fig. 6). Later, we will define a *neighborhood system* for S such that the bonding is nearest-neighbor (relative to σ) in the boundary model, whereas in the region model there are interactions at all scales. This has important consequences for the distribution of local minima in the "energy landscape"; see Section II. Other energy functionals with global interactions can be found in [22], [26], and [47].

The "interaction" between x and y is defined in terms of an energy function

$$U(x, y) = \sum_{\langle s, t \rangle} \Psi_{s,t}(x) \Phi_{s,t}(y).$$

The summation extends over all "neighboring pairs" (or "bonds") $\langle s, t \rangle$, $s, t \in S$; $\Phi_{s,t}(y)$ is a measure of the *disparity* between the two blocks of pixel data associated with the label sites $s, t \in S$. $\Psi_{s,t}(x)$ depends only on the labels x_s and x_t. In fact, we simply take $\Psi_{s,t}(x) = 1 - x_s x_t$ in the boundary model and $\Psi_{s,t}(x) = \delta_{x_s = x_t}$ in the partition model. In this way, in the "low energy states," large disparities ($\Phi > 0$) will typically be coupled with an active boundary ($x_s = x_t = 1$) or with dissimilar region labels ($x_s \neq x_t$) and small disparities ($\Phi < 0$) will be coupled with an inactive boundary ($x_s = 0$ or $x_t = 0$) or with equal region labels ($x_s = x_t$).

The interaction between the labels and the data is based on various *disparity measures* for comparing two (possibly distant) blocks of image data. These measures derive from the raw data as well as from various *transformations*. We experiment with several transformations $y \rightarrow y'$ for texture analysis, for example transformations of the form

$$y_t' = \left| y_t - \sum \alpha_j y_{t_j} \right| \tag{1.1}$$

where $\sum \alpha_i = 1$ and $\{t_j\}$ are pixels nearby to pixel t in the same row, column, or diagonal. We call these "directional residuals," regarding $\sum \alpha_j y_{t_j}$ as a "predictor" of y_t. We have also investigated raw gray levels ($y' = y$), other features based on local measures of gray-level *range*, and isotropic versions of (1.1). In any case, disparity is measured by the *Kolmogorov–Smirnov statistic* (or *distance*), a common tool in nonparametric statistics which has desirable invariance properties. (In particular, using the directional residuals, the disparity measure is invariant to linear distortions ($y_{ij} \rightarrow a y_{ij} + b$) of the raw data, and using the raw data itself for comparisons, the disparity measure is invariant to *all* monotone (data) transformations.) The general form of the Φ term is then

$$\Phi_{s,t}(y) = \phi\left(\max_{1 \leq i \leq m} \rho(y_{(s)}^{(i)}, y_{(t)}^{(i)}) \right)$$

where ϕ is monotone increasing, ρ denotes a distance based on the Kolmogorov–Smirnov statistic (see Section

II), and $y_{(s)}^{(i)}$, $y_{(t)}^{(i)}$ are the data in the two blocks associated with $\langle s, t \rangle$ for the ith transform. Often, we simply take $m = 1$ and $y^{(1)} = y$.

The other component in the model is a *penalty function* $V(x)$ which counts the number of "taboo patterns" in x; states x for which $V(x) > 0$ are "forbidden." For example, boundary maps are penalized for dead-ends, "clutter," density, etc. whereas partitions are penalized for too many transitions or regions which are "too small."

Given the observed image y, our estimate $\hat{x} = \hat{x}(y)$ is then any solution to the constrained optimization problem

$$\text{minimize}_{x: V(x) = 0} U(x, y). \tag{1.2}$$

We seek to minimize the energy of the data-label interaction over all possible nonforbidden label states x.

The rationale for constrained optimization is that our expectations about certain types of labels are quite precise and rigid. For example, most "physical boundaries" are smooth, persistent, and well-localized; consequently it is reasonable to *impose* these assumptions on image boundaries, and corresponding restrictions on partition geometries. Contrast this with other inference problems, for example restoring an image degraded by blur and noise. Aside from constraints derived from scene-specific knowledge, the only reasonable *generic* constraints might be "piecewise continuity," and generally the degree of ambiguity favors more flexible constraints, such as those in U, or in the energy functions used in [20] and [22].

There are no multiplicative parameters in the model, such as the "smoothing" or "weighting" parameters in [3], [17], [19], [20], and [22]; in effect, the energy is $U + \lambda V$ with $\lambda = \infty$. Thresholds must be selected for the disparity measures; this may be done "in the blind," but the performance of the algorithm is certainly increased if these choices are data-driven, either from prior experience with the types of textures, or by extracting training samples, as we have done here (see Section V). Fortunately, the algorithm is not unduly sensitive to these choices within a certain range of values. Other inputs include the label resolution, block sizes, and penalty patterns. The algorithm is robust against these choices provided that modest information is available about the pixel resolution. The selection of penalty patterns is nearly self-evident: as already noted, our expectations about the geometry of regions and boundaries are very concrete and easily encoded in local constraints.

C. Constrained Stochastic Relaxation

The search for $\hat{x}$, the optimal labeling, is based on a version of stochastic relaxation which incorporates hard constraints. The theoretical foundations are laid out in [18], although there is enough information provided here (see Section IV) to keep this paper self-contained. (See also Grenander [28] and Linsker [41] for similar modifications of stochastic relaxation.) We employ two algorithms, both of which are approximations to precise computational procedures, in which convergence is at least

theoretically guaranteed, if not always realized in practice.

Introduce a control parameter t corresponding to "temperature" and another control parameter λ, corresponding to a Lagrange multiplier for the constraint $V = 0$. Let

$$U_k(x) = t_k^{-1}[U(x, y) + \lambda_k V(x)]$$

where y (the data) is fixed, $t_k \searrow 0$, and $\lambda_k \nearrow \infty$. One generates a sequence of states $\hat{x}_k$, $k = 1, 2, \cdots$, by Monte Carlo sampling from the Gibbs measures with energy functions U_k. If t_k is fixed, then under suitable conditions on the growth of λ_k, the sequence x_k converges to a sample from the Gibbs distribution with energy U, but restricted to $V = 0$, i.e., constrained to assign probability zero to the forbidden configurations. We refer to this as "constrained stochastic relaxation." On the other hand, if $\lambda_k \nearrow \infty$ and $t_k \searrow 0$, both at suitably (coupled) rates (see Section IV), then the sequence converges to a solution of (1.2); this is "constrained simulated annealing."

The first algorithm we use is deterministic, with the temperature fixed at zero, and is essentially the ICM ("iterated conditional mode") algorithm of Besag [2], although $\lambda = \lambda_k \nearrow \infty$. The convergence is rapid and the performance is sufficient for all but the most difficult images. The second algorithm is constrained stochastic relaxation with the temperature fixed at some "small" value (but, again, $\lambda = \lambda_k \nearrow \infty$); the idea is simply that the very likely states of the constrained Gibbs distribution should be "close" to the mode, i.e., $\hat{x}$. This "low temperature sampling" is more computationally demanding than ICM, but invariably arrives at a better labeling; see Section V on experiments.

D. Related Work

The subject of edge detection is very active; however, as we have already mentioned, the raw output of standard edge detectors, for example those based on differential operators, would generally require considerable "processing" to produce unique, smooth, persistent boundaries. The work in "regularization theory" [44], [49], as well as our earlier paper [20], is different in that "edge" or "line" process is employed there as a device for suspending continuity constraints in the context of reconstructing an image from sparse data, restoring an image corrupted by blur and noise, and similar tasks not aimed at boundary detection per se.

In contrast, there is much related work in the area of texture segmentation. In Triendl [58] and Triendl and Thompson [59], local property values referred to as "texture parameters" are computed in a neighborhood of each pixel and edge maps are computed for several principal components of these feature images. These maps are integrated across components, and then spatially, resulting in a final boundary map. Similarly, in Thompson [57] (see also Zobrist and Thompson [62]), the idea is to integrate multiple cues into a single "distance function" which is a linear combination of elementary measures, and which

is intended to measure the overall similarity between two regions. Edges are then detected using the Robert's cross gradient applied to this distance; there is no effort to organize the responses. Our experiments compare favorably to those in [57] and [59], both of which utilize collages of natural textures similar to ours. Local filtering is also the focus of Laws' approach [40], although the mechanism for combining cues is entirely different from ours.

Among the statistical approaches to texture segmentation are those in which the image data is regarded as a realization of a two-dimensional stochastic process, or random field. In Simchony and Chellappa [53] and in Derin and Cole [12], the image is modeled as a two-layered Markov random field: the "upper level" is the region process, a simple Ising-type process, and the "bottom level" is the observed intensity process, with some distribution (e.g., Gaussian) conditional on the regions. Both papers employ stochastic relaxation to search for the MAP and related estimators. Modestino *et al.* [45] use random tessellations for the upper level and maximum likelihood estimation based on cooccurrence data for the labeling. In each case, the random field parameters are assumed known in advance or are estimated from training samples. The image data in [12] is generated from the model, and simpler than the actual texture collages in [53] and [45]. Again, our results compare favorably.

Finally, Kashyap and Eom [37] employ a "long correlation" random field model and statistical hypothesis testing to obtain a boundary map. Boundary placements are estimated in each small strip of the image based on least-squares estimates of the model parameters. There is no training data and the results are good, although, naturally, the boundaries are somewhat disorganized.

E. Stochastic Formulation

This work is an extension, incorporating hard constraints, of a Bayesian paradigm in which prior probability models are constructed for both the observed and unobserved scene attributes. The degradation process is modeled as a conditional probability distribution, and the estimates are obtained by a form of Monte Carlo sampling. This framework has been applied by numerous researchers, including applications to image restoration [20], [2], [9], [24], [28], image synthesis [28], computed tomography [22], texture and boundary analysis [17], [53], [19], [21], [25], scene segmentation based on optical flow [47] or shading and texture [11], [13], frame-to-frame matching for computing optical flow and stereo disparity [36], and surface reconstruction [43], [44].

To facilitate placing our model in the general framework, let us write $x = (x^L, x^P)$, where x^L is the vector of boundary or region labels, and x^P is the vector of pixel gray levels. These are the relevant image attributes for the problem at hand. The prior distribution Π is a probability for x: $0 \leq \Pi(x) \leq 1$, $\Sigma_x \Pi(x) = 1$, where Σ_x is the summation over all configurations of x (all assignments of gray levels and boundary placements, for example). Adopting the Gibbs representation in the unconstrained

case we would represent Π as

$$\Pi(x) = \frac{1}{z} \exp\left\{-U(x)\right\}, \qquad z = \sum_x \exp\left\{-U(x)\right\}.$$

The real-valued function U is called the energy, and evidently determines Π. In our case we separate the prior energy into a pixel-label interaction term and a pure label contribution. The former, $U(x^L, x^P)$, promotes placements of boundaries, or assignments of distinct labels, between regions in the image that demonstrate distinct spatial patterns. The pure label contribution is to inhibit "blind" endings of boundaries, redundant boundary representations, excessively small or thin regions, and other unexpected label configurations. Rather than inhibiting these configurations with high energy via large parameters, we have found it convenient, especially when working with boundaries and partitionings, to extend this framework, by allowing "infinite energies" (zero probabilities) in the prior distribution. (See Moussouris [46] for an analysis of Gibbs measures with "forbidden states.") We forbid these configurations by introducing $V = V(x^L) \geq 0$ and concentrating on $\{x: V(x^L) = 0\}$. The prior, then, is of the form

$$\Pi(x) = \frac{1}{z} \delta_{\{V=0\}}(x^L) \exp\left\{-U(x^L, x^P)\right\}.$$

The constraint, $V(x^L) = 0$, amounts to a placement of infinite energy barriers in the "energy landscape." These inhibit the free flow that is essential to the good performance of stochastic relaxation; indeed the theory willl in general break down, and convergence is no longer guaranteed. As indicated above, a simple and effective solution is to introduce the barriers gradually during the relaxation process. Again, the supporting convergence theory is spelled out in [18].

There is usually a problem-specific degradation that precludes directly observing x. It may be the blur and noise introduced by the camera and recording devices, the attenuated Radon transform that figures into emission tomography, or, as in this case, an "occlusion": we assume the gray levels x^P are observed (uncorrupted) whereas the labels x^L are of course unobserved. Thus the data is $y = x^P$ and our only interest is in estimating the unobserved label process x^L. $\Pi(y|x)$ is singular, and the posterior distribution for x^L given the data is

$$\Pi(x^L|y) = \frac{1}{\hat{z}} \delta_{\{V=0\}}(x^L) \exp\left\{-U(x^L, y)\right\}.$$

There is skepticism about the MAP estimator: see, e.g., [2], [15], [44]. It has sometimes been found to be too "global," leading to gross mislabeling in certain classification problems and "over-smoothing" in surface reconstruction and image restoration. (See [7] for a different view.) The discussion paper of Besag [2] has shed much light on the subject; see especially the remarks of Silverman [52] on MAP versus simulations from the posterior $\Pi(x|y)$, and the remarkable comparisons between the

exact MAP estimate and approximations derived from simulated annealing in the commentary of Greig, Porteous, and Seheult [27]. However, pixel-based error measures are too local for boundary analysis. In particular, the Bayes rule based on misclassification error rate, namely the *marginal* (individual component) modes of $\Pi(x|y)$, is unsuitable because this estimator lacks the fine structure we expect of boundary maps; placement decisions cannot be based on the data alone—pending labels (i.e., *context*) must be considered. See [44], [50], and [61] for discussions of alternative loss functions and performance criteria.

II. PARTITION MODEL

A. Partitionings

Denote the pixel (image) lattice $\{(i, j): 1 \leq i, j \leq N\}$ by S_I and let S_L (formally S in Section I) be the label lattice, just a copy of S_I in the case of the partition model. For each experiment, a *resolution* σ is chosen, which determines a sublattice $S_L^{(\sigma)} \subset S_L$, and the coarseness of the partitioning. Larger σ's will correspond to coarser partitionings and give more reliable results (see Section V), but they lose boundary detail. Specifically, let

$$S_L^{(\sigma)} = \left\{(i\sigma + 1, j\sigma + 1): 0 \leq i, j \leq \frac{N-1}{\sigma}\right\}.$$

Recall that the observation process, or data, consists of gray levels y_s, $s \in S_I$. With the usual gray-level discretization, the state, or configuration, space for the data is

$$\Omega_I = \left\{\{y_s\}: s \in S_I, 0 \leq y_s \leq 255\right\}.$$

The configuration space for the partitioning x is determined by σ and by a maximum number of allowed regions P:

$$\Omega_L^{(\sigma, P)} = \left\{\{x_s\}: s \in S_L^{(\sigma)}, 0 \leq x_s \leq P - 1\right\}.$$

Recall that the labels are generic: x defines a partitioning by identifying sites with a given label $(0, 1, \cdots, P - 1)$ as belonging to the same region. Only the sublattice $S_L^{(\sigma)}$ is labeled, and a maximum number of labels (regions) is fixed *a priori*. A prior estimate of the number of distinct (but not necessarily connected) regions must be available, since the model often subdivides homogeneous regions when P is too large (see Section V). The boundary model (Section III) is more robust in this regard.

Each label site $s \in S_L$ is associated with a square block $D_s \in S_I$ of pixel sites centered at s. (Recall that S_L is just a copy of S_I; we sometimes use "s" ambiguously to reference a site in S_L *and* the corresponding site in S_I.) x_s labels the pixels in D_s: $\{\{y_r\}: r \in D_s\}$. As we will see shortly, the partitioning is based on the spatial statistics of these (overlapping) subimages. The size of D_s is therefore important. We have experimented only with textures (the boundary model has been applied more generally), and is obvious that for these the pixel blocks $\{D_s\}_{s \in S_L^{(\sigma)}}$ must be large enough to capture the characteristic pattern of the texture, at least in comparison to the other textures

present. Of course, "large enough" is with respect to the features used, but in the absence of a *multiscale* analysis, an *a priori* choice of scale is unavoidable. In all of our experiments, $|D_s| = 441$, a 21×21 square block of pixels. There is again a resolution issue: larger blocks characterize the textures more reliably, having less within-region variation, but boundary detail is sacrificed.

B. Label-Data Interaction

We establish a neighborhood system on the label lattice $S_L^{(\sigma)}$: each $s \in S_L^{(\sigma)}$ is associated with a set of neighbors $N_s \subset S_L^{(\sigma)}$. The system is symmetric, meaning that $s \in N_r \Leftrightarrow r \in N_s$. As we shall see, the neighborhood system largely determines the computational burden. For now we will proceed as though the neighborhood system is given, but we will have much more to say about it shortly.

Let $\langle s, t \rangle_\sigma$ denote a neighbor pair, meaning s, $t \in S_L^{(\sigma)}$, $s \in N_t$. We will introduce a *disparity measure* $\Phi_{s,t} = \Phi_{s,t}(y)$ for each neighbor pair $\langle s, t \rangle_\sigma$. Roughly speaking, $\Phi_{s,t}$ measures the similarity between the pixel gray levels in the two pixel blocks associated with s, $t \in S_L^{(\sigma)}$. For the partition model, $\Phi_{s,t}$ is simply -1 ("similar") or $+1$ ("dissimilar"); it is more complicated for the boundary model. The interaction energy is then

$$U(x, y) = \sum_{\langle s,t \rangle_\sigma} \delta_{\{x_s = x_t\}} \Phi_{s,t}(y),$$

where $\delta_{\{x_s = x_t\}} = 1$ if $x_s = x_t$, and 0 otherwise. In the low energy states, similar (resp. dissimilar) pairs, $\Phi_{s,t} = -1$ (resp. $\Phi_{s,t} = +1$), are associated with identical (resp. distinct) labels: $x_s = x_t$ (resp. $x_s \neq x_t$). Although $U(x, y)$ is conceived of as the interaction term in a *prior* distribution that is jointly on x and y, only the posterior distribution is actually used, and y is fixed by observation. It would be interesting, and perhaps instructive (see [35]), to sample from the joint distribution, but computationally very expensive.

C. Neighborhood System

A simple example will serve to highlight the issues. Suppose y has R *constant gray-level* (untextured) regions ($y_s \in \{0, 1, \cdots R - 1\}$, $s \in S_I$), and $\sigma = 1$ (full resolution). Of course, in this case y *is* a labeling, so there is no point in bringing in the partition process x; but this is just an illustrative example. The obvious disparity measure is simply $\Phi_{s,t} = -1$ if $y_s = y_t$, and $+1$ otherwise:

$$U(x, y) = \sum_{\langle s,t \rangle_1} \delta_{\{x_s = x_t\}} (\delta_{\{y_s \neq y_t\}} - \delta_{\{y_s = y_t\}}). \quad (2.1)$$

Entertain, for the time being, a nearest neighbor system on S_L, which is the natural choice. To be concrete, take N_s to be the four (two horizontal and two vertical) nearest neighbors of s. There are three essential difficulties with this choice of neighborhood system. Two can be readily appreciated:

- (See Fig. 1.) If $R = 2$ and $P = 3$, and if region "0" (i.e., $\{s \in S_I: y_s = 0\}$) is split into two disjoint pieces by region "1" (i.e., $\{s \in S_I: y_s = 1\}$), then (2.1) has

0	0	0	1	1	0	0
0	0	0	1	1	0	0
0	0	1	1	1	0	0
0	0	1	1	0	0	0
0	1	1	0	0	0	0
0	1	1	0	0	0	0
1	1	1	0	0	0	0

1	1	1	2	2	1	1
1	1	1	2	2	1	1
1	1	2	2	2	1	1
1	1	2	2	1	1	1
1	2	2	1	1	1	1
1	2	2	1	1	1	1
2	2	2	1	1	1	1

2	2	2	1	1	0	0
2	2	2	1	1	0	0
2	2	1	1	1	0	0
2	2	1	1	0	0	0
2	1	1	0	0	0	0
2	1	1	0	0	0	0
1	1	1	0	0	0	0

Fig. 1. Left: original "image" and a correct labeling. Middle: a correct labeling. Right: spurious labeling.

0	0	0	1	1	2	2
0	0	0	1	1	2	2
0	0	1	1	1	2	2
0	0	1	1	2	2	2
0	1	1	2	2	2	2
0	1	1	2	2	2	2
1	1	1	2	2	2	2

1	1	1	2	2	0	0
1	1	1	2	2	0	0
1	1	2	2	2	0	0
1	1	2	2	0	0	0
1	2	2	0	0	0	0
1	2	2	0	0	0	0
2	2	2	0	0	0	0

0	0	0	1	1	0	0
0	0	0	1	1	0	0
0	0	1	1	1	0	0
0	0	1	1	0	0	0
0	1	1	0	0	0	0
0	1	1	0	0	0	0
1	1	1	0	0	0	0

Fig. 2. Left: original "image" and a correct labeling. Middle: a correct labeling. Right: spurious labeling.

two kinds of global minima: correct labelings, in which there are two populations of labels corresponding to the two gray-level regions; and spurious labelings, in which the three regions (two of type "0" and one of type "1") are given three distinct labels.

- (See Fig. 2.) If $R = 3$, and region "0" does not neighbor region "2", then there are again two kinds of global minima: correct labelings have three labels; spurious labelings have only two, incorrectly identifying regions "0" and "2".

Quite obviously, the model requires more *global interactions*. In particular, just a few long range interactions would distinguish the correct from the spurious labelings. Only a correct labeling would achieve the *global* minimum of U in these two examples.

The third difficulty with local neighborhoods is computational, and is already apparent when $R = 1$ and $P = 2$. This time there are only two global minima, and each is a desirable labeling ($\{x_s = 0 \; \forall_s \in S_L\}$ or $\{x_s = 1 \; \forall_s \in S_L\}$). But, with $N = 512$, for example, consider the label configuration in which $x_{ij} = 0$ whenever $1 \leq i \leq 256$ and $x_{ij} = 1$ whenever $257 \leq i \leq 512$, a half "black" and half "white" picture. This is a local minimum, and rather severe in that it would take very many "uphill" or "flat" moves (single site changes) to arrive at either of the global minima. SR is a *local* relaxation algorithm, and despite the various convergence theorems, the practical fact of the matter is that "wide" local minima such as these are impossible to cope with. (But, there are some encouraging results in the direction of *multiscale* relaxation, see [24], [5], [54].) Many readers will be reminded of the Ising model, in the absence of an external field, and the notorious difficulty of finding its (two) global minima by Monte Carlo relaxation. In fact, the $R = 1, P = 2$ energy landscape is *identical* to that of the Ising model, as is readily demonstrated by a suitable transformation of the label variables. (Indeed, the same goes for the $R = 2, P = 2$ case, although this is less obvious. A suitable transformation identifies the two Ising minima with the two acceptable labelings: $x_s = 0 \Leftrightarrow y_s = 0$ and $x_s = 1 \Leftrightarrow y_s = 0$.)

These local minima can be mostly eliminated by introducing long range interactions in the label lattice $S_L^{(\sigma)}$; the same remedy as for the label ambiguities. We will provide a heuristic argument for the important role of long range interactions in creating a favorable energy landscape. In any case, simulations firmly establish their utility. First recall that the *distance* between two sites in a graph is the smallest number of edges that must be crossed in traveling from one site to the other. Notice that in the four nearest neighbor graph (two-dimensional lattice) the average distance between sites is large. Correct partitioning requires *all* pairs of label sites to resolve their relationships ("same" or "different"), as dictated by the statistics of their associated pixel blocks. Of course most pairs are not neighbors. With a *local* relaxation, such as SR, the resolution is achieved by propagating relationships through intervening sites. Thus the task is facilitated by minimizing the number of intervening sites, and a relatively small number of long range connections can drastically reduce the typical number of these.

The largest distance over all pairs of sites is the *diameter* of a graph. In an appropriate limiting (large graph) sense, *random graphs* have minimum diameter among all graphs of fixed degree.[1] In light of our heuristics, this suggests a random neighborhood system for $S_L^{(\sigma)}$. Indeed, random neighborhoods have a remarkable effect on the structure of local minima for these systems. In a series of experiments, with "perfect" disparity data (such as the $\sigma = 1$ gray-level problems discussed above) we could always achieve the global minimum by single-site iterative improvement when adopting a random graph neighborhood configuration, using rather modest degrees for large graphs. *We conjecture, but have been unable to prove, that even with the degree a vanishingly small fraction of the graph size, random graphs (in the "large graph limit") have no local minima, under the Ising potential or the potential $U(x, y)$ with perfect disparity data (2.1).*

Of course the disparity data is not usually perfect. In challenging texture discrimination tasks there will be pixel blocks from the same texture that are measured as dissimilar ($\Phi_{s,t} = 1$) and others from distinct textures that are measured as similar ($\Phi_{s,t} = -1$). Under these circumstances it helps to also have near neighbor interactions, since these tend to bond neighboring label sites and thereby increase the effective number of long range interactions per site. Although near neighbors were not always needed to get the best results, we settled on using four near neighbors, and sixteen random neighbors per site, in each of our experiments (see Section V). By "near neighbors" we mean the closest two horizontal and two vertical neighbors whose associated pixel blocks do not overlap. For example, with $\sigma = 7$, and using 21×21 blocks, the near neighbors have two intervening sites in $S_L^{(\sigma)}$. Details on the generation of the (pseudo) random neighbors can be found in [25]. Overall, perhaps the most effective

---

[1]A graph has *fixed degree* if each site has the same number of neighbors. The degree is then the number of neighbors per site.

neighborhood system would have a gradual fall-off of interaction densities with distance, a system with an equal number of neighbors at each Manhattan distance, for example.

D. Kolmogorov-Smirnov Statistic

At the heart of the partitioning and boundary algorithms is a disparity measure $\Phi_{s,t}$. Recall that if s, t are neighbors in $S_L^{(\sigma)}(\langle s, t \rangle_\sigma)$ then $\Phi_{s,t}$ is a measure of disparity between two corresponding blocks of pixel data, $\{\{y_r\}: r \in D_s\}$ and $\{\{y_r\}: r \in D_t\}$ in the case of the partition model. We base $\Phi_{s,t}$ on the Kolmogorov–Smirnov distance, a measure of separation between two probability distributions, well-known in statistics. When applied to the sample distributions (i.e., histograms) for two sets of data, say $v^{(1)} = \{v_1^{(1)}, v_2^{(1)}, \cdots, v_{n_1}^{(1)}\}$ and $v^{(2)} = \{v_1^{(2)}, v_2^{(2)}, \cdots, v_{n_2}^{(2)}\}$ it provides a test statistics for the hypothesis that $v^{(1)}$ and $v^{(2)}$ are samples from the same underlying probability distribution, meaning that $F_1 = F_2$ where, for $i = 1, 2$, $v^{(i)} = \{v_1^{(i)}, v_2^{(i)}, \cdots, v_{n_i}^{(i)}\}$ are independent and identically distributed with $F_i(t) = P(v_j^{(i)} \leq t)$. The test is designed for continuous distributions and has a powerful invariance property which will be discussed below.

The sample distribution function of a data set $\{v_1, v_2, \cdots, v_n\}$ is

$$\hat{F}(t) = \frac{1}{n} \# \{k: v_k \leq t\}, \quad -\infty < t < +\infty.$$

Thus, $\hat{F}$ is a step function, with jumps occurring at the points $\{v_k\}$. It characterizes the histogram. Now consider two sets of data $v^{(1)}$, $v^{(2)}$ with sample distribution functions $\hat{F}_1$, $\hat{F}_2$. The Kolmogorov–Smirnov distance (or statistic) is the maximum (vertical) distance between the graphs of $\hat{F}_1$, $\hat{F}_2$, i.e.,

$$d(v^{(1)}, v^{(2)}) = \max_{-\infty < t < +\infty} \left| \hat{F}_1(t) - \hat{F}_2(t) \right|. \quad (2.2)$$

We write $d(v^{(1)}, v^{(2)})$ to emphasize the data (which in our case consists of blocks of possibly transformed pixel intensity values); the conventional notation is $d(\hat{F}_1, \hat{F}_2)$.

The invariance property is the following. Suppose $v^{(1)}$, $v^{(2)}$ are samples from *continuous* distributions F_1, F_2. The *under the ("homogeneity") hypothesis $F_1 = F_2$, the probability distribution of d (as a random variable) is independent of the (common) underlying distribution.* Basically, this stems from the fact that d is invariant to strictly monotone transformations of the data, i.e.,

$$d(v^{(1)}, v^{(2)}) = d(q^{(1)}, q^{(2)})$$

where $q_j^{(i)} = \tau(v_j^{(i)})$ and τ is strictly increasing or decreasing. Thus, in two sample tests for homogeneity, one rejects the null hypothesis that $F_1 = F_2$ if $d(v^{(1)}, v^{(2)}) \geq d^*$, *where d^* depends only on n_1 and n_2, and on the significance level of the test.*

For our purposes, the data $v^{(1)}$ and $v^{(2)}$ consist of either the (raw) gray levels, or (in most cases of texture discrimination) transformations of these, restricted to blocks of

pixels; these blocks are adjacent in the boundary model, but may be well separated in the partition model (recall that we employ a largely random topology). In either case, the assumptions made in statistical testing are generally violated: it may be unreasonable to assume that the gray levels in a block of pixels represent independent and identically distributed observations from some underlying probability distribution (although this is occasionally done). Of course, the size of the blocks relative to the image structures is very important. The blocks may contain hundreds of pixels, but if they are still small relative to the image structures, then the formal assumption will be more nearly satisfied. At any rate, the formal theory is primarily motivational. The distance (2.2) is an effective "measure of homogeneity" which is invariant to pointwise (monotone) data transformations induced by lighting and other factors.

E. Disparity Measures

Sometimes, just gray level histograms are enough for good partitionings, as with the SAR image of water and ice (see Section V). In these cases, disparity is measured as follows. Recall that D_s, $s \in S_L^{(\sigma)}$, is a square block of pixel sites (always 21×21 in the partitioning experiments) centered at s. Let $y(D_s) = \{ y_r: r \in D_s \}$. Given (possibly distant) neighbors $s, t \in S_L^{(\sigma)}$, we define $\Phi_{s,t}$ using the Kolmogorov–Smirnov statistic and a *threshold* c:

$$\Phi_{s,t} = 2\delta_{\{d(y(D_s),y(D_t)) > c\}}(y) - 1.$$

In other words, $\Phi_{s,t}$ is 1 or -1 depending on whether the Kolmogorov–Smirnov statistic is above threshold or not.

Of course, many distinct textures have nearly identical histograms (see [21] for some experiments with partitioning *and* classification of such textures, also in the Bayesian framework). In these cases, discrimination will rely on features, or transformations, that go beyond raw gray levels, involving various *spatial* statistics. We use several of these at once, defining $\Phi_{s,t}$ to be 1 if the Kolmogorov–Smirnov statistic associated with *any* of these transformations exceeds a transformation-specific threshold, and -1 otherwise. The philosophy is simple: if enough transformations are employed, then two distinct textures will differ significantly in at least one of the aspects represented by the transformations. Unfortunately, the implementation of this idea is complicated; more transformations mean more thresholds to adjust, and more possibilities for "false alarms" based on "normal" variations within homogeneous regions.

Proceeding more formally, let Λ denote one such data transformation and put $y' = \Lambda(y)$, the transformed image. In general, y'_t is a function of both y_t *and* the gray-levels in a window centered at $t \in S_l$. For example, y'_t might be the mean, range, or variance of y in a neighborhood of t, or a measure of the local "energy" or "entropy." Or, y'_t might be a *directional residual* defined in (1.1); *isotropic residuals*, in which the pixels $\{ t_j \}$ *surround* t, are also effective. Notice that any Λ given by (1.1) is linear in the sense that if $y_s \rightarrow ay_s + b$, $\forall s \in S_l$ then $y'_s \rightarrow |a| y'_s$, $\forall s \in S_l$, and recall that the Kolmogorov–

Smirnov statistic is *invariant* with respect to such changes. This invariance is shared by other features, such as the mean, variance, and range. It should also be noted that these transforms are decidedly *multivariate*, depending (statistically) on the marginal distributions of the data of at least dimension three. Many approaches to texture analysis are based solely on the one- or two-dimensional marginals, i.e., the gray-level histogram and cooccurrence matrices. We were not able to reliably detect some of the boundaries between the Brodatz microtextures with these standard features. Perhaps the jury is still out.

Given a family of transformations, $\Lambda_1, \Lambda_2, \cdots, \Lambda_m$, we define

$$\Phi_{s,t} = \max_{1 \le i \le m} \left[2\delta_{\{d(y^{(i)}(D_s), y^{(i)}(D_t)) > c_i\}}(y^{(i)}) - 1 \right]$$

$$(2.3)$$

where $y^{(i)} = \Lambda_i(y)$, $1 \le i \le m$, and $y^{(i)}(D_r) = \{ y_s^{(i)}, s \in D_r \}$, $r = s, t$. The thresholds $c_1, \cdots, c_m$ are chosen to limit the percentage of "false alarms" (cases of exceeding threshold for pairs of blocks within the same texture); see Section V.

The disparity measure (2.3) inherits the aforementioned invariance to linear shifts for many transforms, including all "differences of averages." More importantly, perhaps, imagine we are comparing two pairs of image blocks, each pair in a different region of the image. Then, roughly speaking, the two distances are automatically calibrated, regardless of the differing statistical properties of the two regions; i.e., *the disparity measure has the same interpretation anywhere in the image*.

F. Penalties

Recall that $V(x)$ counts the total number of "penalties" associated with $x \in \Omega_L^{(\sigma, P)}$. There are two kinds of "forbidden" configurations that give rise to penalties: roughly, these correspond to very *small regions* and very *narrow regions*. Fix σ and $s \in S_L^{(\sigma)}$. Let E_s be the 5×5 block of sites in $S_L^{(\sigma)}$ centered at s. A configuration x is "small at $s \in S_L^{(\sigma)}$" if fewer than nine labels in $\{ x_t: t \in E_s \}$ agree with x_s. Notice that a right corner at s is allowed; there are exactly nine agreements. The total number of penalties for "small regions" is

$$\sum_{s \in S_L^{(\sigma)}} \delta \left\{ \left(\sum_t \in E_s \delta_{\{x_t = x_s\}} \right) < 9 \right\}.$$

$$(2.4)$$

Obviously, the numbers "5" and "9", as well as other penalty parameters below, are quite arbitrary, and could reasonably be scale-dependent.

As for "thin regions," these are regions that have a horizontal or vertical "neck" that is only one label-site wide (at resolution σ). Let τ_h be a one-site horizontal translation within $S_L^{(\sigma)}$, and let τ_v be the analogous vertical translation. Penalties arise when either $\{ x_{s-\tau_h} \ne x_s$ and $x_s \ne x_{s+\tau_h} \}$ or $\{ x_{s-\tau_v} \ne x_s$ and $x_s \ne x_{s+\tau_v} \}$. The number of "thin-region" penalties is therefore

$$\sum_{s \in S_L^{(\sigma)}} \delta_{\{x_{s-\tau_h} \ne x_s, x_s \ne x_{s+\tau_h}\}} + \delta_{\{x_{s-\tau_v} \ne x_s, x_s \ne x_{s+\tau_v}\}}, \quad (2.5)$$

and $V(x)$ is just the sum of (2.4) and (2.5).

G. Summary

We are given

1) a gray-level image $y = \{ y_{ij} \}$;
2) a resolution $\sigma = 1, 2, \cdots$, and a maximum number of labels P;
3) a disparity measure $\Phi_{s,t}(y)$ for each pair $\langle s, t \rangle_\sigma$ in the sublattice $S_L^{(\sigma)}$;
4) a collection of penalty patterns.

The (MAP) partitioning $\hat{x} = \hat{x}(y)$ is then any solution $x \in \Omega_L^{(\sigma, P)}$ of the constrained optimization

$$\text{minimize}_{x: V(x) = 0} \sum_{\langle s, t \rangle_\sigma} \delta_{\{x_s = x_t\}} \Phi_{s,t}(y)$$

where $V(x)$ is the number of penalties in x.

III. BOUNDARY MODEL

A. Boundary Maps

The pixel lattice is again S_I. Let S_B denote another regular lattice interspersed among the pixels (see Fig. 3) and of dimension $(N - 1) \times (N - 1)$; these are the "boundary sites." We will associate $s = (i, j) \in S_B$ with the pixel $(i, j) \in S_I$ to the upper left of s.

Given y, a gray-level image, we wish to assign values to the boundary variables $x = \{ x_s, s \in S_B \}$, where $x_s = 1$ (resp. 0) indicates the presence (resp. absence) of a boundary at site $s \in S_B$. We have already discussed the corresponding interpretation of the boundary map $x = x(y)$ in terms of physical discontinuities in the underlying three-dimensional scene. We establish a boundary resolution or grid size $\sigma \geq 1$, analogous to the resolution used earlier for the partition model. Let $S_B^{(\sigma)} \subset S_B$ denote the sublattice $\{ (i\sigma + 1, j\sigma + 1): 1 \leq i, j \leq (N - 2)/\sigma \}$. Only the variables x_s, $s \in S_B^{(\sigma)}$, interact directly with the data; the remaining variables x_s, $s \in S_B \setminus S_B^{(\sigma)}$, are determined by those on the "grid" $S_B^{(\sigma)}$. Fig. 4 shows the grids $S_B^{(2)}$ and $S_B^{(3)}$ for $N = 8$; the sites off the grid are denoted by dots. The selection of σ influences the interpretation of x, the computational load, the interaction range at the pixel level, and is related to the role played by the size of the spatial filter in edge detection methods based on differential operators. Finally, let Ω_I and $\Omega_B^{(\sigma)}$ denote the state spaces of intensity arrays and boundary maps respectively; that is,

$$\Omega_I = \{ \{ y_s \}: s \in S_I, 0 \leq y_s \leq 255 \},$$
$$\Omega_B^{(\sigma)} = \{ \{ x_s \}: s \in S_B^{(\sigma)}, x_s \in \{ 0, 1 \} \}.$$

Sometimes, we simply write Ω_B for $\Omega_B^{(1)}$.

B. Boundary-Data Interaction

Let $\langle s, t \rangle_\sigma$, $s, t \in S_B^{(\sigma)}$ denote a *nearest-neighbor* pair relative to the grid. Thus, $s = (i\sigma + 1, j\sigma + 1)$, $t = (k\sigma + 1, l\sigma + 1)$ is such a (horizontal or vertical) pair if either $i = k$ and $j = l \pm 1$, or $j = l$ and $i = k \pm 1$. We identify $\langle s, t \rangle_\sigma$ with the elementary boundary segment consisting of the horizontal or vertical string of $\sigma + 1$ sites (in S_B) including s, t and the $\sigma - 1$ sites "in between."

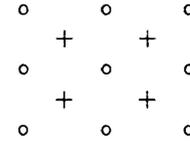The energy function $U(x, y)$ should promote boundary

Fig. 3. Pixel sites ($\bigcirc$) and boundary sites ($+$) for a 3×3 lattice.
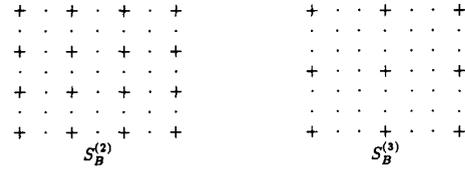

Fig. 4. The boundary grid $S_B^{(\sigma)}$ at resolutions $\sigma = 2$ (left) and $\sigma = 3$ (right).

maps x which are faithful to the data y in the sense that "large" values of $\Delta_{s,t}(y)$ are associated with "on" segments ($x_s x_t = 1$) and "small" values with "off" segments ($x_s x_t = 0$). There are no *a priori* constraints on x at this point; in fact, because of digitization effects, textures, and so-on, the energy U will typically favor maps x with undesirable deadends, multiple representations, high curvature, etc. These will be penalized later on. A simple choice for the x/y interaction is

$$U(x, y) = \sum_{\langle s, t \rangle_\sigma} (1 - x_s x_t) \phi(\Delta_{s,t}(y)) \quad (3.1)$$

where the summation extends over all nearest-neighbor pairs $\langle s, t \rangle_\sigma$; the "weighting function" $\phi(x)$, $x \geq 0$, will be described presently.

The energy in (3.1), which is similar to a "spin-glass" in statistical mechanics, is a variation of the ones we used in our previous work [19], [20]; when $\sigma = 1$, the variable $x_s x_t$ corresponds directly to the "edge" or "line" variables in [20] and [47]. Since y is given, the term $1 - x_s x_t$ can be replaced by $-x_s x_t$ with no change in the resulting boundary interpretation. By contrast, in [20] we were concerned with image *restoration* and regarded *both x and y* as unobservable; the data then consists of some transformation of y, involving for example blur and noise. In that case, or in conceiving U as defining a prior distribution over *both* y and x, the bond between the associated pixels should be broken when the edge is active, i.e., $1 - x_s x_t = 0$. The term $1 - x_s x_t$ is exactly analogous to the "controlled-continuity functions" in [56]. See also [35] for experiments involving simulations from a related Markov random field model; the resulting "fantasies" (y, x) are generated by stochastic relaxation and yield insight into the nature of these layered Markov models.

Returning to (3.1), a little reflection shows that ϕ should be increasing, with $\phi(0) < 0 < \phi(+\infty)$; otherwise, if ϕ were never negative, the energy would always be minimized with $x_s \equiv 1$. The intercept $d^* \doteq \phi^{-1}(0)$ is critical; values of Δ above (resp. below) d^* will promote (resp. inhibit) boundary formation. The influence of this threshold is reduced by choosing $\phi'(d^*) = 0$. We employ

the simple quadratic

$$
\phi(x) = \begin{cases} \left(\dfrac{x - d^*}{d^*}\right)^2, & 0 \leq x \leq d^* \\[2ex] -\left(\dfrac{x - d^*}{\alpha - d^*}\right)^2, & d^* < x. \end{cases} \quad (3.2)
$$

Notice that if we select $\alpha \approx \max \Delta_{s,t}$ then the maximum "penalty" ($\phi(0) = -1$) and "reward" ($\phi(\alpha) = 1$) are balanced.

C. Disparity Measures

We employ one type of measure for depth and shape boundaries and another for the texture experiments. In the former case, the disparity measure involves the (raw) gray-levels only, whereas for texture discrimination we also consider data transforms based on the directional residuals (1.1). Except when $\sigma = 1$, the data sets are compared by the Kolmogorov–Smirnov distance.

The disparity measure should gauge the intensity "flux," $\Delta_{s,t} = \Delta_{s,t}^{(\sigma)}(y) \geq 0$, across $\langle s, t \rangle_\sigma$, i.e., orthogonal to the associated segment. At the highest resolution ($\sigma = 1$), the measure $|y_{s^*} - y_{t^*}|$ (where s^*, t^* are the two pixels associated with the boundary sites s, t—see Fig. 5) can be effective for simple scenes but necessitates a single differential threshold: differences above d^* (resp. below d^*) promote (resp. inhibit) boundary formation. Typically, however, this measure will fluctuate considerably over the image, complicating the selection of d^*. (Such is the case, e.g., for the "cart" scene, see Section V.) Moreover, this measure lacks any invariance properties, as will be explained below. A more effective measure is one of the form

$$
\Delta_{s,t}(y) = \frac{|y_{s^*} - y_{t^*}|}{\gamma + \sum |y_{s_i^*} - y_{t_i^*}|} \quad (3.3)
$$

where the sum extends over parallel edges $\langle s_i, t_i \rangle$ in the immediate vicinity of $\langle s, t \rangle$. Thus the difference $|y_{s^*} - y_{t^*}|$ is "modulated" by adjacent, competing differences. The result is a spatially varying threshold and the distribution of $\Delta_{s,t}(y)$ across the image is less variable than that of $|y_{s^*} - y_{t^*}|$. Choosing $\gamma = \text{const.} \times \overline{\Delta}$, where $\overline{\Delta}$ is the mean (raw) absolute intensity difference over all (vertical and horizontal) bonds, renders $\Delta_{s,t}(y)$ invariant to linear transformations of the data; that is, $\Delta_{s,t}(y) = \Delta_{s,t}(ay + b)$ for any a, b.

At lower resolution, let D_{s^*} and D_{t^*} denote two adjacent blocks of pixels, of equal size and shape. An example is illustrated in Fig. 6 for the case of two square blocks of size $5^2 = 25$ pixels which straddle a vertical boundary segment with $\sigma = 3$. Let $y(D_r) = \{y_s, s \in D_r\}$, $r = s^*$, t^*, be the corresponding gray-levels and set

$$
\Delta_{s,t}(y) = d(y(D_{s^*}), y(D_{t^*})), \quad (3.4)
$$

where d is the Kolmogorov–Smirnov distance discussed in Section II. This is the disparity measure used for the House and Ice Floe scenes (see Section V).

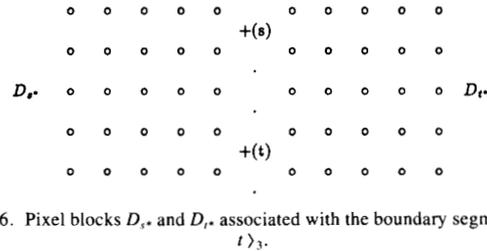Fig. 5. Pixel pairs ($\bigcirc$'s) associated with horizontal and vertical boundary segments.

Fig. 6. Pixel blocks D_{s^*} and D_{t^*} associated with the boundary segment $\langle s, t \rangle_3$.

One difficulty with (3.4) is that the distance between two nonoverlapping histograms is the maximum value, namely 1, regardless of the amount of separation. Thus, two constant regions differing by a single gray-level are as "far apart" as two differing by 255 levels. Thus, it is occasionally necessary to "desensitize" (3.4) for example by "smearing" the data or perhaps adding some noise to it; see Section V.

Raw gray-level data is generally not satisfactory for discriminating textures. Instead, as discussed in Section II, we base the disparity measure on several data transformations, involving higher order spatial statistics, such as the directional residuals defined in (1.1). Given a family $\Lambda_1, \Lambda_2, \cdots, \Lambda_m$ of these transforms (see Section II), a resolution σ, and blocks D_{s^*}, D_{t^*} as above, define

$$
\Delta_{s,t}(y) = \max_{1 \leq i \leq m} \left[c_i^{-1} d\left(y^{(i)}(D_{s^*}), y^{(i)}(D_{t^*}) \right) \right] \quad (3.5)
$$

where $y^{(i)} = \Lambda_i(y)$, $1 \leq i \leq m$, and $y^{(i)}(D_r) = \{y_s^{(i)}, s \in D_r\}$, exactly as in Section II. Then $\Delta_{s,t}(y) > d^*$ (and, hence, $\phi(\Delta_{s,t}(y)) > 0$) if and only if $d(y^{(i)}(D_{s^*}), y^{(i)}(D_{t^*})) \geq d^* c_i$ for some transform i. The thresholds $c_1, \cdots, c_m$ are again chosen to limit "false alarms." Finally, we note that (3.5) has the same desirable invariance properties as the measure constructed for partitioning (Section II).

D. Penalties

$V(x)$ again denotes the total number of "penalties" associated with $x \in \Omega_B^{(\sigma)}$. These penalties are simply local binary patterns over subsets of $S_B^{(\sigma)}$. Fig. 7 illustrates a family of four such patterns; they can be associated with any resolution σ by the obvious scaling. These correspond, respectively, to an isolated or abandoned segment, sharp turn, quadruple junction, and "small" structure. Depending on σ, the pixel resolution, and scene information, we may or may not wish to include the latter three. For example, including the last one with $\sigma = 6$ would prohibit detection of a square structure of pixel size 6×6.
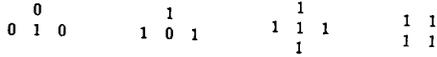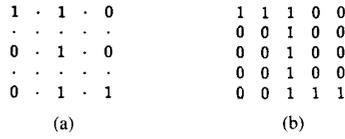
```
    0                1              1            1 1
  0 1 0            1 0 1        1 1 1 1        1 1
                                    1
```

Fig. 7. Forbidden boundary patterns.

```
  1 · 1 · 0            1 1 1 0 0
  · · · · ·            0 0 1 0 0
  0 · 1 · 0            0 0 1 0 0
  · · · · ·            0 0 1 0 0
  0 · 1 · 1            0 0 1 1 1
     (a)                  (b)
```

Fig. 8. (a) A boundary configuration at resolution $\sigma = 2$. (b) Completion of (a) from $\sigma = 2$ to $\sigma = 1$.

Finally, there is a natural extension from $\Omega_B^{(\sigma)}$ to Ω_B which is useful for display and evaluation. Given $x \in \Omega_B^{(\sigma)}$, we define $x_s = 0$ for sites $s \in S_B \setminus S_B^{(\sigma)}$ lying on a row or column disjoint from $S_B^{(\sigma)}$, and $x_s = x_{t_1} x_{t_2}$ if s lies on a segment $\langle t_1, t_2 \rangle_\sigma$. Thus, for example, the state $x \in \Omega_B^{(2)}$ in Fig. 8(a) is identified with the state $x \in \Omega_B$ in 8(b).

E. Summary

We are given
1) a gray-level image $y = \{ y_{ij} \}$;
2) a resolution level $\sigma = 1, 2, \cdots$;
3) a disparity measure $\Phi_{s,t}(y) = \phi(\Delta_{s,t}(y))$ for each neighbor pair $\langle s, t \rangle_\sigma$ in the sublattice $S_B^{(\sigma)}$;
4) a collection of penalty patterns.

The (MAP) boundary estimate $\hat{x} = \hat{x}(y)$ is any solution $x \in \Omega_B^{(\sigma)}$ of the constrained optimization

$$\text{minimize}_{x: V(x) = 0} \sum_{\langle s,t \rangle_\sigma} (1 - x_s x_t) \phi(\Delta_{s,t}(y))$$

where $V(x)$ is the number of penalties in x.

IV. ALGORITHMS

We begin with an abstract formulation of the optimization and sampling problems outlined in Sections I–III. Recall that our search algorithms are motivated by the *stochastic* formulation of the problem, in particular the role of the *posterior* distribution $\Pi(x \mid y)$. However, y is fixed by observation, and can be ignored in this discussion of computational issues. Thus, we are given two functions U and V on a space of configurations $\Omega = \{(x_{s_1}, x_{s_2}, \cdots, x_{s_M}): x_{s_i} \in \Lambda, 1 \le i \le M \}$ where Λ is finite, M is very large, and $S = \{ s_1, s_2, \cdots, s_M \}$ is a collection of "sites," typically a 2-D lattice. Write $x = (x_{s_1}, \cdots, x_{s_M})$ for an element of Ω, and let

$$\Omega^* = \left\{ x: V(x) = v \right\} \quad v = \min_x V(x)$$

$$\Pi^*(x) = \delta_{\Omega^*}(x) \frac{\exp \left\{ -U(x) \right\}}{\sum_{x' \in \Omega^*} \exp \left\{ -U(x') \right\}}.$$

We wish to solve the constrained optimization problem minimize $\{ U(x): V(x) = v \}$ or to sample from the Gibbs distribution Π^*. (Recall that sampling at low temperature allows us to approximate the global minimum or to entertain other estimates, for instance the posterior mean.)

We have studied [20] Monte Carlo site-replacement algorithms for the unconstrained versions of these problems: stochastic relaxation (SR) for sampling, and stochastic relaxation with simulated annealing (SA) for optimization. SA was devised in [8] and [39] for minimizing a "cost functional" U (e.g., the tour length for the traveling salesman problem) by regarding U as the energy of a physical system and simulating the dynamics of chemical annealing. The effect is to drive the system towards the "ground states," i.e., the minimizers of U. This is accomplished by applying the Metropolis (relaxation) algorithm to the Boltzmann distribution

$$\frac{e^{-U(x)/t}}{\sum_{x'} e^{-U(x')/t}} \tag{4.1}$$

at successively lower values of the "temperature" t.

We presented two theorems in [20]: one for generating a sequence $\{ X(k) \}$ which converges in distribution to (4.1) for $t = 1$ (SR), and one for generating a sequence $\{ X(k) \}$ having asymptotic distribution the uniform measure over $\Omega_0 = \{ x \in \Omega: U(x) = u \}, u = \min_x U(x)$, (SR with SA). The essence of the latter algorithm is a "cooling schedule" $t = t_1, t_2, \cdots$ for establishing convergence. SA has been extensively studied recently [4], [10], [23], [24], [30], [32], [34], [55]; see also the comprehensive review [1] and the references therein.

Results concerning *constrained* SR and SA are reported in [18], which was motivated by a desire to find a theoretical foundation for the algorithms used here. We have deviated from the instructions in [18], with regard to the cooling schedule, but at least we know that the algorithms represent approximations to rigorous results.

Both algorithms produce a *Markov chain* on Ω by sampling from the low-order, marginal conditional distributions of the free Gibbs measures

$$\Pi(x; t, \lambda) = \frac{\exp \left\{ -t^{-1} \left(U(x) + \lambda V(x) \right) \right\}}{\sum_{x'} \exp \left\{ -t^{-1} \left(U(x') + \lambda V(x') \right) \right\}}.$$

It is easy to check that

$$\lim_{\lambda \to \infty} \Pi(x; 1, \lambda) = \Pi^*(x) \tag{4.2}$$

and that

$$\lim_{\substack{\lambda \to \infty \\ t \to 0}} \Pi(x; t, \lambda) = \begin{cases} \left| \Omega_0^* \right|^{-1}, & x \in \Omega_0^* \\ 0, & \text{otherwise} \end{cases} \tag{4.3}$$

where $\Omega_0^* = \{ \omega \in \Omega^*: U(\omega) = \xi \}, \xi = \min_{x \in \Omega^*} U(x)$. Let Π_0 denote the uniform measure in (4.3). Sampling directly from $\Pi(x; t, \lambda)$ is impossible due to the size of Ω; otherwise just use (4.2) and (4.3) to generate a sequence of random variables $X(k), k = 1, 2, \cdots$, with values in Ω, and limiting distribution either Π^* or Π_0. However, we can evaluate *ratios* $\Pi(x; t, \lambda)/\Pi(z; t, \lambda)$, $x, z \in \Omega$, and hence *conditional* probabilities. The price

for *indirect* sampling is that we must restrict the rate of growth of λ and the rate of decrease of t.

Fix two sequences $\{t_k\}$, $\{\lambda_k\}$, a "site visitation" schedule $\{A_k\}$, $A_k \subset S$, and let $\Pi_k(x) = \Pi(x; t_k, \lambda_k)$. The set A_k is the cluster of sites to be updated at "time" k; the "centers" of the clusters are addressed in a raster scan. In our experiments we take either $|A_k| \equiv 1$ or $|A_k| \equiv 5$, in which case the A_k's are of the form $\{(i, j), (i + 1, j), (i - 1, j), (i, j + 1), (i, j - 1)\}$.

Define a nonhomogeneous Markov chain $\{X(k), k = 0, 1, 2, \cdots\}$ on Ω as follows. Put $X(0) = \eta$ arbitrarily. Given $X(k) = (X_{s_1}(k), \cdots, X_{s_M}(k))$, define $X_s(k + 1) = X_s(k)$ for $s \notin A_{k+1}$ and let $\{X_s(k + 1): s \in A_{k+1}\}$ be a (multivariate) sample from the conditional probability distribution $\Pi_{k+1}(x_s, s \in A_{k+1} | x_s = X_s(k), s \notin A_{k+1})$. Then, under suitable conditions on $\{t_k\}$ and $\{\lambda_k\}$, either

$$\lim_{k \to \infty} P(X(k) = x | X(0) = \eta) = \Pi^*(x)$$

or the limit is $\Pi_0(x)$. The condition in the former case (constrained SR) is that $t_k \equiv 1$, $\lambda_k \nearrow \infty$, and $\lambda_k \leq \text{const.} \cdot \log k$. The condition for convergence to Π_0 (constrained SA) is that $t_k \searrow 0$, $\lambda_k \nearrow \infty$ and $t_k^{-1}\lambda_k \leq \text{const.} \cdot \log k$. The algorithm yields a solution to the constrained optimization problem (1.2) in the sense that the asymptotic distribution of $X(k)$ is uniform over the solution set: if the solution is unique, i.e., $\Omega_0^* = \{x_0\}$, then $X(k) \to x_0$ in probability. See [18] for proofs.

A. Approximations

The logarithmic rate is certainly slow. Still, we often adhere to it for ordinary annealing; others [36], [47] have as well. We refer the reader to [35] for some interesting comparisons between schedules. It is commonplace to find linear ($t_k = t_0 - ak$) and exponential ($t_k = (1 - \gamma)^k t_0$, γ small) schedules; here k refers to the number of sweeps or iterations of S; in our experiments $S = S_L^{(\sigma)}$ or $S_B^{(\sigma)}$.

We now describe several protocols used in our experiments. One variant we do *not* use is to fix $\lambda_k \equiv \lambda$ very large and do ordinary annealing, which might appear sensible since the solutions to min $\{U(x): V(x) = 0\}$ coincide with those of min $\{U(x) + \lambda V(x)\}$ for all λ sufficiently large (due to the fact that Ω is finite). However this is not practical: unless t_o is very large and t_k is reduced very slowly, the system immediately gets stuck in local energy minima of $U + \lambda V$ which are basically independent of the data, although faithful to the constraints. *It is better to begin with states faithful to the data and slowly impose the constraints*, a standard technique in conventional optimization.

One variation of constrained SR that has been effective is "low-temperature sampling": fix $t_k \equiv \epsilon$ (small) and let $\lambda_k \nearrow \infty$. The idea is to reach a likely state of the posterior distribution $\Pi(x | y)$. In practice, we allow λ_k to grow *linearly*; the details are in Section V.

Another variation is the analog for constrained relaxation of "zero-temperature" sampling, which has been extensively studied by Besag [2] under the name ICM (for "iterated conditional modes"); see also [11], [14], and [22]. Without constraints, this algorithm, which is deterministic, results in a sequence of states $X(k)$ which monotonically decrease the global energy, i.e., increase the posterior likelihood. The constrained version operates as follows. Recall that when the set of sites A_{k+1} is visited for updating, we defined $X(k + 1)$ by replacing the coordinates of $X(k)$ in A_{k+1} by a *sample* drawn from the conditional distribution of Π_{k+1} on $\{x_s, s \in A_{k+1}\}$ given the values $\{x_s = X_s(k), s \notin A_{k+1}\}$. Suppose we replace the sample with the *mode*, i.e., the most likely vector $\{x_s, s \in A_{k+1}\}$ conditional upon $\{x_s = X_s(k), s \notin A_{k+1}\}$. In essence, we fix $t_k \equiv 0$. This generates a *deterministic* sequence $X(k), k = 0, 1, 2, \cdots$ depending only on $X(0)$, Π_k, and $\{A_k\}$. (Notice that the mode is unaffected by t_k since it corresponds to the minimum of $U(x) + \lambda_k V(x)$.) Then, *during the kth sweep*, with $\lambda \equiv \lambda_k$, the energy $U + \lambda_k V$ is successively reduced, just as in ICM where $\lambda_k \equiv 0$. Of course since there is no *fixed* (reference) energy, the algorithm cannot be conceived as one of iterative improvement. Several experiments were run with both the stochastic and deterministic algorithms; see Section V.

B. Summary

Each experiment used one of the following two variations on constrained relaxation (see Section V for details):

- Choose a label resolution σ and an associated pixel block size.
- Select features.
- Choose a site visitation schedule $\{A_k\}$, $A_k \subset S$ $k = 1, 2, \cdots$.
- Fix a *linear* growth schedule for $\{\lambda_k\}$.
- Choose $X(0)$, arbitrary.
- Set $k = 0$.

Deterministic Algorithm:

0) Set $t_k = 1$, $k = 1, 2, \cdots$.

1) Define $X_s(k + 1) = X_s(k)$, $s \notin A_{k+1}$ and define $\{X_s(k + 1): s \in A_{k+1}\}$ to be the (multivariate) mode of

$$\Pi_{k+1}(x_s, s \in A_{k+1} | x_s = X_s(k), s \notin A_{k+1}).$$

2) $k = k + 1$.

3) Go to 1.

Stochastic Algorithm:

0) Set $t_k = \epsilon$, $k = 1, 2, \cdots$ (ϵ "small").

1) Define $X_s(k + 1) = X_s(k)$, $s \notin A_{k+1}$ and define $\{X_s(k + 1): s \in A_{k+1}\}$ to be a (multivariate) sample from

$$\Pi_{k+1}(x_s, s \in A_{k+1} | x_s = X_s(k), s \notin A_{k+1}).$$

2) $k = k + 1$.

3) Go to 1.

Usually, but not always, the deterministic algorithm was sufficient (again, see Section V).

V. Experiments[2]

A. Partition Model

There are three experiments: an L-band synthetic aperture radar (SAR) image[3] of ice floes in the ocean (Fig. 9), a texture mosaic constructed from the Brodatz album [6] (Fig. 10), and another mosaic from pieces of rug, plastic, and cloth (Fig. 11).

Processing: In each experiment the partitioning was randomly initiated; the labels, $x_s s \in S_L^{(\sigma)}$, were chosen independently and uniformly from $0, 1, \cdots, P - 1$. Thereafter, label sites were visited and updated one at a time, by a "raster scan" sweep through the label array. MAP partitionings were approximated by "zero-temperature" sampling (see Section IV), with $\lambda = \lambda_k$ increasing with the number of sweeps. Specifically, λ was held at 0 through the first 10 sweeps, and thereafter was raised by 1 every 5 sweeps: $\lambda_k = 0, k = 1, \cdots 10; \lambda_k = 1, k = 11, \cdots 15; \lambda_k = 2, k = 16, \cdots 20;$ etc. Most probably, λ could have been increased more rapidly, perhaps with every sweep, without substantially changing the results, but this was not systematically investigated. For the three experiments shown in Figs. 9, 10, and 11, between 15 and 50 sweeps sufficed to bring the changes in labels to a halt; see below for more details. Recall that zero-temperature sampling corresponds to choosing the conditional mode. Occasionally there are ties, and these were resolved by choosing randomly, and uniformly, from the collection of modes.

As a general rule, results were less reliable at higher resolutions (lower σ's) and when more labels were allowed (higher values of P). In these cases, repeated experiments, with different initializations, often produced different results. With P too large, homogeneous regions were frequently subdivided, being assigned two or three labels. With σ too small, the tendency was to mislabel small patches within a given texture. It is likely that many of these mistakes correspond to local minima; perhaps some could be corrected by following a proper annealing schedule (see Section IV), and by more careful choices of thresholds (see below). Here again, definitive experiments have not been done.

Measures of Disparity: Recall that the disparity measure is derived from the Kolmogorov–Smirnov distance between blocks of pixel data under various transformations, as defined in Section II, equation (2.3). For the SAR image, good partitionings were obtained using only the raw data: $m = 1$ and $y_{(1)}$ is just y in (2.3). Evidently, gray-level distributions are enough to segment the water and ice "textures," at least when supplemented by the "prior constraints" embodied in the penalty term $V(x)$.

The texture collages in Figs. 10 and 11 are harder. We used four data transformations in addition to the raw pixel data. Hence, for these experiments $m = 5, y^{(1)} = y$, and $y^{(2)}, \cdots y^{(5)}$ are based on various transforms. In particular,

$y_s^{(2)}$ measures the intensity *range* in the 7×7 pixel block V_s centered at s:

$$y_s^{(2)} = \max_{t \in V_s} y_t - \min_{t \in V_s} y_t;$$

$y_s^{(3)}$ is the "residual" (1.1) obtained by comparing y_s to the 24 "boundary pixels" (∂V_s) of V_s (i.e., all pixels on the perimeter of the 7×7 block):

$$y_s^{(3)} = \left| y_s - \frac{1}{24} \sum_{t \in \partial V_s} y_t \right|;$$

and $y^{(4)}$ and $y^{(5)}$ are horizontal and vertical "directional residuals":

$$y_s^{(4)} = \left| y_s - \tfrac{1}{2} (y_{s+(1,0)} + y_{s+(-1,0)}) \right|$$

$$y_s^{(5)} = \left| y_s - \tfrac{1}{2} (y_{s+(0,1)} + y_{s+(0,-1)}) \right|.$$

Parameter Selection: The resolution (σ) was 7 for the SAR picture (Fig. 9); 15 for the Brodatz collage; and 13 for the pieces of rug, plastic, and cloth. These numbers were chosen more or less ad hoc, but are small enough to capture the important detail of the respective pictures while not so small as to incur the degraded performance, seen at higher resolutions and mentioned earlier.

The number of allowed labels is also important; recall that too many usually results in over-segmentation. This was actually used to advantage in the SAR experiment (Fig. 9), where there are evidently two varieties of ice. The best segmentations were obtained by allowing three labels. Invariably, two would be assigned to the ice, and one to the water. Using just two labels led to mistakes within the ice regions, although there was little experimentation with the Kolmogorov–Smirnov threshold, and no attempt was made with the data transforms ($m > 1$) used for the collages. In the other experiments, the number of labels was set to the number of texture species in the scene.

The most important parameters were the thresholds, $\{c_i\}\ 1 \leq i \leq m$, associated with the Kolmogorov–Smirnov statistics [see (2.3)]. For the SAR experiment, $m = 1$, and the threshold was guessed, *a priori*; it was found that small changes are reflected only in the lesser details of the segmentation. For the collages ($m = 5$), the thresholds were chosen by examining histograms of Kolmogorov–Smirnov distances for block pairs *within homogeneous samples of the textures*. Thresholds were set so that no more than 3 or 4% of these intraregion distances would be above threshold (a "false alarm"). Of course, we would have preferred to find more or less universal thresholds, one for each data transform, but this may not be possible. Conceivably, with enough of the "right"
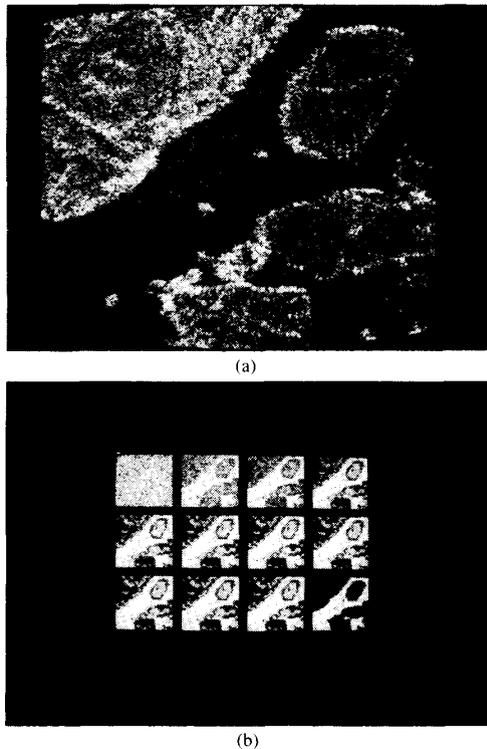
(a)

(b)

Fig. 9. (a) Synthetic aperture radar image of oceanic ice floes. (b) Evolution of the label states from random initialization (upper left) to final partition (lower right).

(a)

(b)

Fig. 10. (a) Collage of five Brodatz textures. (b) Evolution of the labels, upper left to lower right.

transforms, one could set conservative (high) and nearly universal thresholds, and be assured that visibly distinct textures would be segmented with respect to at least one of the transforms. Recall that the disparity measure (2.3) is constructed to signal ''different'' when the distance between blocks, with respect to *any* of the transforms, exceeds threshold.

Fig. 9 (SAR): As mentioned earlier, three labels were used, with the expectation that the ice would segment into two regions (basically, dark and light). The resolution was $\sigma = 7$, and the Kolmogorov–Smirnov statistic was computed only on the raw data, so $m = 1$. The threshold was $c_1 = 0.15$. The original image is 512 × 512 (the pixel resolution is about 4 by 4 m), but to avoid special treatment of the boundary, only the 462 × 462 piece shown in Fig. 9(a) was processed. The label lattice $S_L^{(\sigma)}$ is 64 × 64. Fig. 9(b) shows the evolution of the partitioning during the relaxation. For display, gray levels were arbitrarily assigned to the labels. The upper left panel is the random starting configuration. In successive panels are the states of the labels after each five iterations (full sweeps). In the bottom right panel, the two labels associated with ice are combined, ''by hand.''

Fig. 10 (Brodatz Textures): The Kolmogorov–Smirnov thresholds were $c_1 = 0.40$, $c_2 = 0.53$, $c_3 = 0.26$, $c_4 = 0.28$, and $c_5 = 0.19$, corresponding to the transforms $y^{(1)}$, $\cdots y^{(5)}$ discussed above. A 246 × 246 piece of the orig-

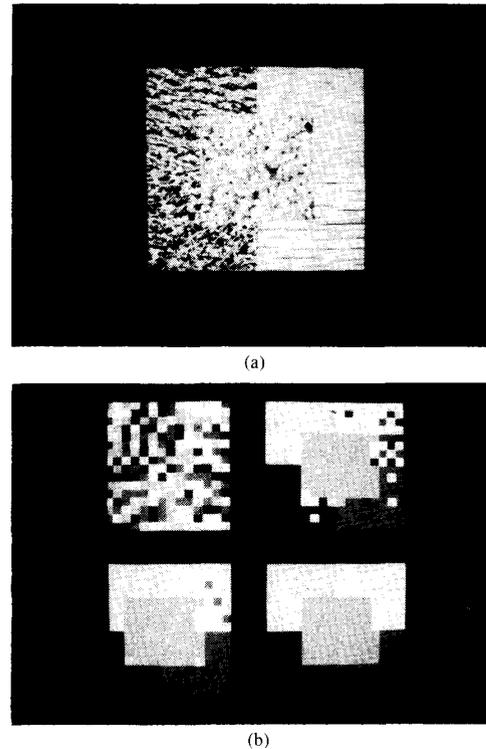inal 256 × 256 image was processed, and is shown in Fig. 10(a). Leather and water are on top, grass and wood on the bottom, and sand is in the middle. The resolution was $\sigma = 15$, which resulted in a 16 × 16 label lattice $S_L^{(\sigma)}$. Fig. 10(b) shows the random starting configuration (upper left panel), the configuration after 5 iterations (upper right panel), after 10 iterations (lower left panel), and after 15 iterations (lower right panel), by which point the labels had stopped changing.

Fig. 11 (Rug, Plastic, Cloth): The 216 × 216 image in Fig. 11(a) was partitioned at resolution $\sigma = 13$, with a 16 × 16 label lattice. The Kolmogorov–Smirnov thresholds were $c_1 = 0.90$, $c_2 = 0.49$, $c_3 = 0.20$, $c_4 = 0.11$, and $c_5 = 0.12$, corresponding to the same data transforms used for the Brodatz textures (Fig. 10). The experiment makes apparent a *hazard* of long range bonds: the gradual but marked lighting variation across the top of the image produces a large Kolmogorov–Smirnov distance when raw pixel blocks from the left and right sides are compared. This makes it necessary to essentially ignore the raw data Kolmogorov–Smirnov statistic, and base the partitioning on the four data transformations; hence the threshold $c_1 = 0.9$. The transformed data are far less sensitive to lighting gradients. Fig. 11(b) displays the evolution of the partitioning during relaxation. The layout is the same one used in the previous figures, showing every 5 iterations, except that there are 10 iterations between the final two
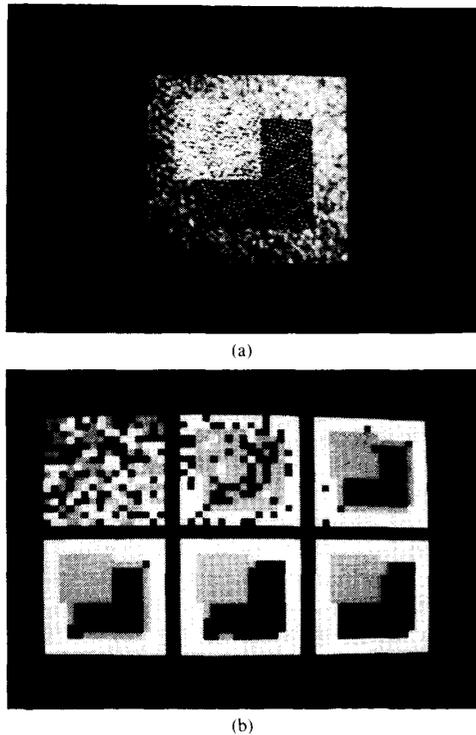
(a)

(b)

Fig. 11. (a) Texture collage: rug, plastic, cloth. (b) Evolution of the labels, upper left to lower right.

panels. The lower right panel is the partitioning after the 30th sweep, by which time the pattern was "frozen."

B. Boundary Model

There are five test images: one made indoors from tinker toys ("cart"), an outdoor scene of a house, another of ice floes in the ocean (the same SAR image used above), and two texture mosaics constructed from the Brodatz album.

Processing: All the experiments were performed with the same site-visitation schedule. Given the resolution σ, which varies among experiments, the sites of the sublattice $S_B^{(\sigma)}$ were addressed in a raster-scan and *five sites* were simultaneously updated. Specifically, at each visit to the site $(i\sigma + 1, j\sigma + 1)$, the values of the boundary process at this site and its four nearest neighbors, $\{((i \pm 1)\sigma + 1, (j \pm 1)\sigma + 1)\}$, were replaced based on the conditional distribution of these five boundary variables given the variables at the other sites and the data y. Of course this distribution is concentrated on the $2^5 = 32$ possible configurations for these five variables.

Two update mechanisms were employed: stochastic relaxation and the "zero-temperature," deterministic variation discussed earlier. In the former case, the updated binary quintuple is a sample from the aforementioned conditional distribution, which varies depending on the penalty weight λ_k for the kth sweep of the lattice $S_B^{(\sigma)}$. Of course "0-temperature" refers to replacing the sample by

the conditional model. In both cases we let λ_k grow *linearly*.

Stochastic relaxation at low temperature is more effective than at zero temperature (essentially iterative improvement). However, deterministic relaxation sufficed for all but two scenes, the ice floes and the four texture collage; these results could not be duplicated with deterministic relaxation. In one case, we present both results for comparison.

Generally, deterministic relaxation stabilizes in 5-10 sweeps whereas stochastic relaxation requires more sweeps, perhaps 20-60. We provide several pictures showing the evolution of the algorithm.

Penalties: All the experiments were conducted with the same forbidden patterns, namely those in Fig. 12, with the exception of the house scene, for which the last pattern was omitted. (At the resolution used for the house, namely $\sigma = 3$, the inclusion of that pattern would inhibit the formation of structures at the scale of six pixels; many such nontrivial structures appear in that scene.) Thus, the penalty function $V(x)$ records a unit penalty for each occurrence in the boundary map $x = \{x_s, s \in S_B^{(\sigma)}\}$ of any of the five patterns depicted in Fig. 12. It is interesting to note that in no case was the final labeling completely free of penalties, i.e., $V(\hat{x}) = 0$. Perhaps this could be achieved with a proper annealing schedule, or with updates of more than five sites.

Measures of Disparity: All the experiments are based on instances of the measures (3.3)-(3.5) described in Section III.

1) For the first experiment, the cart scene, the boundary resolution is $\sigma = 1$ and we employed the measure given in (3.3) with $\gamma = 10\overline{\Delta}$ and the raw difference $|y_{s^*} - y_{t^*}|$ modulated by the four nearest differences of the same orientation as $\langle s^*, t^* \rangle$. Thus, for the horizontal pair $\langle s, t \rangle$ of adjacent boundary sites,

$$\Delta_{s,t}(y) = \frac{|y_{s^*} - y_{t^*}|}{10\overline{\Delta} + \sum_{k = \pm 1, \pm 2} |y_{i,j+k} - y_{i+1,j+k}|}$$

where $s^* = (i, j)$, $t^* = (i + 1, j)$, and $\overline{\Delta}$ is the mean absolute intensity difference over the image. The utility seems largely impervious to the choice of the scaling constant (here $= 10$) for the mean as well as to the range of the modulation.

2) We used the Kolmogorov-Smirnov measure (3.4) for both the house and ice floes scenes. For the house, we chose $\sigma = 3$ and blocks of size 25; the setup is depicted in Fig. 6. Due to the uniform character of the background (e.g., the sky) the distance (3.4) was computed based on the transformed data $y'_{ij} = y_{ij} + \eta_{ij}$, where $\{\eta_{ij}\}$ are independent variables, and distributed with a triangular density, specifically that of $10(u_1 + u_2)$, u_1, u_2 uniform (and independent) on $[0, 1]$.

The boundary resolution for the radar experiment is $\sigma = 8$, reflecting the larger important structures there; the image is 512×512. The dynamic range is very narrow

```
        1           1               1       1
  0                 1 0 1       1 1 1     1 1           1     1
0 1 0     1 0 1       1         1 1 1     1 1
        1           1                           1     1
```
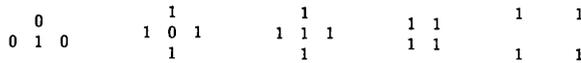
Fig. 12. Forbidden patterns for the boundary experiments.

and the difference between the dark water and somewhat less dark ice is essentially one of texture, due in part to the customary speckle noise. In particular, the ice cannot be well-differentiated from the water based on shading alone. The disparity measure is (3.4), applied to the raw image data over 24 × 24 blocks. The problem encountered in the house scene is actually *alleviated* by the speckle.

3) The texture mosaic experiments are based on the measure (3.5) for a particular family $\Lambda_1, \cdots, \Lambda_5$ of five data transformations or "features." In each case, the resolution is $\sigma = 5$ and block size is 21 × 21. Recall that these five features are combined into a single measure of change according to the formula in (3.5). The transformations used are the *range*

$$z_s^{(1)} = \max_{t \in V_s} y_t - \min_{t \in V_s} y_t$$

over a 7 × 7 window V_s centered at pixel s, and the four *directional residuals*

$$z_s^{(2)} = \left| y_s - \tfrac{1}{2} \left(y_{s+(0,1)} + y_{s+(0,-1)} \right) \right|$$

$$z_s^{(3)} = \left| y_s - \tfrac{1}{2} \left(y_{s+(1,0)} + y_{s+(-1,0)} \right) \right|$$

$$z_s^{(4)} = \left| y_s - \tfrac{1}{2} \left(y_{s+(1,1)} + y_{s+(-1,-1)} \right) \right|$$

$$z_s^{(5)} = \left| y_s - \tfrac{1}{2} \left(y_{s+(-1,1)} + y_{s+(1,-1)} \right) \right|.$$

These residuals were then uniformly averaged over V_s, yielding the final features $y^{(1)}, \cdots y^{(5)}$.

It is instructive to compare the Kolmogorov–Smirnov differences for the raw and transformed data over these texture mosaics. Typically, if one looks at the resulting two histograms of differences for a give transform, on finds that, whereas the raw (Kolmogorov–Smirnov) differences are actually larger at the texture borders, the *transitions* between the borders and interiors are sharper for the transformed data. Detecting the boundaries with the raw data necessitates an unacceptable number of "false alarms" in the sense of interior "microedges."

Finally, the values of the constants $c_1, \cdots, c_5$ used in the construction of $\Delta_{s,t}$ [see (3.5)] are selected by restricting the percentage of false alarms. The details are given in the following section.

Parameter Selection: Recall that the total change across the boundary segment $\langle s, t \rangle$ is measured by $\phi(\Delta_{s,t}(y))$, where ϕ is given in (3.2). Given Δ, there are two parameters to choose: a normalizing constant α and the intercept $d^* = \phi^{-1}(0)$.

For the object boundary experiments, namely the cart, house, and ice floes, the parameters α and d^* were chosen as follows. Find the *mean* disparity over all (vertical and horizontal) values of $\Delta_{s,t}$ for relevant bonds $\langle s, t \rangle$; take α equal to the 99th percentile of those above the mean and d^* equal to the 70th percentile of those above the mean. This yields the values $\alpha = 150$, $d^* = 42$ for the cart scene; recall that for this experiment, both the grid and block sizes are unity. For the house scene ($\sigma = 3$) the Kolmogorov–Smirnov statistics were computed over 5 × 5 blocks, and the resulting parameters are then $\alpha = 1$ and $d^* = 0.7$. (The number of distances at (the maximum) value $\Delta = 1$ was considerable.) Finally, for the ice floes, the recipe above yielded $\alpha = 0.33$, $d^* = 0.13$.

Turning to the experiments with texture mosaics, let $c_{i,k}$ denote the normalizing constant in (3.5) for feature i, $1 \le i \le 5$, and texture k, $1 \le k \le K$, where K is the number of textures in the mosaic. For each feature i and texture type k, we computed the histogram of the (combined vertical and horizontal) Kolmogorov–Smirnov distances and selected $c_{i,k} = 100(1 - \gamma)$ percentile of that histogram. Specifically, we took $\gamma = 0.01$ for the two Brodatz collages. (Other experiments indicated that *any* (small) value of γ will suffice, say $0 \le \gamma \le 0.03$.) Thus, $100(1 - \gamma)$ % of the distances $d(y^{(i)}(D_1), y^{(i)}(D_2))$ are below $c_{i,k}$ *within* each texture type k. Now set $c_i = \max_{1 \le k \le K} c_{i,k}$, ensuring that at most $k\gamma(100)$% of the *interior* differences $\Delta_{s,t}(y)$ within the entire collage will exceed the threshold $d^* = 1$. Finally, we put $\alpha = 2$.

Fig. 13 (Cart Scene): Sixty sweeps of stochastic relaxation were run with $t_k \doteq 0.05$ and $\lambda_k \nearrow 3$. Actually, all the boundaries were "in place" after about 10 sweeps, as illustrated in Fig. 13(b), which shows every third sweep up to the 46th. The image is 110 × 110. Not shown is a run with the deterministic algorithm; the results are virtually indistinguishable.

Fig. 14 (House Scene): This 256 × 256 monochrome image was supplied to us by the Visions group at the University of Massachusetts. The update is by deterministic relaxation with λ_k increasing linearly from $\lambda_0 = 0$ to $\lambda_{10} = 2$.

Fig. 15 (Ice Floes): The image in Fig. 15(a) is 512 × 512. We did 60 sweeps of stochastic relaxation with $t_k \doteq = 0.1$ and $\lambda_k \nearrow 2$. Fig. 15(b) shows 16 "snapshots"— every third sweep as in Fig. 13, as well as the final (60th) sweep.

Fig. 16 (Brodatz Collage 1): The collage is composed of nine Brodatz textures [Fig. 16(a)]: leather, grass, and pigskin (top row), raffia, wool, and straw (middle row), and water, wood, and sand (bottom row). Two of the textures, leather and water, are repeated in the two circles. The image size is 384 × 384, the individual textures all being 128 × 128. We show the results [Fig. 16(b)] of both the deterministic (left) and stochastic (right) algorithms; they are roughly comparable. Other false alarm rates ($\gamma = 0.005, 0.02$, and 0.03) yield the same overall quality.

Fig. 17 (Brodatz Collage 2): There are four textures
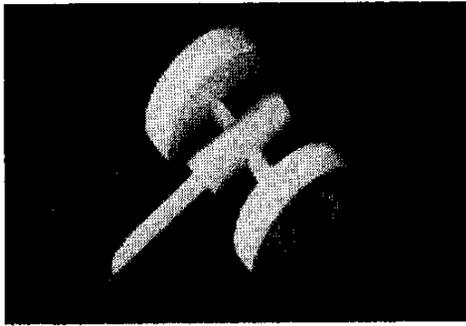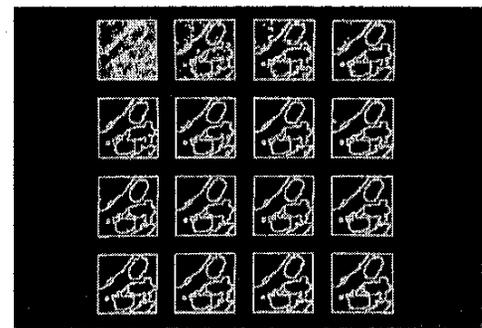
(a)
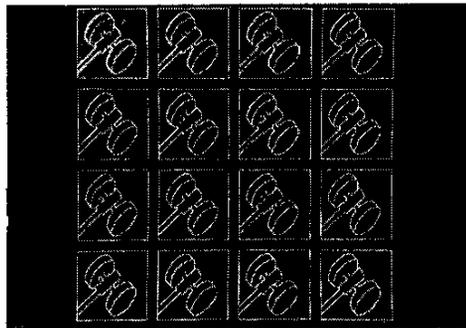
(b)

Fig. 13. (a) Image of a tinker toy cart. (b) Evolution of the boundaries (upper left to lower right) with multiple-site stochastic relaxation.
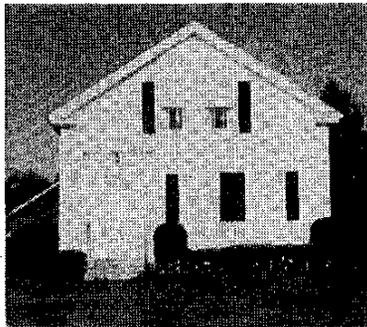

(a)

(b)

Fig. 15. (a) Ice floes image [see Fig. 9(a)]. (b) Evolution of the boundaries (at resolution $\sigma = 8$) during stochastic relaxation.
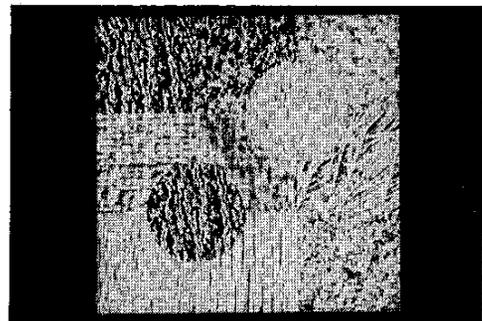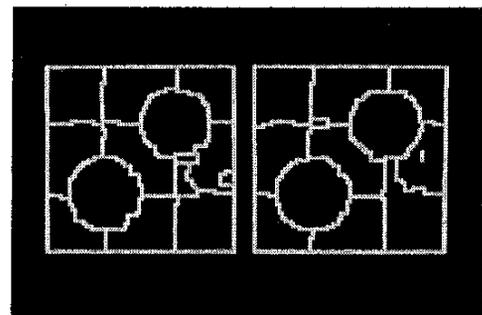

(a)

(b)

Fig. 14. (a) House scene. (b) Final boundary placements at resolution σ = 3 with deterministic relaxation.

(a)

(b)

Fig. 16. (a) Collage of nine Brodatz textures. (b) Final boundary placements with deterministic relaxation (left) and stochastic relaxation (right).
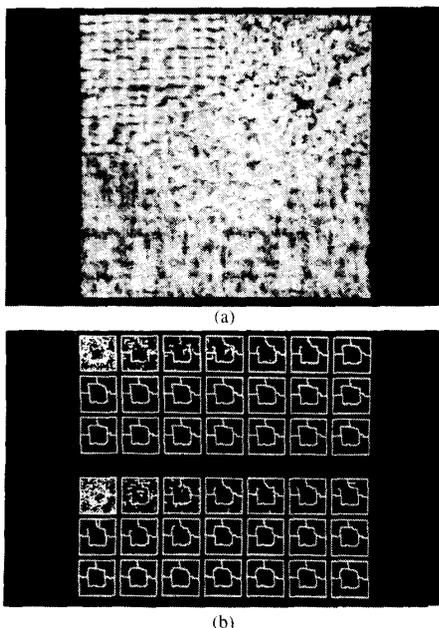
(a)

(b)

Fig. 17. (a) Collage of four Brodatz textures. (b) Two runs of stochastic relaxation showing every third sweep.

[Fig. 17(a)]: raffia (upper left), sand (upper right), wool (bottom), and pigskin (center). Two runs (different seeds) are shown [Fig. 17(b)] individual frames representing every third sweep.

VI. GENERALIZATIONS

These models may be extended in many directions, or combined into a single model. What follows is a brief description of three such generalizations.

A. Other Boundary Primitives

The elementary boundary unit or primitive is a horizontal or vertical segment whose length is resolution-dependent, namely $\sigma + 1$ in pixel units. As a result, a discontinuity running at $45°$ is detected and localized with less reliability than one at $0°$ or $90°$, where the disparity measure has maximum sensitivity. An obvious remedy is to replace the pairs $\langle s, t \rangle_\sigma$ by other "primitives," e.g., a distinguished family of triples, such as the six represented (up to translates) in Fig. 18.

More generally, for any such family $\mathcal{Q}$ consider the interaction term

$$\sum_{A \in \mathcal{Q}} \left(1 - \Pi_{s \in A} b_s\right) \phi\left(\Delta_A(y)\right)$$

where $\Delta_A(y)$ is a measure of the disparity in the "A-direction." In Fig. 18, these directions are, respectively, $\pi/4$, $\pi/4$, $-\pi/4$, $-\pi/4$, $\pi/2$, and 0. One can imagine a variety of ways to situate two appropriately shaped sets of pixels to straddle (and abut) segments such as these.

B. Multivariate Data

Data may be available from several sensors, e.g., multispectral satellite data, or an optical camera and laser ra-
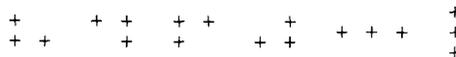

Fig. 18. Family of possible boundary primitives.

dar. One might integrate these with a composite disparity measure of the type used in the texture studies, viz. $\Delta = \max \Delta^{(k)}$, in order to improve discrimination.

Another possibility is to attempt to classify boundaries, especially if range data is also available. Let $y^{(1)}$ and $y^{(2)}$ be range and brightness values, and let x_s assume three values, say 0, 1, 2, corresponding to "off," "occluding (= depth) boundary" and "other boundary" (for instance a crease or shadow). Now rig the energy function to couple $\Delta(y^{(1)})$ with $\xi = \{\xi_s\}$, $\xi_s = \delta_{\{1\}}(x_s)$ and $\Delta(y^{(2)})$ with $\eta = \{\eta_s\}$, $\eta_s = \delta_{\{1,2\}}(x_s)$; for example, just add the two corresponding terms to form U. The penalty patterns are the usual ones regarding dead-ends etc. (with 0's and 2's as well as 0's and 1's), and mixtures of 0, 1, 2 corresponding to physically implausible (or impossible) transitions.

C. Region-Boundary Model

Put the boundary and region labels into a single model, for example of the form $U_1(x^b, y) + U_2(x^l, y)$. Now penalize improper local configurations in the pair (x^b, x^l), for example "type 1" errors (a boundary "between" like region labels) and "type 2" errors (no boundary "between" unlike labels). The problem may be that there are deep local minima which are unfaithful to the data but difficult to escape from, at least without updating many sites.

VII. SUMMARY

We have developed algorithms for partitioning an image, possibly textured, into homogeneous regions and for locating boundaries at significant transitions. Both are based on a scale-dependent notion of disparity, or gradient, and both incorporate prior expectations about regular boundary or region configurations.

The disparity measure scores the difference between the statistical structures of two scale-dependent blocks of pixels. We have experimented with several measures. Ideally, the disparity will be large when there is an apparent difference, either in gray-level or in texture, between the blocks. Usually, it was necessary to tune the measure to the particular textures or structures involved; a more universal measure may require both better preprocessing (e.g., first extracting reflectance from intensity [33]) and better use of "high-level" information about expected macrostructures and shapes. For texture discrimination, by either partitions or boundary placement, we introduce a class of features, or transformations, that are decidedly multivariate, depending on the spatial distribution of large numbers of pixel gray levels. Our disparity measure is then a composite of measures of differences in the histograms of the block data, under the various transformations. Low-order features, such as those derived solely

from raw gray-level histograms and cooccurrence matrices, were not as effective in our framework.

Disparity measures between pairs of pixel blocks drive the segmentations or boundary placements through a "label model," that specifies likely label configurations conditional on disparity data. For partitioning, labels are generic and associated with local blocks of the image. Two labels are the same if their respective regions are judged to be instances of the same texture. For boundary placement, the labels are zero or one, and interpreted as indicating, respectively, the absence or presence of boundary elements. *A priori* knowledge about acceptable label configurations, which, for example, may preclude very small or thin regions, or cluttered boundary elements, is applied by restricting labels to an appropriate subset of all possible configurations. The result of modeling disparity-label interactions and of defining restricted configurations can be regarded as a Gibbs distribution jointly on pixel gray levels and label configurations, with the marginal label distribution supported on a subset of the configuration space.

Partitioning and boundary finding is accomplished by approximating the maximum *a posteriori* (MAP) label configuration, conditioned on observed pixel data. Because certain configurations are forbidden, MAP estimation amounts to constrained optimization. Stochastic relaxation and simulated annealing are extended to accommodate constraints by introducing a nonnegative constraint function that is zero only for allowed label configurations. The constraint function, with a multiplicative constant, is added to the posterior energy, and the constant is slowly increased during relaxation. Straightforward calculations establish an upper bound on the rate of increase of this multiplicative constant that ensures convergence of the relaxation and annealing algorithms to the desired limits. In a series of partitioning and boundary-finding experiments, deterministic and other fast variations of the constrained relaxation algorithm are found to be effective.

The partitioning model is appropriate when a small number of homogeneous regions are present. Disjoint instances of a common texture are automatically identified. The boundary model can be effective in complex, multitextured, scenes. Both models sometimes require prior training to adjust parameters.

REFERENCES

[1] E. Arts and P. van Laarhoven, "Simulated annealing: A pedestrian review of the theory and some applications," in *NATO Advanced Study Institute on Pattern Recognition: Theory and Applications*, Spa, Belgium, 1986.

[2] J. Besag, "On the statistical analysis of dirty pictures" (with discussion), *J. Roy. Statist. Soc.*, series B, vol. 48, pp. 259–302, 1986.

[3] A. Blake, "The least disturbance principle and weak constraints," *Pattern Recog. Lett.*, vol. 1, pp. 393–399, 1983.

[4] E. Bonomi and J.-L. Lutton, "The *n*-city travelling salesman problem: Statistical mechanics and Metropolis algorithm," *SIAM Rev.*, vol. 26, pp. 551–568, 1984.

[5] A. Brandt, "Multi-level approaches to large scale problems," in *Proc. Int. Congr. Mathematicians 1986*, A. M. Gleason, Ed., Amer. Math. Soc., Providence, RI, 1987.

[6] P. Brodatz, *Texture: A Photographic Album for Artists and Designers*. New York: Dover, 1966.

[7] T. M. Cannon, H. J. Trussell, and B. R. Hunt, "Comparison of image restoration methods," *Appl. Opt.*, vol. 17, pp. 3385–3390, 1978.

[8] V. Cerny, "A thermodynamical approach to the travelling salesman problem: An efficient simulation algorithm," Inst. Phys. Biophys., Comenius Univ., Bratislava, Czechoslovakia, 1982.

[9] B. Chalmond, "Image restoration using an estimated Markov model," Dep. Math., Univ. Paris, Orsay, France, 1987.

[10] T.-S. Chiang and Y. Chow, "On eigenvalues and optimal annealing rate," Inst. Math., Academia Sinica, Taipei, Taiwan, preprint 1987.

[11] F. S. Cohen and D. B. Cooper, "Simple parallel hierarchical and relaxation algorithms for segmenting noncausal Markovian random fields," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, pp. 195–219, 1987.

[12] H. Derin and W. S. Cole, "Segmentation of textured images using Gibbs random fields," *Comput. Vision, Graphics, Image Processing*, vol. 35, pp. 72–98, 1986.

[13] H. Derin and H. Elliott, "Modeling and segmentation of noisy and textured images using Gibbs random fields," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, pp. 39–55, 1987.

[14] H. Derin and C. Won, "A parallel image segmentation algorithm using relaxation with varying neighborhoods and its mapping to array processors," Dep. Elec. Comput. Eng., Univ. Massachusetts, Tech. Rep., 1986.

[15] P. A. Devijver and M. M. Dekesel, "Learning the parameters of a hidden Markov random field image model: A simple example," in *Pattern Recognition Theory and Applications*, P. A. Devijver and J. Kittler, Eds. Heidelberg: Springer-Verlag, 1987, pp. 141–163.

[16] A. Gagalowicz and Song De Ma, "Sequential synthesis of natural textures," *Comput. Vision, Graphics, Image Processing*, vol. 30, pp. 289–315, 1985.

[17] D. Geman, "A stochastic model for boundary detection," *Image Vision Comput.*, May 1987.

[18] D. Geman and S. Geman, "Relaxation and annealing with constraints," Division Appl. Math., Brown Univ., Complex Systems Tech. Rep. 35, 1987.

[19] D. Geman, S. Geman, and C. Graffigne, "Locating texture and object boundaries," in *Pattern Recognition Theory and Applications*, P. A. Devijver and J. Kittler, Eds. Heidelberg: Springer-Verlag, 1987.

[20] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 721–741, 1984.

[21] S. Geman and C. Graffigne, "Markov random field image models and their applications to computer vision," in *Proc. Int. Congr. Mathematicians, 1986*, A. M. Gleason, Ed., Amer. Math. Soc., Providence, RI, 1987.

[22] S. Geman and D. E. McClure, "Statistical methods for tomographic image reconstruction," in *Bull. ISI (Proc. 46th Session Int. Statistical Institute)*, vol. 52, 1987.

[23] B. Gidas, "Nonstationary Markov chains and convergence of the annealing algorithm," *J. Statist. Phys.*, vol. 39, pp. 73–131, 1985.

[24] —, "A renormalization group approach to image processing problems," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 164–180, Feb. 1989.

[25] C. Graffigne, "Experiments in texture analysis and segmentation," Ph.D. dissertation, Division Appli. Math., Brown Univ., 1987.

[26] P. J. Green, "Discussion on "On the statistical analysis of dirty pictures," J. Besag. *J. Roy. Statist. Soc.*, vol. B-48, pp. 259–302, 1986.

[27] D. M. Greig, B. T. Porteous, and A. H. Seheult, Discussion on "On the statistical analysis of dirty pictures," by J. Besag, *J. Roy. Statist. Soc.*, vol. B-48, pp. 259–302, 1986.

[28] U. Grenander, "Tutorial in pattern theory," Division Appl. Math., Brown Univ., Lecture Notes Volume, 1984.

[29] W. E. L. Grimson and T. Pavlidis, "Discontinuity detection for visual surface reconstruction," *Comput. Vision, Graphics, Image Processing*, vol. 30, pp. 316–330, 1985.

[30] B. Hajek, "A tutorial survey of theory and applications of simulated annealing," in *Proc. 24th IEEE Conf. Decision and Control*, 1985, pp. 755-760.

[31] R. M. Haralick, K. Shanmugam, and I. Denstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. 6, pp. 610-621, 1973.

[32] R. Holley and D. Stroock, "Simulated annealing via Sobolev inequalities," preprint, 1987.

[33] B. K. P. Horn and M. J. Brooks, "The variational approach to shape from shading," *Comput. Vision, Graphics, Image Processing*, vol. 33, pp. 174-208, 1986.

[34] C.-R. Hwang and S.-J. Sheu, "Large time behaviors of perturbed diffusion Markov processes with applications," I, II, and III, Inst. Math., Academic Sinica, Taipei, Taiwan, preprint 1987.

[35] A. Kashko and B. F. Buxton, "Markov random fantasies," preprint, 1986.

[36] A. Kashko, H. Buxton, and B. F. Buxton, "Parallel stochastic optimization in computer vision," preprint, 1987.

[37] R. L. Kashyap and K. Eom, "Texture boundary detection based on the long correlation model," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-2, pp. 58-67, 1989.

[38] R. L. Kashyap and A. Khotanzad, "A model-based method for rotation invariant texture classification," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, pp. 472-481, 1986.

[39] S. Kirkpatrick, C. D. Gellatt, Jr., and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671-680, 1983.

[40] K. Laws, "Textured image segmentation," Ph.D. dissertation, Univ. Southern California, Los Angeles, USCIPI Rep. 940, 1980.

[41] R. Linsker, "An iterative-improvement penalty-function-driven wire routing system," *IBM J. Res. Develop.*, vol. 28, pp. 613-624, 1984.

[42] C. von der Malsburg and E. Bienenstock, "Statistical coding and short-term synaptic plasticity: A scheme for knowledge representation in the brain," in *Disordered Systems and Biological Organization* (NATO ASI Series), E. Bienenstock, F. Fajelman Soulié, and G. Weisbuch, Eds. Berlin: Springer-Verlag, 1986.

[43] J. L. Marroquin, "Surface reconstruction preserving discontinuities," Artificial Intell. Lab., Massachusetts Inst. Technol., Memo 792, 1984.

[44] J. L. Marroquin, S. Mitter, and T. Poggio, "Probabilistic solution of ill-posed problems in computational vision," *J. Amer. Statist. Assoc.*, vol. 82, pp. 76-89, 1987.

[45] J. W. Modestino, R. W. Fries, and A. L. Vickers, "Texture discrimination based upon an assumed stochastic texture model," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-3, pp. 557-580, 1981.

[46] J. Moussouris, "Gibbs and Markov random systems with constraints," *J. Statist. Phys.*, vol. 10, pp. 11-33, 1974.

[47] D. W. Murray and B. F. Buxton, "Scene segmentation from visual motion using global optimization," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, pp. 220-228, 1987.

[48] T. Pavlidis, "A critical survey of image analysis methods," Expanded version of paper delivered to 8th Int. Conf. Pattern Recognition, Paris, France, Oct. 1986.

[49] T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory," *Nature*, vol. 317, pp. 314-319, 1985.

[50] B. D. Ripley, "Statistics, images, and pattern recognition," *Canadian J. Statist.*, vol. 14, pp. 83-111, 1986.

[51] A. Rosenfeld and L. S. Davis, "Image segmentation and image models," *Proc. IEEE*, vol. 67, pp. 764-772, 1979.

[52] B. W. Silverman, Discussion on "On the statistical analysis of dirty pictures," by J. Besag, *J. Roy. Statist. Soc.*, vol. B-48, pp. 259-302, 1986.

[53] T. Simchony and R. Chellappa, "Stochastic and deterministic algorithms for texture segmentation," Dep. EE—Syst., Univ. Southern California, 1988.

[54] R. H. Swendsen and J.-S. Wang, "Nonuniversal critical dynamics in Monte Carlo simulations," *Phys. Rev. Lett.*, vol. 58, pp. 86-88, 1987.

[55] H. Szu, "Non-convex optimization," in *Proc. SPIE Conf. Real Time Signal Processing IX*, vol. 698, 1986.

[56] D. Terzopoulos, "Regularization of inverse visual problems involving discontinuities," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, pp. 413-424, 1986.

[57] W. B. Thompson, "Textural boundary analysis," *IEEE Trans. Comput.*, vol. C-26, pp. 272-276, 1977.

[58] E. E. Triendl, "Texturerkennung and texturreproduktion," *Kybernetik*, vol. 13, pp. 1-5, 1973.

[59] E. Triendl and T. Henderson, "A model for texture edges," in *Proc. 5th Int. Conf. Pattern Recognition*, Miami Beach, FL, Dec. 1-4, 1980, pp. 1100-1102.

[60] H. L. Voorhees, Jr., "Finding texture boundaries in images," Master's thesis, Dep. Elec. Eng. Comput. Sci., Massachusetts Inst. Technol., 1987.

[61] W. A. Yasnoff, J. K. Mui, and J. W. Bacus, "Error measures for scene segmentation," *Pattern Recog.*, vol. 9, pp. 217-231, 1977.

[62] A. I. Zobrist and W. B. Thompson, "Building a distance function for Gestalt grouping," *IEEE Trans. Comput.*, vol. C-24, pp. 718-728, 1975.

Donald Geman (M'84) received the B.A. degree in English literature from the University of Illinois, Urbana, in 1965 and the Ph.D. degree in mathematics from Northwestern University, Evanston, IL, in 1970.

Since 1970 he has been a member of the Department of Mathematics and Statistics, University of Massachusetts, Amherst, where he is currently Professor of Mathematics and Statistics. His visiting appointments include those at the University of North Carolina (1976-1977), Brown University (1983), and Université de Paris-Sud (1986). His research interests are in stochastic processes, image processing, and computer vision.
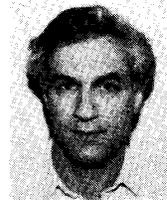
Stuart Geman (M'84) received the B.A. degree in physics from the University of Michigan, Ann Arbor, in 1971, the M.S. degree in physiology from Dartmouth College, Hanover, NH, in 1973, and the Ph.D. degree in applied mathematics from the Massachusetts Institute of Technology, Cambridge, in 1977.

Since 1977 he has been a member of the Division of Applied Mathematics at Brown University, Providence, RI, where he is currently Professor of Applied Mathematics. His research interests include stochastic processes, image and speech processing, and computer vision.

Dr. Geman is a Fellow of the Institute of Mathematical Statistics and was a recipient of the Presidential Young Investigator Award.

Christine Graffigne was born in Liévin, France, on December 1, 1959. She received the Master's degree in mathematics from Lille University, France, in 1982, the French Ph.D. degree from Paris-Sud University, Orsay, France, in 1986, and the Ph.D. degree in applied mathematics from Brown University, Providence, RI, in June 1987.

From 1987 until 1988, she was an Assistant Professor at Paris-Sud University, Orsay, France. Since 1988, she has been a Research Scholar with the French National Center for Scientific Research (CNRS), and has worked at Paris-Sud University. Her research interests are focused on stochastic processes, image processing, pattern recognition, and parallel computing.

Ping Dong (M'88) was born in Zhuzhou, China, on December 2, 1954. He received the B.S. and M.S. degrees in electrical engineering from Northwest Telecommunication Engineering Institute, Xian, China, in 1979 and 1981, respectively, and the Ph.D. degree in electrical engineering from University of Massachusetts at Amherst in 1987.

He joined the Codex, Motorola research and advanced development department, Canton, MA, in 1987. His current research interests are image segmentation and estimation algorithms, random signal processing and estimation algorithms, and communication theory.