Contour Grouping with Prior Models

James H. Elder, Member, IEEE, Amnon Krupnik, and Leigh A. Johnston, Member, IEEE

Abstract—Conventional approaches to perceptual grouping assume little specific knowledge about the object(s) of interest. However, there are many applications in which such knowledge is available and useful. Here, we address the problem of finding the bounding contour of an object in an image when some prior knowledge about the object is available. We introduce a framework for combining prior probabilistic knowledge of the appearance of the object with probabilistic models for contour grouping. A constructive search technique is used to compute candidate closed object boundaries, which are then evaluated by combining figure, ground, and prior probabilities to compute the maximum a posteriori estimate. A significant advantage of our formulation is that it rigorously combines probabilistic local cues with important global constraints such as simplicity (no self-intersections), closure, completeness, and nontrivial scale priors. We apply this approach to the problem of computing exact lake boundaries from satellite imagery, given approximate prior knowledge from an existing digital database. We quantitatively evaluate the performance of our algorithm and find that it exceeds the performance of human mapping experts and a competing active contour approach, even with relatively weak prior knowledge. While the priors may be task-specific, the approach is general, as we demonstrate by applying it to a completely different problem: the computation of human skin boundaries in natural imagery.

Index Terms—Perceptual organization, grouping, contours, edges, graph search, Bayesian probabilistic inference, segmentation, remote sensing, skin detection.

1 INTRODUCTION

Perceptual organization is the problem of grouping local image features that project from a common structure (e.g., an object) in the scene. The main tradition in both biological and computational vision research is to treat perceptual organization as a problem that can be solved in complete generality, in bottom-up fashion [1], [2], [3], [4]. In this view, it is assumed that visual context and history and the higher-level knowledge and goals of the perceiver have no role to play. To support this view, Kanizsa [3] has demonstrated that humans are able to perceptually organize artificial images composed of unfamiliar, nonsense shapes.

On the other hand, to say that perceptual organization can proceed in some fashion without higher-level knowledge or top-down constraints is not to say that it always does and that such constraints, when available, are not exploited. Convincing arguments have been made over the years that the human visual system does make use of higher-level knowledge and goals of the perceiver to speed and simplify the task of perceptual organization and to resolve ambiguities [5], [6], [7].

Our focus in this paper is to explore how such prior knowledge might be combined with more general grouping expertise, within a modern probabilistic framework. We argue that there are many practical scenarios in both human and machine vision where such prior knowledge may be available:

- Visual search. In searching for an object, an observer has some knowledge of what she/he/it is looking for: sometimes only approximate information about shape and appearance, and, typically, poor knowledge of the object’s position and attitude.
- Object recognition. In many recognition applications, the rough location of an object can be computed before recognition. This prior knowledge may then be useful in perceptually organizing the object to facilitate recognition.
- Recurrent organization. One approach to locating objects of potential interest is to search for locations that maximize the correlation with a simple parametric template that can then serve as a very simple prior model to guide detailed grouping. The existence of high-bandwidth feedback pathways at multiple stages of visual cortex [8] suggests that this style of computation is biologically feasible.
- Database Revision. There are numerous applications where existing models must be updated and/or refined. For example, an out-of-date Geographic Information System (GIS) model may be updated using recent high-resolution satellite imagery.

In many machine vision applications not all objects are of equal interest. Some applications may target faces, others hands, or cars, or shadows, etc. A vision system for a robotic manipulator may be interested only in things of the scale of the end-effector. But, in most or all of these applications, perceptual grouping is an important part of the problem. We argue that it is important to have a general framework for grouping which accommodates prior domain knowledge as a kind of “plug-in” module.

In this paper, we will focus on the particular problem of contour grouping, that is, of integrating the local edges or curve tangents that lie on a common contour in the image. Although the goal of contour grouping is sometimes defined as the inference of subsets of curve elements belonging to distinct curves [9], we define contours as sequences of local curve elements. This stronger inference directly supports the computation of higher-order curve properties (e.g., curvature) and leads naturally to standard representations (e.g., polygons, splines).
There are numerous existing algorithms for partially grouping contour elements into disjoint contour fragments \[10, 11, 12, 13, 14, 15, 16, 17\] and such partially-grouped representations have been shown to be useful for object recognition \[15\]. However, the goal of computing complete bounding contours has proven to be more elusive. While recent approaches exploiting the global property of contour closure have yielded limited success \[18, 19\], the general problem of computing the complete bounding contour of an object of arbitrary shape in a complex natural image remains essentially unsolved.

We pose the problem of contour grouping as a problem of probabilistic inference: The goal of the computation is to compute the most probable closed contour that completely bounds the object of interest. We formulate a Bayesian inference problem based on a combination of probabilistic object cues providing information about whether individual tangents lie on the object boundary, with probabilistic grouping cues providing evidence about whether sequences of tangents should be grouped together. In order to use this information to estimate the most probable object boundary, we introduce an approximate, constructive search algorithm that allows critical global constraints to be applied throughout the search.

Perceptual organization algorithms are often evaluated informally on somewhat arbitrary natural images. One goal of the present work is to objectively and quantitatively evaluate algorithm performance on real data. To this end, we apply our approach to the computation of exact lake boundaries from satellite imagery. We compile a ground truth database derived from hand segmentations of lakes from remote-sensed imagery by eight human mapping experts and use this database to compare the performance of our algorithm both to sensed imagery by eight human mapping experts and use this derived from hand segmentations of lakes from remote-satellite imagery. We compile a ground truth database to compare the performance of our algorithm to computing the maximum likelihood contour connecting two interactively-selected edges within an image editing application \[32\]. This probabilistic approach has been adapted by Crevier \[16, 17\] to a multiscale framework that allows grouping of circular arcs as well as straight contour segments. Liu and Geiger \[33\] have also used a shortest-path approach to compute maximum-likelihood paths connecting automatically selected keypoints.

In all of these approaches, the underlying probability distributions have been chosen with a combination of common sense and trial-and-error and so do not necessarily fit the natural image data on which the algorithms are expected to work. Clearly the probabilistic approach would benefit if these distributions were based on real empirical statistical data from natural images.

Almost 50 years ago, Brunswik and Kamiya \[34\] proposed that the classical Gestalt principles of perceptual organization should be quantitatively related to the statistics of the natural world. However, their suggestion remained largely untested until 1998, when Kruger \[35\] first reported data on the second-order spatial statistics of Gabor filter responses to natural images and Elder and Goldberg first described the ecological statistics of Gestalt laws for contour grouping \[36, 37, 38\]. More recently, there has been an interesting study of natural image statistics relevant to the problem of image segmentation \[39\] and three studies of natural image statistics relevant to the perceptual organization of contours \[40, 41, 42\].

In spite of this progress, very few probabilistic computer vision algorithms for contour grouping use real natural image statistics. A notable exception is Geman and Jedynak’s work on road tracking in satellite images \[43\]. In their system, simple nonlinear filters are used to locally test for roads and response distributions are learned from real data. Assuming independence and a uniform prior, the problem of road tracking is expressed as the maximization of a product of local likelihood ratios. Exact solution is intractable: instead, they build outward from their starting point a tree of possible tracks and use an entropy testing heuristic to extend the tree where uncertainty is maximal.

This problem and probabilistic framework have been studied further by Yuille and Coughlan \[44\], who focus on the computational complexity of the probabilistic search task, determining bounds on algorithm performance. Like Geman and Jedynak, they seek to maximize a product of likelihood
rather than an entropy testing heuristic they define a reward function involving the likelihood ratio for each partial contour hypothesis, partially corrected with a heuristic estimate of the reward for completing the contour. This approach has recently been applied to the grouping of contours between detected junctions [45], although this work was not based on real natural image statistics.

3.2 Computing Complete Contours
Most contour grouping algorithms (e.g., [14], [27], [28], [9], [49], [19], [16], [17], [45]) group only partial contours: They are not able to compute complete boundaries. One key to computing complete contours is the effective use of the constraint of closure. Contour closure has been shown to be a powerful cue to perceptual grouping in human vision [2], [50], [51]. Jacobs [13], [15] has studied the problem of inferring highly-closed convex cycles of line segments from an image, to be used as input for a part-based object recognition strategy. More recently, closure has been demonstrated as a very potent cue in computer vision algorithms for grouping similar contours [18], [19]. However, in this paper, we seek to compute highly complex contours with very rough, fractal-like boundaries, up to 1,000 segments in length, in images of up to 40,000 segments.

3.3 Application of Global Constraints
Prior methods for computing complete contours [18], [19] use a constraint of closure, but their formulations prohibit the application of other important global constraints. For example, although the shortest path approach used by Elder and Zucker [18] guarantees an exact solution in polynomial time, the algorithm cannot accommodate constraints of completeness, simplicity (no self-intersections), or nontrivial scale priors. In this paper, we propose as an alternative an approximate, constructive search technique which finds a good (not necessarily optimal) solution, but which can accommodate these important constraints.

3.4 Heterogeneous Objects
Although segmentation approaches (e.g., [48], [49]) guarantee a closed boundary, they require an assumption of stationary (or, at the very least, smoothly varying) statistics over the interior of an object. This is simply not realistic for many objects in the world, e.g., the human body (clothed or unclothed). In such cases, it is critical to accurately capture and exploit the statistical regularities of the object boundary. In our case, we are interested in computing the boundaries of lakes in satellite imagery and skin patches in natural imagery. In the first case, the statistics are not stationary because of islands and swampy areas. In the second case, hair and facial features break the homogeneity of skin patches. While segmentation approaches are unlikely to successfully group entire structures under these conditions, we will demonstrate that our contour-based approach can.

3.5 Weak versus Strong Priors
Active contour approaches typically require good initial conditions to converge correctly. Even in simplified conditions [49] an explicit, initial polygon that substantially overlaps the target object must be provided. In our approach, the nature of the prior knowledge is much more flexible. While we will demonstrate the superiority of our approach for relatively strong priors providing both position and shape information, we will also demonstrate that our algorithm performs well with only very coarse size and/or photometric priors, making it much more useful for
“visual search” scenarios where the location of the object of interest is completely unknown. We will show how these priors can be rigorously combined with more generic grouping cues in a consistent probabilistic framework.

3.6 Quantitative Evaluation

Contour grouping systems are typically evaluated subjectively and qualitatively. We are aware of only two exceptions. Borra and Sarkar [52] evaluated three existing edge grouping systems in terms of their utility for object recognition in aerial imagery. Williams and Thornber [53] evaluated six saliency measures for the detection of natural shapes in synthetic noisy line element images. However, to our knowledge, ours is the first contour grouping study in which the algorithm is directly evaluated (without reference to a specific high-level task), quantitatively and objectively, on natural image data, and compared against other systems, both human and machine.

4 FRAMEWORK

4.1 Definitions

We assume an input representation in which contours are locally represented by variable length tangents. These tangents are formed in a two-stage process. First, edges are detected in the image using an adaptive multiscale edge detection algorithm [54]: An example edge map is shown in Fig. 1b. Second, a greedy algorithm is employed to group local edges into relatively short linear segments of variable length that closely approximate the underlying curve and whose position and orientation are represented to floating-point accuracy. The primary purpose of this second stage is to reduce artifacts induced by the discrete pixel representation of the edge map. In addition to carrying position and orientation information, these tangents are augmented with estimates of the luminance intensity on both sides of the contour. Details of the representation can be found in [18].

\[ T = \{ t_1, \ldots, t_N \} \]

...}

\[ S = \{ s_1, \ldots, s_m \} \]

...}

This definition restricts the mapping to be injective (tangents cannot be repeated), with the exception that the first and last tangent may be the same. In this case, the contour is closed.

We assume that there exists a correct organization of the image \( C \subseteq S \). \( C \) includes all tangent sequences that correspond to actual contours in the image and all subsequences of these. In other words, contours in \( C \) need not be complete. \( C \) includes the contour bounding the object(s) of interest, but also many other “distractor” contours that may exist in the image. We do not restrict these contours to be disjoint since there are many instances in real images where contours bounding distinct structures may share one or more contour elements. For example, in our GIS application, the visible boundary of a lake is in fact composed largely of trees and bushes that surround the lake. Thus there are many contour elements that may be

1. In our formulation, the local detection and grouping processes are decoupled: Contours are restricted to pass through precomputed tangents.
We assume that the object boundary \( e^o \) and therefore all object contours are simple (not self-intersecting).

### 4.2 Constructive Algorithm

While the shortest path algorithm used in previous grouping approaches [18], [16], [17] is guaranteed to find an optimal solution, it cannot accommodate the constraints of simplicity (no self-intersections), completeness, or nontrivial global priors, all of which are important parts of the problem we address. We therefore propose as an alternative a heuristic algorithm that uses a greedy, monotonic strategy. While not guaranteed to find an optimal solution, it can handle these important global constraints.

The proposed constructive method estimates the complete object boundary \( e^o \) from contours that are likely to be object contours. We begin by constructing the set \( S_2 = \{ s_2, \ldots, s_{2m} \} \) of contours \( s_2 \) of length \( m = 2 \) tangents and estimating the posterior probability \( p(s_2) \in C^o | D \) that each is an object contour, given observed data \( D \). The means for estimating these posterior probabilities are discussed in Section 4.4. The data cues we use in our particular application are detailed in Section 5.2.

To limit memory and time complexity, we discard all but the \( N_2 \) most probable contours \( C_2 \) and from these we construct a new set \( S_3 = \{ s_{31}, \ldots, s_{3n_3} \} \) of length \( m = 3 \) contours. This process is iterated to some maximum contour length \( M = \max \{ m \} \). Self-intersecting and closed contours may form at all stages \( m > 2 \); these are subtracted from \( C_0 \) and the closed contours \( C_m \) are stored. Note that self-intersection is a global property that could not be handled easily using a shortest-path algorithm.

In order to make the memory requirements for our algorithm invariant with respect to contour length, we fix the total number \( N \) of tangents represented at each stage (after pruning) to \( N = 4,000 \). This means that the number of contours \( N_m \) that are represented declines as the contour length \( m \) increases: \( N_m = N/m \). Approximate prior knowledge about the object of interest allows the maximum contour length \( M \) to be chosen in a principled manner (Section 5.3).

In order to limit the time requirements of our algorithm, we have applied two additional constraints in our implementation. First, we only consider tangent pairs whose endpoints are within 20 pixels of each other. This constraint limits the time required to compute the pairwise grouping probabilities. Second, we select and store only the top 10 local grouping hypotheses for each tangent endpoint. This constraint limits the computation time required for the constructive search.

### 4.3 Computing Posterior Probabilities of Contour Hypotheses

We seek to estimate

\[
p(s \in C^o | D) = \frac{P(D | s \in C^o) P(s \in C^o)}{P(D)}. \tag{1}
\]

Here, \( D = D^c \cup D^f \) represents the set of observable cues available, where \( D^c \) represents the set of object cues determining the likelihood that tangents lie on the boundary of the target object and \( D^f \) represents the grouping cues determining the likelihood that a sequence of tangents should be grouped together. These can, in turn, be broken down into local components: \( D^c = \{ d_1, \ldots, d_N \} \), where each \( d_i \) is a set of \( l_i \) observations about tangent \( t_i; d_i = \{ d_i^1, \ldots, d_i^{l_i} \} \). Similarly, \( D^f = \{ d_j \}, i, j \in \{ 1, \ldots, N \}, i \neq j \), where each \( d_{ij} \) is a set of \( l_i \) observations about the relationship between tangents \( t_i \) and \( t_j; d_{ij} = \{ d_{ij}^1, \ldots, d_{ij}^{l_i} \} \).

We assume that local cues \( d_i, d_{ij} \) pertaining to a particular tangent or tangent pair are independent of cues pertaining to other tangents or tangent pairs when conditioned upon a contour hypothesis \( s \in C^o \). We also assume that local cues are independent of all components of a contour hypothesis \( s \) except the tangent or tangent pair to which they directly pertain.

Let \( I^o \) represent the set of indices of all tangents in the image: \( I^o = \{ 1, \ldots, N \} \). Let \( I^o_i = \{ \alpha_1, \alpha_2, \ldots, \alpha_{m-1} \} \) represent the indices of tangents on a particular contour hypothesized to be on the boundary of the object. Let \( I^o_f \) represent the indices of tangents not on the contour: \( I^o_f = I^o \setminus I^o_i \).

Similarly, let \( I^f \) represent the set of index pairs of all tangent pairs in the image: \( I^f = \{ 1, \ldots, N \} \times \{ 1, \ldots, N \} \). Let \( I^f_i = \{ \alpha_1, \alpha_2, \ldots, \alpha_{m-1} \} \) represent the indices of directly successive tangent pairs on the hypothesized contour. Let \( I^f_f \) represent all other tangent pairs: \( I^f_f = I^f \setminus I^f_i \).

We observe that the normalizing term \( P(D) \) in (1) can be replaced by \( \prod_{i \in I^o_f} P(d_i) \prod_{(i,j) \in I^f_f} P(d_{ij}) \) without affecting the computation since both are constant over all hypotheses. Thus, we can expand (1) as:

\[
p(s \in C^o | D) \propto \prod_{i \in I^o_f} P(d_i | s \in C^o) \prod_{(i,j) \in I^f_f} P(d_{ij} | s \in C^o) P(s \in C^o).
\]

Observing that \( d_i, i \in I^o_f \) and \( d_{ij}, (i, j) \in I^f_f \) are independent of the current hypothesis \( s \in C^o \), the likelihood terms for the tangents not on the contour hypothesis cancel and we have:

\[
p(s \in C^o | D) \propto \prod_{i \in I^o_f} P(d_i | t_i \in T^c) \prod_{(i,j) \in I^f_f} P(d_{ij} | \{ t_i, t_j \} \in C) P(s \in C^o).
\]

Now, recall that the purpose of computing the posterior probabilities of contour hypotheses in the constructive phase is to decide which contours to prune and which to continue. In making this decision, the contours we compare all consist of the same number of tangents. Since the prior term \( p(s \in C^o) \) depends only upon the number of tangents in \( s^2 \), it can be replaced by the term \( \prod_{i \in I^o_f} P(t_i \in T^c) \prod_{(i,j) \in I^f_f} P(\{ t_i, t_j \} \in C) \), as

2. Generally in this paper we use the terms “prior models” and “prior knowledge” to refer to task, domain, or application-specific information that may be applied to assist the grouping process. This must be distinguished from the technical use of the word “prior” in this section to refer to the unconditioned probabilities of the tangents or tangent sequences that we are trying to infer. In this technical usage, the “priors” do not involve any properties such as tangent position or orientation that must be measured from the image. Thus, most of the “prior knowledge” in the colloquial sense of the term is actually incorporated in the conditioned and unconditioned probability distributions for the observable cues \( d_i \) and \( d_{ij} \).
both are constant over all hypotheses at each stage of the constructive computation. Substituting into (2):
\[
p(s \in C^o|D) \propto F(s, D) = \prod_{i \in I_1} \prod_{(i,j) \in I_1} p_{ij}^o,
\]
reflecting the probability of the contour computed in the constructive phase, 2) a background factor, reflecting the probability that all other tangents lie in the background, and 3) a prior factor, that accounts for the number of tangents on the hypothesis. It is the background factor that embodies the completeness constraint: if the hypothesized contour does indeed represent the entire object boundary, it must be the case that all tangents that are not part of the contour do not lie on the object boundary. This constraint turns out to be the crucial factor in preventing a bias to short contours inherent in the foreground term.

Specifically, the maximum a posteriori estimate \( \hat{c}^o \) of the object boundary \( c^o \) is given by
\[
\hat{c}^o = \arg \max_{s \in \mathcal{C}^o} p(s = c^o|D),
\]
where
\[
p(s = c^o|D) = \frac{p(D|s) p(s = c^o)}{p(D)}.
\]
Again, we observe that we can replace the normalizing term \( p(D) \) by \( \prod_{i \in I_1} p(d_i) \prod_{(i,j) \in I_1} p(d_{ij}) \), without affecting the computation since both are constant over all hypotheses. Then (6) can be written as a product of three factors:
\[
p(s = c^o|D) = F^*(s, D) B^*(s, D) P^*(s)
\]
where
\[
F^*(s, D) = \prod_{i \in I_1} \frac{p(d_i|s = c^o)}{p(d_i)} \prod_{(i,j) \in I_1} \frac{p(d_{ij}|s = c^o)}{p(d_{ij})}
\]
and
\[
P^*(s) = p(s = c^o).
\]
The foreground term \( F^*(s, D) \) is derived from the posterior \( F(s, D) \) computed in the constructive phase. Since \( I_1 \) consists only of tangents on the contour hypothesis, and the current hypothesis is assumed to be the true object boundary \((s = c^o)\), we have that
\[
\prod_{i \in I_1} \frac{p(d_i|s = c^o)}{p(d_i)} = \prod_{i \in I_1} \frac{p(d_i|t_i \in T^o)}{p(d_i)}.
\]
Similarly, since \( I_1 \) consists only of successive tangent pairs on the contour hypothesis and the current hypothesis is assumed to be the true object boundary \((s = c^o)\), we have that
\[
\prod_{(i,j) \in I_1} \frac{p(d_{ij}|s = c^o)}{p(d_{ij})} = \prod_{(i,j) \in I_1} \frac{p(d_{ij}|t_i, t_j \in C)}{p(d_{ij})}.
\]
Thus, for a contour hypothesis of length \( m \), we have
\[
F^*(s, D) = (p^o p^f)^{-m} F(s, D)
\]
where
\[
p^o = p(t_i \in T^o)
\]
and
\[
p^f = p(t_i, t_j \in C).
\]
To compute the background term \( B^*(s, D) \), we continue to assume that the contour hypothesis only influences grouping cues relating successive tangents on the hypothesized contour. Thus, the second product over local grouping likelihoods vanishes and \( B^*(s, D) \) simplifies to

---

3. See [38] for evidence of mutual independence for our encoding of the Gestalt grouping cues.
\[ B'(s, D) = \prod_{i \in I^c_2} \frac{p(d_i | t_i \not\in T')}{p(d_i)}. \]

However, we can no longer make the same assumption for the object cues: given a complete contour hypothesis, all tangents not on the hypothesis must be explained by the background model. Thus, we have:

\[ B'(s, D) = (1 - p^a)^{m-N} \prod_{i \in I^c_2} (1 - p_i^o) = (1 - p^a)^{m-N} \prod_{i \in I_2} (1 - p_i^o). \]

Finally, we must consider the prior term \( P^*(s) \) which embodies both our prior knowledge about the number of tangents \( m \) likely to be on the object boundary and the number of possible contours of length \( m \) that could be formed from the \( N \) tangents in the image.

Our prior \( p(m^o) \) on the number of tangents on the object boundary will be derived from training data, modeled as a normal distribution, and the number of possible closed contours of length \( m \) in the image is given by \( N!/(2m(N-m))^{-1} \). Thus,

\[ P^*(s) = p(m^o = m) \frac{2m(N-m)!}{N!}. \]

In practice, we find that the prior term has very little effect on the computation: It is the foreground and background terms that together determine the contour selected to represent the object boundary.

5 GIS APPLICATION

We first test our approach on the problem of computing exact lake boundaries from panchromatic satellite (IKONOS) imagery. We have approximate polygonal models of the boundaries of these lakes from an existing digital database. A prior registration stage is required to register these polygonal models to the IKONOS data. This registration was based on the observation that lakes encompass relatively few contours (save for those generated by islands, marshlands, etc.). Registration thus involves a search over the set of 2D translations and rotations to minimize the number of edge pixels within the polygon. Fig. 1a shows an example model before (blue) and after (red) registration.

5.1 Estimating the Priors

We derived statistical models for the prior terms and likelihood distributions using data from seven lakes for which the bounding contours were hand-traced by a mapping expert on the edge map. The prior ratios \( P^o_i \) and \( P^g_{ij} \) for object membership and grouping are determined solely from the number of tangents in the image and the length of the prior model.

The prior ratio for object membership \( P^o_i \) is simply the ratio of the number of tangents on the object boundary to the number off the object boundary. Of course we do not know the exact number of tangents on the boundary, but we can use the prior model to estimate this. From our training data, we find that the relationship between the length of the polygonal model and the number of tangents on the human-traced object boundary for our training data is nearly linear, with a Pearson correlation of 0.97. Thus, we use the training data to estimate the mean \( \bar{a} \) and standard deviation \( \sigma_a \) of the ratio of the number of tangents on the lake boundary to the length of the polygonal model: \( \bar{a} = 0.28 \) tangents/pixel, \( \sigma_a = 0.06 \) tangents/pixel. Given a novel image, we estimate the expected number of tangents \( m^o \) on the lake boundary to be \( m^o = \bar{a}L \), and the standard deviation to be \( \sigma_m = \sigma_aL \), where \( L \) is the length of the prior model (in pixels). We then have that \( P^o_i = m^o/(N-m^o) \), where \( N \) is the total number of tangents in the image and \( m^o \) and \( \sigma_m \) define the prior on contour length used in the evaluation stage (Section 4.5).

We estimate the prior ratio \( P^g_{ij} \) for grouping by assuming that, on average, each tangent groups with one other tangent in each direction. Note that this is simply an estimate of the expected case and does not imply that every tangent in the image will group with exactly one tangent in both directions. This assumption implies that \( P^g_{ij} = 1/(N-2) \), where \( N \) is the total number of tangents in the image.

5.2 Estimating Likelihoods and Posteriors

It is common to use Gaussian, log Gaussian, or exponential models for probability density functions (PDFs) in the computational modeling of perceptual processes. However, we have found that statistical distributions of grouping cues are typically kurtotic [38]. To model these distributions, we employed a generalized Laplacian distribution that has been used in the past to model kurtotic wavelet response histograms [55], [56], [57]:

\[ p(x) = Ae^{-(|x|/\mu)/\sigma}, \]

where

\[ c = \sqrt{\Gamma(3/\gamma) / \Gamma(1/\gamma)} \]

and

\[ A = \frac{\Gamma(\gamma/\gamma)}{\Gamma(1/\gamma)} \text{ if } x \text{ is defined on } (-\infty, \infty) \]

\[ A = \frac{\Gamma(\gamma/\gamma)}{\sigma^\gamma} \text{ if } x \text{ is defined on } [0, \infty). \]

This distribution is symmetric and unimodal. \( \mu \) is the mean of the distribution, \( \sigma \) is the standard deviation, and \( \gamma \in (0, \infty) \) controls the kurtosis. If \( \gamma = 2 \), the distribution is Gaussian. If \( \gamma < 2 \), the distribution has positive kurtosis. If \( \gamma > 2 \), the distribution has negative kurtosis, approaching a uniform distribution as \( \gamma \to \infty \). Given a target kurtosis calculated from the data, the required \( \gamma \) is found using standard nonlinear optimization techniques.

Each of the object and grouping cues is assumed to be independent\(^5\) so that the joint likelihood for the cue vector for a given tangent or pair of tangents is simply the product of the individual likelihoods.

In this application, as object cues \( d^k_i \) for tangent \( t_i \) we use:

1. The intensity on the dark side of \( t_i \): Lakes tend to be dark in panchromatic IKONOS imagery.
2. The distance between \( t_i \) and the nearest tangent in the direction opposite to the local intensity gradient

\(^4\) The factor of \( 2m \) arises because each closed contour can be represented \( 2m \) ways, starting at any one of the \( m \) tangents and traversing in either clockwise or counterclockwise directions.

\(^5\) For evidence of the independence of the classical Gestalt grouping cues, see [38].
(i.e., toward the interior of the lake if \( t_i \) is on the lake boundary): Lakes tend to have few interior contours.

3. The distance between \( t_i \) and the nearest point on the model.

4. The angle of \( t_i \) with respect to the nearest model segment.

Likelihood distributions for tangents on and off the lake boundaries and the resulting posterior distributions for all four object cues are shown in Fig. 2. The difference between the distributions for tangents on and off the boundary attests to the useful information in each of these cues. Fig. 1c shows a map of the posterior probability of object membership \( p(t_i \in T^o(d^k)) \) for the four object cues combined. The intensity of the tangents is inversely proportional to the probability that they lie on the lake boundary. It is clear that prior knowledge of the object of interest focuses the problem considerably.

As grouping cues \( d_{ij} \) between tangents \( t_i \) and \( t_j \), we use:

1. The distance (tip to tail) between the tangents (proximity cue).

2. A measure of good continuation [1], [2], [18] between the two tangents, approximated as the absolute value of the sum of the angles formed by a linear interpolant.

3. The change in the mean intensity (brightness) between the two tangents.

4. The change in edge contrast between the two tangents.

Likelihood distributions for grouped and nongrouped tangent pairs and the resulting posterior distributions for all four grouping cues are shown in Fig. 2. Again, the degree to
which the distributions for grouped and nongrouped tangents differ indicates the strength of the cue.

5.3 Selecting the Maximum Contour Length

The approximate polygonal model can be used to select the maximum contour length $M$ considered in the search for the object boundary. In Section 5.2, we noted the linear relationship between the number $\bar{m} = L$ of tangents on the lake and the length $L$ of the prior model: $\bar{m} = \alpha L$, $\alpha = 0.28$, $\sigma_n = 0.06$ tangents/pixel. Given a new image and new prior model of length $L$ pixels, we estimate the 95th percentile upper bound $M$ on the number of tangents on the boundary as $M = L(\alpha + 2\sigma_n)$.

An interesting pattern is observed in the number of closed contours generated at each contour length in our constructive algorithm: Fig. 3c shows an example. A large number of closed contours less than 10 tangents in length are generated; this is presumably due to the large number of small contours represented and the numerous complete boundaries of trees located near the lakeshore. A smaller number of closed contours (including the maximum probability contour) between 280 and 300 tangents in length are also found. No closed contours between 10 and 280 tangents in length are computed. In a typical case, we find between 1,000 and 2,000 very short closed contours and on the order of 100 closed contours of the scale of the object. The preponderance of relatively short closed contours underlines the importance of the completeness criterion (Section 4.5) in avoiding a bias toward short contours.

5.4 Inferential Power of Object and Grouping Cues

We selected cues based upon our intuitions about which observable properties would be most powerful in discriminating object tangents from nonobject tangents and tangents that should be grouped from those that should not. One advantage of our probabilistic framework and the ground-truth database is that they allow us to actually quantify the inferential power of each of these cues. We describe our method for measuring statistical power for the grouping cues; the object cues are treated in similar fashion.

To measure the statistical power of each cue for the grouping of two tangents $\{t_i, t_j\}$, we define a random variable $G_{ij} \in \{G_1, G_2\}$ to represent the actual grouping relationship between tangents $t_i$ and $t_j$, where $G_1$ represents the event $\{t_i, t_j\} \in C$ and $G_2$ represents the event $\{t_i, t_j\} \not\in C$. Prior to making any observations about the relationship between the two tangents, there is an inherent uncertainty in the decision variable $G_{ij}$ that can be formally characterized by its informational entropy $H(G_{ij})$ [58]:

$$H(G_{ij}) = -p(G_1) \log_2 p(G_1) - p(G_2) \log_2 p(G_2).$$
If it were the case that two arbitrary tangents were as likely to be grouped as not, the prior entropy of the grouping decision would be exactly one bit. However, since it is far more likely that two arbitrary tangents are not grouped, the prior entropy is much lower. If we assume that, on average, a tangent group with one other tangent in each direction, then $p(G_{1}) \approx 1/n$ and $p(G_{2}) \approx 1 - 1/n$, where $n$ is the number of tangents in the image. For our sample of images, in which the mean number of tangents per image is roughly 5,000, the prior entropy $H(G_{ij})$ of the grouping decision is roughly 0.0027 bits.

Grouping cues provide knowledge about the relationship between tangents that can reduce the uncertainty in the grouping decision. The size of this reduction in uncertainty can be characterized by the mutual information $I(G_{ij};d_{ij})$ between the grouping decision $G_{ij}$ and the cue $d_{ij}$:

$$I(G_{ij};d_{ij}) = \int p(G_{1}, d_{ij}) \log_2 \frac{p(G_{1}, d_{ij})}{p(G_{1})} dd_{ij} + \int p(G_{2}, d_{ij}) \log_2 \frac{p(G_{2}, d_{ij})}{p(G_{2})} dd_{ij}.$$

Given observation of the grouping cue $d_{ij}$, the remaining uncertainty in the grouping decision $G_{ij}$ is given by

$$H(G_{ij}|d_{ij}) = H(G_{ij}) - I(G_{ij};d_{ij}).$$

Thus, the mutual information measures the information, in bits, that the cue $d_{ij}$ provides about the decision $G_{ij}$ of whether to group tangents $t_i$ and $t_j$. Normalizing the mutual information $I(G_{ij};d_{ij})$ by the prior uncertainty $H(G_{ij})$ yields the percentage reduction in informational uncertainty provided by the cue.

Fig. 3a shows the information provided by each object cue in deciding whether tangents are on or off the lake boundary. Interestingly, the two most important cues are appearance cues (1 and 2): the intensity and distance to the nearest neighboring tangent on the dark side of a tangent. Of less, but still significant importance, are the position cues (3 and 4): the distance to the prior model and the relative orientation of the tangent with respect to the nearest model segment.

Fig. 3b shows the information provided by each grouping cue in deciding whether two tangents should be grouped. Proximity is by far the most powerful cue. Good continuation and brightness similarity cues are of less value and the contrast cue is of almost no value.

### 6 Performance Evaluation

We tested the algorithm on five novel lakes. These images are relatively complex: The number of tangents in an image averaged 19,000, ranging from roughly 3,000 to 42,000. The number of tangents on the lake boundary averaged 490, ranging from roughly 100 to 1,000. Computation time is roughly linear with the number of tangents in the image: The algorithm takes an average of 15 minutes on a 1 GHz Pentium to process each of these relatively complex test images.

Fig. 1d shows an example result of our algorithm (red), together with a contour traced by a human observer (blue). Note that, for most of the contour, the human and machine observers agree almost exactly, but there are key ambiguous regions where the contours diverge significantly. Results for three of the test lakes are shown in Fig. 4 (labeled “Strong Prior”).

#### 6.1 Ground-Truth Database

An ideal ground-truth database would be obtained by a field survey of the actual lake perimeters. To obtain such data at the resolution of our satellite imagery (1 m) is impractical. Typically, such high-resolution contours are mapped by hand from remote-sensed imagery. Thus, for this study, we constructed a ground-truth database from hand segmentations of lake boundaries on the image data by eight human mapping experts. Each expert was instructed to trace the perimeter of each of the test lakes using the software tool of their choice.

We found that human experts varied considerably in their segmentations, and that these variations were not uniformly distributed, but concentrated at ambiguous areas, such as marshland. Given this nonuniform variability in human consistency, we used the following method for evaluating machine performance: For each lake, we computed the deviation between each point placed by the observer and the nearest point on the contour computed by the algorithm. We then normalized this deviation by the mean deviation between the point and the contours traced by all other human observers. The median of these normalized deviations over all points and all observers was then computed for each lake. The resulting median relative error is less than 1 if the algorithm is more consistent with the human experts than they are with each other and greater than 1 if the algorithm is less consistent.

Fig. 3d shows the quantitative results of the algorithm (labeled Grouping (Strong Prior)) for all five test lakes. We found a median relative error of less than 1 for all lakes, indicating that the algorithm is more consistent with the human experts than they are with each other.

#### 6.2 Contour Grouping with Weak Priors

It is of interest to know how the algorithm would fare when given far weaker prior knowledge. In particular, we are interested in the case when the two cues based on the distance and angle between each tangent and the GIS model (cues 3 and 4) are not available, so that the system has no prior knowledge of the shape, position, or pose of the object. To test this scenario, we turn off these cues, but continue to use the first two cues (the intensity on the dark side of the tangents and the distance to the nearest tangent on that side). Note that these two cues are general to all lakes and do not depend upon any prior knowledge of the specific lake to be grouped.

We also augment these two general cues with a new third cue that embodies weak prior knowledge about the size of the object, obtained by estimating the length of the bounding contour from the length of the GIS polygon. This gives the length of the contour to within roughly 20 percent accuracy but provides no information about location or termination.
shape. This knowledge of the approximate size of the object allows the distance between each pair of tangents on the bounding contour to be very roughly predicted and when incorporated into the contour probability (3), has the effect of pruning paths that wander off with no sign of returning.

Fig. 4 shows results of the Weak Prior formulation of the algorithm for three of the test lakes. The quantitative results are shown in Fig. 3d. While the median error is greater than when a strong prior is used, the algorithm still performs about as well as the average human expert.

6.3 A Competing Approach: Active Contours

Active contour models are widely used for approximating contours in computer vision [46], [59] and remote sensing applications [60], [61]. One advantage of the active contour approach is that one can initialize the model as a closed contour and thus maintain the constraint of closure throughout the computation. In our constructive approach, on the other hand, we begin only with a large number of open contour fragments and then search for promising closed contours.

We tested a public-domain implementation of active contours, using the prior GIS model as an initial contour. The algorithm is governed by 14 parameters, such as resistance to tension and bending, attraction to image features, feature threshold, image smoothing, etc.; these parameters were tuned by trial-and-error to optimize performance.

8. Available from http://www.s2.chalmers.se/ghassan/phd/illus/snakes/index.htm. We selected this implementation on the basis of its availability and completeness. Other active contour algorithms may produce different results.
Fig. 4 shows results of this active contour algorithm for three of the test lakes. Quantitative performance is shown in Fig. 3d. We find that, on average, the active contour performance is about 68 percent worse than the strong-prior formulation of our grouping algorithm and about 29 percent worse than the weak-prior formulation. This is in spite of the fact that the GIS polygon is given to the active contour implementation as an initial condition, but is not given to the weak-prior version of our grouping algorithm.

It is possible that a region-based approach [49] might perform better on this application, although, due to the presence of islands and marshes within the lakes, the homogeneity condition required by such an approach is not exactly satisfied in this application. However, such an approach would still require good initialization (strong position and shape priors). Also, our approach has the potential to generalize to the demarcation of much less homogeneous objects.

7 GENERALIZING TO OTHER APPLICATIONS

One of the main goals of this work is to determine how prior knowledge about the grouping target can be integrated with more general grouping cues within a rigorous probabilistic framework. While the priors may change over time, depending upon the goals of the perceiver, the probabilistic machinery and algorithm remain the same. To illustrate this, we demonstrate the approach on a completely different application: that of finding the boundaries of major skin regions in natural images. Due to space limitations, we describe this application in less detail: we wish only to illustrate the generality of the approach.

The test and training images are a selection from the Corel Photo Library. As object cues, we use only the RGB color on either side of the tangents [62]. The likelihood functions for skin and nonskin color channels are modeled by fourth order Gaussian mixture models, trained on a set of five images from the database.
We derive the statistics for proximity and good continuation by manually tracing the skin region boundaries on the set of five training images. Note that this is weak prior knowledge: The algorithm is given no information on the location, shape, pose, or size of the objects to be grouped.

To compute multiple contours, after each maximum a posteriori contour is computed, the probabilities of the associated tangents are set to zero and the algorithm is iterated. Fig. 5 shows the top two bounding contours computed for the example images. Complete contours bounding the two skin regions in each image are computed, despite the heterogeneities within these regions. Skin-colored background regions are ignored (e.g., the wood surfaces in the background of Fig. 5b). This was made possible by combining the color object cue with the grouping cues that exploit the geometric regularity of the boundaries.

An alternative to using separate statistical models for separate types of imagery (e.g., satellite data, natural imagery) is to develop adaptive “bootstrapping” models, capable of estimating grouping statistics online. We hope to explore this idea in future work.

8 Summary
We have introduced the problem of grouping complete bounding contours using strong or weak prior knowledge of the objects of interest and have shown how the problem can be formulated as a problem of Bayesian inference. We developed an approximate constructive search technique that computes candidate object boundaries. This approach allows important global constraints, such as simplicity, completeness, and nontrivial priors to be applied. Candidate boundaries are evaluated rigorously using foreground and background data as well as prior probabilistic knowledge of the object size.

We demonstrated the effectiveness of the method on the problem of computing exact lake boundaries from high-resolution satellite imagery. We found that the algorithm outperformed both human mapping experts and a competing active contour model. We also demonstrated the generality of the approach by applying it to the problem of skin segmentation. The algorithm is capable of combining object and grouping cues to segment complete regions of human skin in natural images despite heterogeneities within these regions.

The approach is essentially parameter-free: all distributions are based upon the measured statistics of natural images.

Acknowledgments
This work was supported by the Natural Sciences and Engineering Research Council of Canada, the Canadian GEOIDE Network of Centres of Excellence, and the Ontario Centre for Research in Earth and Space Technology.

References


James H. Elder received the BSc degree in electrical engineering from the University of British Columbia in 1987 and the PhD degree in electrical engineering from McGill University in 1995. From 1987 to 1989, he was with Bell-Northern Research in Ottawa, Canada. From 1995 to 1996, he was a senior research associate in the Computer Vision Group at the NEC Research Institute in Princeton, New Jersey. He joined the faculty of York University, Canada, in 1996, where he is presently an associate professor. His current research addresses problems in computer vision, biological vision, and virtual reality. He is a member of the IEEE and the IEEE Computer Society.

Amnon Krupnik received the BSc degree in geodetic engineering and the BA degree in computer science in 1988 and the MSc degree in geodetic engineering in 1990, all from the Technion-Israel Institute of Technology. He received the PhD degree from Ohio State University in 1994. He has been with the department of Civil Engineering at the Technion since 1994. His major research activities are in the areas of computer vision, digital photogrammetry, and remote sensing.

Leigh A. Johnston received the BSc and BE degrees in computer science and electrical engineering, respectively, in 1997 and the PhD degree in electrical engineering in 1999 from the University of Melbourne, Australia. From 2000 to 2001, she was a research associate with the Centre for Systems Engineering and Applied Mechanics at the Université Catholique de Louvain, Belgium, following which she returned to the University of Melbourne as a research associate in the Centre for Ultra-Broadband Information Networks. Since 2002, she has been a research associate in the Centre for Vision Research at York University, Canada, where she is interested in statistical signal processing algorithms for computer vision applications. She is a member of the IEEE.