Int J Comput Vis (2008) 77: 47–63
DOI 10.1007/s11263-007-0092-6

# Learning Probabilistic Models for Contour Completion in Natural Images

**Xiaofeng Ren · Charless C. Fowlkes · Jitendra Malik**

Received: 17 August 2005 / Accepted: 11 September 2007 / Published online: 18 October 2007


**Abstract** Using a large set of human segmented natural images, we study the statistics of region boundaries. We observe several power law distributions which likely arise from both multi-scale structure within individual objects and from arbitrary viewing distance. Accordingly, we develop a scale-invariant representation of images from the bottom up, using a piecewise linear approximation of contours and constrained Delaunay triangulation to complete gaps. We model curvilinear grouping on top of this graphical/geometric structure using a conditional random field to capture the statistics of continuity and different junction types. Quantitative evaluations on several large datasets show that our contour grouping algorithm consistently dominates and significantly improves on local edge detection.

**Keywords** Grouping · Natural images · Boundary detection · Scale invariance · Conditional random fields · Machine learning

X. Ren (✉)
Toyota Technological Institute at Chicago, 1427 E. 60th Street, Chicago, IL 60637, USA
e-mail: xren@tti-c.org

C.C. Fowlkes
School of Information and Computer Science University of California, Irvine, CA 92697-3425, USA
e-mail: fowlkes@ics.uci.edu

J. Malik
Computer Science Division, University of California at Berkeley, Berkeley, CA 94720, USA
e-mail: malik@cs.berkeley.edu

## 1 Introduction

Finding the boundaries of objects and surfaces in a scene is a problem of fundamental importance for computer vision. For example, there is a large body of work on object recognition which relies on boundary detection to provide information about object shape (e.g. Borgefors 1988; Huttenlocher et al. 1993; Belongie et al. 2002; Felzenszwalb 2001). Even in cases where simple intensity features are sufficient for object detection, e.g. faces, it is still desirable to incorporate boundary detection in order to provide precise object segmentation (e.g. Borenstein and Ullman 2002; Tu et al. 2005; Yu et al. 2002). The availability of high quality estimates of boundary location will ultimately govern whether these algorithms are successful in real-world scenes where clutter and texture abound.

The problem of boundary detection has been attacked at several different levels:

1. *Local edge detection*: this is the traditional approach to boundary detection and has been an area of central research since the early days of computer vision. A local edge detector typically considers a small patch centered at each image location, measures oriented brightness gradients, and outputs this contrast as the boundary likelihood. The Canny edge detector (Canny 1986) is perhaps the most influential and popular algorithm in this category. More recent approaches (Konishi et al. 1999; Martin et al. 2004) train a classifier on a collection of natural images in order to combine local gradient cues.
   By nature of its locality, edge detection is necessarily limited in how much information it can extract about boundary presence. Recent psychophysical studies (Martin et al. 2003) have suggested that current techniques may not be far from this limit: when presented with only

local patches, human observers perform no better than a gradient based classifier (Martin et al. 2004). It seems that utilizing greater contextual information is necessary to further improve boundary detection.

2. *Mid-level inference*: since boundaries and regions interact over large parts of the image, it makes sense to define global criteria for determining the correct boundaries (e.g. Geman and Geman 1984). Work in this category is often motivated by classical studies in human visual perception (Palmer 1999), particularly the Gestalt movement which emphasized many non-local phenomena in perception. Region grouping or segmentation focuses on the association of similar regional elements. The dual approach, *curvilinear continuity*, focuses on exploiting the spatial continuity of bounding contours.
3. *High-level knowledge*: a recent trend in computer vision is to combine object knowledge with low- and mid-level processing (Borenstein and Ullman 2002; Tu et al. 2005; Yu et al. 2002). If by using shape or appearance-based templates, one could recognize objects in an image, it becomes relatively straightforward to locate the boundaries of the detected objects. High-level knowledge is a very powerful cue and quite effective where it applies. However, it is statistically and computationally expensive to model multiple object categories and poses. One may also argue that, as human subjects can easily detect boundaries of novel objects without explicitly recognizing them, boundary detection should primarily be an input to object recognition, not vice versa.

Our work falls in the category of mid-level approaches to boundary detection. We utilize curvilinear continuity, a well-studied contextual cue, to integrate local measurements of edge contrast. Our motivation is both to understand the extent to which curvilinear continuity informs boundary detection, if at all, and to build a mid-level boundary detector that achieves better performance than local approaches by exploiting this information.

### 1.1 Related Work on Curvilinear Continuity

The study of curvilinear continuity has a long and influential history in psychophysics as well as neurophysiology, starting with *good continuation*, described in Wertheimer's work (Wertheimer 1938); *illusory contours*, studied and popularized by Kanizsa (1979); the discovery of related neural mechanisms in V2 (von der Heydt et al. 1984); and the more recent theory of relatability (Kellman and Shipley 1991), which characterizes geometric conditions under which contour completion takes place.

In computer vision there also exists a rich literature on how to model curvilinear continuity (e.g., Shashua and Ullman 1988; Parent and Zucker 1989; Heitger and von der Heydt 1993; Mumford 1994; Williams and Jacobs 1997). A typical approach consists of detecting contour fragments based on brightness contrast and then linking up fragments using some measure of continuity. A popular family of approaches model continuity by a stochastic motion of tangents (Mumford 1994; Williams and Jacobs 1997; Thornber and Williams 2001) and are fairly successful in modeling illusory contours. More recent developments focus on finding salient *closed* contours (Elder and Zucker 1996; Williams and Thornber 1999; Jermyn and Ishikawa 2001) and make use of *closure* to locate the most salient cycle in an image.

Unfortunately, most existing approaches have only been demonstrated on a few synthetic or simple real images. Such limited experiments are far from convincing: while we may be able to complete low-contrast edges using continuity, spurious completions are often introduced in the process. Is the net effect positive or negative? This question can only be answered empirically by quantitative measurements. To the best of our knowledge, *no such measurements have been made on a large, diverse set of real-world images.* We expect this may explain why, despite the large amount of work on contour completion in the last 15 years, completion has not become a standard pre-processing step every time a boundary map is needed.

### 1.2 Our Approach

Central to our approach is the use of human-segmented images of natural scenes. Three such datasets are utilized in this work: a set of 30 news photos of baseball players (Mori et al. 2004), 344 images of horses (Borenstein and Ullman 2002) and the *Berkeley Segmentation Dataset* (Martin et al. 2002), containing 1000 images of various natural scenes. These datasets are used for two distinctive purposes:

1. To gain insights about the problem: we study the statistical properties of human-marked boundaries in these datasets. Empirical work along this line has confirmed the ecological validity of curvilinear continuity (Geisler et al. 2001; Ren and Malik 2002; Elder and Goldberg 2002). An important finding from our study is that the statistics of boundary contours obey power-law distributions suggestive of *scale invariance* (see Sect. 2).
2. To quantitatively evaluate models of continuity: we use human-marked ground-truth boundaries to train the models, and evaluate them on held out test data. We measure precision and recall of the models and compare them to local edge operators.

The contribution of this paper is a probabilistic model of continuity and closure built on a scale-invariant geometric structure which yields boundary estimates that are quantifiably better than the input provided by a local detector. Our algorithm starts with local estimates of the probability of boundary (*Pb*) based on brightness and texture gradients (Martin et al. 2004). The resulting edge map is discretized into a set of piecewise linear segments and potential

completions are generated utilizing the *constrained Delaunay triangulation*, as described in Sect. 3. We show how to carry out this construction in a scale-invariant manner and that the resulting geometric structure empirically captures a vast majority of the true image boundaries, while typically reducing the complexity from hundreds of thousands of pixels to a thousand candidate line segments.

In Sect. 4 we develop two models of curvilinear continuity on top of this geometric structure: a simple local model which classifies each pair of neighboring edges independently and a global model which uses a *conditional random field* (Lafferty et al. 2001) in order to enforce both continuity and long-range probabilistic constraints on junctions. Performance evaluations of both the local and global models are presented in Sect. 5 along with exemplary results.[1]

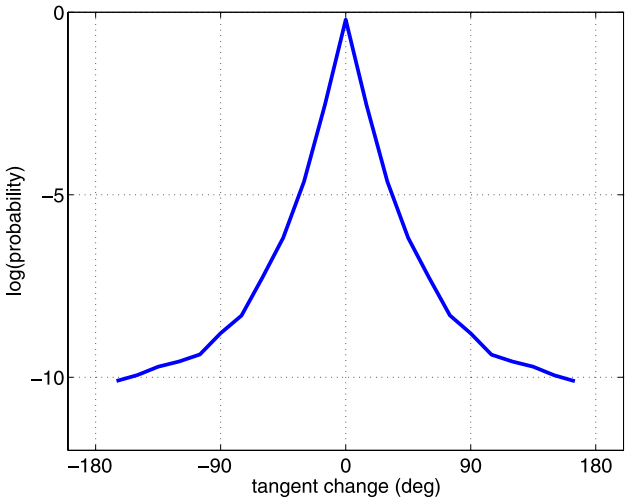



**Fig. 1** The empirical distribution of tangent change between two adjacent locations along a contour $C$. This distribution is sharply peaked around zero and has a heavy tail, suggesting that boundaries are mostly smooth but make occasional sharp turns

## 2 Empirical Studies of Contours in Natural Images

We begin with a study of the statistical properties of boundary contour shape in natural scenes. For this purpose, we use the Berkeley Segmentation Dataset (Martin et al. 2002) which consists of 1000 images of resolution 320-by-480 from the Corel stock photo database, covering a wide variety of natural scenes containing humans and other animals, landscapes and architecture.

### 2.1 Testing the Markov Model

An influential probabilistic model of contours is the first order Markov model explored in Mumford (1994), Williams and Jacobs (1997). Suppose a contour $C$ is parameterized by arc-length $s$, then if its curvature $\kappa(s)$ is random and independent from location to location, the tangent direction $\mathbf{t}(s)$ is a random walk. The Markov property states that the distribution of the tangent $\mathbf{t}(s+1)$ at the next location $s+1$ is dependent only on $\mathbf{t}(s)$. This model is appealing since the Markov property enables efficient computation (i.e. dynamic programming) (Williams and Jacobs 1997; Zhu 1999; Ren and Malik 2002).

To empirically study contours in natural images, we trace out human marked segment boundaries in the dataset and numerically estimate the derivatives of each resulting contour $C$.[2] Figure 1 shows the empirical distribution of the tangent angle change, $P(\mathbf{t}(s+1) - \mathbf{t}(s))$. This distribution exhibits two non-Gaussian properties often observed in natural images: (1) the distribution is sharply peaked around zero, i.e. in most cases boundaries are smooth; and (2) the distribution has a heavy tail, i.e. occasionally boundaries do make sudden turns.[3] This is consistent with empirical findings in other related studies (Zhu 1999; Elder and Goldberg 2002).

In the Markov model, curvature $\kappa(s)$ is assumed to be independent from location to location. As $s$ increases, at each step there is a constant probability $p$ that a high curvature event $\{\kappa(s) > \kappa_0\}$ occurs. Therefore, if we look at a contour segment $L$ between two high curvature locations, the waiting time between these two independent events is given by the length $|L|$ and should obey an exponential distribution: $P(|L| = t) = p(1-p)^t$.

This observation leads to an empirical test of the validity of the Markov contour model. We estimate the curvature along a contour $C$ and break it at locations where the curvature is above a threshold and is at a local maximum. Thus a contour $C$ is decomposed into segments where each contour segment $L$ is smooth. Figure 2 shows a few examples of this decomposition.

Figure 3 shows the empirical distribution of contour segment length $|L|$, on a log-log scale in (a) and a semi-log scale in (b). We fit a line in both cases, for a range of lengths between 10 and 200 pixels. The $r^2$ statistic, a measure of the goodness of fit, is 0.9917 for the power law fit, and 0.9391 for the exponential fit.

Limited by image resolution, the empirical data we study spans only one order of magnitude. Nevertheless, a power law fits the data almost perfectly,[4] much better than an ex-

[1] The software and datasets used in this work have been made available through the following link: http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/ijcv2006/.

[2] We apply a simple 1-D Gaussian smoothing along $C$ before estimating derivatives.

[3] Although there are man-made objects in the dataset, there seems to be no special role associated with 90-degree angles.

[4] Deviations from this power law at fine scales may be attributed to the resolution limits of the pixel grid and patience of human segmenters. Deviations at large scales are inevitable due to finite image size.

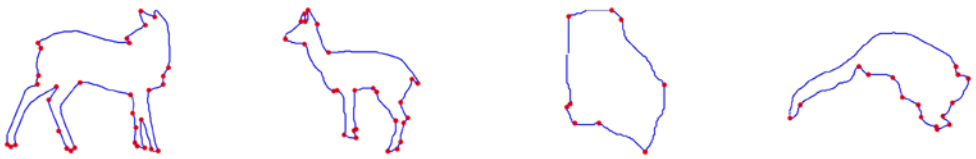


**Fig. 2** We take each boundary contour $C$ and break it up at local curvature maxima (corners). Shown are a few examples of this decomposition. The contour segment length $|L|$ is large for large-scale features such as the back of an animal, and is small in the vicinity of fine-scale details such as the head

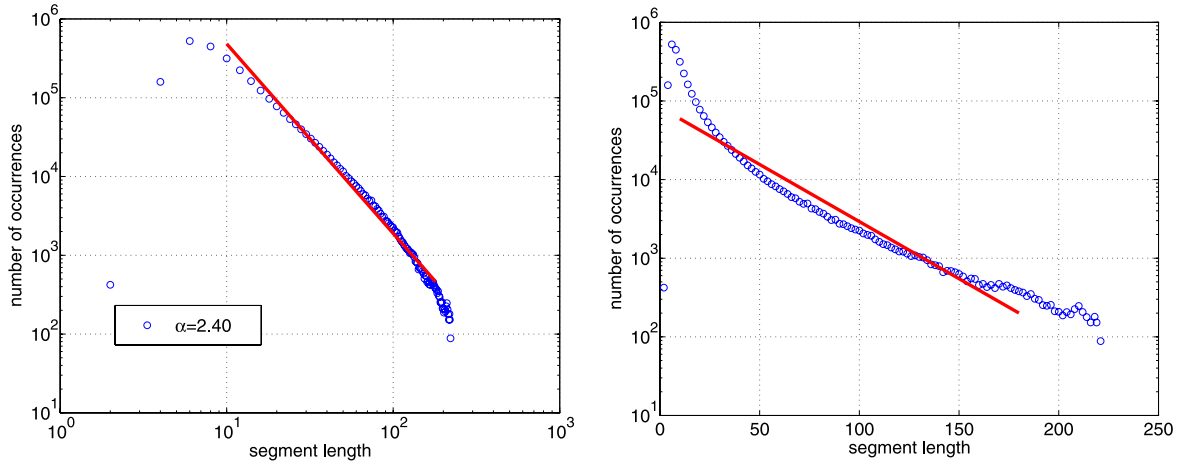



**Fig. 3** Empirical distributions of approximately straight contour segment length $|L|$. (**a**) The marginal distribution of $|L|$ on a log-log scale. It closely follows a power law with $\alpha = 2.40$, in direct contradiction with the Markov model that predicts an exponential distribution. (**b**) As a comparison, the same distribution on a semi-log scale and an exponential fit. The power law fits the data much better than the exponential

ponential distribution. This is a fairly robust result: if we change the smoothing and/or the curvature threshold, the exponent may drift a bit, but the power law still holds. This is in direct contradiction with the exponential prediction under the Markov model. We therefore conclude that the Markov assumption is empirically *false*.

In some ways, the failure of the first-order Markov contour model for natural images is not so surprising since the model cannot enforce long range structure. In particular, straight lines are very *uncommon* in the Markov model, as they require tangent changes to be small for a long period of time. Typical contours sampled from this model are random walks and lack the large-scale structures evident in natural contours.

## 2.2 Sources of Multi-scale Structure

Power law distributions are typically associated with *scale-invariance* phenomena, and have been discovered in many studies of natural image statistics. For example, the power spectra and wavelet coefficients of natural images typically follow a power law distribution (Ruderman and Bialek 1994; Huang and Mumford 1999), as do geometric quantities such as the area of homogeneous regions (Alvarez et al. 1999). These power laws have confirmed the near scale-invariance of natural images and have inspired many efforts in developing scale-invariant statistical models for generic images (e.g. Ruderman 1997; Lee et al. 2001). The power law in Fig. 3 is no exception: it suggests that contours are multi-scale in nature.

Introspection suggests there are at least two potential sources of multi-scale structure in natural scenes:

*Arbitrary viewing distance* an object may appear at any distance to the observer. Figure 4(a) shows a line-drawing illustration of an object viewed at two distances, hence appearing at two different scales.

*Hierarchy of parts in a single object* objects in the natural world are themselves multi-scale as they consist of parts at various scales. For example, Fig. 4(b) shows the boundary of a human body which has a hierarchy of parts: head, torso, arms, legs, fingers, hair, each with its own distinct scale.

Arbitrary viewing distance implies that the collection of all natural images should exhibit invariant statistics over a continuum of scales. However, this imposes few constraints on the appearance of any specific image. In contrast, hierarchical or "fractal-like" structure typically involves a discrete

set of relevant scales rather than a continuum, but can exist within a single image.

If scaling behavior is due solely to arbitrariness in viewing distance (e.g. Ruderman 1997), a nearby object would have smoother boundaries and therefore a higher probability for long straight segments $|L|$, while a small, far away object would have more high-curvature points along its boundary and shorter segments. We can test this prediction by using $|C|$, the length of the entire boundary, as an indication of the object scale. In Fig. 5(a), we show the empirical distributions of $|L|$, where the contour segments are divided into four groups, from small $|C|$ to large $|C|$. We find that the distributions are almost identical; there is no indication that larger $|C|$ leads to significantly higher probability for large $|L|$. A similar phenomenon is observed in the case of contour curvature, as shown in Fig. 5(b).

We take these results as a strong evidence that arbitrariness of viewing distance is not the sole source of scale-invariance; the multi-scale nature of individual objects is also important. While scale invariance has been studied mostly as a phenomenon in large ensembles of images, properties that hold for individual objects have more immediate implications in practice. Any model of boundaries should clearly take both factors into consideration.

(a) (b)

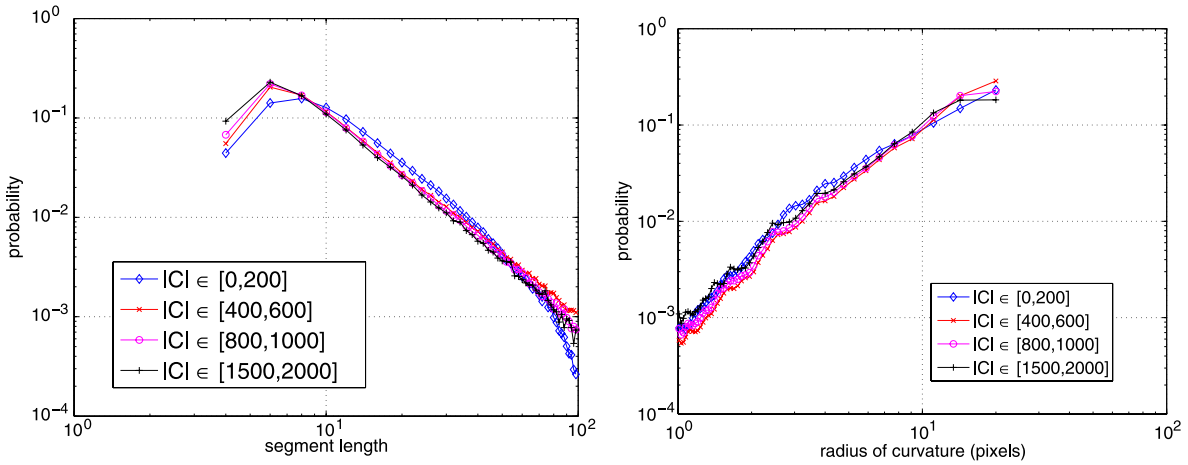
**Fig. 4** Two sources of multi-scale structure in natural scenes: (**a**) Imaging geometry is free to vary so the same object can appear at any distance from the observer, resulting in scaled versions appearing in the image. (**b**) Objects consist of a collection of parts which have different characteristic scales. Here the torso, leg, head and fingers consist of roughly parallel boundaries at vastly different scales. In either case, it is desirable that vision algorithms adapt to the appropriate scale at each image location

## 3 A Scale-Invariant Representation of Contours

The empirical studies in the previous section provide guidance in designing a model of contour completion. Since contour segments show non-Markov, multi-scale structure, an appropriate model for the contours of an object should allow for long range dependencies that simultaneously capture both large and small scale features. Furthermore, since that object can appear at any distance from the camera, these connections should co-vary with the scale of the object in the scene. Similarly, the representation of contours in an entire scene should be independent of image resolution (at least over a range of scales where this is possible).

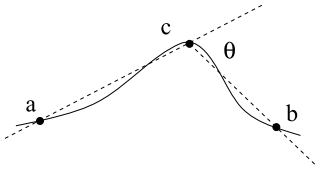
### 3.1 Decomposition into Linear Segments

Consider the contour decomposition shown in Fig. 2. Here contours are decomposed in such a way that each contour segment is approximately straight and the collection

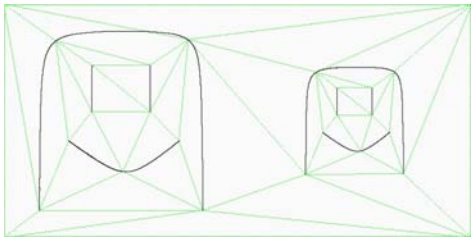



(a) (b)

**Fig. 5** We use $|C|$, the length of an entire boundary, as an indication of the object scale. (**a**) The distributions of smooth segment lengths $|L|$ conditioned on $|C|$. If *arbitrary viewing distance* were the only source of scale-invariance, larger $|C|$ would be correlated with smoother contours and higher probabilities of large $|L|$. Instead, we find that the empirical distributions are very similar for different ranges of $|C|$, showing no such correlation. This indicates that multi-scale nature of *individual* objects also plays an important role. In (**b**) a similar phenomenon is observed in the conditional distributions of $R = 1/\kappa$, the radius of curvature. The distributions of $R$ are also very similar for different ranges of $|C|$




(a)



(b)

**Fig. 6** Building a discrete graph structure from bottom-up. (**a**) We take a contour detected by a local edge operator *Pb* and recursively split it until the angle $\theta$ (a scale-invariant quantity) is below a threshold. (**b**) We complete this piecewise linear structure with *constrained Delaunay triangulation*. *G*-edges (gradient edges detected by *Pb*) are in *black* and *C*-edges (completed by CDT) in *dashed/green*. The resulting triangulation graph is an efficient representation of the image and covaries with the object scale, as illustrated in this example: the internal triangulation of the two objects (in particular the "true" completions) are the same, regardless of the scale change

of straight line approximations efficiently describes different scale structures within the depicted objects. Large-scale features (smooth contour segments), such as the back of an animal, are conserved and not broken into fragments. On the other hand, fine-scale parts are further sub-divided; hence the "sampling density" is high in places where local curvature maxima abound. Interestingly, this is consistent with the classical theory that information along contours is concentrated in regions of high magnitude of curvature (Attneave 1954).

While such a decomposition captures the multi-scale structure within an object, it does not covary with the viewing distance as it depends on curvature (measured in camera coordinates). However, replacing curvature with a truly scale-invariant quantity, e.g. turning angle, yields a representation that satisfies our criteria above.

We construct our contour segment based representation as follows: given an input image, first we use the local *Pb* operator (Martin et al. 2004) to detect low-level edges. Canny's hysteresis thresholding is used to trace out the detected contours. We set the threshold fairly low such that most edges with low-level contrast are retained. We then recursively split these contours until each segment is approximately straight.

Figure 6(a) shows an illustration of this splitting procedure: for a given contour, let $\theta$ be the angle between segments $\overline{ca}$ and $\overline{cb}$. Pick the set of points $\{a, b, c\}$, in a coarse-to-fine search, such that the angle $\theta$ is maximized. If $\theta$ is above a threshold, split the contour by adding a vertex at $c$ and continue.[5] In our implementation, we fix $a$ and $b$ to be the two end points of the curve, and vary the point $c$ to maximize $\theta$.

[5]Due to the non-maximum suppression in *Pb*, edges near T-junctions are sometimes missing. A heuristic is added to handle T-junctions: when a vertex is very close to another line (relative to line length), we split this line and insert an additional vertex.

In general, this angle-based decomposition could be "unstable" when processing a curve with near constant curvature (i.e. an arc of a circle) close to the threshold. However, as we have observed in the empirical statistics, contours in natural images tend to have sharp turns or corners which anchor the decomposition and make it stable under normal circumstances.

### 3.2 Completion with Constrained Delaunay Triangulation

In real images, a local edge operator may miss some low-contrast contours, leaving gaps in our representation. We use the *constrained Delaunay triangulation* algorithm to span these gaps, completing the piecewise linear approximation. The *Delaunay triangulation* (DT) is the dual of the Voronoi diagram and is the unique triangulation of a set of vertices in the plane such that no vertex is inside the circumcircle of any triangle. The constrained Delaunay triangulation (CDT) is a variant of the DT in which a set of user-specified edges must lie in the triangulation. The CDT retains many nice properties of DT and is widely used in geometric modeling and finite element analysis. It was used to find convex contours in Huttenlocher and Wayner (1992) and also recently applied to image segmentation by Wu and Yu (2003).

In the context of contour completion, the CDT's ability to complete contours across gaps in the local detector output is of particular interest. A typical scenario of contour completion is one low-contrast contour segment (missed by *Pb*) in between two high-contrast segments (both detected by *Pb*). If the low-contrast segment is short in length, it is likely that no other vertices lie in the circumcircle and CDT will correctly complete the gap by connecting the two high-contrast segments. This is closely related to the medial axis and ligature analysis used in August et al. (1999). We show some example completions in Fig. 7.

We used the TRIANGLE program (Shewchuk 1996) to produce CDTs as shown in Fig. 7. The linearized edges extracted from the *Pb* contours, which we refer to as **gradient**

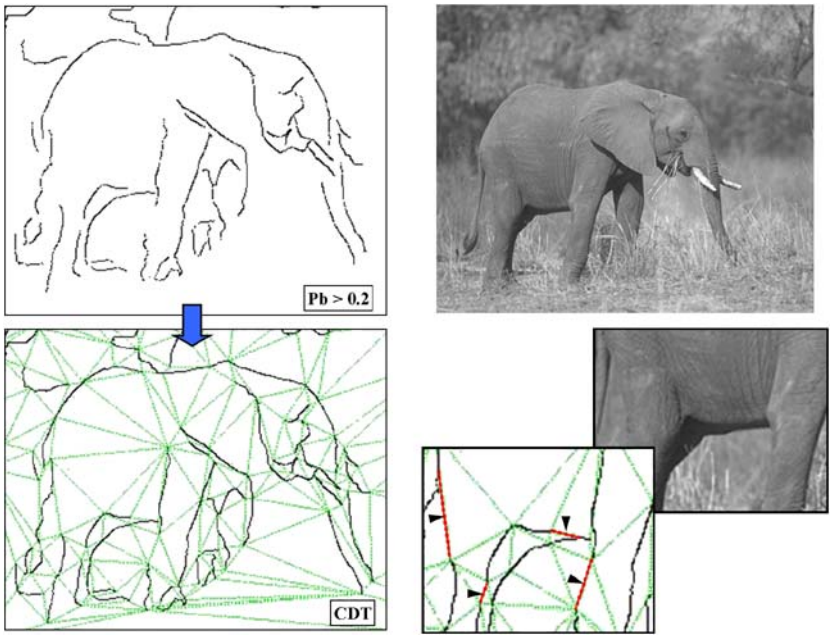



**Fig. 7** An example of the CDT completion. $G$-edges are in *black* and $C$-edges in *dashed/green*. Note how the CDT manages to complete gaps on the front legs of the elephant (*red* highlight on the inset at *right*). These gaps are commonly formed when an object contour passes in front of a background whose appearance (brightness/texture) is similar to that of the object

**edges** or $G$**-edges**, become constrained edges in the triangulation. The remaining **completion edges** or $C$**-edges** are filled in by the CDT. The process also partitions the image into disjoint triangular regions.

Figure 6(b) shows the output triangulations for our toy example: although the two views of the "object" greatly differ in scale, the line approximations of their contours are almost identical and so are the triangulations. This invariance leads to interesting applications of this representation for object recognition (Ren et al. 2005; Ren 2007). We note, however, that the edge detector we use, *Pb*, is not scale invariant. The computed edge maps may differ at different image resolutions; hence the scale invariance of our representation is only approximate over some range of scales.

### 3.3 Using the Triangulation Graph

In the remainder of the paper, we will describe models of curvilinear continuity built on top of the triangulation graph. The CDT graph has many advantages over using the pixel grid:

- It is (approximately) invariant to image resolution and object scale, while efficiently representing multi-scale structure within individual objects (i.e. both long straight lines and fine-scale details).
- By moving from $\approx$100,000 pixels to $\approx$1000 edges, it achieves high computational efficiency, independent of image resolution.
- By using contour segments (instead of pixels) as tokens, the CDT graph allows modeling of longer-range interactions.
- By restricting completions to the edges in the graph, it partially solves the problem of having too many spurious completions.

The triangulation graph is an image-dependent structure built from low-level signals. It is far more concise than generic, image-independent multi-scale representations (e.g. multi-scale Markov random fields Leuttgen et al. 1993; Li 1995 or Beamlets Donoho and Huo 2002). For example, existing multi-scale contour grouping algorithms (Sharon et al. 2000; Ren and Malik 2002) operate on a pyramid of fixed pixel grids and combine completions across scales. In comparison, the triangulation graph provides a single mid-level structure which makes inference (and learning) much more efficient; no propagation of information is needed inside each edgel. It also allows us to explicitly represent junction geometry and connectivity.

Before we proceed, a natural question to ask is how good the triangulation graph is as an approximation of the input image? How much important image structure is lost in the triangulation process? To answer this, as well as perform quantitative training and testing, we first need to establish a technique for transferring our ground-truth human segmentations from the pixel grid onto the triangulated graph.

For training, we set binary ground-truth labels on CDT edges by running a maximum-cardinality bipartite matching between the human marked boundaries and the CDT edges with a fixed distance threshold. We label a CDT edge as boundary if 75% of the pixels lying under the edge are matched to human boundaries; otherwise we label it as non-boundary. For testing, we always project the boundary estimate of CDT edges back to the pixel-grid, allowing direct comparison with traditional edge detectors.

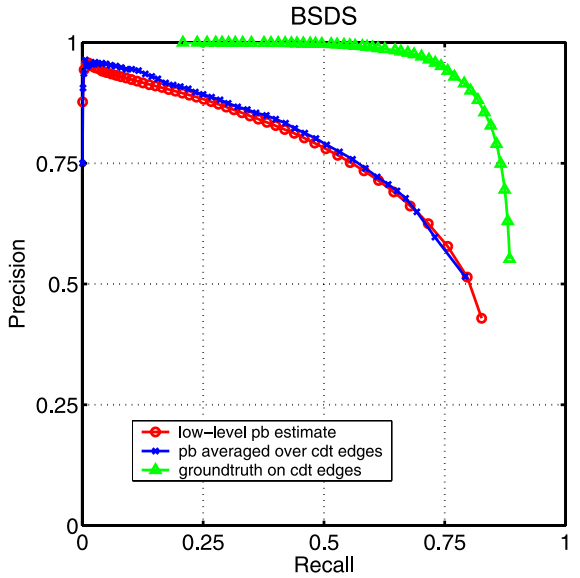



**Fig. 8** This Precision-Recall curve verifies that moving from pixels to the CDT completion doesn't give up any boundaries found by the local measurement. For comparison, we also show an upper-bound curve given by assigning labels to the CDT edges using ground-truth. The upper bound curve has a precision near 1 even at high recall and achieves a greater asymptotic recall than the local boundary detector, indicating it is completing some gradientless gaps

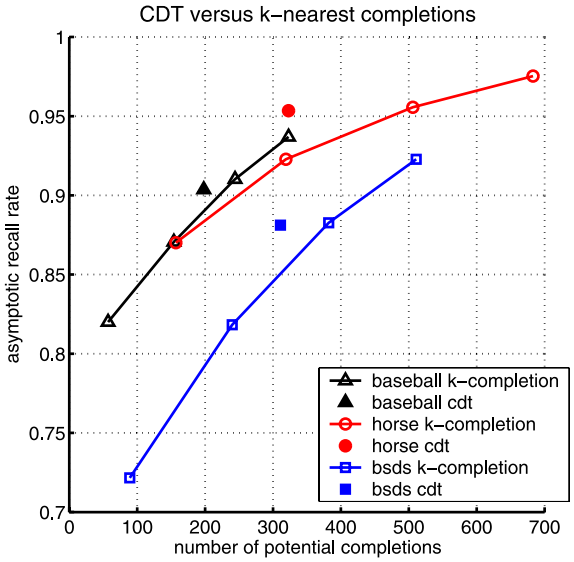



**Fig. 9** This figure shows the relative merits of the CDT versus an alternate completion scheme based on connecting each vertex to the $k$-nearest visible vertices for $k = \{1, 3, 5, 7\}$. The plot shows the asymptotic recall rate (the number of illusory contours found) versus the average number of potential completions which need to be considered. An ideal algorithm would achieve asymptotic recall of 1 with very few potential completions. The *single filled marker* shows performance of the CDT based completion while the curve shows performance over a range of choices of $k$. For each dataset, we find that the CDT based completion gives better recall rate at a given number of potential completions the than $k$-nearest visible neighbor algorithm

Throughout the rest of this paper, performance is evaluated with using a **precision-recall curve** which shows the trade-off between false positives and missed detections. For each given thresholding $t$ of the algorithm output, boundary points above threshold are matched to human-marked boundaries $H$ and the precision $P = P(H(x, y) = 1|Pb(x, y) > t)$ and recall $R = P(Pb(x, y) > t|H(x, y) = 1)$ are recorded (see Martin et al. 2004 for more discussion).

Using this performance measure, we can show empirically that very little of the true boundary structure is lost in this discretization process. Figure 8 gives the performance of the local boundary detector *Pb* as well as the performance when we assign the average underlying *Pb* to each CDT edge. In addition, we show an upper-bound curve by assigning soft ground-truth labels to the CDT edges: similar to the evaluation process, we run a bipartite matching between ground-truth boundary pixels and pixels on the CDT edges. The soft ground-truth labels are the percentage of pixels on the CDT edges that are matched to human marked pixels. This gives the "best" possible boundary labeling constrained by the CDT representation.

These empirical results show that assigning a single probability of boundary to an entire CDT edge does not hurt performance and that the CDT graph contains most edges found by *Pb*. Moreover, the precision of the upper-bound curve is close to 100% even at high-recall, which clearly indicates that little image structure is lost in the triangulation process. The gap between the asymptotic recall of *Pb* and the upper-bound curve shows that the CDT is completing some contours that are completely illusory (i.e. contours with no local evidence).

### 3.4 Alternative Completion Schemes

The CDT is not the only scale-invariant scheme for generating completions from a set of detected edgels. One likely alternative is to connect each vertex with up to $k$ of its nearest neighbors subject to the visibility constraint that added completion edges do not cross gradient edges detected locally. We refer to this as $k$-nearest visible neighbor ($k$-NVN) completion.

As $k$ increases, chances are higher that boundary gaps are correctly completed, at the cost of more spurious completions. To quantitatively compare these completion schemes, we test them on the human-annotated datasets, estimating two quantities for each scheme: $\bar{m}$, average number of completions per image; and $R_0$, the asymptotic recall rate (percentage of ground-truth boundary pixels being detected if all the edges are turned on). A good completion scheme should have a small $\bar{m}$ and a high $R_0$. The results are shown in Fig. 9. For all three datasets, we find that the CDT based completion is more efficient than k-NVN. In addition, the CDT yields a planar graph which is conceptually simple to work with.

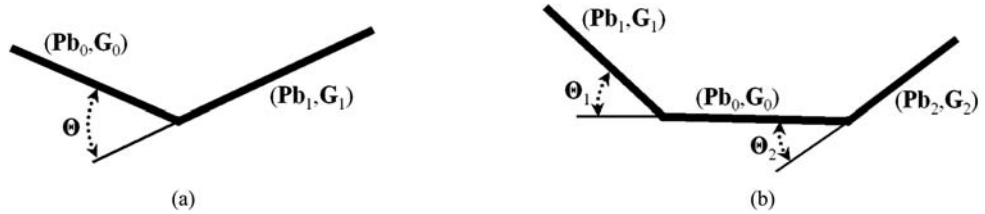



**Fig. 10** (**a**) A simple 2-edge model of local curvilinear continuity. Each edge has associated features: the average $Pb$ and whether it is a $G$-edge or $C$-edge. Continuity is measured by the angle $\theta$. The edge pair is either classified as a continuation ($X_{e_0} = 1, X_{e_1} = 1$) or not. In the 2-edge product model, (**b**), evidence of continuity come from both ends of edge $e_0$. The new "probability of boundary" for $e_0$ is the product of the 2-edge model on both pairs $(e_0, e_1)$ and $(e_0, e_2)$

## 4 Modeling Curvilinear Continuity

In the previous section we described how to construct a discrete, scale-invariant structure, the CDT graph, from low-level edges. We will associate a random variable $X_e$ to every edge $e$ in the CDT graph, where $X_e = 1$ if $e$ corresponds to a true boundary contour and $X_e = 0$ otherwise. The variables $\{X_e\}$ interact with each other through vertices or junctions in the graph. Our goal is to build a probabilistic model of continuity and junction frequency on the CDT graph and make inferences about $\{X_e\}$.

We will introduce two models: one for local continuity and one for global continuity and junctions. The local continuity model considers only the immediate neighbors of each edge $e$, estimating the corresponding $X_e$ independently based on its local context. This serves as a baseline model and is a convenient place to study the relevant features (which are all defined locally). The global continuity model uses *conditional random fields* in order to build a joint probabilistic model over the collection of all edges in the CDT. We fit model parameters and estimate the marginals of the random field on $X_e$ using loopy belief propagation.

### 4.1 A Baseline Model of Local Continuity

Each edge $e$ in a CDT graph is naturally associated with a set of features including the average $Pb$ estimated for pixels along the edge $e$ and whether it is a $G$-edge (detected in $Pb$) or $C$-edge (completed by the CDT). The local context of $e$ includes these features and those of neighboring edges in the CDT graph.

Consider the simplest case of context: a pair of connected edges (shown in Fig. 10(a)). Each edge can be declared a true boundary or not, yielding four possible labellings of the pair. We observe that the ground-truth contours in our datasets are almost always closed; line endings and junctions are rare. Therefore we make the simplifying assumption that there are only two possible labellings: either both edgels are on, or both are off. Modeling local continuity thus becomes a binary classification problem.

Our best local model uses as features $\overline{Pb}$, the average $Pb$ over the pair of edges; $G$, an indicator variable whether both of the edges are $G$-edges, and $\theta$, the angle formed at the connection of the pair. We use logistic regression to fit a linear model to these features. We found that logistic regression performed as well as other classifiers (we also tested support vector machines and hierarchical mixture of experts). It has the advantage of being computationally efficient and simple to interpret.

To apply the local continuity model, we use the classifier to assign a new "probability of boundary" value to each edge $e$. Consider Fig. 10(b): evidence of continuity comes from both ends of an edge $e_0$, as a contour at $e_0$ would have to continue in both directions. We assume that these two sources of information are independent and take a product of their probabilities. Recall that $X_e = 1$ if the pixels corresponding to $e$ lie on a true boundary and 0 otherwise. The logistic model gives an estimate of $P(X_{e_0} = 1, X_{e_1} = 1)$, the posterior probability that the pair of edges $(e_0, e_1)$ are both true. Let $S_1$ and $S_2$ be the two sets of edges connecting to $e_0$ at the two ends respectively. We define the new boundary operator $Pb_L$ under the 2-edge product model to be:

$$Pb_L = \max_{e_1 \in S_1} P(X_{e_0} = 1, X_{e_1} = 1) \cdot \max_{e_2 \in S_2} P(X_{e_0} = 1, X_{e_2} = 1).$$

We experimented with various additional features, such as: maximum $Pb$ and minimum $Pb$ of the edge pair, length of the edges, length of $C$-edges, region-based similarity (i.e. brightness and texture similarity between the regions on the two sides of the edge pair), and whether one or both edges have an alternative continuation that is better in terms of $\theta$. These features gave at most a marginal improvement in performance. We do not consider them further in the remainder of this paper.

We also considered several classifier variants such as: a second-layer classifier, trained to combine information from the two ends of $e_0$; a 3-edge classifier which directly takes as input the features from triples of edges, hence removing the independence assumption between two pairs; and simultaneous 4-way classification on pairs of edges. The simplest 2-edge product model described above performed as well as any of these variants.

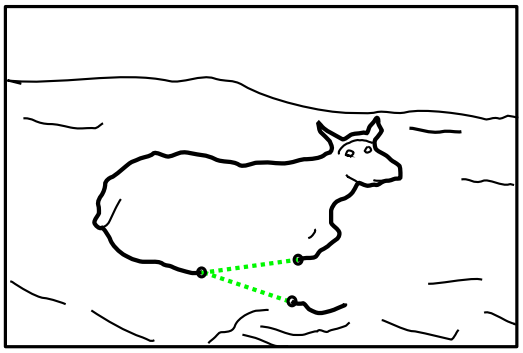


**Fig. 11** Local Ambiguity: A scenario in which completion cannot be resolved locally. The edge detector reports the same contrast for the boundary of a closed figure and some background clutter. The local continuity model can correctly suppress isolated clutter but potential completions (*two dashed/green lines*) cannot always be distinguished locally. Our global model builds a random field over the set of potential edges which allows for interdependence of neighboring edge decisions

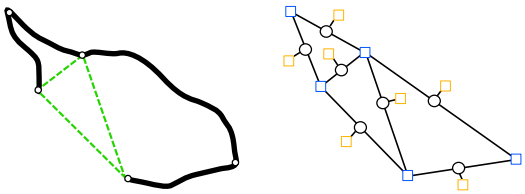


**Fig. 12** The factor graph representing our conditional random field model for global continuity. The CDT provides both the geometric and graphical structure for our model. Each edge in the CDT corresponds to a variable $X_e$ (represented by *circles*) which is 1 if pixels corresponding to edge $e$ constitute a true boundary. Potential functions (*squares*) consist of a singleton potential for each edge which encodes the underlying *Pb* measurement and continuity/closure potentials at the vertices whose values are dependent on both the angles between incoming line segments and the numbers of $C$- and $G$-edges entering a junction. Note that unlike the graphical models for typical MRF models on a lattice where potentials correspond to graph edges, this random field includes potentials involving more than two variables. The factor graph is a bipartite representation of the resulting hypergraph (see Kschischang et al. 2001)

## 4.2 A Random Field Model of Global Continuity

The local model described in the previous section makes an independent decision for each edge. This model has the short-coming that the decision of whether an edge $e$ is turned on or off is only dependent on the local pattern of gradients but does not depend on whether some neighboring edge, $e'$, is turned on. As a result, information about boundary probability cannot be propagated between CDT edges. Figure 11 gives an example where completion based on such local decisions is ambiguous. In this section we describe a global probabilistic model that is based on the same image measurements as the local model ($Pb$,$G$,$\theta$) but also captures the statistical dependencies between neighboring boundary variables.

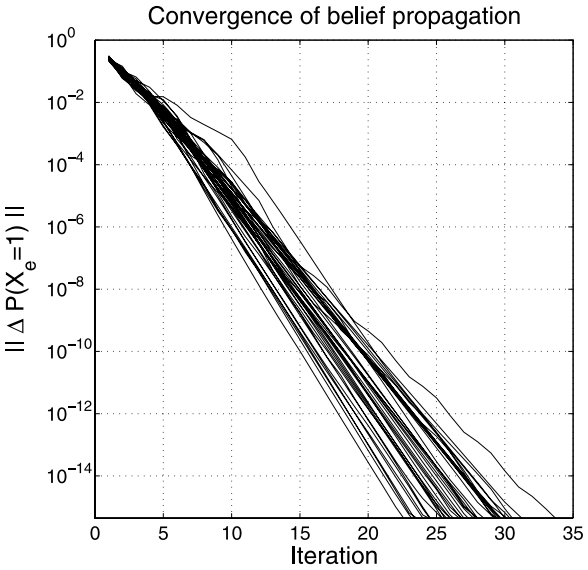



**Fig. 13** Convergence of Loopy Propagation on CDT Graphs: Here we show the change in marginal beliefs versus iteration for belief propagation on the BSDS. Each curve shows the maximum absolute change over all edge marginals for a given image ($L_\infty$ norm). The marginals quickly converge, reaching acceptable accuracies in 10 iterations. Convergence behavior was similarly uniform across the baseball and horse dataset

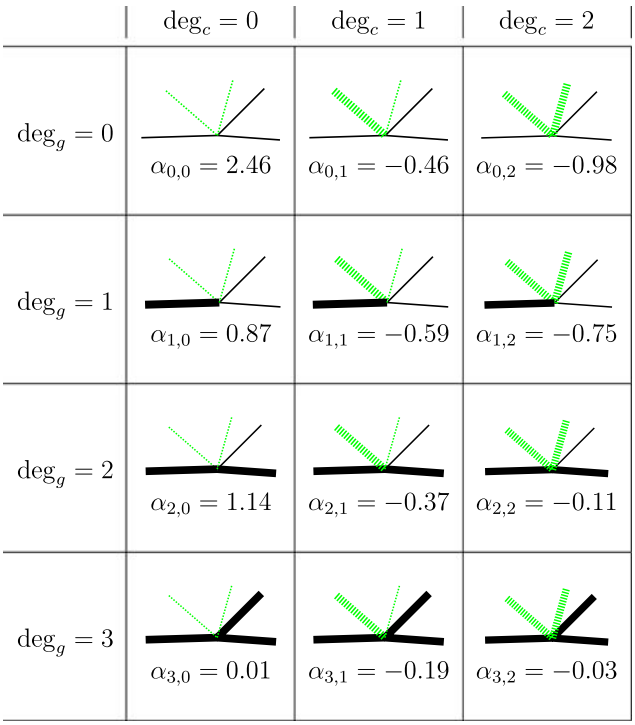



**Fig. 14** Learning about junctions: a weight is associated with each junction type $(\deg_g, \deg_c)$, where $\deg_g$ is the number of $G$-edges (gradient edges, shown in *black*) being turned on (shown as *thick lines*), and $\deg_c$ the number of $C$-edges (completion edges, shown in *dashed/green*) being turned on. $\deg_g + \deg_c = 1$ corresponds to the case of a line ending; $\deg_g + \deg_c = 2$ a continuation of contour; and $\deg_g + \deg_c = 3$ a T-junction. Compare with hand set potentials of Geman and Geman (1984)

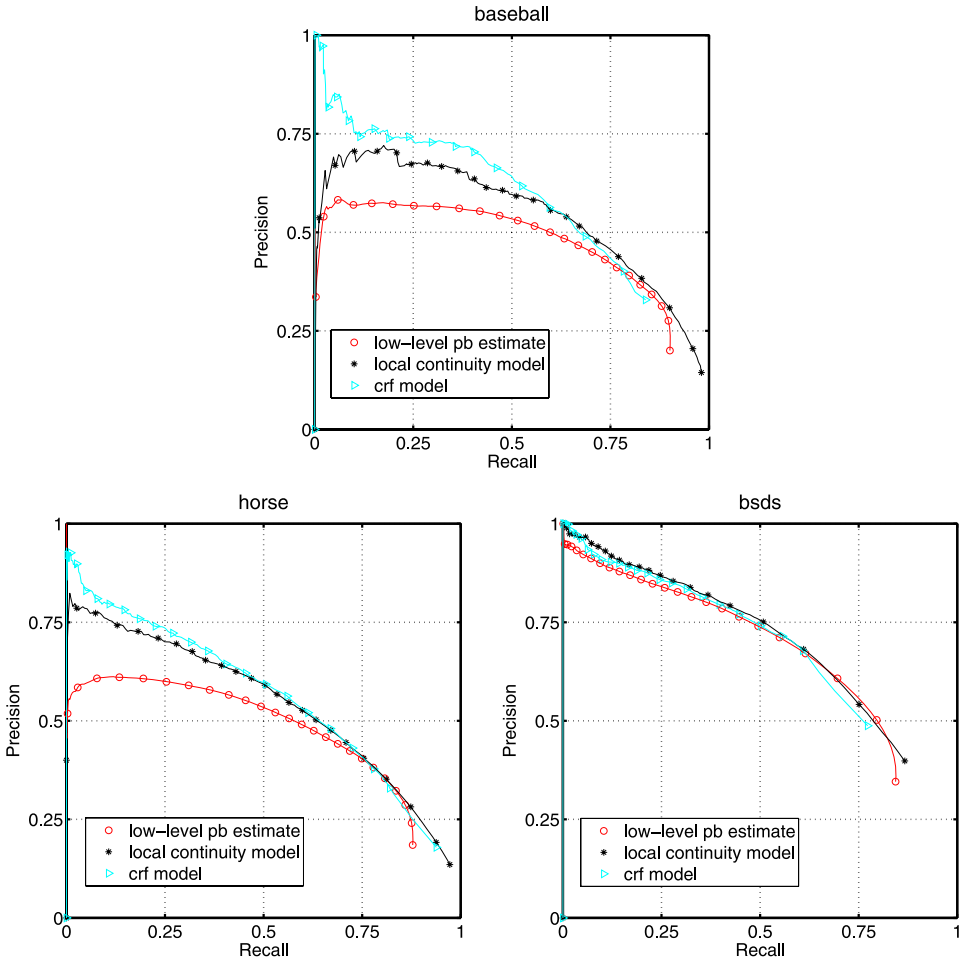



**Fig. 15** Pixel-based precision-recall evaluations comparing the local classifier ($Pb_L$), global CRF ($Pb_G$) and raw $Pb$. Both techniques improve boundary detection on all three datasets and the overall ordering of the curves is generally preserved

A natural probabilistic structure for modeling spatial dependencies is the Markov Random Field (MRF) in which random variables, representing for example true pixel gray values, are independent conditioned on some immediate spatial neighborhood (Besag 1974). In the computer vision literature, there is a large body of research applying MRF models to vision problems such as texture, segmentation, de-noising, super-resolution and stereo (see for example Geman and Geman 1984; Li 1995; Zhu et al. 1998; Freeman et al. 2000).

We use a variant on the MRF, introduced by Lafferty et al. (2001), which is referred to as a conditional random field (CRF). Unlike generative MRF models that attempt to capture the joint distribution of image measurements and hidden labels, a CRF focus directly on the conditional distribution of labels given the observations. One key advantage from our perspective is that the observed variables need not be conditionally independent given the hidden variables. This allows much greater freedom in choosing model features.

Conditional random fields have been considered as a method for image segmentation by Kumar and Hebert (2006), Shental et al. (2003), He et al. (2004), however, the focus has been on pixel-level labeling which is tied directly to the image resolution. Moving to scale free, mid-level tokens is a key ingredient to handling objects at a continuum of scales. For example, the multi-scale approach of He et al. (2004) uses features at three different scales relative to the image, but the features are different for each scale. To make such a model scale-invariant requires adding a set of features for every possible object scale and tying the parameters together across scales. In our case, because the CDT structure and input features are approximately scale-invariant, a simple random field suffices.

We base the independence structure of our edge variables on the topology given by the CDT. Recall the random variable $X_e$ which is 1 if the pixels corresponding to $e$ lie on a true boundary and 0 otherwise. Let $X_V$ be the collection of variables for all edges that intersect at a vertex $V$ of the CDT. We consider log-linear distributions over the collection of edges of the form

$$P(X|I,\Theta) = \frac{1}{Z(I,\Theta)} e^{-\{\sum_e \phi(X_e|I,\Theta)+\sum_V \psi(X_V|I,\Theta)\}}.$$

Note that the partition function

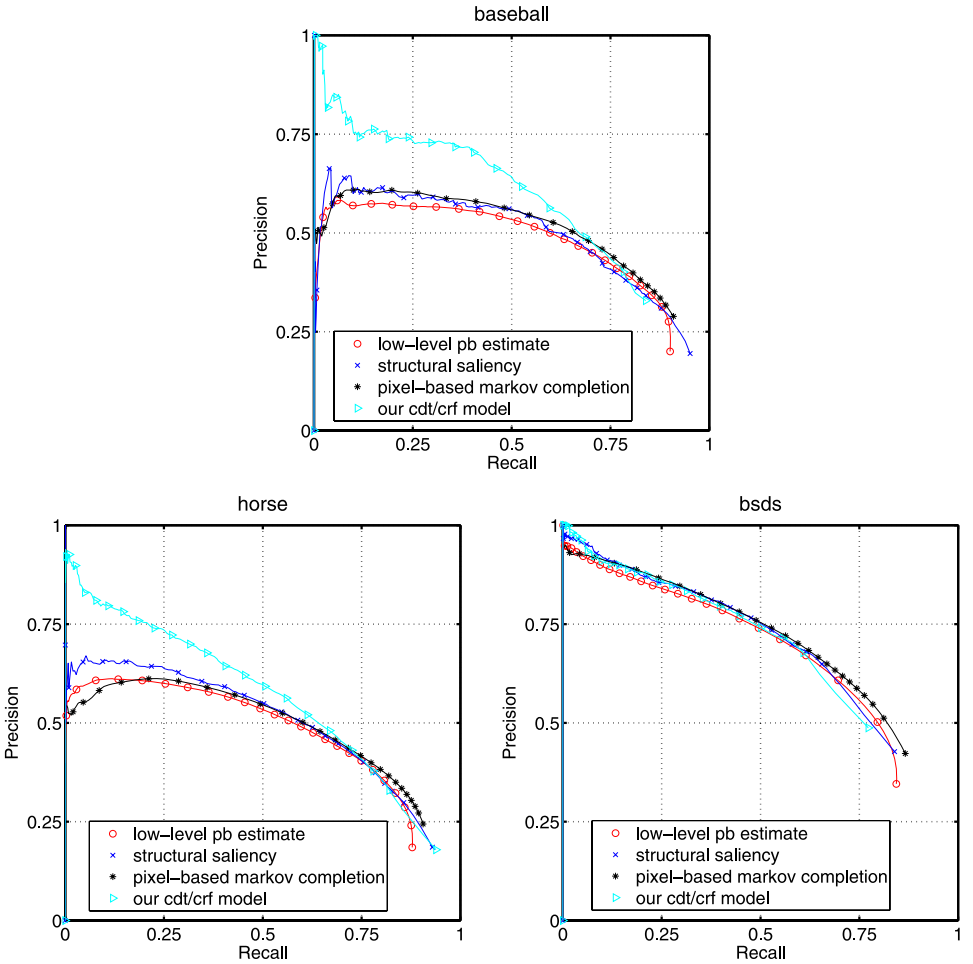



**Fig. 16** Comparing to two representative saliency algorithms: the Markov stochastic completion model of (Williams and Jacobs 1997), adapted to natural images as in Ren and Malik (2002), which operates on the pixel-grid; and the structural saliency algorithm of (Shashua and Ullman 1988), which we apply on the CDT edgels. Our CRF algorithm performs much better on the baseball and horse datasets; none of the algorithms stands out on the BSDS

$$Z(I,\Theta)=\sum_{X\in\mathcal{X}} e^{-\{\sum_e \phi(X_e|I,\Theta)+\sum_V \psi(X_V|I,\Theta)\}}$$

is dependent on the observed image $I$ as well as the collection of model parameters, $\Theta$. Figure 12 shows a CDT and the corresponding factor graph displaying the conditional independence structure in this random field.

The energy function $\phi$ captures the extent to which the image evidence $I$ supports the presence of a boundary under edge $e$. The edge energy function is given by

$$\phi(X_e|I,\Theta)=-\beta\log(Pb_e)X_e$$

where $Pb_e$ is the average $Pb$ recorded over the pixels corresponding to edge $e$. Continuity conditions at a junction between contour segments are described by the vertex energy

$$\psi(X_V|I,\Theta)=-\sum_{i,j}\alpha_{i,j}\mathbf{1}_{\{\deg_g=i,\deg_c=j\}} - \gamma\mathbf{1}_{\{\deg_g+\deg_c=2\}}f(\theta)$$

where $\deg_g$ is the number of $G$-edges at vertex $V$ which are turned on (i.e. $X_e=1$) and similarly $\deg_c$ is the number of $C$-edges which are turned on. $(\deg_g,\deg_c)$ indexes the "junction type" for every possible assignment of $X_V$: for example, a total degree of one, $\deg_g+\deg_c=1$, corresponds to a line ending, 2 is a continuation and 3 is a T-junction. The model associates a separate weight $\alpha_{i,j}$ to each possible junction type. Figure 14 shows examples of different junction types and their associated indices. Since high degree junctions do not occur in the training data, we do not consider configurations $X$ which result in junctions with $\deg_g>4$ or $\deg_c>4$.

When the total degree of a vertex is 2 (exactly two edges are turned on at a junction), $\gamma$ weights the continuity of the two edges. Let $\theta$ be the angle formed between them as in the local model (Fig. 10). To penalize sharp turns, $f$ is a smooth function which is 1 when $\theta=0$ and falls off to $-1$ as $\theta\to\pi$. If the angle $\theta$ between the two edges is close to 0, they form a good continuation, $-\gamma f(\theta)$ is small and they are more likely to both be turned on. We use the $f$ given by the logistic fit for 2-edge local continuity based on angle (as described in the previous section).

The resulting sum of these terms specifies an energy for each configuration of the edge variables (e.g. $X=\{1,0,0,1,1,\ldots\}\in\mathcal{X}$). Low energy (high probability) con-

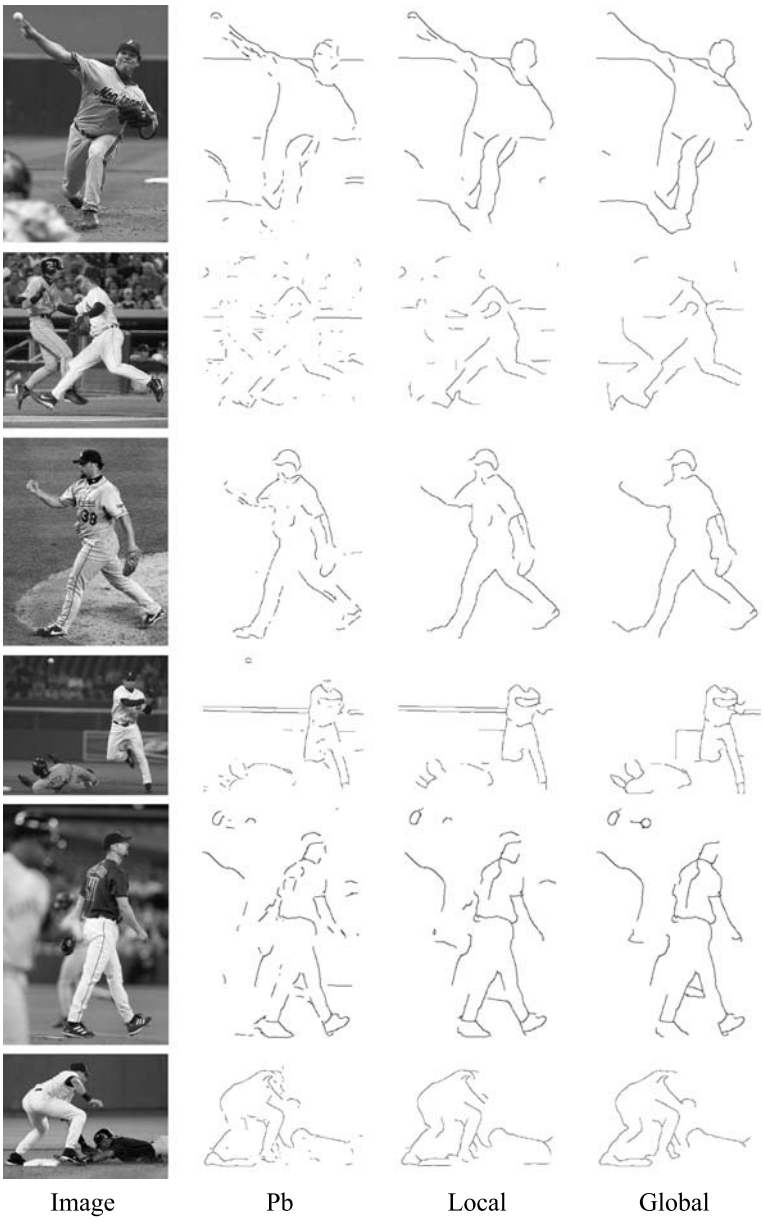



**Fig. 17** Examples from the baseball data set. The three columns of edge maps show the local boundary detector *Pb*, our local model, and our CRF global model respectively. For better visualization, the output of each algorithm has been thresholded at a level which yields 2000 boundary pixels

figurations use edges with high *Pb*, smooth continuations, and appropriate junction frequencies. In order to gauge the probability that a given edge $e$ corresponds to a true boundary, we would like to compute its marginal probability under this model,[6] that is

$$P(X_e = 1|I, \Theta) = \sum_{\substack{X \in \mathcal{X} \\ X_e = 1}} P(X|I, \Theta).$$

[6] Another possibility would be to find the most probable configuration $X^\star = \arg\max P(X|I, \Theta)$. We opt instead for use of the marginals in order to get a "soft" value that can be directly compared to the low-level *Pb* over a range of thresholds. Maintaining uncertainty in boundary detection can also prove useful during further high-level processing such as recognition.

Unlike the sequence modeling tasks where conditional random fields were first investigated our graph is not tree structured. It contains many triangles (among other loops) resulting in a high tree-width. Performing exact inference of marginals in such a model has extremely high computational complexity.

We opt to approximate the edge and vertex degree expectations using loopy belief propagation (Pearl 1988; Weiss 2000). Loopy belief propagation (LBP) has been applied to many vision problems (Weiss 1997; Freeman et al. 2000; Sun et al. 2002), typically on the 4-connected pixel lattice. The connectivity of the CDT graphs we study is much less uniform than a regular grid. However, in our experiments belief propagation appears to converge quickly ($< 10$ iterations) to a reasonable solution. Figure 13 shows the rate

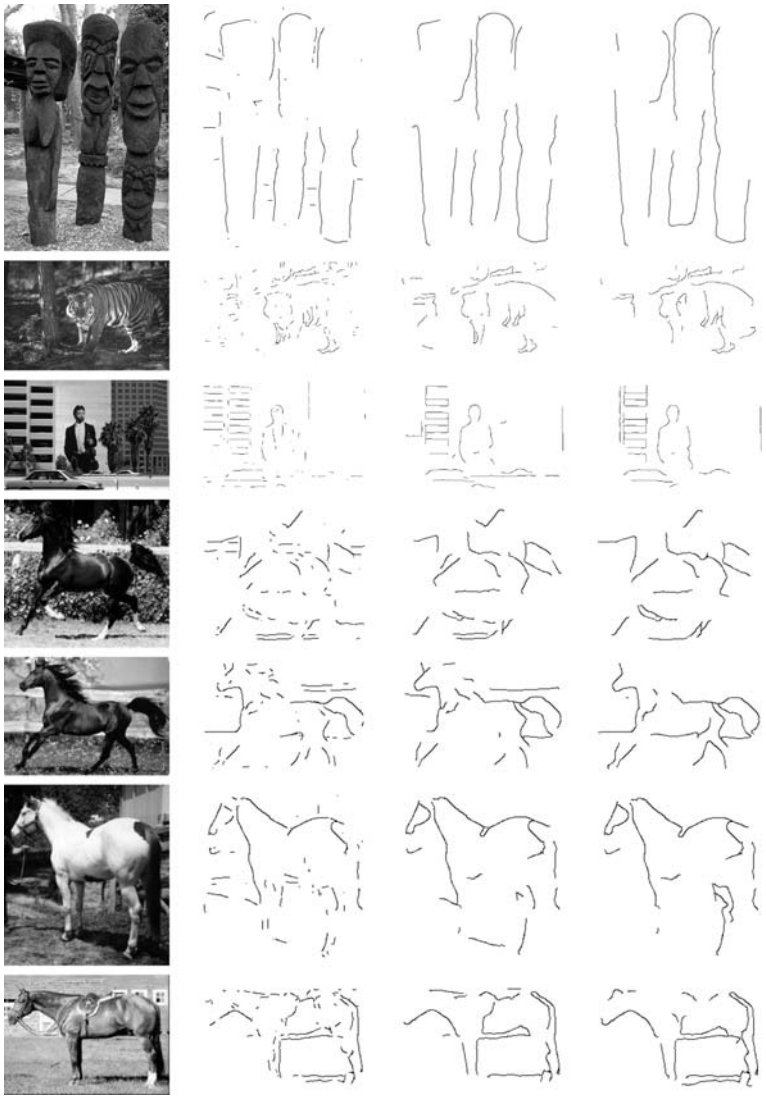



**Fig. 18** More examples from the BSDS and horse data sets. Our algorithm outputs smooth, extended contours on the foreground and suppresses noise in the background

of convergence of marginal estimates for the BSDS images. We suspect that there is relatively little frustration in our potentials and we have never observed any limit-cycle behavior (Murphy et al. 1999).

### 4.3 Learning Parameters

Our model has the collection of parameters $\Theta = \{\alpha, \beta, \gamma\}$. A single generic set of parameter values are learned from a collection of training images and used on all testing images. Since we exclude high degree junctions, we only need to set $\alpha_{i,j}$ for $i, j < 4$. This leaves a total of 18 free parameters which we fit by maximum likelihood.

Since our model lies in the exponential family, the likelihood is convex in the parameters. Taking a derivative of the log likelihood with respect to any parameter yields a difference of two expectations. For example, the derivative with respect to the continuation parameter $\gamma$ for a single training image/ground truth labeling, $(I, X)$ is:

$$\begin{aligned}
&\frac{\partial}{\partial\gamma}\log\left(\frac{e^{-\{\sum_e \phi(X_e|I,\Theta)+\sum_V \psi(X_V|I,\Theta)\}}}{Z(I,\Theta)}\right)\\
&\quad=\sum_V \frac{\partial}{\partial\gamma}\{\gamma \mathbf{1}_{\{\deg_g+\deg_c=2\}} f(\theta)\}\\
&\qquad-\frac{\partial}{\partial\gamma}\log Z(I,\Theta)\\
&\quad=\sum_V \mathbf{1}_{\{\deg_g+\deg_c=2\}} f(\theta)\\
&\qquad-\left\langle\sum_V \mathbf{1}_{\{\deg_g+\deg_c=2\}} f(\theta)\right\rangle_{P(X|I,\Theta)}
\end{aligned}$$

where we have used the fact that derivatives of the log partition function generate the moment sequence of $P(X|I,\Theta)$. The first term is the observed sum of $f(\theta)$ on degree 2 vertices while the second term is the expectation under the model given our current setting of the parameters (which can be estimated using LBP). When the model expectations match those observed in the training data, we have found the maximum likelihood setting of the parameters.

Optimization of the parameters can be achieved with any of several iterative numerical procedures. State-of-the-art codes based on quasi-Newton schemes typically require the value of a function in addition to gradients in order to perform line searches in promising directions. Although evaluating $P(X|I,\Theta)$ for a given $X$ is intractable, its value can also be estimated using the converged beliefs to compute the Bethe free energy in place of the true partition function. For the experiments described here, we found it was sufficient to use simple gradient descent with a momentum term. Parameters typically converged after a few hundred iterations.

Figure 14 shows the weights learned for different junction types. We find that the parameters learned from ground-truth boundary data match our intuition well. For example, the weight $\alpha_{1,0}$ is smaller than $\alpha_{2,0}$, indicating that line endings are less common than continuation and reflecting the prevalence of closed contours. For degree 2 vertices, we find $\alpha_{2,0} > \alpha_{1,1} > \alpha_{0,2}$, indicating that continuation along $G$-edges is preferable to invoking $C$-edges. Since high degree junctions are very uncommon in the training data, their corresponding $\alpha$ values tend to be uninformative. However, their impact on the test performance is limited for the same reason.

## 5 Results: Is Curvilinear Continuity Useful?

We have described two different algorithms, each outputting a new estimate of the boundary probability at each pixel: $Pb_L$, the local model on the CDT, and $Pb_G$, the global random field model. In order to evaluate these, we use three human-segmented datasets: a set of news photos of baseball players (Mori et al. 2004) split into 15 training and 15 testing images, images of horses (Borenstein and Ullman 2002) split into 172 training and 172 test images, and the Berkeley Segmentation Dataset (Martin et al. 2002) (BSDS300) which contains 200 training images and 100 test images of various natural scenes.

Performance on these datasets is quantified using the precision-recall framework as in Martin et al. (2004). Figure 15 shows the precision-recall curves for the three datasets. These quantitative comparisons clearly demonstrate that mid-level information *is* useful in a generic setting. Both models of curvilinear continuity outperform *Pb*. The global model, which is able to combine local evidence of continuity and global constraints such as closure, performs the best.

The improvement is most noticeable in the low-recall/high-precision range which corresponds to the case of boosting the most prominent boundaries and suppressing background noise. These boundaries are typically smooth; thus continuity helps suppress false positives in the background. This is evident in the examples shown in Fig. 17. We also find that our models push the asymptotic recall rate much closer to 100% (without loss in precision), reflecting their abilities to complete gradient-less gaps.

We observe that the benefit of continuity on the baseball player and horse dataset is much larger than that on the BSDS dataset. One reason may be that the baseball and horse datasets are harder (note the low precision rates for *Pb*) which makes the role of continuity more important.

In Fig. 16 we compare to two representative saliency algorithms: the pixel-based Markovian contour completion algorithm of Williams and Jacobs (1997), adapted to the setting of natural images as in Ren and Malik (2002); and the iterative updating saliency algorithm of Shashua and Ullman (1988). The Williams-Jacobs algorithm operates on the pixel-grid, with transition probabilities learned from data. The Shashua-Ullman algorithm is applied on the edgels in the CDT triangulation. We see that for the baseball and horse datasets, both algorithms improve edge detection performance only marginally, much less effective than our CRF-based algorithm. For the case of the BSDS, none of the algorithms stands out; the Williams-Jacobs algorithm does slightly better at the high-recall end.

If we look at these curves, there still exists a huge gap in performance between our models and the upper bound given by the ground-truth labels. It may be that the remaining gap will ultimately be bridged by incorporating high-level knowledge, i.e. detecting objects in the scene using the mid-level boundary map and then "cleaning up" the boundaries in a top-down fashion. In the case of general datasets like the BSDS, this will require recognizing thousands of object categories.

## 6 Conclusion

We have described two probabilistic models of curvilinear continuity which have a verifiably favorable impact on the problem of boundary detection. The local model, though quite simple, yields a significant performance gain. The global model, by making long-range inferences over local continuity constraints, is the most successful in utilizing mid-level information.

The key step in our approach to modeling continuity is moving from pixels to the piecewise linear approximations of contours. Motivated by the ecological statistics of

human-marked boundaries, this scale-invariant geometric representation of images shows promising potential for both mid-level and high-level visual processing. The shift from 100,000 pixels to 1000 Delaunay edges is also important as it yields huge gains in both statistical and computational efficiency.

We have shown that the outputs of our algorithms are quantifiably better than a low-level edge detector on a wide variety of natural images. We feel strongly that continued progress in mid-level vision rests on being able to make such quantitative comparisons between algorithms. Many recent exciting results in object recognition and stereo reconstruction could not have occurred without the firm grounding provided by quantitative evaluation. Work on perceptual organization should be held to the same high standards.

## References


Alvarez, L., Gousseau, Y., & Morel, J. (1999). Scales in natural images and a consequence on their bounded variation norm. In *Scale-space theories in computer vision*.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, *61*, 183–193.

August, J., Siddiqi, K., & Zucker, S. W. (1999). Contour fragment grouping and shared, simple occluders. *Computer Vision and Image Understanding*, *76*(2), 146–162.

Belongie, S., Malik, J., & Punicha, J. (2002). Shape matching and object recognition using shape contexts. *IEEE Transactions on PAMI*, *24*(4), 509–522.

Besag, J. E. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of Royal Statistical Society: Series B*, *36*(2), 192–236.

Borenstein, E., & Ullman, S. (2002). Class-specific, top-down segmentation. In *Proceedings of 7th European conference on computer vision* (Vol. 2, pp. 109–124).

Borgefors, G. (1988). Hierarchical chamfer matching: A parametric edge matching algorithm. *IEEE Transactions on PAMI*, *10*(6), 849–865.

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on PAMI*, *8*, 679–698.

Donoho, D., & Huo, X. (2002). Beamlets and multiscale image analysis. In *Lecture notes in computational science and engineering: Vol. 20. Multiscale and multiresolution methods* (pp. 149–196). New York: Springer.

Elder, J., & Goldberg, R. (2002). Ecological statistics of gestalt laws for the perceptual organization of contours. *Journal of Vision*, *2*(4), 324–353.

Elder, J. H., & Zucker, S. W. (1996). Computing contour closures. In *Proceedings of 4th European conference on computer vision* (Vol. I, pp. 399–412).

Felzenszwalb, P. (2001). Learning models for object recognition. In *Proceedings of IEEE conference on computer vision and pattern recognition*.

Freeman, W. T., Pasztor, E. C., & Carmichael, O. T. (2000). Learning low-level vision. *International Journal of Computer Vision*, *40*(1), 25–47.

Geisler, W. S., Perry, J. S., Super, B. J., & Gallogly, D. P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research*, *41*, 711–724.

Geman, S., & Geman, D. (1984). Stochastic relaxation, gibbs distribution, and the bayesian retoration of images. *IEEE Transactions on PAMI*, *6*, 721–741.

He, X., Zemel, R., & Carreira-Perpinan, M. (2004). Multiscale conditional random fields for image labelling. In *Proceedings of IEEE conference on computer vision and pattern recognition* (Vol. 2, pp. 695–702).

Heitger, F., & von der Heydt, R. (1993). A computational model of neural contour processing. In *Proceedings of 4th international conference on computer vision* (pp. 32–40). Berlin, Germany, May 1993.

Huang, J., & Mumford, D. (1999). Statistics of natural images and models. In *Proceedings of IEEE conference on computer vision and pattern recognition* (pp. 541–547).

Huttenlocher, D., & Wayner, P. (1992). Finding convex edge groupings in an image. *International Journal of Computer Vision*, *8*(1), 7–27.

Huttenlocher, D. P., Klanderman, G., & Rucklidge, W. (1993). Comparing images using the Hausdorff distance. *IEEE Transactions on PAMI*, *15*(9), 850–863.

Jermyn, I., & Ishikawa, H. (2001). Globally optimal regions and boundaries as minimum ratio weight cycles. *IEEE Transactions on PAMI*, *23*(10), 1075–1088.

Kanizsa, G. (1979). *Organization in vision: essays on gestalt perception*. New York: Praeger.

Kellman, P. J., & Shipley, T. F. (1991). A theory of visual interpolation in object perception. *Cognitive Psychology*, *23*, 141–221.

Konishi, S., Yuille, A. L., Coughlan, J. M., & Zhu, S. C. (1999). Fundamental bounds on edge detection: An information theoretic evaluation of different edge cues. In *Proceedings of IEEE conference on computer vision and pattern recognition* (pp. 573–579).

Kschischang, F. R., Frey, B. J., & Loeliger, H.-A. (2001). Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory*, *47*, 498–519.

Kumar, S., & Hebert, M. (2006). Discriminative random fields. *International Journal of Computer Vision*, *68*(2), 179–202.

Lafferty, J., McCallum, A., & Pereira, F. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of 18th international conference on machine learning*.

Lee, A., Mumford, D., & Huang, J. (2001). Occlusion models for natural images: A statistical study of a scale-invariant dead leaves model. *International Journal of Computer Vision*, *41*(1/2), 35–59.

Leuttgen, M., Karl, W. C., Willsky, A. S., & Tenney, R. R. (1993). Multiscale representation of markov random fields. *Special Issue of the IEEE Transactions on Signal Processing on Wavelet Transforms*, *41*(12), 3377–96.

Li, S. Z. (1995). *Markov random field modeling in computer vision*. New York: Springer.

Martin, D., Fowlkes, C., & Malik, J. (2002). Berkeley segmentation dataset. http://www.cs.berkeley.edu/projects/vision/bsds.

Martin, D., Fowlkes, C., Walker, L., & Malik, J. (2003). Local boundary detection in natural images: Matching human and machine performance. In *European conference on visual perception* (Perception, 32 supp., p. 55).

Martin, D., Fowlkes, C., & Malik, J. (2004). Learning to detect natural image boundaries using local brightness, color and texture cues. *IEEE Transactions on PAMI*, *26*(5), 530–549.

Mori, G., Ren, X., Efros, A., & Malik, J. (2004). Recovering human body configurations: Combining segmentation and recognition. In *Proceedings of IEEE conference on computer vision and pattern recognition* (Vol. 2, pp. 326–333).

Mumford, D. (1994). Elastica and computer vision. In C. Bajaj (Ed.), *Algebraic geometry and its applications* (pp. 491–506). New York: Springer.

Murphy, K., Weiss, Y., & Jordan, M. (1999). Loopy belief propagation for approximate inference: An empirical study. In *Proceedings of of UAI* (pp. 467–475).

Palmer, S. (1999). *Vision science: Photons to phenomenology*. Cambridge: MIT Press.

Parent, P., & Zucker, S. W. (1989). Trace inference, curvature consistency, and curve detection. *IEEE Transactions on PAMI*, *11*(8), 823–39.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Palo Alto: Morgan Kaufmann.

Ren, X. (2007). Learning and matching line aspects for articulated objects. In *Proceedings of IEEE conference on computer vision and pattern recognition*.

Ren, X., & Malik, J. (2002). A probabilistic multi-scale model for contour completion based on image statistics. In *Proceedings of 7th European conference on computer vision* (Vol. 1, pp. 312–327).

Ren, X., Berg, A., & Malik, J. (2005). Recovering human body configurations using pairwise constraints between parts. In *Proceedings of 10th international conference on computer vision* (Vol. 1, pp. 824–831).

Ruderman, D. L. (1997). Origins of scaling in natural images. *Vision Research*, *37*(23), 3385–98.

Ruderman, D. L., & Bialek, W. (1994). Statistics of natural images: Scaling in the woods. *Physics Review Letters*, *73*(6), 814–817.

Sharon, E., Brandt, A., & Basri, R. (2000). Completion energies and scale. *IEEE Transactions on PAMI*, *22*(10), 1117–1131.

Shashua, A., & Ullman, S. (1988). Structural saliency: the detection of globally salient structures using a locally connected network. In *Proceedings of 2nd international conference on computer vision* (pp. 321–327).

Shental, N., Zomet, A., Hertz, T., & Weiss, Y. (2003). Learning and inferring image segmentations with the gbp typical cut algorithm. In *Proceedings of 9th international conference on computer vision* (pp. 1243–1250).

Shewchuk, J. (1996). Triangle: Engineering a 2d quality mesh generator and delaunay triangulator. In *First workshop on applied computational geometry* (pp. 124–133).

Sun, J., Shum, H.-Y., & Zheng, N.-N. (2002). Stereo matching using belief propagation. In *Proceedings of 7th European conference on computer vision* (pp. 510–524).

Thornber, K. K., & Williams, L. R. (2001). Orientation, scale, and discontinuity as emergent properties of illusory contour shape. *Neural Computation*, *13*(8), 1683–1711.

Tu, Z. W., Chen, X. R., L Yuille, A., & Zhu, S. C. (2005). Image parsing: Unifying segmentation, detection, and recognition. *International Journal of Computer Vision*, *63*(2), 113–140.

von der Heydt, R., Peterhans, E., & Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science*, *224*, 1260–1262.

Weiss, Y. (1997). Interpreting images by propagating bayesian beliefs. *Advances in Neural Information Processing Systems*, *10*, 908–915.

Weiss, Y. (2000). Correctness of local probability propagation in graphical models with loops. *Neural Computation*, 1–41.

Wertheimer, M. (1938). Laws of organization in perceptual forms (partial translation). In W. B. Ellis (Ed.), *A sourcebook of gestalt psychology* (pp. 71–88). New York: Harcourt Brace.

Williams, L. R., & Jacobs, D. W. (1997). Stochastic completion fields: A neural model of illusory contour shape and salience. *Neural Computation*, *9*(4), 837–858.

Williams, L. R., & Thornber, K. K. (1999). A comparison of measures for detecting natural shapes in cluttered backgrounds. *International Journal of Computer Vision*, *34*(2/3), 81–96.

Wu, Q., & Yu, Y. (2003). Two-level image segmentation based on region and edge integration. In *Proceedings of DICTA* (pp. 957–966).

Yu, S., Gross, R., & Shi, J. (2002). Concurrent object segmentation and recognition with graph partitioning. *Advances in Neural Information Processing Systems*, *15*.

Zhu, S. C. (1999). Embedding gestalt laws in markov random fields. *IEEE Transactions on PAMI*, *21*(11), 1170–1187.

Zhu, S. C., Wu, Y. N., & Mumford, D. B. (1998). Frame: Filters, random field and maximum entropy: Towards a unified theory for texture modeling. *International Journal of Computer Vision*, *27*(2), 1–20.